Identification and estimation of ecologically relevant flow indices for non-perennial rivers, South Africa

Bernhard Stuart Schacht

A thesis submitted in fulfilment of the requirements for the degree of Magister Scientiae in Environmental and Water Science,
Department of Earth Science,
University of the Western Cape

Supervisor:
Professor Dominic Mazvimavi

March 2019
KEY WORDS

Ecologically relevant flow indices

Environmental flow

Environmental flow assessments

Aquatic ecosystems

Hydrological regionalisation

Water resources management

Ungauged catchments

Prediction of flow characteristics

Redundancy analysis

Multiple regression

Artificial neural networks

Cluster analysis
ABSTRACT

Identification and estimation of ecologically relevant flow indices for non-perennial rivers, South Africa

B.S Schacht

MSc Environmental and Water Science Thesis, Department of Earth Science, University of the Western Cape

March 2019

The importance of environmental flows and establishing the balance between hydrological response and ecological functioning of rivers has been globally recognised. Methods for determining environmental flows range from simplistic hydrological methods, to complex holistic approaches. Extensive research has been conducted to understand the dynamics of perennial rivers. However, non-perennial rivers remain among the least studied freshwater systems. Although these rivers only flow during certain periods of the year, these river systems are still of ecological significance. This study therefore aims to characterise the flow regime of non-perennial rivers in an ecologically meaningful way and to assess the spatial variation of these flow attributes. This study identified a subset of ecologically relevant flow indices that provided the opportunity to characterise the flow regime in an ecologically meaningful way, which was based on recommendations of previous studies. Redundancy analysis was used in the study to assess factors that account for the spatial variation of ecologically relevant flow indices of non-perennial rivers. The results identified that mean annual rainfall and the slope equalled or exceeded 90% of the time were the only catchment characteristics found to be significant at explaining the variance of flow characteristics. The final objective of the study explored methods that can be used to predict flow characteristics in ungauged catchments. Most catchments around the globe are ungauged or poorly gauged, which is a core issue in hydrology. Multiple regression and artificial neural networks were used to determine which approach is more suitable for prediction of selected flow characteristics of non-perennial rivers. The results identified that the use of multiple regression is more suitable due to higher model performance. However, the use of artificial neural networks was shown to be invaluable where relationships between flow and...
characteristics were non-linear. Cluster analysis was also used in the study to evaluate whether or not the clustering of catchments improves the prediction performance of flow characteristics. The results showed that the prediction performance was not improved. However, the transfer of information from gauged to ungauged catchments can be based on determining the cluster to which the ungauged catchment belongs to and then an average value of flow characteristics can be predicted for the ungauged catchment. The outcome of the study provided a subset of flow characteristics that can be used to characterise the flow regime in an ecologically meaningful way. The development of predictive models could also improve the management of these river systems and improve environmental flow management of ungauged catchments.
PLAGIARISM DECLARATION

I Bernhard Stuart Schacht declare the Identification and estimation of ecologically relevant flow indices for non-perennial rivers, South Africa is my own work. This work has not been submitted for any other degree or examination at any other university and that all the sources used or quoted have been indicated and acknowledged with complete references.

Full Name: Bernhard Stuart Schacht
Date: March 2019
Signed:
ACKNOWLEDGEMENTS

There are several individuals that I need to express my gratitude to for their contribution towards the completion of my MSc degree studies:

- Firstly my family and friends for their continued support, especially my mother and father, Mark and Helen Schacht, whose support throughout my life has not been unnoticed.
- Thanks needs to be given to various colleagues from the Institute for Water Studies, Dr Timothy Dube, Dr Jaco Nel, Dr Michael Grenfell, Clinton Andries and Dylan Seaton for their assistance during my research. It was most appreciated to have had the opportunity to discuss my work and issues with you all.
- Mr Paul Besener for the impact he has made on my life. I am forever grateful.
- I would like to thank Richard Nell for assisting in reviewing my thesis. Your assistance and input was greatly appreciated.
- Thanks need to be given to the Non-Perennial Rivers Project, The Southern African Science Service Centre for Climate Change and Adaptive Land Management (SASSCAL), and ACCESS for the financial assistance for my studies. Without these contributions I would not have had the opportunity to do my MSc and fortunately I was able to attend the 19th Waternet Symposium in Zambia.
- Lastly and most importantly, my supervisor and co-author Professor Dominic Mazvimavi. He has taught me more than I could have asked for and his continuous support and motivation was most appreciated. I was extremely fortunate to work alongside someone with such experience and knowledge that I could feed off from. This thesis would not have been possible without his assistance!
TABLE OF CONTENTS

KEY WORDS .................................................................................................................. i
ABSTRACT .................................................................................................................. ii
PLAGIARISM DECLARATION ....................................................................................... iv
ACKNOWLEDGEMENTS .............................................................................................. v
TABLE OF CONTENTS ............................................................................................... vi
LIST OF FREQUENTLY USED SYMBOLS ................................................................. xi
LIST OF FIGURES ..................................................................................................... xiii
LIST OF TABLES ......................................................................................................... xv
CHAPTER 1: INTRODUCTION .................................................................................. 1
  1.1 Background .......................................................................................................... 1
  1.2 Research Problem ................................................................................................ 4
  1.3 Research Question ............................................................................................... 5
  1.4 Research Aim ....................................................................................................... 5
  1.5 Research Objectives ............................................................................................ 5
  1.6 Rationale ............................................................................................................... 6
CHAPTER 2: LITERATURE REVIEW .......................................................................... 7
  2.1 Introduction .......................................................................................................... 7
  2.2 Perennial and non-perennial rivers ..................................................................... 7
  2.3 Global issues regarding non-perennial rivers .................................................... 9
  2.4 Importance of rivers to aquatic ecosystems ....................................................... 10
  2.5 Variable flow regime of non-perennial rivers and their ecological significance .. 11
  2.6 Human induced alterations of the flow regime and its impact on aquatic ecosystems .... 13
  2.7 Impact of climate change on river flow and aquatic ecosystems ....................... 14
2.8 Environmental flow ....................................................................................................... 16

2.8.1 Hydrological methodologies ................................................................................... 17

2.8.2 Hydraulic rating methodologies .............................................................................. 19

2.8.3 Habitat simulation methodologies .......................................................................... 19

2.8.4 Holistic Methods ..................................................................................................... 20

2.8.4.1 Building Block Methodology (BBM) .................................................................. 21

2.8.4.2 Downstream Response to Instream Flow Transformations (DRIFT) ................. 21

2.9 Hydrological regionalisation .......................................................................................... 22

2.9.1 Hydrological model-dependent methods .................................................................. 23

2.9.2 Hydrological model-independent method ................................................................. 25

2.10 Catchment characteristics used for hydrological regionalisation ............................. 26

2.10.1 Rainfall ......................................................................................................................... 27

2.10.2 Evaporation ................................................................................................................ 27

2.10.3 Slope and topography ............................................................................................... 28

2.10.4 Geology ....................................................................................................................... 29

2.10.5 Soil ................................................................................................................................ 30

2.10.6 Drainage density ........................................................................................................ 31

2.10.7 Land cover .................................................................................................................. 31

2.11 Summary ......................................................................................................................... 32

CHAPTER 3: METHODOLOGY ....................................................................................... 34

3.1 Introduction ................................................................................................................... 34

3.2 Selection of gauging stations ......................................................................................... 34

3.3 Delineation of catchments ............................................................................................. 35

3.4 Catchment descriptors ................................................................................................... 37

3.4.1 Elevation and slope ................................................................................................... 38

3.4.2 Rainfall ......................................................................................................................... 38
CHAPTER 5: ORDINATION ........................................................................................................ 62
5.1 Introduction ................................................................................................................... 62
5.2 Methodology ................................................................................................................. 62
5.3 Results .......................................................................................................................... 66
5.3.1 Relationships between catchment characteristics .................................................. 66
5.3.2 Relationship between flow characteristics ............................................................. 67
5.3.3 Relationship between flow and catchment characteristics ...................................... 68
5.4 Discussion and conclusion ............................................................................................. 73
CHAPTER 6: PREDICTION OF FLOW CHARACTERISTICS ........................................... 75
6.1 Introduction ................................................................................................................... 75
6.2 Methodology ................................................................................................................. 75
6.2.1 Multiple linear regression ....................................................................................... 75
6.2.2 Artificial neural networks ......................................................................................... 76
6.2.3 Model performance evaluation ............................................................................... 78
6.3 Correlation between flow and catchment characteristics ........................................... 78
6.4 Mean annual runoff ....................................................................................................... 79
6.5 Flow duration curve ....................................................................................................... 82
6.6 Baseflow Index .............................................................................................................. 85
6.7 Number of zero flow days ............................................................................................. 87
6.8 3-day minima and 3-day maxima of daily flows ......................................................... 88
6.9 Discussion and conclusion ............................................................................................. 90
CHAPTER 7: IDENTIFICATION OF CATCHMENTS WITH SIMILAR HYDROLOGICAL PROCESSES ........................................................................................................ 92
7.1 Introduction ................................................................................................................... 92
7.2 Methodology ................................................................................................................ 92
7.2.1 Selection of catchment descriptors for the derivation of clusters .............................. 93
7.2.2 Determination of the number of clusters and validation ........................................... 94

7.3 Results and Discussion ................................................................................................. 96

7.3.1 Classification of clusters using catchment characteristics .................................. 96

7.3.2 Classification of clusters using flow characteristics ............................................ 96

7.3.3 Number of clusters .................................................................................................. 99

7.3.4 Catchment characteristics of clusters ..................................................................... 103

7.3.5 Flow characteristics of clusters ................................................................................ 105

7.3.6 Prediction of flow characteristics based on clusters ............................................. 106

7.3.6.1 Prediction of mean annual runoff .................................................................... 107

7.3.6.2 Prediction of baseflow index (BFI) ................................................................. 109

7.3.6.3 Prediction of the number of zero flow days ..................................................... 109

7.4 Conclusion .................................................................................................................. 110

CHAPTER 8: CONCLUSIONS AND RECOMMENDATIONS ........................................ 112

8.1 Introduction ................................................................................................................. 112

8.2 Factors accounting for the spatial variation of selected ecologically relevant flow characteristics ........................................................................................................ 112

8.3 Prediction of flow characteristics using multiple regression and artificial neural networks ................................................................................................................. 113

8.4 Identification of catchments with similar hydrological responses .............................. 113

8.5 Prediction of flow characteristics based on hydrologically similar catchments ......... 114

8.6 Recommendations ...................................................................................................... 114

REFERENCES ................................................................................................................. 116
**LIST OF FREQUENTLY USED SYMBOLS**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q</td>
<td>Mean annual runoff</td>
<td>(mm/year)</td>
</tr>
<tr>
<td>CV</td>
<td>Coefficient of variation of annual flows</td>
<td>%</td>
</tr>
<tr>
<td>q_{90}</td>
<td>Dimensionless daily flow with a 90 % exceedence</td>
<td></td>
</tr>
<tr>
<td>q_{75}</td>
<td>Dimensionless daily flow with a 75 % exceedence</td>
<td></td>
</tr>
<tr>
<td>q_{25}</td>
<td>Dimensionless daily flow with a 25 % exceedence</td>
<td></td>
</tr>
<tr>
<td>q_{10}</td>
<td>Dimensionless daily flow with a 10 % exceedence</td>
<td></td>
</tr>
<tr>
<td>IC</td>
<td>Concavity Index</td>
<td>%</td>
</tr>
<tr>
<td>CVB</td>
<td>Hydrological Index</td>
<td>ratio</td>
</tr>
<tr>
<td>3-day min</td>
<td>3-day minimum of daily means of discharge</td>
<td>mm/year</td>
</tr>
<tr>
<td>3-day max</td>
<td>3-day maximum of daily means of discharge</td>
<td>mm/year</td>
</tr>
<tr>
<td>ZFD</td>
<td>Number of zero flow days</td>
<td>days</td>
</tr>
<tr>
<td>BFI</td>
<td>Baseflow Index</td>
<td>ratio</td>
</tr>
<tr>
<td>A</td>
<td>Catchment area</td>
<td>km²</td>
</tr>
<tr>
<td>MAP</td>
<td>Mean annual rainfall</td>
<td>mm/year</td>
</tr>
<tr>
<td>ET</td>
<td>Mean annual pan evaporation</td>
<td>mm/year</td>
</tr>
<tr>
<td>Dd</td>
<td>Drainage density</td>
<td>km/km²</td>
</tr>
<tr>
<td>RL</td>
<td>River length</td>
<td>km</td>
</tr>
<tr>
<td>S_{20}</td>
<td>Slope exceeded or equalled 20 % of the time</td>
<td>degrees</td>
</tr>
<tr>
<td>S_{50}</td>
<td>Slope exceeded or equalled 50 % of the time</td>
<td>degrees</td>
</tr>
<tr>
<td>S_{90}</td>
<td>Slope exceeded or equalled 90 % of the time</td>
<td>degrees</td>
</tr>
<tr>
<td>S_{ave}</td>
<td>Average catchment slope</td>
<td>degrees</td>
</tr>
<tr>
<td>E_{min}</td>
<td>Minimum catchment elevation</td>
<td>m</td>
</tr>
<tr>
<td>E_{range}</td>
<td>Range of catchment elevation</td>
<td>m</td>
</tr>
<tr>
<td>E_{max}</td>
<td>Maximum catchment elevation</td>
<td>m</td>
</tr>
<tr>
<td>GL_{TM}</td>
<td>Proportion of catchment underlain by Table Mountain Group</td>
<td>%</td>
</tr>
<tr>
<td>GL_{KD}</td>
<td>Proportion of catchment underlain by Karoo Dolerite</td>
<td>%</td>
</tr>
<tr>
<td>GL_{CG}</td>
<td>Proportion of catchment underlain by Cape Granite</td>
<td>%</td>
</tr>
<tr>
<td>GL_{BF}</td>
<td>Proportion of catchment underlain by Beaufort Group</td>
<td>%</td>
</tr>
<tr>
<td>LC_{S}</td>
<td>Proportion of catchment under shrublands</td>
<td>%</td>
</tr>
<tr>
<td>LC_{G}</td>
<td>Proportion of catchment under grasslands</td>
<td>%</td>
</tr>
<tr>
<td>LC_{BG}</td>
<td>Proportion of catchment under bare ground</td>
<td>%</td>
</tr>
<tr>
<td>LC&lt;sub&gt;P&lt;/sub&gt;</td>
<td>Proportion of catchment under plantations</td>
<td>%</td>
</tr>
<tr>
<td>LC&lt;sub&gt;CL&lt;/sub&gt;</td>
<td>Proportion of catchment under cultivated lands</td>
<td>%</td>
</tr>
<tr>
<td>LC&lt;sub&gt;T&lt;/sub&gt;</td>
<td>Proportion of catchment under thicket</td>
<td>%</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure 4.1: Location of selected non-perennial gauging stations ............................................. 46
Figure 4.2: Study catchments delineated using SRTM 1 arc second 30 m DEM ....................... 47
Figure 4.3: Variation of elevation within the study area derived from SRTM 1 arc second 30m DEM ........................................................................................................................... 48
Figure 4.4: Cumulative frequency distribution of slopes on 5 selected catchments ................. 49
Figure 4.5: Median slopes of the study area derived from STRM 1 arc second 30 m DEM .. 49
Figure 4.6: Mean annual rainfall within the study area derived from WR2012 (http://waterresourceswr2012.co.za/resource-centre/; Accessed June 2017) ....................... 50
Figure 4.7: Reference evapotranspiration within the study area derived from WR2012 (http://waterresourceswr2012.co.za/resource-centre/; Accessed June 2017) .......... 51
Figure 4.8: Drainage density of the study area derived from NGI 1:50 000 river map ......... 52
Figure 4.9: Lithology of the catchments in the study area derived from the 1:1 000 000 geological map of South Africa ........................................................................................................................... 53
Figure 4.10: Land cover within the study area derived from the South African Land Cover map 2014 ............................................................................................................................ 54
Figure 4.11: Soil textural classes in the study area derived from WR2012 1: 250 000 soil map of South Africa.................................................................................................................... 54
Figure 4.12: Variation of mean annual runoff within the study area ....................................... 56
Figure 4.13: Variation of $q_{10}$, $q_{25}$, $q_{75}$ and $q_{90}$ within the study area ............................... 57
Figure 4.14: Variation of baseflow index within the study area ............................................ 58
Figure 4.15: Variation of coefficient of variation of annual flows within the study area ...... 58
Figure 4.16: Variation of the number of zero flow days within the study area ....................... 59
Figure 4.17: Variation of 3-day minima and maxima of daily means of discharge within the study area ............................................................................................................................ 60

Figure 5.1: Variation of catchment characteristics among selected catchments .................... 66
Figure 5.2: Variation of flow characteristics among selected catchments ............................. 68
Figure 5.3: Relationship between flow and catchment characteristics among selected catchments .......................................................................................................................... 70
Figure 6.1: Simple illustration of a multi-layer perceptron with 5 input variables, 3 units in the hidden layer and 1 output variable ................................................................. 77
Figure 6.2: Comparison between the performance of multiple regression and an MLP 7-9-1 artificial neural network in predicting mean annual runoff .................................................. 81
Figure 6.3: Comparison of the model performance for predicting flow percentiles using an exponential model and neural networks based on the RSR and $R^2$ ........................................ 83
Figure 6.4: Comparison of observed flow duration curves with those predicted using an exponential model and neural network for catchments H5H004, G1H036, K6H001 and Q9H017 .......................................................................................................................... 84
Figure 6.5: Comparison between the performance of multiple regression and a Linear 3-1 neural network for predicting baseflow index ................................................................. 86
Figure 6.6: Comparison between the performance of multiple regression and a MLP 5-6-1 artificial neural network for predicting number of zero flow days ...................................... 88

Figure 7.1: Variation of $R_g$ statistics with increases in the number of clusters .................. 99
Figure 7.2: Comparison of the average flow duration curve for cluster membership when the number of clusters is 4, 5, 6 and 7 ................................................................. 100
Figure 7.3: Andrew’s curves for the clusters based on the catchment characteristics that significantly explain the variance of flow characteristics .................................................. 102
Figure 7.4: Cluster membership of the selected catchments in the study area .................. 103
Figure 7.5: Range of variation of catchment characteristics for the derived clusters ........ 104
Figure 7.6: Range of variation of flow characteristics for the derived clusters ............... 106
Figure 7.7: Relationship between mean annual runoff and mean annual rainfall for the derived clusters ................................................................................................................. 107
Figure 7.8: Relationship between mean annual runoff and proportion of grasslands for the derived clusters ............................................................................................................. 108
LIST OF TABLES

Table 3.1: Catchment descriptors selected in the study ........................................................ 37
Table 3.2: Selected ecologically relevant flow indices of non-perennial rivers for the study 42

Table 4.1: Distribution of catchment area for selected river gauging stations in the study area ........................................................................................................................................... 47
Table 4.2: Selected ecologically relevant flow indices of non-perennial rivers for the study 55

Table 5.1: Flow and catchment characteristics selected for redundancy analysis .............. 65
Table 5.2: Proportion of variance of flow characteristics explained by catchment characteristics in the derived ordination axes ..................................................................................................................... 71
Table 5.3: Proportion of variance of flow characteristics explained by catchment characteristics ..................................................................................................................... 72

Table 6.1: Correlation between flow and catchment characteristics .................................... 79
Table 6.2: Neural networks used for predicting mean annual runoff from catchment characteristics ..................................................................................................................... 81
Table 6.3: Neural networks used for predicting number of zero flow days (ZFD) from catchment attributes ..................................................................................................................... 87
Table 6.4: Predictive equations for predicting 3-day minima and maxima of daily means of discharge ............................................................................................................................. 89
Table 6.5: Predicting 3-day minima and maxima of daily means of discharge using neural networks ............................................................................................................................. 89

Table 7.1: Cluster membership for 2 to 10 clusters based on cluster analysis of catchment characteristics ..................................................................................................................... 97
Table 7.2: Cluster membership for 2 to 10 clusters based on cluster analysis of flow characteristics ............................................................................................................................. 98
Table 7.3: Mean values for catchment characteristics of the derived clusters .................... 104
Table 7.4: Average values of flow characteristics for the derived clusters ....................... 105
This page has been left blank intentionally.
CHAPTER 1: INTRODUCTION

1.1 Background

Water is essential for supporting life on Earth. Without access to water, humans will cease to exist (White et al., 2010). The availability of freshwater has been recognized as a global issue in the 21\textsuperscript{st} century (Döll et al., 2003; Srinivasan, 2011; Bigas et al., 2012). Many of the world’s arid and semi-arid regions are facing challenges due to the increase in water demand that decreases the available water resource (Grafton et al., 2011). The decrease in available water is a result of increases in both population and economic growth, which subsequently increase the water demand. Without examining alternative measures for ensuring water availability, livelihood options of particularly, the poor in water-scarce and low-income countries will be limited (Grafton et al., 2011). Ensuring the availability of freshwater for both human and ecological water requirements is critical in the management and assessment of water resources.

River systems are utilised by humans as a source of water for irrigation, food, energy, waste disposal and industry, for both economic and societal needs (Desai, 2012). However, the fears of floods and droughts, coupled with the need to ensure food security and energy have resulted in the implementation of various measures to increase the available water for human consumption. Abstractions, dam construction, inter-catchment transfers and hydropower schemes potentially overcome the threats of water, food and energy shortages, and these interventions consequently influence the hydrological cycle (Desai, 2012; Rolls et al., 2012).

Climate change and anthropogenic activities including land-use change, hydraulic abstraction and artificial water abstraction have been identified as the main factors that have significantly influenced the hydrological cycle (Nan et al., 2011; Zhang et al., 2012; Wang & Qin, 2017). Previous studies have observed changes in both global average surface temperature and rainfall as a result of climate change and global warming (Zhang et al., 2012; Ren et al., 2013; Thoeun, 2015; Dahlman, 2017). An increase in the global average surface temperature and a decrease in rainfall consequently results in a reduction in river flows (Rolls et al., 2012; Zhang et al., 2012; Zhang et al., 2012, Schneider et al., 2013; Duong et al., 2017). However, the effects of climate change may benefit some regions with an increase in flow, while other
regions will experience a reduction in flows, depending on the variation between surface
temperature and rainfall distribution (Martins et al., 2016). The extent to which these
anthropogenic activities affect river flows varies in terms of their specific processes, such as
flow regulation, abstraction or land use change, which may increase or decrease hydrological
responses (Zhang et al., 2012; Martins et al., 2016). Natural and anthropogenic activities
affect aquatic and terrestrial ecosystems through altering the natural flow regime. Understanding
the effects of both climate change and anthropogenic activities on river flows
is critical in terms of managing the water resource, especially in arid and semi-arid regions
where the availability of water is limited.

The importance of rivers in influencing the functioning and operation of terrestrial and
aquatic ecosystems has been globally recognised (Monk et al., 2007; Kennard et al., 2010;
Belmar et al., 2011; Roll et al., 2012; Sponseller et al., 2013; Worrall et al., 2013; Stagl &
Hattermann, 2016). Riverine systems are habitats, vectors for connectivity, and agents for
physical distribution and geomorphic change that enable the functioning of aquatic
ecosystems (Sponseller et al., 2013). Many aquatic species develop life history adaptations
and strategies that are synchronized in relation to the flow regime (Bejarano et al., 2010;
Kennard et al., 2010). For example, the natural timing of low and high flow events provides
environmental cues for the initiation of reproductive cycles, such as spawning, egg hatching,
migration through the system, or even access to floodplains (Bejarano et al., 2010).
Alterations of the natural flow regime due to climatic or anthropogenic processes cause
disruptions to these ecosystems. Hydrological and ecological processes within a river system
are sensitive to flow alterations that affect the natural balance and may lead to the
degradation of the river (Rossouw, 2011). Policies and laws have therefore been implemented
to reduce the adverse impacts of flow modification as part of achieving environmental, social
and economic sustainability.

There is an increasing awareness around the globe regarding the flow regime that needs to be
allocated and reserved to maintain and support the functioning and operation of riverine
ecosystems (Matthews & Richter, 2007; Mazvimavi et al., 2007; Matthews et al 2014;
Acreman et al., 2014). This is regarded as the environmental flow. Quantifying the
environmental flow of rivers involves determining the quantity and quality of both societal
and ecological needs in a sustainable approach (Desai, 2012). Environmental awareness has
resulted in policies and laws that have been implemented for the protection and conservation

http://etd.uwc.ac.za/
of riverine ecosystems. The Earth Summit in Rio de Janeiro in 1992 promoted conservation of ecosystems as public goods, whereby water rights were granted to ecosystems and not just the needs of mankind (Acreman & Dunbar, 2004). The 2002 Johannesburg World Summit on Sustainable Development introduced the protection of the environment as a key pillar of sustainable development, where policies and laws were to be implemented to give priorities to ecosystems once the basic human needs were met (Acreman & Dunbar, 2004).

In the context of South African legislation, the National Water Act (NWA) of 1998 and National Environmental Management: Biodiversity Act 10 of 2004 were enacted to meet the increasing human demand for water while reducing the adverse impacts on rivers and aquatic ecosystems, due to the realisation of the need to allocate water to riverine systems (Desai, 2012). The NWA of 1998 stipulates that the utilisation of water resources should be conducted in a sustainable manner in order to maintain river integrity, and an acceptable level of ecological functioning. The implementation of these laws emphasises the awareness and need to protect these systems and signifies the social, economic and environmental importance of river systems.

As river flows are depleted or altered through natural or anthropogenic processes, ecological degradation ensues and society loses the benefits that are provided by ecosystems such as commercial and subsistent fisheries, water purification, flood storage and recreation (Matthews & Richter, 2007). Studies are therefore being conducted to understand the links between the flow regime and the functioning of ecosystems and to determine the flow regime that is required to support the health of aquatic ecosystems (Matthews & Richter, 2007; Rolls et al., 2012; Vis et al., 2015). The environmental flow is therefore an essential component of integrated water resource management, which recognises the importance of these flows in terms of providing a resource for mankind as well as for the world’s biodiversity (Acreman & Dunbar, 2004). The environmental flow, however, varies between different river systems, and determining the characteristics of individual catchments is essential for evaluating the ecological water requirements of river systems.

Perennial and non-perennial rivers differ on the basis of their spatial and temporal flow regimes (Rossouw et al., 2011; Avenant et al., 2014; Cid et al., 2017). Non-perennial rivers are distinguished from perennial rivers by higher flow variability and periods of zero flow (Rossouw, 2011; Avenant et al., 2014). More than 55% of Africa’s land surface is characterised as being arid or semi-arid and surface flows do not occur throughout the year.
(Avenant et al., 2014). Previous studies have emphasized that some of the most variable flow regimes occur within arid environments in South Africa and Australia (Hughes, 2005; Poff et al., 2006; Avenant et al., 2014).

The unpredictability and complexity of non-perennial rivers are a result of the variability of climatic characteristics, including precipitation and evaporation, which vary spatially and temporally (Avenant et al., 2014). Climate change is currently increasing the complexity and unpredictability of the natural flow regime due to changes in rainfall and evaporation rates (Zhang et al., 2012; Ren et al., 2013; Avenant et al., 2014; Thoeun, 2015; Dahlman, 2017), and the pressure to provide water for both human consumption, as well as environmental needs will therefore increase (Avenant et al., 2014).

The importance of non-perennial rivers has generally been disregarded in the past, as these river systems are generally perceived to have low economic and ecological significance (Skoulikidis et al., 2017; Datry et al., 2017). The formulation of policies and legislation therefore needs to take into account non-perennial rivers, as these systems are often associated with regions of high water use and environmental needs are often neglected (Rossouw, 2011).

1.2 Research Problem

Methods currently available for determining environmental flows for South African rivers were developed using data mainly from perennial rivers. In South Africa, two-thirds of the rivers are considered non-perennial (Rossouw et al., 2005). Despite the critical importance of non-perennial rivers, these systems remain poorly understood because current research tends to focus primarily on perennial rivers, this is typical being a worldwide trend (Rossouw et al., 2005; Rossouw, 2011; Seaman et al., 2016; Skoulikidis et al., 2017). The flow regime of non-perennial rivers is highly variable both spatially and temporally compared to perennial rivers, which therefore requires focused attention in terms of research and management (Rossouw et al., 2005). It is therefore crucial to develop methods explicitly designed for non-perennial rivers to maintain the hydrological and ecological balance and conserve their biodiversity (Rossouw, 2011).

The availability of river flow, also referred to as streamflow, data is an important requirement for water resource development projects, such as dam construction and identifying the
adverse effects of such development on riverine systems (Laaha & Blöschl, 2004; Blöschl et al., 2013; Razavi & Coulibaly, 2013; Viglione et al., 2013; Li et al., 2015). Previous studies have identified that most catchments around the globe are ungauged or inadequately gauged (Schmocker-Fackel et al., 2007; Blöschl et al., 2013; Razavi & Coulibaly, 2013; Xie et al., 2014). Ungauged catchments have been defined as those having inadequate hydrological records, in terms of quantity and/or quality (Sivapalan et al., 2003; Blöschl et al., 2013; Razavi & Coulibaly, 2013; Razavi, 2014; Li et al., 2015). A core issue in hydrology is therefore to estimate flow characteristics within ungauged regions to provide the opportunity to effectively manage river systems (Parajka et al., 2005; Jin et al., 2008; He et al., 2011; Razavi & Coulibaly, 2013; Abimbola et al., 2017; Singh, 2018).

1.3 Research Question

How can we characterise the flow regime of non-perennial rivers in an ecologically meaningful way?

1.4 Research Aim

To characterise flow regimes of non-perennial rivers in an ecologically meaningful way and assess the spatial variation of these flow attributes.

1.5 Research Objectives

The objectives of this research project are:

1. To identify ecologically relevant river flow indices for non-perennial rivers.
2. To determine factors that account for the spatial variations of ecologically relevant river flow indices of non-perennial rivers.
3. To explore methods for predicting ecologically relevant flow indices of ungauged non-perennial rivers.
1.6 Rationale

Water is an essential resource not only for the benefit of humankind, but is also critical for the functioning of ecosystems (Monk et al., 2007; Naiman et al., 2008; Kennard et al., 2010; White et al., 2010; Belmar et al., 2011). Environmental flows are therefore an important concept in terms of protecting the functioning of aquatic ecosystems. The study attempts to improve the understanding of non-perennial rivers and understanding the links between hydrological and ecological balance, which can be used to effectively manage these systems for the protection and conservation of aquatic ecosystems.

Most catchments, also referred to as basins or watersheds, around the globe are ungauged or inadequately gauged (Razavi & Coulibaly, 2013; Xie et al., 2014). Ungauged or inadequately gauged catchments lack river gauging stations, due to site inaccessibility, financial and/or maintenance issues (Visessri & McIntyre, 2016; Blöschl, 2016; Atieh et al., 2017; Singh, 2018). The study assessed the use of hydrological regionalisation approaches to extrapolate information from gauged stations for those that are ungauged. This provides the opportunity to gain insight and knowledge within catchment areas that are ungauged.
CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

The chapter evaluates the importance of non-perennial rivers, in terms of ecological requirements and anthropogenic activities, as well as distinguishing physiographic and climatic characteristics between perennial and non-perennial rivers. The impacts of anthropogenic and climate change on the flow regime is also evaluated. This chapter also highlights the importance of conserving and protecting the natural flow regime of river systems. Finally, methods available for hydrological regionalisation are examined for determining a suitable method to transfer information from gauged to ungauged catchments.

2.2 Perennial and non-perennial rivers

Perennial and non-perennial rivers differ in terms of spatial and temporal variation of flows (Rossouw et al., 2011; Avenant et al., 2014; Cid et al., 2017). Non-perennial rivers are characterised by highly variable seasonal and inter-annual flow regimes and low levels of flow predictability, whereas perennial rivers exhibit more stable conditions which are more predictable (Cid et al., 2017). Perennial rivers have continuous flow throughout the year. (McDonough et al., 2011; Avenant et al., 2014; Berhanu et al., 2015). The continuous flow within these rivers may be due to the presence of baseflow from groundwater discharge or all year rainfall (Berhanu et al., 2015). To the contrary, non-perennial rivers are characterised by higher flow variability, low flows during the recession periods immediately after rainfall events and periods of zero flow (Hughes, 2005; Watson & Dallas, 2012; Avenant et al., 2014). Non-perennial rivers are considered to be the most dynamic, complex and diverse of freshwater systems (Larned et al., 2010; McDonough et al., 2011; Skoulikidis et al., 2017).

Stewart et al (2012) suggested that non-perennial rivers are a global phenomenon compared to perennial rivers and representative of most of the world’s river systems. Non-perennial rivers account for more than 50% of global river networks (McDonough et al., 2011; Arthington & Balcombe, 2011; Stewart et al., 2012; Skoulikidis et al., 2017). Non-perennial river systems are located in arid, semi-arid and dry sub-humid regions of the earth (Seely et al., 2003; Rossouw et al., 2011; Skoulikidis et al., 2017). The spatial and temporal extent of
non-perennial rivers is likely to further increase due to the combined effects of climate change and anthropogenic activities including water abstraction and land-use change (Larned et al., 2010; McDonough et al., 2011; Stewart et al., 2012).

More than 55% of the land in southern Africa has been characterised as arid and semi-arid (Avenant et al., 2014). The occurrence of perennial, seasonal and non-perennial river systems in southern Africa signifies the extreme diversity of climatic and associated flow characteristics (Hughes, 2005). Studies have shown that some of the most variable river flow regimes occur in the arid areas within southern Africa and Australia (Hughes, 2005; Poff et al., 2006; Avenant et al., 2014). In South Africa, non-perennial rivers are located in the arid and semi-arid regions within the western and central regions of the country (Watson & Dallas, 2012; Seaman et al., 2016). Within these regions, rainfall is highly sporadic with some areas receiving less than 500 mm annually of rainfall, and high evaporation rates further intensify the complexity and unpredictability of non-perennial rivers (Watson & Dallas, 2012; Avenant et al., 2014; Seaman et al., 2016). The functioning and characteristic features of non-perennial rivers may be distinguishable from perennial rivers based on climatic factors, and thus requires focused attention in terms of research and management (Rossouw et al., 2011).

Non-perennial rivers are therefore mainly characterised by aridity and high rainfall variability (Rossouw et al., 2005; Hughes, 2005; Avenant et al., 2014). Regions can be classified using the aridity index, which refers to the degree of dryness at a given location on the earth’s surface, such as humid, sub-humid, arid, semi-arid and hyper-arid, based on the precipitation and evaporation. The problem becomes more severe in arid and hyper-arid regions where the aridity is high, due to low annual rainfall and high evaporation rates, for example the evaporation rates are six times greater than the mean annual rainfall in Namibian catchments (Rossouw et al., 2005). The consequence of high rates of evaporation and limited water flow can cause salt build up, which has a detrimental effect upon biodiversity, which lack resilience to altered habitat conditions (Rossouw et al., 2005). Furthermore, climatic characteristics are also a reason as to why dams are not efficient within these dryland regions, as these regions are often characterised with poor vegetation cover and associated highly erodible soils. These factors subsequently result in the build-up of sediment, which reduces the storage of surface water bodies, such as dams and wetlands, thus further impacting upon water availability (Rossouw et al., 2005). Although there is limited water within non-
perennial river systems, these rivers are important sources of water for both ecological requirements and anthropogenic activities, and thus require appropriate assessment and management of the water resource. With the climate predicted to become increasingly warmer and drier over large regions of southern Africa, combined with an increased inter and intra-annual variation in rainfall, the pressure to provide water for both human and ecological requirements will increase (Avenant et al., 2014).

2.3 Global issues regarding non-perennial rivers

In the past, the importance of non-perennial rivers for social, economic and ecological needs have been underestimated, mainly as these systems are regarded as having low economic and ecological significance (McDonough et al., 2011; Skoulikidis et al., 2017; Datry et al., 2017). Although such rivers are common worldwide, a method for establishing environmental flows is still lacking and the knowledge of the ecological functioning within these river systems is generally poor (Seaman et al., 2016).

Non-perennial rivers remain among the least studied freshwater systems (Arthington & Balcome, 2011; Skoulikidis et al., 2017), and these river systems are utilised as waste disposal sites and sources for sand and gravel (Skoulikidis et al., 2017). Commercial and subsistent farmers in dryland regions recognise the importance of these river systems and understand the significance of these systems for their survival. However policy and management measures do not emphasize the importance of these river systems, which therefore restricts the effective management of river systems for both human well-being and the functioning of aquatic ecosystems. Without an effective framework for managing the resource, these river systems may be adversely affected by human-induced flow modifications.

The hydrological and ecological balance of non-perennial rivers is sensitive to alterations of the natural flow regime that often cause the degradation of the river system. River degradation refers to the alteration of the river channel cross-section and width, such as the scouring of river beds, and erosion of river banks (Mugade & Sapkale, 2015). This is generally influenced by climatic and human disturbances, river discharges, sediment load, and morphological characteristics of the channel (Mugade & Sapkale, 2015). Over extraction of the water resource has led to an increase in the spatial and temporal extent of non-
perennial rivers, as well as the exposure to various stressors, which threaten biodiversity, ecological integrity and the functioning of ecosystems (Skoulikidis et al., 2017). The alteration of the natural flow regime can reduce or increase low flows during dry seasons, and can temporarily result in perennial rivers changing to non-perennial rivers. It is therefore important to assess the flow parameters within the river system and determine the flow regime needed to maintain ecological integrity. The formulation of policy therefore needs to take into account non-perennial rivers, as these systems are often located within regions of high water demand and environmental needs are often neglected (Rossouw, 2011).

2.4 Importance of rivers to aquatic ecosystems

The importance of the flow regime and its inherent variability has been widely recognised in influencing the functioning of riverine ecosystems (Monk et al., 2007; Naiman et al., 2008; Kennard et al., 2010; Belmar et al., 2011; Roll et al., 2012; Worrall et al., 2013; Stagl & Hattermann, 2016; Ge et al., 2018). The flow regime has been described as a “master variable” that is essential for shaping the physical structure of the river system, through erosional and depositional processes, that in turn rearranges ecological communities and governs the long-term structural and physical characteristics of the habitat (Larned et al., 2010; McDonough et al., 2011; Cunjak et al., 2012; Sponseller et al., 2013; Arthington et al., 2013; Stagl & Hatterman, 2016). Varying levels of spatial and temporal variability are key factors in determining patterns and processes within aquatic ecosystems (Cid et al., 2017), and subsequently the adaptations of aquatic biota are dictated by the inter- and intra-annual variability of the flow regime (Kennard et al., 2010; Seaman et al., 2013).

Temporal variation of flow dictates the structure and functioning of riverine ecosystems, as well as adaptations and life histories developed by riparian and aquatic species (Rossouw et al., 2005; Naiman et al., 2008; Cid et al., 2017). Variation of river flows control essential habitat conditions within river channels, floodplains, and even stream-influenced groundwater zones, which are critical for ecological processes (Stagl & Hatterman, 2016). The frequency, magnitude, duration, timing and spatial extent of flow events are common drivers of ecological integrity in riverine ecosystems (Rolls et al., 2012). For example, high and low flows are common drivers for maintaining the biological and physical characteristics within a river system.
Perennial and non-perennial rivers differ based on spatial and temporal variability of flow as discussed previously. Hence, processes and patterns of aquatic ecosystems within these systems differ and the aquatic ecosystems living within these two different river systems develop the necessary adaptations to survive and recover from the flow conditions that they are exposed to (Rossouw et al., 2011; Seaman et al., 2013; Avenant et al., 2014; Cid et al., 2017).

2.5 Variable flow regime of non-perennial rivers and their ecological significance

Non-perennial rivers are unpredictable and complex due to the highly variable flow regime, with no-flow during one day and the occurrence of flooding event on another (Seaman et al., 2013). As already mentioned, the full range of variability of flows, including high and low-flows are important for aquatic ecosystems. The onset and duration of wet and dry periods within non-perennial river catchments are highly variable, which has significant influences on ecological patterns and processes (Cid et al., 2017). Understanding the influence of different flow conditions on aquatic ecosystems is therefore important for the management of non-perennial river systems.

Although periods of no-flow conditions may be detrimental in perennial rivers, such conditions are natural within non-perennial rivers that aquatic biota experience in arid and semi-arid regions (Rossouw et al., 2005). Periods of no-flow are important for non-perennial rivers for maintaining the health of the river system (Rossouw et al., 2005). Examples of low flow events, including duration, timing and seasonality, and their relative importance for aquatic ecosystems is briefly discussed.

The duration of low flow events have been expressed as a critical driver of riverine ecosystems, as native species are adapted to survive within these stressful flow conditions (Yarnell et al., 2015). The connectivity of water allows the mixing of gene pools, movement of organisms and the transport of sediments through the system (Seaman et al., 2016). The duration of low flows dictate the extent and quality of physical habitat, therefore influencing the composition and distribution of species inhabiting the region (Yarnell et al., 2015). As low flows restrict the connectivity of flow within the river system and reduces available habitats, native species use regions of refugia or exhibit adaptive life strategies to persist through stressful periods of low flow (Yarnell et al., 2015; Cid et al., 2017). The effects of
low-flow events on aquatic biota tends to increase with increasing duration, where low-flow events of longer duration have a greater impact on the ecosystems than those events of shorter duration (Rolls et al., 2012).

The timing and seasonality of flows has also been recognized as important in influencing aquatic ecosystems, such as recruitment, migration and germination of aquatic species (Rolls et al., 2012), having substantial effects on the recruitment and migration of aquatic biota, especially if these low-flow events occur during periods of migration that may restrict the dispersal of aquatic organisms (Rolls et al., 2012). Studies have been conducted to establish this theory, which investigated the effects on different trout species in the USA. The studies revealed that low-flow events during winter had minimal effects on the aquatic habitat and movement of trout species; however, low-flow events that occurred during summer caused a reduction in the growth of specific trout species (Rolls et al., 2012).

In non-perennial rivers the occurrence and characteristics of high-flow events are highly variable (Stromberg et al., 2007). Flood events in these river systems may exhibit varying characteristics, including events of high magnitude and short duration, or events of higher magnitude and longer duration (Stromberg et al., 2007). The occurrence of different high-flow events produces a range of effects on riverine systems, including sediment flux, channel morphology and ecological functioning. Floods that are categorized by large magnitude and long duration have a substantial impact on the river system, such as shaping the structure of vegetation, and providing water availability for short-term hydrological processes (e.g. groundwater recharge) and long-term geomorphic processes (e.g. channel incision and the deposition of sediments) (Stromberg et al., 2007). The rapid onset and end of small floods cause minor disturbances within river systems, and also provide a water source for a temporary period (Stromberg et al., 2007).

For aquatic biota, high flow conditions may alter flow conditions in the river for a short-term period, and allow aquatic organisms a time frame of respite from low-flow conditions (Matthews & Richter, 2007). These flow conditions may relieve the stresses that organisms are exposed to, such as high water temperatures and low availability of dissolved oxygen due to low-flows, as well as remove wastes and provide additional organic material within the system (Matthews & Richter, 2007). These events typically improve the ability of organisms to access upstream and downstream regions within the river channel, which may be critical during periods of growth and development of aquatic species (Matthews & Richter, 2007).
The access to these regions provides additional food resources, with less competition, and thus increases the growth rate of aquatic species (Matthews & Richter, 2007).

The full range of flow variability, including both high and low-flows, is therefore important for the functioning of aquatic ecosystems. However, the balance between hydrological and ecological processes is disturbed through anthropogenic activity and climate change which can significantly influence the natural flow regime.

2.6 Human induced alterations of the flow regime and its impact on aquatic ecosystems

Climate change and anthropogenic activities have been regarded as the main factors influencing the variability of the hydrological cycle in recent years (Zhang et al., 2012; Wang et al., 2012; Yang et al., 2017; Duong et al.; 2017; Ge et al., 2018). Flow regulation, surface and groundwater abstraction, land-use change and urbanisation are some of the common anthropogenic activities impacting the hydrological cycle (Rolls et al., 2012). Alterations of the natural flow regime due to anthropogenic activities have significantly impacted the movement of water and sediment within river systems, and therefore have adversely impacted upon the aquatic community inhabiting the region (Magoba, 2014). The effects that are associated with flow modification include changes in river discharge, decreased suspended sediment, flooding, channel and floodplain narrowing, and a decrease in the diversity of the riparian habitat (Magoba, 2014). Changes in the flow regime due to anthropogenic activity are expected to persist with an increase in human population growth and the demand for municipal, agricultural and industrial use (Wang et al., 2012).

Water is a necessity for life and humans introduce measures for protection against floods and droughts (Naiman et al., 2008). As a consequence, many of the world’s rivers have been modified to satisfy human demands for water, while reducing the adverse effects of floods and droughts (Naiman et al., 2008). The issue that arises, for example, is that the regulation of floods reduces downstream flow, resulting in many rivers exhibiting little of their natural flow variability and ecological needs are therefore heavily constrained (Naiman et al., 2008). As discussed earlier, although flooding events are often seen as destructive, such occurrences have important roles within aquatic ecosystems.
Alterations of the natural flow regime, due to anthropogenic activities introduce various threats to freshwater ecosystems and biodiversity. Consequently, these alterations have been associated with ecosystem instability and have thus resulted in a decline in the biomass and abundance of aquatic species (Cunjak et al., 2012; Leigh et al., 2012). Examples of these impacts include changes in land-use and land cover, such as processes of afforestation and deforestation that have significantly altered the runoff generation within catchments, through the disturbance of the balance between rainfall and evaporation (Wang et al., 2012). Processes of flow regulation and water abstraction have significantly altered ecologically relevant attributes of flow, including the magnitude, frequency, duration, timing and seasonality, and the rate of change of flow events (Leigh et al., 2012). The alteration of ecologically relevant components of flow has subsequently resulted in adverse effects on the functioning and operation of aquatic ecosystems (Leigh et al., 2012).

It is important to determine the main anthropogenic activities affecting the flow regime to enable the sustainable development of the water resource and implement measures for the conservation and protection of biodiversity (Wang et al., 2012). Climate change, along with anthropogenic activity are factors that increase the complexity of the balance between hydrology and ecological requirements (Cunjak et al., 2012). Quantifying the link between hydrology and ecology is a major challenge, especially in the context of climate change, due to inadequate knowledge and the lack of ecological and physical data over spatial and temporal dimensions (Cunjak et al., 2012).

2.7 Impact of climate change on river flow and aquatic ecosystems

Climate change has been considered as the one of the major challenges facing humankind in the 21st century and studies have shown extensive adverse impacts on both human and natural systems around the globe (Duong et al., 2017). Future changes in climate have been widely researched (Zhang et al., 2012; Yang et al., 2017; Duong et al., 2017) and many regions are expected to experience significant changes of flow regimes. Some scientists have reported that the effects of climate change on river flows may be more important in altering flow regimes than anthropogenic activity, such as hydraulic structures and abstraction (Döll & Zhang, 2010; Yang et al., 2017; Ge et al., 2018).
According to observed data, the last decade has been recorded as the warmest in the last 100 years, where the global surface temperature displayed an increase of 0.85°C over the period of 1880 - 2012. As a result, the increase in global surface temperature coupled with the increase in spatial and temporal variation of rainfall, is likely to lead to changes in the hydrological cycle (Duong et al., 2017). This will lead to changes of flow regimes, for example, river flow in tropical regions are expected to increase due to a higher frequency and magnitude of rainfall events (Duong et al., 2017). On the contrary, drought conditions during dry periods may lead to pressure of water availability for human and ecological requirements (Duong et al., 2017). The results presented by Duong et al (2017) show that river flows are likely to increase in the wet season and decline in the dry season. More severe flooding conditions are expected in the wet season, while water shortages and a decline in water availability are expected in the dry season. Oguntunde & Abiodun (2013) and Aich et al (2014) assessed the influence of climate change on river flow in the Niger River catchment. The results of these studies showed that temperatures are expected to increase and rainfall is expected to decrease in the headwater regions. The prediction of climate change therefore shows an increase in temperature as well as spatial and temporal variability of rainfall, which will subsequently result in a reduction of flow (Rolls et al., 2012; Zhang et al., 2012; Duong et al., 2017). A decrease in river flow decreases the available water supply for both human and ecological requirements, and further increases the frequency and duration of low-flow events (Rolls et al., 2012).

The influence of climate change on river flow has been widely researched; however, less attention has been applied to the potential impacts on ecological functioning in rivers. In terms of aquatic ecosystems, changes of flows may result in the conversion of perennial rivers to non-perennial rivers and vice versa, which may/will have a substantial influence on the patterns and process of the aquatic biota and riparian vegetation (Döll & Schmied, 2012). A change in river dynamics from perennial to non-perennial rivers will result in the decline and loss of aquatic biota; as some species do not possess the necessary adaptations to persist and provide resilience against harsh flow conditions, and will therefore be replaced by those species that are more tolerant to specific flow events, such as droughts and floods (Cid et al., 2017). As a result, previously established balances between hydrological and ecological processes may be disturbed (Wrzesiński & Sobkowiak, 2018).
2.8 Environmental flow

Understanding the relationship between the flow regime and ecological processes has occupied scientific research for an extensive period of time (Rolls et al., 2012; Vis et al., 2015). Maintaining environmental flows in river systems is a global concept, as flow altering is regarded as the primary contributor to ecological degradation, and consequently leads to a loss of biodiversity (USGS, 2013). Environmental flow refers to the flow required to achieve desired ecological objectives, which is also termed as the instream flow, environmental allocation or ecological flow requirements (Acreman & Dunbar, 2004; Hughes et al., 2014; Acreman, 2016). Initially environmental flows were based on the concept of minimum flow, which states that as long as river flows are kept above a critical threshold, the river ecosystem will be conserved (Acreman & Dunbar, 2004; Belmar et al 2011). However, studies have identified that the flow regime and its inherent range of variability is important when assessing the functioning of ecosystems (Acreman & Dunbar, 2004; Worrall et al., 2013; Vis et al., 2015; Stagl & Hatterman, 2016). Quantifying the full range of flow variability within a river system may therefore be used to link the hydrological characteristics and ecological processes, thus providing an opportunity to effectively manage these river systems in a sustainable manner (Vis et al., 2015).

Environmental flows research has focused on assessing the importance of different flow components and evaluating the influence on the functioning of riverine ecosystems (Worrall et al., 2013). The flow regime can be characterised by a number of ecologically relevant flow indices, which have been grouped into five categories namely; magnitude, frequency, duration, timing and rate of change of flow events (Richter et al., 1996; Olden & Poff, 2003; Kennard et al., 2010). The importance of environmental flows has resulted in the development of numerous environmental flow assessments, which range from fairly simplistic approaches to complex holistic approaches.

From a global perspective, the development of environmental flow methodologies began as early as the 1950’s in western USA, with noticeable progress during the 1970’s as a result of the implementation of environmental legislation (King et al., 2008; Jain, 2012). Other countries such as England, Australia, New Zealand and South Africa began to make swift progress in this field during the 1980’s (King et al., 2008). Currently, other countries that have not made progress within the context of environmental flows, either did not recognise
the importance of managing and protecting riverine ecosystems or did not make these assessments a priority (King et al., 2008).

The importance of environmental flows led to the development of several environmental flow assessment (EFA) methods, ranging from simple to complex holistic approaches. Environmental flow assessments aim to protect the integrity of the riverine systems due to increased pressures from both anthropogenic activity and climate change (Avenant, 2010). Environmental flow assessments are based on determining the amount of water that can be harvested within a river system, whilst preventing adverse ecological impacts (Avenant, 2010). These methods aim to establish the condition of riverine ecosystems under specific flow conditions, to identify responses to specific flow and to evaluate where flow-ecology relationships can be modelled (Rolls et al., 2012). Environmental flow methodologies can be categorised into four groups namely; hydrological, hydraulic rating, habitat simulation and holistic methods (Rossouw et al., 2005; King et al., 2008; Karimi et al., 2012; Jain, 2012; Hao et al., 2016; Tegos et al., 2017).

2.8.1 Hydrological methodologies

The simplest environmental flow methodologies are the hydrological methods, which are based on the analysis of available flow data (Rossouw et al., 2005; King et al., 2008; Karimi et al., 2012; Linnansaari et al., 2013). Hydrological methods are also referred to as fixed percentage or standard-setting methodologies, whereby a percentage of flow, either based on monthly or annual flows, is regarded as the environmental flow to maintain the ecological integrity at some specific level (Rossouw et al., 2005; King et al., 2008; Tegos et al., 2017). Examples of hydrological methods include the Tennant or Montana Method and Flow Duration Curve Analysis (FDCA) (King et al., 2008; Avenant, 2010).

Initial EFAs, such as the Tennant Method were designed to protect specific fish species that had recreational or commercial significance (King et al., 2008; Avenant, 2010; Kawde et al., 2016). Historically, fish were regarded as the central theme of initial EFAs (Avenant, 2010). The Tennant method determined the ecological condition of a river based on the relationship between the percentage of flow and the quality of the instream fish habitat (Avenant, 2010; Kawde et al., 2016). The Tenant method therefore used a percentage of flow to propose
environmental flow recommendations (Karimi et al., 2012; Jain, 2012; Kawde et al., 2016). The FDCA is also a common method for prescribing environmental flows, and specific flows, such as $q_{90}$ and $q_{95}$, which are frequently used as indicators to represent the minimum flow required for ecological functioning (Karimi et al., 2012; Jain, 2012). The $q_{90}$ and $q_{95}$ represent flow percentiles of the flow regime. For example $q_{90}$ represents the flow that is equalled or exceeded 90% of the time.

More recent methods, such as the Range of Variability Approach (RVA), have been developed to characterise the full range of variability of the flow regime (Rossouw et al., 2005; King et al., 2008; Karimi et al., 2012; Kawde et al., 2016). The RVA is based on 32 flow statistics of the flow regime, which represents different aspects of the flow regime, calculated from long-term time series of flow (Rossouw et al., 2005; King et al., 2008; Karimi et al., 2012; Jain, 2012; Kawde et al., 2016). The flow indices reflect the magnitude of both high and low flows, the timing and frequency of various flows, and their duration (Rossouw et al., 2005). Each hydrological index is calculated based on the annual averages for each year in the time series of flow, which provides an opportunity to assess the inter-annual variability of flow (Rossouw et al., 2005). An acceptable range of variation is set for individual hydrological indices to maintain ecological functioning, for example $\pm 1$ standard deviation from the mean or between the 25th and 75th percentile (Rossouw et al., 2005). This approach provides the opportunity to set preliminary flows to maintain the integrity of riverine ecosystems, which can be monitored and revised (Rossouw et al., 2005).

Hydrological methods are typically used for environmental flow assessments, as these approaches are inexpensive and rapid, requiring only historical flow records (King et al., 2008; Linnansaari et al., 2013; Saniruzzaman et al., 2015). However, limitations associated with such methodologies are from an ecological perspective, therefore this approach is considered simplistic and does not cover the nature and variability of the flow regime (King et al., 2008; Saniruzzaman et al., 2015). The method has often been criticised for the lack of ecological validity, arising from the uncertainty between flow-ecology relationships (Linnansaari et al., 2013), and as a result, hydraulic rating methods were developed (Jain, 2012).
2.8.2 Hydraulic rating methodologies

Hydraulic rating methodologies use hydrological, hydraulic and ecological data, as well as incorporation of professional experience (Jain, 2012; Kawde et al., 2016). The assumption of this approach is that riverine habitat availability is closely related with various hydraulic properties of the river such as wetted perimeter, maximum depth, river width and flow velocity (Rossouw et al., 2005; King et al., 2008; Jain, 2012; Hao et al., 2016). These methods measure changes in hydraulic properties and develop relationships between habitat availability and discharge to determine environmental flow recommendations (King et al., 2008; Tegos et al., 2017). The values are plotted against discharge, and a break point is identified where there is a change in slope on the curve. The assumption of this approach is that when flow falls below the break point, the quality of the habitat and the functioning of an ecosystem would severely deteriorate (Rossouw et al., 2005).

The Wetted Perimeter Method is an example of a hydraulic rating method, which develops relationships between changes in wetted perimeter at sections of a river with discharge, as the basis for environmental flow recommendations (King et al., 2008; Jain, 2012; Kawde et al., 2016). The analysis of the habitat-discharge response curve provides information whereby habitat quality significantly declines with specific flows (Jain, 2012). The hydraulic rating methodologies have been considered an improvement from the hydrological methods, as the method incorporates various ecological characteristics such as instream physical habitat of biota (King et al., 2008). However, these methods rely on a simplistic assumption that a single or group of hydraulic variables adequately represent the requirements of riverine ecosystems and that environmental flow recommendations are generally based on target species (King et al., 2008; Saniruzzaman et al., 2015).

2.8.3 Habitat simulation methodologies

The habitat simulation methodologies attempt to characterise the environmental flow recommendation on the basis of biotic responses to flow at the level of instream habitats (King et al., 2008; Tegos et al., 2017). Such methodologies model the changes in physical habitat with discharge using data based on one or more hydraulic variables such as flow.
depth, velocity and substratum composition (Rossouw et al., 2005; King et al., 2008). This type of data is collected at various cross-sections across a river channel. Simulated available habitat conditions are then linked to ecological information on the basis of suitable living conditions for a target species, including life histories, assemblages and activities (Rossouw et al., 2005; King et al., 2008; Tegos et al., 2017). The output of the method produces habitat-discharge curves, which indicate the optimum discharge as the proposed environmental flow (King et al., 2008). The advantage of these methodologies is that being computer-based, a large amount of hydrological, hydraulic and biological data can be efficiently processed (King et al., 2008; Saniruzzaman et al., 2015). However, the concern surrounding this approach is that environmental flows are based on target species, which may not be representative of the entire ecological community (King et al., 2008; Saniruzzaman et al., 2015).

2.8.4 Holistic Methods

The development of holistic methods in the 1990’s emphasised the need to progress towards protecting the entire riverine ecosystem, rather than focusing on individual species (Rossouw et al., 2005; Avenant, 2010). Holistic methodologies differ from other EFAs on the basis that they consider the flow needs of the entire riverine ecosystem (Rossouw et al., 2005; Avenant et al., 2012). Holistic approaches integrate various components of the river, such as hydrology, fluvial geomorphology, water quality, water temperature, riparian vegetation, fish and aquatic invertebrates, and understanding the influence of each component for effectively determining the environmental flow of a river (King et al., 2008; Avenant, 2010; Avenant et al., 2012; Hao et al 2016). Examples of holistic methods that have been developed include the Building Block Methodology (BBM), Downstream Response to Instream Flow Transformation (DRIFT) and Flow Stress-Response Method (Avenant, 2010; Avenant et al., 2012).
2.8.4.1 Building Block Methodology (BBM)

The basic approach of BBM is that the functioning of riverine ecosystems is reliant on basic elements (building blocks) of the flow regime, including both low and high flows (Rossouw et al., 2005; Gopal, 2013). BBM recognises that some flows within the flow regime are more important than others for riverine ecosystems, and such flows can be identified and described in terms of their magnitude, duration, frequency and timing (King et al., 2008; Gopal, 2013). The components, which are referred to as ‘building blocks’ of the flow usually fall into the following categories; dry-season base flows, wet-season low flows, wet-season floods, dry-season freshes and dry-season subsurface flows. Each block represents an important part of the flow regime for ecological and geomorphological processes; however, the importance of each block will differ between different rivers (Gopal, 2013). The desired minimum flow that is required for each block is described, and the modified flow regime is obtained by combining the blocks in a way that mimics the natural flow conditions of the river (Rossouw et al., 2005; Gopal, 2013).

2.8.4.2 Downstream Response to Instream Flow Transformations (DRIFT)

DRIFT is a holistic approach, which was developed in South Africa and used for water resource development projects on the Palmiet and Breede River, as well as the Lesotho Highland Water Project (Rossouw et al., 2005; King et al., 2008; Gopal, 2013; Kawde et al., 2016). DRIFT is essentially a data management tool, which allows the input of data and knowledge to assess the potential impacts of flow modification on social, economic and ecological aspects (King et al., 2003; King et al., 2008; Gopal, 2013). The underlying principle of DRIFT is that different components of the flow regime give rise to different ecological responses (King et al., 2003). Therefore a change in a specific component of the flow regime would affect the riverine ecosystem differently than a change in another (King et al., 2003).

DRIFT focuses on the identification of a specific flow associated with a particular set of ecological functioning and hydrological and hydraulic characteristics (King et al., 2008). Specialists within each discipline provide the potential impacts of reducing flow within the
river, and relate these impacts in terms of social, economic and ecological effects (King et al., 2008). DRIFT allows water managers to assess social and ecological consequences of flow modification using different scenarios and therefore identifying water developments that have minimal consequences based on specified flow recommendations (King et al., 2003; Rossouw et al., 2005; Gopal, 2013).

Although there are a wide variety of environmental flow assessment methods, the current issue is that the methods that have been developed and applied are mainly in the context of perennial rivers with adequate data (Karimi et al., 2012). In almost all developing countries, there is a lack of knowledge regarding environmental flows, which reduces the confidence to allocate environmental flows for riverine ecosystems (Karimi et al., 2012). In the context of non-perennial rivers, a method for establishing environmental flows is still lacking, and no formal method has therefore emerged (Seaman et al., 2016). A study by Seaman et al (2016) suggested that DRIFT could possibly be used to determine the environmental flows for non-perennial rivers; however, there were some challenges and modifications that needed to be addressed first, such as determining the reference condition, connectivity and understanding the presence of isolated pools.

2.9 Hydrological regionalisation

The availability of continuous and reliable flow data is an important requirement for water resource development projects, such as dam construction and assessing their impacts upon riverine ecosystems (Laaha & Blöschl, 2004; Blöschl et al., 2013; Razavi & Coulibaly, 2013; Viggio et al., 2013; Li et al., 2015; Javeed & Apoorva, 2015; Visessri & McIntyre, 2016; Blöschl, 2016; Singh, 2018). Flow data can easily be obtained within gauged catchments and if necessary the flow record can be extended by using a locally calibrated rainfall-runoff model (Visessri & McIntyre, 2016). Most catchments around the globe are ungauged or inadequately gauged (Schmocker-Fackel et al., 2007; Blöschl et al., 2013; Razavi & Coulibaly, 2013; Xie et al., 2014). Ungauged or inadequately gauged catchments lack flow gauging networks, as often these regions are inaccessible (He et al., 2011; Blöschl et al., 2013; Razavi, 2014; Visessri & McIntyre, 2016; Blöschl, 2016; Atieh et al., 2017; Singh, 2018). Previous studies have established that ungauged catchments are those having inadequate records of hydrological data, in terms of both quantity and quality (Sivapalan et
A core issue of hydrology is therefore to define flow characteristics in areas that are ungauged and is one of the major challenges in hydrology (Parajka et al., 2005; Jin et al., 2008; He et al., 2011; Razavi & Coulibaly, 2013; Abimbola et al., 2017; Singh, 2018).

Hydrological regionalisation is defined as any process of transferring information from one region of known hydrological parameters to other regions, usually of unknown hydrological parameters (Parajka et al., 2005; Merz et al., 2006; Javeed & Apoorva, 2015; Elesbon et al., 2015; Abimbola et al., 2017). This transfer can be based on data series, or specific parameters of flow, including maximum, minimum, average, variance or even equations related to these statistical parameters (Elesbon et al., 2015; Abimbola et al., 2017). The classification process primarily focuses on understanding the link between catchment structure, climate and functioning (Begou et al., 2015). The concept of this approach is that catchments with similar climate, geology, topography, vegetation and soils would usually share similar hydrological processes (Oueslati et al., 2010). However, the complexity of predicting flow increases with diversity in climate and sparse meteorological data observations (Spence et al., 2013). Hydrological regionalisation techniques can be classified into either hydrological model-dependent or hydrological model-independent groups (Razavi, 2014).

2.9.1 Hydrological model-dependent methods

These approaches transfer parameters of the rainfall-runoff model from gauged to ungauged catchments to simulate a continuous river flow for a catchment of interest (Razavi & Coulibaly, 2013). The hydrological model-dependent methods include arithmetic mean, spatial proximity, physical similarity, scaling relationships and regression-based methods (Razavi & Coulibaly, 2013). Each approach has a unique method for deriving quantitative relationships between model parameters and physical attributes of the catchment. Approaches that transfer model parameters from gauged to ungauged catchments assume that the catchments are the same and will have the exact response to the same input (Parajka et al., 2003; Razavi & Coulibaly, 2013). The limitation with using the hydrological model-dependent approaches is that this requires physical attributes for parameter estimation, as well as relevant expertise and associated knowledge.
In the arithmetic mean approach, the model parameters from the rainfall-runoff model are averaged, either based on the surrounding catchments (local) or all catchments (global approach) (Razavi & Coulibaly, 2013; Razavi, 2014). Spatial proximity approaches transfer the parameters based on a specific distance between catchments, either based on the catchment centroids or outlets (Parajka et al., 2005; Razavi & Coulibaly, 2013; Abimbola et al., 2017). This approach assumes that catchments close to each other will have similar responses (Oudin et al., 2008; He et al., 2011; Abimbola et al., 2017). This approach uses an interpolation approach that is linked to the geographic location, with kriging being the most commonly used interpolation method. The debate of whether or not homogenous catchments are linked to spatial proximity has been a major topic of interest. A study by Shu and Burn (2003) suggested that catchments within a specific proximity to each other are not necessarily homogenous in terms of their hydrological response (Parajka et al., 2003).

The third method is physical similarity. The concept of this approach is the transfer of hydrological parameters of interest from gauged to ungauged catchments according to their similarity of catchment characteristics (Parajka et al., 2005; Oudin et al., 2008; Razavi & Coulibaly, 2013; Li & Zhang, 2016). Donor catchments are selected based on similarities of climatic and physical characteristics of the catchment between the ungauged or catchment of interest (Parajka et al., 2005; Abimbola et al., 2017). An example of this approach was conducted by Oudin et al. (2008) and Samuel et al. (2011). These studies initially grouped catchments based on physical and non-hydrological attributes. This approach is generally conducted using a multivariate statistical analysis approach to group these catchments. Once these catchments are grouped, the rainfall-runoff parameters for the gauged catchments are computed and the parameters that are in the same group are used to create a regional rainfall-runoff model. The river flow for ungauged catchments is then generated based on the physical similarities between the gauged catchments (Razavi, 2014). Regression based approaches, including linear and non-linear regression methods are used to estimate river flow in ungauged catchments. This approach is based on developing relationships between the model parameters and catchment attributes (Razavi, 2014; Li & Zhang, 2016). The relationship between parameter values and catchment characteristics are established and calibrated in gauged catchments, and the parameters are therefore estimated in ungauged catchments based on the relationship with gauged catchments (Li & Zhang, 2016). Due to the variety of catchment characteristics available, this approach is commonly used for regionalisation studies (Merz et al., 2006). However, in multiple regression approaches, one
may encounter multicollinearity between the selected catchment characteristics, which can result in the regression being highly unstable and unreliable (Merz et al., 2006). Generally studies limit the number of catchment characteristics used, due to this constraint, and select catchment characteristics based on scientific knowledge and the influence of such characteristics on runoff generation (Merz et al., 2006).

2.9.2 Hydrological model-independent method

These approaches develop and employ equations representing parameters of the rainfall-runoff model, for example precipitation and temperature are used as the input and river flow as the output (Razavi & Coulibaly, 2013; Razavi, 2014). One of the advantages of using hydrological model-independent methods is reduced data requirements compared to hydrological model-dependent methods, as well as the simplicity of the approach and the limited knowledge and expertise required (Razavi & Coulibaly, 2013; Razavi, 2014). These techniques do not simulate the actual rainfall-runoff process, and therefore this approach is not affected by uncertainties relating to the physical process of the model (Razavi, 2014). However, the approach is still affected by other sources of uncertainty, including the estimation method used and its parameterization (Razavi & Coulibaly, 2013; Razavi, 2014). A study by Goswami et al (2007) applied methods from both groups for regionalisation of river flow, where the conceptual hydrological model and the artificial neural network (ANN) were used to simulate river flow in 12 French catchments that were considered ungauged. The results of this study suggested that the ANN approach had a better performance than the conceptual hydrological model (Razavi & Coulibaly, 2013; Razavi, 2014). It can be noted based on literature that the application of data-driven methods is often advantageous, especially when the availability of hydrological data is inadequate for regionalisation using hydrological models (Razavi & Coulibaly, 2013; Razavi, 2014). Hydrological model-independent methods, which are mainly data driven, can be categorized into three groups; regression based analysis, time series models and scaling methods.

Regression-based analysis, which includes linear and non-linear regressions, develops linear and non-linear relationship between river flow and catchment characteristics (Razavi & Coulibaly, 2013). Chiang et al (2002) used a multiple regression analysis approach to determine relationships between river flow and catchment attributes. The results from the
MRA equations are then used to generate a synthetic hydrograph under the time series model. The next category is time series models, e.g. AutoRegressive (AR) models. The last category is scaling methods. A study by Buttle and Eimers (2009) identified that estimation of annual maximum and mean daily flow can be well predicted using scaling relationships. The scaling relationship between river flow parameters and catchment attributes has the potential to extrapolate data from gauged to ungauged catchments.

2.10 Catchment characteristics used for hydrological regionalisation

Catchment characterization is typically focused on those physical and ecological attributes that influence the storage, movement and release of water to runoff and evaporation within a catchment (Blöschl et al., 2013). For hydrological regionalisation studies, the primary focus is on geology, soil characteristics, topography, stream network geometry, land-use and land cover (Blöschl et al., 2013). The variation of these attributes is generally associated with the long-term hydrological and geomorphic processes within the catchment (Blöschl et al., 2013). The characterization of specific catchment characteristics can be derived via remotely sensed data, such as topography and land cover classification, or through the process of field assessments. Understanding the influence of these catchment characteristics is crucial in identifying the influence on hydrological processes (Blöschl et al., 2013). For example, geological information can provide valuable insight into groundwater contributions to runoff, or even evaluating the effect of vegetation cover on runoff production mechanisms (Blöschl et al., 2013).

The use of catchment attributes, however, varies between different studies (Razavi & Coulibaly, 2013). A study by Kokkenon et al (2003) indicated that catchment attributes that are used for regionalisation studies should characterise the factors that drive the hydrological response (Razavi & Coulibaly, 2013). In general, the most widely used catchment attributes include catchment area, elevation, slope of catchments, and mean annual or daily rainfall and temperature (Razavi & Coulibaly, 2013). The use of suitable catchment attributes for a given study area would increase the reliability of the estimation of river flow and/or other flow characteristics.
2.10.1 Rainfall

Rainfall has been identified as the main climatic forcing of the hydrological catchment response (Blöschl et al., 2013; Drogue & Khediri, 2016; Mbali, 2016). Studies have emphasized that without reliable measurements or estimates of rainfall, the use of hydrological models becomes meaningless and the understanding of the influence on the water balance is therefore limited (Blöschl et al., 2013; Drogue & Khediri, 2016). Information about precipitation, in terms of both spatial and temporal characteristics, is an essential component for hydrological studies.

Rainfall data are therefore an important input in hydrological modelling (Keblouti et al., 2012). Rain gauges provide accurate measurements of rainfall for that specific location of the gauge, however, inaccuracies can occur when estimating rainfall between them (Visessri, 2014). In most situations, the network of rainfall measuring stations is sparse and the available data are insufficient to characterise the variability and distribution of rainfall within a given region (Keblouti et al., 2012; Visessri, 2014). For this particular reason, it is necessary to develop methods to estimate rainfall in regions where rainfall data are unavailable (Keblouti et al., 2012; Visessri, 2014). Spatial interpolation is an approach that estimates rainfall over specific regions, by utilising point rainfall data and interpolating rainfall over an area of interest. However, the advantages and disadvantages of the interpolation strongly depend on the characteristics of the input data, topography and orographic effects and the density of the gauge network (Keblouti et al., 2012; Visessri, 2014).

2.10.2 Evaporation

Evaporation has been identified as having a substantial influence on the water balance and the generation of runoff (Bouwer et al., 2008; Love et al., 2010; Shoko, 2013; Bloschl et al., 2013; Zhao et al., 2013). Bloschl et al (2013) expressed that the annual water balance and runoff variability is controlled mainly by rainfall and evaporation. Differences between rainfall and evaporation can explain much of the runoff variability. Evaporation dominates the water balance and governs the availability of water in arid and semi-arid regions, where small changes in evaporation rates can consequently lead to large changes in volume of
surface flows (Bouwer et al., 2008; Love et al., 2010). In humid catchments, the amount of rainfall that is received is greater than the amount of water lost to evaporation. In arid catchments, however, the amount of water lost to evaporation exceeds the amount of water received (Bloschl et al., 2013). According to statistics, evaporation rates in humid areas are approximately 50% of the annual precipitation, whereas in arid regions, evaporation is approximately 90% of annual precipitation (Mazvimavi, 2003; Zhao et al., 2013). Camacho Saruez et al (2015) expressed that the rates of evaporation in arid and semi-arid regions are limited to the amount of water available, whereas in humid areas, evaporation rates are limited by the amount of energy available. The influence of evaporation therefore becomes an important component in understanding the runoff variability in arid and semi-arid areas.

Different methods have been developed to estimate evaporation rates at the catchment scale including meteorological ground-based point data and remote sensing techniques (Shoko, 2013). Meteorological methods such as Penman Monteith and Thornthwaite are based on point measurements, such as wind speed, radiation, relative humidity and temperature, which are used to estimate evaporation (Shoko, 2013). The limitation of using meteorological methods is that the meteorological stations are often unevenly distributed and the reference evaporation is derived through the interpolation of point evaporation data. Micro-meteorological approaches such as the pan evaporation method, have been extensively used to estimate evaporation and have shown to be invaluable in this regard (Gokool et al., 2016). However, these approaches are often complex to apply, data intensive and often only applied in small scales with homogenous land cover (Gokool et al., 2016). The advancement of remote sensing techniques provides the opportunity to estimate evaporation over large areas, especially within inaccessible regions (Shoko, 2013; Gokool et al., 2016). The use of remote sensing techniques for estimating evaporation provides a relatively timeous and cost effective approach (Gokool et al., 2016). The limitations foreseen with remote sensing techniques are often related to the required data and the resolution and quality of the imagery (Bouwer et al., 2008; Shoko, 2013; Gokool et al., 2016).

2.10.3 Slope and topography

Topography has been described as the dominant catchment characteristic governing the movement of water within a catchment (Jencso & McGlynn, 2011; Malan, 2016). Studies
have shown a strong correlation between elevation with rainfall, ET, soil patterns and vegetation cover (Parajka et al., 2013; Hallema et al., 2016). The ability of hillslopes to retain and release water affects runoff generation (Hallema et al., 2016). Regions of higher elevation with steep slopes, higher rainfall, lower ET, and shallow soils have high surface runoff (Guzha et al., 2018).

Studies have shown that an increase in slope results in a reduction in soil infiltration, therefore increasing runoff (Mu et al., 2015). Nassif et al (1975) conducted a study which evaluated the influence of slope on runoff and infiltration. The results of the study suggested that there is a threshold slope value beyond which the peak runoff is relatively unaffected (Mu et al., 2015). Cerdá & García-Fayos (1997) conducted a study to evaluate the influence of the angle of the slope on runoff generation and the results indicated that the slope angle did not have any influence on the volume of runoff, but there was a clear indication of the influence the angle has on runoff initiation (Mu et al., 2015). Slope explains the rate of movement of water by kinetic energy to the catchment outlet (Mazvimavi, 2003; Masoudian, 2009; Mirus & Loague, 2013). The slope will influence the catchment response time to rainfall events, which has a substantial influence on the concentration and travel times of flow (Masoudian, 2009; Gericke & du Plessis, 2012). Slope is highly variable between catchments and no single slope index has been commonly agreed upon. The reason for this is due to the fact that a single slope index may not be representative of the catchment that affects runoff processes (Mazvimavi, 2003). Etekhabi (2001) was of the view that the average slope was unrepresentative and recommended median slope to be more representative of the catchment (Blöschl et al., 2013).

2.10.4 Geology

Geology has been shown to have considerable influence on the infiltration and subsurface water storage (Larocque et al., 2010; Jencso & McGlynn, 2011; Rumsey et al., 2015; Ries et al., 2017). The type of lithology has shown to have considerable influence on groundwater recharge, which subsequently influences baseflow.

The representation of geological characteristics in a quantitative manner is a major challenge for hydrological regionalisation (Mazvimavi, 2003; Tallakan & Van Lanen, 2004; Hughes,
2006; Bloomfield, 2009; Blöschl et al., 2013; Pfister et al., 2017). The reason for this is due to the fact that hydrogeological characteristics, such as depth to the water and the permeability of the soil are highly variable in space (Blöschl et al., 2013). Such data are not readily available and the data that is provided is often a generalised coverage of various geological characteristics (Mazvimavi, 2003; Hughes, 2006). Hughes (2006) expressed that in many developing countries the derivation of geological characteristics is based on coarse geological maps with low spatial resolution and not generated using hydrologically meaningful sources. Based on this inherent constraint, regionalisation studies, particularly in developing countries, focus on the proportion of different lithologies within the study area (Mazvimavi, 2003; Laaha & Blöschl, 2006).

2.10.5 Soil

The characteristic features of soils have been described in studies as having a substantial control on rainfall infiltration, percolation and moisture storage (Mohamoud, 2004; Larocque et al., 2010; Schmocker-Fackel et al., 2007; Malan, 2016). On some soils, infiltration of rainfall can be low which leads to higher runoff. On the other hand, some soils promote infiltration of rainfall, which results in reduction of flow (Schmocker-Fackel et al.). In arid and semi-arid regions, much of the water is infiltrated before the water reaches a river (Perrin & Tournoud, 2009). Soil properties, therefore have a critical influence on the movement of water within a catchment, controlling the direction and speed of flow (Malan, 2016). The importance of soil as a descriptor in hydrological studies is well documented, which is evident in the inclusion of soil parameters in hydrological models (Mohamoud, 2004).

Studies have shown relationships between underlying bedrock geology and topographic position, which has a significant influence on the rate of formation and properties of the soil (Mohamoud, 2004; Mu et al., 2015; Guzha et al., 2018). These factors are therefore important in identifying the hydrological influence of soil parameters on runoff generation (Mohamoud, 2004). For example, soil depth has been found to decrease with increasing elevation (Malan, 2016). Therefore when analysing hydrological signatures of catchments, the ability of catchments to retain water is low in regions of steep slopes compared to flat terrains, which therefore results in a low moisture storage capacity (Mohamoud, 2004). Moisture storage capacity has a critical influence on the generation of runoff, as when
precipitation occurs, water infiltrates and fills the available storage, whereby the soil becomes highly responsive and quickly releases a large quantity of the rainfall event as quickflow runoff (Mohamoud, 2004). The storage and distribution of water in the soil is, however, also strongly influenced by vegetation cover.

An issue that arises in South Africa and other developing countries is that soil data is often based on coarse maps which have not been produced in a hydrologically meaningful manner (Hughes, 2010). Soil data available are relevant for agricultural purposes, but difficult to derive hydrologically meaningful information (Hughes, 2010). However, although detailed data does not exist on the hydraulic properties of the soil, the soil textural class can be used to derive hydraulic properties of the soil (Kapangaziwiri et al., 2012).

2.10.6 Drainage density

Drainage density (Dd) is an important catchment characteristic which influences runoff processes, such as the intensity of flood events, the distribution of sediment and the availability of water within a catchment (Mazvimavi, 2003; Pallard et al., 2009; Zavoianu, 2011). Catchments characterized by high Dd show increasing peak flow. Decreasing Dd is typically associated with decreasing peak flow, due to longer travel times (Pallard et al., 2009; Love et al., 2011).

2.10.7 Land cover

Land cover has considerable influence on interception, infiltration, evapotranspiration and runoff (Farley & Jobbágy, 2005; Blöschl et al., 2013; Zhao et al., 2015; Ries et al., 2017). Arid and semi-arid regions are characterised by extreme temperatures and rainfall events of short duration and high intensity, which leads to low infiltration and thus large volumes of runoff (Zhang et al., 2014).

The interception of rainfall by tree canopy can substantially reduce the volume of net rainfall reaching the ground, which would subsequently increase rates of transpiration (Farley & Jobbágy, 2005; Zhao et al., 2015; Malan, 2016). The interception rates tend to be inversely
proportional to rainfall intensity and directly proportional to the density of plant canopy (Blöschl et al., 2013). Therefore catchments with low rainfall intensities and high density of plant cover, interception rates will thus be high.

Vegetation cover also significantly influences the rate of infiltration. For example, forests generally have deep roots and associated high infiltration rates (Farley & Jobbágy, 2005; Zhao et al., 2015; Malan, 2016). The rate of infiltration is influenced by the type of vegetation cover. Grasses have been shown to increase or decrease the rate of infiltration, which influences runoff (Mazvimavi, 2003; Chikodzi, 2013; Duan et al., 2017).

2.11 Summary

The importance of rivers in providing water for both ecological and societal needs has been globally recognised. Although non-perennial rivers have not been widely researched, these rivers are still important for the functioning of aquatic ecosystems and remain among the least studied freshwater systems. The influence of climate change and anthropogenic activity on river flows further increases the complexity of effectively managing non-perennial rivers. Environmental flows have become an important concept in recent years, which identifies the flow that is required to achieve desired objectives. Various methods have been developed ranging from fairly simplistic approaches to complex holistic approaches. In the context of non-perennial rivers, no formal method has emerged for environmental flow assessments which introduce doubt as to whether or not these rivers are being managed to maintain ecological integrity.

This chapter explored hydrological indices that could be used to characterise non-perennial rivers in an ecologically meaningful way. Although a large number of hydrological indices exist, the concern that arises is that these hydrological indices have been developed in the context of perennial rivers. This chapter has outlined the hydrological differences between perennial and non-perennial rivers, which emphasize the need to introduce and develop hydrological indices that characterise the flow regime explicitly for non-perennial rivers. This would provide the opportunity to improve the management of non-perennial rivers and allow water managers to maintain ecological integrity.

The chapter also explored methods available for predicting flow characteristics in ungauged catchments. Various methods have been developed to predict flow characteristics in
ungauged catchments based on hydrological model-dependent and hydrological model-independent methods. Literature has highlighted the limitations of hydrological model-dependent methods as the approach requires physical attributes for parameter estimation, as well as relevant expertise. Hydrological model-independent methods such as multiple regression and artificial neural networks seem to be most promising. The use of cluster analysis has also been identified as being a feasible approach for transferring information between gauged and ungauged catchments.
CHAPTER 3: METHODOLOGY

3.1 Introduction

This chapter provides the criteria for selecting non-perennial river gauging stations in the study. The chapter presents the catchment characteristics that have been considered for the study, as well as the methods used for derivation and analysis of data. The catchment characteristics include topographic, land cover and climate variables that can be used to explain the variability of flow characteristics. The chapter also presents the criteria for selecting ecologically relevant flow indices for the study. Lastly, the chapter highlights the methods that have been selected to determine catchment attributes that account for the spatial variation of ecologically relevant flow indices and the prediction of such flow indices in ungauged catchments.

3.2 Selection of gauging stations

The study investigates non-perennial rivers in the Western, Northern, and Eastern Cape. Time series of daily flow data for South African rivers are available from the Department of Water and Sanitation (DWS) online data archive (http://www.dwa.gov.za/Hydrology). The time series of daily flow data provided the opportunity to distinguish between perennial and non-perennial rivers in the study area. Rivers that have periods of no flow during the year were considered as non-perennial. Non-perennial rivers in the study area were only selected for further analysis depending whether or not the gauging station complied with the specified criteria.

The selection of gauging stations for this study was based on the accuracy and availability of continuous flow data, and the absence of flow regulation. Time series of flow collated from DWS were considered to be accurate data, as the data has undergone various accuracy control checks. The gauging stations are adequately maintained and the rating curves are constantly checked to evaluate their accuracy and improve where necessary. The gauging stations are equipped with electronic data loggers and backup data loggers are also installed at the sites to ensure data integrity (ORASECOM, 2011). Gauging stations were only selected in the study if the missing data in the time series of flow was less than 10 % of the total time series of
flow used for the analysis, which was based on general guidelines and recommendations for the selection of gauging stations in hydrological studies (Sinclair Knight Merz, 2010; Chiverton, 2015). Missing data in time series of flow may be due to equipment failure and/or computational errors (Gao, 2017). When the period with missing data is large, flow statistics may not be representative of the catchment. Network model data for each primary catchment were extracted from WR2012, which identifies various land-uses and the presence of upstream dams within each catchment. Google Earth Pro was also used as an aid to identify the presence of upstream dams and various land-uses which significantly influence natural flows. Gauging stations that were clearly influenced by anthropogenic activity were not considered in the study. Flow regulation and catchment urbanization change the natural conditions of the river system, by way of altering drainage patterns and the hydrology of the catchment (Olden & Poff, 2003; Naiman et al., 2008; Rolls et al., 2012). It is important for the study to include river gauging stations that have minimal anthropogenic influence for the purpose of understanding how these systems operate in their natural state. River gauging stations were included in the study for further analysis if the above criteria were met.

3.3 Delineation of catchments

The purpose for delineating catchment boundaries was for the extraction of spatial datasets of catchment characteristics, such as the variation of land cover and geology. The traditional approach for delineating catchments was based on the use of topographic maps that indicate stream channels and contours. However, this approach is extremely time-consuming. Recently catchment delineation has been achieved using automated techniques, through the use of software such as ArcGIS, Surfer and ILWIS (Abed, 2013). In automated techniques, the basic input data is a raw digital elevation model (DEM).

The main factors that account for the accuracy of the DEM include the source resolution and the spatial resolution of the dataset (Mason & Maidment, 2000). The source resolution determines the level of content that may be extracted during digitization (Mason & Maidment, 2000). In most standard DEM’s, surface characteristics are generalised by being reduced to an individual node within the dataset. The issue of generalisation is that features that are smaller than the internal spacing of the nodes are not distinguished, and this results in the ‘smoothing’ of the surface within the DEM. Therefore, the general assumption is that
elevation data sources with higher spatial resolution more accurately represent terrain features, and thus results in a higher accuracy for delineating catchment boundaries (Mason & Maidment, 2000). Studies have also shown that the overall accuracy of the DEM data depends on the location and the type of land cover (Tighe & Chamberlain, 2009; Mehta, 2017).

Several DEM’s have been developed such as USGS National Elevation Data (NED) 10 m and 30 m, Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) 30 m, and Intermap’s STAR technology with a resolution of 5 m (Tighe & Chamberlain, 2009). Most of these DEMs are not freely available and are often only available for conterminous United States and Western European countries. The study evaluated SRTM and ASTER GDEM, which is a product of METI and NASA. A study by Tighe & Chamberlain (2009) concluded that the ASTER GDEM does not meet the published specifications, regarding the initial proposed accuracy of the DEM. For this particular reason, the SRTM 1 arc second 30 m DEM obtained from USGS Earth Explorer (https://earthexplorer.usgs.gov/, Accessed: June 2017) was used.

ArcGIS was used to delineate the catchment boundary for the selected gauging stations. A DEM is used as an input dataset and contains data relevant to the land surface, which is then filled to fill voids that are present in the DEM (Mason & Maidment, 2000). Voids may be a result of missing data, and can therefore affect the accuracy of the delineation process through misrepresentation of the land surface. Once the DEM has been filled, the flow direction and flow accumulation tool (ESRI, 2018) was run in ArcGIS. A pour point is then assigned to the region, which indicates the outlet of the catchment (ESRI, 2018). In this study, the pour point was represented by the geographical location of the gauging station. Once the pour point was created, the catchment tool (ESRI, 2018) in ArcGIS could be run to generate the catchment boundary, which uses the flow direction and pour point as the input data, and produces a raster output layer (ESRI, 2018). The raster layer needs to be converted to a vector shapefile, to create a polygon which represents the catchment boundary.
3.4 Catchment descriptors

Catchment characteristics were selected in the study on the basis of the results and recommendations of previous studies, which were identified as influencing flow as presented in Chapter 2. Thus the catchment characteristics selected include those reflecting the influence of topography, soil properties, land use and climatic factors (Table 3.1).

Table 3.1: Catchment descriptors selected in the study

<table>
<thead>
<tr>
<th>Catchment Characteristic</th>
<th>Description and Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Mean annual rainfall</td>
<td>Derived from the WR2012 online data archive</td>
</tr>
<tr>
<td>2. Mean annual evaporation</td>
<td>Derived from the WR2012 online data archive</td>
</tr>
<tr>
<td>3. Proportion of the catchment with different lithologies</td>
<td>Derived from the 1:1 000 000 geological map of South Africa produced by CGS</td>
</tr>
<tr>
<td>4. Slope</td>
<td>Derived from 30m SRTM 1-arc second DEM</td>
</tr>
<tr>
<td>5. Elevation</td>
<td>Derived from 30m SRTM 1-arc second DEM</td>
</tr>
<tr>
<td>6. Soil</td>
<td>Derived from the 1: 250 000 soil map produced by WR2012</td>
</tr>
<tr>
<td>7. Area</td>
<td>DWS online data archive</td>
</tr>
<tr>
<td>8. Drainage density</td>
<td>Derived from the catchment area and 1:50 000 river shapefile produced by NGI</td>
</tr>
<tr>
<td>9. Proportion of catchment with different land cover types</td>
<td>2014 Land Cover map of South Africa produced by NGI</td>
</tr>
</tbody>
</table>

Chapter 2 has already discussed the importance of the factors given in Table 3.1. In this chapter, the methods for collecting and analysing data for each of the above are explained.
3.4.1 Elevation and slope

The SRTM 1-arc second 30 m DEM (https://earthexplorer.usgs.gov/, Accessed: June 2017) was used to derive elevation and slope statistics for the selected catchments in the study. ArcGIS uses a DEM as the input layer and then generates a surface slope as an output layer for the catchment. The technique used in ArcGIS to calculate surface slope determines the rate of change (delta) of the surface in horizontal (dz/dx) and vertical (dz/dy) from the centre cell. The algorithm can be expressed as (ESRI, 2018):

\[
\text{slope\_degrees} = \arctan \left( \sqrt{\left(\frac{dz}{dx}\right)^2 + \left(\frac{dz}{dy}\right)^2} \right) \times \frac{180}{\pi}
\]  

where; \( \arctan \) returns the inverse tangent of the elements of \( x \), \( \frac{dz}{dx} \) is the rate of change in the \( x \) direction from the centre cell, \( \frac{dz}{dy} \) is the rate of change in the \( y \) direction from the centre cell, \( \pi \) is pi.

The maximum rate of change in elevation between the centre cell and its neighbouring cell identifies the steepest descent. The surface tool creates a 3 x 3 grid of cells with associated z-values (ESRI, 2018). The slope values are calculated using the average maximum technique.

Once the surface slope layer has been generated, the layer provides slope values for each individual pixel in the catchment. Individual pixel values can then be used to construct a cumulative frequency distribution of slope from which slope indices \( S_\Psi \) are derived. \( S_\Psi \) refers to the slope value for which \( \Psi \) % of the pixels in the catchment is equal to or less than this value. The study represented slope in degrees, and thus the slope values range between 0 to 90°. The study used several slope indices for values of \( \Psi \) from 10 % to 100 %. The inclusion of several slope values provides the opportunity to identify the slope value that best explains the variance of flow characteristics.

3.4.2 Rainfall

Rainfall data used in the study were extracted from the WR2012 online data archive (http://waterresourceswr2012.co.za/resource-centre/, Accessed: June 2017). This data were considered to be of an acceptable quality since the WR2012 project undertook the following quality controls. WR2012 used all useful individual rainfall station data from DWS, which
was updated to 2010 and the data was patched where necessary. DWS used a screening process to validate the data and evaluate the consistency of the data by means of visual assessment and standard validation tests known as ‘single mass plots’ and ‘cusum plots’ to identify anomalies in the data.

Rainfall data from WR2012 were available in the form of point measurements and pixel averages generated through interpolated point data. For this study, the pixel averages were used as the spatial variation of catchment rainfall was important to the study. Rainfall data derived from WR2012 were examined through a classification process in ArcGIS, by evaluating the distribution of individual pixels of rainfall values to derive various statistical parameters such as minimum, mean and maximum rainfall within the catchment.

3.4.3 Evaporation

Pan evaporation data were used in the study and is referred to as ET in the study. The reason for selecting pan evaporation over other sources of estimating evaporation has been identified in Chapter 2. ET data were extracted from WR2012 (http://waterresourceswr2012.co.za/resource-centre/, Accessed: June 2017), which provided monthly evaporation data for each quaternary catchment. Quaternary evaporation data was used to derive evaporation at the catchment scale. If the catchment was located in more than one quaternary catchment, the quaternary data would be averaged to derive ET at the catchment scale.

3.4.4 Drainage density

The drainage density was derived by dividing the total river length by the catchment area. The 1: 50 000 river shapefile from the Land Cover Map of South Africa 2013 produced by the National Geo-spatial Information (NGI) was used to determine the length of the drainage network within individual catchments. The data was collected from the Department of Rural Development- National Geo-spatial Information, Mowbray in June 2018.
3.4.5 Geology

The classification of geological characteristics for the study was based on the 1: 1 000 000 geological map of South Africa, which was produced by the Council of Geoscience (CGS) in 2013 (http://www.geoscience.org.za/index.php/publication/downloadable-material). The dataset contains the lithostratigraphic, chronostratigraphic and lithologic data, which are the essential attributes used to examine their influence on flow characteristics.

3.4.6 Soils

The 1: 250 000 soil data in the WR2012 (http://waterresourceswr2012.co.za/resource-centre/) were used to derive the influence of soil on flow characteristics of interest. The soil data from WR2012 contained information of the textural class of the soil, including sandy, sandy loam and clayey loam, as well as the depth of the soil. The textural class was used in the study as a surrogate to identify the influence of soil on flow characteristics. The proportional area covered by each textural class in individual catchments was calculated in ArcGIS, by determining the proportional area covered by a textural class in comparison to the total area of the catchment.

3.4.7 Land cover

The 1: 100 000 National Land Cover maps of South Africa 2014, produced by NGI, were used to identify the variation of land cover types in the study area. The study evaluated the influence of land cover on selected flow characteristics through examining the influence of different land cover types within the catchment, which is expressed as the proportional area covered within the catchment.
3.5 Flow characteristics

This section introduces the flow characteristics selected in the study, the basis upon which they were selected and the derivation of these flow characteristics.

3.5.1 Selection and derivation of flow characteristics

River flow data were extracted from the DWS online data archive (http://www.dwa.gov.za/Hydrology/ Accessed: March 2017) for the selected gauging stations. The use of long-term time series of flow is important in hydrological studies to ensure representative estimates of river flow statistics. The recommendation of previous studies was used to select an appropriate record length (Razavi, 2014; Chiverton, 2015; Jian et al., 2016). A study by Kennard et al (2010) examined the uncertainty in the estimation of hydrological indices. The study suggested that the bias in the estimation of flow indices substantially decreased and the overall precision and accuracy noticeably increased with increasing record length. Based on the recommendations of previous studies the hydrological indices were derived from time series of daily flow based on 30 years of hydrological data from 1986-2016.

Monk et al (2007) expressed that the selection and derivation of hydrological indices is time consuming and is not a simple procedure due to the large number of ecologically relevant hydrological indices that exist, and the inherent redundancy that exists among them. Olden & Poff (2003) and Kennard et al (2010) concluded that it is important to select hydrological indices on the basis of the ecological question of interest. Monk et al (2007) suggested that future research should employ a refined number of hydrological indices based on the Indicators of Hydrologic Alteration (IHA), where redundant indices have been removed using hydrological understanding, rather than being dependent on statistical approaches. This approach allows the analysis of the full range of the flow regime, and maximises the understanding of the relationship between flow and ecology. Taylor et al (2003) concluded that the use of the IHA is beneficial with the absence of ecological data and expertise, and preliminary environmental flows can be set using historical or simulated hydrological data.
The absence of ecological expertise, as well as observed runoff, is a common issue for South African catchments (Taylor et al., 2003).

The selection of hydrological indices in the study was based on recommendation of previous studies. Monk et al (2007) proposed the use of hydrological indices based on the IHA, as the number of hydrological indices has been refined. The study therefore used the hydrological indices based on the IHA and redundancy analysis was used to remove redundant hydrological indices. Studies have also emphasized the importance of selecting hydrological indices that are relevant for the objectives of the study. Thus alternative hydrological indices were explored to efficiently characterise the flow regime of non-perennial rivers in an ecologically meaningful way, which includes annual flow percentiles, hydrological index (CVB), concavity index (IC) and coefficient of variation of annual flows (CV). These hydrological indices were considered as they are representative of the shape of the flow duration curve, as well as an indication of the degree of variability between long-term annual flows. The selected ecologically relevant flow indices in the study are presented in Table 3.2.

Table 3.2: Selected ecologically relevant flow indices of non-perennial rivers for the study

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Flow Index</th>
<th>Flow Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q</td>
<td>Mean annual runoff</td>
<td>CVB</td>
</tr>
<tr>
<td>q_{10}</td>
<td>Flow exceeded 10% of the time</td>
<td>IC</td>
</tr>
<tr>
<td>q_{25}</td>
<td>Flow exceeded 25% of the time</td>
<td>BFI</td>
</tr>
<tr>
<td>q_{75}</td>
<td>Flow exceeded 75% of the time</td>
<td>ZFD</td>
</tr>
<tr>
<td>q_{90}</td>
<td>Flow exceeded 90% of the time</td>
<td>3-day min</td>
</tr>
<tr>
<td>CV</td>
<td>Coefficient of variation of annual flows</td>
<td>3-day max</td>
</tr>
</tbody>
</table>

This study estimates BFI as the ratio between baseflow to total flow and therefore an alternative approach was required for deriving BFI. This study used digital filtering approach
to estimate BFI based on the time-series of flow. Alternative techniques such as chemical and isotope tracing techniques, electrical resistivity imagery, geophysical measurements and simulation modelling were not considered in the study, as these techniques are often marked as being immensely time consuming and requiring extensive resources (Smaktin, 2001; Hughes et al., 2002; Wenninger et al., 2010).

The use of a digital filtering approach has been widely used for the separation of baseflow and quickflow from a hydrograph. The digital filtering algorithm can be expressed as (Hughes et al., 2002):

\[ q_i = \alpha q_{i-1} + \beta (1+\alpha) (Q_i-Q_{i-1}) \]  

\[ QB_i = Q_i-q_i; \text{ where} \]

\[ Q_i = \text{total flow time series} \]
\[ q_i = \text{high flow time series component} \]
\[ QB_i = \text{baseflow time series component} \]
\[ i = \text{daily time step} \]
\[ \alpha, \beta = \text{separation parameters} \]

Hughes et al (2002) stated that there is no need to change the value of \( \beta \) (0.5), as there is more than enough flexibility for adjusting the \( \alpha \) value to achieve a desired result. The \( \alpha \) parameter essentially controls the volume of baseflow. The higher the \( \alpha \) value the lower the baseflow volume. Smaktin and Watkins (1997) conducted a study on South Africa rivers and concluded that an \( \alpha \) value of 0.995 can be considered suitable for estimating baseflow using digital filtering. This method was used for the separation of baseflow and quickflow in their study to derive BFI.

The prediction of selected hydrological indices in the study can also be used to develop an understanding of the catchment and derive various parameters of the channel relating to various disciplines including fluvial geomorphology and hydrogeology. For example, the prediction of flow characteristics can provide information relating to the wetted perimeter, bankfull discharge and flow depth.
3.6 Prediction of flow indices

This section briefly describes the multivariate analysis methods used in the study. The section identifies methods that can be used for developing predictive equations of flow characteristics, methods available for determining the effects of different combination of catchment characteristics on flow characteristics, and methods that are suitable for grouping catchments into homogenous groups.

3.6.1 Multivariate analysis

Redundancy analysis (RDA) (Mazvimavi, 2003; Chikodzi, 2013) was chosen for this study because the approach is suitable in determining effects of different combinations of catchment characteristics on sets of flow characteristics. RDA was also chosen for the study because the majority of the responses and explanatory variables have some form of linear relationship, for example, runoff increases with increasing precipitation and slope, and decreases with evaporation. The main purpose of redundancy analysis was to determine which catchment characteristics significantly explain the variance of flow characteristics, which are important for predicting flow characteristics later in the study. The identification of significant catchment characteristics can be used as the basis for grouping catchments into clusters that share similar hydrological responses. RDA was performed using Canonical Community Ordination (CANOCO), which is a general package for multivariate data analysis, with an emphasis on dimension reduction (ordination) and regression analysis or the integrated combination of the two, called constrained ordination (ter Braak & Šmilauer, 2018). A more detailed description on redundancy analysis is discussed in Chapter 5.

Multiple linear regression (MLR) (Mazvimavi, 2003; Lacombe et al., 2014) and artificial neural networks (ANNs) (Mazvimavi, 2003; Riad et al., 2004; Aichouri et al., 2015) were used in Chapter 6 to identify which approach would be more suitable for predicting selected flow characteristics of non-perennial rivers. MLR was used in the study, due to the general assumption of the linear relationship that exists between flow and catchment characteristics. However, the presence of linear relationships is not always evident between flow and catchment characteristics, and the use of ANN may be more suitable for the study to predict...
flow characteristics, due to complex non-linear relationships. A more detailed description of MLR and ANN are discussed in Chapter 6.

Cluster analysis (Mazvimavi, 2003; Chikodzi, 2013) is an approach of grouping objects into homogenous groups based on similarities and differences. The results from redundancy analysis in Chapter 6 are used for cluster analysis. The results from Chapter 6 can be used to classify and group catchments into homogenous groups that share similar hydrological responses. This approach was applied towards the end of the study, which attempted to group catchments into homogenous groups based on similarities of both flow and catchment characteristics. The purpose of cluster analysis was to determine whether or not the grouping of catchments improves the prediction of flow characteristics. A more detailed description of cluster analysis is presented in Chapter 7.
4.1 Introduction

This chapter presents the gauging stations that were selected in the study, as well as a description in the variation of catchment and flow characteristics.

4.2 Selected catchments

Based on the selection criteria presented in Chapter 3, 36 gauging stations were selected in the study area. Figure 4.1 shows the locations of the selected gauging stations of non-perennial rivers within the study area in the Western, Northern and Eastern Cape Provinces of South Africa.

Figure 4.1: Location of selected non-perennial gauging stations
The catchments were delineated (Figure 4.2). Table 4.1 shows the distribution of selected catchments based on their catchment size, with the smallest and largest catchment areas being 23 and 8232 km$^2$ respectively.

![Figure 4.2: Study catchments delineated using SRTM 1 arc second 30 m DEM](http://etd.uwc.ac.za/)

Catchments with an area of less than 200 km$^2$ were the most dominant (36 %) whilst 3 catchments had areas of greater than 4000 km$^2$ (Table 4.1).

Table 4.1: Distribution of catchment area for selected river gauging stations in the study area

<table>
<thead>
<tr>
<th>Catchment size (km$^2$)</th>
<th>Number of catchments</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 200</td>
<td>13</td>
</tr>
<tr>
<td>201-500</td>
<td>9</td>
</tr>
<tr>
<td>501-1000</td>
<td>6</td>
</tr>
<tr>
<td>1001-4000</td>
<td>5</td>
</tr>
<tr>
<td>&gt; 4000</td>
<td>3</td>
</tr>
</tbody>
</table>
4.3 Elevation and slope

Elevation ranges between 11 and 2243 m above mean sea level, with an overall mean of 1186 m (Figure 4.3). Catchments that are characterised by high elevations are situated within the central and northern parts of the study area. Catchments with low elevations are generally located along the coastal parts of the study area, within the south-western, southern and south-eastern parts.

Figure 4.3: Variation of elevation within the study area derived from SRTM 1 arc second 30m DEM

Slope is highly variable among the selected catchments. Figure 4.4 shows the variation of slope based on five representative catchments. H1H006, which is a flow gauging station on the Breede River located in the south-western part of the study area, shows the largest difference between maximum and minimum slope within its catchment, while flow gauging station R2H009 shows the smallest variation between the maximum and minimum slope.
Catchments with the highest median slopes are located mainly in the south-western regions of the study area, with slope values ranging between 2.23° and 27.6° and an overall mean of 10.42° (Figure 4.5). These catchments are located within mountainous regions and therefore the median slope can be expected to be higher. Most of the catchments had a median slope of between 4°-10°.
4.4 Rainfall

Rainfall varies considerably between the Western, Eastern and Northern Cape. The Western Cape has the highest spatial variation in rainfall compared to other provinces in South Africa, with a low of 60 mm/year and a peak of 3345 mm/year, which most areas receive 350 to 1000 mm/year (CSIR, 2014). Rainfall is strongly influenced by the cold Benguela current and coastal winds. The influence of both the Atlantic and Indian Oceans results in coastal temperatures differing over short distances, with temperatures typically ranging between 10º to 35 ºC (CSIR, 2014). The Eastern Cape is situated between the Mediterranean and Subtropical regions, while the northern regions of the province are arid. Summer temperatures range between 16º to 26 ºC, while winter temperatures range between 7º to 20 ºC. Temperature increases more inland, with a decrease in mean annual rainfall, resulting in increasing aridity. The Northern Cape has semi-arid to arid conditions with high temperatures, with extremes of 47.8 ºC and sparse rainfall, and a mean annual rainfall between 50 and 400 mm/year (CSIR, 2014). The Western Cape has a Mediterranean climate, characterised by dry, hot summers (October to April) and cold, wet winters (June to August) (CSIR, 2014).

Figure 4.6: Mean annual rainfall within the study area derived from WR2012 (http://waterresourceswr2012.co.za/resource-centre/; Accessed June 2017)
Mean annual rainfall (MAP) of selected catchments varies from 80 to 3312 mm/year (Figure 4.6). This is mainly due to the influence of altitude, whereby there is a strong correlation between rainfall and elevation. Lowest mean annual rainfall are found in the central and north-eastern parts, whereas the highest mean annual rainfall are found within the south-western parts of the study area.

4.5 Evaporation

ET ranges between 1370 to 2042 mm/year, with an overall mean of 1603 mm/year (Figure 4.7). Low ET rates are found within the south-western, southern and south-eastern parts of the study area, with ET dramatically increasing from the coastal parts towards to the central parts of the study area.

Figure 4.7: Reference evapotranspiration within the study area derived from WR2012 (http://waterresourceswr2012.co.za/resource-centre/: Accessed June 2017)
4.6 Drainage density

Catchments with the lowest drainage density (Dd) are predominantly located in the north-eastern and south-western parts of the study area (Figure 4.8), whereas the highest Dd is found predominantly in the central parts of the study area. The lowest and highest drainage density for the study area is 1.22 and 6.39 km/km² respectively with an overall mean of 2.70 km/km². Catchments with higher Dd are expected to have increased runoff and peak flow, as these catchments are more dissected than catchments characterised by a lower Dd.

Figure 4.8: Drainage density of the study area derived from NGI 1:50 000 river map

4.7 Geology

The study area is mainly comprised of the Beaufort Group (37 %), including mudstone, sandstone and mudrock, which are found in the northern and north-eastern parts of the study area. The Table Mountain Group, including quartzitic sandstone, shales, alluviums, sand and calcrete, covers 18 % of the study area. The Bokkeveld Group underlies 17 % of the study area, including sandstone, shales and siltstone. The Bokkeveld and Table Mountain Groups
are mostly found in the southern and south-western parts of the study area. The remainder of the study area (28%) has other lithologies (Figure 4.9).

![Lithology of the catchments in the study area derived from the 1:1 000 000 geological map of South Africa](Figure 4.9)

4.8 Land cover

Shrubland is the dominant land cover type covering 54% of the study area, mainly in the north-eastern and south-western parts of the study area (Figure 4.10). Forests and cultivated land cover 16 and 13% of the study area respectively. Forests are found within the south-western and southern parts of the study area, with noticeable concentration of forests in the south-eastern parts (Figure 4.10). Cultivated land is mostly found within the south-western parts of the study area, as well as occurring in the south-eastern parts.
Figure 4.10: Land cover within the study area derived from the South African Land Cover map 2014

4.9 Soils

Figure 4.11: Soil textural classes in the study area derived from WR2012 1: 250 000 soil map of South Africa
Sandy loam is the dominant textural class in the study area, covering 57% of the area (Figure 4.11). Clayey-loam and sandy soils occur on 35% and 8% respectively of the study area. Sandy soils are predominantly found in the central parts of the study area, whilst clayey loam soils are generally found in the south-western and north-eastern parts.

4.10 Flow characteristics

This section presents the selected ecologically relevant flow indices in the study (Table 4.2). The selected flow characteristics for the study have characterised various aspects of the flow regime of non-perennial rivers and have been selected based on their ecological significance. This section also presents the spatial variation of selected flow indices within the study area.

Table 4.2: Selected ecologically relevant flow indices of non-perennial rivers for the study

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Flow Characteristics</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q</td>
<td>Mean annual runoff (mm/year)</td>
<td>mm/year</td>
</tr>
<tr>
<td>CV</td>
<td>Coefficient of variation of annual flows</td>
<td>%</td>
</tr>
<tr>
<td>q90</td>
<td>Dimensionless daily flow with a 90% exceedence</td>
<td>%</td>
</tr>
<tr>
<td>q75</td>
<td>Dimensionless daily flow with a 75% exceedence</td>
<td>%</td>
</tr>
<tr>
<td>q25</td>
<td>Dimensionless daily flow with a 25% exceedence</td>
<td>%</td>
</tr>
<tr>
<td>q10</td>
<td>Dimensionless daily flow with a 10% exceedence</td>
<td>%</td>
</tr>
<tr>
<td>IC</td>
<td>Concavity Index</td>
<td>%</td>
</tr>
<tr>
<td>CVB</td>
<td>Hydrological Index</td>
<td>Ratio</td>
</tr>
<tr>
<td>3-day min</td>
<td>3-day minimum of daily means of discharge</td>
<td>mm/year</td>
</tr>
<tr>
<td>3-day max</td>
<td>3-day maximum of daily means of discharge</td>
<td>mm/year</td>
</tr>
<tr>
<td>ZFD</td>
<td>Number of zero flow days</td>
<td>Days</td>
</tr>
<tr>
<td>BFI</td>
<td>Baseflow Index</td>
<td>Ratio</td>
</tr>
</tbody>
</table>
4.10.1 Mean annual runoff and flow percentiles

Figure 4.12 illustrates the variation of mean annual runoff within the study with the lowest being 1.8 mm/year, highest 498 mm/year and the overall mean annual runoff of 96 mm/year. The highest mean annual runoff is found within the south-western and southern parts of the study area. The lowest mean annual runoff is predominantly found within the central and north-eastern parts of the study area. These trends are expected due to the distribution of rainfall within the study area (Figure 4.6).

![Legend](http://etd.uwc.ac.za/)

**Figure 4.12: Variation of mean annual runoff within the study area**

Figure 4.13 presents the variation of flow percentiles $q_{10}$, $q_{25}$, $q_{75}$, and $q_{90}$ within the study area. Catchments that have high mean annual runoff are expected to also have high flow percentiles. Catchments with high values of $q_{10}$ and $q_{25}$ are found within the south-western, southern and south-eastern parts of the study area. This relationship is expected, as these parts of the study area have high rainfall and steep slopes. These catchments are also characterised by higher $q_{75}$ and $q_{90}$ compared to catchments within the central parts of the study area.
Catchments within the central and south-eastern parts of the study area are characterised by prolonged periods of no-flow, as the $q_{75}$ and $q_{90}$ values are 0 (Figure 4.13). These catchments show a considerable contrast between high and low-flow conditions and are representative of a steep flow duration curve compared to catchments within the south-western parts of the study area.

4.10.2 Baseflow Index

BFI varies in the study area between 0.05 and 0.35 (Figure 4.14), with an overall mean of 0.19. High BFI is predominantly found within the south-western parts of the study area, which is expected as these regions are characterised by high rainfall and elevation (Figure 4.3 and 4.6). Rainfall is the major source of recharge to groundwater systems and this relationship is shown where high BFI coincides with high rainfall. Catchments that are characterised with low BFI have a low mean annual rainfall.
4.10.3 Coefficient of variation of annual flows

The coefficient of variation (CV) of annual flows ranges between 133 and 900 % within the study area (Figure 4.15), with an overall mean of 414 %.
As expected, catchments within the mountainous areas in the south-western parts of the study area with consistent rainfall have lower CV values. Catchments within the central and south-eastern parts of the study area tend to have high CV values and therefore exhibit higher annual variation.

4.10.4 Number of zero flow days

The number of zero flow days varies between 1 and 320 days (Figure 4.16), with an overall mean of 88 days/year. The number of zero flow days is lowest within the south-western parts of the study area due to high rainfall, as well as high BFI (Figure 4.14).

![Figure 4.16: Variation of the number of zero flow days within the study area](http://etd.uwc.ac.za/)

The number of zero flow days greatly increases within the central and south-eastern parts of the study area. Rainfall is highly sporadic within these parts of the study area, along with high evaporation rates. These parts of the study area are characterised by high aridity and consequently high number of zero flow days.

59
4.10.5 3-day minima and maxima of daily means of discharge

3-day min ranges between 0 and 29 mm/year, with an overall mean of 3.30 mm/year (Figure 4.17). High 3-day min is found within the south-western parts of the study area, which is expected due to high rainfall and BFI. Catchments in the central and south-eastern parts of the study area show the lowest 3-day min due to low rainfall, low BFI, and high rates of evaporation.

![Figure 4.17: Variation of 3-day minima and maxima of daily means of discharge within the study area](http://etd.uwc.ac.za/)

3-day max ranges between 62 and 12138 mm/year, with an overall mean of 2273 mm/year (Figure 4.17). Catchments that are characterised by high 3-day max are found within the south-western parts, whereas catchments in the central parts of the study area show low 3-day max.
4.11 Summary

This chapter has shown the variation of flow and catchment characteristics within the study area. Rainfall varies considerably within the study area from 80 to 3312 mm/year with an overall mean of 872 mm/year. Parts of the study area receiving high rainfall are typically associated with high slope. The relationship between rainfall and slope is expected since steep slopes occur in areas with high elevation, and which also receive high rainfall. Evaporation varies from 1370 to 2157 mm/year in the study area. High rainfall and low evaporation rates are typically found within the south-western, southern and south-eastern parts of the study area. The central regions of the study area are characterised by catchments with low rainfall and high evaporation rates, and consequently high aridity.

Mean annual runoff varies considerably from 1.8 to 498 mm/year within the study area. High runoff is found within the south-western parts of the study area, which is expected due to high rainfall and steep slopes, and low evaporation. Catchments within the south-western parts of the study area are also characterised by high flow percentiles (q_{10}, q_{25}, q_{75} and q_{90}), 3-day minima, and 3-day maxima of daily means of discharge. The central parts of the study area are characterised by high aridity, where evaporation rates exceed rainfall and consequently result in high number of zero flow days and low BFI. The coefficient of variation of annual flows differs considerably within the study area, with the south-western parts exhibiting low annual variation. Catchments found within the central and south-eastern parts of the study area have a high annual variability of flow, which is expected due to sporadic rainfall.
CHAPTER 5: ORDINATION

5.1 Introduction

Multivariate analysis methods such as ordination are suitable for determining the effects of different combinations of catchment characteristics on flow characteristics. These methods enable the explanation of the variance of flow characteristics by a set of catchment characteristics. The aim of this chapter is to identify catchment characteristics that are important in explaining the variance of flow characteristics. The identification of significant catchment characteristics explaining the variance of flow characteristic can be used as a basis for predicting flow characteristics and grouping catchments into homogenous groups that share similar hydrological responses.

5.2 Methodology

Ordination is a class of techniques for relating multiple response variables to explanatory variables (ter Braak & Šmilauer, 2018). In this study, ordination helps to relate flow and catchment characteristics and the ordination diagrams show how flow characteristics vary with catchment characteristics. Ordination techniques can be classified as unconstrained or canonical ordination (Borcard et al., 2011). Unconstrained ordination techniques analyse a set of variables and represent the major variation in these variables generated by a reduced set of orthogonal axes. Unconstrained ordination is considered as a passive form of analysis and the user interprets the results of the ordination. Unconstrained ordination techniques aim to explain the variance of variables, such as flow characteristics, by a relatively small number of components. The issue that arises with this approach is that the variation of flow characteristics is explained by unknown variables. This is a major weakness with unconstrained ordination, such as principle component analysis, which does not consider the explanatory variables such as catchment characteristics. Examples of unconstrained ordination techniques include principle component characteristics, correspondence analysis and principle coordinate analysis (Borcard et al., 2011).

Canonical or constrained ordination, however, analyses two or more sets of variables in the process (Borcard et al., 2011). Canonical ordination aims to identify the underlying structure
in a dataset by considering the relationship between response and explanatory variables. These techniques describe the variation of response variables that can be explained by explanatory variables (Koplin et al., 2012). Canonical ordination techniques are suitable for this study which aims to identify the explanatory variables that best explain the variance of response variables. Canonical ordination techniques can be classified as being symmetrical or asymmetrical. In symmetrical techniques, such as multiple factor analysis and canonical correlation analysis, the two data matrices play the same role in the analysis and can be interchanged (Legendre et al., 2010). Asymmetric analysis means that the two data sets that are used in the analysis do not play the same role; there is a matrix of response variables Y and a matrix of explanatory variables X, which is used to describe the variation in Y. Examples of asymmetrical techniques include redundancy analysis (RDA) and canonical correspondence analysis (CCA) (Borcard et al., 2011).

Asymmetrical techniques combine multiple regression and correlation to identify the relationship between response and explanatory variables (Borcard et al., 2011). The difference between CCA and RDA is that CCA assumes a unimodal relationship, whilst the latter assumes linear relationship between response and explanatory variables. CCA would therefore show increases in flow to a specific threshold, and thereafter the flow would decrease as catchment characteristics continue to increase. RDA is therefore suitable for the study, as the approach assumes flow increases with increasing precipitation and slope, and decreasing evaporation rates (Mazvimavi, 2003; Chikodzi, 2013). Mazvimavi (2003), Kaplin et al (2012) and Chikodzi (2013) used redundancy analysis to explain the variance of flow characteristics using a set of catchment characteristics. RDA was performed using CANOCO, which is a general package for multivariate data analysis, with an emphasis on dimension reduction (ordination) and regression analysis or the integrated combination of the two, called constrained ordination (ter Braak & Šmilauer, 2018).

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

$$y_{ik} = a_k + b_k x_i$$

(5.1)

where $a_k$ and $b_k$ are unknown regression coefficients.
Constrained ordination techniques aim to identify the underlying structure in a data set by considering the relationships between response and explanatory variables. In the case of RDA, $x_i$ is a linear combination of explanatory variables, $z_{ij}$, where $j = 1, 2, \ldots, n_p$ is the number of explanatory variables. $z_i$ refers to the standardized catchment characteristics. For example, if $p = 2$, $x_i$ is given by:

$$x_i = c_1 z_{i1} + c_2 z_{i2}$$  \hspace{1cm} (5.2)

Where, $c_1$ and $c_2$ are the weights, also called the canonical coefficients, to measure explanatory variables $z_{i1}$ and $z_{i2}$ to derive $x$, the theoretical explanatory variable (ter Braak & Šmilauer, 2018). Substituting for $x_i$ in Eq 5.1 gives the RDA model

$$y_{ik} = a_k + b_k c_1 z_{i1} + b_k c_2 z_{i2}$$  \hspace{1cm} (5.3)

RDA aims to estimate the unknowns in this model, $a_k$ and $b_k$ of the response variables and $c_1$ and $c_2$, the weights from the explanatory variables (ter Braak & Šmilauer, 2018).

CANOCO has the ability to statistically test the significance of the constrained ordination model, using Monte-Carlo permutation tests. CANOCO uses the false discovery rate, Holm's correction or Bonferroni correction. The Bonferroni correction method multiplies each $p$ value by the number of performed tests, which has often been identified as being too conservative and leads to large loss of power (Roback & Askins, 2005; Gordon et al., 2007). The Holm's correction is an improvement of the Bonferroni correction, in which once the first hypothesis has been rejected, the second hypothesis is treated as a completely new test. In Holm's correction the tests are first sorted on the basis of their $p$ values from smallest to largest. If the first hypothesis is rejected, we no longer deal with $q$ but rather $q-1$, where $q$ refers to the number of tests. The false discovery rate (FDR) is a popular approach for controlling Type 1 error, which is defined as the expected proportion of incorrectly rejected $H_0$ among all rejections. The FDR allows the occurrence of Type 1 error under a reasonable proportion by taking the number of rejections into consideration. The false discovery rate was therefore used in the study, due to the fact that the Bonferroni and Holm's corrections tend to be too conservative (ter Braak & Šmilauer, 2018). The permutation tests were used in the study to identify the catchment characteristics that are significant at a 5% level in explaining the variance of flow characteristics.
Table 5.1 presents the catchment characteristics selected for redundancy analysis that are likely to influence the flow characteristics presented in the same table. The flow characteristics were standardized by converting the initial flow data, given as m³/sec to mm/year. CV of annual flow, IC, CVB and BFI are given as a ratio and the number of zero flow days are represented as days.

Table 5.1: Flow and catchment characteristics selected for redundancy analysis

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Flow Characteristics</th>
<th>Catchment Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q</td>
<td>Mean annual runoff</td>
<td>Mean annual rainfall (MAP)</td>
</tr>
<tr>
<td>CV</td>
<td>Coefficient of variation of annual flows</td>
<td>Mean annual evaporation (ET)</td>
</tr>
<tr>
<td>q₉₀</td>
<td>Dimensionless daily flow with a 90 % exceedence</td>
<td>River length (RL)</td>
</tr>
<tr>
<td>q₇₅</td>
<td>Dimensionless daily flow with a 75 % exceedence</td>
<td>Drainage density (Dd)</td>
</tr>
<tr>
<td>q₂₅</td>
<td>Dimensionless daily flow with a 25 % exceedence</td>
<td>Slope (Sₐᵥₑ, S₂₀, S₅₀, S₉₀)</td>
</tr>
<tr>
<td>q₁₀</td>
<td>Dimensionless daily flow with a 10 % exceedence</td>
<td>Elevation (Eₘᵢₙ, Eₘₐₓ, Eₖₑₙₜₜₑ)</td>
</tr>
<tr>
<td>IC</td>
<td>Concavity Index</td>
<td>Proportion of each catchment under different lithologies (GL)</td>
</tr>
<tr>
<td>CVB</td>
<td>Hydrological Index</td>
<td>Proportion of each catchment under different land cover types (LC)</td>
</tr>
<tr>
<td>3-day min</td>
<td>3-day minimum of daily discharge</td>
<td>Proportion of each catchment under different soil textures (Ss)</td>
</tr>
<tr>
<td>3-day max</td>
<td>3-day maximum of daily discharge</td>
<td>Proportion of each catchment under different soil textures (Ss)</td>
</tr>
<tr>
<td>ZFD</td>
<td>Number of zero flow days</td>
<td>Proportion of each catchment under different soil textures (Ss)</td>
</tr>
<tr>
<td>BFI</td>
<td>Baseflow Index</td>
<td>Proportion of each catchment under different soil textures (Ss)</td>
</tr>
</tbody>
</table>
5.3 Results

5.3.1 Relationships between catchment characteristics

Figure 5.1 shows which catchment characteristics vary in a similar manner among the selected catchments in the study. Mean annual rainfall and slope tends to vary in a similar manner from one catchment to another, which suggests that catchments that have high rainfall tend to occur within catchments with steep slopes. The relationship between rainfall and slope is expected since steep slopes occur in areas with high elevation which also receive high rainfall (Jencso & McGlynn, 2011; Parajka et al., 2013; David & Davidova, 2014; Hallema et al., 2016). The variation of land cover types (thicket and plantations) and geology (Cape Granite and Table Mountain) with rainfall and slope suggests that these land cover types typically occur within catchments with high rainfall, steep slope and low evaporation (Figure 5.1).
Figure 5.1 shows that evaporation varies in an opposite manner to rainfall. This reveals that as rainfall increases among catchments evaporation rates reduce. River length (RL), area (A) and minimum elevation (E_{min}) show a similar variation with evaporation among catchments, which indicates that catchments with high evaporation rates tend to have increased river lengths, large catchment areas and low elevation. Large catchment areas tend to occur in low-lying areas with high temperatures, and therefore high evaporation values.

Bare ground (LC_{BG}), grasslands (LC_{G}), Karoo dolerite (GL_{KD}) and Beaufort (GL_{BF}) tend to vary in a similar manner with evaporation. This relationship indicates that catchments characterized by land cover types (LC_{BG} and LC_{G}) and geology (GL_{KD} and GL_{BF}) typically occur within catchments characterised by low rainfall and flat slopes, and consequently high evaporation rates.

5.3.2 Relationship between flow characteristics

Figure 5.2 shows which flow characteristics vary in a similar manner among the selected catchments. Mean annual runoff tends to vary in a similar manner with q_{10}, q_{25}, q_{75}, q_{90}, BFI, IC, 3-day min and 3-day max (Figure 5.2). The relationship between mean annual runoff and these flow characteristics suggests that catchments that have high runoff will also have large values of these flow characteristics.

The number of zero flow days varies in an opposite way to mean annual runoff and BFI (Figure 5.2). This suggests that catchments characterised by high runoff and BFI tend to have a low number of zero flow days. This relationship is expected as an increase in runoff and BFI reduces the number of zero flow days. Catchments that are characterized by high BFI therefore have low number of zero flow days. A large baseflow means flow throughout the year and therefore a reduction in the number of zero flow days.

CV of annual flows and CVB tend to vary in an opposite way to runoff, BFI and IC among the catchments. Mean annual runoff and BFI are inversely related to the coefficient of variation of annual flows. Catchments characterised by high BFI and runoff have a low CV of annual flows. The CVB represents the ratio of CV of annual flows to BFI. Catchments that are characterised by low CVB values therefore exhibit low inter-annual variability and large groundwater storages with higher baseflow contribution.
Figure 5.2: Variation of flow characteristics among selected catchments

The concavity index (IC) and BFI tend to vary in a similar manner among the catchments (Figure 5.2). The concavity index measures the contrast between high and low flow conditions, thus representing the shape of the flow duration curve (Bloschl et al., 2013). Catchments IC values close to 1 do not have a large difference between low and high flows. Catchments that are dominated by contrasting climates and poor groundwater storage typically have values close to 0, resulting in severe flow conditions from low-flow to quickflow due to rainfall events (Bloschl et al., 2013). The variation of BFI and IC is therefore expected, as catchments characterised by higher BFI tend to exhibit lower contrast between high and low-flow conditions and a flatter flow duration curve. Figure 5.2 illustrates that IC and BFI are inversely related to CV and CVB, which is expected as catchments characterised by higher IC and BFI show lower CV and CVB values.

5.3.3 Relationship between flow and catchment characteristics

Figure 5.3 shows the relationship between catchment characteristics with the ordination axes of flow characteristics. This identifies catchment characteristics that account for the variance of flow characteristics represented by each ordination axis. Figure 5.3 shows that MAP,
slope, GL\textsubscript{TM}, GL\textsubscript{CG}, LC\textsubscript{P} and LC\textsubscript{T} varies in a similar way with flow characteristics in the first axis amongst catchments. Slope and elevation have been found in previous studies to significantly influence mean annual runoff, through governing the movement of water to the catchment outlet (Mazvimavi, 2003; Masoudian, 2009; Jencso & McGlynn, 2011; Mirus & Loague, 2013).

GL\textsubscript{TM} varies in a similar manner with flow characteristics in the first axis. This suggests that the occurrence of GL\textsubscript{TM} among catchments tends to be associated with high runoff, flow percentiles, and BFI. Catchment geology has a substantial influence on the subsurface storage and drainage network (Jencso & McGlynn, 2011; Rumsey \textit{et al.}, 2015; Ries \textit{et al.}, 2017). GL\textsubscript{TM} consists of faults and fractures which promote the direct recharge of rainfall into the subsurface storage. Previous studies have shown baseflow contribution to be high within the Table Mountain Group (Sun, 2005 and Le Maitre & Colvin, 2008). The results of the ordination show that GL\textsubscript{TM} is typically found within catchments of high rainfall and steep slopes, which therefore results in high rates of recharge. As a result, the baseflow contribution within these regions would be high, which is evident as BFI shows a moderate correlation with GL\textsubscript{TM} (Figure 5.3). However, the relationship between the Table Mountain Group and flow characteristics in the first axis may merely be as a result that these rock types are associated with mountainous areas receiving high rainfall. GL\textsubscript{KD} tends to vary in an opposite way with BFI (Figure 5.3). Although previous studies have identified Karoo dolerite as having a high potential for groundwater storage (Molaba, 2017), the relationship was not identified in the study as Karoo dolerite typically underlies regions with low rainfall and flat slopes. This relationship is expected as Sun (2005) pointed out that low recharge rates coincide with low rainfall and high evaporation rates.

The variation of thicket, plantations and forest with flow characteristics in the first axis, such as runoff, is a result of these land cover types being mostly found in catchments of high rainfall and steep slopes, and subsequently high runoff. For example, plantations have been found to decrease mean annual runoff due to higher interception, infiltration and transpiration rates (Farley & Jobbágy, 2005; Zhao \textit{et al.}, 2015). The occurrence of plantations reduces net rainfall reaching the surface due to interception by tree canopy, which increases ET due to transpiration from vegetation. Plantations also generally have deep roots and infiltration is generally high under areas characterised by plantations. However, the degree to which plantations influence runoff is dependent on the characteristics of plants. For example, some

http://etd.uwc.ac.za/
trees may be characterised by dense tree canopy or shallow root systems which affects interception and infiltration of rainfall differently. Farley and Jobbágy (2005) and Zhao et al (2015) found that afforestation substantially influences catchment hydrology, through the reduction of runoff. The relationship between flow characteristics in the first axis and plantations is thus more related to the climatic and physiographic characteristics of areas covered by plantations. Plantations are typically found within the mountainous regions of the study area, with higher elevation, slope and MAP, which consequently increases runoff.

Figure 5.3: Relationship between flow and catchment characteristics among selected catchments

The occurrence of bare ground and grasslands, which are typically underlain by Karoo Dolerite and Beaufort Group (Figure 5.3), tends to vary in an opposite way with flow characteristics in the first axis among catchments. This relationship indicates that catchments covered by these land cover types are characterised by low MAP, flat slopes, high evaporation rates, and consequently low mean annual runoff. Previous studies have shown grasslands to influence mean annual runoff (Mazvimavi, 2003; Chikodzi, 2013; Duan et al., 2017). Grasslands have been identified in previous studies as generating runoff in the channel, due to higher infiltration rates and contribution to the river channel (Chikodzi, 2013). However, Mazvimavi (2003) and Duan et al (2017) identified that mean annual runoff decreases with the proportion of the catchment covered by grasslands, due to the presence of
wetlands which increases the rate of ET. This relationship is further explained with the number of zero flow days (ZFD) varying in a similar manner with land cover types of bare ground and grasslands. LC_G shows a negative correlation with flow characteristics in the first axis, including runoff, flow percentiles and BFI.

ZFD varies in a similar manner with ET, RL and A among catchments (Figure 5.3). The results suggest that an increase in evaporation, river length and area consequently increases the number of zero flow days among catchments. The reason for this phenomenon could be due to longer travel time of flow to the main channel, resulting in high infiltration and evaporation rates (Love et al., 2011; Chikodzi, 2013; Huang et al., 2017). Ries et al (2017) pointed out that runoff coefficients were observed to decrease with increasing river length due to increased infiltration rates. Ries et al (2017) also pointed out that increasing catchment area is often associated with decreasing runoff among catchments, as smaller catchments are generally found within mountainous areas with high runoff rates per unit area. Catchments with large areas are typically found within areas of gentle slopes, low rainfall and high evaporation in the study area (Figure 5.3), which subsequently result in longer travel times due to low flow velocities. Catchments characterised with small areas are typically found in mountainous areas, promoting runoff with high flow velocity, which reduces the rate of infiltration and travel time to the main channel. Love et al (2011) and Chikodzi (2013) showed similar results to this study where actual runoff was shown to decrease with increasing catchment area. Table 5.2 presents the total proportion of variance of the flow characteristics explained by the derived ordination axes of catchment characteristics.

Table 5.2: Proportion of variance of flow characteristics explained by catchment characteristics in the derived ordination axes

<table>
<thead>
<tr>
<th></th>
<th>Axis 1</th>
<th>Axis 2</th>
<th>Axis 3</th>
<th>Axis 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigenvalues</td>
<td>58.20</td>
<td>15.40</td>
<td>5.80</td>
<td>2.30</td>
</tr>
<tr>
<td>Explained variation</td>
<td>58.20</td>
<td>73.55</td>
<td>79.32</td>
<td>81.65</td>
</tr>
<tr>
<td>Pseudo-canonical</td>
<td>0.97</td>
<td>0.89</td>
<td>0.91</td>
<td>0.81</td>
</tr>
<tr>
<td>Explained fitted</td>
<td>67.84</td>
<td>85.73</td>
<td>92.46</td>
<td>95.17</td>
</tr>
</tbody>
</table>

http://etd.uwc.ac.za/
The first axis explains 58.2% of the variance of flow characteristics, 15.4% by the second axis, 5.8% by the third axis and 2.3% by the fourth axis. The total variance explained is represented by the explained variation (Table 5.2) with the total variance of flow characteristics explained by 81.65%. The pseudo-canonical correlation in Table 5.2 describes the linear correlation between the case scores of the axis derived from flow and the corresponding case scores derived from the catchment characteristics (ter Braak & Šmilauer, 2018). The first flow characteristics axis has a high correlation coefficient of 0.97 with the first axis of catchment characteristics. A high correlation coefficient indicates a strong linear relationship with the derived catchment characteristics. However, this does not necessarily mean that the catchment characteristics in the derived axis explain the variance of the flow characteristics. It may merely mean that the catchment characteristics and flow characteristics are closely correlated and vary in a similar way. The “explained fitted variation (cumulative)” in Table 5.2 shows the contribution of each catchment characteristics axis compared to the total variance of flow characteristics explained by all the catchment characteristics axes, with the first axis of catchment characteristics explaining 67.84% of the variance of flow characteristics (ter Braak & Šmilauer, 2018).

Monte Carlo permutation test was used to determine those catchment characteristics that are significant at a 5% level explaining the variance of flow characteristics. Table 5.3 presents the catchment characteristics that were found to be significant at a 5% level in explaining the variance of flow characteristics and identifies the amount of variance explained for each catchment characteristic. Mean annual rainfall and slope equalled or exceeded 90% of the time ($S_{90}$) were the only catchment characteristics that were found to significantly explain the variance of flow characteristics. The results reveal that the cumulative percentage explained by MAP and $S_{90}$ is 57%.

Table 5.3: Proportion of variance of flow characteristics explained by catchment characteristics

<table>
<thead>
<tr>
<th>Catchment Characteristics</th>
<th>Percentage Explained (%)</th>
<th>Cumulative Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAP (mm)</td>
<td>46</td>
<td>46</td>
</tr>
<tr>
<td>$S_{90}$ (degrees)</td>
<td>11</td>
<td>57</td>
</tr>
</tbody>
</table>
5.4 Discussion and conclusion

Redundancy analysis has shown that MAP and $S_{90}$ are the only catchment characteristics that are significant in explaining the variance of flow characteristics in this study. The importance of rainfall in explaining the variance of flow characteristics has been shown in several studies and has been identified as the main climatic forcing of the hydrological catchment response (Mazvimavi, 2003; Mazvimavi et al., 2005; Chikodzi, 2013; Blöschl et al., 2013; Drogue & Khediri, 2016).

Slope is an important characteristic of a catchment as it explains the rate of movement of water by kinetic energy to the catchment outlet (Mazvimavi, 2003; Masoudian, 2009; Mirus & Loague, 2013; David & Davidova, 2014). The slope influences the catchment response time to rainfall events, which has a substantial effect on the concentration and travel times of flow (Masoudian, 2009; Gericke & du Plessis, 2012; David & Davidova, 2014; Huang et al., 2017). David and Davidova (2014) showed that high slope led to fast concentrations and consequently to high volumes of peak discharge. Mu et al (2015) and Malan (2016) identified that soil depth decreases with increasing slope, which would consequently reduce rates of infiltration and enhance the generation of runoff.

The results of redundancy analysis show that 18% of the variation of flow characteristics cannot be accounted for by the catchment characteristics used in the study. The unexplained variation of 18% may be due to random hydrological behaviour which cannot be distinguished, or due to the limitation of the data sources used in the study. The use of more hydrologically-meaningful data for geological and soil data sets, such as depth to the water table, permeability and porosity may increase the variance explained for the flow characteristics used in the study. Such data, in developing countries especially, are not readily available and the data provided is often a generalised coverage of various geological characteristics (Mazvimavi, 2003; Hughes, 2006). Hughes (2006) expressed that in many developing countries; the derivation of geological characteristics is based on coarse geological maps with low resolution that have not been generated using hydrologically meaningful source data. This limitation may urge the development of more hydrologically meaningful data in developing countries, which would be more suitable for hydrological studies.
Catchment characteristics that have been identified in this chapter as significant in explaining the variance of flow characteristics can be used as a basis to group catchments into hydrologically homogenous groups. MAP and $S_{90}$ were found to significantly explain the variance of flow characteristics. Although other catchment characteristics are important, such as the Table Mountain Group and ET, they are not significant at explaining the variance of flow characteristics. The results of ordination were also used as the basis for cluster analysis for grouping catchments into hydrologically similar groups based on similarities of catchment characteristics in Chapter 7.
CHAPTER 6: PREDICTION OF FLOW CHARACTERISTICS

6.1 Introduction

The aim of this chapter is to explore the model performance of predicting flow characteristics through the use of univariate statistical methods and artificial neural networks. The development of regression models and neural networks provides the opportunity to predict flow characteristics in ungauged catchments and determines which approach is more suitable for predicting selected flow characteristics of non-perennial rivers.

6.2 Methodology

The study used multiple linear regression (MLR) and artificial neural networks (ANN) to predict flow characteristics from catchment attributes (Mazvimavi, 2003; Riad et al., 2004; Mazvimavi et al., 2005; Aichouri et al., 2015).

6.2.1 Multiple linear regression

Multiple linear regressions is an extension of simple linear regression, in which multiple explanatory variables are used to establish the relation between the response and explanatory variables, which can be expressed as (Patel et al., 2016):

\[ Y = \alpha + \beta_1 X_1 + \beta_{i+1} X_{i+1} + \ldots + \beta_p X_p + e \]  

where; \( Y \) = dependent variable, \( \alpha \) = constant or intercept, \( \beta_i \) = slope (beta coefficient), \( X_i \) = independent variables, \( i = 1, 2, 3, \ldots, p \), and \( p \) = number of independent variables.

Multiple regression methods are only applicable if there is a linear relationship between flow and catchment characteristics, and the variables approximate a normal distribution (Mazvimavi, 2003). The study used forward selection to produce regression models of flow characteristics, which was based on data from 36 gauging stations. The forward selection approach starts with no independent variables in the model. At the first step, the independent variable that shows the highest coefficient of determination is selected in the model. At each
step, the independent variable that increases the coefficient of determination of the model is selected. The process is terminated once the remaining independent variables are not significant. Once a variable has been selected in the model, the variable cannot be removed. The selection of independent variables for the prediction of flow characteristics was based on previous studies, which identified influential catchment characteristics governing hydrological response.

Numerous techniques have been developed to predict flow duration curves (FDCs) in ungauged catchments. The use of exponential (Eq 6.2), logarithmic (Eq 6.3), and power models (Eq 6.4) tend to be popular for predicting FDCs in ungauged catchments (Mazvimavi, 2003; Mazvimavi et al., 2005; Sauquet & Catalogne, 2011). The advantage of these approaches is the reduction of computational effort for regionalisation of FDCs in ungauged catchments.

\[
q_p(i) = \beta(i) \exp(-\alpha p) \tag{6.2}
\]

\[
q_p(i) = \beta(i) + \alpha(i) \ln(p) \tag{6.3}
\]

\[
q_p(i) = \beta(i) p^{\alpha(i)} \tag{6.4}
\]

where \(q_p\) is dimensionless daily flows, \(p\) is the percentage exceedence and \(\alpha\) and \(\beta\) are coefficients.

6.2.2 Artificial neural networks

ANN’s have been widely used due to the ability to model both linear and non-linear relationships (Riad et al., 2004; Aichori et al., 2015). Generally there are four steps that are required for developing an ANN (Londhe & Charhate, 2010). Firstly, the data needs to be transformed or scaled. Large variation in input data can slow down or prevent the training of the network, thus data is scaled using linear, logarithmic or normal transformation. The second step is the network architecture definition, where the number of layers is set (Londhe & Charhate, 2010). The ANN consists of three layers (Figure 6.1), the input, hidden and output layer, with each layer comprising of a series of nodes that are interconnected (Londhe & Charhate, 2010; Elsafi, 2014; Aichouri et al., 2015). Figure 6.1 illustrates the network architecture of a simple multi-layer perceptron neural network.
The number of nodes that is required in the input and hidden layer depends on the complexity of the problem being studied. For example, if the number of nodes used in the hidden layer is too small, the network may be insufficient to characterise the process correctly. On the other hand, if the number of nodes is too high, the training of the ANN may take too long and the network may sometimes over-fit the data. The next step involves the training of the ANN, through determining the weights of the ANN, which forms the connection between the neurons. The ANN is trained with a set of input and known output data (Londhe & Charhate, 2010; Elsafi, 2014). At the beginning, the weights of the neurons may be assigned randomly or based on experience. The learning algorithm of the ANN changes the weights of the neurons accordingly, so that the difference between the ANN output and the actual value is small. Once the difference between the ANN output and the actual value is within a specified range, the ANN can be considered as trained.

The last process is the validation of the performance of the ANN. Training of the ANN was based on using a sub-sample of catchments. The validation of the ANN is based on using a sub-sample that was not used during training to evaluate the performance of the ANN. The validation process determines whether the ANN needs to be re-trained or if the network can be applied for the intended use (Riad et al., 2004; Londhe & Charhate, 2010; Aichouri et al., 2015). The model performance for predicting flow characteristics can be determined by a range of statistical criteria, including root mean square error (RMSE), coefficient of correlation (r), and coefficient of determination ($R^2$).
6.2.3 Model performance evaluation

The model performance of both multiple regression and neural networks were evaluated using the coefficient of determination ($R^2$), root mean square error (RMSE) and the standard deviation ratio (RSR). The coefficient of determination (Eq 6.2), which ranges between 0 and 1, describes the proportion of variance in the observed data, which is explained by the model, with higher values indicating a better model performance (Golmohammadi et al., 2014). Previous studies have identified that the model performance of $R^2$ greater than 0.50 can be considered as being acceptable (Golmohammadi et al., 2014). The RMSE (Eq 6.3) indicates a perfect match between the observed and predicted data when the value is 0, with increasing RMSE increasing the error in prediction. Singh et al (2004) identified that when the RMSE is less than half the value of the observed standard deviation of the population, the model can be considered low and indicating a good model performance for prediction. The RSR (Eq 6.4) is calculated based on the ratio of the RMSE and standard deviation of the observed data, where the lower the value the more reliable the performance of the model (Golmohammadi et al., 2014). The equations for $R^2$, RMSE and RSR are presented below (Golmohammadi et al., 2014):

$$R^2 = \frac{\sum_{i=1}^{n} (O_i - O)(P_i - \bar{P})}{\sqrt{\sum_{i=1}^{n} (O_i - O)^2} \sqrt{\sum_{i=1}^{n} (P_i - \bar{P})^2}}$$ \hspace{1cm} (6.5)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (P_i - O_i)^2}{n}}$$ \hspace{1cm} (6.6)

$$RSR = \frac{\sum_{i=0}^{n} (O_i - P_i)^2}{\sqrt{\sum_{i=0}^{n} (O_i - \bar{O})^2}}$$ \hspace{1cm} (6.7)

6.3 Correlation between flow and catchment characteristics

Table 6.1 presents the correlation between flow and catchment characteristics. Catchment characteristics are only presented if they show a significant correlation coefficient at 5% confidence level with flow characteristics.
Table 6.1: Correlation between flow and catchment characteristics

<table>
<thead>
<tr>
<th></th>
<th>Q</th>
<th>CV</th>
<th>Q90</th>
<th>Q75</th>
<th>Q25</th>
<th>Q10</th>
<th>IC</th>
<th>3-day min</th>
<th>3-day max</th>
<th>ZFD</th>
<th>BFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAP</td>
<td>0.73</td>
<td>-0.59</td>
<td>0.54</td>
<td>0.59</td>
<td>0.74</td>
<td>0.75</td>
<td>0.61</td>
<td>0.49</td>
<td>0.63</td>
<td>-0.69</td>
<td>0.55</td>
</tr>
<tr>
<td>ET</td>
<td>-0.27</td>
<td>0.03</td>
<td>-0.21</td>
<td>-0.21</td>
<td>-0.21</td>
<td>-0.22</td>
<td>-0.14</td>
<td>-0.21</td>
<td>-0.36</td>
<td>0.23</td>
<td>-0.03</td>
</tr>
<tr>
<td>Dd</td>
<td>0.03</td>
<td>0.25</td>
<td>0.32</td>
<td>0.24</td>
<td>-0.06</td>
<td>-0.07</td>
<td>-0.17</td>
<td>0.39</td>
<td>0.17</td>
<td>-0.11</td>
<td>0.13</td>
</tr>
<tr>
<td>Emin</td>
<td>-0.34</td>
<td>0.08</td>
<td>-0.21</td>
<td>-0.24</td>
<td>-0.32</td>
<td>-0.31</td>
<td>-0.13</td>
<td>-0.19</td>
<td>-0.36</td>
<td>0.18</td>
<td>-0.27</td>
</tr>
<tr>
<td>Erange</td>
<td>0.30</td>
<td>-0.33</td>
<td>0.12</td>
<td>0.19</td>
<td>0.38</td>
<td>0.38</td>
<td>0.34</td>
<td>0.10</td>
<td>0.17</td>
<td>-0.33</td>
<td>0.34</td>
</tr>
<tr>
<td>Save</td>
<td>0.45</td>
<td>-0.30</td>
<td>0.57</td>
<td>0.63</td>
<td>0.37</td>
<td>0.35</td>
<td>0.22</td>
<td>0.59</td>
<td>0.50</td>
<td>-0.44</td>
<td>0.56</td>
</tr>
<tr>
<td>S20</td>
<td>0.51</td>
<td>-0.37</td>
<td>0.47</td>
<td>0.54</td>
<td>0.49</td>
<td>0.48</td>
<td>0.32</td>
<td>0.46</td>
<td>0.49</td>
<td>-0.47</td>
<td>0.53</td>
</tr>
<tr>
<td>S50</td>
<td>0.38</td>
<td>-0.22</td>
<td>0.57</td>
<td>0.61</td>
<td>0.26</td>
<td>0.25</td>
<td>0.13</td>
<td>0.61</td>
<td>0.48</td>
<td>-0.38</td>
<td>0.51</td>
</tr>
<tr>
<td>S90</td>
<td>0.39</td>
<td>-0.21</td>
<td>0.63</td>
<td>0.67</td>
<td>0.28</td>
<td>0.24</td>
<td>0.09</td>
<td>0.67</td>
<td>0.50</td>
<td>-0.36</td>
<td>0.50</td>
</tr>
<tr>
<td>LCS</td>
<td>-0.13</td>
<td>0.16</td>
<td>-0.18</td>
<td>-0.16</td>
<td>-0.07</td>
<td>-0.07</td>
<td>-0.08</td>
<td>-0.18</td>
<td>-0.20</td>
<td>0.21</td>
<td>0.07</td>
</tr>
<tr>
<td>LCG</td>
<td>-0.35</td>
<td>0.04</td>
<td>-0.26</td>
<td>-0.30</td>
<td>-0.35</td>
<td>-0.34</td>
<td>-0.02</td>
<td>-0.25</td>
<td>-0.33</td>
<td>0.02</td>
<td>-0.41</td>
</tr>
<tr>
<td>LCBG</td>
<td>-0.17</td>
<td>0.56</td>
<td>-0.08</td>
<td>-0.08</td>
<td>-0.16</td>
<td>-0.18</td>
<td>-0.34</td>
<td>-0.08</td>
<td>-0.16</td>
<td>0.43</td>
<td>-0.35</td>
</tr>
<tr>
<td>LCCL</td>
<td>0.05</td>
<td>-0.30</td>
<td>-0.20</td>
<td>-0.17</td>
<td>0.11</td>
<td>0.13</td>
<td>0.21</td>
<td>-0.24</td>
<td>-0.04</td>
<td>0.01</td>
<td>0.17</td>
</tr>
<tr>
<td>LCT</td>
<td>0.34</td>
<td>0.03</td>
<td>0.47</td>
<td>0.38</td>
<td>0.18</td>
<td>0.23</td>
<td>0.03</td>
<td>0.49</td>
<td>0.47</td>
<td>-0.29</td>
<td>0.06</td>
</tr>
<tr>
<td>LCP</td>
<td>0.36</td>
<td>-0.12</td>
<td>0.45</td>
<td>0.56</td>
<td>0.35</td>
<td>0.25</td>
<td>0.03</td>
<td>0.47</td>
<td>0.44</td>
<td>-0.21</td>
<td>0.28</td>
</tr>
<tr>
<td>GLBF</td>
<td>-0.41</td>
<td>0.24</td>
<td>-0.26</td>
<td>-0.32</td>
<td>-0.42</td>
<td>-0.42</td>
<td>-0.23</td>
<td>-0.25</td>
<td>-0.37</td>
<td>0.19</td>
<td>-0.50</td>
</tr>
<tr>
<td>GLBV</td>
<td>-0.10</td>
<td>0.13</td>
<td>-0.21</td>
<td>-0.17</td>
<td>-0.06</td>
<td>-0.05</td>
<td>-0.05</td>
<td>-0.20</td>
<td>-0.14</td>
<td>0.19</td>
<td>0.03</td>
</tr>
<tr>
<td>GLKD</td>
<td>-0.34</td>
<td>0.24</td>
<td>-0.23</td>
<td>-0.28</td>
<td>-0.36</td>
<td>-0.36</td>
<td>-0.31</td>
<td>-0.22</td>
<td>-0.30</td>
<td>0.25</td>
<td>-0.46</td>
</tr>
<tr>
<td>GLTM</td>
<td>0.47</td>
<td>-0.25</td>
<td>0.56</td>
<td>0.60</td>
<td>0.40</td>
<td>0.39</td>
<td>0.15</td>
<td>0.59</td>
<td>0.51</td>
<td>-0.38</td>
<td>0.56</td>
</tr>
<tr>
<td>GLCG</td>
<td>0.40</td>
<td>-0.21</td>
<td>0.21</td>
<td>0.21</td>
<td>0.37</td>
<td>0.40</td>
<td>0.21</td>
<td>0.15</td>
<td>0.37</td>
<td>-0.21</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Note: Bold values indicate correlation coefficients that are significant at the 5 % level

6.4 Mean annual runoff

Table 6.1 shows a strong correlation between mean annual runoff (Q) and mean annual rainfall (MAP) \((r = 0.73)\), which indicates an increase in runoff with increasing rainfall. ET shows a negative correlation with Q \((r = -0.27)\).

Mean annual runoff shows a moderate correlation with \(S_{ave}, S_{20}, S_{50}, \) and \(S_{90}\) (Table 6.1), which is expected as Chapter 5 identified that high runoff tends to occur within catchments of
steep slopes and high rainfall. Soil depth has also been found to decrease with increasing elevation, which consequently reduces infiltration in headwater regions and promotes runoff (Mu et al., 2015; Malan, 2016). The correlation between runoff and slope decreases with decreasing slope (Table 6.1). $S_{20}$ shows a correlation of 0.51, whereas $S_{90}$ shows a correlation of 0.39 with runoff.

Q shows a moderate correlation with GL$_{TM}$ ($r = 0.47$) (Table 6.1). A moderate correlation between runoff and the Table Mountain Group is expected as Chapter 5 identified that catchments underlain by the Table Mountain Group are typically found within parts of the study area characterised by high rainfall and steep slopes, and consequently high runoff. GL$_{KD}$ shows a negative correlation with Q ($r = -0.34$) (Table 6.1), which is a result of the lithology being found within parts of the study area characterised by low rainfall and slope, and consequently low runoff. Plantations (LC$_P$) also show a moderate correlation with runoff (Table 6.1). Chapter 5 identified that plantations are typically found within mountainous parts of the study area, with high elevation, steep slopes and high rainfall, and consequently high runoff. Grasslands (LC$_G$) show a negative correlation with runoff (Table 6.1). Grasslands are typically found within parts of the study area characterised by low elevation, high evaporation and low rainfall, which was identified in Chapter 5.

The predictive equation for estimating mean annual runoff based on catchment attributes using step-wise regression and the model performance is presented in Figure 6.2. The observed results show a good model performance based on the $R^2$ (0.73), % RMSE (60) and RSR (0.51). The results suggest that mean annual runoff can be predicted with a good model accuracy and low predictive error based on guidelines of previous studies (Moriasi et al., 2007).

Neural networks were used to compare the model performance for predicting mean annual runoff with multiple regression. Table 6.2 presents the neural networks that were selected for predicting mean annual runoff. The neural networks were selected on the basis of the $R^2$, % RMSE and RSR of the model. Table 6.2 also presents the characteristics of the neural networks, including the type of neural network, and the number of input and hidden layers used to predict the output variable. Neural networks have the ability to model non-linear relationships between flow and catchment characteristics, and the inclusion of $S_{20}$, $S_{90}$ and GL$_{TM}$ in the neural networks (Table 6.2) suggests that these catchment characteristics are non-linearly related to Q, which cannot be identified using multiple linear regression.
Table 6.2: Neural networks used for predicting mean annual runoff from catchment characteristics

<table>
<thead>
<tr>
<th>Type of network</th>
<th>Number of units in the hidden layer</th>
<th>Catchment characteristics</th>
<th>$R^2$</th>
<th>% RMSE</th>
<th>RSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLP 7-9-1</td>
<td>9</td>
<td>$S_{20}$, $S_{90}$, MAP, $GL_{TM}$, ET, LC, BFI</td>
<td>0.72</td>
<td>66.16</td>
<td>0.56</td>
</tr>
<tr>
<td>Linear 6-1</td>
<td>None</td>
<td>MAP, LC, $S_{20}$, $GL_{TM}$, $S_{90}$, ET</td>
<td>0.64</td>
<td>70.22</td>
<td>0.60</td>
</tr>
</tbody>
</table>

Notes:

1. MLP 7-9-1: multi-layer perceptron with 7 input variables, 9 units in the hidden, and 1 output unit
2. Linear 6-1: Linear model with 6 input variables and 1 output unit

MLR and MLP 7-9-1 demonstrate a good model performance for predicting mean annual runoff (Figure 6.2). These models have a low RMSE and RSR, which suggest a low predictive error for predicting mean annual runoff.

Figure 6.2: Comparison between the performance of multiple regression and an MLP 7-9-1 artificial neural network in predicting mean annual runoff
The use of multiple regression and MLP 7-9-1 are suitable for predicting mean annual runoff of non-perennial rivers based on their model performance. Although multiple regression is recommended as the method is simpler and has a lower predictive error compared to the MLP 7-9-1. The use of multiple regression was also shown to be suitable for predicting runoff in previous studies (Mazvimavi, 2003; Mazvimavi et al, 2005; Lacombe et al, 2014). The model for predicting mean annual runoff using multiple regression can be recommended for future studies.

6.5 Flow duration curve

An exponential model (Eq. 6.2) was found to be suitable for modelling the relationship between dimensionless daily flows and the exceedence probability of non-perennial rivers. The α coefficient in Eq. 6.2 can be related to BFI, with the following equation being derived:

\[ a = 32.96 \exp(-6.07 \text{BFI}) \quad R^2 = 0.73 \]  

\[(6.8)\]

The coefficient, β, is not related to any catchment attribute used in the study. The exponential model equation for predicting FDCs in ungauged basins therefore becomes:

\[ q_p = \beta \exp\left\{ -[32.96 \exp(-6.07 \text{BFI})p] \right\} \]  

\[(6.9)\]

The prediction of flow percentiles using an exponential model shows a good model performance based on the \( R^2 \) ranging between 0.99 to 0.61 and RSR ranging between 0.66 to 0.1 (Figure 6.3). Predicting high flow percentiles, such as \( q_{10} \) to \( q_{60} \), shows a high model performance where the RSR value is less than 0.5 (Moriasi et al, 2007). Although the prediction of low flow percentiles, such as \( q_{90} \) and \( q_{80} \), shows a poor performance in comparison to high flow prediction (Figure 6.3), the accuracy of predicting low flow percentiles can be considered acceptable as \( R^2 \) is greater 0.6 and RSR is less than 0.66 (Moriasi et al, 2007). This trend is expected as the accuracy for predicting flow percentiles has been shown to decrease from high to low flow percentiles (Hope & Bart, 2012).
The possibility of using neural networks to predict flow percentiles from catchment attributes was explored. The most suitable prediction of flow percentiles was made by a multi-layer perceptron with input variables including MAP, $S_{20}$, $S_{50}$, $S_{90}$, ET, Dd and BFI. Prediction of flow percentiles using neural networks show opposing trends compared to the exponential model, where low flow percentiles, $q_{70}$ to $q_{90}$, are predicted with higher model accuracy compared to high flow percentiles, such as $q_{10}$ (Figure 6.3). This suggest that neural networks have a greater ability to predict low flows, $q_{90}$, $q_{80}$ and $q_{70}$, from catchment attributes than flood events, $q_{10}$. Although the prediction of high flows has a considerable lower predictive accuracy compared to low flows, these models can be regarded as acceptable, as the $R^2$ and RSR value is within a specific range of acceptance.

Figure 6.4 compares the flow duration curve derived using observed flows with those predicted using a neural network and exponential model. For the catchments presented, the prediction of the flow duration curve using the exponential model has a lower predictive error in comparison to the use of neural networks (Figure 6.4). The results show that the prediction of high flow percentiles, $q_{10}$ and $q_{20}$, using neural networks has a higher RMSE in comparison to the exponential model.
The results suggest that the prediction of flow duration curves of non-perennial rivers in ungauged catchments reveal a good model performance using an exponential model. The prediction of flow duration curves in ungauged catchments may be improved by increasing the sample size of catchments used in the study, which may improve in deriving the β coefficient. The study used 36 gauged non-perennial rivers which may not have been sufficient to derive the β coefficient using catchment attributes. In the case of this study, the β coefficient was shown to have no relationship with catchment attributes. An alternative approach could be recommended by estimating the β coefficient as a cluster average through clustering catchments into homogenous groups. We could assume that catchments belonging to the same cluster have similar β coefficients.
6.6 Baseflow Index

The use of multiple regression for predicting BFI shows a good model performance based on $R^2$, RSR and RMSE (Figure 6.5). This model can be considered satisfactory as the $R^2$ and RSR is within a specific range (Moriaisi et al, 2007). Mazvimavi et al (2005) predicted BFI with a higher coefficient of determination ($R^2 = 0.75$) compared to model performance of this study ($R^2 = 0.58$). This may be due to the fact that the correlation between BFI with mean annual rainfall and slope were not found to be as significant in this study (Table 6.1). This study observed a correlation coefficient of $(r = 0.55)$ between BFI and rainfall and $(r = 0.51)$ between BFI and $S_{50}$, whereas Mazvimvi et al (2005) observed a correlation coefficient of $(r = 0.60)$ between BFI and rainfall and $(r = 0.74)$ between BFI and $S_{50}$. Rainfall tends to be high in mountainous areas; however, there are parts of the study area that are characterised by arid and semi-arid conditions with high elevation, steep slopes and low rainfall. This would therefore decrease the correlation between rainfall and slope with BFI. Previous studies have shown a strong correlation between BFI with rainfall and slope (Mazvimavi, 2003; Rumsey et al, 2015), which observed an increase in BFI with increasing rainfall and slope.

The neural networks that were recognised as most promising for predicting BFI based on the statistical performance indices used in the study are presented in Table 6.3. The results indicate that the use of neural networks shows a poor model performance as the $R^2$ value is less than 0.5 and the RSR value is in the range that is considered as an unsatisfactory model (Moriaisi et al, 2007).

Table 6.3: Neural networks used for predicting baseflow index (BFI) based on catchment characteristics

<table>
<thead>
<tr>
<th>Type of Network</th>
<th>Number of units in the hidden layer</th>
<th>Catchment characteristics</th>
<th>$R^2$</th>
<th>% RMSE</th>
<th>RSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear 3-1</td>
<td>None</td>
<td>MAP, LC&lt;sub&gt;G&lt;/sub&gt;, GL&lt;sub&gt;Tm&lt;/sub&gt;</td>
<td>0.42</td>
<td>31.04</td>
<td>0.75</td>
</tr>
<tr>
<td>MLP 5-4-1</td>
<td>4</td>
<td>MAP, GL&lt;sub&gt;Tm&lt;/sub&gt;, LC&lt;sub&gt;G&lt;/sub&gt;, S&lt;sub&gt;50&lt;/sub&gt;, LC&lt;sub&gt;P&lt;/sub&gt;</td>
<td>0.40</td>
<td>37.37</td>
<td>0.91</td>
</tr>
</tbody>
</table>

Notes:

1. Linear 3-1: Linear model with 3 input variables and 1 output unit
2. MLP 5-4-1: multi-layer perceptron with 5 input variables, 4 units in the hidden, and 1 output unit

The use of a Linear 3-1 neural network has a higher predictive error for predicting BFI compared to the use multiple regression (Figure 6.5). Multiple regression is therefore more suitable for predicting baseflow index of non-perennial rivers compared to neural networks due to higher model performance.

![Figure 6.5: Comparison between the performance of multiple regression and a Linear 3-1 neural network for predicting baseflow index](https://etd.uwc.ac.za/)

The quantification of hydrogeological characteristics in terms of the depth to the water table, permeability, and porosity may improve the prediction performance of BFI. The proportion of catchments underlain by different lithology types may not have been sensitive in describing the variation of BFI in the study, which was shown to have to have a satisfactory model performance. Such catchment attributes are important in controlling the water storage and groundwater flow routing in the catchment, which are essential in explaining low flows (Abebe & Foerch, 2006; Cervi et al, 2017). Studies have suggested that characterising the
fracture behaviour and geometry of geological units within the catchment may improve the prediction of low flows (Cervi et al., 2017).

6.7 Number of zero flow days

The prediction of the number of zero flow days using multiple regression shows a satisfactory model performance based on the $R^2$ and RSR value (Figure 6.6). Table 6.4 presents the neural networks that were found to be most suitable for predicting number of zero flow days. The results show that the MLP 5-6-1 neural network can be considered as a satisfactory performance for predicting the number of zero flow days, based on the $R^2$ and RSR.

Table 6.3: Neural networks used for predicting number of zero flow days (ZFD) from catchment attributes

<table>
<thead>
<tr>
<th>Type of Network</th>
<th>Number of units in the hidden layer</th>
<th>Catchment attributes</th>
<th>$R^2$</th>
<th>% RMSE</th>
<th>RSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLP 5-6-1</td>
<td>6</td>
<td>BFI, MAP, $E_{min}$, ET, Dd</td>
<td>0.58</td>
<td>72.70</td>
<td>0.67</td>
</tr>
<tr>
<td>MLP 5-7-1</td>
<td>7</td>
<td>MAP, BFI, $E_{min}$, Dd, ET</td>
<td>0.56</td>
<td>71.59</td>
<td>0.66</td>
</tr>
</tbody>
</table>

Notes:

1. MLP 5-6-1: multi-layer perceptron with 5 input variables, 6 units in the hidden layer, and 1 output unit
2. MLP 5-7-1: multi-layer perceptron with 5 input variables, 7 units in the hidden layer, and 1 output unit

MLR and MLP 5-6-1 underestimates catchments with high number of zero flow days, as well as over-estimating for catchments with low number of zero flow days (Figure 6.6). Although there are several catchments where the number of zero flow days is under and over-estimated, these models can be considered as satisfactory in terms of the model performance.
The use of multiple regression would be considered as more suitable for prediction of the number of zero flow days of non-perennial rivers, as the approach has a higher model performance based on the coefficient of determination, RMSE and RSR compared to the use of neural networks.

6.8 3-day minima and 3-day maxima of daily flows

Table 6.4 presents the model performance for predicting 3-day min and 3-day max using multiple regression. The results indicate the prediction of 3-day min shows a higher model performance compared to predicting 3-day max. The prediction of 3-day min shows a moderate $R^2$; however, the results indicate a high % RMSE.
Table 6.4: Predictive equations for predicting 3-day minima and maxima of daily means of discharge

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Predictive equation</th>
<th>$R^2$</th>
<th>% RSME</th>
<th>RSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-day min</td>
<td>$1.46S_{90} + 0.02\text{MAP} - 0.41S_{20} - 0.10S_S$</td>
<td>0.65</td>
<td>137.90</td>
<td>0.58</td>
</tr>
<tr>
<td>3-day max</td>
<td>$5.27\text{MAP} + 20.86GL_{TM}$</td>
<td>0.58</td>
<td>0.31</td>
<td>0.64</td>
</tr>
</tbody>
</table>

Table 6.5: Predicting 3-day minima and maxima of daily means of discharge using neural networks

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>$R^2$</th>
<th>% RMSE</th>
<th>RSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-day min</td>
<td>MLP 4-8-1</td>
<td>0.80</td>
<td>113.90</td>
</tr>
<tr>
<td>3-day max</td>
<td>MLP 4-8-1</td>
<td>0.48</td>
<td>80.42</td>
</tr>
</tbody>
</table>

The results suggest that the use of neural networks seems to be more suitable for the prediction of 3-day min due to a higher coefficient of determination. The results suggest that the margin of error for predicting 3-day min is high based on the % RMSE. Multiple regression is more suitable for the prediction of 3-day max than the use of neural networks due to higher model performance based on the coefficient of determination and lower % RMSE.
6.9 Discussion and conclusion

In this chapter, two data driven modelling techniques, namely multiple linear regression (MLR) and artificial neural networks (ANNs) were used for the prediction of flow characteristics of non-perennial rivers. The aim of the chapter was to identify which approach would be more suitable for predicting selected flow characteristics of non-perennial rivers. Overall MLR seems to be more suitable for predicting flow characteristics of non-perennial rivers compared to the use of neural networks. The chapter found that the prediction of mean annual runoff, BFI, number of zero flow days and 3-day max show a good model performance using multiple regression and these equations can be recommended for future studies. Mazvimavi (2003), Mazvimavi et al (2005) and Lacombe et al (2014) showed that multiple regression models perform well for predicting discharge of flow, which was also shown in this study.

An exponential model was found to be a promising approach for predicting the flow duration curve of non-perennial rivers with a lower predictive error than neural networks. The prediction of high flow percentiles, $q_{10}$ and $q_{20}$, was shown to have a higher model performance compared to low flow percentiles, such as $q_{80}$ and $q_{90}$. The use of neural networks seems to be a feasible approach for predicting low flow percentiles, which was shown to outperform the exponential model. The results urge future studies to increase the sample size of catchments used which may improve the derivation of the $\beta$ coefficient of the exponential model equation.

The chapter also highlighted the importance of using artificial neural networks for predicting selected flow characteristics, such as 3-day min and low flow percentiles, $q_{90}$ and $q_{80}$. Riad et al (2004) and Aichouri et al (2015) showed that artificial neural networks perform better than the traditional regression model in predicting flow. The results of these studies showed that neural networks are capable of modelling complex non-linear rainfall-runoff relationships in arid and semi-arid regions, where the relationship between rainfall and runoff is typically irregular. The chapter revealed that the prediction of 3-day min using neural networks shows a higher model performance compared to multiple regression. This suggests that some catchment characteristics have a non-linear relationship with these flow characteristics and cannot be suitable identified using multiple regression. In instances where the model performance of prediction using ANNs are similar or marginally better compared to multiple

http://etd.uwc.ac.za/
regression, multiple regression should be considered as more suitable, as the approach is simpler. Previous studies have identified that the network architecture of the neural network can have a substantial influence on the model performance, as well as issues surrounding over-parameterization, which reduces the predictive powers of the model. Although the results show that multiple regression and neural networks have the capability of predicting selected flow characteristics of non-perennial rivers based on the coefficient of determination, some flow characteristics show a high margin of error for prediction based on the RMSE, such as 3-day min, which needs to be taken into consideration when predicting these flow characteristics in ungauged catchments.

The results of the study urge the use of hydrogeological characteristics such as depth to the water table, permeability and porosity which is important catchment attributes controlling the water storage and groundwater flow routing in the catchment. This may improve the prediction of low flow percentiles, number of zero flow days and BFI. The results of the study recommend future studies to increase the number of catchments used, which may improve the accuracy of prediction. This would increase the variability of flow characteristics and catchment attribute which may allow to the relationship between flow indices and catchment attributes to be more discernible.
CHAPTER 7: IDENTIFICATION OF CATCHMENTS WITH SIMILAR HYDROLOGICAL PROCESSES

7.1 Introduction

Chapter 6 developed predictive equations for predicting flow characteristics in ungauged catchments. This chapter explores whether or not the prediction of flow characteristics is improved by clustering catchments into homogenous groups. A homogenous group refers to grouping of catchments that have a similar hydrological response. The primary objective of clustering catchments into homogenous groups is to assess the membership of ungauged catchments, thus predicting hydrological responses of such catchments.

7.2 Methodology

Cluster analysis is the process of grouping similar catchments according to one or more chosen hydrological signatures, where catchments belonging to the same group are similar (Sawicz et al., 2011; Olden et al., 2015; Elesbon et al., 2015; Begou et al., 2015; Singh et al., 2016; Rahmat et al., 2017). The aim of cluster analysis in the context of the study is to delineate hydrologically homogenous groups, which would allow the transfer of information between gauged and ungauged catchments (Toth, 2013; Singh et al., 2016). Flow characteristics of ungauged catchments can be estimated based on the cluster to which the catchment belongs to, whereby average flow characteristics can be given. The assumption is that catchments within the same group have similar hydrological responses and information can be transferred between them.

Cluster analysis methods are broadly classified into hierarchical and partitioning clustering methods (Ahmad et al., 2013; Zhou et al., 2017; Li et al., 2018). Hierarchical clustering algorithm is the most widely used algorithm (Demirel & Kahya, 2007; Rahmat et al., 2017), whereas k-mean clustering is the one of the most commonly used partitioning clustering method (Li et al., 2018). The Euclidean distance is commonly used as the distance metric in hydrological studies (Blöschl et al., 2013; Latt et al., 2015; Rahmat et al., 2017).
The main issue with cluster analysis in the context of hydrological studies is the availability of different clustering algorithms and distance metrics (Olden et al., 2015). Unfortunately, different clustering algorithms and distance metrics used on the same dataset produce different results (Ahuja, 2012; Olden et al., 2015; Begou et al., 2015). The selection of clustering methods is therefore subjective, and for this particular reason there is no generally agreed clustering method in hydrological studies (Sawicz et al., 2011, Singh et al., 2016).

Cluster analysis was used in the study to group catchments in hydrological homogenous groups. The study used agglomerative hierarchical clustering algorithms and euclidean distance to define homogenous classes between the flow characteristics, as this algorithm and distance metric is widely used in hydrological studies (Demirel & Kahya, 2007; Blöschl et al., 2013; Latt et al., 2015). The use of agglomerative clustering has become more common due to the less time complexity and computational stability (Zhou et al., 2017). The Euclidean distance commonly gives the similarity between two catchments and a distance can be represented by the difference between analytical values from the catchments (Rahmat et al., 2017). Agglomerative clustering is displayed as a dendrogram and starts with n clusters, each of which contains a single object in the data (Yan, 2005; Ahmad et al., 2013; Singh et al., 2016). In the second step, the catchments that are most similar are fused and create a new cluster (Yan, 2005; Ahmad et al., 2013). Eventually, the final result of agglomerative clustering shows all the subgroups fused into one group, where the vertical axis on the dendrogram shows the level of similarity to increase with increasing the number of clusters (Yan, 2005; Li et al., 2018; Ahmad et al., 2013; Singh et al., 2016).

7.2.1 Selection of catchment descriptors for the derivation of clusters

Fraiman et al (2008) pointed out that the determination of which variables are important in cluster analysis can be a difficult task. The inclusion of insignificant and redundant variables introduces ‘noise’ in cluster analysis and the results of the clusters may not reveal the classification objectives, such as classifying catchments into hydrologically homogenous groups (Fraiman et al., 2008; Marlini & Zani, 2013). Redundancy analysis in Chapter 5 identified catchment characteristics that significantly explain the variance of flow characteristics, which included MAP and $S_{90}$. These catchment characteristics are
standardized to ensure that the analysis is independent of measurement units used for the above mentioned catchment characteristics, which can be expressed as (Mazvimavi, 2003):

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}$$  \hspace{1cm} (7.1)

Where; \(i = 1, \ldots, n_c\) catchments, \(j = 1, \ldots, n_e\) 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 across all catchments, and \(s_j\) = standard deviation for variable \(j\) over all catchments.

7.2.2 Determination of the number of clusters and validation

The final step of cluster analysis is to determine a suitable estimate for the number of clusters, which has a deterministic effect on the results obtained (Yan, 2005). One of the main difficulties with using this approach is that the correct number of clusters is often unknown (Yan, 2005; Kodinariya & Makwana, 2013; Zhou et al., 2017).

Cluster validation is an approach of assessing the validity of the classifications that have been obtained in the clustering algorithm (Yan, 2005; Yu et al., 2014). Cluster validation provides a mean of checking the quality of the cluster results and the optimum number of clusters from the clustering algorithm (Yu et al., 2014). Higher cluster validity reflects a higher agreement between the clustering results and the members associated with clusters (Yan, 2005). The validation of clustering algorithms can be done by simple visual inspection. For example, if a catchment with high mean annual runoff and high rainfall is grouped within a cluster associated with high rates of evaporation and low mean annual runoff. The groupings of these two catchments are therefore not hydrologically sensible. An alternative approach to validate the results obtained in the clustering algorithm is the use of Andrew’s curve.

Andrews curve or plot provides a graphical observation of the homogenous groups (Rahmat et al., 2017). The use of Andrews curve can be used to determine the homogeneity of catchments within clusters and catchments that show distinct differences based on the curves can be placed into another group. One of the limitations with Andrews curve is that they are not able to preserve the order (Rahmat et al., 2017). For example, the curve will be
completely different if the variables used in the clustering algorithm are altered (Rahmat et al., 2017). The Andrews curves are produced as (Mazvimavi, 2003):

\[ f(v) = \frac{z_{i1}}{\sqrt{2}} + z_{i2} \sin(v) + z_{i3} \cos(v) + z_{i4} \sin(v) + z_{i5} \cos(v) \] (7.2)

Where; \( z_{i1}, z_{i2}, z_{i3} \) are the standardized catchment characteristics. The shape of the curves are affected by the order in which the catchment characteristics are presented. Studies have identified that it is important to represent the most important catchment characteristic explaining the variance of flow characteristics as \( z_{i1} \). The reason for the ordering of variables is that the variables at the beginning have low frequency cycles and are readily discerned; whilst variables at the end show higher frequency cycles and may not being easily discerned (Mazvimavi, 2003; Gharibnezhad et al., 2011). The results of redundancy analysis identified which catchment characteristics are significant in explaining the variance of flow characteristics. The following order of catchment characteristics was used to generate Andrews curve, MAP (\( z_{i1} \)), \( S_{90} \) (\( z_{i2} \)) and GL\textsubscript{TM} (\( z_{i3} \)). Andrews curve identifies whether or not catchments have been grouped correctly based on the shape of the curves and if catchments have been placed in the wrong cluster, the catchment is removed from the cluster (Rahmat et al., 2017).

Mazvimavi (2003) pointed out that it is important in the context of hydrological regionalisation studies to group catchments that explain the variability of flow characteristics. An important validation technique is to compare the clustering results from those derived from catchment characteristics and those derived from flow characteristics. An agreement between clusters derived from both catchment and flow characteristics validates the delineation of homogenous groups. The Rand Index (\( R_g \)) can be used to determine the level of agreement between clusters derived from catchment and flow characteristics. \( R_g \) is the ratio of the total number of pairs of catchments that are grouped in both clustering algorithms and those which occur in different groups, to the total number of possible combinations. The \( R_g \) can be expressed as (Mazvimavi, 2003):

\[
R_g = \left[ T_g - \frac{U_g}{2} - \frac{V_g}{2} + \frac{n(n-1)}{2} \right] / \frac{n(n-1)}{2} \] (7.3)

\[
T_g = \sum_{i=1}^{g} \sum_{j=1}^{g} mij^2 - n \] (7.4)
\[ U_g = \sum_{i=1}^{g} m_{ij}^2 - n \]  \hspace{1cm} (7.5)

\[ V_g = \sum_{i=1}^{g} m_{ij} - n \]  \hspace{1cm} (7.6)

Where \( g \) is the number of clusters, \( m_{ij} \) is the number of catchments in common between the \( i \)th cluster (catchment characteristics) and \( j \)th cluster (flow characteristics), which forms a matrix \( M \). The matrix calculates \( m_{ij} \) which is the marginal column total of \( M \) and \( m_i \) which is the marginal row total of \( M \). The \( R_g \) index ranges between 0 and 1, where values closer to 1 show a strong agreement between clusters derived from catchment and flow characteristics.

7.3 Results and Discussion

7.3.1 Classification of clusters using catchment characteristics

Table 7.1 shows the membership of catchments for 2 to 10 clusters. There are no significant changes in the composition of clusters when the number of clusters increases from 4 to 6. Catchments that are affected by an increase in clusters from 4 to 6 include; E2H007, J3H017, H3H005, J2H005 and H4H015 which form a new cluster when the number of clusters is 5 and K3H001, K5H002 and K3H004 form a new cluster when the number of clusters is 6. An increase in the number of clusters from 6 to 10 results in minor subdivisions.

7.3.2 Classification of clusters using flow characteristics

Table 7.2 shows the membership of catchments for 2 to 10 clusters, which was derived from catchment characteristics. Increasing the number of clusters from 3 to 5 shows no significant change in the composition of members, with H4H015, K3H001 and K3H004 forming a new cluster. K3H004 is an outlier and forms an individual cluster when the number of clusters is 6. Cluster 1 has a major subdivision of cluster membership when the number of clusters increases from 6 to 7 clusters.
Table 7.1: Cluster membership for 2 to 10 clusters based on cluster analysis of catchment characteristics

<table>
<thead>
<tr>
<th></th>
<th>10</th>
<th>9</th>
<th>8</th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>D3H015</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>D5H011</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>G5H008</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>J1H018</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>E3H001</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>J1H017</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>E2H007</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>J3H017</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>H3H005</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>J2H005</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>H4H016</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>G1H031</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>K3H003</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>G1H041</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>G1H036</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>G4H014</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>H1H006</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>H5H004</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>G2H012</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>S3H006</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>R2H009</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>H8H001</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>J3H018</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Q6H003</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>P4H001</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>S3H012</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>R3H001</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>R3H003</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Q9H002</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Q9H017</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>H4H015</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>J3H016</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>K6H001</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>K3H001</td>
<td>10</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>K5H002</td>
<td>10</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>K3H004</td>
<td>10</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>
Table 7.2: Cluster membership for 2 to 10 clusters based on cluster analysis of flow characteristics

<table>
<thead>
<tr>
<th></th>
<th>10</th>
<th>9</th>
<th>8</th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>D3H015</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>S3H006</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>D5H011</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>E3H001</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>G5H008</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>J1H018</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Q6H003</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Q9H017</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>S3H012</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>G2H012</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>H3H005</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>J1H017</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>J3H017</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>J2H005</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>J3H018</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Q9H002</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>E2H007</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>G1H031</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>G4H014</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>K6H001</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>P4H001</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>R3H003</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>G1H036</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>H4H016</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>H5H004</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>J3H016</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>R2H009</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>R3H001</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>G1H041</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>K5H002</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>H8H001</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>H1H006</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>K3H003</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>H4H015</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>K3H001</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>K3H004</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
7.3.3 Number of clusters

Figure 7.1 presents the results of the Rg statistic which determines the level of agreement between cluster memberships derived from catchment characteristics and flow characteristics.

![Rg statistic vs Number of clusters](http://etd.uwc.ac.za)

Figure 7.1: Variation of Rg statistics with increases in the number of clusters

Figure 7.1 shows that increasing the number of clusters from 2 to 3 causes Rg statistic to decrease from 0.71 to 0.59. An increase in the number of clusters from 4 to 6 causes Rg to increase from 0.67 to 0.70. An increase in the number of clusters from 6 to 7 causes Rg statistic to increase from 0.70 to 0.78. No significant change in the Rg statistic occurs when the number of clusters increases from 7 to 10. Rg statistic suggests that the number of clusters should be 7. An alternative approach that was used to select the number of clusters was the analysis of the flow duration curve of the clusters. For each cluster, the average flow duration curve was plotted based on q_{10}, q_{25}, q_{75}, and q_{90} (Figure 7.2). This identifies whether or not there are distinct hydrological responses between different clusters and whether or not they are hydrologically different. The analysis was undertaken for 4 to 7 clusters based on the results of the Rg statistic, which would identify the optimum number of clusters for the study.
Figure 7.2 illustrates that when the number of clusters is 4, cluster 2 and 4 seem similar in terms of $q_{10}$ and $q_{25}$. However cluster 2 represents a steeper flow duration curve compared to cluster 1 and thus is hydrologically different. Similar trends are found between cluster 1 and 3, however cluster 3 is representative of higher discharge. When the number of clusters increases from 5 to 7 there does not seem to be noticeable differences between the flow duration curves amongst the clusters and therefore the hydrological responses within clusters are no longer unique. The results suggest that the number of clusters should be 4, which is representative of hydrologically similar clusters. Andrew’s curves were plotted for these clusters, which are presented in Figure 7.3.
Cluster 1

Cluster 2

Cluster 3

http://etd.uwc.ac.za/
Cluster 4

Figure 7.3: Andrew’s curves for the clusters based on the catchment characteristics that significantly explain the variance of flow characteristics.

H3H005, J2H005 and H4H016 show differing curves in cluster 1 compared to the rest of the members. These catchments have higher $S_{90}$ compared to the cluster average of 3.43°. These catchments were therefore removed from the cluster.

G4H014 has a curve that differs from the rest of the members of cluster 2. G4H014 has similar catchment characteristics in terms of rainfall and geology in comparison to the rest of the group, however, G4H014 has a very high $S_{90}$ of 5.43° in comparison to the cluster mean of 2.63°. G4H014 was therefore removed from the cluster.

J3H018, Q6H003 and Q9H002 show curves that differ to the other members of cluster 3. These catchments are characterised higher $S_{90}$ in comparison to cluster mean of 3.80°. These catchments were therefore removed from the cluster.

H4H015 shows a curve that differs from the other members in cluster 4. H4H015 has a very high $S_{90}$ of 17.32° in comparison to the cluster mean of 12.93°. H4H015 also has low mean annual rainfall in comparison to the other members of the clusters. H4H015 was therefore removed from the cluster.
7.3.4 Catchment characteristics of clusters

Figure 7.4 presents the spatial distribution of cluster memberships within the study area. Cluster 1 shows the largest diversity in terms of geographic coverage within the study area. Catchments belonging to cluster 2 and 4 are found within the same geographic area. Cluster 3 is found predominantly within the south-eastern regions of the study area.

Figure 7.4: Cluster membership of the selected catchments in the study area

Table 7.3 presents the average catchment characteristics for the derived clusters. Cluster 1 shows the lowest mean annual rainfall and highest mean annual evaporation and is therefore characterised as the most arid. Cluster 4 shows the highest mean annual rainfall and lowest evaporation. Rainfall increases from cluster 1, 3 to 4 and there are no overlapping clusters. Although clusters 2 and 4 display similar mean annual rainfall, the hydrological response between the clusters is expected to be different as cluster 4 is characterised by steeper slopes, as well as being underlain by higher proportion of the Table Mountain Group. Cluster 1, 2 and 3 show similar $S_{90}$, with cluster 4 showing the highest average of 11.47°. Cluster 3 has no proportion of the catchment underlain by the Table Mountain Group, cluster 1 and 2 have
moderate proportions and cluster 4 has a high proportion of the catchment underlain by the Table Mountain Group.

Table 7.3: Mean values for catchment characteristics of the derived clusters

<table>
<thead>
<tr>
<th></th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>Cluster 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAP (mm)</td>
<td>333</td>
<td>841</td>
<td>603</td>
<td>866</td>
</tr>
<tr>
<td>ET (mm)</td>
<td>1688</td>
<td>1575</td>
<td>1513</td>
<td>1509</td>
</tr>
<tr>
<td>$S_{90}$ (degrees)</td>
<td>2.17</td>
<td>2.16</td>
<td>3.01</td>
<td>11.47</td>
</tr>
<tr>
<td>$GL_{TM}$ (%)</td>
<td>16.11</td>
<td>27.52</td>
<td>0</td>
<td>86.08</td>
</tr>
<tr>
<td>$LC_{T}$ (%)</td>
<td>1.74</td>
<td>13.11</td>
<td>22.34</td>
<td>45.03</td>
</tr>
</tbody>
</table>

In terms of land cover, cluster 4 shows the highest proportion of thicket, with cluster 1 showing small proportions. Figure 7.5 provides a graphical representation of the range of catchment characteristics between the derived clusters and the variation amongst them.

Figure 7.5: Range of variation of catchment characteristics for the derived clusters
7.3.5 Flow characteristics of clusters

Table 7.4 presents the average values of flow characteristics of the clusters derived from catchment characteristics. This identifies whether or not the grouping of catchments can be considered as hydrologically homogenous. Cluster 1 shows the lowest mean annual runoff, which is expected as Table 7.3 identified cluster 1 as being the most arid. Cluster 4 shows the highest mean annual runoff, as well as the highest flow percentiles of \( q_{10} \), \( q_{25} \), \( q_{75} \), \( q_{90} \). Cluster 2 and 4 show the highest BFI and as a result these clusters have the lowest number of zero flow days. Cluster 1 and 3 show lowest BFI and mean annual runoff and as a result have the highest number of days of zero flow. This identifies that catchments belonging to cluster 1 and 3 would have prolonged periods of no flow.

<table>
<thead>
<tr>
<th></th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>Cluster 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q (mm)</td>
<td>9.52</td>
<td>198.73</td>
<td>32.61</td>
<td>319.21</td>
</tr>
<tr>
<td>CV (%)</td>
<td>624</td>
<td>258</td>
<td>465</td>
<td>309</td>
</tr>
<tr>
<td>CVB</td>
<td>82.47</td>
<td>12.71</td>
<td>39.30</td>
<td>12.17</td>
</tr>
<tr>
<td>( q_{90} )</td>
<td>0</td>
<td>4.12</td>
<td>0.01</td>
<td>36.40</td>
</tr>
<tr>
<td>( q_{75} )</td>
<td>0</td>
<td>10.29</td>
<td>0.13</td>
<td>53.41</td>
</tr>
<tr>
<td>( q_{25} )</td>
<td>1.60</td>
<td>173.26</td>
<td>9.59</td>
<td>196.02</td>
</tr>
<tr>
<td>( q_{10} )</td>
<td>8.77</td>
<td>446.79</td>
<td>45.74</td>
<td>541.10</td>
</tr>
<tr>
<td>IC</td>
<td>0.04</td>
<td>0.19</td>
<td>0.10</td>
<td>0.13</td>
</tr>
<tr>
<td>3-day min (mm)</td>
<td>0</td>
<td>1.22</td>
<td>0.20</td>
<td>27.38</td>
</tr>
<tr>
<td>3-day max (mm)</td>
<td>477.39</td>
<td>3454.14</td>
<td>1260.82</td>
<td>7525.15</td>
</tr>
<tr>
<td>ZFD (days)</td>
<td>228</td>
<td>27</td>
<td>96</td>
<td>1</td>
</tr>
<tr>
<td>BFI</td>
<td>0.11</td>
<td>0.23</td>
<td>0.14</td>
<td>0.27</td>
</tr>
</tbody>
</table>

The concavity index (IC) is a measure of the contrast between high and low flow events, which represents the shape of the flow duration curve. Catchments having an IC value close to 1 are represented by large aquifers with moderate daily variations of flow. Clusters 2 and 4 are characterised by the highest IC value, which is expected as these clusters also have high
BFI. These clusters also exhibit the lowest CV of annual flows and CVB values, which is expected as a result of stable flow conditions and higher baseflow contribution. Cluster 1 and 3 can therefore be characterised as the most variable in terms of the flow regime. Cluster 1 has the lowest IC of 0.04, which means catchments belonging to the cluster are represented by steeper flow duration curves, where poor groundwater storage does not moderate the daily variation of flow. Figure 7.6 presents the variation of flow characteristics within the derived clusters.

Figure 7.6: Range of variation of flow characteristics for the derived clusters

7.3.6 Prediction of flow characteristics based on clusters

Chapter 6 identified catchment characteristics that can be used to predict flow characteristics in ungauged catchments. The regional regression models and predictive equations developed were based on data of all 36 catchments. One of the assumptions of regionalisation is that the

http://etd.uwc.ac.za/
grouping of catchments into hydrologically homogenous groups can improve the prediction of flow characteristics in ungauged catchments.

7.3.6.1 Prediction of mean annual runoff

Chapter 6 identified that MAP and the proportion of catchment under grasslands are important for predicting mean annual runoff. Chapter 6 also identified a strong correlation between mean annual runoff and rainfall \((r = 0.73)\). Figure 7.7 presents the relationship between runoff and rainfall for the derived clusters.

Figure 7.7: Relationship between mean annual runoff and mean annual rainfall for the derived clusters
Figure 7.7 shows that only cluster 4 have a significant correlation between rainfall and runoff, whilst cluster 1 shows a moderate correlation between rainfall and runoff. Clusters 2 and 3 display poor relationships between rainfall and runoff. Therefore the grouping of catchments into hydrologically homogenous groups has not improved the prediction of mean annual runoff.

The proportion of grasslands were also found to be important for predicting mean annual runoff. Figure 7.8 illustrates the relationship between mean annual runoff and the proportion of the catchment covered by grasslands.

Figure 7.8: Relationship between mean annual runoff and proportion of grasslands for the derived clusters
Chapter 6 identified a moderate negative correlation between mean annual runoff and the proportion of the catchment covered by grasslands ($r = -0.35$). Only cluster 3 shows a significant correlation between mean annual runoff and proportion of grasslands. The grouping of catchments has not improved the prediction of mean annual runoff, as there is no significant relationship between mean annual runoff and proportion of grasslands (Figure 7.8).

### 7.3.6.2 Prediction of baseflow index (BFI)

Chapter 6 identified that mean annual rainfall is important for predicting BFI. Rainfall was shown to have a moderate correlation with BFI ($r = 0.55$). The relationship between rainfall and BFI for the derived clusters is presented below:

- Cluster 1: $r = 0.58$
- Cluster 2: $r = 0.42$
- Cluster 3: $r = 0.16$
- Cluster 4: $r = 0.61$

When rainfall was correlated with BFI based on all catchments without clustering, the correlation coefficient was 0.55. Therefore, clustering has not improved the prediction of BFI, except for cluster 1 and 4 which show a higher correlation coefficient.

### 7.3.6.3 Prediction of the number of zero flow days

Chapter 6 identified that the prediction of the number of zero flow days can be predicted using rainfall. Rainfall was identified to have a strong negative correlation with number of zero flow days ($r = -0.69$). The relationship between rainfall and number of zero flow days is presented below:

- Cluster 1: $r = -0.61$
- Cluster 2: $r = -0.64$
- Cluster 3: $r = -0.32$
- Cluster 4: $r = 0.57$

The grouping of catchments into clusters has not improved the prediction of the number of zero flow days. The correlation between rainfall and the number of zero flow days is lower within the derived clusters compared to the results in Chapter 6.
7.4 Conclusion

This chapter has demonstrated the importance of using redundancy analysis as the basis for grouping catchments into clusters that share similar hydrological response. The use of $R_g$ statistic and analysis of the flow duration curve was used to determine the optimum number of clusters. Andrew’s curves were then used to validate the clusters and ensure that there are no outliers present within the derived clusters.

The chapter identified that the clusters have unique flow characteristics and therefore can be regarded as having hydrologically similar response. The most distinguishable difference between clusters in terms of catchment characteristics are mean annual rainfall, slope and proportion underlain by Table Mountain Group. Rainfall is similar between two clusters; however, these clusters are differentiated by differences in slope and proportion underlain by the Table Mountain Group. Land cover types of thicket do not significantly vary between clusters.

One of the aims of the chapter was to establish whether or not the prediction of flow characteristics could be improved by grouping catchments into hydrologically homogenous groups. The chapter has shown that the prediction of flow characteristics was not improved. The grouping of catchments into hydrologically homogenous groups has narrowed the range of variability of the physiographic characteristics and the relationship between flow and catchment characteristics are no longer discernible. Although the predictive equations for estimating flow characteristics has not been improved, the average values of the derived clusters can be used to predict flow characteristics in ungauged catchments. Identifying the Andrews curve of catchment characteristics between ungauged catchments and clusters can provide a means to place an ungauged catchment within a cluster. This would provide the opportunity to the transfer of information from gauged to ungauged catchments based on determining the cluster to which the ungauged catchment belongs to and then an average value of flow characteristics can be predicted for the ungauged catchment.

A major issue in hydrology is that there is no commonly agreed method for cluster analysis and the optimum number of clusters is often unknown. This limitation causes doubt within the results of cluster analysis by means of whether or not the correct number of clusters was selected and if the groups share similar hydrological response. It is therefore important to
develop a clustering algorithm and distance metric that would enhance the reliability of obtained results and reduce the current subjectivity present within cluster analysis.
CHAPTER 8: CONCLUSIONS AND RECOMMENDATIONS

8.1 Introduction

The aim of the study was to characterise the flow regime of non-perennial rivers in an ecologically meaningful way and to determine factors that account for the spatial variations of ecologically relevant river flow indices. A case study was adopted for 36 river gauging stations in the Western, Eastern and Northern Cape Provinces of South Africa. The flow characteristics in the study included mean annual runoff, coefficient of variation of annual flows, $q_{10}$, $q_{25}$, $q_{75}$, $q_{90}$, baseflow index, concavity index, hydrological index, number of zero flow days, 3-day minima and 3-day maxima of daily means of discharge. These flow characteristics were selected on the recommendations of previous studies with the attempt to characterise the flow regime in an ecologically meaningful way that can be used to efficiently manage non-perennial rivers. Assessing the spatial variation of these flow characteristics was based on the relationship between flow and catchment characteristics. The selection of catchment characteristics was based on the recommendations of previous studies on the characteristics that influence hydrological response which included rainfall, evaporation, slope and elevation, land cover, geology, soil, drainage density, river length and catchment area.

8.2 Factors accounting for the spatial variation of selected ecologically relevant flow characteristics

Redundancy analysis provided the opportunity to identify catchment characteristics that significantly explain the variance of flow characteristics at a 5 % confidence level. Mean annual rainfall and the slope exceeded or equally 90 % of the time ($S_{90}$) were the only catchment characteristics that were identified as significantly explaining the variance of flow characteristics. The total variance of flow characteristics explained by MAP and $S_{90}$ accounted for 57 %, with MAP accounting for 46 % of variance. The results of the chapter identified catchment characteristics that account for the spatial variation of flow characteristics, which was used to predict these flow characteristics in Chapter 6 using multiple regression and artificial neural networks. The results were also used as a basis for
grouping catchments that share similar hydrological responses and to determine whether or not the grouping of catchments improves the prediction of flow characteristics.

8.3 Prediction of flow characteristics using multiple regression and artificial neural networks

The study used multiple regression and artificial neural networks to determine which approach would be more suitable for the prediction of selected flow characteristics of non-perennial rivers. The model performance of both multiple regression and neural networks was based on $R^2$, % RMSE and RSR. Overall, multiple regression shows a higher model performance and accuracy for predicting flow characteristics in the study compared to neural networks. The prediction of mean annual runoff, BFI, number of zero flow days and 3-day max show a good model performance using multiple regression. Generally the results show good performance for predicting flow characteristics based on the coefficient of determination, however, the prediction of some flow characteristics illustrates a higher degree of error based on the % RMSE. An exponential model was shown to be feasible for predicting the flow duration curves of non-perennial rivers which was shown to have a lower predictive error than artificial neural networks. The prediction of flow duration curves in ungauged catchments may be improved by increasing the sample size of catchments used in the study. The study used 36 gauged non-perennial rivers which were insufficient to derive the $\beta$ coefficient of the exponential model equation using catchment attributes. In the case of this study, the $\beta$ coefficient was shown to have no relationship with catchment attributes. The use of artificial neural networks shows a higher model performance for predicting 3-day min, which suggests that some catchment characteristics have non-linear relationships with flow characteristics which cannot be discerned using multiple regression. In cases where the model performance is similar between neural networks and multiple regression, the use of multiple regression is recommended as the approach is simpler.

8.4 Identification of catchments with similar hydrological responses

The results of ordination in Chapter 5 provided a set of catchment characteristics that were used to classify catchments into homogenous groups, which was mean annual rainfall and $S_{90}$. The rand index ($R_g$) was used to determine the number of clusters in the study, which identifies the level of agreement between clusters derived based on catchment characteristics
and those derived by flow characteristics. Previous studies have elaborated on the limitation of determining the optimum number of clusters and the approach is often subjective. The study assessed the flow duration curves of the derived clusters to assist in determining the optimum number of clusters. The approach identified whether or not the clusters could be regarded as sharing similar hydrological response and hydrologically different between clusters. Andrews curves were then used to validate the membership of clusters and removed outliers of the cluster that showed differing curves.

8.5 Prediction of flow characteristics based on hydrologically similar catchments

One of the assumptions of regionalisation is that the grouping of catchments into hydrologically homogenous groups can improve the prediction of flow characteristics in ungauged catchments. The prediction of flow characteristics based on clusters provided the opportunity to determine whether or not the grouping of catchments improved the model performance of the predictive equation. However, the grouping of catchments into homogenous group did not improve the prediction of flow characteristics. The grouping of catchments into homogenous groups resulted in narrow ranges of physiographic attributes and the relationship between flow and catchment characteristics were no longer discernible. The prediction of flow characteristics may be improved with increasing the number of memberships within each cluster and increasing the variability of catchment characteristics. Although the predictive equations were not improved, flow characteristics can be estimated in ungauged catchments by placing the ungauged catchment in a cluster that share similar catchment characteristics. The transfer of information from gauged to ungauged catchments can be based on determining the cluster to which the ungauged catchment belongs to and then an average value of flow characteristics of the cluster can be assigned to the ungauged catchment.

8.6 Recommendations

The study identified that there is limited knowledge regarding the functioning and operation of non-perennial rivers. In the context of perennial rivers, a range of environmental flow assessments have been developed ranging from simple hydrological methods, to complex holistic approaches. Non-perennial rivers differ with regards to their flow regime compared
to perennial rivers and it is therefore important to design approaches for the determination of
environmental flow that are explicit for these river systems. Studies need to identify an
adequate set of flow indices that can be used efficiently to characterise the flow regime of
non-perennial rivers in an ecologically meaningful way. Studies have also pointed out the
issue of inherent redundancy that exists among hydrological indices. An approach should be
carried out to efficiently characterise the flow regime, limiting redundant indices and
adequately representing the flow regime in an ecologically meaningful way.

One of the main issues with reference to assessing the spatial variation of flow characteristics
was the use of high resolution geological data and hydrologically meaningless data. In the
context of hydrology, hydrologically meaningful data such as water depth, porosity and
permeability are important for understanding the relationship between subsurface storage and
river flow. The use of more hydrologically meaningful data may improve the understanding
of these rivers systems. Studies have argued that this is a major concern in developing
countries where such data is not readily available.
REFERENCES


http://etd.uwc.ac.za/
40. CSIR. 2014. Environmental screening study for a proposed LNG terminal at Saldanha and associated pipeline infrastructures to Atlantis and Cape Town, Western Cape, South Africa.
42. Dahlman L. 2017. Climate Change: Global Temperature


53. Duah AA. 2010. Sustainable Utilisation of Table Mountain Group Aquifers University of the Western Cape


61. ESRI. 2018. ArcGIS Desktop


86. Hughes DA. 2006. Water resource estimation in less developed regions-issues of uncertainty associated with a lack of data
87. Hughes DA. 2008. Hydrological information requirements and methods to support the determination of environmental water requirements in ephemeral river systems. Water Research Commission


125. Magoba RF. 2014. Influences of river flows on recruitment success of Fynbos Riparian Vegetation along selected high gradient streams in the Western Cape, South Africa. University of the Western Cape
127. Malan GJ. 2016. Investigating the suitability of land cover type information for hydrologically modelling in the mountain regions of Hessqua, South Africa. University of Stellenbosch
129. Mason D, Maidment DR. 2000. An analysis of a Methodology for Generating Watershed Parameters using GIS. CRWR Online Report 00-3. Center for Research in Water Resources. The University of Texas at Austin
140. Merwade V. 2012. Watershed and Stream Network Delineation using ArcHydro Tools. School of Civil Engineering, Purdue University
146. Molaba GL. 2017. Investigating the possibility of targeting major dolerite intrusions to supplement municipal water supply in Bloemfontein. A geophysical approach. Faculty of Natural and Agricultural Sciences (Institute for Groundwater Studies). University of the Free State.

http://etd.uwc.ac.za/


http://etd.uwc.ac.za/


197. Sinclair Knight Merz. 2010. Developing guidelines for the selection of streamflow gauging stations. SKM


222. Tighe ML, Chamberlain D. 2009. Accuracy comparison of SRTM, ASTER, NED, NEXTMAP USA Digital Terrain Model over several USA Study Sites

223. Tocachi BO. 2014. Regionalisation of hydrological indices to assess land-use change impacts in the Tropical Andes. Department of Civil and Environmental Engineering. Imperial College London


136


241. WCSS Statistical Software. Hierarchical Clustering/Dendrograms


248. Yan M. 2005. Methods of Determining the Number of Clusters in a Data Set and a New Clustering Criterion. Faculty of the Virginia Polytechnic Institute and State University


138