AUSTRALIAN RAINFALL AND RUNOFF
REVISON PROJECT 7: BASEFLOW FOR CATCHMENT SIMULATION

STAGE 2 REPORT

AUGUST, 2011

Project Project 7: Baseflow for Catchment Simulation	**AR&R Report Number** P7/S2/017
Date 31 August 2011	**ISBN** 978-0-85825-916-4
Contractor Sinclair Knight Merz	**Contractor Reference Number** VW04648
Authors Rachel Murphy Zuzanna Graszkiewicz Peter Hill Brad Neal Rory Nathan	**Verified by**

ACKNOWLEDGEMENTS

Contractor Details

Sinclair Knight Merz (SKM)
PO Box 2500
Malvern
Victoria 3144

Tel: (03) 9248 3100
Fax: (03) 9500 1180
Web: www.skmconsulting.com

FOREWORD

AR&R Revision Process

Since its first publication in 1958, Australian Rainfall and Runoff (ARR) has remained one of the most influential and widely used guidelines published by Engineers Australia (EA). The current edition, published in 1987, retained the same level of national and international acclaim as its predecessors.

With nationwide applicability, balancing the varied climates of Australia, the information and the approaches presented in Australian Rainfall and Runoff are essential for policy decisions and projects involving:

- infrastructure such as roads, rail, airports, bridges, dams, stormwater and sewer systems;
- town planning;
- mining;
- developing flood management plans for urban and rural communities;
- flood warnings and flood emergency management;
- operation of regulated river systems; and
- Prediction of extreme flood levels.

However, many of the practices recommended in the 1987 edition of ARR now are becoming outdated, and no longer represent the accepted views of professionals, both in terms of technique and approach to water management. This fact, coupled with greater understanding of climate and climatic influences makes the securing of current and complete rainfall and streamflow data and expansion of focus from flood events to the full spectrum of flows and rainfall events, crucial to maintaining an adequate knowledge of the processes that govern Australian rainfall and streamflow in the broadest sense, allowing better management, policy and planning decisions to be made.

One of the major responsibilities of the National Committee on Water Engineering of Engineers Australia is the periodic revision of ARR. A recent and significant development has been that the revision of ARR has been identified as a priority in the Council of Australian Governments endorsed National Adaptation Framework for Climate Change.

The update will be completed in three stages. Twenty one revision projects have been identified and will be undertaken with the aim of filling knowledge gaps. Of these 21 projects, ten projects commenced in Stage 1 and an additional 9 projects commenced in Stage 2. The remaining two projects will commence in Stage 3. The outcomes of the projects will assist the ARR Editorial Team with the compiling and writing of chapters in the revised ARR.

Steering and Technical Committees have been established to assist the ARR Editorial Team in guiding the projects to achieve desired outcomes. Funding for Stages 1 and 2 of the ARR revision projects has been provided by the Federal Department of Climate Change and Energy Efficiency. Funding for Stages 2 and 3 of Project 1 (Development of Intensity-Frequency-Duration information across Australia) has been provided by the Bureau of Meteorology.

Project 7: Baseflow for Catchment Simulation

An important aspect of flow estimation as distinct from flood estimation is the relative importance of the baseflow component of a hydrograph. Whereas the quickflow component is the most significant component of a hydrograph for flood estimation and the baseflow component is neglected, this is not always the case for general flow estimation. In recent years the need to estimate small flood flows (in-bank floods) has arisen and, therefore, estimation of baseflow needs to be considered within Australian Rainfall and Runoff.

This project focuses on the development of appropriate techniques for estimating the baseflow component of a hydrograph. It is expected that both statistical and deterministic approaches be developed to meet the various needs of the industry.

This project will result only in preliminary guidance in a form suitable for inclusion in Australian Rainfall and Runoff. It is expected that further developments will occur post this edition of Australian Rainfall and Runoff.

The aim of Project 7 is to identify and test techniques for estimation of the baseflow component of a flood hydrograph for situations where the baseflow cannot be neglected as a significant component of the flood hydrograph.

Mark Babister
Chair Technical Committee for
ARR Research Projects

Assoc Prof James Ball
ARR Editor

AR&R REVISION PROJECTS

The 21 AR&R revision projects are listed below :

ARR Project No.	Project Title	Starting Stage
1	Development of intensity-frequency-duration information across Australia	1
2	Spatial patterns of rainfall	2
3	Temporal pattern of rainfall	2
4	Continuous rainfall sequences at a point	1
5	Regional flood methods	1
6	Loss models for catchment simulation	2
7	**Baseflow for catchment simulation**	**1**
8	Use of continuous simulation for design flow determination	2
9	Urban drainage system hydraulics	1
10	Appropriate safety criteria for people	1
11	Blockage of hydraulic structures	1
12	Selection of an approach	2
13	Rational Method developments	1
14	Large to extreme floods in urban areas	3
15	Two-dimensional (2D) modelling in urban areas.	1
16	Storm patterns for use in design events	2
17	Channel loss models	2
18	Interaction of coastal processes and severe weather events	1
19	Selection of climate change boundary conditions	3
20	Risk assessment and design life	2
21	IT Delivery and Communication Strategies	2

AR&R Technical Committee:

Chair: Mark Babister, WMAwater

Members: Associate Professor James Ball, Editor AR&R, UTS

Professor George Kuczera, University of Newcastle

Professor Martin Lambert, University of Adelaide

Dr Rory Nathan, SKM

Dr Bill Weeks, Department of Transport and Main Roads, Qld

Associate Professor Ashish Sharma, UNSW

Dr Bryson Bates, CSIRO

Michael Cawood, Engineers Australia Appointed Technical Project Manager, MCA

Related Appointments:

ARR Project Engineer: Monique Retallick, WMAwater

Assisting TC on Technical Matters: Dr Michael Leonard, University of Adelaide

PROJECT TEAM

Project Team Members:

- Dr Rory Nathan (ARR TC Project Manager and SKM)
- Rachel Murphy (SKM)
- Zuzanna Graszkiewicz (SKM)
- Peter Hill (SKM)
- Brad Neal (SKM)
- Dr Tony Ladson (SKM)
- Jason Wasik (SKM)

This report was independently reviewed by:

- Erwin Weinmann
- Trevor Daniell (University of Adelaide)

EXECUTIVE SUMMARY

ARR Update Project 7 aims to develop a method for calculating and adding baseflow contribution to design flood estimates. Phase 1 of the project focussed on the physical processes of groundwater-surface water interaction and theoretical approaches to baseflow separation. The identified methods were applied to eight case study catchments across Australia in order to develop a suitable approach for more wide scale application.

Phase 2 of Project 7 covers the analysis of 236 catchments across Australia, the development of prediction equations to estimate baseflow parameters and the development of a method for the application of these to design estimates for catchments across Australia.

Streamflow can be considered to comprise of two main components based on the timing of response in a river after a rainfall event. Water that enters a stream rapidly is termed "quickflow" and is sourced from direct rainfall onto the river surface and rainfall-runoff across the land surface. Water which takes longer to reach a river is termed "baseflow" and is sourced primarily from groundwater discharge into the river. Different locations have varying degrees of baseflow contribution to streamflow based on regional hydrogeological conditions.

This study consisted of the analysis of baseflow characteristics for the 236 study catchments. A baseflow series was separated from the streamflow series for each catchment using the method developed in Phase 1 of the project. The method used the Lyne-Hollick filter with a filter parameter value of 0.925 and 9 passes across the hourly data. At each site, the number of events analysed was equal to four times the number of years of data. The baseflow of each event was characterised by three ratios, defined below based on the characteristics presented in Figure 1:

1. **Baseflow Peak Ratio**: Ratio of the peak baseflow (C) to the peak streamflow (A), given by C/A.

2. **Baseflow Volume Ratio**: The event baseflow index (BFI), which is given by the total baseflow volume for the duration of the event divided by the total streamflow volume. This is the ratio of the shaded areas in the example hydrograph.

3. **Baseflow Under Peak Ratio**: Ratio of the baseflow at the time of the streamflow peak (B) to the peak streamflow (A), given by B/A.
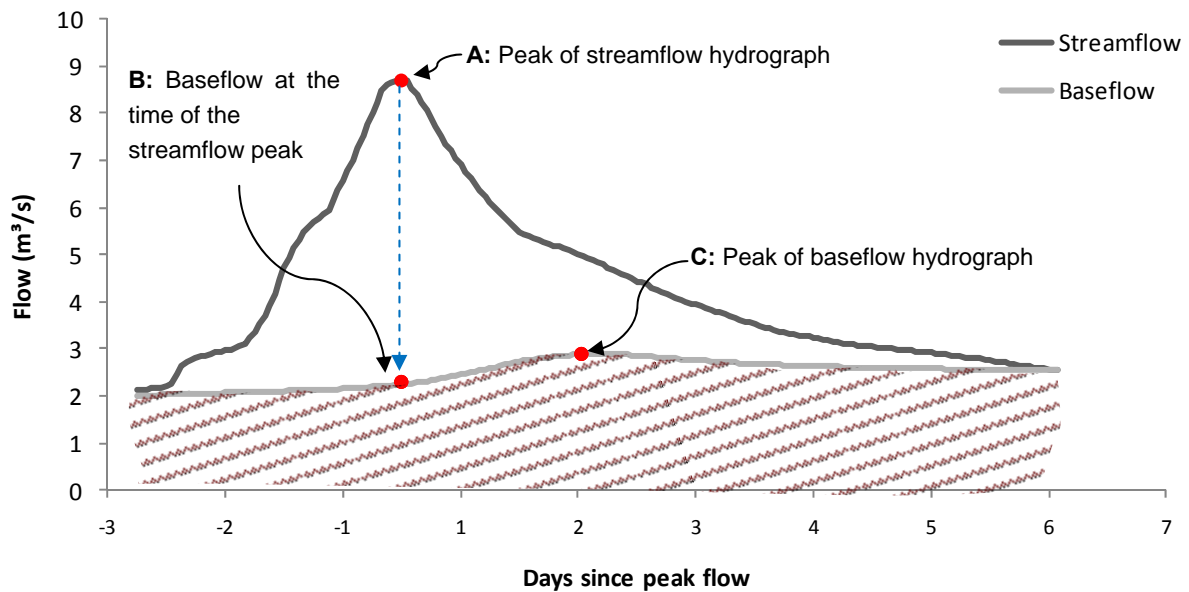
Figure 1 Key characteristics for calculation in a flood hydrograph

The Baseflow Peak Ratio and Baseflow Volume Ratio values for each event were summarised using a power relationship with Average Recurrence Interval (ARI). From this relationship, the ratio value at the 10 year ARI for each site was used as the key statistic in further analysis.

The analysis of baseflow for the study catchments aimed to provide information that could be used in determining the baseflow in ungauged catchments. Catchment characteristics were used to develop a regression model to predict the Baseflow Peak Ratio and Baseflow Volume Ratio for the fitted 10 year ARI event. The model consisted of a regression tree and multiple linear regression equations.

Regression tree analysis classifies catchments into clusters based on the values of particular catchment characteristics. Multiple linear regression produces a continuous linear function for estimating ratio values based on independent variables.

A regression tree was used to group like catchments on the basis of the Baseflow Peak Ratio at the 10 year ARI. Each cluster was defined by the values of particular catchment characteristics identified by the analysis as statistically significant. A multiple linear regression analysis was then conducted on each cluster identified by the regression tree. This produced an equation that could be used to predict the value of the Baseflow Peak Ratio at the 10 year ARI for an ungauged catchment.

The value of the Baseflow Volume Ratio at the 10 year ARI was predicted using the same clusters identified using the Baseflow Peak Ratio at the 10 year ARI. A second multiple regression analysis was conducted on the catchments in each group using the Baseflow Volume Ratio at the 10 year ARI as the dependent variable. The calculated value for the Baseflow Peak Ratio at the 10 year ARI was used as an independent variable in the Baseflow Volume Ratio regression.

The complete regression model for both the Baseflow Peak Ratio and Baseflow Volume Ratio

explained approximately 60% of the variation in baseflow characteristics across all study catchments.

To allow the application of baseflow characteristics to design flood estimation the proportion of total streamflow that was estimated to be baseflow was linked to the calculated value of surface runoff. This produced two factors which are directly related to the ratios analysed since:

$$R_{BFn} = \frac{Ratio_i}{1 - Ratio_i}$$

Equation 1

Where $Ratio_i$ represents either the Baseflow Peak Ratio or the Baseflow Volume Ratio, both measures of the baseflow contribution relative to the total streamflow

R_{BFn} where *n* represents either the flow or volume baseflow factor that relates baseflow contributions to the surface runoff, representing either:

1. **Baseflow Peak Factor**. This factor is applied to the estimated surface runoff peak flow to give the value of peak baseflow for a 10 year ARI event.

2. **Baseflow Volume Factor**. This factor is applied to the estimated surface runoff volume to give the volume of the baseflow for a 10 year ARI event.

The Baseflow Peak Factor and Baseflow Volume Factor are presented in Figure 2 and Figure 3, which cover the whole of Australia. A national catchment dataset was used to define spatial areas for the application of factors in the maps. The total upstream catchment area was used in this analysis, rather than the interstation catchment areas. This reflects the standard method used to define contributing catchment areas for design flood assessments.
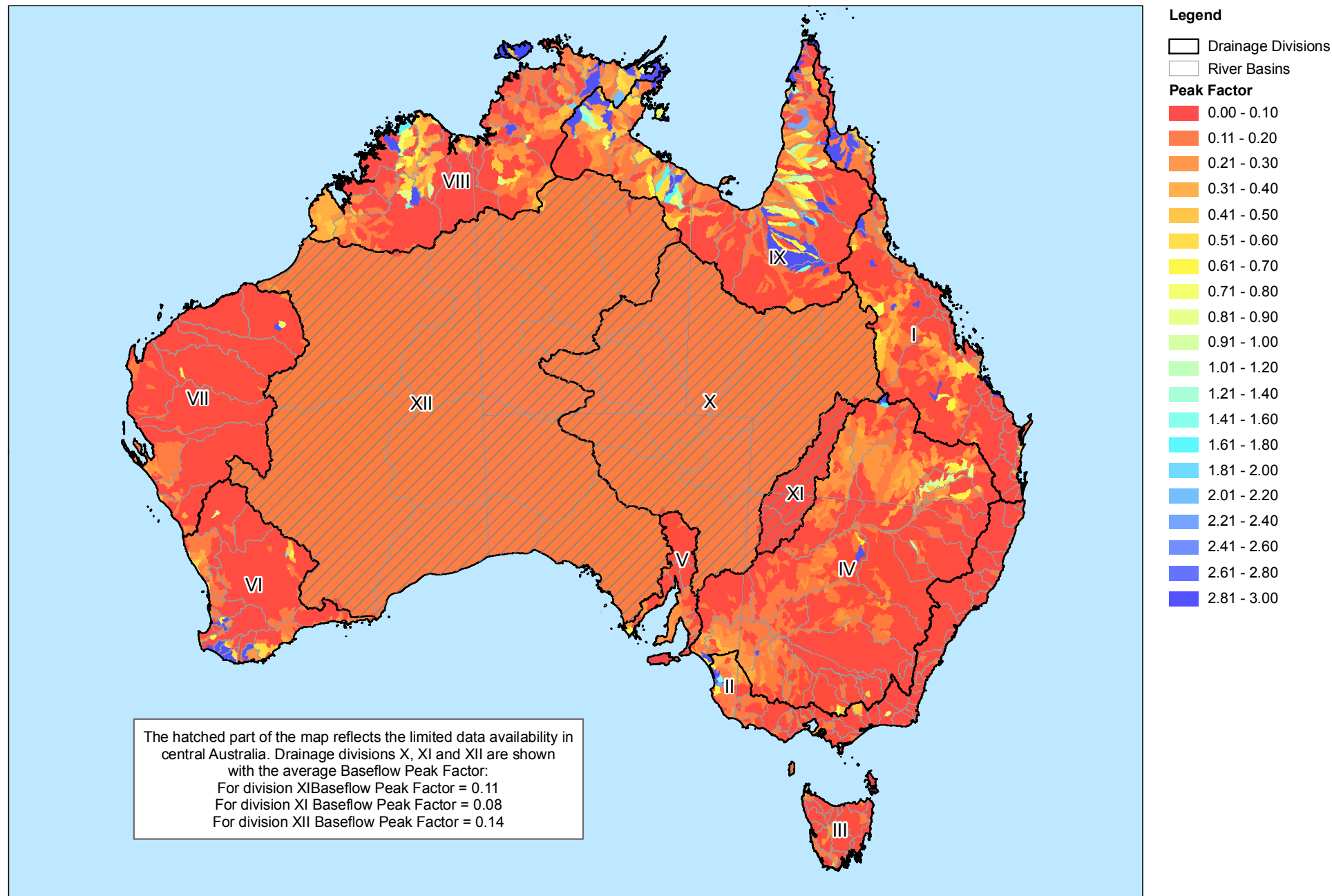
SKM

Legend

☐ Drainage Divisions
☐ River Basins

Peak Factor

■	0.00 - 0.10
■	0.11 - 0.20
■	0.21 - 0.30
■	0.31 - 0.40
■	0.41 - 0.50
■	0.51 - 0.60
■	0.61 - 0.70
■	0.71 - 0.80
■	0.81 - 0.90
■	0.91 - 1.00
■	1.01 - 1.20
■	1.21 - 1.40
■	1.41 - 1.60
■	1.61 - 1.80
■	1.81 - 2.00
■	2.01 - 2.20
■	2.21 - 2.40
■	2.41 - 2.60
■	2.61 - 2.80
■	2.81 - 3.00

The hatched part of the map reflects the limited data availability in central Australia. Drainage divisions X, XI and XII are shown with the average Baseflow Peak Factor:
For division XIBaseflow Peak Factor = 0.11
For division XI Baseflow Peak Factor = 0.08
For division XII Baseflow Peak Factor = 0.14

■ **Figure 2 - Map of Baseflow Peak Factor for ARI of 10 years**

SINCLAIR KNIGHT MERZ

0   500
Kilometres
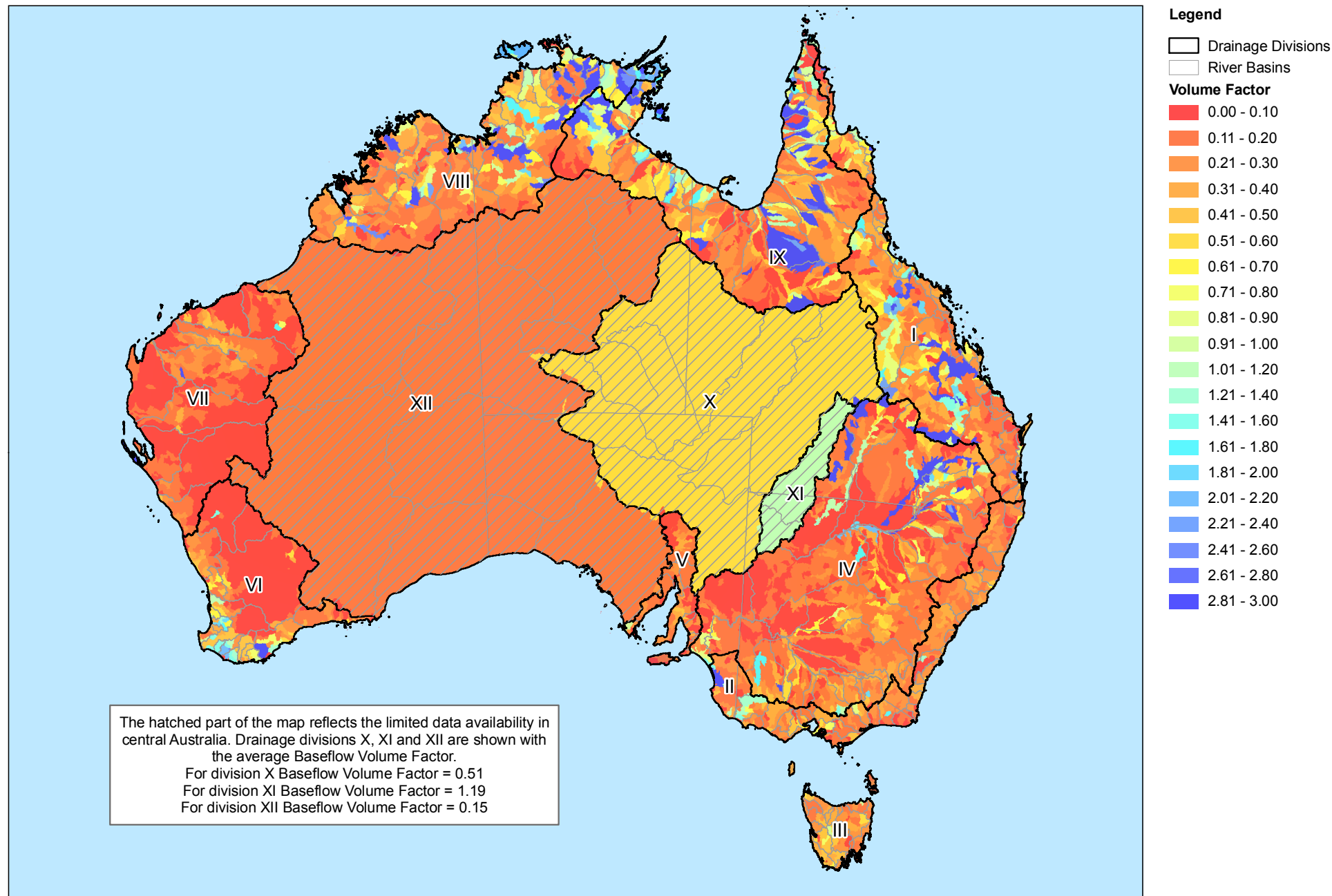
Figure 3 - Map of Baseflow Volume Factor for ARI of 10 years

SINCLAIR KNIGHT MERZ

I:\VWES\Projects\VW04648\Technical\Spatial\ArcGIS\Fig47-Baseflow_Volume_Factor.mxd. Date: 06/09/2010

The maps presented in Figure 2 and Figure 3 can be used by practitioners to immediately determine the Baseflow Peak Factor and the Baseflow Volume Factor to apply to surface runoff estimates for an ARI of 10 years. A third factor is able to be calculated from the relationship between the Baseflow Peak Ratio and Baseflow Under Peak Ratio:

3. **Baseflow Under Peak Factor**. This factor is applied to the estimated surface runoff peak flow to give the baseflow under the peak surface runoff. It can be determined from the final value of the Baseflow Peak Factor for an event based on the relationship Baseflow Under Peak Factor = 0.7 x Baseflow Peak Factor

These factors provide information on the baseflow contribution to design flood events for 1 in 10 year event magnitudes. Table 1 shows the ARI factors that should be applied to the 10 year ARI Baseflow Peak Factor and Baseflow Volume Factor to scale the relevant factor to reflect events of other magnitudes.

Table 1 ARI Factors, F_{ARI}, to be applied to the 10 year Baseflow Peak Factor and the Baseflow Volume Factor to determine the Baseflow Peak Factor for events of various ARIs

ARI (years)	ARI Factor for Baseflow Peak Factor	ARI Factor for Baseflow Volume Factor
0.5	3.0	2.6
1	2.2	2.0
2	1.7	1.6
5	1.2	1.2
10	1.0	1.0
20	0.8	0.8
50	0.7	0.7
100	0.6	0.6

For events of ARIs not shown in Table 1, Figure 4 can be used to determine an appropriate ARI factor. This is to be multiplied by the 1 in 10 year Baseflow Peak Factor or the Baseflow Volume Factor as relevant to determine the factor for other event magnitudes as specified in Equation 2 and Equation 3.
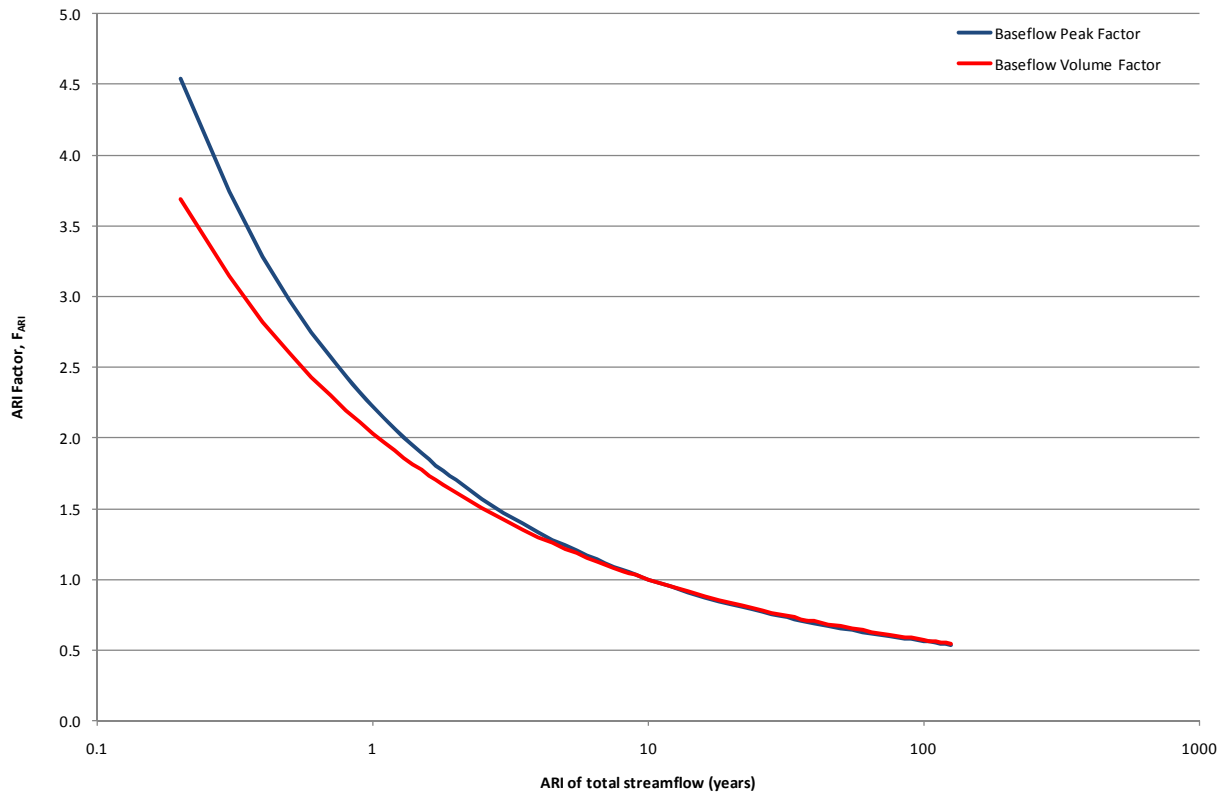
Figure 4 ARI Factors, F$_{ARI}$, to be applied to the 10 year Baseflow Volume Factor to determined the Baseflow Volume Factor for events of various ARIs

Equation 2 shows the final factor to be applied to the calculated surface runoff to determine the event peak baseflow.

$$R_{BPF} = F_{ARI}R_{BPF,10yrARI}$$

Equation 2

Equation 3 shows the final factor to be applied to the calculated surface runoff volume to determine event baseflow volume.

$$R_{BVF} = F_{ARI}R_{BVF,10yrARI}$$

Equation 3

This information should be applied to design flood estimation using the procedure outlined in the relationships below which relate to the typical flood hydrograph in Figure 1.

To calculate the peak baseflow (point C in Figure 1):

1. Determine the Baseflow Peak Factor for a 10 year ARI ($R_{BPF,10yrARI}$) from Figure 2.

2. Determine the ARI factor corresponding to the event ARI using Table 1 or Figure 4. Apply to the 10 year Baseflow Peak Factor as in Equation 2 to determine the Baseflow Peak Factor for the event magnitude of interest.

3. Apply the Baseflow Peak Factor to the calculated peak surface runoff as in Equation 4.

$$Q_{Peak\ baseflow} = R_{BPF} Q_{Peak\ surface\ runoff}$$

Equation 4

To calculate the baseflow under the peak streamflow (point B in Figure 1):

1. The Baseflow Peak Factor (R_{BPF}) calculated for the appropriate event ARI as above should be used in Equation 5 to calculate the Baseflow Under Peak Factor (R_{BUPF}).

$$R_{BUPF} = 0.7 \times R_{BPF}$$

Equation 5

2. R_{BUPF} should be used as in Equation 6 to calculate the baseflow under the peak streamflow.

$$Q_{Baseflow\ under\ peak\ streamflow} = R_{BUPF} Q_{Peak\ surface\ runoff}$$

Equation 6

To calculate the total streamflow peak (point A in Figure 1):

1. Calculate the baseflow under the streamflow peak for the appropriate ARI as above

2. Add the baseflow under the streamflow peak calculated using Equation 6 to the calculated peak surface runoff as in Equation 7.

$$Q_{Peak\ streamflow} = Q_{Peak\ surface\ runoff} + Q_{Baseflow\ under\ peak\ streamflow}$$

Equation 7

To calculate the total baseflow volume for an event (red hashed in Figure 1):

1. Determine the Baseflow Volume Factor for a 10 year ARI ($R_{BVF,10yrARI}$) from Figure 3

2. Determine the ARI factor corresponding to the event ARI using Table 1 or Figure 4. Apply to the 10 year Baseflow Volume Factor as in Equation 3 to determine the Baseflow Volume Factor (R_{BVF}) for the event.

3. Apply the Baseflow Volume Factor to the calculated surface runoff volume as in Equation 8.

$$V_{Baseflow} = R_{BVF} V_{Surface\ Runoff}$$

Equation 8

To calculate the total streamflow volume for an event (blue hashed in Figure 1):

1. Calculate the baseflow volume for the event using the appropriate ARI factors.

2. The baseflow volume calculated using Equation 8 should be added to the calculated surface runoff as in Equation 9.

$$V_{Total\ streamflow} = V_{Surface\ runoff} + V_{Baseflow}$$

Equation 9

This approach can be directly applied to the 1 in 10 year Baseflow Peak Factor or Baseflow Volume Factor to readily scale the value to reflect a variety of event magnitudes. This enables the method to be applied on a wide scale, for any event size between 1 in 0.5 years and 1 in 100 years. As expected, these scaling factors indicate the baseflow contribution to flood events is largest for small events. For rare events, baseflow is only a small proportion of the total surface runoff.

Table of Contents

1.   Introduction

Guidelines for rainfall-based design flood estimation are contained in Australian Rainfall and Runoff (Institution of Engineers, 1999). The procedure for estimating a design flood hydrograph with specified annual exceedance probability (AEP) for a catchment starts with a design rainfall of the desired AEP. As indicated in Figure 5, the probability of the calculated design flood peak will depend upon the choice of the critical storm duration, areal reduction factor, storm temporal pattern, design losses, runoff routing model, model parameters, and baseflow.
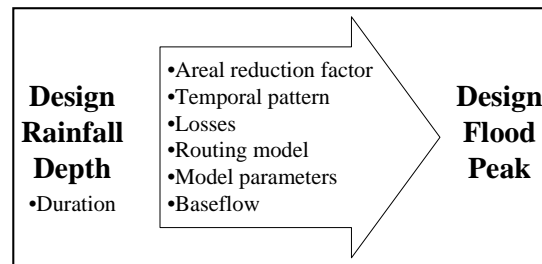
Design Rainfall Depth
•Duration

•Areal reduction factor
•Temporal pattern
•Losses
•Routing model
•Model parameters
•Baseflow

Design Flood Peak

Figure 5 Event Based Design Flood Estimation

Each of these components has a distribution of possible values, so the probability of the calculated flood peak should theoretically account for the effect of the combined probabilities. In the light of the current lack of information on the true distribution of each of the components, and the complexity involved, the recommendation in Australian Rainfall and Runoff is to take some 'central' or 'typical' value for each of the key inputs. Of all of the inputs shown in Figure 5, there is least guidance available in Australian Rainfall and Runoff on appropriate values for the baseflow contribution to design flood estimates.

Book V, Section 2 of Australian Rainfall and Runoff (Cordery, 1998) provides methods for estimating surface runoff during flood events, but does not currently provide any guidance on estimating the component of the flood hydrograph sourced from baseflow. Baseflow is generally a minor component in extreme flood events, but can potentially be significant in smaller flood events. This is particularly the case where the catchment geology consists of high yielding aquifers with large baseflows.

The focus of Australian Rainfall and Runoff Update Project 7 (Baseflow for Catchment Simulation) is to recommend practical yet technically robust preliminary advice on the estimation of baseflow in design flood events for inclusion in Australian Rainfall and Runoff. This report presents the findings of Stage 2 of the project, and summarises the extensive data analysis tasks to obtain baseflow statistics for flood events across 236 catchments, the development of prediction equations to relate the baseflow statistics to catchment characteristics, and the application of these relationships across Australia. The following provides a summary of the report structure:

- Section 2 contains a summary of Phase 1 findings;
- Section 3 describes how the baseflow analysis of each study catchment was conducted. It includes information about the preparation of streamflow data, the methods for separating baseflow and identifying events and flood frequency analysis;

- Section 4 summarises the results of baseflow analysis for all the study catchments and describes the key characteristics identified for use in application to design;

- Section 5 describes the development of the regression tree and prediction equations to calculate baseflow statistics for ungauged catchments;

- Section 6 introduces the concepts required to develop the analysis results into a method for application to design. It also covers the variation of baseflow statistics with ARI;

- Section 7 describes the application of the analysis across Australia and summarises the method for including baseflow into design calculations;

- Section 8 discusses the validation of the sensitivity of the approach to the spatial scale;

- Section 9 provides conclusions from the study;

- Acknowledgements and References are provided at the end of the report; and

- Appendices are provided with further details of the analysis tasks. This includes a discussion around accounting for the differences in event length as a result of the event analysis approach, flood frequency distributions for each catchment, presentation of the variation in Baseflow Peak Factor and Baseflow Volume Factor for each catchment, and diagnostics summarising the validity of regression relationships.

2.   Summary of Phase 1 of ARR Project 7

ARR Update Project 7 aims to develop a method for estimating baseflow contribution to different sized flood events across Australia. Phase 1 of the project included preliminary tasks, with the intent to develop a suitable approach to baseflow separation and analysis that could be utilised in Phase 2 of the study. To support this, a thorough literature review was undertaken to consolidate the understanding of the physical processes of groundwater-surface water interaction. This informed decision making to enable a practical approach to baseflow separation to be identified.

A method for estimating baseflow contribution to flood peaks was subsequently tested in case study catchments across a range of Australian conditions. The outcomes of the work provided a method for more widescale application across Australia. The following discussion summarises the key findings of Phase 1 of the project in more detail and outlines how these findings relate to this current phase of the study.

2.1.   Baseflow separation theory

Streamflow can be considered to comprise of two main components based on the timing of response in a river after a rainfall event. Water that enters a stream rapidly is termed "quickflow" and is sourced from direct rainfall onto the river surface and rainfall-runoff across the land surface. Water which takes longer to reach a river is termed "baseflow" and is sourced primarily from groundwater discharge into the river. Different locations have varying degrees of baseflow contribution to streamflow based on regional hydrogeological conditions.

Baseflow has been the subject of much investigation in the past, and a range of techniques are available to estimate its behaviour. Through Phase 1 of the study, it was necessary to quantify the magnitude of baseflow associated with flood events, regardless of the source of the water or the detailed and often complex physical processes which generate it. For this reason, the study used automated baseflow separation techniques as an investigative tool, rather than more detailed models or field based chemical tracer studies of groundwater and surface water interaction.

The magnitude of the peak and the shape of a baseflow hydrograph in flood events are subjective because baseflow is not readily measureable. However, some common features capture the general understanding of the physical processes in action:

- Low flow conditions prior to the commencement of a flood event typically consist entirely of baseflow.

- The rapid increase in river level relative to the surrounding groundwater level results in an increase in bank storage. The delayed return of this bank storage to the river causes the baseflow recession to continue after the peak of the total hydrograph.

- Baseflow will peak after the total hydrograph peak, due to the storage-routing effect of the sub-surface stores.

- The baseflow recession will most likely follow an exponential decay function (a master recession curve).

- The baseflow hydrograph will rejoin the total hydrograph as quickflow ceases.

Various techniques are available to separate baseflow from gauged streamflow data. These include graphical and automated procedures. The advantage of an automated technique is that it provides an objective, repeatable estimate of baseflow that is comparable over time and between locations. The absolute magnitude of baseflow at individual sites may vary from estimates derived using different separation techniques. Studies which require accurate estimates of baseflow magnitude should therefore be supplemented with detailed at-site investigations of both aquifer and streamflow characteristics, where available.

2.2.   Selected baseflow separation technique

Historically, most baseflow separation approaches have been developed and applied to daily streamflow data. As the focus of ARR Project 7 is on flood events, it was necessary to identify a method that is suitable for the analysis of hourly streamflow data. A number of different baseflow separation methods were reviewed and trialled at several case study locations to evaluate the suitability of techniques using hourly data.

The outcomes of this testing process identified that the most plausible baseflow hydrographs were produced when the Lyne-Hollick (1979) filter was applied using 9 passes across the hourly data with a filter parameter value of 0.925. Based on analysis of flood series from case study catchments, this method produced a plausible baseflow hydrograph for a range of event sizes at seven of the eight case study catchments. The results obtained at one case study site in Western Australia were considered less plausible, however regionalisation tasks in Phase 2 of the project were considered to present an opportunity to consider the unique characteristics of this region in more detail.

2.3.   Data analysis and comparison at case study locations

At each case study catchment location, analysis was undertaken to extract flood events using the peaks over threshold approach. In order to capture a full spectrum of flood event sizes, the number of events extracted from the streamflow series was equal to four times the number of years of streamflow data available. This produced a collection of events that represents the average of four events per annum. The range in event sizes extracted depended on the availability of streamflow data at the particular location, however event ARIs typically ranged between less than 0.5 years to greater than 50 years.

To enable the calculation of statistics for each flood event, it was necessary to identify the start and end of the event. These were defined via an automated approach, so as to minimise the subjectivity associated with the decision and to ensure a consistent approach was applied across all case study catchments and events. This automated process calculated the difference between the baseflow and streamflow series, and identified instances where this difference was minimised. This was combined with a 23 hour moving average of the difference between baseflow and streamflow to prevent any small local minima from being misinterpreted as the start or end of the event. This method was largely successful across the case study sites, and a manual review of the hydrograph for each event was undertaken to confirm the outcomes. In a few instances, manual manipulation of the start and/or end dates was necessary to achieve a more technically correct definition of the event.

For each event identified at each of the case study locations, baseflow related statistics were calculated. These statistics were selected to provide a means to assess baseflow contributions to streamflow events, and include:

- ratio of the peak baseflow to the peak streamflow (termed Baseflow Peak Ratio in Phase 2 of the study);

- the ratio of the event baseflow volume to the event streamflow volume, equivalent to the event BFI (termed Baseflow Volume Ratio); and

- ratio of the baseflow under the streamflow peak to the peak streamflow (Baseflow Under Peak Ratio).

The outcomes from the case study analysis demonstrated that the baseflow contribution to the total flood peak varies depending on event size and location. The proportional contribution of baseflow to the event peak tended to decrease as the magnitude of the total flood event increases. This trend was observed for all three of the baseflow measures noted above.

The variability associated with the estimates of baseflow relative to the total streamflow generally decreased with ARI both at individual sites and between sites.

The findings of Phase 1 of ARR Project 7 summarised in Sections 2.1 to 2.3 above are captured in a technical report (SKM, 2009).

2.4.   Data collation and catchment characteristics

Phase 1 of the project also separately involved the selection of catchments across Australia to be included in the study. A number of criteria were established for this task, including the absence of regulating structures and diversions, and the availability of hourly streamflow data for over 20 years.

Data sources were again contacted and requested to confirm that the selection criteria were met. In addition, advice was sought on any specific sites which exhibited unusual hydrograph features. As a result of this correspondence, a small number of sites were removed from the analysis.

A further check of catchment impairment was conducted using a list of the referable dams of Australia (Boughton, 1999). Catchments which were found to contain significant dams affecting the streamflow were removed from the analysis. Clarification was sought from the relevant agencies when required.

Excessively large catchments were also eliminated from the analysis. A catchment area limit of 10,000 km^2 was used since the eventual outputs of the project relate to rainfall based design and it is less likely that this approach will be applied over such large areas.

As a result of this process, the site selection phase identified 236 catchments that fulfilled the criteria for this study. The location of these catchments is presented in Figure 7. Hourly streamflow data for each of these sites was obtained from the relevant agency as described in the data collation report (SKM, 2010).

The hourly streamflow data was sourced from the relevant agency and reviewed in light of the appropriate data quality codes. Periods of poor quality data were eliminated from the data set. At the conclusion of this process, a timeseries of suitable streamflow data, in m^3/s, was obtained for each catchment.

In addition, a range of catchment, hydrologic and climate characteristics were identified as potential drivers for regional variation in baseflow. These characteristics include information relating to streamflow, climate, vegetation cover, soil type, geological conditions and topography. The key features of interest and the source data sets are summarised in Table 2.

Details of each of these characteristics were extracted for the selected catchments for application in Phase 2 of the study.

Further details about the streamflow and catchment characteristics data sets and the extracted characteristics are provided in a database and supporting report (SKM, 2010).

Table 2 Catchment characteristics

Characteristic	Statistics	Units	Source
CLIMATE CHARACTERISTICS			
Precipitation	Mean annual precipitation	mm/yr	BOM mean monthly and mean annual rainfall data. Climatic Atlas of Australia (BOM, 2000).
	Minimum annual precipitation	mm/yr	
	Maximum annual precipitation	mm/yr	
	Range (difference between minimum and maximum)	mm/yr	
	Standard deviation	mm/yr	
Evapotranspiration	Mean annual evapotranspiration	mm/yr	BOM mean monthly and mean annual evapotranspiration data. Climatic Atlas of Australia (BOM, 2001).
	Minimum annual evapotranspiration	mm/yr	
	Maximum annual evapotranspiration	mm/yr	
	Range	mm/yr	
	Standard deviation	mm/yr	
CATCHMENT CHARACTERISTICS			
Catchment Area	Catchment Area	km^2	Catch 2000 and existing SDL and National Land and Water Resources Audit catchments
Location	Latitude of catchment centroid	Degrees	GEODATA 9 second DEM version 3 (Geoscience Australia)
	Longitude of catchment centroid	Degrees	
Elevation	Maximum	m	GEODATA 9 second DEM version 3 (Geoscience Australia)
	Minimum	m	
	Mean	m	
	Range	m	
	Standard deviation	m	
Slope	Maximum	Degrees	GEODATA 9 second DEM version 3 (Geoscience Australia)
	Minimum	Degrees	
	Mean	Degrees	
	Range	Degrees	
	Standard deviation	Degrees	

Characteristic	Statistics	Units	Source
Aspect	Proportion of catchment facing north, south, east and west	% of catchment	GEODATA 9 second DEM version 3 (Geoscience Australia)
Woody vegetation	Proportion of catchment with woody vegetation	% of catchment	Forest extent and change (v4), Department of Climate Change
Soil type	Average soil depth across catchment	m	Digital Atlas of Australian Soils, BRS and CRC Catchment Hydrology interpretation. (CRC for Catchment Hydrology, 2004)
	Average plant available water holding capacity across catchment	mm	
	Top soil layer thickness	m	
	Top soil layer saturated hydraulic conductivity	mm/hr	
	Top soil layer saturated volumetric water content	m	
	Top soil layer nominal field water capacity	m	
	Top soil layer nominal wilting point water capacity	m	
	Lower soil layer thickness	m	
	Lower soil layer saturated hydraulic conductivity	mm/hr	
	Lower soil layer saturated volumetric water content	m	
	Lower soil layer nominal field water capacity	m	
	Lower soil layer nominal wilting point water capacity	m	
Geology Relevant geological classifications identified, refer to comments below	Number of stream junctions per geology classification	Number/km^2	Surface geology of the states of Australia 1:1,000,000 scale, prepared by Geoscience Australia. Geological classifications based on interpretation as discussed below.
	Reach length intersecting each geology classification	m	
	Stream density in each geology classification	m/km^2	
	Area of each geology classification within each catchment	m^2	
	Percentage of catchment area intersecting each geology classification within a catchment	% of catchment area	
	Weighted average conductivity based on proportion of catchment with each geology classification	m/day	
	Weighted storage ranking based on proportion of catchment with each geology type	% of aquifer volume	
	Weighted average conductivity based on proportion of river reach intersecting each geology classification	m/day	
	Weighted storage ranking based on proportion of river reach intersecting each geology type	% of aquifer volume	

Characteristic	Statistics	Units	Source
Impervious area	Area of catchment in urban area Proportion of catchment in urban area	m^2 % of catchment area	Built up areas dataset, GEODATA TOPO 250k series 3 topographic data, GeoScience Australia
STREAM CHARACTERISTICS			
Stream length	Stream length	m	GEODATA TOPO 250k series 3 topographic data, GeoScience Australia
Stream frequency	Number of stream junctions per unit catchment	Number/km^2	GEODATA TOPO 250k series 3 topographic data, GeoScience Australia
Stream density	Stream length within a catchment Stream length per unit catchment area	m m/km^2	GEODATA TOPO 250k series 3 topographic data, GeoScience Australia

The distribution of catchment sizes associated with the selected catchments is shown in Figure 6. Note that a non-linear axis scale is applied in this figure to capture the broad range of catchment areas included in the data set.
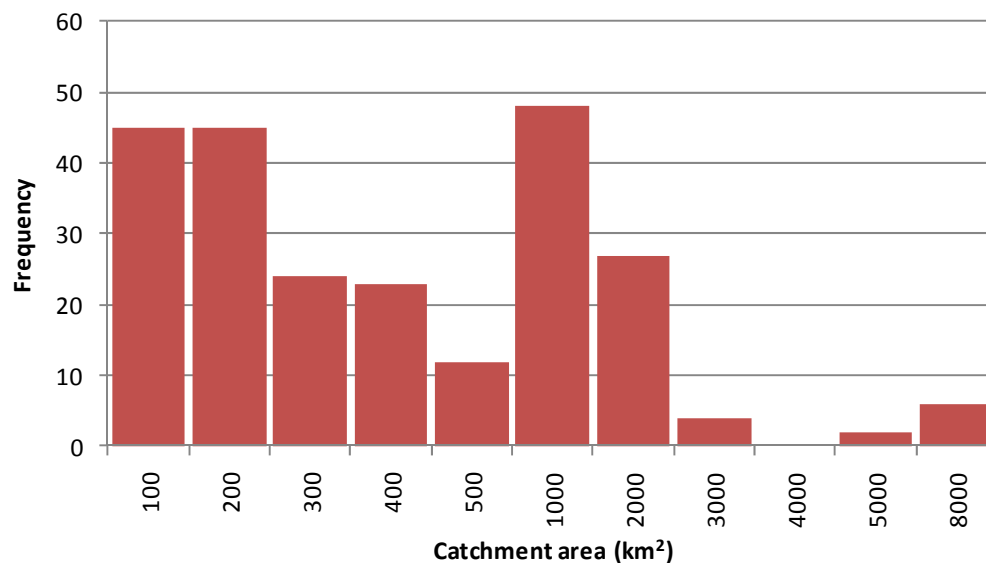

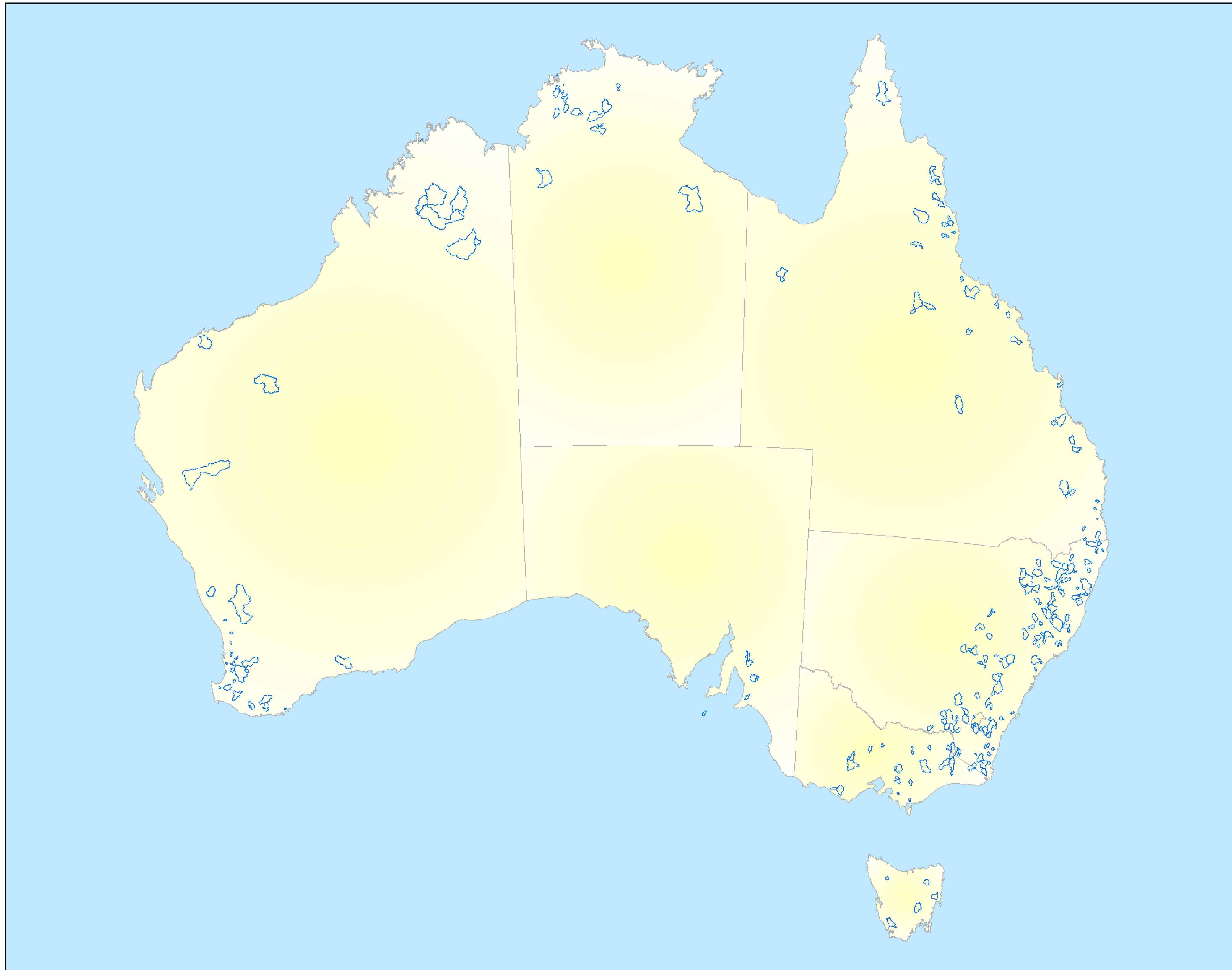**Figure 6 Distribution of catchment sizes of selected sites**

SKM

Legend

☐ Selected Catchments

Figure 7 - Catchment boundaries of selected sites

SINCLAIR KNIGHT MERZ

I:\VWES\Projects\VW04648\Technical\Spatial\ArcGIS\Fig2-Selected_Sites.mxd. Date: 06/09/2010

0   500
Kilometres

2.5.   Application of Phase 1 findings

The approach developed and trialled in Phase 1 of the study can be summarised into a number of key steps, outlined below:

1) Preliminary data review
2) Analysis of events for each case study catchment
    ■ Check and review of hourly streamflow data;
    ■ Separate baseflow from the hourly streamflow series;
    ■ Identify 4N events from the hourly streamflow series;
    ■ Estimate event start and end;
    ■ Generate baseflow statistics for each event; and
    ■ Identify Average Recurrence Interval (ARI) of each event.
3) Analysis at the catchment scale
    ■ Consolidate event results and baseflow statistics for each case study catchment;
    ■ Determine relationship between baseflow statistics and ARI at each case study catchment; and
    ■ Predict baseflow statistics for a given ARI.

Through Phase 2 of Project 7, the outcomes from these tasks were reviewed, revised where necessary, and replicated across the complete data set of 236 catchments. Extending on this analysis, Phase 2 of the project also involved the following additional tasks to develop tools to support the estimation of baseflow contribution to design flood events across Australia:

4) Regionalisation and development of prediction equations
    ■ Consolidate baseflow statistics for given ARI across all study catchments;
    ■ Determine key catchment characteristics that drive baseflow behaviour in various regions of Australia; and
    ■ Generate prediction equations that enable baseflow statistics to be estimated in each defined region.
5) Application of prediction equations across a set of national catchments that cover all areas of Australia
6) Development of a method that utilises the predicted baseflow estimates that is suitable for application in design flood estimation
7) Validation of the method developed to calculate baseflow contribution in design flood events. This validation step tests the predictive capacity of the developed regression relationships. Validation of the approach for providing unbiased flood quantile estimates when applied with flood estimation methods is not considered in this current study.

The following sections of this report outline each of these tasks in more detail.

3. Catchment analysis

3.1. Streamflow data accuracy

When applying an automated process that partitions the baseflow component from a streamflow series, the accuracy of the estimate of baseflow is dependent upon the streamflow data, which itself is a function of:

- the accuracy of the measurement of the water level; and

- uncertainty in the rating curve.

Errors in water level measurements may also show a consistent measurement bias referred to as drift. The other source of water level measurement error is uncertainty in the gauge zero. Standards Australia recommends an uncertainty of ±3mm for uncertainty in gauge zero and between 3mm and 5mm for measured values of certain types of recorders (Lowe, 2009).

The measurement error associated with streamflow gauges is typically in the order of 5-10%, depending on the stability of the channel cross section. For concrete V-notch weirs during low flow conditions, errors in streamflow could be as low as 1% but will increase at high flows. These relatively low errors at low flows indicate that streamflow data accuracy does not inhibit the accurate estimation of baseflow during periods of low flows, other than at streamflow gauging stations with highly variable geomorphic conditions.

A rating curve is produced using the velocity-area method. The main sources of error for gaugings used to generate the rating curve are:

- The subsections the river cross section is divided into to produce the gauging for the velocity area method. Using few sections results in high uncertainties.

- The measurement of the depth of the cross section sections.

- There is a separate uncertainty associated with the current meter used to measure velocity.

- Uncertainty exists in the calculations of the variation in velocity with depth in a river.

- The changes in velocity over short periods of time introduce additional uncertainty to the velocity measurement.

Additional issues exist with the establishment of the stage-discharge relationship for a gauge site. The relationship can be affected by scour, fill or growth of vegetation (Pelleiter, 1988). These errors associated with the rating curve impact all recorded streamflows to some degree. Extremely low or high flows pose additional problems associated with the stage-discharge relationship. In such instances, the flow can be estimated based on extrapolation of the available gauging records. This may be a significant source of uncertainty for

streamflow records associated with large flood events. The infrequency in occurrence of these large floods makes them less likely to be captured in the rating curve. Low flows may have insufficient gaugings caused by low water depth or low velocity that don't allow the use of a current meter (Scanlon, 2007). There are no recognised methods available for estimating the uncertainty associated with these two factors.

Uncertainty in streamflow data varies considerably between sites, depending on (Lowe, 2009):

- The number of historical gauging;

- The measurement error associated with water level measurements; and

- The sensitivity of streamflow to changes in water level.

Errors in streamflow measurements can be expected to vary between ±2% and ±24% (Lowe, 2009).

The quality code for streamflow records should be reviewed before data is analysed for baseflow to better understand the uncertainty in the stramflow records.

3.2. Streamflow data preparation

Streamflow data often contains missing data due to hydrographic equipment failures or events occurring outside of the range for which the streamflow recording equipment has been rated. Equipment failures are to a large extent unavoidable and some missing data is likely to occur in most streamflow records. Typically, baseflow separation requires processing using a continuous data series. This may involve the use of an infilling process to eliminate periods of missing data.

In order to prepare the streamflow data for analysis, issues in the data sets were identified using a number of steps. The data received from each agency was reviewed to identify any obvious problems such as long periods of missing data, a clear change in magnitude or frequency of flow events or other unusual behaviour.

Given the focus of this study is on the analysis of baseflow during flood events, it was considered imperative that the streamflow data being used was at a suitable frequency. In some cases, early years of data (generally before approximately the 1970s, although this varied) consisted of interpolated daily or sub-daily data, which could be identified in the streamflow hydrograph by 'blocky' or 'pointy' events. Periods of data that could be seen to have fewer than approximately eight measurements per day were removed from the datasets.

The data received from each agency was also treated to remove values associated with inadequate quality codes, as defined in consultation with the agency. As each provider uses a different approach to assign data quality, the filtering criteria applied are specific to the source and type of the data. These are specified in the data collation report (SKM, 2010).

In general, streamflow data infilling should be minimised because any baseflow signal identified in the infilled data will generally reflect the infilling technique and not the actual baseflow processes. An example of this is the use of regressions to infill streamflow data. If streamflow data is infilled from an adjacent streamflow gauge, then the baseflow properties of the infilled data will reflect those of the streamflow data used to infill the missing data, and may or may not reflect the behaviour in the catchment itself. Similarly, the use of a rainfall-runoff model will have a pre-defined baseflow recession constant, which may or may not reflect actual baseflow behaviour, depending on the suitability of the model calibration and the range of flows over which it is calibrated. If any infilling of data is undertaken, consideration should be given to the likely compatibility of baseflow properties between the raw data at the site of interest and the infilled data.

There can be advantages in infilling streamflow data in baseflow studies if only short periods of data are missing. Having a complete record allows baseflow statistics to be prepared for different seasons using a comparable length of record. Neal et.al. (2004) adopted a maximum extent of infilled data of 5% of the record for use in regional baseflow assessment, which is a reasonable guide for local investigations as well.

For this study, periods of missing data were infilled using linear interpolation to produce a continuous time series for the separation of baseflow. These infilled periods were subsequently removed from the data series for the analysis stages of the project. This ensured that flood events extracted from the data did not include periods of missing and poor quality data. Essentially, the infilling process was undertaken to ensure that the separated baseflow was not influenced by the boundary conditions associated with the missing periods. The available length of record was then used to confirm that a site was appropriate for analysis, rather than just the period of missing data.

An initial summary of period of record, missing data proportion and extreme streamflow values was also prepared to provide a quick overview of the data provided. This review process was used to pick up readily identifiable issues such as inconsistent units for the data.

In instances where other hydrogeological factors mimic or interfere with the baseflow signal (as discussed in the Phase 1 report; SKM, 2009), baseflow separation should not be attempted without treatment of the data and even then with caution. This can be achieved by either accounting for those upstream influences in the streamflow data at the gauging station location, or by undertaking baseflow separation on inflows between streamflow gauging stations. In both cases, flow monitoring errors are likely to compound and hence there is much scope for variability and uncertainty in baseflow estimates. Baseflow should be relatively stable (or follow an exponential decay function in the absence of catchment rainfall) on successive days, so comparison of baseflow estimates on successive days will give an indication of the uncertainty in those estimates. To attempt to pick up these issues, further checks of the data were made during the analysis process itself, when the streamflow and baseflow hydrographs for each event were visualised. This provided an opportunity to review the data in detail for any further anomalies.

3.3. Baseflow separation

Phase 1 of Project 7 involved the development of a method and selection of parameters to separate baseflow from hourly streamflow time series data. The process of selecting the preferred approach considered the baseflow series produced using combinations of different parameter values and number of passes of the Lyne-Hollick (1979) algorithm shown in Equation 10.

$$q_f(i) = kq_f(i-1) + \frac{[q(i) - q(i-1)] \times (1+k)}{2}, \text{ subject to } q_f(i) \geq 0 \text{ and } q_b(i) = q(i) - q_f(i)$$

Equation 10

Where
$q_b(i)$ = the filtered baseflow response for the ith sampling instant

$q_f(i)$ = the filtered quickflow response for the ith sampling instant

$q(i)$ = the original streamflow for the ith sampling instant for the first pass

k = filter parameter, equivalent to the recession constant

Four plausible combinations were identified in Phase 1. These were the combination of either seven or nine passes of the filter over the hourly streamflow data and parameter values of either 0.925 or 0.95 (Figure 8). Based on visual assessment of hydrographs from a number of sites, the combination of nine passes and a parameter value of 0.925 were determined to produce the most plausible baseflow series. Assessment considered the following features:

- Rise of the baseflow hydrograph – a steep rise in baseflow at the commencement of the streamflow event may signify the inclusion of quickflow in the baseflow hydrograph.
- The timing of the peaks in the baseflow hydrograph – the baseflow hydrograph should peak after the streamflow hydrograph due to the storage-routing of the sub-surface storages.
- The steepness and magnitude of the peaks in the baseflow hydrograph should appear plausible relative to the total streamflow series.
- The baseflow recession behaviour in the log domain – the baseflow hydrograph will most likely follow an exponential decay function (a master recession curve), which should appear linear in the log domain.
- General baseflow hydrograph behaviour in high and low flow periods, including the extent of interflow and quickflow in the baseflow hydrograph.
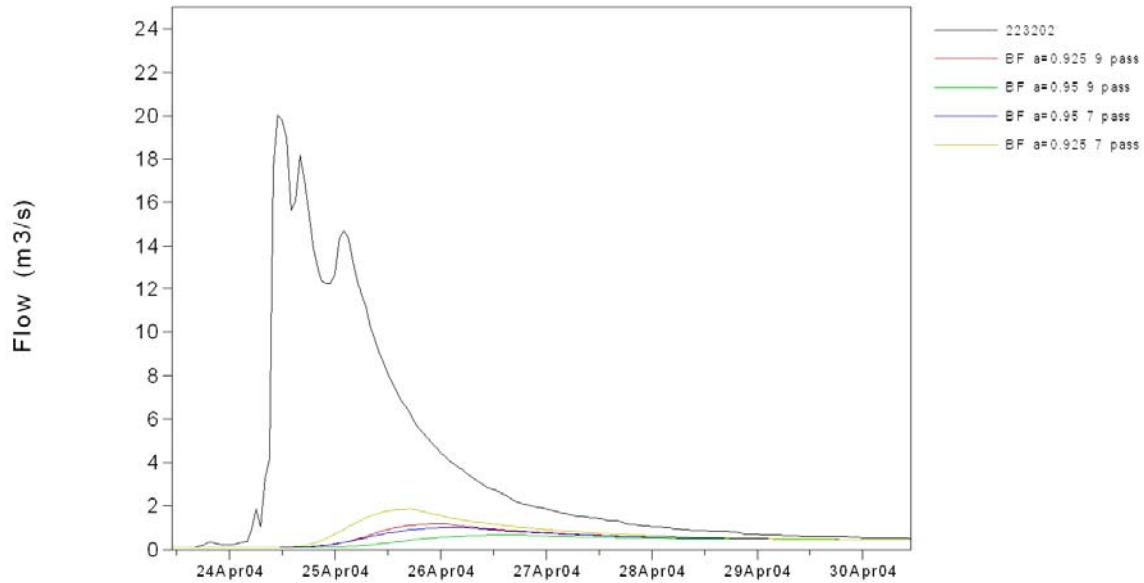
Figure 8 Example hydrograph showing baseflow separation options for site 223202 in Victoria

The selected baseflow separation method was applied to eight case study catchments around Australia in Phase 1 of the study. It was found to give a reasonable representation of baseflow characteristics across most areas. South-western Western Australia was identified as posing challenges due to the unique hydrogeological conditions.

To further assess the suitability of this method, the original four combinations of parameters were re-applied to an additional 13 catchments across Australia in Phase 2 of the study. These locations are presented in Table 3 and Figure 9.

Table 3 Complete set of catchments used in the testing of the Lyne-Hollick separation method

Site Number	Site Name	Catchment Area (km^2)	State	Project Stage
110003A	BarronRiver @ PicnicCrossing	231	Qld	Stage 1 case study catchment
145018A	Burnett Creek @ U/S Maroon Dam	81	Qld	
204041	Orara River @ Bawden Bridge	1637	NSW	
421018	Bell River @ Newrea	1248	NSW	
223202	Tambo River @ Swifts Creek	903	Vic	
229214	Little Yarra River @ Yarra Junction	154	Vic	
803002	Lennard River @ Mt Herbert	441	WA	
606001	Deep River @ Teds Pool	437	WA	
4201	Jordan River @ Mauriceton	744	Tas	Stage 2 method check
143110A	Bremer River @ Adams Bridge	123	Qld	
236212A	Brucknell Creek @ Cudgee	243	Vic	
407253	Piccaninny Creek @ Minto	652	Vic	
410057	Goobarragandra River @ Lacmalac	660	NSW	
410705	Molonglo River @ Burbong Bridge	484	ACT	
608151	Donnelly River @ Strickland	777	WA	
706003	Ashburton River @ Nanutarra	70385	WA	
803003	Fletcher River @ Dromedary	69	WA	
913006	Gunpowder Creek @ Gunpowder	1173	Qld	
A4260504	Finniss River @ 4km East Of Yundi	191	SA	
G8140001	Katherine River @ Railway Bridge	1064	NT	
G8170066	Coomalie Creek @ Stuart Highway	84	NT	

The application of these parameters to a wider range of sites confirmed that the approach could be considered reasonable to apply across Australia. In visually assessing the separated baseflow series it was concluded that a k-value of 0.925 and the use of nine passes was the most appropriate combination to produce reasonable baseflow hydrographs. This is consistent with the findings of Phase 1 of the study and confirms the preliminary evaluation made in that part of the project.

An automated tool was developed to apply the filter in a simple and rapid manner across the 236 catchments of interest. Using this tool, a baseflow series was generated for each site using the available hourly streamflow data.

Legend

● Gauges
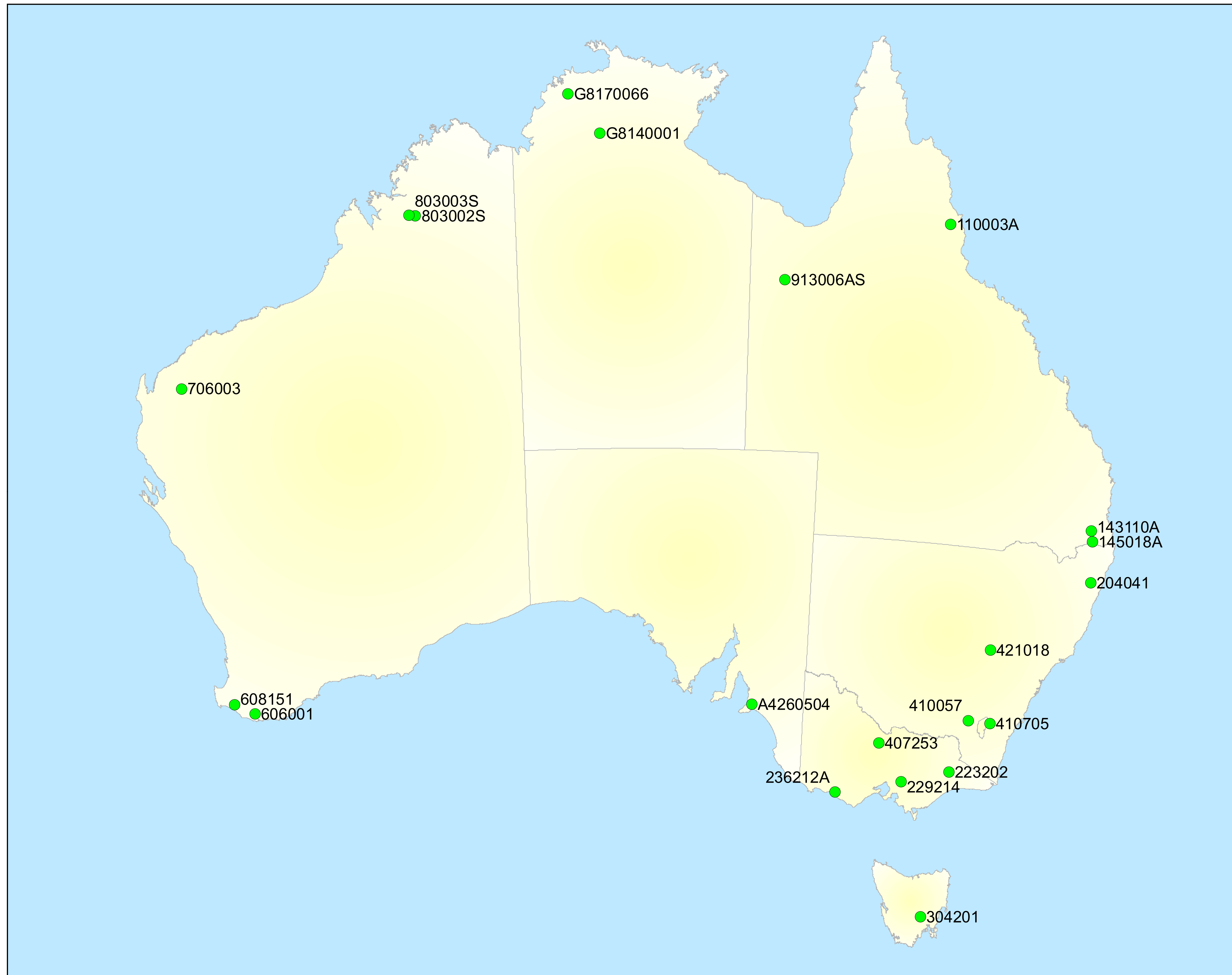
Figure 9 - Location of catchments used in the testing of the Lyne-Hollick algorithm separation method

SINCLAIR KNIGHT MERZ

I:\VWES\Projects\VW04648\Technical\Spatial\ArcGIS\Fig4-Location_of_catchments.mxd. Date: 06/09/2010
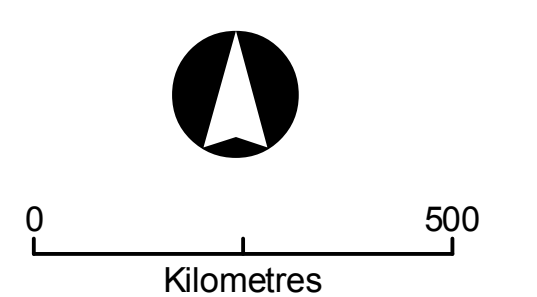
3.4. Identification of flood events

Using the hourly streamflow data at each catchment location, analysis was undertaken to extract flood events using the peaks over threshold approach. In order to capture a full spectrum of flood event sizes, the number of events extracted was equal to four times the number of years of streamflow data available. Independence between events was defined as a minimum interval of 7 days and a minimum difference in the magnitude of successive events of 75%. This analysis produced the date and magnitude of the relevant flood events. Over 30,000 flood events were identified across the 236 catchments.

To enable the calculation of statistics for each flood event, it was necessary to identify the start and end of the events identified in the above process. There is some difficulty and subjectivity associated with this, as identifying the point at which quickflow is assumed to cease is not entirely certain. Even the most sophisticated baseflow separation techniques are constrained by this uncertainty.

For the purposes of this study, the start and end of each event were defined via an automated approach to minimise the subjectivity associated with the decision and to ensure a consistent approach was applied across all case study catchments and events. This automated process calculated the difference between the baseflow and streamflow series and identified instances where this difference was minimised. A moving average with a duration of 23 hours was also applied to the difference between the baseflow and streamflow. Minimums in the moving average were also identified. The start and end of each event was determined by the occurrence of both a local and moving average minimum within a consecutive 23 hour period. The moving average was incorporated into this process to prevent any small local minima from being misinterpreted as the start or end of the event. This approach ensured that instances of slight fluctuations in the streamflow or baseflow data did not trigger the start or end of the event.

Figure 10 demonstrates the outcomes of this approach for a multi-peaked event on the Bell River at Newrea in NSW (streamflow gauge site 421018). It can be seen that each fluctuation in streamflow that occurs during the main component of the event takes place over such a short period of time that the variations in flow do not trigger the criteria for the start or end of the event. This demonstrates the value of the moving average approach to identifying the minimum difference between the streamflow and baseflow series.
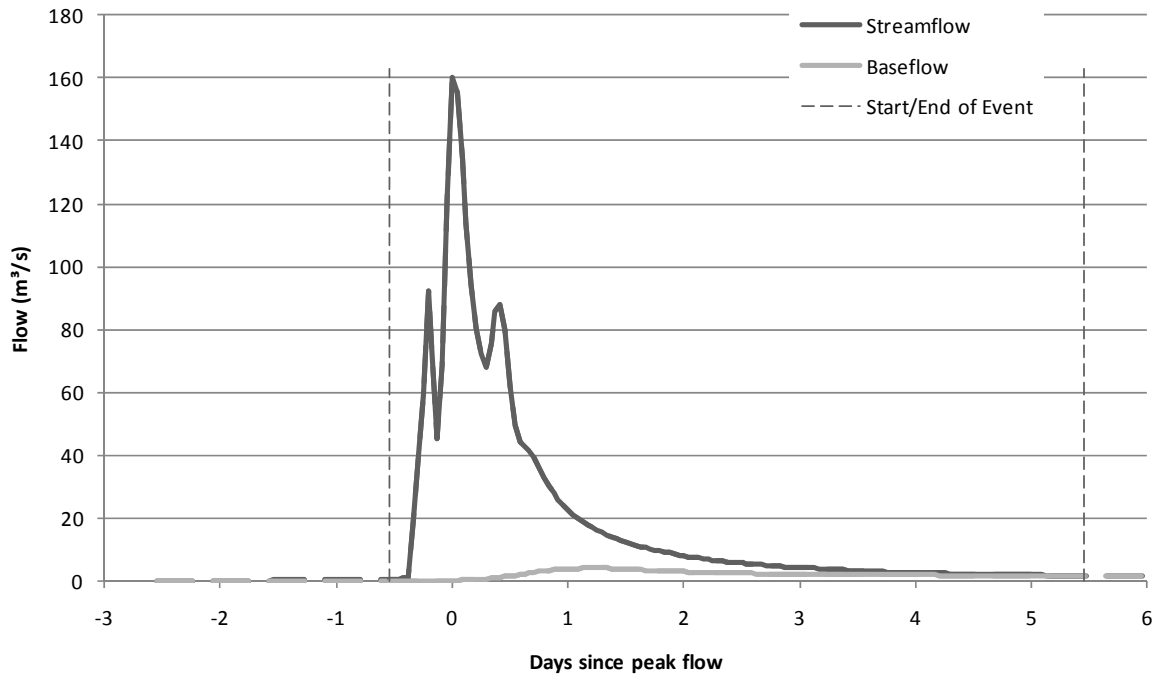
Figure 10 Hydrograph demonstrating the automated approach to baseflow separation and identification of start and end of a multi-peaked runoff event on the Bell River at Newrea, NSW for January 1984

In contrast, fluctuations that occur over longer durations do tend to trigger the criteria for the start and end of events. For example, the automated approach considered the events in Figure 11 to be independent.
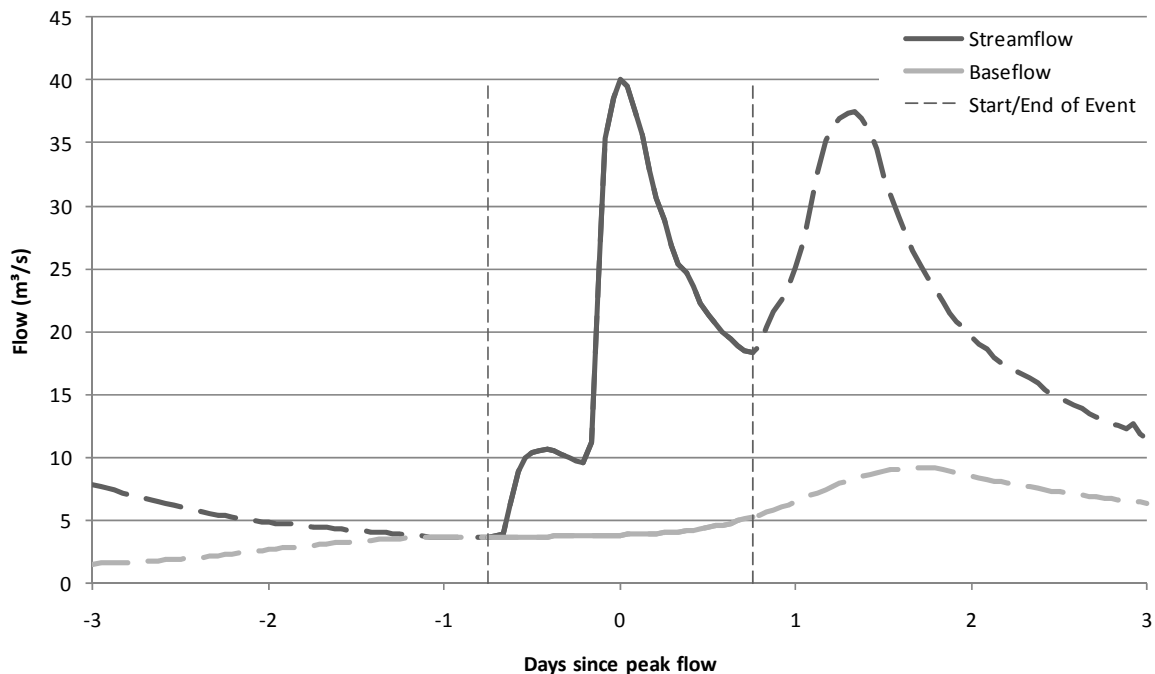

Figure 11 Hydrograph demonstrating the automated approach to baseflow separation and identification of start and end of a runoff event on the Bell River at Newrea, NSW for September 1986

This approach was applied to each of the flood events identified through the peaks over threshold analysis. The method was largely successful across catchments, but manual

reviews of each event hydrograph were undertaken to confirm the outcomes. In some instances, it was determined that a more technically correct definition of the event could be achieved through manual manipulation of the start and/or end dates. In Phase 1 of the study (refer to SKM, 2009), no manual changes were made to alter the automated definitions of event start and end points.

In this phase of the project it was determined that accepting and accounting for any subjectivity was preferable to including automatically defined events that were not realistic portrayals of streamflow behaviour at a site. In addition, given that the contribution of both streamflow and baseflow near the start and end of events tends to be relatively small, particularly in less frequent events, the effect of manually defining the start and end of the event was considered to have minimal influence on parameters calculated for analysis.

Given the quantity of data involved, the process of identifying the start and end of 4N discrete events for all 236 catchments was undertaken by a number of operators. Training and relevant background information was provided to minimise the subjectivity in this task and a series of guidelines were used to ensure consistency between individuals and across sites. The operators had the ability to override the automatic definition of the event start and end. Appropriate start and end times were determined by a visual assessment of the streamflow and baseflow hydrographs for each event.

The guidelines for adjusting the start and end times of events are broadly outlined below:

- An event was considered to start where the streamflow hydrograph separated from the baseflow hydrograph. The event was considered to continue until the conclusion of the runoff recession;
- The occurrence of a significant recession period between two peaks was used to indicate separate events; and
- The occurrence of another peak, however small, within approximately 24 hours of the event was considered to represent a single event with multiple peaks.

The use of these guidelines across sites was considered important since events at different sites often varied significantly in length, peakiness, recession length and time between events.

The application of these criteria was most successful in the less frequent, larger events, since these tend to be discrete and easily distinguishable from any smaller event that may occur around the same time. In contrast, small events often occur during the streamflow recession of previous events, which leads to the identification of an event that actually incorporates multiple smaller peaks.

Figure 11 and Figure 12 show an example of the outcome of the manual review process that was applied. Figure 11 shows an event hydrograph where a local minimum within the event was automatically selected as the end point of the event. In considering the baseflow hydrograph and the occurrence of the recession arm, the event was adjusted as in Figure 12.
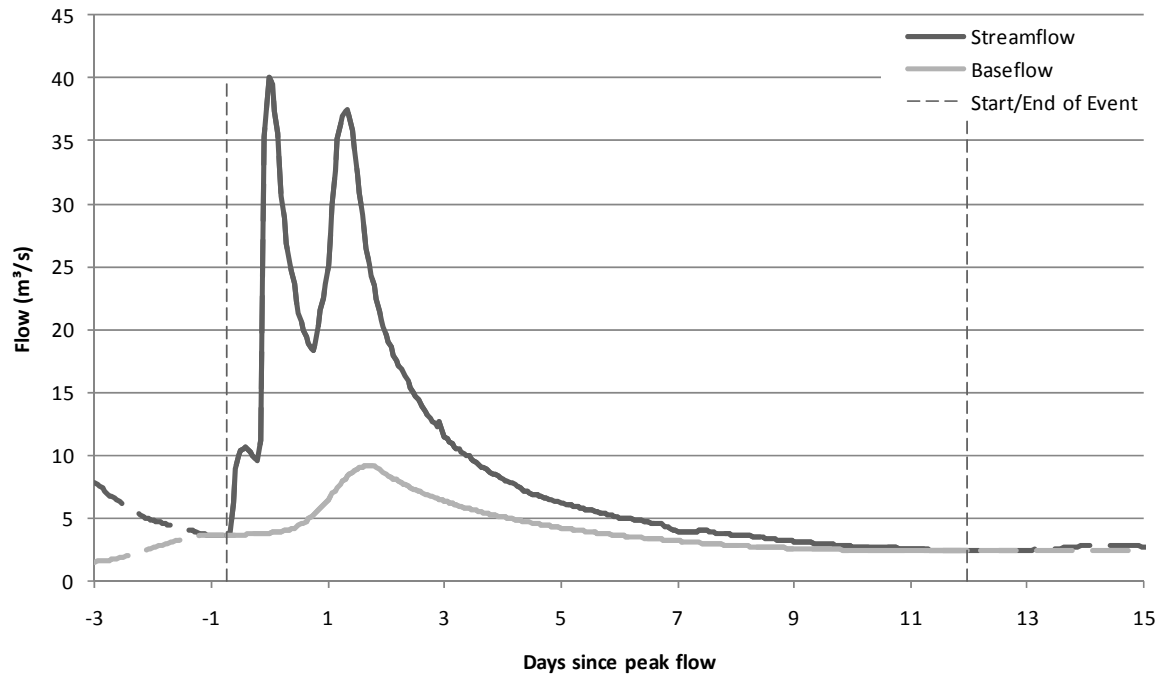
Figure 12 Event start and end adjusted to reflect criteria

Following the completion of this analysis, a review and consolidation process was undertaken by a single operator to provide a consistent overview of the outcomes of the event identification task. This review involved checking each of the 30,000 events. In some cases, only a brief visual assessment was required to confirm that the event was suitably identified. In other cases, the reviewer reassessed and manipulated the start and/or end dates of an event. Events and sites with a high contribution of baseflow were checked to ensure the overall results were reasonable. The review process also identified a small number of sites that exhibited unusual streamflow behaviour, which prompted further investigation into the gauge location and characteristics.

This final manual review of all event hydrographs ensured overall consistency in the definition of events initially analysed by a number of operators.

Further details and comparison of this operator influence is provided in Appendix A.

3.5.   Flood frequency analysis

In Phase 1 of the study, the Generalised Pareto Distribution was fitted to the total streamflow peak for each event by L-moments to identify the Average Recurrence Interval (ARI) associated with each flood. The Generalised Pareto Distribution works well for larger events that occur infrequently, however does not provide an adequate distribution for more frequent events. Other methods that work well for frequent events do not tend to fit well to rare events. Given the range of event magnitudes captured in this analysis, it was considered necessary to re-evaluate the flood frequency distribution approach to ensure that the full spectrum of event sizes was adequately fitted.

A tailored method was developed to determine the ARI for total streamflow for each event. This approach used three different functions applied based on event magnitude to fit a distribution to the data. Maintaining consistency with the method applied in Phase 1, a Generalised Pareto Distribution was fitted to events with ARI greater than one year. The N events (where N = years of streamflow data record) with greatest magnitude were identified for this purpose. A second order polynomial was fitted to events with ARI less than one year (the remaining 3N events). A sixth order polynomial was fitted to the quantiles derived from the Pareto and second order polynomial.

A final distribution was developed by combining these three functions such that:

- The Generalised Pareto fit was applied to events having an event ARI greater than three years.
- The second order polynomial was applied to events having an event ARI smaller than one year.
- The sixth order polynomial was applied to events with an ARI between 1 and 3, as a flexible interpolation to link the other two distributions.

This combined approach generated a distribution that produced a reasonable fit to both the frequent and rare events extracted at each site. Figure 13 presents an example flood frequency distribution displaying the application of this approach, and equivalent plots are provided for all catchments in Appendix B.
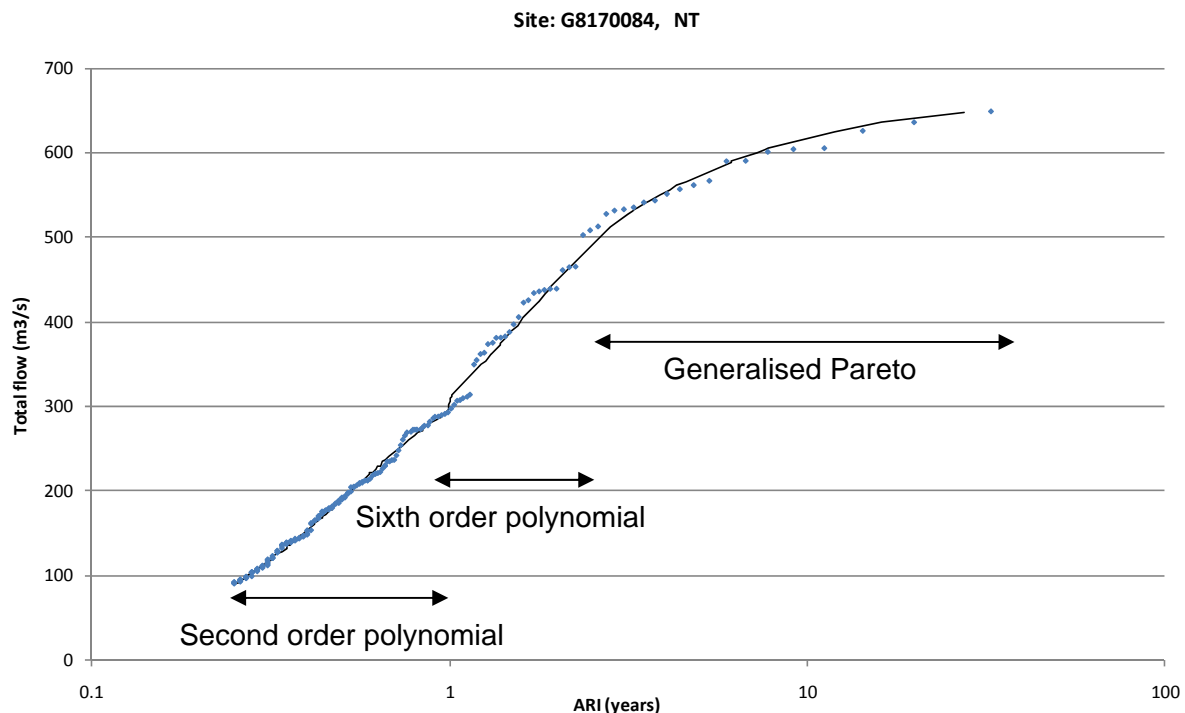
Figure 13 Flood frequency distribution for G8170084 (Adelaide River at Tortilla Flats, NT) displaying the application of the combined approach to fit a flood frequency distribution

This project provides information that is applicable for a range of flood event magnitudes, ranging from an ARI of 0.5 years to an ARI of 100 years. This measure of flood frequency does not assume that a particular event will definitively occur with a frequency as designated by the ARI, but rather provides a probabilistic estimate of the average recurrence interval of a particular sized flood event. Seasonality can influence the timing of flood events, however this has not been incorporated in the analysis of this study.

3.6. Characterising baseflow contribution to flood events

For each event identified at each site, a range of baseflow related statistics were calculated. These statistics were selected to provide a means to assess baseflow contributions to streamflow events. Considering the example hydrograph presented in Figure 14, the statistics of interest include:

1. **Baseflow Peak Ratio**: Ratio of the peak baseflow (C) to the peak streamflow (A), given by C/A.

2. **Baseflow Volume Ratio**: The event baseflow index (BFI), which is given by the total baseflow volume for the duration of the event divided by the total streamflow volume. This is the ratio of the shaded areas in the example hydrograph.

3. **Baseflow Under Peak Ratio**: Ratio of the baseflow at the time of the streamflow peak (B) to the peak streamflow (A), given by B/A.

The timing associated with these key features was also collated for each event. The statistics above were extracted for flow events at each site. It is important to note that these statistics reflect the baseflow contribution relative to the total flow event (ie: runoff + baseflow).
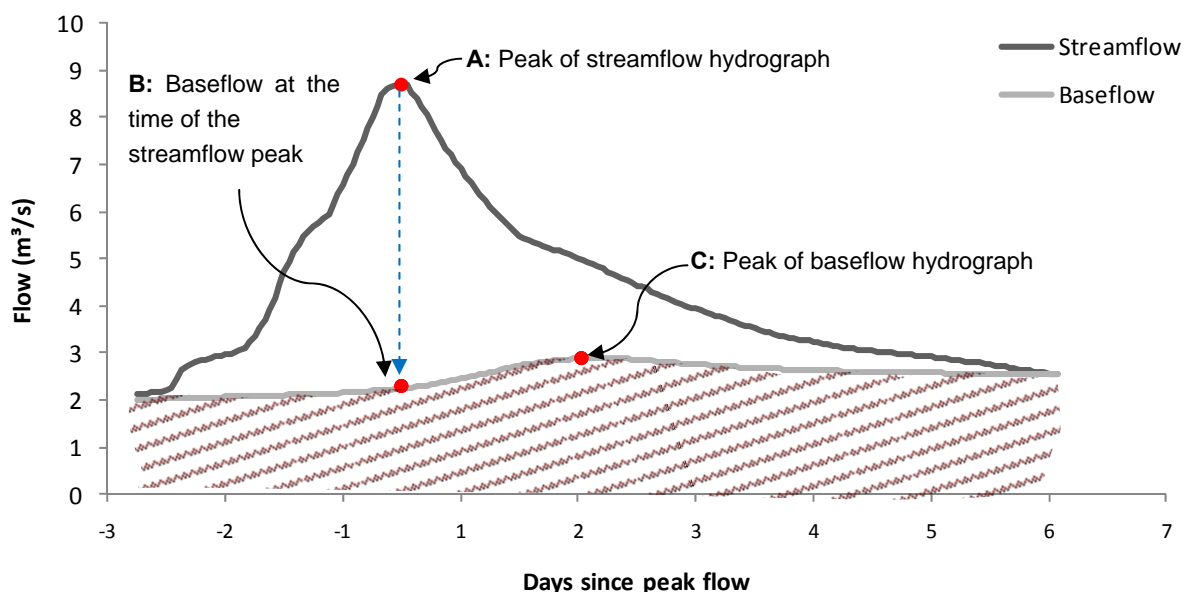
Figure 14 Flow hydrograph for an event at Victorian site 229214 (Little Yarra River at Yarra Junction) displaying key features

Of particular relevance to the Baseflow Peak Ratio and the Baseflow Under Peak Ratio is the hydrograph behaviour prior to the event itself. This behaviour reflects the streamflow and baseflow recessions from previous rainfall events, which combines with the response during the current event to influence the total streamflow and baseflow behaviour. To some degree, the difference between the baseflow peak and the baseflow under the peak is a reflection of the relative contribution of the current and prior event characteristics in characterising the baseflow.

The Baseflow Volume Ratio represents the relative contribution of baseflow to total streamflow on a standardised basis for comparison between sites along a river with different upstream catchment areas and sites in different catchments. For a particular event, the BFI is given by the ratio of the lower shaded area to the total shaded area in the example hydrograph in Figure 14.

Whilst long-term average baseflow index is generally regarded as an indicator of hydrogeologic conditions, it is important to remember that baseflow index is a measure relative to total streamflow. Over short time intervals, baseflow index will generally reflect fluctuations in quickflow rather than changes in baseflow. During extended dry periods the baseflow index will be equal to 1.0, whilst during major flood events, the baseflow index may be close to zero. This is important to note for the purposes of this study, as different sized flood events will respond with different BFI ratios. As such, the BFI values calculated in this study (Baseflow Volume Ratio) may vary widely from those calculated elsewhere that represent long-term average conditions.

The absolute magnitude of baseflow for given event sizes was considered in Phase 1 of this study. This approach has not been extended in Stage 2 as it was considered to be too heavily dependent upon conditions prior to the flood event itself, which are not captured through this approach. Instead, the ratios described above have been analysed as these consider the baseflow characteristics based upon standardised behaviours.

4. Analysis Summary

A significant amount of data was generated from the catchment analysis tasks. This information was consolidated for the 4N events across the 236 sites to provide a collection of data to be used for further analysis and development of prediction equations.

4.1. Average Recurrence Interval

The ARI for each event was determined as described in Section 3.5 and used to produce a flood frequency distribution characterising the total flow regime (surface runoff and baseflow) for each catchment. Across the country, the range in flow magnitudes vary widely, making it difficult to compare these flood frequency curves directly. By standardising the flow using catchment area to produce a curve that reflects depth of runoff, the flood frequency distribution for each of the 236 catchments can be compared. Figure 15 presents this information. Due to the large number of catchments analysed in this study, it is difficult to identify the curve for any given catchment within this figure. However, a number of general observations can be made:

1. There is variation in the shape of the flood frequency curves. Some catchments show changes in flow magnitude with ARI, particularly for smaller event sizes. Other catchments appear to have flow magnitude that is less variant with ARI.

2. With the exception of a small number of catchments, the flood frequency curves tend to be grouped together. The event sizes that relate to given ARIs tend to lie within a range of approximately two orders of magnitude.

3. Across all catchments, the magnitude of the smallest events extracted (reflecting an ARI of around 0.2 years) ranges between approximately 0.0001 m^3/s per unit area (in km^2) and 2 m^3/s per unit area (in km^2).

4. For an ARI=100 years, events range in magnitude between 0.1 m^3/s per unit area (in km^2) and 50 m^3/s per unit area (in km^2).

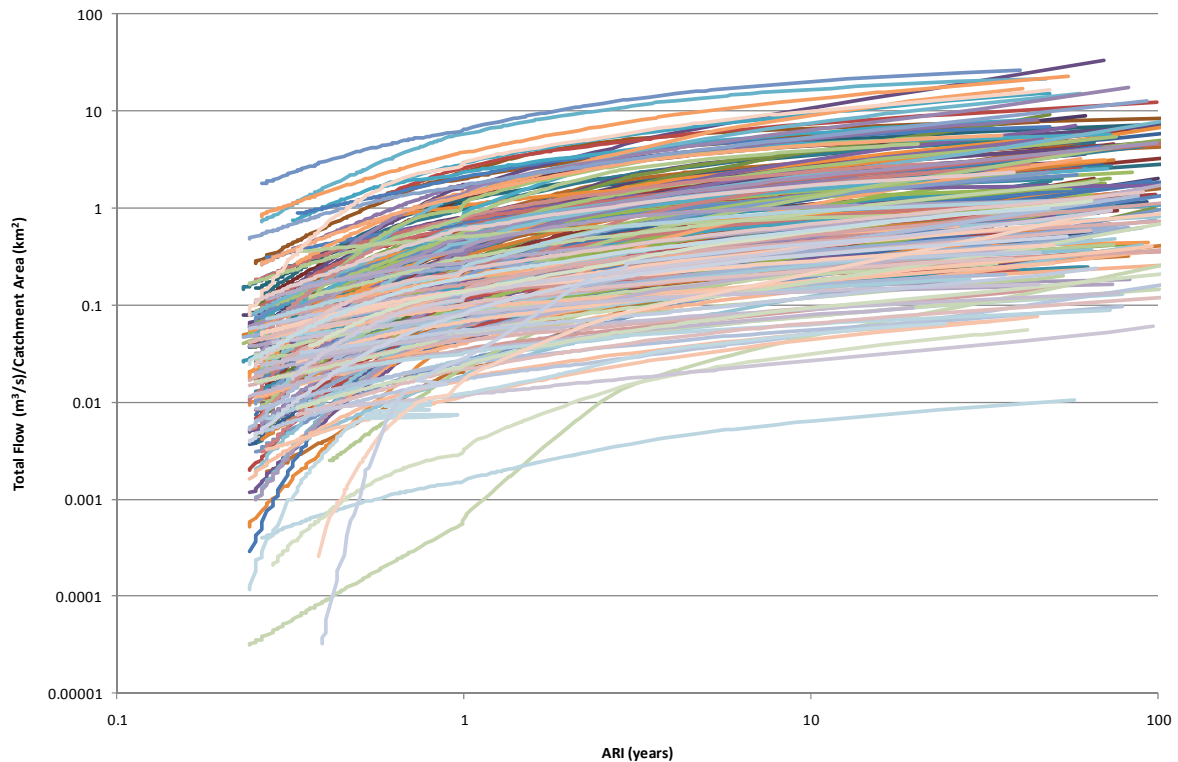5. Other catchments tend to display sharp reductions in event magnitude for the smaller ARIs.

Figure 15 Flood frequency distribution, standardised by catchment area, for study catchments

The flow peak for specific ARIs was interpolated from this flood frequency information for further analysis. ARIs calculated included:

- 0.5 years
- 1.25 years
- 2 years
- 5 years
- 10 years

Using this information, the relationship between ARIs and catchment area was investigated across the collection of catchments. This comparison is presented in Figure 16 for ARI=2 and ARI=10. While there is a high degree of scatter in the data, the comparison confirms that larger catchments typically generate larger flood peaks for a given recurrence interval, and that this behaviour is consistent for different flood event sizes.
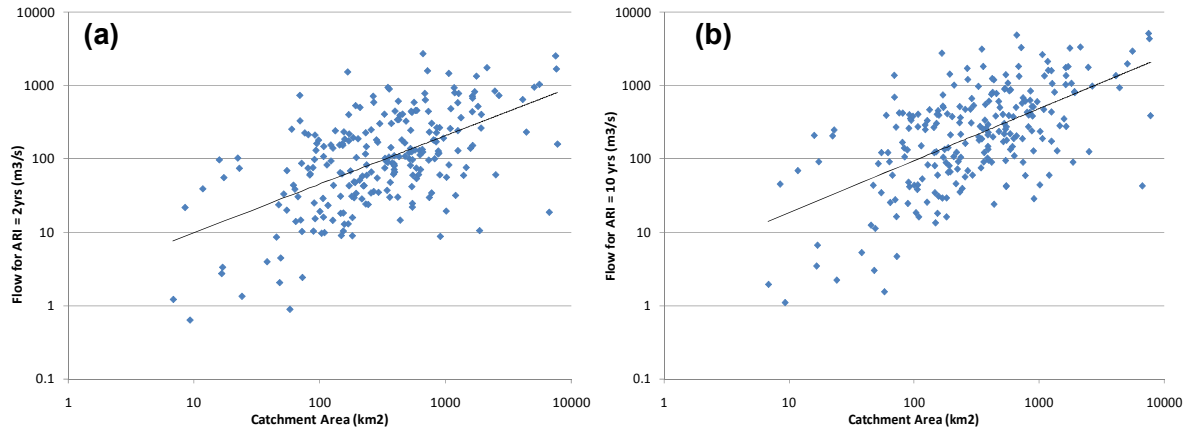
Figure 16 Relationship between catchment area and flow peak for all catchments for (a) ARI=2; (b) ARI=10

Comparing the magnitude of the flood events for different ARIs provides further information on the flow regime at each catchment. For example, the ratio of flow for ARI=0.5 to flow for ARI=10 is presented in Figure 17a against catchment mean annual rainfall. This shows that the magnitude of frequent events as a proportion of the 10 year event varies between catchments. The variability of the ratio value also increases with the catchment mean annual rainfall. A similar comparison is presented in Figure 17b for the ratio of flow for ARI=5 to flow for ARI=10. In this figure the catchment mean annual rainfall does not appear to have a strong influence on the catchment runoff regime. Across all catchments, events with an ARI=5 tend to have a magnitude of approximately 75% of the ARI=10 year events regardless of the rainfall conditions.
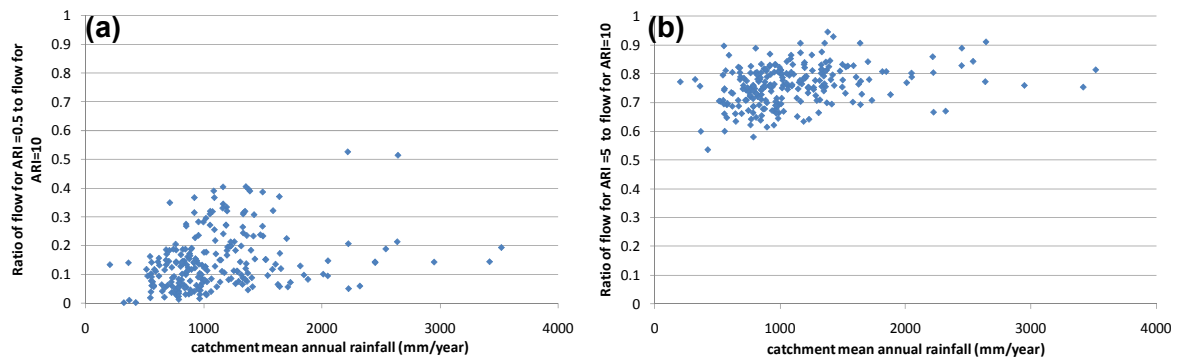
Figure 17 Comparison between flow for (a) ARI=0.5 and ARI=10 and (b) ARI=5 and ARI=10

It was observed that some events extracted at some sites were so small in magnitude that they were not considered to represent a flood event at all. The ratio of very frequent events (such as ARI=0.5 and ARI=1.25) compared to ARI=10 was used to identify catchments where there was a sharp drop in the flood frequency distribution. In these locations, small events were manually excluded from further analysis where the flow for a given ARI was very low compared to the ARI=10 event. This ensured that the events captured in the assessment were 'real' flood events rather than artefacts of the identification process.

4.2.   Understanding the Baseflow Peak Ratio

The value of the Baseflow Peak Ratio (the ratio of the peak baseflow to the peak streamflow) was extracted for each event at all 236 catchments of interest. For each catchment, the ratio values were plotted against the total streamflow ARI to understand the relationship between baseflow contribution and flood size. This analysis is consistent with that undertaken in Phase 1 of the study. An example plot is shown in Figure 18. Similar plots were generated for all catchments. In Figure 18, each point on the chart reflects a single flood event extracted at the site. A high degree of scatter is evident for small events, reflecting the high frequency of occurrence of these events under a variety of conditions. Fewer data points are available for larger event sizes, which is a function of the length of the streamflow record at each location and the infrequent nature of these events. These general characteristics were typical across all catchments.

In the example displayed in Figure 18, the Baseflow Peak Ratio value tends to decrease with increasing event magnitude. This relationship is consistently observed at most catchments considered in Phase 2 of this study. In most locations, the shape of this relationship is best described by a power function, shown as the black solid line through the data points. The slope of the relationship varies between catchments. Some catchments display steep reductions in the Baseflow Peak Ratio with ARI while others are less sensitive to event magnitude. Across the collection of sites, this approach yielded R^2 values up to 0.67, and a median value was 0.16. Appendix C provides an indication of the response observed at each catchment, although should be reviewed in combination with the discussion of Section 6.
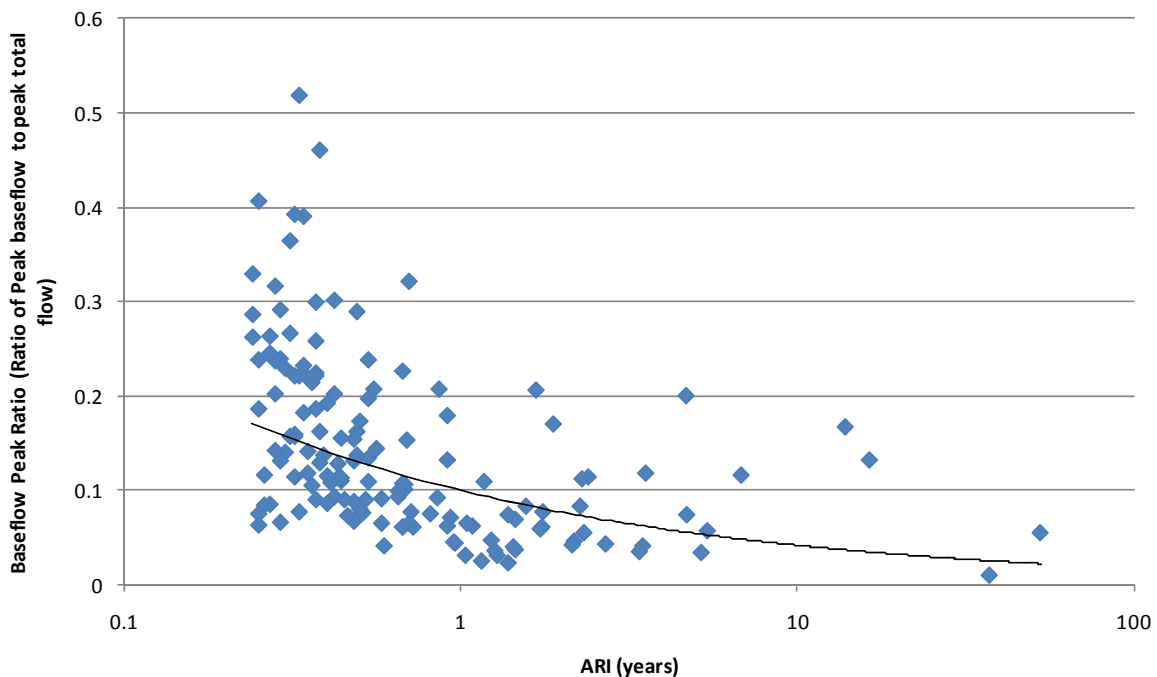

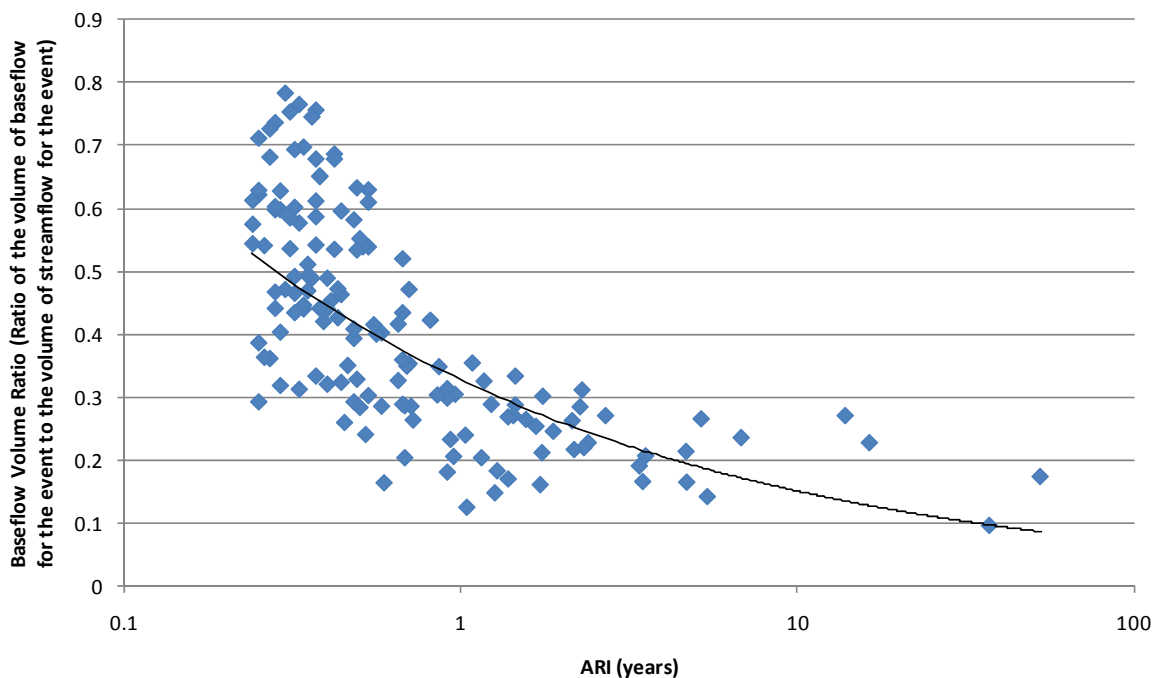**Figure 18 Variation in Baseflow Peak Ratio with ARI of total flow peak for catchment 218001 (Tuross River @ Tuross Vale, NSW)**

Exceptions to this negative power relationship were observed for some catchments where the value of the Baseflow Peak Ratio appeared invariant or showed a slightly increasing relationship with ARI. Since it was considered unrealistic that baseflow contribution would increase for larger flood events, these instances were considered to be a reflection of the period of data available for analysis and the sensitivity of the power function. For example, Figure 19 shows a catchment where the power function is predicted to have a positive slope. Visual inspection suggests that the data generally conforms to the more commonly observed negative relationship between the Baseflow Peak Ratio and ARI. In this case, the density of data points for frequent events with low ratio values is dominating the predicted power function, and the function is not considered to reasonably describe the baseflow behaviour.
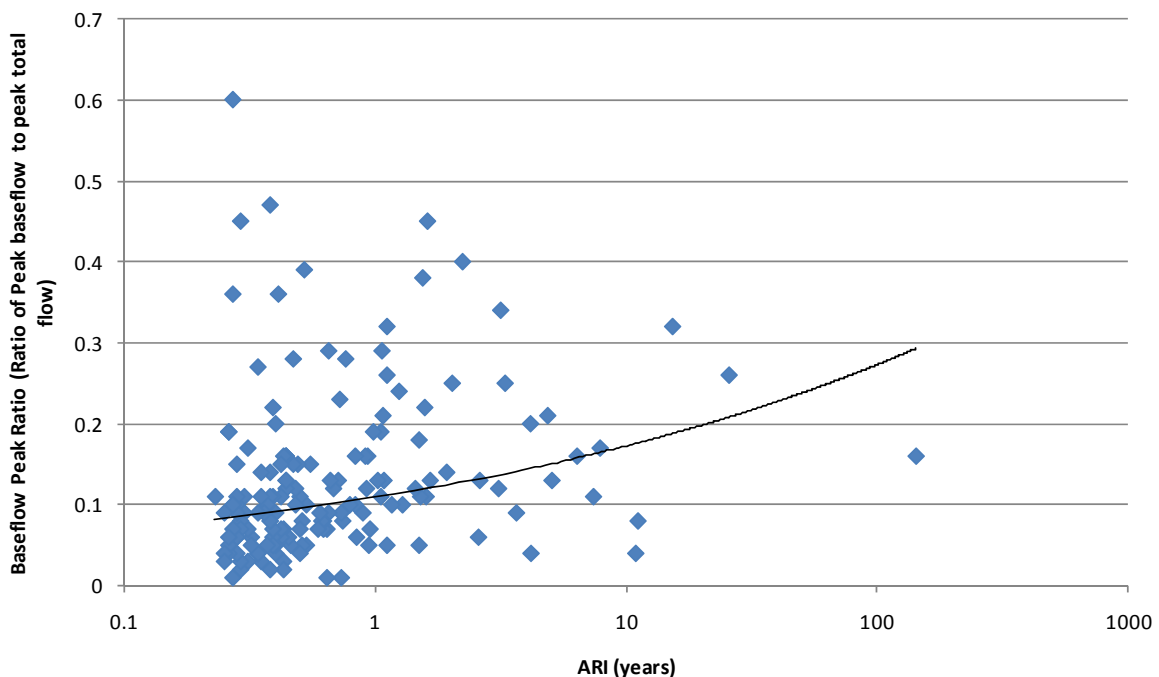
Figure 19 Variation in the Baseflow Peak Ratio with ARI of total flow peak for catchment 120305A (Native Companion Creek @ Violet Grove, Queensland)

To prevent such occurrences from unreasonably impacting on the nationwide assessment, the slope of the power term was investigated in more detail. For each catchment, the Baseflow Peak Ratio value for an event with an ARI of 5 years was compared to the Baseflow Peak Ratio value for an ARI of 10 years. Where the reduction in the ratio value was less than five percent (i.e. the slope of the power function was close to zero or had positive gradient), it was assumed that the power function was not suitable to describe the baseflow behaviour with increasing ARI. These instances tended to have poor R^2 values for the power trend. Instead, the mean Baseflow Peak Ratio value was calculated for the catchment and used to describe the baseflow contribution regardless of event size. This approach was applied to 37 catchments located in Queensland, Victoria, Western Australia and Northern Territory.

4.3.   Understanding the Baseflow Volume Ratio

The volume of baseflow relative to the volume of streamflow (Baseflow Volume Ratio) was calculated for each event. Similar to the analysis undertaken for the Baseflow Peak Ratio, this information was collated for each catchment and the trend with ARI was identified. Figure 20 presents an example plot showing the variation in Baseflow Volume Ratio with event size for the Tuross River catchment in NSW. A power function has been fitted to the data points to describe the baseflow contribution to different sized flood events for each catchment.

Figure 20 Variation in the Baseflow Volume Ratio with ARI of total flow peak for catchment 218001 (Tuross River @ Tuross Vale, NSW)

As observed for the Baseflow Peak Ratio, the Baseflow Volume Ratio data did not always fit well to a power function. Consistent with the approach described for the Baseflow Peak Ratio, the average ratio value was used to describe the baseflow behaviour for catchments that appeared invariant or showed a slightly increasing relationship with ARI. A change in ratio value of less than 5% relative to the value for an ARI of 10 years was used to define when the mean ratio value would be applied. Instances of this behaviour were infrequent. Appendix D displays the resulting relationship for each location. As for the Baseflow Peak Ratio, these figures should be considered in light of the information provided in Section 6.

4.4.   Understanding the 1 in 10 year event

For this study, the 1 in 10 year ARI event has been used for preliminary interpretation and analysis of the data. This event size was selected since it corresponds to a reasonable flood magnitude which is well represented within the streamflow data record. Events which are

larger than this may occur infrequently in the available streamflow records and smaller events may occur too frequently to draw meaningful conclusions.

The Baseflow Peak Ratio value for an ARI = 10 years is shown for each study catchment in Figure 21. Most of the sites have a Baseflow Peak Ratio value less than 0.2. There is little variation in the magnitude of the ratio for many of the catchments in eastern Australia, with the exception of some of the alpine catchments in Victoria. Catchments in the south west of Western Australia and the Northern Territory show higher ratio values.

Figure 22 presents the Baseflow Volume Ratio for each catchment. There is much greater variation in this ratio value across the country. Low values tend to occur in NSW and Queensland, while larger values occur in Victoria, Western Australia and the Northern Territory.
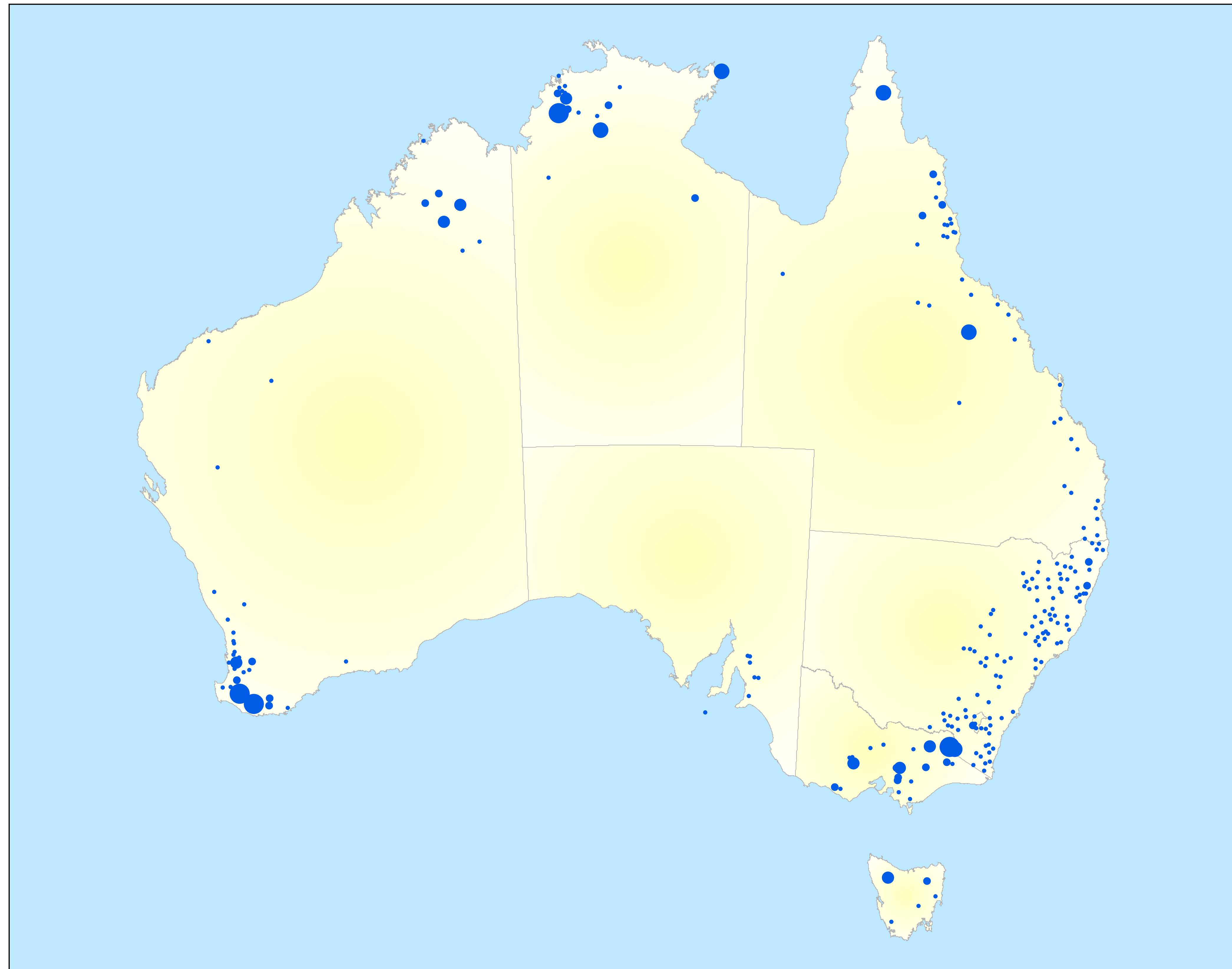
Legend

Peak Ratio
- · 0.00 - 0.15
- • 0.16 - 0.30
- ● 0.31 - 0.45
- ● 0.46 - 0.60
- ● 0.61 - 0.75

■ **Figure 21 - Baseflow Peak Ratio for ARI = 10 years for study catchments**

SINCLAIR KNIGHT MERZ

0   500

Kilometres
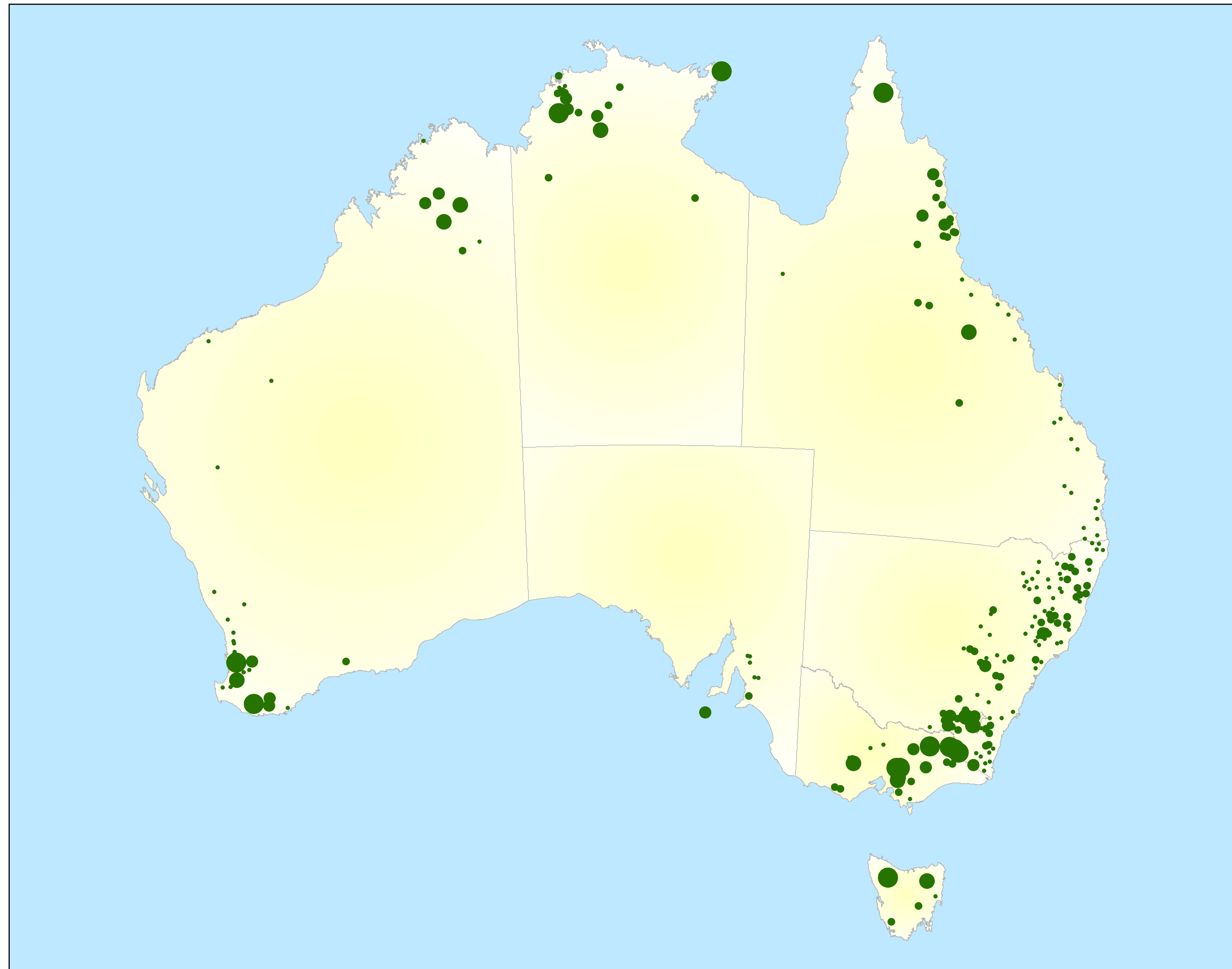
Legend

Volume Ratio

- · 0.00 - 0.15
- · 0.16 - 0.30
- ● 0.31 - 0.45
- ● 0.46 - 0.60
- ● 0.61 - 0.75

■ **Figure 22 - Baseflow Volume Ratio for ARI = 10 years for study catchments**

SINCLAIR KNIGHT MERZ

0   500

Kilometres

5. Development of the prediction equations

The development of a system of prediction equations aims to allow the calculation of the Baseflow Peak Ratio and Baseflow Volume Ratio for an ungauged catchment based on catchment characteristics. This section of the report describes the development of a model comprising a regression tree and regression equations that can be used to estimate the baseflow contribution to design flood events for any catchment across Australia. This section summarises the techniques used, details of the derived relationships and the accuracy of the model.

5.1. Overview of method for fitting prediction equations

A total of 236 streamflow gauge records were analysed to determine the baseflow behaviour in catchments across Australia, as outlined in the previous sections. To enable the prediction of baseflow behaviour for ungauged catchments, catchment characteristics for the study sites were used to predict the value of the Baseflow Peak Ratio and the Baseflow Volume Ratio at the 10 year ARI.

The two techniques used to create a predictive model were regression tree analysis and multiple linear regression. Regression tree analysis classifies catchments into groups based on catchment characteristics. A regression tree was used to group like catchments on the basis of the Baseflow Peak Ratio at the 10 year ARI. Each cluster was defined by the values of particular catchment characteristics identified by the analysis as significant.

Multiple linear regression produces a continuous linear function for estimating ratio values based on more than one independent variable. A multiple linear regression analysis was conducted on each cluster identified by the regression tree. This produced an equation that could be used to predict the value of the Baseflow Peak Ratio at the 10 year ARI for an ungauged catchment. The value of the Baseflow Volume Ratio at the 10 year ARI was predicted using the same clusters identified using the Baseflow Peak Ratio at the 10 year ARI. A second multiple regression analysis was conducted on the catchments in each cluster using the Baseflow Volume Ratio at the 10 year ARI as the dependent variable. The calculated value for the Baseflow Peak Ratio at the 10 year ARI was used as an independent variable in the Baseflow Volume Ratio regression.

The catchment characteristics previously extracted, as described in Section 2.4, were used to develop the predictive model. As a large number of catchment characteristics were originally extracted for the 236 catchments, some were determined to be less appropriate for use in regression tree and multiple linear regression analyses. Characteristics that related to ranges and standard deviations of a measure were considered to be represented in other available variables and were excluded from the analysis. Characteristics related to aspect were also excluded, as were characteristics that were influenced by the spatial scale over which data was extracted. This included features such as the number of stream junctions and the stream length within a catchment.

The catchment characteristics that were considered as predictive variables for the regressions are listed in Table 4.

Table 4 List of variables considered for use in regression equations

Variable	Abbreviation
Top soil layer nominal field water capacity (m)	A_FCP_MN
Top soil layer saturated hydraulic conductivity (mm/hr)	A_KSAT_MN
Top soil layer saturated volumetric water content (m)	A_SAT_MN
Top soil layer thickness (m)	A_THICK_MN
Lower soil layer nominal field water capacity (m)	B_FCP_MN
Lower soil layer saturated hydraulic conductivity (mm/hr)	B_KSAT_MN
Lower soil layer saturated volumetric water content (m)	B_SAT_MN
Lower soil layer thickness (m)	B_THICK_MN
Average soil depth (m)	SOLDEPTH_MN
Average plant available water holding capacity (mm)	SOLPAWHC_MN
Maximum elevation (m)	ELEV_MAX
Minimum elevation (m)	ELEV_MIN
Mean elevation (m)	ELEV_MN
Maximum annual evapotranspiration (mm/yr)	EVAP_MAX
Minimum annual evapotranspiration (mm/yr)	EVAP_MIN
Mean annual evapotranspiration (mm/yr)	EVAP_MN
Proportion of relevant geology (%): Alluvial - coarse grained (gravels/sands)	PCGEOL_AC
Proportion of relevant geology (%): Alluvial - medium grained (fine to med-grained sands)	PCGEOL_AS
Proportion of relevant geology (%): Alluvial ('general' or undifferentiated- sands, silts, clays or fine-grained)	PCGEOL_AU
Proportion of relevant geology (%): Basalt	PCGEOL_B
Proportion of relevant geology (%): Colluvial	PCGEOL_C
Proportion of relevant geology (%): Fractured sandstone in GAB Basin, WA and Canning Basin, WA	PCGEOL_FSS
Proportion of relevant geology (%): Igneous & metamorphic rocks, conglomerates, mudstones, siltstones, conglomerate, shale, phyllite, chert, BIF	PCGEOL_IM
Proportion of relevant geology (%): Limestone	PCGEOL_L
Proportion of relevant geology (%): Sandstone	PCGEOL_SS
Maximum annual rainfall (mm/yr)	RAIN_MAX
Minimum annual rainfall (mm/yr)	RAIN_MIN
Mean annual rainfall (mm/yr)	RAIN_MN
Maximum slope (degrees)	SLOPE_MAX

Variable	Abbreviation
Minimum slope (degrees)	SLOPE_MIN
Mean slope (degrees)	SLOPE_MN
Proportion of woody vegetation (%)	PCVEG
Weighted average conductivity based on proportion of catchment with each geology type	PC_GEOL_ WeightedConductivity
Weighted storage ranking based on proportion of catchment with each geology type	PC_GEOL_ WeightedStorageRanking

The following sections outline this approach in more detail.

5.2. Development of the regression tree

5.2.1. Regression tree theory

A regression tree classifies data into clusters. It is a directed graph that divides a whole data sample into many subsets. The tree is fitted to data sets that contain continuous dependant variables. Regression trees are an alternative method of prediction which is not based on an algebraic model.

The regression tree is fitted using the Automatic Interaction Detection (AID) method (Morgan and Sonquist, 1963). The complete set of data is split into two clusters by finding the predictor variable and the cut point that minimises the sum of squares around the mean of the dependant variable in the resultant clusters.

The AID method, like other tree-clustering methods, will continue to split the data until an exhaustive tree consisting of one data point per terminal cluster is produced. To produce a tree with more than one data point in each terminal node, a number of restrictions were placed on the splitting mechanism. These included specifying:

- the minimum number of catchments in terminal nodes
- the maximum number of splits allowed
- the minimum proportion reduction in error (PRE) for the tree allowed at any split

The physical justification of each split was also considered to ensure that it was not the result of random correlation.

The success of a regression tree is assessed by the value of the proportion reduction in error (PRE). This is a scale similar to the coefficient of determination (R^2) and allows the comparison of the predictive ability of different regression trees. Additionally, an assessment of the statistics of each cluster gives an indication of the validity of the split. Each cluster is expected to show a mean value of the dependent variable that is different from the mean values of other clusters and a relatively small standard deviation within the cluster.

5.2.2. Baseflow Peak Ratio regression tree

Initial attempts at clustering the data identified a number of catchment characteristics that on first investigation appeared to be proxies suitable for regionalising catchments. Some of the regions in Australia known to have distinct baseflow behaviour were examined to determine if there was any basis for splitting the catchments under these criteria.

Regionalisation options examined were:

- Basins 1 and 2 (eastern coast) compared to the rest of Australia
- Basin 6 (south west Western Australia) compared to the rest of Australia
- Basins 8 and 9 (tropical north) compared to the rest of Australia

There was no justification found for manually separating these regions out to characterise baseflow. It was determined that other catchment characteristics would adequately capture the behaviour of baseflow in these regions.

Instead, a regression tree was used to split the sample into groups of data that are more easily described within the groups than between the groups. Determining whether a tree will be useful in describing the variability in a sample is done by examining the mean and standard deviations of each cluster. When looking at two groups created by a split, the means of each group should be significantly different and the standard deviations should be small. This indicates that the catchments in a group are both sufficiently similar to each other and sufficiently different from the other groups to justify splitting them into groups.

The regression tree ultimately developed for this study was identified using the Baseflow Peak Ratio as the dependent variable. Clusters were split based on the catchment characteristics that can best group catchments into ratio groups. This tree is shown in Figure 23.

Given the strong relationship between the Baseflow Peak Ratio and the Baseflow Volume Ratio (further described in Section 6), it was considered reasonable to use this clustering approach to regionalise the dataset regardless of whether the baseflow peak or volume was being considered.
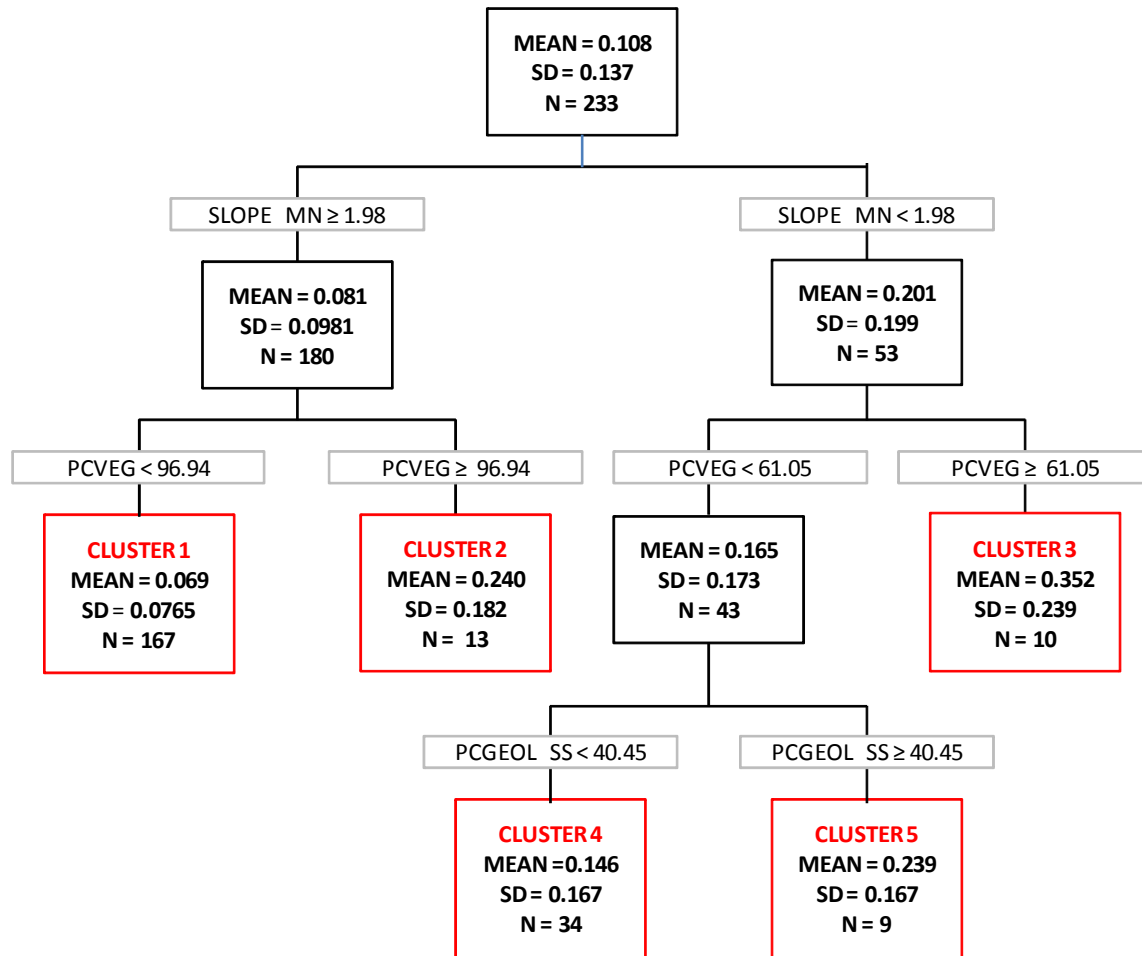
Figure 23 Regression Tree developed based on analysis of study catchments. Refer to Table 4 for a definition of terms.

The first characteristic splitting the variables is the degree of slope. Catchments with a high slope (greater than 1.98 degrees) were identified to have baseflow characteristics different from those catchments with lower slope. The degree of slope that corresponds with these differences is determined automatically through the regression tree development process. Catchments with a high slope are then split into two groups, one with a very high vegetation cover (~97%) and the other with less woody vegetation cover. For catchments with lower slope values (less than 1.98 degrees), catchments are also separated by the degree of vegetation cover. For these less steep catchments, those with less than approximately 60% of vegetation cover are further split according to the proportion of the catchment that contains the sandstone geology type.

It can be seen in Figure 23 that splitting clusters 1 and 2 creates one group with a higher mean (average Baseflow Peak Ratio = 0.23) than the other (average Baseflow Peak Ratio = 0.07). The large group in cluster 1 (made up of more than 160 catchments) is shown to have a much lower mean value for the Baseflow Peak Ratio. This indicates a low peak baseflow relative to peak streamflow. This is consistent with knowledge on baseflows across Australia where in many cases baseflow is not a significant contributor to streamflow, and indicates

that this behaviour is influenced by the location of the catchment in relation to elevated landscapes. The cluster 1 catchments considered in this study tend to be located along the Great Dividing Range and other mountainous areas.

Cluster 2 catchments are also located in mountainous areas, but tend to be in more forested locations, such as alpine Victoria and Tasmania.

Cluster 3 consists of catchments with low slope and a high proportion of vegetation cover. This cluster also contains the catchments with the highest values for peak baseflow relative to peak streamflow. This is considered reasonable as catchment characteristics such as low slope and high proportion of vegetation cover would act to reduce the magnitude of surface runoff, making baseflow a higher proportion of the total observed streamflow. These catchments primarily occur in south-west Western Australia.

For catchments where both slope and vegetation cover are relatively low, the proportion of the catchment that contains sandstone determines the baseflow in the regression tree. A high proportion of sandstone in such a catchment generally exhibits baseflow peaks higher than those with a low proportion of sandstone. This is likely due to the hydraulic properties of sandstone geology types. Cluster 4 catchments are sparsely distributed across most parts of the country. In contrast, Cluster 5 catchments considered in this study were focused in northern Australia.

The tree produced in Figure 23 has a PRE value of 0.376. This indicates that the separation of catchments into these clusters accounts for approximately 38% of the variation in Baseflow Peak Ratio values at the 10 year ARI flood.

5.2.3. Robustness of the regression tree

To determine the robustness of the regression tree, a commercial statistical package (SYSTAT 12) was used to run a jack-knife analysis (essentially a bootstrap approach without replacement). This is a type of resampling methodology, in which one data point is removed from the sample without replication at a time. For our purposes, 235 samples of a 236 catchment dataset were sampled without replication. This method tests how sensitive the clusters are to changes in the sample population. In the case of a good fit, the removal of one catchment from the sample should not significantly change the arrangement of the tree.

In this analysis, the variable used to split the tree and the value of the variable was examined and compared to the tree developed in Section 5.2.2. The regression tree in Figure 23 could be considered to be a robust tree if the resampling process produced generally consistent results.

Figure 24 shows the outcomes of this resampling analysis. Two pieces of information are captured in this figure, namely variables used to split the tree and the values at which the splits occur. The charts at each level of the tree indicate the proportion of the resampled data that produced a particular result.

The first split variable, SLOPE_MN occurs for 93% of replicates. Other possible variables occur infrequently at this level of the tree. Of note, another possible variable at this level of the tree was the ELEV_MAX. The occurrence of this characteristic was considered to further confirm the robustness of the slope term in the original tree. Whilst not explicitly representing the slope, this characteristic captures a similar physical process in the regression tree, namely the importance of elevated areas within the catchment that generate runoff.

At all other levels of the tree, both the variable selected and the value at which the split occurs is replicated in over 95% of the samples. The use of catchment vegetation to differentiate between cluster 1 and 2 is reinforced in Figure 24, with more than 99% of the samples which used slope in the first split also used vegetation at this level of the tree. In these cases, the value of the vegetation coverage was also generally consistent with that in Figure 23.

For catchments with low degrees of slope, the vegetation was also the defining characteristic for the second split in the tree. In all cases, the value of vegetation coverage was the same as that in Figure 23. The split into clusters four and five was also generally consistent with that presented for the total data set, with sandstone geology types being able to describe the differences in catchments for 95% of samples.

These results indicate that the tree developed in Figure 23 is robust and was not simply produced as a result of the catchment sample used in the original analysis. The clusters identified in Figure 23 were thus considered an appropriate form of regionalisation which grouped catchments with consistent baseflow behaviour. Within each of these clusters, it was considered relevant to develop prediction equations to describe the baseflow behaviour based on catchment characteristics.
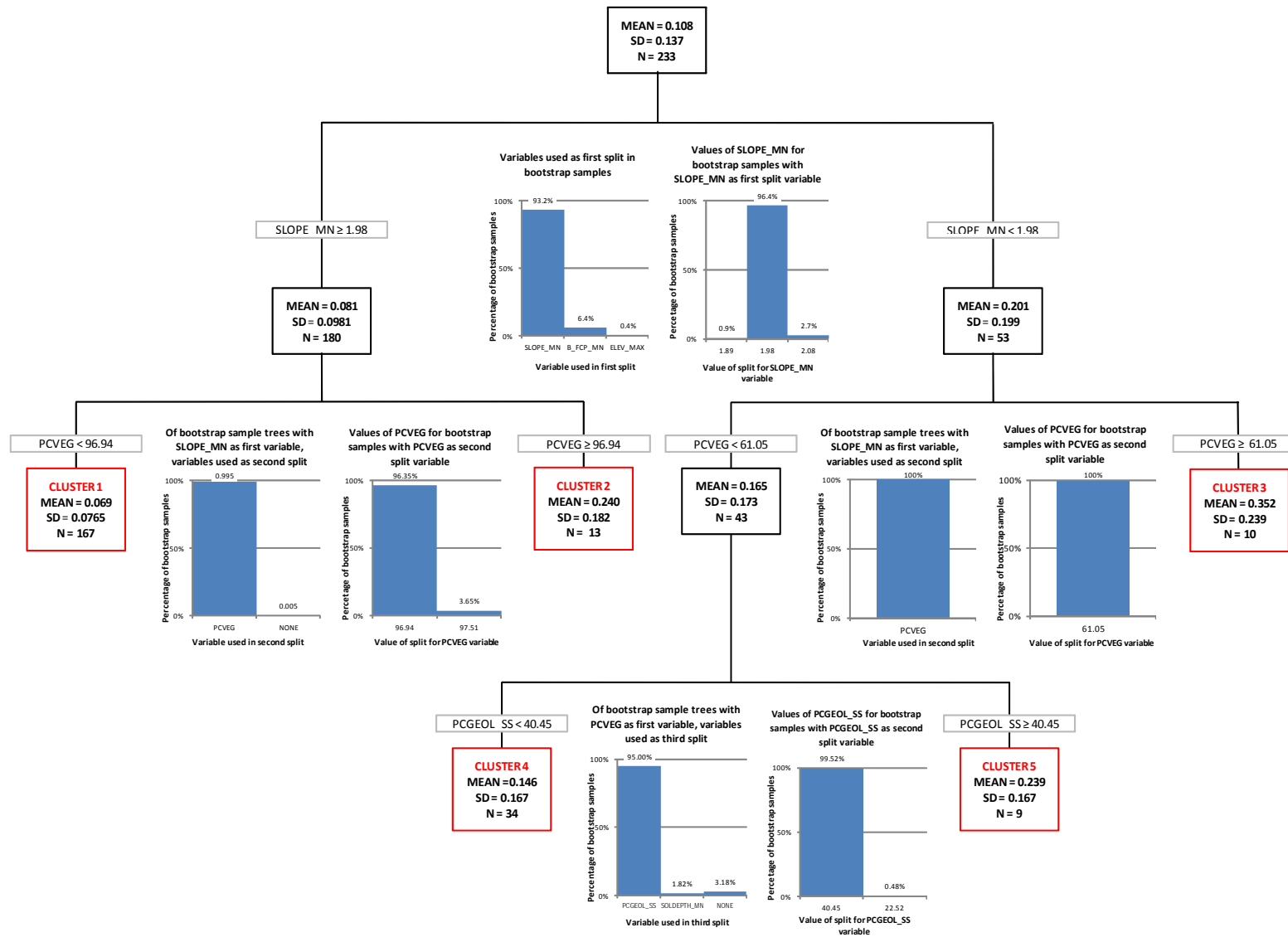
Figure 24 Results of regression tree jack-knifing

5.3. Development of regression relationships

5.3.1. Multiple linear regression approach

Multiple linear regression was used on each cluster of the regression tree to produce a prediction equation for the values of the Baseflow Peak Ratio and the Baseflow Volume Ratio based on catchment characteristics. The multiple linear regression model is of the form:

$$Y = a_0 + a_1X_1 + a_2X_2 + ... + a_nX_n$$

Equation 11

where the dependent variable Y is expressed as a linear function of n independent variables X_1, X_2 , ..., X_n. The regression coefficients a_0, a_1, a_2, ..., a_n are estimated from the sample data using the least squares method. The degree of leverage indicated by the F-statistic was used as the criteria for including independent variables in the regression. Instances of high leverage indicated that the variable was a strong predictor potentially suitable for inclusion in the prediction equation.

A forward step-wise selection method was used to select variables for inclusion in the regression. This involved first adding the best explanatory catchment characteristic at each step. Each independent variable in the regression was then cycled out to determine whether a different variable was a better addition given the variables already included.

In some instances, it is necessary to transform some or all of the dependent or independent variables to produce a valid model. Transforming variables aims to improve the model fit and ensure that the model assumptions are appropriate.

The multiple linear regression models were assessed using the coefficient of determination R^2. R^2 describes the proportion of the original variance that is explained by the model. This was used throughout each stage of building the model to evaluate the effect of each change on predictive power.

5.3.2. Selection of independent variables

It is necessary to ensure variables incorporated into regression relationships are independent. A cross-correlation matrix can be used to show the degree of correlation between a pair of variable. A matrix was compiled using on the catchment characteristics, to ensure that only independent variables were used in the development of regressions. For this study, variables with correlation values greater than 0.7 were considered to demonstrate a level of dependence and were not included in the same regression relationship. The cross correlation matrix is shown in Figure 25, with the level of grey shading indicating variables that were not independent.

It was considered logical that a number of the soil properties displayed some level of dependence, since characteristics in the upper levels of the soil profile would influence the

properties at lower profile depths. Similarities were also noticed between the elevation and slope characteristics, which was also considered logical. Other variables that were considered dependent tend to be related to the same characteristic. For instance, the mean elevation was found to be related to both the minimum and maximum elevation variables.

Those characteristics with no shading in the matrix were considered independent and were used in the development of the regression relationships.

	A_FCP_MN	A_KSAT_MN	A_SAT_MN	A_THICK_MN	B_FCP_MN	B_KSAT_MN	B_SAT_MN	B_THICK_MN	SOLDEPTH_MN	SOLPAWHC_MN	ELEV_MAX	ELEV_MIN	ELEV_MN	EVAP_MAX	EVAP_MIN	EVAP_MN	PCGEOL_AC	PCGEOL_AS	PCGEOL_AU	PCGEOL_B	PCGEOL_C	PCGEOL_FSS	PCGEOL_IM	PCGEOL_L	PCGEOL_SS	RAIN_MAX	RAIN_MIN	RAIN_MN	SLOPE_MAX	SLOPE_MIN	SLOPE_MN	PCVEG	PC_GEOL_WeightedConductivity	PC_GEOL_WeightedStorageRanking
A_FCP_MN	1.0																																	
A_KSAT_MN	0.2	1.0																																
A_SAT_MN	0.9	0.5	1.0																															
A_THICK_MN	-0.7	0.2	-0.4	1.0																														
B_FCP_MN	0.8	-0.2	0.5	-0.8	1.0																													
B_KSAT_MN	-0.1	0.7	0.3	0.4	-0.5	1.0																												
B_SAT_MN	0.8	0.4	0.9	-0.5	0.5	0.3	1.0																											
B_THICK_MN	0.8	0.4	0.8	-0.5	0.6	0.0	0.7	1.0																										
SOLDEPTH_MN	0.6	0.4	0.7	-0.3	0.5	0.0	0.6	0.9	1.0																									
SOLPAWHC_MN	0.6	0.7	0.9	-0.1	0.2	0.4	0.8	0.8	0.8	1.0																								
ELEV_MAX	0.4	0.0	0.4	-0.3	0.5	-0.2	0.4	0.3	0.4	0.3	1.0																							
ELEV_MIN	0.1	0.0	0.1	-0.1	0.2	-0.1	0.1	0.1	0.1	0.0	0.6	1.0																						
ELEV_MN	0.2	0.0	0.2	-0.2	0.3	-0.1	0.3	0.2	0.2	0.2	0.9	0.9	1.0																					
EVAP_MAX	0.1	0.3	0.2	0.2	-0.2	0.3	0.1	0.2	0.1	0.3	-0.3	-0.3	-0.3	1.0																				
EVAP_MIN	0.1	0.4	0.3	0.1	-0.1	0.4	0.2	0.2	0.2	0.4	-0.3	-0.3	-0.3	0.9	1.0																			
EVAP_MN	0.1	0.4	0.3	0.2	-0.1	0.4	0.2	0.2	0.2	0.3	-0.3	-0.3	-0.3	1.0	1.0	1.0																		
PCGEOL_AC	-0.1	-0.1	-0.1	0.0	0.0	-0.1	-0.1	0.0	0.0	-0.1	0.0	0.0	0.0	0.0	-0.1	-0.1	1.0																	
PCGEOL_AS	-0.2	0.0	-0.2	0.2	-0.2	0.1	-0.2	-0.1	0.0	-0.1	-0.2	-0.1	-0.2	-0.1	-0.2	-0.1	0.0	1.0																
PCGEOL_AU	-0.1	0.0	-0.1	0.0	-0.1	-0.1	-0.1	0.0	0.0	-0.1	-0.3	-0.3	-0.4	0.3	0.3	0.3	0.1	0.0	1.0															
PCGEOL_B	0.6	0.2	0.4	-0.4	0.4	0.0	0.4	0.5	0.3	0.4	0.2	0.3	0.2	0.1	0.1	0.1	0.0	-0.1	-0.1	1.0														
PCGEOL_C	-0.2	0.0	-0.2	0.1	-0.2	0.0	-0.2	-0.1	-0.1	-0.2	-0.3	-0.2	-0.3	0.0	0.0	0.0	0.0	0.1	0.1	-0.1	1.0													
PCGEOL_FSS	0.2	-0.2	0.1	-0.3	0.1	-0.1	0.2	0.1	0.0	0.0	0.0	0.0	0.0	-0.1	-0.1	-0.1	0.0	0.0	0.0	0.2	0.0	1.0												
PCGEOL_IM	-0.2	0.1	-0.1	0.3	-0.2	0.2	-0.1	-0.2	-0.1	0.0	0.2	0.0	0.2	0.2	0.2	0.2	0.0	-0.1	-0.2	-0.4	-0.3	-0.2	1.0											
PCGEOL_L	-0.1	0.0	-0.1	0.0	-0.1	0.0	-0.1	-0.1	-0.1	-0.1	-0.1	0.0	-0.1	-0.1	-0.1	-0.1	0.0	0.0	0.0	-0.1	0.0	0.0	-0.1	1.0										
PCGEOL_SS	-0.1	-0.2	-0.1	-0.1	0.0	-0.2	-0.1	-0.1	-0.1	-0.1	0.0	0.0	0.0	-0.3	-0.3	-0.3	0.0	-0.1	0.0	-0.2	0.0	-0.1	-0.6	0.1	1.0									
RAIN_MAX	0.5	0.3	0.6	-0.2	0.3	0.1	0.5	0.5	0.4	0.5	0.2	-0.2	0.0	0.6	0.6	0.6	0.0	-0.1	0.1	0.3	-0.1	-0.1	0.1	-0.1	-0.2	1.0								
RAIN_MIN	0.4	0.4	0.5	0.0	0.1	0.2	0.4	0.5	0.4	0.5	-0.1	-0.2	-0.2	0.6	0.7	0.7	-0.1	-0.2	0.2	0.3	-0.1	-0.1	0.0	-0.1	-0.2	0.7	1.0							
RAIN_MN	0.5	0.4	0.6	-0.1	0.2	0.2	0.5	0.6	0.4	0.6	0.1	-0.2	-0.1	0.6	0.7	0.7	-0.1	-0.2	0.1	0.3	-0.1	-0.1	0.0	-0.1	-0.2	0.9	0.9	1.0						
SLOPE_MAX	0.5	0.0	0.4	-0.3	0.4	-0.2	0.4	0.4	0.4	0.3	0.7	0.0	0.3	0.0	0.0	0.0	0.0	-0.2	-0.2	0.1	-0.2	0.0	0.0	-0.1	0.1	0.5	0.2	0.3	1.0					
SLOPE_MIN	0.3	0.1	0.3	-0.1	0.2	0.2	0.4	0.2	0.1	0.3	0.3	0.1	0.2	0.0	0.0	0.0	-0.1	-0.2	-0.3	0.1	-0.2	0.2	0.0	-0.1	0.2	0.2	0.1	0.2	0.8	1.0				
SLOPE_MN	0.5	0.1	0.5	-0.2	0.4	0.0	0.6	0.4	0.4	0.4	0.7	0.0	0.4	-0.1	0.0	-0.1	-0.1	-0.2	-0.3	0.0	-0.3	0.0	0.2	-0.1	0.1	0.5	0.2	0.4	0.8	0.5	1.0			
PCVEG	0.1	0.3	0.3	0.2	-0.1	0.4	0.3	0.2	0.3	0.4	0.3	-0.2	0.1	0.2	0.2	0.2	-0.1	-0.1	-0.2	-0.3	-0.2	0.3	-0.1	0.0	0.3	0.2	0.3	0.4	0.4	0.6	0.6	1.0		
PC_GEOL_WeightedConductivity	0.2	0.0	0.1	-0.2	0.1	-0.1	0.1	0.2	0.1	0.1	-0.2	0.0	-0.1	0.0	0.0	0.0	0.0	0.2	0.1	0.6	0.7	0.4	-0.6	0.0	-0.1	0.0	0.1	0.0	-0.2	-0.2	-0.3	-0.4	1.0	
PC_GEOL_WeightedStorageRanking	0.1	0.0	0.0	-0.1	0.0	-0.1	0.0	0.1	0.0	0.1	-0.3	-0.1	-0.2	0.1	0.0	0.0	0.0	0.4	0.3	0.4	0.7	0.3	-0.6	0.1	-0.2	0.0	0.1	0.0	-0.3	-0.2	-0.4	-0.4	1.0	1.0

Figure 25 Cross-correlation matrix for variables used in regression

5.3.3. Model assumptions

A number of assumptions need to be met to ensure the validity of a regression model, including:

- the residuals do not contain outliers
- the residuals are not highly correlated
- the residuals are normally distributed
- the residuals exhibit constant variance

To ensure that these assumptions hold, diagnostics were used to evaluate each regression following the process recommended by Sheather (2009), including:

1. A scatter plot matrix of continuous predictor variables. This displays the general pattern between the independent and dependent variables.

2. A plot of fitted values against observed dependent values. For a valid model, a plot comparing the fitted values to the original dependent variable will produce an approximately straight line.

3. Diagnostic plots including residuals and fitted values, a normal Q-Q plot and a plot of residuals and leverage. For a valid model, the residuals versus fitted values should display a random distribution of points around the horizontal axis and constant variability along the horizontal axis. Points in the normal Q-Q plot should lie close to a straight line for a valid model. Leverage points are identified using the residuals versus leverage plot, which makes note of the Cook's distance, a further measure of the actual influence of data points. Points which are overly influential lie outside of the value calculated by the following relationship:

$$2 \times \frac{(p + 1)}{n}$$

Equation 12

Where p = number of independent variables in the regression

n = number of data points

4. A plot of standardised residuals against each predictor. This diagnostic test was used to confirm whether the model provides an adequate fit to the data. When a valid model has been produced, a plot of the standardised residuals against any of the independent or dependent variables (i.e. the fitted values) will display a random scatter of points around the horizontal axis and constant variability along the horizontal axis. The presence of any pattern in the standardised residuals indicates that the model is invalid.

   This plot is also used to determine whether any points are outliers which do not follow the general pattern of the rest of the data set. Data points with standardised residual values outside of a particular range of values may be considered to be an outlier. For moderately sized data sets, a standardised residual value outside of the interval from -2 to 2 is commonly used to define outlier points. For larger data sets, this range becomes -4 to 4 to ensure reasonable points are not inadvertently excluded.

5. A marginal model plot which shows how the model compares with a non-parametric solution. A valid solution will show a close relationship between the data and the model.

6. Calculated variance inflation factors to determine the extent of collinearity. A high level of collinearity can result in the wrong sign being attributed to a coefficient or the identification of non-significant predictor variables. Collinearity can be determined using variance inflation factors. A factor value greater than five is suggested as unacceptable (Sheather, 2009).

5.3.4. Baseflow Peak Ratio regressions

The approach described in Section 5.3.1 was used to develop regression relationships for this study. As described earlier, a forward step-wise selection method was used to select variables for inclusion in the regression. The best explanatory catchment characteristic was added into the regression model at each step. Each independent variable in the regression was then cycled out to determine the best predictor variables subject to the variables already included. This process was used to establish a relationship that incorporated up to three variables for each cluster described in Section 5.2.2. Where required, these relationships were transformed to ensure the predictive models were valid, as per the assumptions described in Section 5.3.3. The result of this process was a regionalised outcome for predicting baseflow contribution to flood events.

The prediction equations developed through this process were evaluated by checking the diagnostics and the goodness of fit. All goodness of fit statistics are reported in the arithmetic domain. Following this process, the overall goodness of fit was evaluated by incorporating outcomes of the relationships for each cluster. This provides a measure of the ability to predict the complete dataset, when the data is regionalised based on the regression tree clusters.

A reference regression was also developed for the full collection of catchments to determine the ability to predict baseflow behaviour without incorporating regionalisation into the analysis. This provides a reference baseline from which the other predictive relationships could be compared.

Table 5 shows the suite of regression equations that were developed to estimate the Baseflow Peak Ratio. The reference regression equation shows the best regression possible if no clustering is employed. The value of R^2 for the reference regression is a baseline comparison for the success of the cluster regression fits. This relationship displays a poor fit, with an R^2 of 0.19.

The results for each cluster are also presented in Table 5 . The variables in each equation were selected because they provided the greatest explanation of the variation in Baseflow Peak Ratio values. The R^2 values for the clustered relationships range between 0.18 and 0.96. The best fits are obtained for the cluster groups with fewer data points.

The applied regression for all clusters is the result when all clustering and regression equations are applied to the whole dataset. The R^2 value shows the overall proportion of variation that can be explained by the application of the regression tree and multiple regression equations together. The regression relationships developed to describe the Baseflow Peak Ratio have an overall goodness of fit of over 0.6.

Table 5 Regression relationships for Baseflow Peak Ratio

SLOPE_MN	PCVEG	PCGEOL_SS	Cluster Number	Prediction equation	R²	N
			REFERENCE REGRESSION	Baseflow Peak Ratio @ 10yr ARI = 0.0446 + 0.000841 x **B_KSAT_MN** - 0.506 x **SLOPE_MIN** + 0.000792 x **PCVEG**	0.191	233
>=1.98	<96.94		1	Log(Baseflow Peak Ratio @ 10yr ARI) = -2.534 + 0.418 x (**B_KSAT_MN**)$^{0.2}$ + 0.0115 x **PCGEOL_AU** + 0.00411 x **PCVEG**	0.186	167
	>=96.94		2	Baseflow Peak Ratio @ 10yr ARI$^{0.3}$ = 0.0198 + 0.612 x **A_FCP_MN** + 0.00207 x **A_KSAT_MN** - 0.0143 x **PCGEOL_AU**	0.956	13
<1.98	>=61.05		3	Log(Baseflow Peak Ratio @ 10yr ARI) = -1.837 + 0.0721 x **PCGEOL_AU** - 0.0761 x **PCGEOL_SS** + 0.00114 x **RAIN_MN**	0.766	10
	<61.05	<40.45	4	Log(Baseflow Peak Ratio @ 10yr ARI) = -2.14 + 0.000963 x **EVAP_MAX** + 0.0101 x **PCGEOL_AS** + 0.0147 x **PCGEOL_SS**	0.226	34
		>=40.45	5	SqRt(Baseflow Peak Ratio @ 10yr ARI) = -0.556 + 0.00641 x **B_KSAT_MN** + 0.223 x (**PCGEOL_IM**)$^{0.2}$ + 0.164 x (**PCVEG**)$^{0.2}$	0.846	9
			APPLIED REGRESSION - ALL CLUSTERS, ALL REGRESSIONS		0.606	233

The performance of these regression relationships was assessed using model validity diagnostics. The following discussion provides a summary of this performance evaluation, and a complete package of diagnostic plots is provided in Appendix E.

Cluster 1 consists of steep catchments with mean slope greater than 1.98 degrees and vegetation coverage lower than 96.94%. This cluster includes the majority of catchments in New South Wales and many along the east coast of Australia, including Victoria, Queensland and Tasmania.

The large number of catchments included in this cluster provided challenges in identifying catchment characteristics that are able to adequately define baseflow characteristics. The catchment characteristics best able to describe the variability in this cluster of catchments are:

- mean saturated hydraulic conductivity of the lower soil layer (B_KSAT_MN)
- proportion of the catchment containing general alluvial geology type including undifferentiated sands, silts, clays or fine-grained soils (PCGEOL_AU)
- proportion of the catchment covered in woody vegetation (PCVEG)

Figure 26 shows the fit of the regression for cluster 1. The values calculated from the regression are generally lower than the observed values, highlighting the weak power of the regression to predict the baseflow characteristics for this cluster.

Figure 27 provides further information on the data in this cluster. This box and whisker plot summaries the spread of the observed Baseflow Peak Ratio values, with the central line within the box representing the median ratio value, the ends of the box reflecting the first and third quantiles of the data and the extent of the whiskers showing the range in the data values. This indicates that the majority of the catchments have a Baseflow Peak Factor between 0.025 and 0.075 and that an average Baseflow Peak Factor value would be sufficient to approximate much of this data set. Consequently, the poor fit of the regression relationship described above is not considered significant since it is relevant for only a small number of catchments.

In order to maintain consistency between the Baseflow Peak Ratio and the Baseflow Volume Ratio, the predictive relationship above was adopted to describe the Baseflow Peak Ratio for cluster 1 catchments.
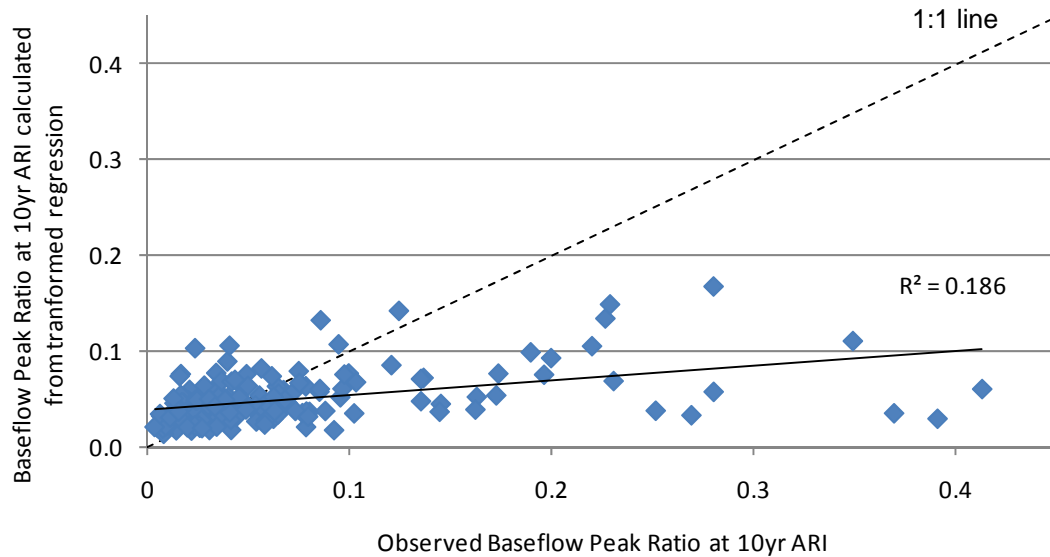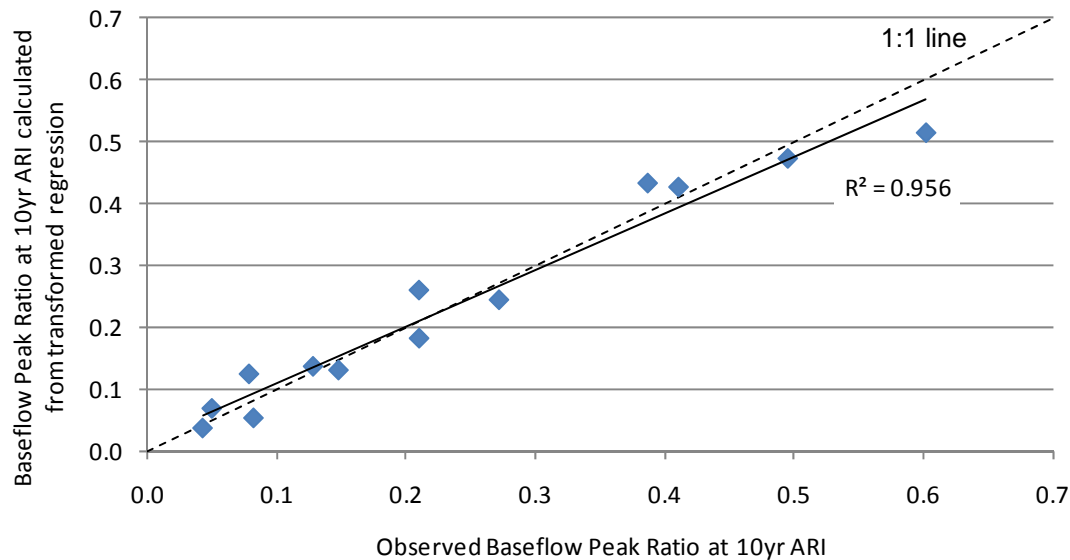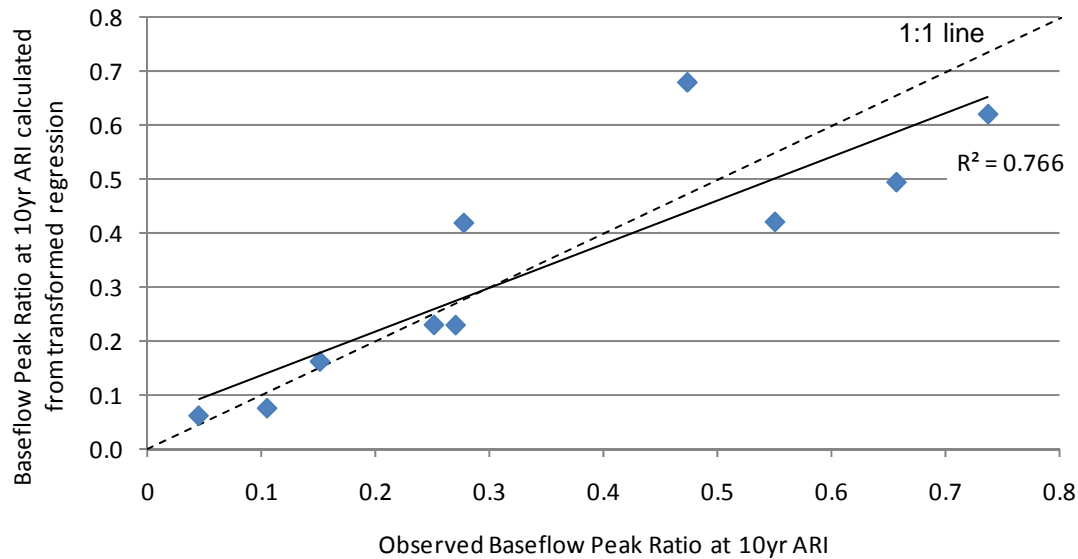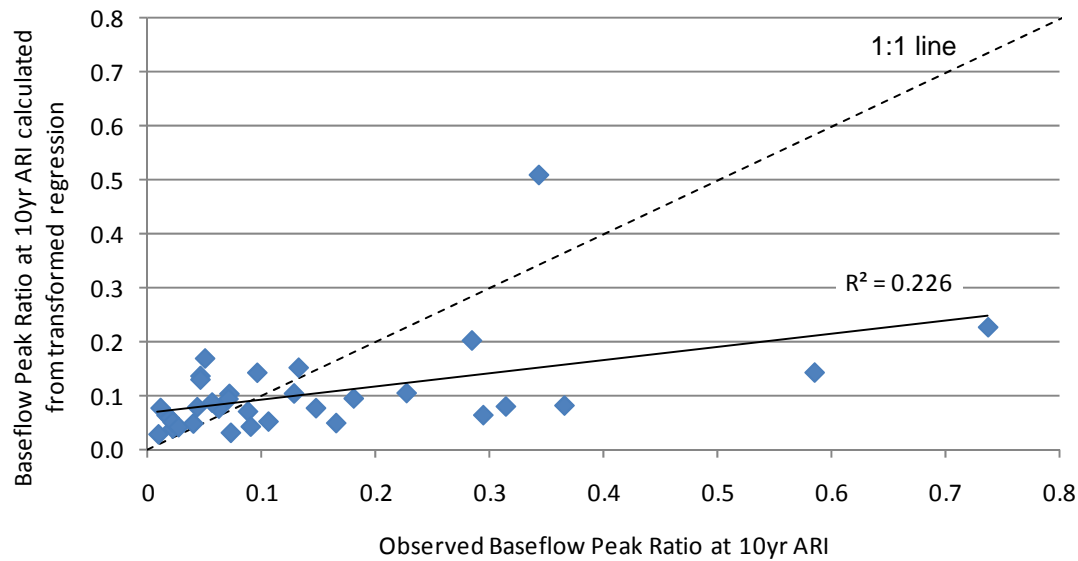
Figure 26 Observed value of the Baseflow Peak Ratio and the value calculated from final transformed regression for cluster 1. The solid line shows the fitted regression.
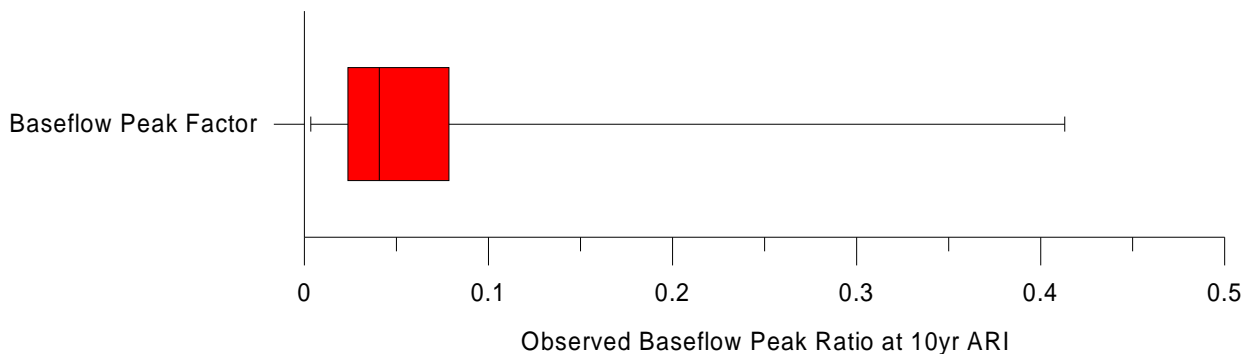
Figure 27 Box and whisker plot of the observed Baseflow Peak Ratio for cluster 1. The central line in the box represents the median ratio value, the ends of the box reflect the first and third quantiles of the data and the extent of the whiskers shows the range in the data values.

Cluster 2 consists of a small number of catchments with high values of slope and very high vegetation cover, located in alpine regions of Victoria and NSW. The regression for cluster 2 shows an excellent fit to the catchments (Figure 28). The catchment characteristics that are best able to describe the variation in cluster 2 catchments are:

- mean saturated hydraulic conductivity of the upper soil layer (A_KSAT_MN)
- proportion of the catchment containing general alluvial geology type including undifferentiated sands, silts, clays or fine-grained soils (PCGEOL_AU)
- top soil layer nominal field water capacity (A_FCP_MN)

Over half (55%) of the variation in Cluster 2 catchments is explained by the mean hydraulic conductivity of the upper soil layer (A_KSAT_MN). A further 25% is explained by the addition of the PCGEOL_AU variable to the regression. The inclusion of the top soil nominal field water

capacity accounts for close to another 10% of variability. Transforming the dependent variable (Baseflow Peak Ratio at 10 year ARI) further improves the goodness of fit in the arithmetic domain for cluster 2. This final relationship is presented in Figure 28.

The diagnostics for this regression shown in Appendix E display that one data point has a high leverage on the resulting regression. However, since the model fit in Figure 28 and the other diagnostics indicate that the model is valid, this leverage point was not considered significant.
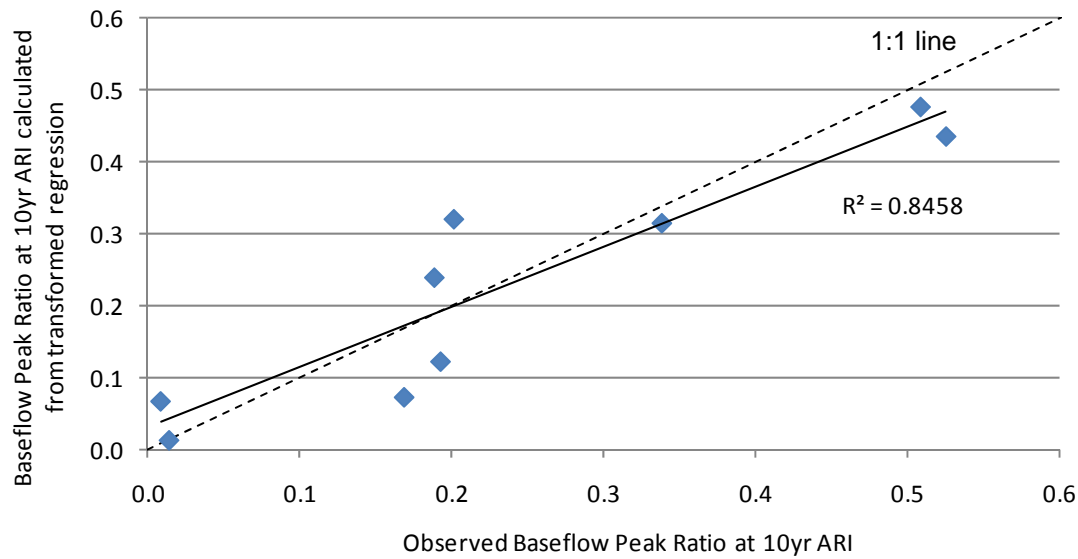

Figure 28 Observed value of the Baseflow Peak Ratio and the value calculated from final transformed regression for cluster 2. The solid line shows the fitted regression.

Cluster 3 includes catchments with mean slope less than 1.98 degrees and with proportions of woody vegetation cover in the catchment greater than 61%. The majority of cluster 3 catchments are located in the south west corner of Western Australia, although other examples are scattered throughout Queensland and inland locations of south eastern Australia. Baseflow Peak Ratio values in this cluster are relatively high.

The characteristics used to describe the variation in Baseflow Peak Ratio values for cluster 3 are:

- proportion of the catchment containing general alluvial geology type including undifferentiated sands, silts, clays or fine-grained soils (PCGEOL_AU)
- proportion of the catchment containing a sandstone geology type (PCGEOL_SS)
- mean annual rainfall (RAIN_MN)

This relationship is displayed in Figure 29. Approximately 50% of the variation in cluster 3 is explained by the proportion of the catchment containing general alluvial geology type including undifferentiated sands, silts, clays or fine-grained soils (PCGEOL_AU). Almost a further 30% is explained by the combined impact of the mean annual rainfall (RAIN_MN) and the proportion of the catchment containing the sandstone geology type (PCGEOL_SS).

The diagnostics for this regression are shown in Appendix E. These display that some data

points have a high leverage on the resulting regression. However, since the model fit in Figure 29 and the other diagnostics indicate that the model is valid, this was not considered significant.

Figure 29 Observed value of the Baseflow Peak Ratio and the value calculated from final transformed regression for cluster 3. The solid line shows the fitted regression.

Cluster 4 includes catchments with:

- mean slope less than 1.98 degrees
- proportion of woody vegetation cover less than 61.05%
- proportion of the catchment containing a sandstone geology type less than 40.45%

These occur in most states of Australia and are geographically widely distributed. The majority of catchments from Western Australia, South Australia and Northern Territory fall into this group. The variation in cluster 4 is explained by:

- maximum annual evapotranspiration (EVAP_MAX)
- proportion of the catchment containing sandstone geology (PCGEOL_SS)
- proportion of the catchment containing an alluvial geology type including medium grained particles (fine to medium-grained sands) (PCGEOL_AS)

This relationship is displayed in Figure 30. The maximum annual evaporation (EVAP_MAX) is able to explain only 9% of the variation in Baseflow Peak Ratio values in cluster 4. The variable PCGEOL_SS adds a further 9% explanation and the variable PCGEOL_SS another 10%.

The values of the Baseflow Peak Ratio calculated from the regression are generally lower than those measured during the analysis. The outlying value which over-estimates the value of the ratio and causes the low R^2 is found in Western Australia. This catchment has a much higher proportion of the medium grained alluvial geology type (94%) than any of the other catchments in this cluster.

Figure 30 Observed value of the Baseflow Peak Ratio and the value calculated from final transformed regression for cluster 4. The solid line shows the fitted regression.

Cluster 5 includes catchments with:

- mean slope less than 1.98 degrees
- proportion of woody vegetation cover less than 61.05%
- proportion of the catchment containing a sandstone geology type more than 40.45%

Catchments in cluster 5 are located across the north of Australia and in Western Australia and South Australia concentrated around the Great Australia Bight. The characteristics that explain the variation in cluster 5 are:

- mean saturated hydraulic conductivity of the lower soil layer (B_KSAT_MN)
- proportion of the catchment containing igneous and metamorphic rock geology type, including conglomerates, mudstones, siltstones and others (PCGEOL_IM)
- proportion of the catchment covered in woody vegetation (PCVEG)

This relationship is displayed in Figure 31. The mean saturated hydraulic conductivity of the lower soil layer explains approximately 50% of the variation in calculated Baseflow Peak Ratio values. The variable PCGEOL_IM explains approximately an additional 25% of the variability. The proportion of the catchment covered in woody vegetation explains a further 15%.

The regression equation for cluster 5 displays a very high goodness of fit. The predicted Baseflow Peak Ratio values correspond well with the observed. There is minimal overestimation for low values of the Baseflow Peak Ratio and minimal underestimation for high values.

The diagnostics in Appendix E show two points of high leverage. These occur for values at the very top and very bottom of the range of observed Baseflow Peak Ratio values. The small number of observations in this cluster means that most extreme values would be expected to have a significant effect on the regression.
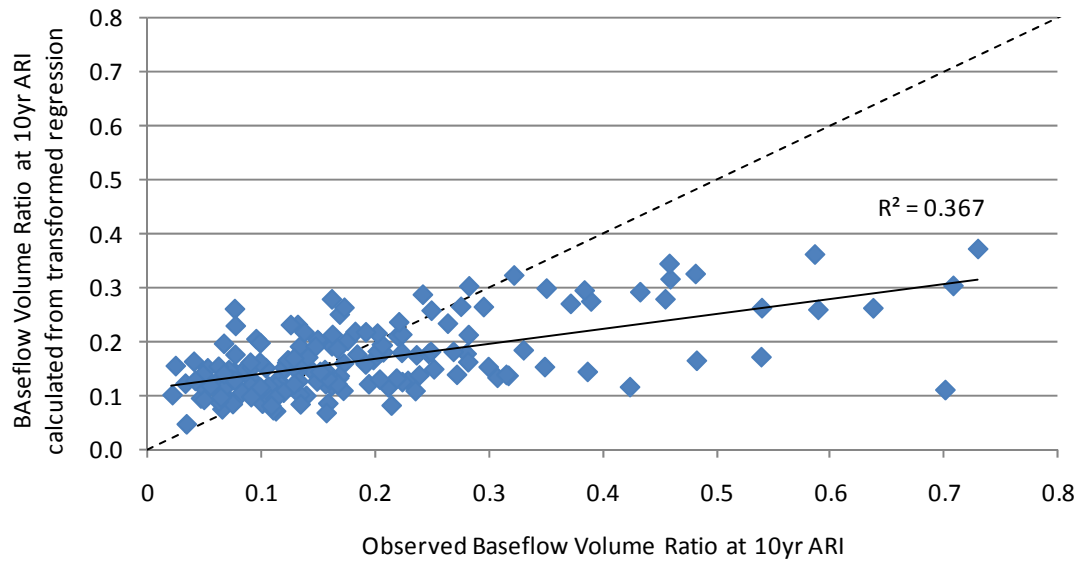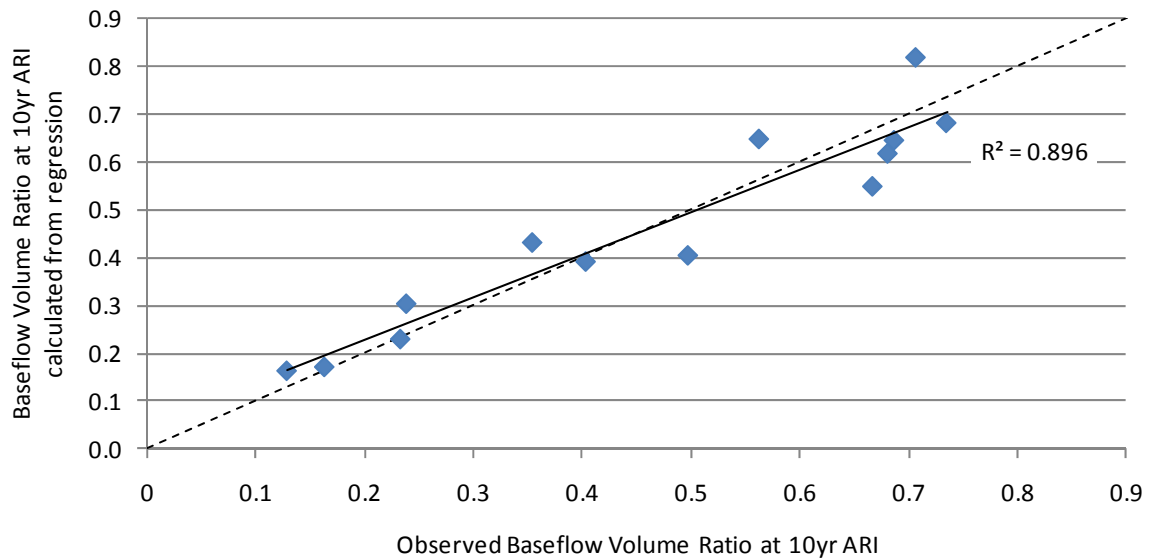
Figure 31 Observed value of the Baseflow Peak Ratio and the value calculated from final transformed regression for cluster 5. The solid line shows the fitted regression.

5.3.5. Baseflow Volume Ratio regressions

The Baseflow Volume Ratio describes the ratio of baseflow volume to total streamflow volume. To maintain consistency with the Baseflow Peak Ratio, the regression tree developed using the Baseflow Peak Ratio was used as the basis for the Baseflow Volume Ratio regression equations. Therefore catchments were kept in the same clusters as described in Section 5.2.2.

The calculated value for the Baseflow Peak Ratio was used as a variable in each of the Baseflow Volume Ratio regression equations. This was done for a number of reasons, in particular because the calculated Baseflow Peak Ratio value was the best predictor for the Baseflow Volume Ratio in all cases, and because this approach ensures consistency between the two calculated ratio values.

A summary of regression equations developed to predict the Baseflow Volume Ratio from calculated Baseflow Peak Ratio values and catchment characteristics is shown in Table 6.

Table 6 Regression relationships for Baseflow Volume Ratio

SLOPE_MN	PCVEG	PCGEOL_SS	Cluster Number	Prediction equation	R^2	N
REFERENCE REGRESSION				Baseflow Volume Ratio @ 10yr ARI = 0.0926 + 0.000931 x **B_KSAT_MN** + 0.321 x **PCGEOL_AC** + 0.86 x **BFR_CALC**	0.520	233
>=1.98	<96.94		1	Log(Baseflow Volume Ratio @ 10yr ARI) = -0.355 + 0.243 x **A_KSAT_MN** - 0.00035 x **EVAP_MAX** + 0.601 x Log(**BFR_CALC**)	0.367	167
	>=96.94		2	Baseflow Volume Ratio @ 10yr ARI = -1.272 + 2.295 x **A_SAT_MN** + 0.00470 x **PCGEOL_IM** + 0.913 x **BFR_CALC**	0.897	13
<1.98	>=61.05		3	Baseflow Volume Ratio @ 10yr ARI = 0.718 - 0.0127 x **PCGEOL_AS** + 0.000517 x **PCGEOL_IM** + 0.375 x Log(**BFR_CALC**)	0.937	10
	<61.05	<40.45	4	Log(Baseflow Volume Ratio @ 10yr ARI) = -0.932 + 0.000514 x **EVAP_MAX** + 0.145 x SqRt(**SLOPE_MAX**) + 0.592 x Log(**BFR_CALC**)	0.488	34
		>=40.45	5	SqRt(Baseflow Volume Ratio @ 10yr ARI) = -2.015 + 2.614 x **B_THICK_MN** + 0.00816 x **PCGEOL_SS** + 1.0268 x SqRt(**BFR_CALC**)	0.956	9
APPLIED REGRESSION - ALL CLUSTERS, ALL REGRESSIONS					0.613	233

Whilst consistent clusters were used for both the Baseflow Peak Ratio and the Baseflow Volume Ratio, the regression relationships developed for these two variables are different. The regression relationships for the Baseflow Volume Ratio were developed based on the most significant variables that could describe the variation in baseflow volume for each of the clusters. Some catchment characteristics are repeated in Table 5 and Table 6, however not all.

The variation in Baseflow Volume Ratio values is relatively well described by the calculated Baseflow Peak Ratio values when the whole dataset is considered together collectively in the reference regression, with an R^2 of 0.52. The mean saturated hydraulic conductivity of the lower soil layer and the proportion of the catchment containing course grained alluvial geology provide the greatest explanation of variation in the reference regression.

The proportion of variation in the Baseflow Volume Ratio that can be explained by the regressions in each of the clusters varies. The proportion of variation that is able to be accounted for by the regression is lower in clusters 1 (0.367) and 4 (0.488) than in the overall reference regression (0.520). The application of all the regression equations for cluster groups, however, resulted in an R^2 value (0.613) that is higher than that produced for all catchments together in the reference regression (0.520).

The proportion of variation that can be described in each of the clusters is higher for the Baseflow Volume Ratio than the Baseflow Peak Ratio in most cases. That is, the R^2 values in Table 6 tend to be greater than those in Table 5. This is because the calculated value for Baseflow Peak Ratio already accounts for a significant proportion of the variation in the Baseflow Volume Ratio. The final R^2 values when all clusters and regressions are applied are comparable for the Baseflow Peak Ratio (0.606) and Baseflow Volume Ratio (0.613).

The Baseflow Volume Ratio calculated for cluster 1 (Figure 32) is relatively uniform and consistently underestimates the observed ratio value for high observed values. The regression also overestimates the lower observed values of the Baseflow Volume Ratio.

The catchment characteristics that explain the variation in the Baseflow Volume Ratio values in cluster 1 are:

- mean saturated hydraulic conductivity of the upper soil layer (A_KSAT_MN);
- maximum annual evaporation (EVAP_MAX); and
- Baseflow Peak Ratio value calculated from Baseflow Peak Ratio clusters and regressions (BPR_CALC).

Approximately 30% of the variation in Baseflow Volume Ratio values in this cluster is explained by the calculated value of the Baseflow Peak Ratio. Only an additional 3% of variation explanation is added with the addition of each variable EVAP_MAX and A_KSAT_MN.
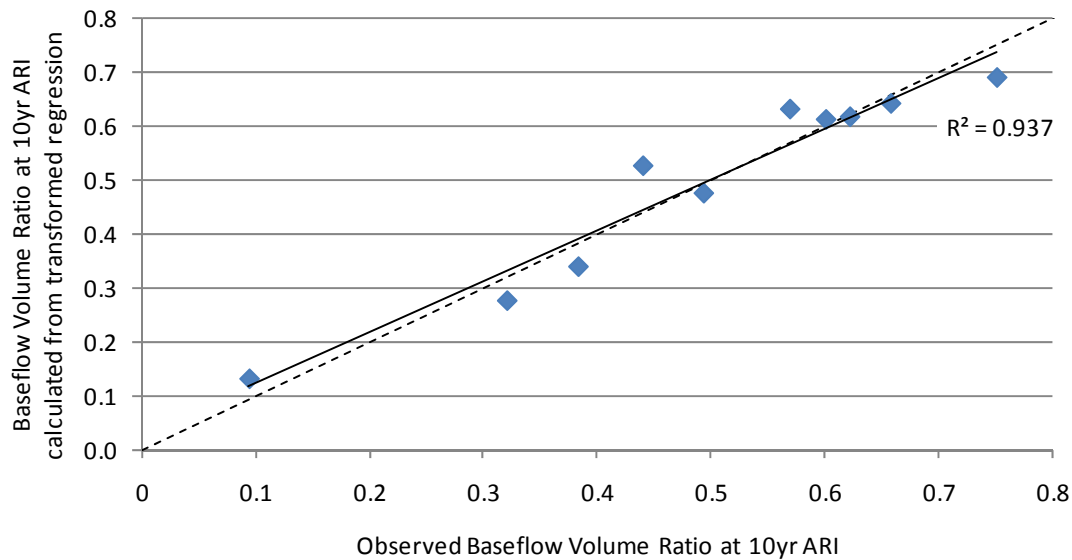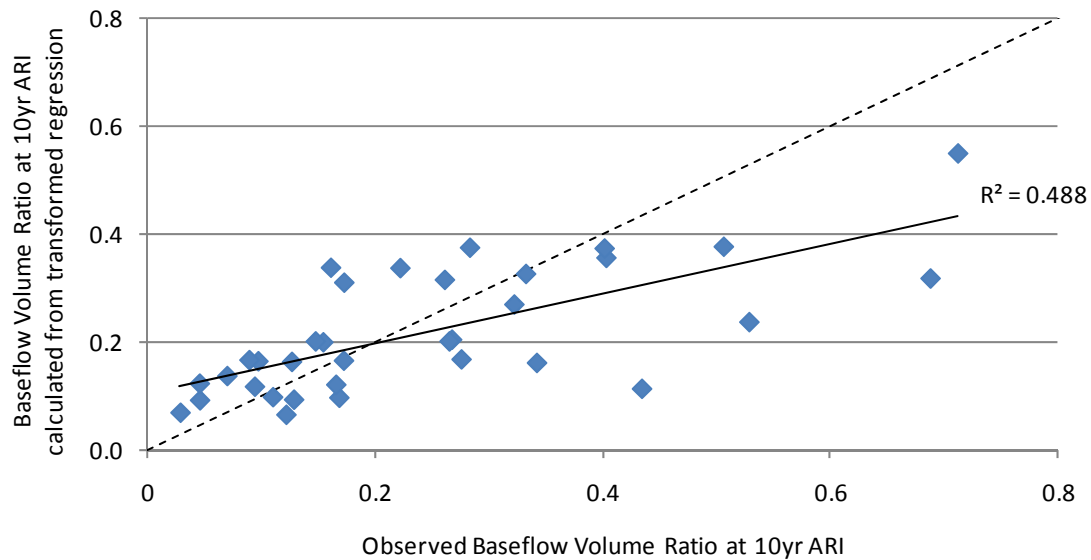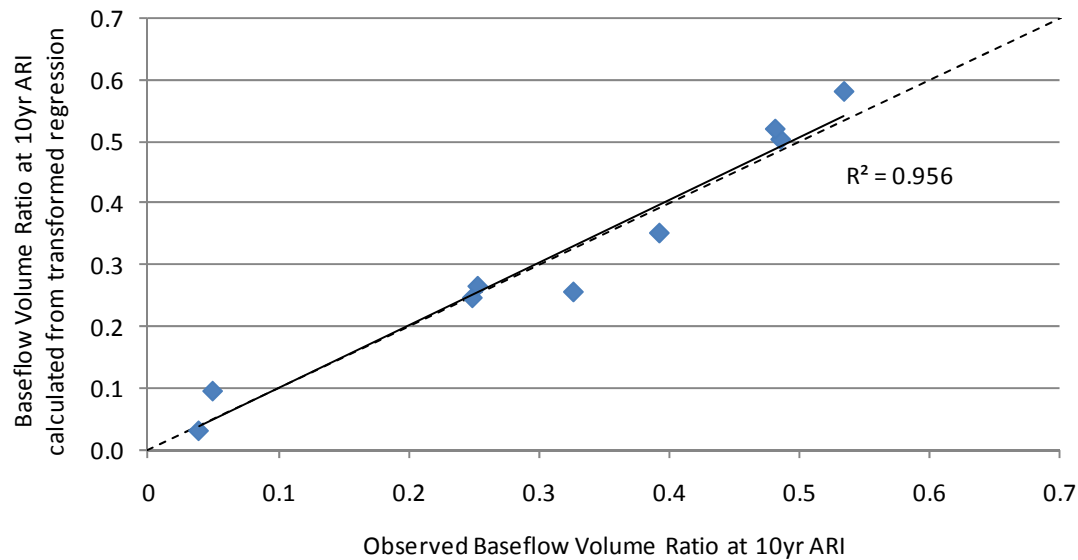
Figure 32 Observed value of the Baseflow Volume Ratio and the value calculated from final transformed regression for cluster 1. The dotted line shows the one-to-one relationship. The solid line shows the fitted regression.

The Baseflow Volume Ratio values in cluster 2 are well described by the regression, as shown in Figure 33. The regression uses:

- mean top soil layer saturated volumetric water content (A_SAT_MN);
- proportion of the catchment containing igneous and metamorphic rock geology type, including conglomerates, mudstones, siltstones and others (PCGEOL_IM); and
- Baseflow Peak Ratio value calculated from Baseflow Peak Ratio clusters and regressions (BPR_CALC).

The calculated value of the Baseflow Peak Ratio describes 62% of the variation in the Baseflow Volume Ratio values. The combination of variables PCGEOL_IM and A_SAT_MN accounts for an additional 28% of variation in Baseflow Volume Ratio values. The cluster 2 regression did not require transformation.

The diagnostics for this regression shown in Appendix E display that one data point has a high leverage on the resulting regression. However, since the model fit in Figure 33 and the other diagnostics indicate that the model is valid, this leverage point was not considered of concern.

Figure 33 Observed value of the Baseflow Volume Ratio and the value calculated from final transformed regression for cluster 2. The dotted line shows the one-to-one relationship. The solid line shows the fitted regression.

The prediction of Baseflow Volume Ratio for cluster 3 catchments is shown in Figure 34. This relationship is described by:

- proportion of the catchment containing an alluvial geology type including medium grained particles (fine to medium-grained sands) (PCGEOL_AS);
- proportion of the catchment containing igneous and metamorphic rock geology type, including conglomerates, mudstones, siltstones and others (PCGEOL_IM); and
- Baseflow Peak Ratio value calculated from Baseflow Peak Ratio clusters and regressions (BPR_CALC).

These characteristics describe a very large proportion of the variation in Baseflow Volume Ratio values. The calculated value of the Baseflow Peak Ratio describes 70% of the variation in the Baseflow Volume Ratio. The inclusion of variables PCGEOL_AS and PCGEOL_IM explain approximately 6% additional variation each. Transformation of the regression equation was required to improve model validity, and this increased the goodness of fit.

Figure 34 Observed value of the Baseflow Volume Ratio and the value calculated from final transformed regression for cluster 3. The dotted line shows the one-to-one relationship. The solid line shows the fitted regression.

The regression for Cluster 4 is presented in Figure 35 and includes:

- maximum annual evaporation (EVAP_MAX);
- maximum slope recorded for the catchment (SLOPE_MAX); and
- Baseflow Peak Ratio value calculated from Baseflow Peak Ratio clusters and regressions (BPR_CALC).

The calculated values of the Baseflow Peak Ratio explain 25% of the variation in the Baseflow Volume Ratio. The variable SLOPE_MAX explains a further 9% of the variation. The variable EVAP_MAX explains another 13%. Transformation of dependent and predictor variables increases the model fit and validity. Despite this, the regression developed for Cluster 4 overestimates the observed value of the Baseflow Volume Ratio for low values of the Baseflow Volume Ratio and underestimates it for high values of the Baseflow Volume Ratio.

Figure 35 Observed value of the Baseflow Volume Ratio and the value calculated from final transformed regression for cluster 4. The dotted line shows the one-to-one relationship. The solid line shows the fitted regression.

The regression developed for cluster 5 predicts Baseflow Volume Ratio values extremely well, as shown in Figure 36. This regression includes:

- mean lower soil layer thickness in the catchment (B_THICK_MN);
- proportion of the catchment containing sandstone geology (PCGEOL_SS); and
- Baseflow Peak Ratio value calculated from Baseflow Peak Ratio clusters and regressions (BPR_CALC).

The calculated values of Baseflow Peak Ratio explain 72% of the variation in Baseflow Volume Ratio values. Another nine per cent of the variation is explained by the mean thickness of the lower soil layer. The final 15% of variation is explained by the variable PCGEOL_SS.

Figure 36 Observed value of the Baseflow Volume Ratio and the value calculated from final transformed regression for cluster 5. The dotted line shows the one-to-one relationship. The solid line shows the fitted regression.

5.4. Predictive capability of overall models

The overall predictive capability of the regression tree and multiple linear regression equations are similar for the Baseflow Peak Ratio and Baseflow Volume Ratio. Figure 37 shows the regression for the Baseflow Peak Ratio that can be fitted to the whole catchment dataset without any clustering.

Comparing Figure 37 with Figure 38 shows the improvement in predictive ability for the Baseflow Peak Ratio that is achieved using a cluster tree and then developing regressions for each cluster. In particular, the degree of underestimation in Figure 38 is significantly lower than in Figure 37. This is especially important for the events that have a high observed Baseflow Peak Ratio value. There is also a greater ability to predict the Baseflow Peak Ratio values where the observed value is small.
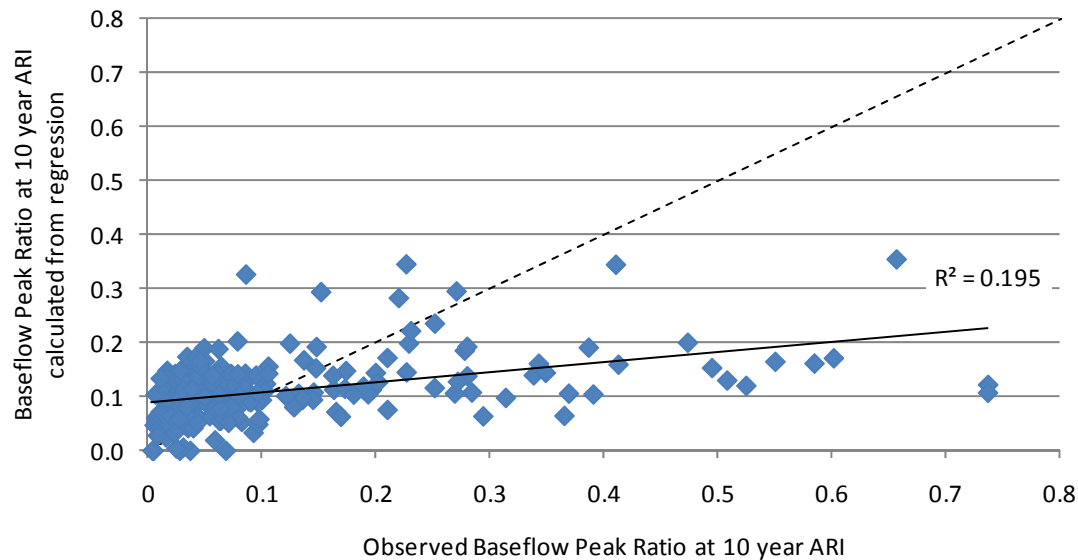
Figure 37 Reference regression fit to Baseflow Peak Ratio values
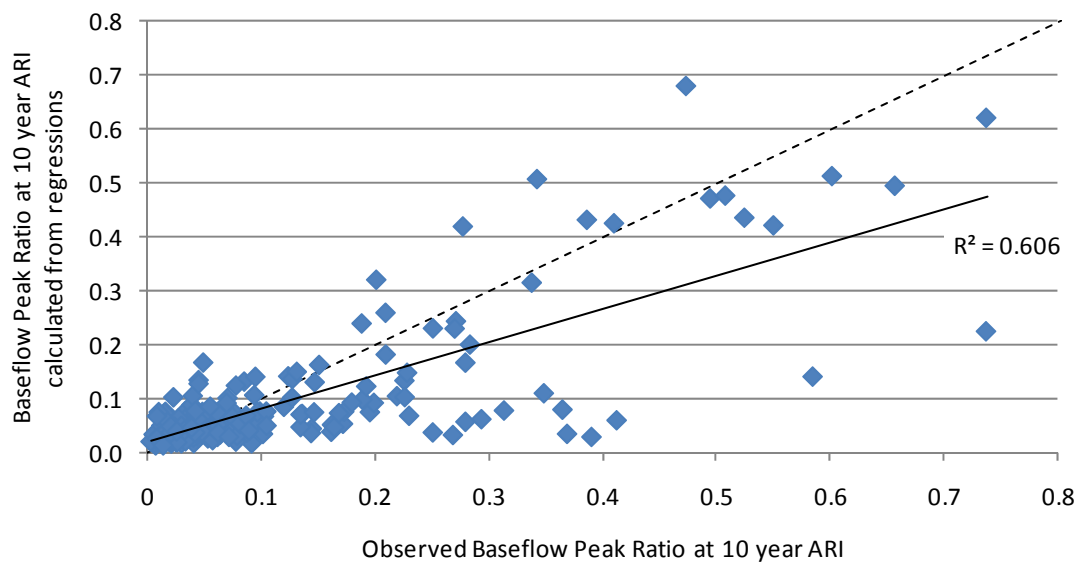
Figure 38 Application of regressions to clusters for Baseflow Peak Ratio

Figure 39 shows the regression that can be fitted to the whole catchment dataset without any clustering for the Baseflow Volume Ratio. Figure 40 displays the application of the regressions to clusters for comparison. Whilst the improvement is smaller than that observed for the Baseflow Peak Ratio, there is still clear benefit in using a regionalisation approach for the development of prediction equations. A significant part of the variation in the Baseflow Volume Ratio is accounted for by the calculated values of the Baseflow Volume Ratio, hence the addition of further variables add less benefit than for the Baseflow Peak Ratio.
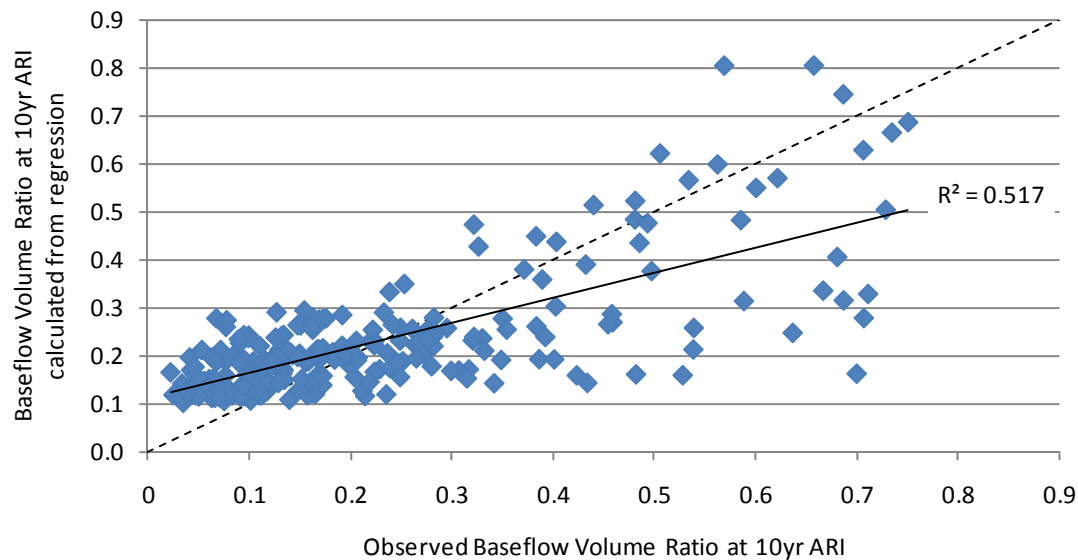
Figure 39 Reference regression fit to Baseflow Volume Ratio values
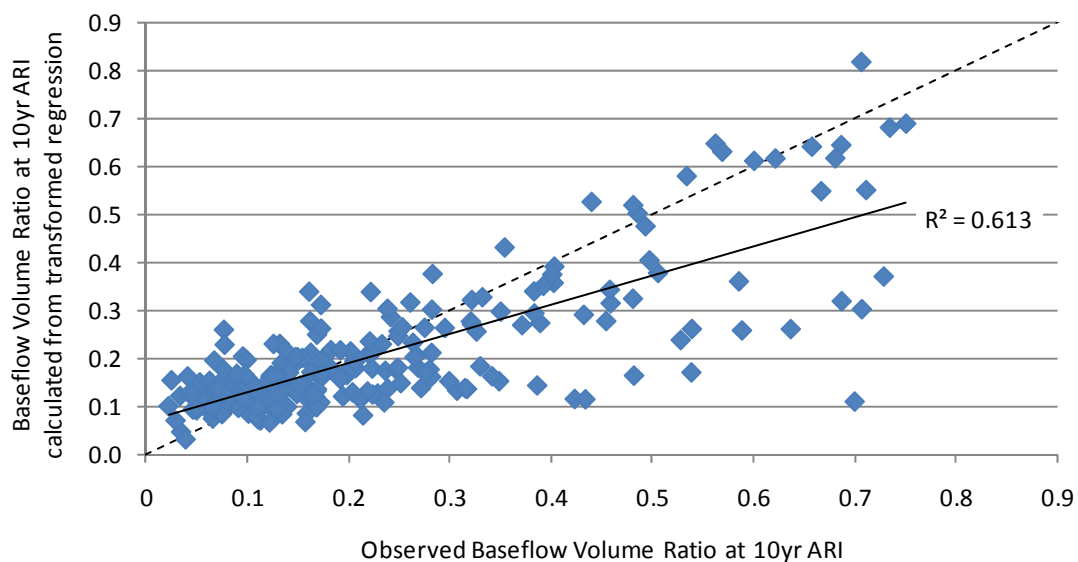
Figure 40 Application of regressions to clusters for Baseflow Volume Ratio

Evidence presented in this section displays that the predictive models developed are reasonably robust and can predict baseflow contribution to total streamflow (considering either the peak baseflow or volumetric contribution) with some confidence. The ability of these regression relationships to reasonably predict the Baseflow Peak Ratio and the Baseflow Volume Ratio based solely on catchment characteristics provides an opportunity to better understand baseflows on a wider scale. This approach separates the baseflow characteristics from the streamflow regime, making it suitable for application in catchments regardless of streamflow data availability. As such, these relationships were considered suitable for application across all regions of Australia, beyond the study catchments used in this work.

6. Relating baseflow contribution to surface runoff estimates

The method described in the previous sections has focussed on the determination of the proportion of the total measured streamflow that can be considered baseflow. The prediction equations allow the ratio between peak baseflow and peak total flow and the ratio between baseflow volume and total event volume to be estimated for ungauged catchments with the use of catchment characteristics.

Calculation of design flood peaks and volumes involves calculation of the surface runoff from, among other things, design rainfall intensity, losses and runoff routing assumptions. Design flood estimates tend to have information on the surface runoff component of the flood event rather than the total streamflow. For practical application, the estimation of a total design flood peak requires the addition of baseflow to the calculated surface runoff.

As such, the ratios documented in Section 5 are suitable to understand baseflow contribution relative to total streamflow but are not directly applicable for design purposes. To relate the estimated catchment ratios to a baseflow peak and/or volume that can readily be added to surface runoff calculations to determine a design flood peak, further relationships are introduced. Baseflow factors, represented by the generalised R_{BF} term in the relationship below, provide a measure of the proportion of the calculated surface runoff that represents baseflow.

$$R_{BFn} = \frac{Ratio_i}{1 - Ratio_i}$$

Equation 13

Where $Ratio_i$ represents either the Baseflow Peak Ratio or the Baseflow Volume Ratio, both measures of the baseflow contribution relative to the total streamflow

R_{BFn} where *n* represents either the flow or volume baseflow factor that relates baseflow contributions to the surface runoff

Equation 13 is relevant to both the Baseflow Peak and Baseflow Volume Ratio and converts the ratio value into a measure that is directly relevant for design flood applications.

Extending on the Baseflow Peak Ratio and the Baseflow Volume Ratio concepts, two factors are calculated as shown in Equation 14 and Equation 15. These are based on the baseflow statistics used for analysis in previous sections but are now presented in a way that is directly applicable for use in design calculations.

- **Baseflow Peak Factor (R_{BPF}):** calculated using Equation 13 and the Baseflow Peak Ratio. This factor is applied to the estimated surface runoff peak flow to give the baseflow peak flow. The relationship is shown in Equation 14.

$$R_{BPF} = \frac{Baseflow\ Peak\ Ratio}{1 - Baseflow\ Peak\ Ratio}$$

Equation 14

- **Baseflow Volume Factor (R_{BVF}):** calculated using Equation 13 and the Baseflow Volume Ratio. This factor is applied to the estimated surface runoff volume to give the event baseflow volume. The relationship is shown in Equation 15.

$$R_{BVF} = \frac{Baseflow\ Volume\ Ratio}{1 - Baseflow\ Volume\ Ratio}$$

Equation 15

Figure 41 provides a visual interpretation of this conversion process, where the relationship between total streamflow, surface runoff and baseflow is displayed with sample hydrographs. In the upper panel, the total streamflow hydrograph is presented, based on data from a gauged streamflow record. This represents the data used in much of the analysis is this study. The second panel presents a surface runoff hydrograph. Mathematically, this is equivalent to the total streamflow for the event less the event baseflow. The event baseflow is shown in the lower panel. Following this logic, the baseflow contribution to the total streamflow event can be calculated with knowledge of the surface runoff. In terms of the peak values, this is calculated simply as the multiplication of the Baseflow Peak Factor and the surface runoff peak flow. A similar relationship applies for the Baseflow Volume Factor.

For this purpose, the Baseflow Peak Ratio values, as estimated by the prediction equations, were converted into Baseflow Peak Factors to further develop the approach in a manner suitable for application in design flood studies. Similarly, the Baseflow Volume Ratio was converted into the Baseflow Volume Factor. These factors are carried forward in the following sections of this report.
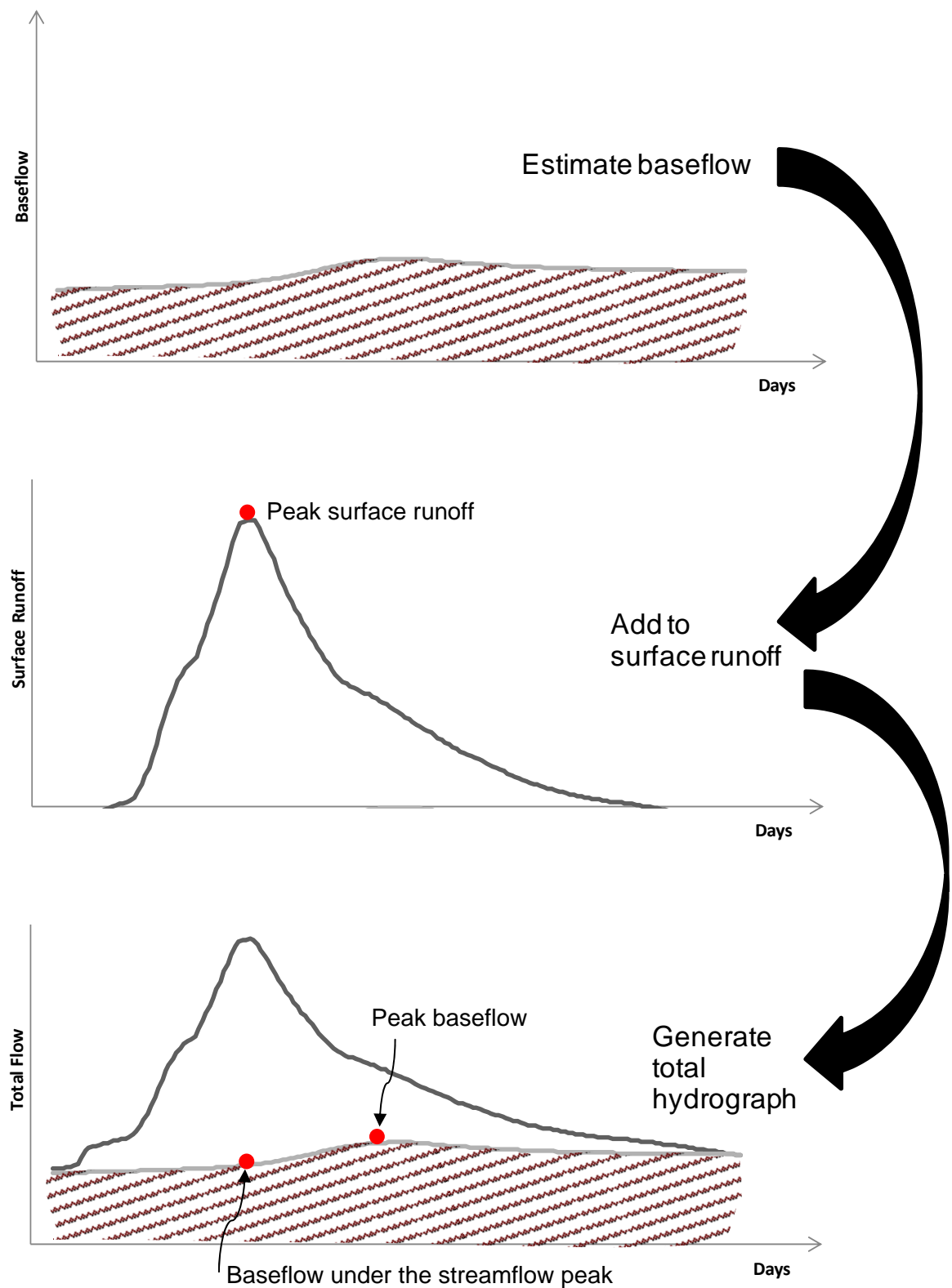
Figure 41 Relationship between total streamflow, surface runoff and baseflow

At this stage, it is also relevant to reintroduce the third baseflow measure extracted during the event analysis process. In Section 3.6, the Baseflow under Peak Ratio is described as the ratio of the baseflow at the time of the streamflow peak relative to the peak streamflow. Using the generic hydrograph in Figure 14, this is given by B/A. To generate a statistic that is useful for design flood purposes, this ratio is also converted in the equivalent factor relative to surface runoff, such that:

- **Baseflow Under Peak Factor (R_{BUPF}):** calculated using the Baseflow Under Peak Ratio. This factor can be applied to the estimated surface runoff peak flow to give the baseflow under the peak surface runoff. The relationship to the Baseflow Under Peak Ratio is shown in Equation 16.

$$R_{BUPF} = \frac{Baseflow\ Under\ Peak\ Ratio}{1 - Baseflow\ Under\ Peak\ Ratio}$$

Equation 16

Since the Baseflow Peak Factor reports on the baseflow peak while the Baseflow Under Peak Factor measures the baseflow under the streamflow peak, it is reasonable to consider that these statistics should be related. That is, an event with a high baseflow peak would likely also have a reasonably high baseflow value at the time of the streamflow peak. The converse is also considered reasonable. At each catchment, the relationship between these two factors was investigated with a scatter plot showing the Baseflow Peak Factor and the Baseflow Under Peak Factor for each event. A linear trendline was fitted to the points, as displayed in Figure 42.
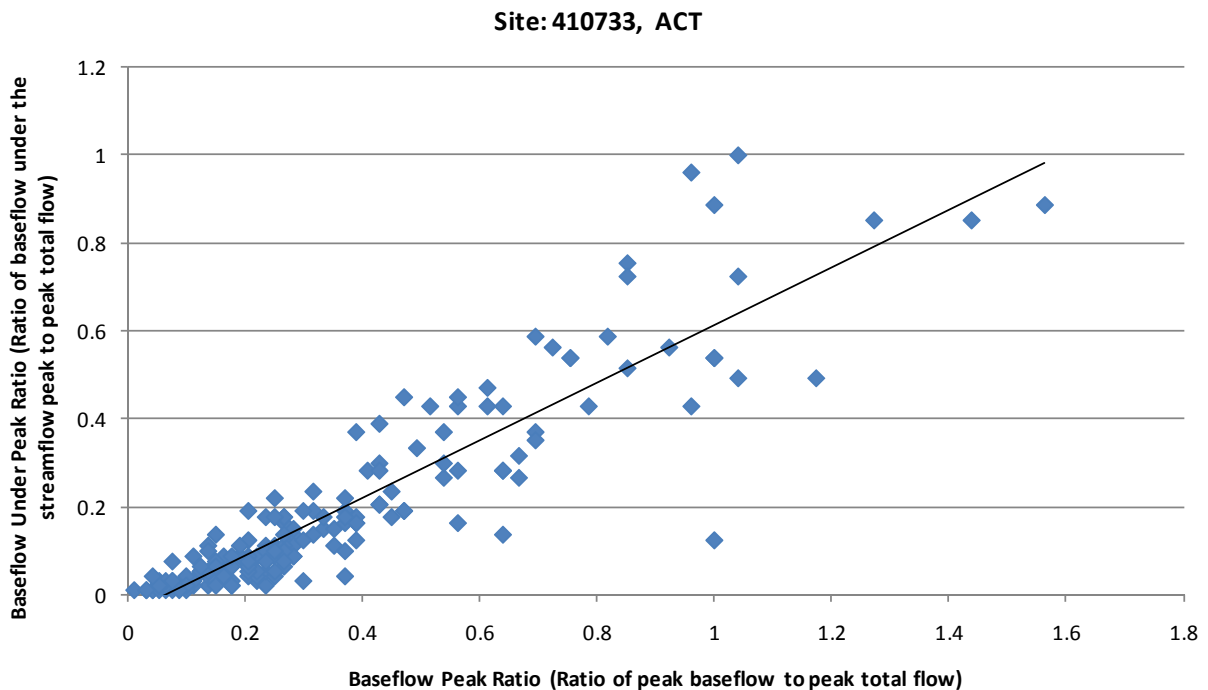
Figure 42 Relationship between Baseflow Peak Factor and Baseflow Under Peak Factor for catchment 410733 (Coree Creek at Threeways, ACT)

This approach was used to investigate the relationship between the Baseflow Peak Factor and the Baseflow Under Peak Factor for all catchments. The line of best fit for each catchment was

then consolidated into Figure 43, which shows that there is reasonable correlation between the two factors.
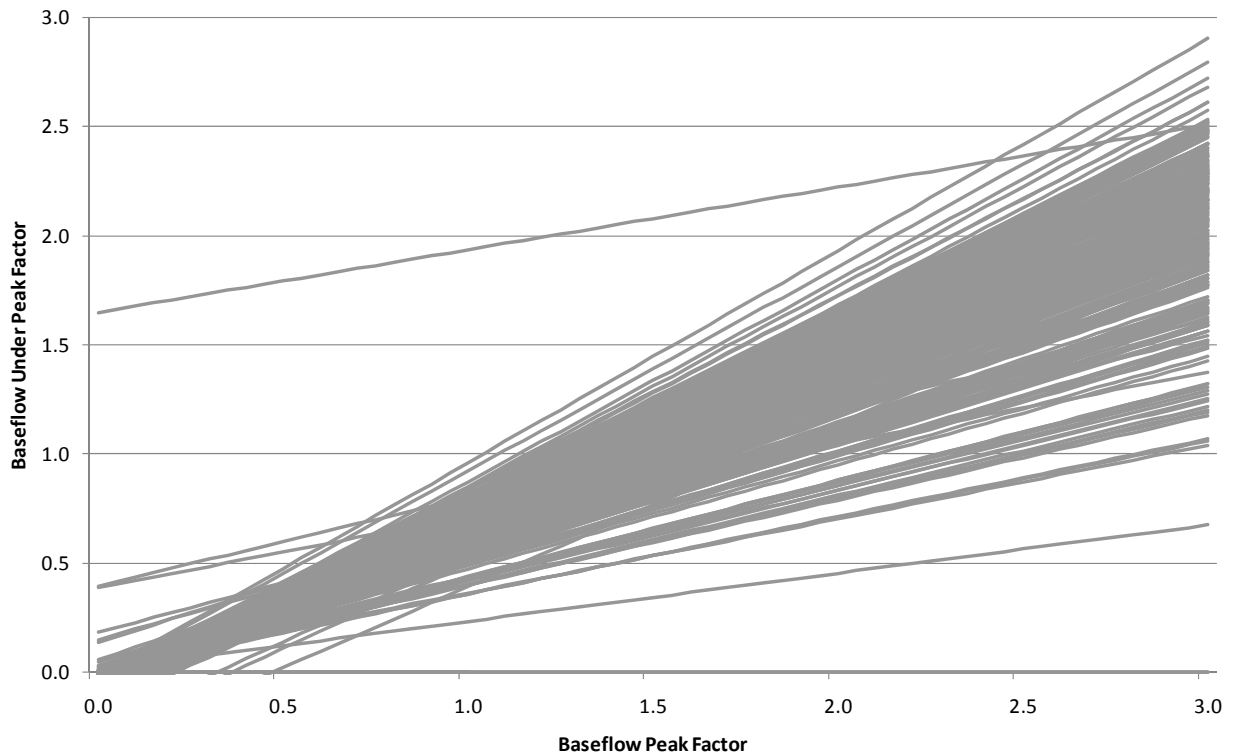
Figure 43 Relationship between the Baseflow Peak Factor and Baseflow Under Peak Factor for all study catchments

A similar comparison was also undertaken for the Baseflow Peak Factor and Baseflow Volume Factor (Figure 44). This figure displays significant variation between the trend lines, indicating a weaker overall relationship between these factors. In terms of physical processes, Figure 44 indicates that the magnitude of the peak baseflow (measured in the Baseflow Peak Factor) does not have a strong influence on the total baseflow volume during an event (measured in the Baseflow Volume Factor). This response is reasonable due to the importance of the event duration in determining the baseflow volume. Events with a high baseflow volume (high Baseflow Volume Factor) may occur as a result of either a long event duration or a baseflow peak of high magnitude. Only the later occurrence would correlate with a high Baseflow Peak Factor.
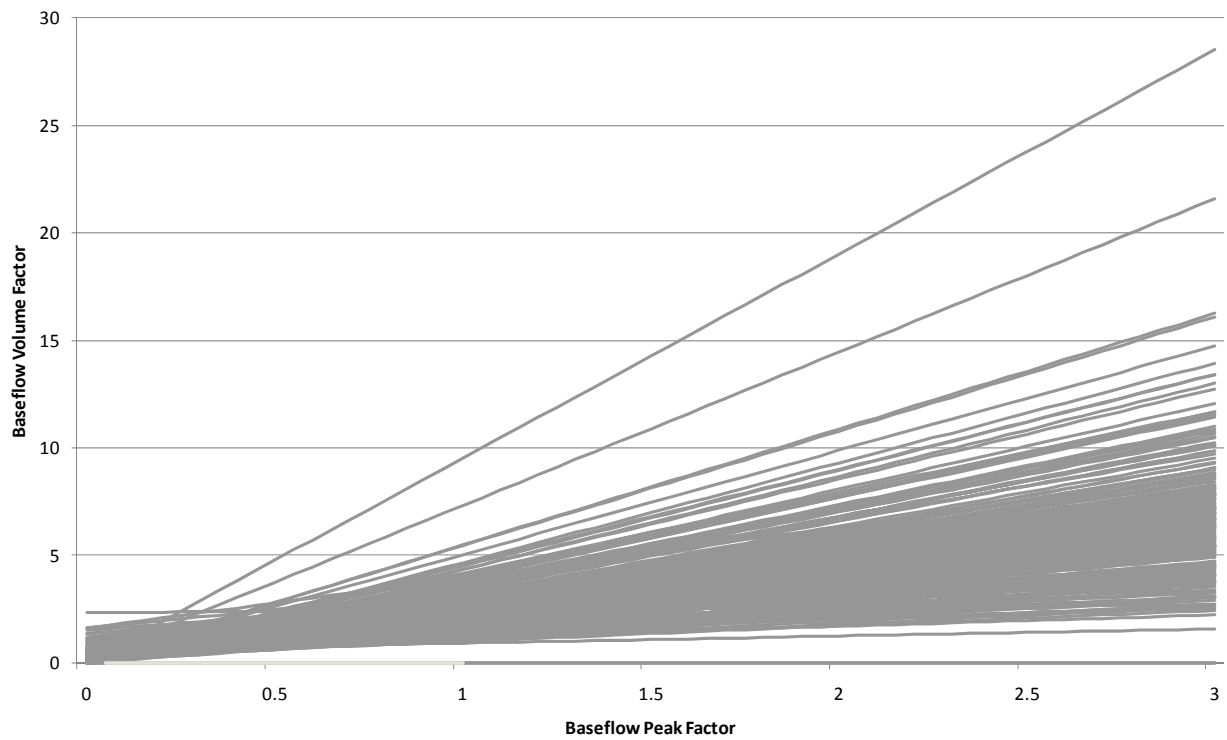
Figure 44 Relationship between the Baseflow Peak Factor and Baseflow Volume Factor for all study catchments

Figure 45 summarises the lines presented in Figure 43 and Figure 44. The average relationship was calculated across the full suite of catchments and is presented as a solid line. The dashed lines show the variability in the underlying data. This is presented as the standard deviation in the lines from Figure 43 and Figure 44. The blue lines represent the relationship between the Baseflow Peak Factor and Baseflow Under Peak Factor. The standard deviation lines lie reasonably closely to the average, reflecting the general consistency between these ratios. In contrast, the data demonstrating the relationship between the Baseflow Peak Factor and Baseflow Volume Factor (shown in red) displays a wider variation, particularly for high ratio values. Thus, these factors are less related.
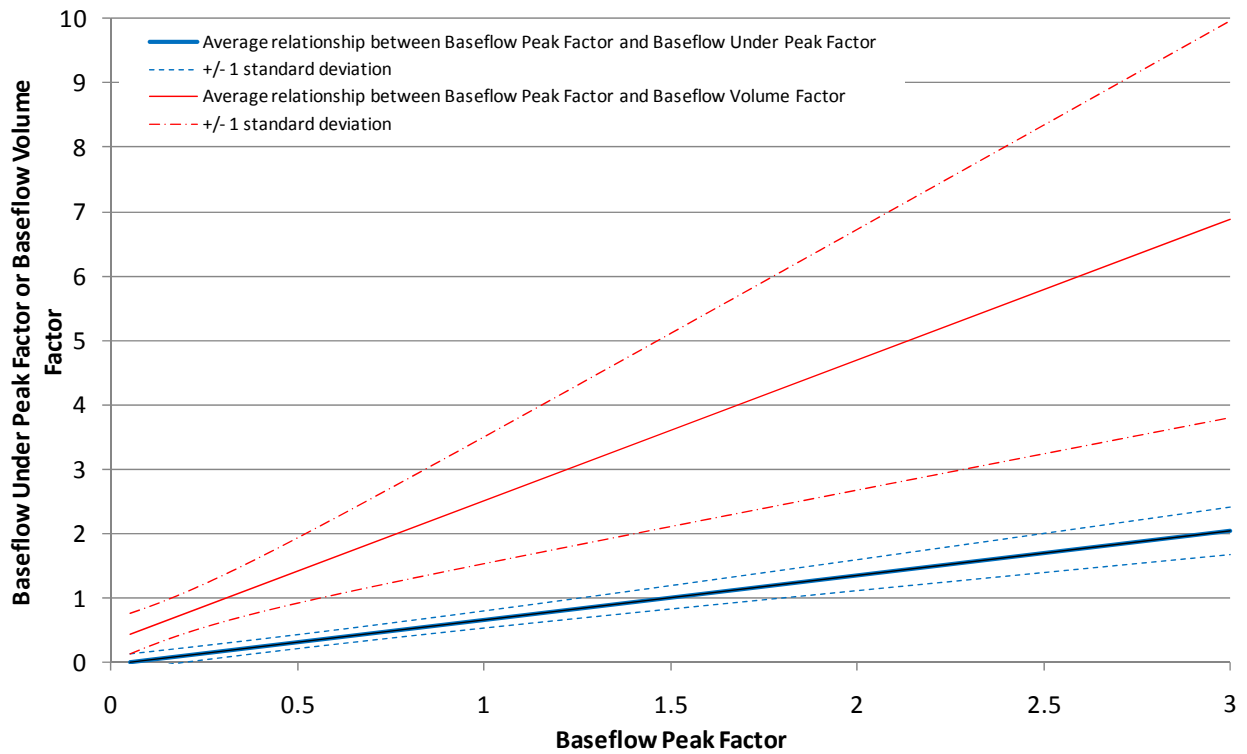
Figure 45 Average relationship between ratio values for all study catchments, based on Figure 43 and Figure 44

Investigation of the link between the Baseflow Peak Factor and the Baseflow Under Peak Factor yields an interesting relationship that can be used to simplify the approach for design flood studies. Equation 17 shows the average relationship between Baseflow Peak Factor and Baseflow Under Peak Factor from Figure 45. This relationship is valid across the full range of factor values and for all event sizes and ARIs.

<div align="center">Baseflow Under Peak Factor = 0.7 x Baseflow Peak Factor</div>

<div align="right">**Equation 17**</div>

The Baseflow Under Peak Factor calculated from Equation 17 yields an estimate of the baseflow at the time of the surface runoff peak flow, which is of relevance when wanting to understand the total streamflow peak. This relationship enables this baseflow statistic to be calculated directly from the Baseflow Peak Factor, such that the total streamflow peak can be calculated by:

$$Q_{Peak\ total\ streamflow} = 0.7 \times R_{BPF} \times Q_{Peak\ surface\ runoff} + Q_{Peak\ surface\ runoff}$$

<div align="right">**Equation 18**</div>

7. Application of the method

7.1. National application of regressions

The prediction equations described in Section 5 were developed using study catchments' characteristics. For these relationships to be applicable across Australia, it is necessary that the study catchments capture characteristics that represent the range of conditions observed nation-wide.

The Nested Catchments and Sub-Catchments for the Australian Continent data set (Australian Surveying and Land Information Group, 2000) were used to define spatial areas for application of the predictive equations. It is noted that the catchment characteristics that define the prediction equations are applicable across any spatial scale, and these catchments were selected simply to provide a meaningful representation at a reasonable spatial scale. The catchment characteristics across Australia were extracted from this national catchment dataset based on the total upstream catchment area rather than the individual catchment area. This reflects the standard method used to define contributing catchment areas for design flood assessments.

The catchment characteristics used in the cluster analysis and to define the predictive equations are listed in Table 7. The range of values observed for these characteristics within the study catchments and the nation-wide catchment data set are provided for comparison. For most characteristics, the range of conditions captured within the selected study catchments reasonably represents the range of conditions across the rest of the country.

Table 7 Range of catchment characteristics

Catchment characteristic	Study catchments		Nation-wide catchments	
	Minimum	Maximum	Minimum	Maximum
Top soil layer nominal field water capacity (m)	0.2	0.46	0	0.5
Top soil layer saturated hydraulic conductivity (mm/hr)	3.7	300	0	300
Top soil layer saturated volumetric water content (m)	0.4	0.62	0	0.63
Lower soil layer saturated hydraulic conductivity (mm/hr)	1.1	300	0	300
Lower soil layer thickness (m)	0.4	1.2	0	1.5
Maximum annual evapotranspiration (mm/yr)	267	1326	97	1443
Proportion of course-grained alluvial geology (%)	0.0	1.2	0	24
Proportion of fine-medium grained alluvial geology (%)	0.0	94	0	100
Proportion of undifferentiated alluvial geology (%)	0.0	100	0	100
Proportion of igneous and metamorphic geology (%)	0.0	100	0	100
Proportion of sandstone geology (%)	0.0	100	0	100
Proportion of woody vegetation (%)	0.05	100	0	100
Maximum slope (degrees)	2	54	0	54
Minimum slope (degrees)	0	0	0	5
Mean slope (degrees)	0	14	0	15

Each catchment in the national dataset was assigned to a cluster using the regression tree rules established in Figure 23. This was used to separate catchments based on mean slope, proportion of woody vegetation and the proportion of sandstone within the total upstream catchment area. Using this approach, it is acknowledged that the assigned cluster will change as the total upstream catchment area increases. That is, an individual catchment in the upper reaches of a system may be assigned to cluster 1 based on the average slope and proportion of woody vegetation in the catchment. Moving downstream, the next catchment may be assigned to cluster 2 whilst also incorporating the area of the first catchment. Figure 46 presents this upstream catchment area concept.
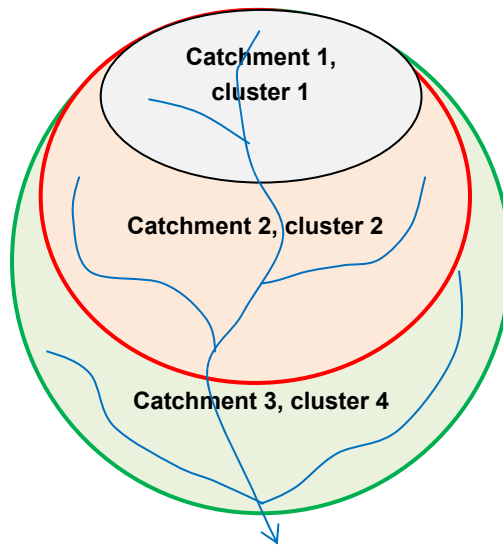
Figure 46 Catchment reporting areas

The spatial distribution of clusters across Australia is displayed in Figure 47. Catchments that are included in cluster 1 tend to be located within the mountainous part of the country, specifically the Great Dividing Range (Victoria, ACT, NSW and Queensland), MacDonnel Ranges (central NT), Darling Range (south-west WA), Hamersley Range (central-WA), Kimberley Plateau (northern WA and NT), Flinders Ranges (SA) and most of Tasmania. Cluster 2 catchments are also located within the mountainous parts of the country, however these have a higher proportion of woody vegetation cover than the catchments in cluster 1. Most of the cluster 2 catchments are located in Victoria and Tasmania, with some occurrences in the ACT and NSW.

Cluster 3 catchments are located across most states, and are characterised by relatively flat catchment slopes and high vegetation coverage. These tend to be in locations that have not been cleared, surrounding agricultural areas such as in south-west WA. Cluster 4 covers much of the central area of Australia and is defined by flat catchments with reasonably low vegetation cover and low proportions of sandstone geology.

Cluster 5 covers areas in northern Australia, west Queensland and around the Great Australian Bight coastline. These areas are characterised by relatively flat catchments with little vegetation with a high sandstone geology presence.
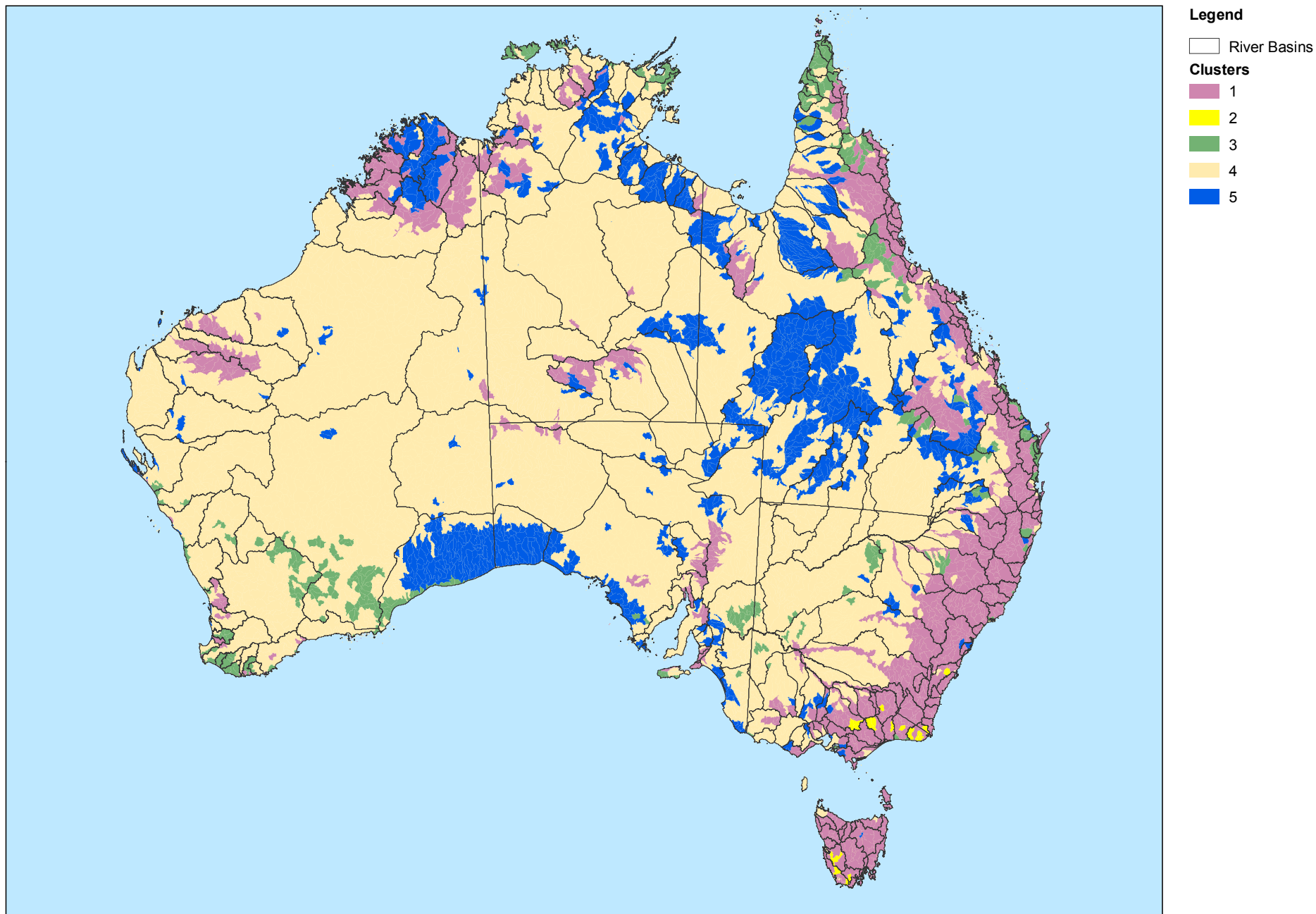
Figure 47 - Location of clusters, defined within total upstream catchment area (spatial distribution of clusters across Australia)

SINCLAIR KNIGHT MERZ

I:\VWES\Projects\VW04648\Technical\Spatial\ArcGIS\Fig42-Cluster_Locations.mxd. Date: 06/09/2010

The catchment characteristics were also used to evaluate the prediction equations in these cluster regions. This resulted in an estimate of the Baseflow Peak Ratio and the Baseflow Volume Ratio for each catchment across Australia, which relates to an ARI of 10 years. The conversion approach described in Section 6 was then applied to generate a Baseflow Peak Factor and Baseflow Volume Factor for all catchments, also relating to an ARI of 10 years. These factors were then mapped to display the baseflow contribution for design flood purposes for all locations across Australia. Figure 48 presents the Baseflow Peak Factor and Figure 49 shows the Baseflow Volume Factor at a national scale.

These maps present the factors that should be applied to surface runoff estimates and added to the runoff to determine the total design flood magnitude for a 10 year ARI flood. These hard copy maps can be used to provide an estimate of the baseflow contribution when the catchment of interest lies within an area where the relevant factor is readily identifiable. This map requires the user to simply identify the location of the catchment on the map to read off the factor of interest. This can then be directly applied to known runoff estimates to produce an estimate of baseflow contribution to the 1 in 10 year event. The relationships noted in Section 6 can be used to estimate the total streamflow peak flow or volume.

This approach provides a detailed estimate of the baseflow in most areas of Australia. In central Australia, where catchments suitable for analysis were limited in availability, it was considered inappropriate to provide this level of detail. The lack of suitable data means that the prediction equations cannot confidently be applied to these areas. Instead, a single Baseflow Peak Factor value is provided for each of the Lake Eyre, Bulloo-Bancannia and Western Plateau basins (Baseflow Peak Factors of 0.11, 0.08 and 0.14 respectively). These values were calculated as the average factors across each of these basins based on the predicted estimates. A similar approach was taken for the Baseflow Volume Factor estimates in central Australia. For the Lake Eyre drainage division, the average Baseflow Volume Factor is 0.51. In the Bulloo-Bancannia drainage division the factor is 1.19, while in the Western Plateau division a value of 0.15 is suggested for application. This approach ensures that the increased uncertainty in these parts of the country is taken into account, and users should consider the level of accuracy required in their analysis when applying the method to design flood studies in this region.

Users with access to GIS software can make use of a shapefile version of Figure 48 and Figure 49. This approach enables the catchment boundary for the study area to be directly overlayed, so that the appropriate factor can be extracted. As the reported factor is calculated using the whole area upstream of a catchment, as in Figure 46, the user needs only to identify and use the factor associated with most downstream overlapping part of their catchment. Aggregation of catchments is not required. This method is preferable in areas with a high degree of variation in factor values between catchments, especially where the baseflow contribution to streamflow may be high.

SKM

Legend

☐ Drainage Divisions
☐ River Basins

Peak Factor

- 0.00 - 0.10
- 0.11 - 0.20
- 0.21 - 0.30
- 0.31 - 0.40
- 0.41 - 0.50
- 0.51 - 0.60
- 0.61 - 0.70
- 0.71 - 0.80
- 0.81 - 0.90
- 0.91 - 1.00
- 1.01 - 1.20
- 1.21 - 1.40
- 1.41 - 1.60
- 1.61 - 1.80
- 1.81 - 2.00
- 2.01 - 2.20
- 2.21 - 2.40
- 2.41 - 2.60
- 2.61 - 2.80
- 2.81 - 3.00

The hatched part of the map reflects the limited data availability in central Australia. Drainage divisions X, XI and XII are shown with the average Baseflow Peak Factor:
For division XI Baseflow Peak Factor = 0.11
For division XI Baseflow Peak Factor = 0.08
For division XII Baseflow Peak Factor = 0.14

- **Figure 48 - Map of Baseflow Peak Factor for ARI of 10 years**

SINCLAIR KNIGHT MERZ

I:\VWES\Projects\VW04648\Technical\Spatial\ArcGIS\Fig46-Baseflow_Peak_Factor.mxd. Date: 06/09/2010

0   500
Kilometres

Legend

- ☐ Drainage Divisions
- ☐ River Basins

Volume Factor

🟥	0.00 - 0.10
🟧	0.11 - 0.20
🟧	0.21 - 0.30
🟧	0.31 - 0.40
🟧	0.41 - 0.50
🟨	0.51 - 0.60
🟨	0.61 - 0.70
🟨	0.71 - 0.80
🟨	0.81 - 0.90
🟩	0.91 - 1.00
🟩	1.01 - 1.20
🟩	1.21 - 1.40
🟩	1.41 - 1.60
🟦	1.61 - 1.80
🟦	1.81 - 2.00
🟦	2.01 - 2.20
🟦	2.21 - 2.40
🟦	2.41 - 2.60
🟦	2.61 - 2.80
🟦	2.81 - 3.00

The hatched part of the map reflects the limited data availability in central Australia. Drainage divisions X, XI and XII are shown with the average Baseflow Volume Factor.
For division X Baseflow Volume Factor = 0.51
For division XI Baseflow Volume Factor = 1.19
For division XII Baseflow Volume Factor = 0.15

Figure 49 - Map of Baseflow Volume Factor for ARI of 10 years

SINCLAIR KNIGHT MERZ

I:\VWES\Projects\VW04648\Technical\Spatial\ArcGIS\Fig47-Baseflow_Volume_Factor.mxd. Date: 06/09/2010

0   500
Kilometres

7.2. Variation in baseflow peak with ARI

The map presented in Figure 48 provides an approach to estimate the baseflow peak for events that have an ARI of 10 years. However, for practical purposes a method is required that relates these to a broad range of event magnitudes. Consistent with the method described in Section 4.2 for the total streamflow related ratios, trendlines describing the variation in Baseflow Peak Factor with ARI were developed for each catchment. An example plot is presented in Figure 50, and similar plots were generated for all catchments. In Figure 50, each point on the chart reflects a single flood event extracted at the site. A high degree of scatter is evident for small events, reflecting the high frequency of occurrence of these events under a variety of conditions. Fewer data points are available for larger event sizes, which is a function of the length of the streamflow record at each location and the infrequent nature of these events. These general characteristics were typical across all catchments.

Appendix C presents these plots for all catchments considered as a part of this study.
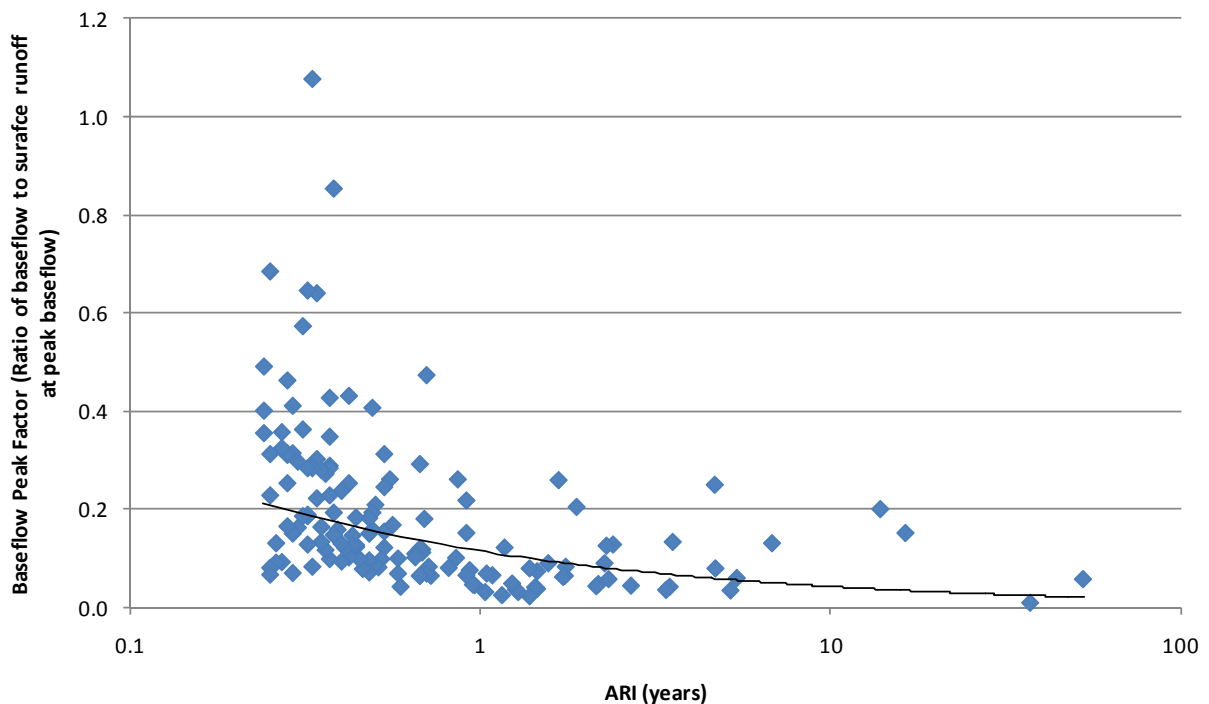
Figure 50 Variation in Baseflow Peak Factor with ARI of total flow peak for catchment 218001 (Tuross River @ Tuross Vale, NSW)

Exceptions to the negative power relationship were observed for some catchments where the value of the Baseflow Peak Factor appeared invariant or showed a slightly increasing relationship with ARI. In these cases, the slope of the power function was modified to be invariant with ARI based on the average Baseflow Peak Factor at the site.

Figure 51 summarises the power functions for all catchments investigated in this study. For infrequent flood events, most of the trend lines tend to converge to Baseflow Peak Factor values less than one. There are a few exceptions to this, mostly for catchments that are invariant in the

factor value with ARI. The variability in factor value is much greater for smaller events, with values ranging between 0.01 and 3.02. Except for three sites, most factor values are below two.

Given the large number of catchments summarised in this figure, it is difficult to distinguish clear patterns in the variation in the Baseflow Peak Factor with increasing ARI. However some general patterns can be observed. The most obvious is the high degree of variation in the lines presented, indicating significantly different baseflow behaviour during the range of flood events across the catchments considered. This variation is evident in two aspects:

- The variation in factor value for any given ARI between the catchments; and
- The different gradients of the lines representing the variation with ARI.
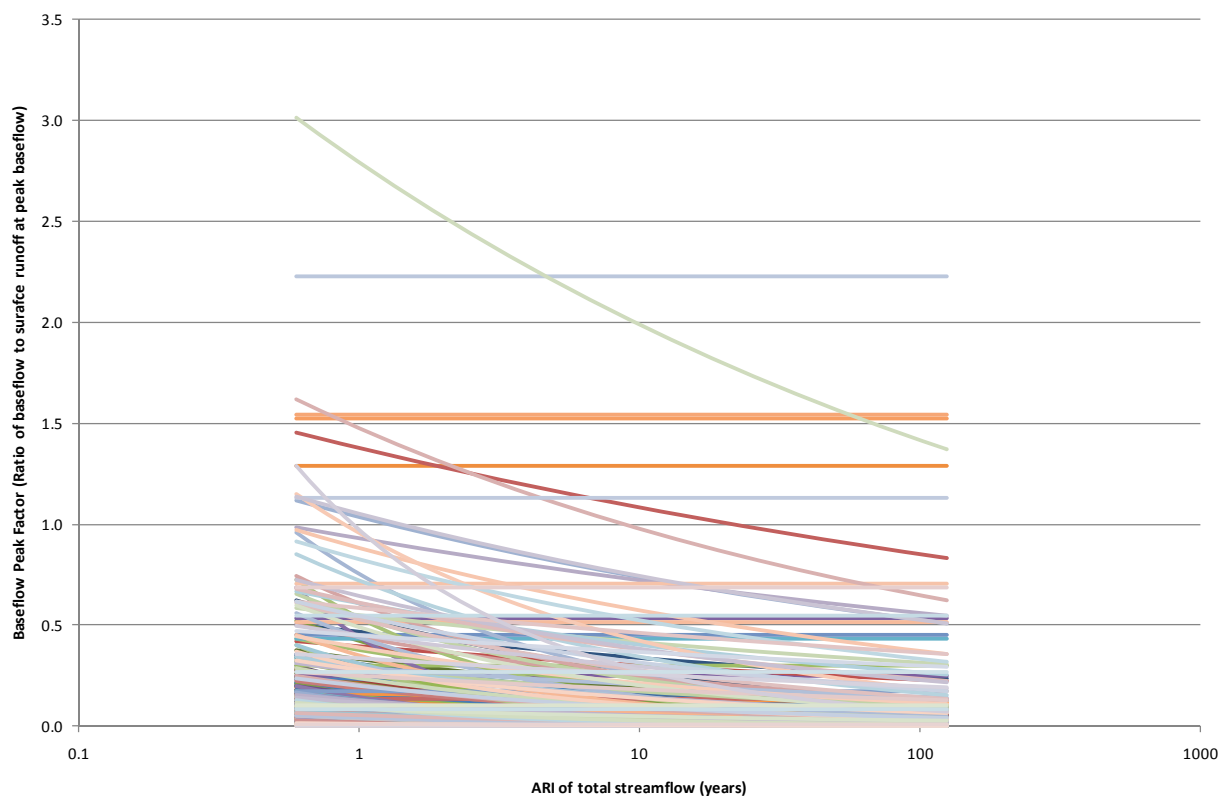
Figure 51 Summarising the variation in the Baseflow Peak Factor with ARI of the total streamflow

Figure 51 can be further summarised to provide a more meaningful comparison across a range of catchments. Figure 52 presents data for the sixty catchments with more than 40 years of record. The availability of data over such a long duration provides more confidence around the behaviour for more infrequent flood events. In this figure, the trendlines from Figure 51 have been standardised by the Baseflow Peak Factor value for an ARI of 10 years. This provides a consistent basis from which to understand how the contribution of baseflow varies across the range of event magnitudes.
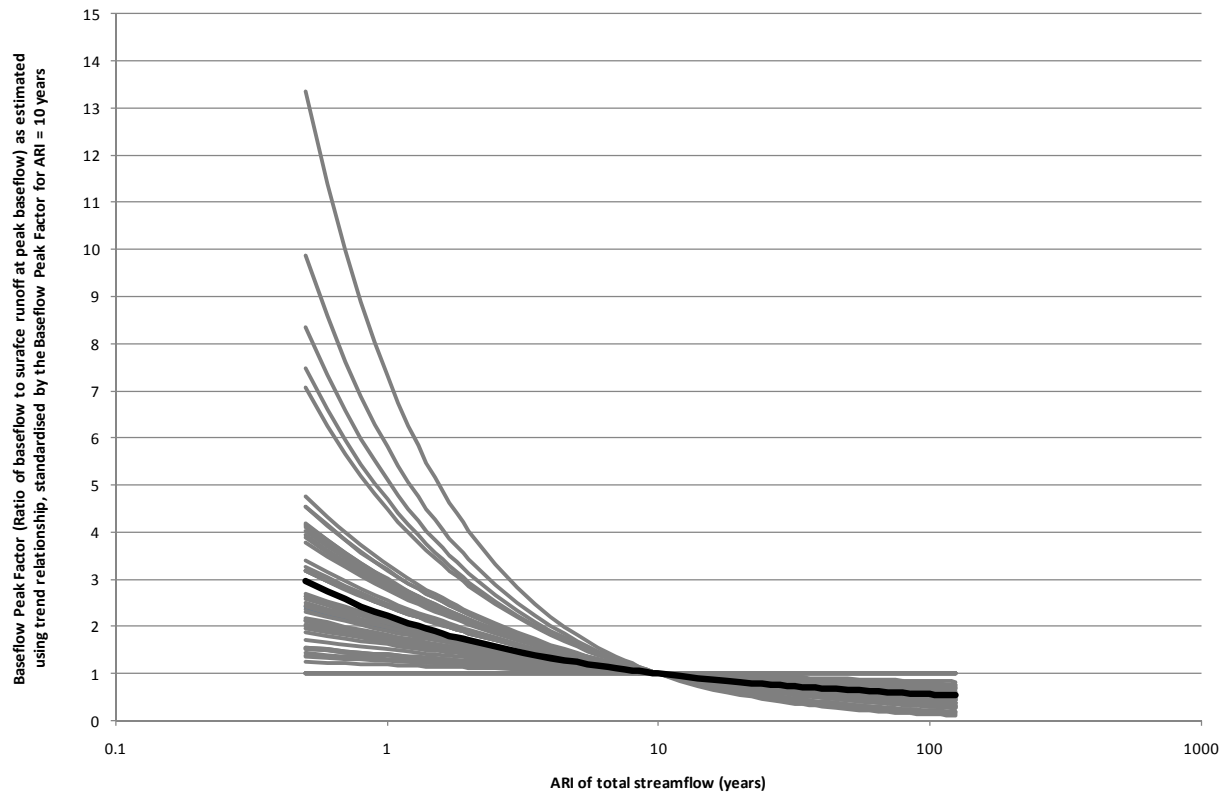
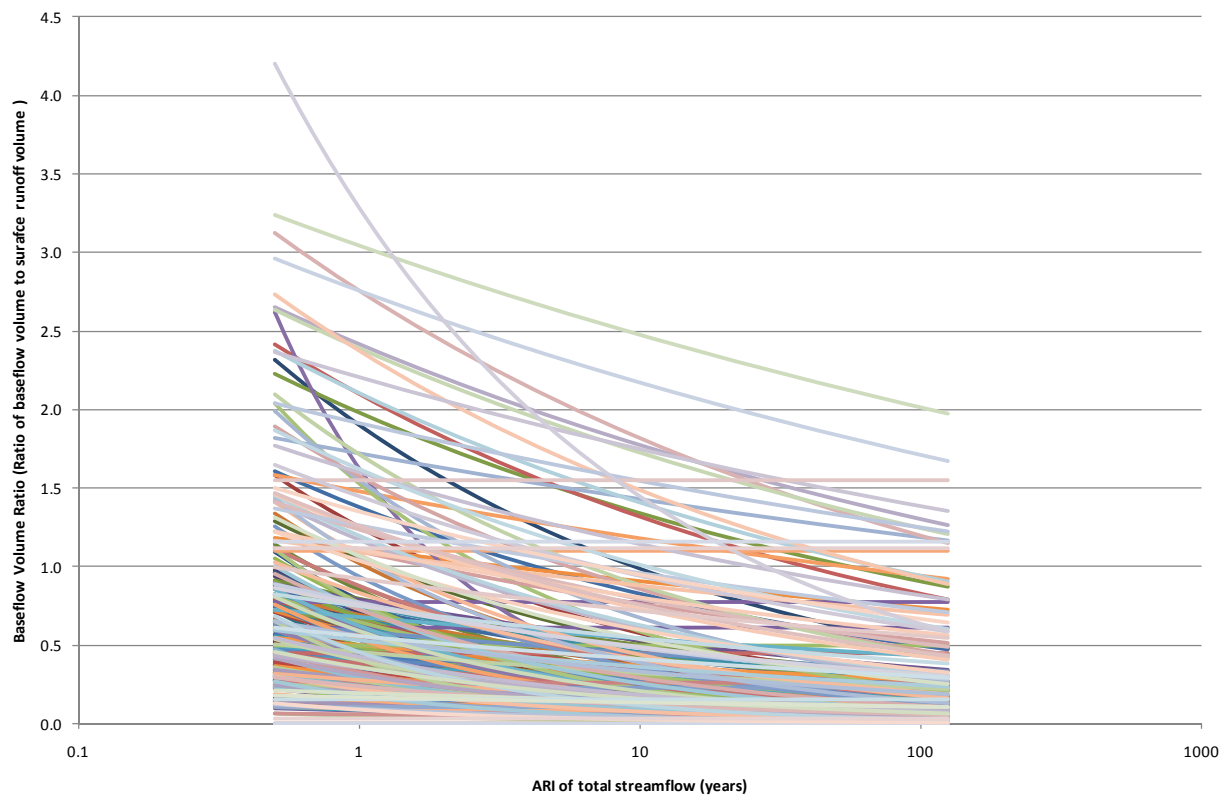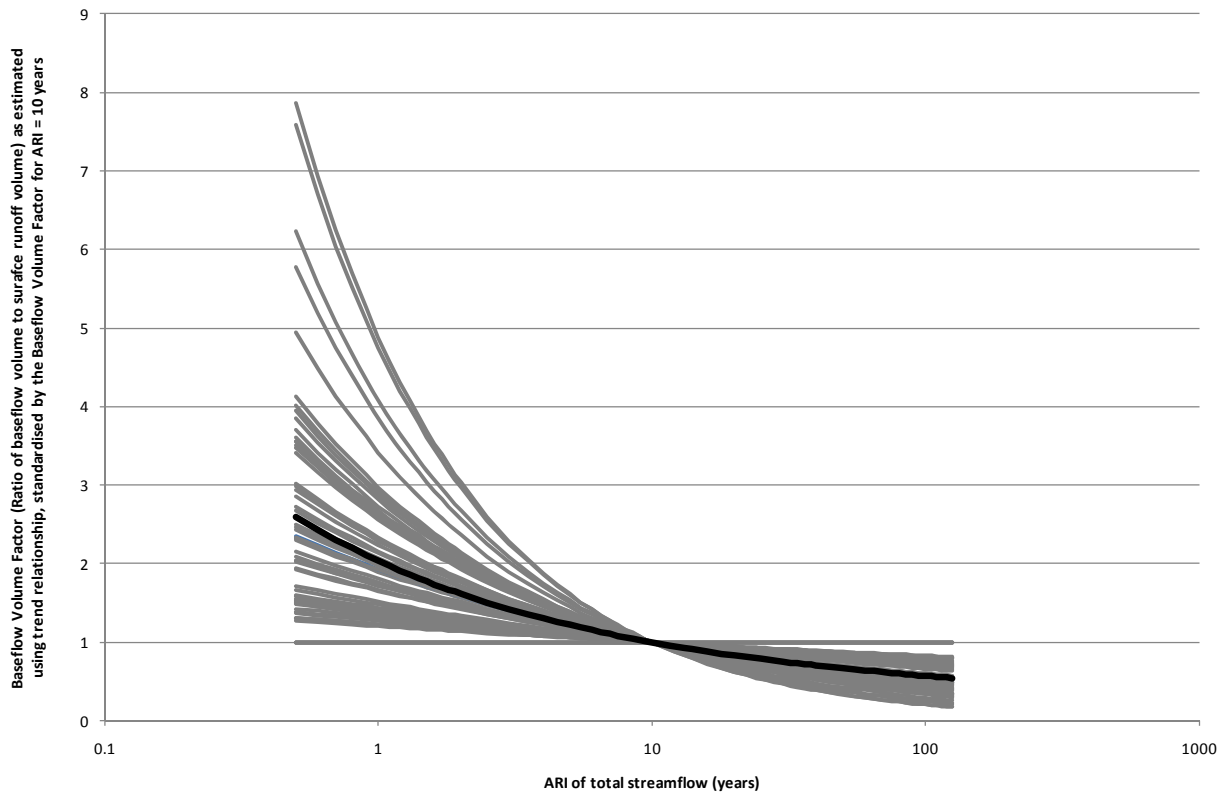**Figure 52 Summarising the variation in the Baseflow Peak Factor with ARI of the total streamflow for catchments with more than 40 years of streamflow data. The relationship for each catchment has been standardised by the factor value for an ARI of 10 years.**

The information on the left side of Figure 52 highlights the variation in behaviour for frequent flood events across the different catchments. The value of the standardised factor ranges between 1 and 30, indicating that the baseflow contribution to an event with ARI = 0.5 years can be up to 14 times greater than that for a 1 in 10 year event. In contrast, the contribution for a 1 in 100 year event can be as low as 14% of that of a 1 in 10 year event.

The solid black line in Figure 52 presents the average response across the 60 catchments with more than 40 years of data. For frequent events with ARI of 0.5, this average response indicates that the baseflow contribution is around three times the contribution for an event with ARI of 10 years. Events with ARI of 100 years have a baseflow contribution of approximately half of the 1 in 10 year contribution.

The frequency histogram of the Baseflow Peak Factor values across the same sixty sites with long (> 40 years) streamflow records is presented in Figure 53. At each location, the average factor value for all events with ARI of less than 1 year was calculated. Similarly, at each location all events with an ARI between 1 and 10 were used to determine the average Baseflow Peak Ratio value for this range of floods. The larger events were used to calculate the average ratio value for ARIs greater than 10 years. The average ratio values for each collection of events were standardised using the Baseflow Peak Ratio value for an ARI of 10 years, as calculated using the relevant relationship from Figure 51. Figure 53 displays the variation in standardised ratio values across all sites.
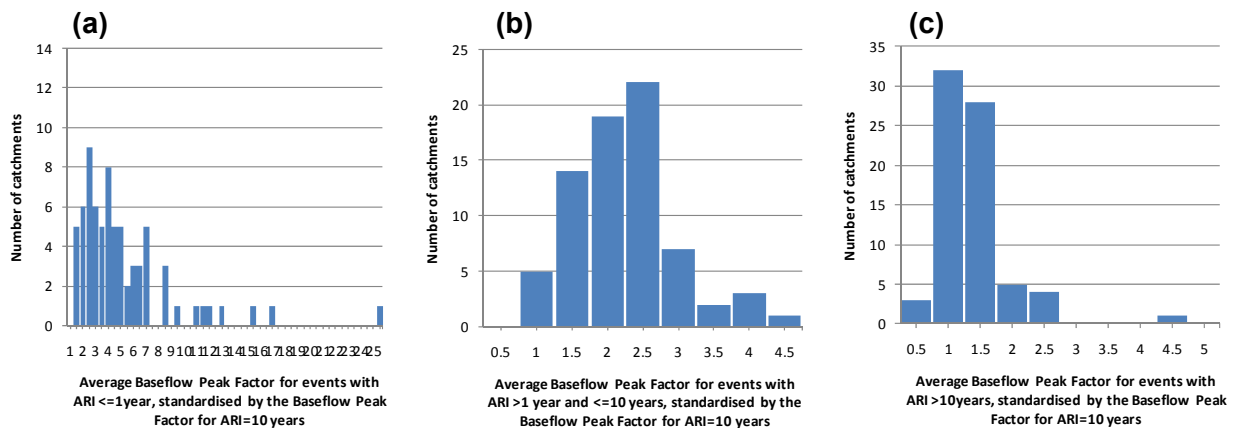
Figure 53 Frequency histogram of Baseflow Peak Factor for (a) ARI ≤ 1; (b) 1 < ARI ≤ 10; and (c) ARI > 10

The variation in the Baseflow Peak Factor for frequent flood events is demonstrated in Figure 53a, with the average ratio value ranging between half and 25 times the Baseflow Peak Factor value for an ARI of 10 years. However, the median of the catchments analysed in this study have an average factor value of up to 3.4 times the ARI of 10 years ratio value. The average standard deviation in factor values for these frequent events is 3.8. Medium sized events (ARIs between 1 and 10) are shown in Figure 53b, and indicate that there is less variation in the factor value compared to more frequent events. The median ratio value is 1.9 times that of events with an ARI of 10 years and the standard deviation is 1.3. Larger events (with ARI greater than 1 in 10) show even less variability. The median ratio value across all events at all catchments is 1.06 times that of events with ARI of 10 years, with a standard deviation of 0.62.

7.3. Variation in baseflow volume with ARI

The map presented in Figure 49 provides an approach to estimate the baseflow volume for events that have an ARI of 10 years. Again, a method is required that can be applied across a broad range of event magnitudes. Trendlines describing the variation in Baseflow Volume Factor with ARI were developed for each catchment. An example of this analysis is shown in Figure 54. Appendix D presents a consolidated package of these plots for each catchment considered in this study.
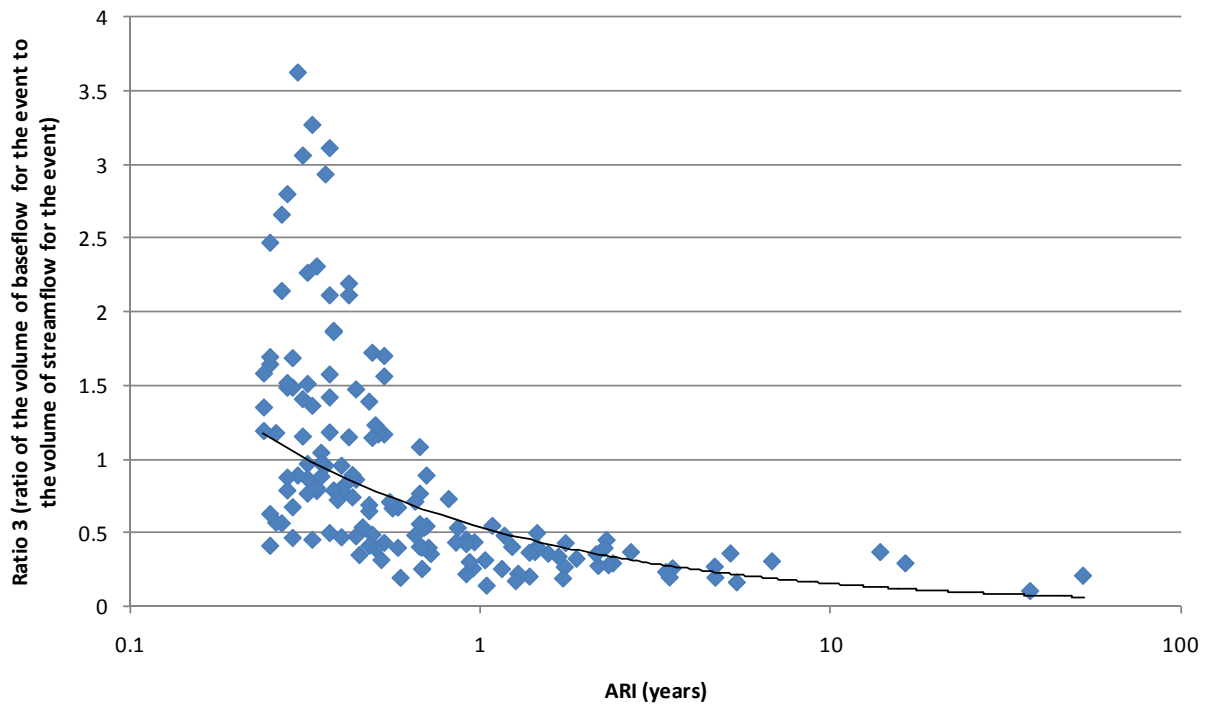
Figure 54 Variation in Baseflow Volume Factor with ARI of total flow peak for catchment 218001 (Tuross River @ Tuross Vale, NSW)

Figure 55 summarises the trendlines from all catchments. The behaviour displayed in Figure 55 is generally consistent with that presented in Figure 51. A high degree of variation across the catchments is demonstrated for frequent events. The factor values for the smallest events can be up to 4.20, indicating that the volume of baseflow for the event is more than four times that of surface runoff. In other catchments, the baseflow volume can be as little as three percent of the surface runoff volume for these frequent events. The baseflow contribution tends to decrease for infrequent events, however this is significant as observed for the Baseflow Peak Factor.

Figure 55 Summarising the variation in the Baseflow Volume Factor with ARI of the total streamflow

A sub-set of catchments (those with more than 40 years of record) are presented in Figure 56. In these, the relationships have been standardised by the Baseflow Volume Factor value for an ARI of 10 years. The variation in the Baseflow Volume Factor value for frequent events is observed by the spread in the lines on the left side of this figure. The value of the standardised factor ranges between 1 and 8, indicating that the baseflow contribution for an event with ARI of 0.2 can be up to eight times greater than that for a 1 in 10 year event. This behaviour for frequent events is less variable than that observed for the Baseflow Peak Factor in Figure 52. For events with an ARI of 100 years, the standardised Baseflow Volume Factor value can be as low as 20% of the 10 year event value.

The solid black line in Figure 56 represents the average response across the catchments with more than 40 years of data. For frequent events, the Baseflow Volume Factor can be considered to be approximately 3.6 times that for events with an ARI of 10 years. The factor for more infrequent events is approximately 57% of the 1 in 10 year factor value.

Figure 56 Summarising the variation in the Baseflow Volume Factor with ARI of the total streamflow for catchments with more than 40 years of streamflow data. The relationship for each catchment has been standardised by the factor value for an event with ARI of 10 years.

Frequency histograms of the Baseflow Volume Factor values are presented in Figure 57 for the sites with more than 40 years of data. The variation in ratio values for frequent events is further displayed in Figure 57a. The Baseflow Volume Ratio for events with an ARI less than 1 year can be up to 5.5 times the Baseflow Volume Ratio for events with ARI of 10 years. The median value is 1.8 with average standard deviation of 0.7. Medium sized events are less variable in nature, with the median Baseflow Volume Ratio value 1.4 times the value for an ARI of 10 years, with average standard deviation 0.5. Larger events have a median ratio value of 0.9, and average standard deviation 0.3.
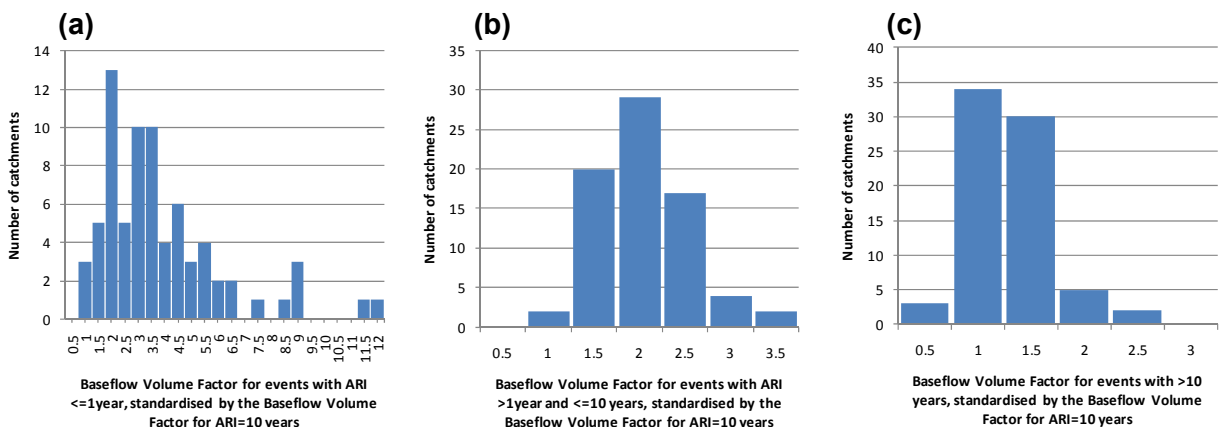
Figure 57 Frequency histogram of the Baseflow Volume Factor for (a) ARI ≤ 1; (b) 1 < ARI ≤ 10; and (c) ARI > 10

7.4. Estimating Baseflow Peak Factor and Baseflow Volume Factor for a range of event magnitudes

The maps presented in Figure 48 and Figure 49 provide information on the Baseflow Peak Factors and Baseflow Volume Factors for events with an ARI of 10 years. Based on the information presented in the previous sections, Table 8 shows the ARI factors that should be applied to the 10 year ARI Baseflow Peak Factor for events of other magnitudes. These have been calculated based on the average response in Figure 52. Similarly, Table 9 shows the ARI factors that should be applied to the 10 year ARI Baseflow Volume Factor for events of other magnitudes based on the average behaviour in Figure 56.

Table 8 ARI Factors, F_{ARI}, to be applied to the 10 year ARI Baseflow Peak Factor to determine the Baseflow Peak Factor for events of various ARIs

ARI (years)	ARI Factor, F_{ARI}
0.5	3.0
1	2.2
2	1.7
5	1.2
10	1.0
20	0.8
50	0.7
100	0.6

Table 9 ARI Factors, F_{ARI}, to be applied to the 10 year ARI Baseflow Volume Factor to determine the Baseflow Volume Factor for events of various ARIs

ARI (years)	ARI Factor, F_{ARI}
0.5	2.6
1	2.0
2	1.6
5	1.2
10	1.0
20	0.8
50	0.7
100	0.6

For events of ARIs not shown in Table 8 and Table 9, Figure 58 can be used to determine an appropriate ARI factor. The relationships shown in Figure 58 represent the average of the study catchments shown in Figure 52 and Figure 56.

These factors can be directly applied to the 1 in 10 year ARI Baseflow Peak Factor or Baseflow Volume Factor to readily scale the value to reflect a variety of event magnitudes. This enables the method to be applied on a wide scale, for any event size between 1 in 0.5 year and 1 in 100

year ARIs. As expected, these scaling factors indicate the baseflow contribution to flood events is largest for the most frequent events. For rare events, baseflow is only a small proportion of the total surface runoff.

i
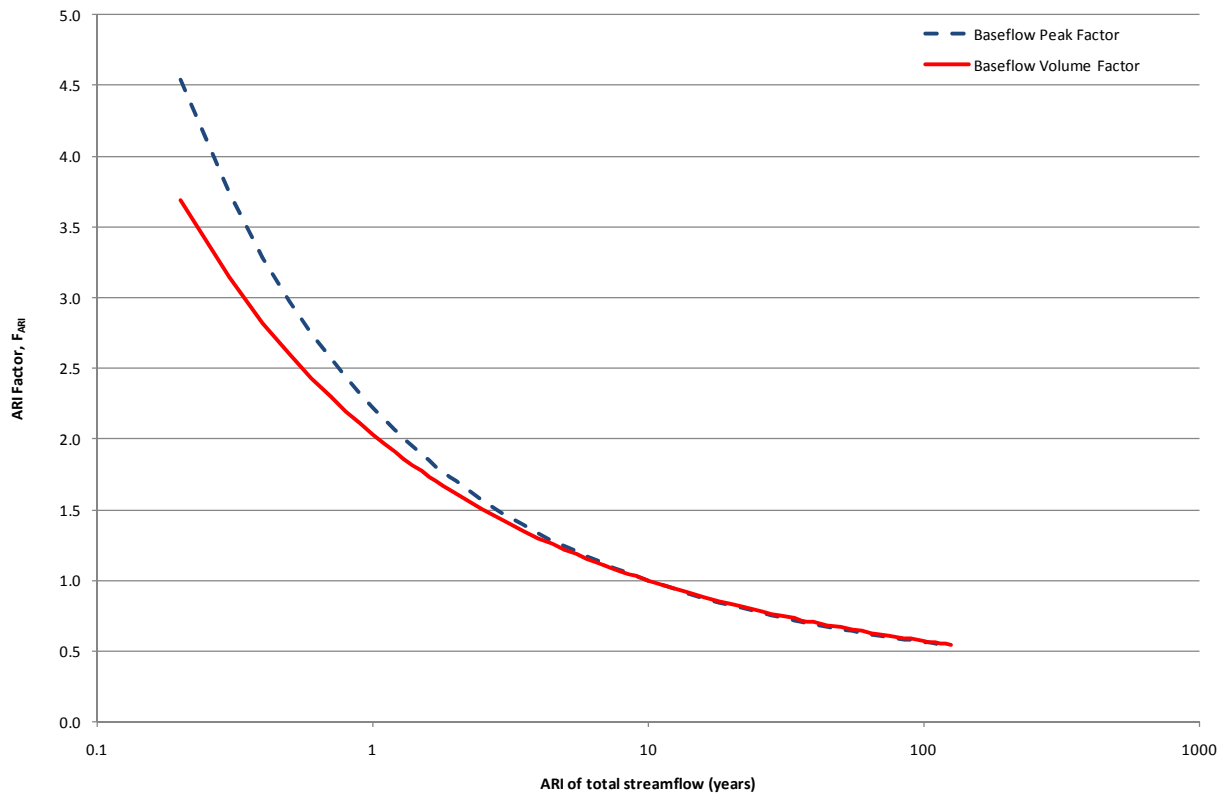
Figure 58 ARI Factors, F$_{ARI}$, to be applied to the 10 year ARI Baseflow Volume Factor to determined the Baseflow Volume Factor for events of various ARIs

Equation 19 shows the final factor to be applied to the calculated surface runoff to determine the event peak baseflow.

$$R_{BPF} = F_{ARI}R_{BPF,10yrARI}$$

Equation 19

Equation 20 shows the final factor to be applied to the calculated surface runoff volume to determine event baseflow volume.

$$R_{BVF} = F_{ARI}R_{BVF,10yrARI}$$

Equation 20

Where further details of the variation are required, for instance to undertake Monte Carlo analysis, the histograms presented in Figure 53 and Figure 57 provide information about the range of factors for different event magnitudes.

7.5.   Summary of application method

The Baseflow Peak Factors and Baseflow Volume Factors determined from the maps for a particular catchment should be applied to design flood estimation using the following procedures.
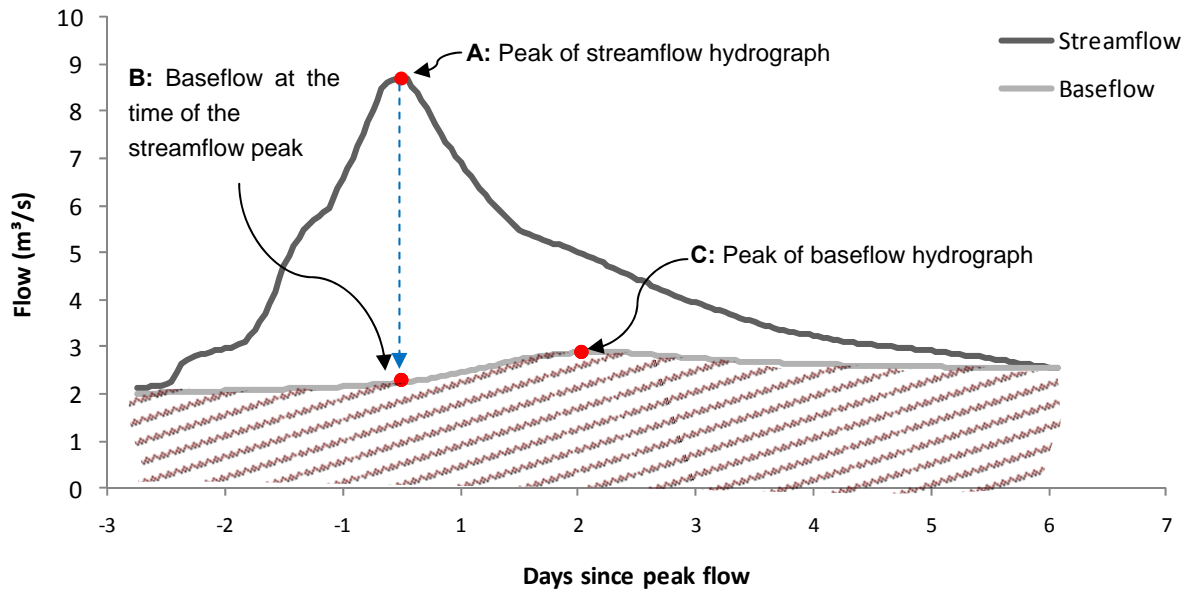
Figure 59 Key characteristics for calculation in a design flood hydrograph

To calculate the peak baseflow (point C in Figure 59):

1. Determine the Baseflow Peak Factor for a 10 year ARI ($R_{BPF,10yrARI}$) from Figure 48.

2. Determine the ARI factor corresponding to the event ARI using Table 8, Table 9 or Figure 58. Apply to the 10 year Baseflow Peak Factor as in Equation 19 to determine the Baseflow Peak Factor for the event magnitude of interest.

3. Apply the Baseflow Peak Factor to the calculated peak surface runoff as in Equation 21.

$$Q_{Peak\ baseflow} = R_{BPF} Q_{Peak\ surface\ runoff}$$

Equation 21

To calculate the baseflow under the peak streamflow (point B in Figure 59):

1. The Baseflow Peak Factor (R_{BPF}) calculated for the appropriate event ARI as above should be used in Equation 22 to calculate the Baseflow Under Peak Factor (R_{BUPF}).

$$R_{BUPF} = 0.7 \times R_{BPF}$$

Equation 22

2. R_{BUPF} should be used as in Equation 23 to calculate the baseflow under the peak streamflow.

$$Q_{Baseflow\ under\ peak\ streamflow} = R_{BUPF} Q_{Peak\ surface\ runoff}$$

Equation 23

To calculate the total streamflow peak (point A in Figure 59):

1. Calculate the baseflow under the streamflow peak for the appropriate ARI as above

2. Add the baseflow under the streamflow peak calculated using Equation 23 to the calculated peak surface runoff as in Equation 24.

$$Q_{Peak\ streamflow} = Q_{Peak\ surface\ runoff} + Q_{Baseflow\ under\ peak\ streamflow}$$

Equation 24

To calculate the total baseflow volume for an event (red hashed in Figure 59):

1. Determine the Baseflow Volume Factor for a 10 year ARI ($R_{BVF,10yrARI}$) from Figure 49

2. Determine the ARI factor corresponding to the event ARI using Table 8, Table 9 or Figure 58. Apply to the 10 year Baseflow Volume Factor as in Equation 20 to determine the Baseflow Volume Factor (R_{BVF}) for the event.

3. Apply the Baseflow Volume Factor to the calculated surface runoff volume as in Equation 25.

$$V_{Baseflow} = R_{BVF} V_{Surface\ Runoff}$$

Equation 25

To calculate the total streamflow volume for an event (blue hashed in Figure 59):

1. Calculate the baseflow volume for the event using the appropriate ARI factors.

2. The baseflow volume calculated using Equation 25 should be added to the calculated surface runoff as in Equation 26.

$$V_{Total\ streamflow} = V_{Surface\ runoff} + V_{Baseflow}$$

Equation 26

8.   Sensitivity of the method to scale

A comparison of the factors obtained for each study catchment can be compared to the outputs from the national dataset to understand the sensitivity to catchment area. Using the spatial data used in Figure 48 and Figure 49, the Baseflow Peak Factor and the Baseflow Volume Factors for each of the 236 study catchments were extracted from the national dataset. These extracted factor values were directly compared to the factors predicted based on the local catchment characteristics.

Figure 60 presents this comparison for the Baseflow Peak Factor, shown using a log scale to visualise the low values in the data. Most of the data points are grouped around low values of the Baseflow Peak Factor, close to the 1:1 line which indicates that the upscaling of the regression tree clusters and the application of the prediction equations on a national scale produces comparable results to when the analysis is undertaken at the catchment scale. Whilst there are a larger number of results occurring slightly above the 1:1 line, the actual deviation from the 1:1 line is small and it is not considered that these points cause a bias towards either overestimating or underestimating the overall ability of the approach to achieve a consistent result. A small number of catchments display behaviour that deviates more significantly away from the 1:1 line. These occurrences are scattered either side of the 1:1 line. This indicates that where information on the local conditions is available (such as a reasonable streamflow record) it should be utilised in preference to the estimation approach developed in this study.
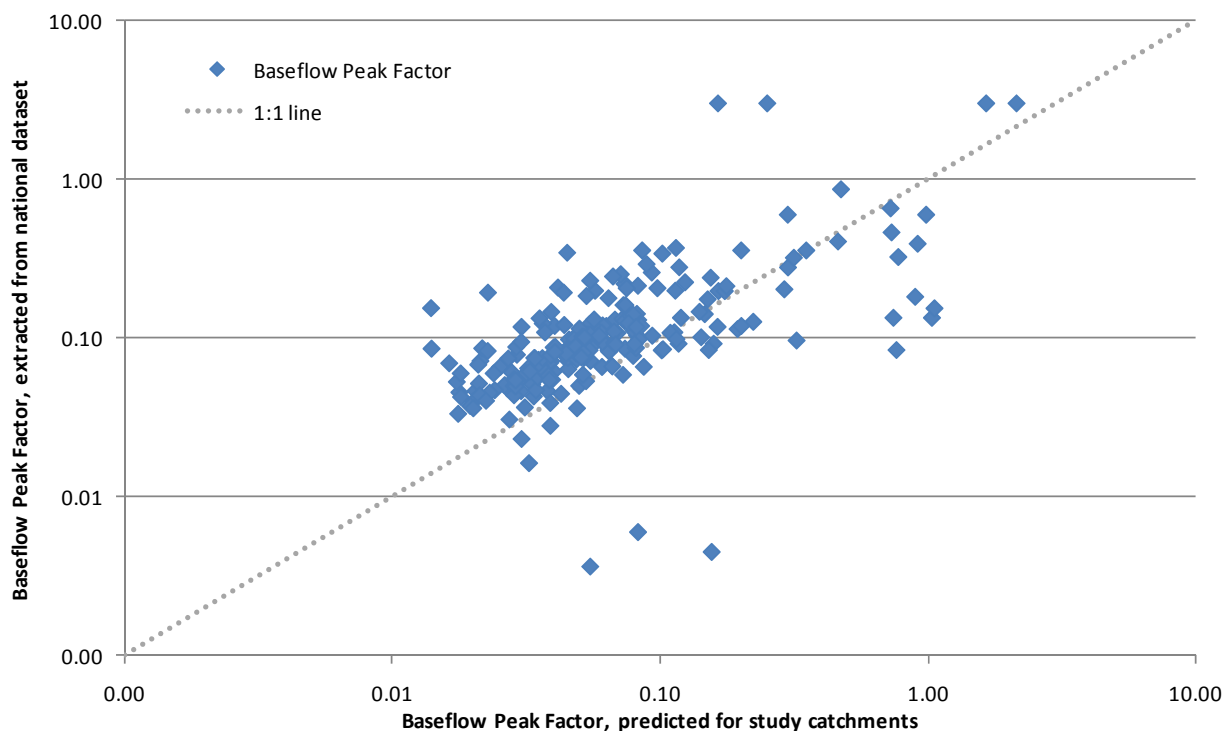
Figure 60 The Baseflow Peak Factor as predicted for the individual study catchments compared to the Baseflow Peak Factor extracted from the national dataset

Figure 61 presents a similar plot for the Baseflow Volume Factor. Most of the data points have a factor value less than 1 and are grouped together around the 1:1 line. There is some scatter around the 1:1 line for the highest and lowest factor values.
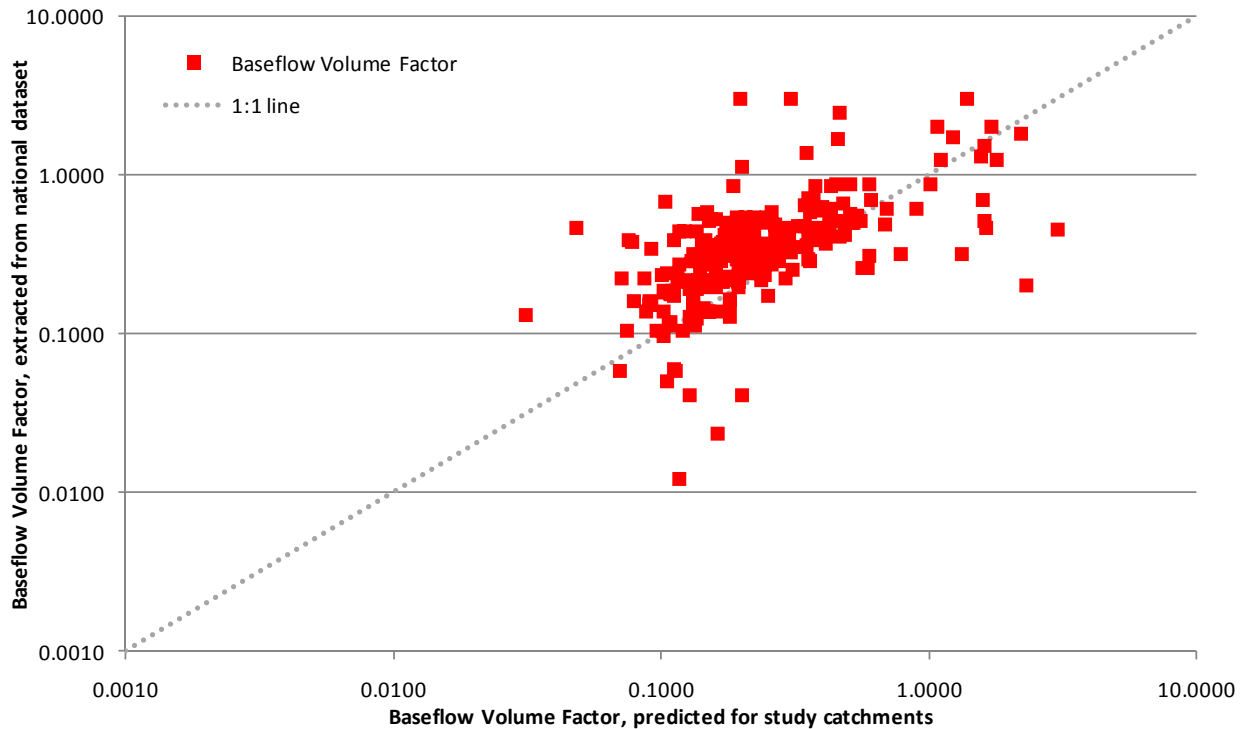
Figure 61 The Baseflow Volume Factor as predicted for the individual study catchments compared to the Baseflow Volume Factor extracted from the national dataset

These differences in the predicted Factors are not considered significant when considered in the context of the uncertainty associated with characterising the baseflow from streamflow records and then subsequently predicting baseflow using local catchment characteristics. In a practical sense, it is proposed that this approach be applied to ungauged catchments where recorded data is not available. This would provide the most appropriate method to estimate baseflow in such situations.

Visual inspection of the data indicates that the uncertainty in the prediction of the Baseflow Peak Factor and Baseflow Volume Factor increases for catchments with a catchment area much smaller than the scale of the catchment data used in the application of this approach at the national scale (as described in Section 7.1). Users should be conscious of the importance of the spatial scale where this approach is to be applied to catchments significantly smaller than those represented in the national catchment dataset. If considered relevant, the prediction equations presented in Table 5 and Table 6 could be re-evaluated on a more appropriate spatial scale. Note that this would require sourcing the relevant datasets, and recalculation of the Baseflow Peak/Volume Ratio and Baseflow Peak/Volume Factor.

This comparison provides information on the reliability of the prediction equations when applied outside of the study catchments. It also provides a measure of the influence of the scale of the application, since the catchment sizes used for the national analysis do not relate specifically to the size of the catchment of interest. This later point is particularly relevant when considering the use of this data in day-to-day applications, since design flood estimation could be undertaken for catchments of any size. In particular, a given catchment of interest could be only a small fraction of the size of the catchments used in Figure 48 and Figure 49. In some locations, this would

influence the underlying catchment characteristics that define the prediction equations, and may consequently result in discrepancies in the predicted factors for application. However, as displayed in Figure 60 and Figure 61 above, this impact is reasonably minor and in most locations the national dataset provides a reasonable estimate of the baseflow contribution to the design flood event.

9. Conclusions

In most cases, baseflow is a minor component of extreme floods but can potentially be a significant component of smaller flood events. Larger contributions of baseflow occur where the catchment geology naturally produces high yielding aquifers with large baseflows. The current version of Australian Rainfall and Runoff provides guidance for estimating surface runoff, but there is little advice on the incorporation of baseflow into design flood events.

ARR Update Project 7 aimed to develop a method for calculating and adding baseflow contribution to design flood estimates. Phase 1 of the project focussed on the physical processes of groundwater-surface water interaction and theoretical approaches to baseflow separation. The identified methods were applied to various case study catchments across Australia in order to develop a suitable approach for more wide scale application.

Phase 2 of Project 7 covered the analysis of 236 catchments across Australia, the development of prediction equations to estimate baseflow parameters and the development of a method for the application of these to design estimates for catchments across Australia.

Baseflow series were produced for 236 study catchments in Phase 2, using the Lyne-Hollick filter method developed in Phase 1. The baseflow characteristics of each of the catchments was analysed in the Phase 2 study catchments using two main parameters and a third secondary parameter:

1. **Baseflow Peak Ratio**: Ratio of the peak baseflow (C) to the peak streamflow (A) is given by C/A (shown in Figure 14).

2. **Baseflow Volume Ratio**: The event baseflow index (BFI) is given by the total baseflow volume for the duration of the event divided by the total streamflow volume. This is the ratio of the shaded areas in the example hydrograph in Figure 14.

3. **Baseflow Under Peak Ratio**: Ratio of the baseflow at the time of the streamflow peak (B) to the peak streamflow (A), given by B/A in Figure 14.

The relationship between the Baseflow Peak Ratio and the Baseflow Volume Ratio with ARI is generally observed to be approximately described by a negative power function. The 10 year ARI event was used as the basis for characterising baseflow at the sites.

A regression model was developed to allow the estimation of the 10 year ARI ratios for ungauged catchments. The model included a regression tree that separated the study catchments into clusters of similar catchments on the basis of catchment characteristics. A regression for each of these clusters was developed using catchment characteristics as independent variables.

The ratios calculated from the regression models describe the proportion of total streamflow that can be considered baseflow. The application of the predictive models to design flood estimation requires the transformation of Baseflow Peak Ratio to a factor. This factor can be applied to a calculated surface runoff peak to determine the baseflow peak of a flood event. Similarly, the Baseflow Volume Factor is calculated from the Baseflow Volume Ratio. It is applied to the calculated surface runoff volume to determine the baseflow volume for an event.

The baseflow under the peak runoff is calculated from the peak baseflow factor. The use of these three factors with surface runoff calculations allow the addition of a flow or volume to flood estimates calculated from parameters such as critical storm duration, areal reduction factor, storm temporal pattern and design losses.

The factors developed in this way were calculated and applied across the whole of Australia using the Nested Catchments and Sub-Catchments for the Australian Continent data set (Australian Surveying and Land Information Group, 2000). Factors can be read from hard copy maps provided for less precise estimates or use in areas of uniform baseflow characterisation. Alternatively, a catchment boundary can be laid over the shape file to determine a more precise factor value. The reported factor is calculated using the whole area upstream of a catchment so only the factor associated with most downstream overlapping part of the catchment is required.

Calculation of the baseflow contributions to events with ARIs other than 10 years is based on the relationships between the ARIs and analysed baseflow ratios developed for all study catchments. These are summarised into a standardised relationship with ARI for both factors. The ARI factors based on all the study catchments are applied to the 10 year ARI factors.

The use of the factors in design flood estimation will allow more accurate estimates of:

- peak baseflow for an event;
- baseflow under the peak streamflow for an event;
- total streamflow peak of an event;
- total baseflow volume for an event; and
- total streamflow volume for an event .

10. Acknowledgements for data supply

Where available, the Bureau of Meteorology provided data from their consolidated collection of streamflow data from across Australia.

Data from South Australia has been sourced from the Department of Water, Land and Biodiversity Conservation. The following disclaimer is provided with the data: the information/data is based on current records of water information. It is presented in good faith as the best available record in this format. The Minister for Environment and Conservation is not obliged to issue revised information/data. The Minister for Environment and Conservation accepts no liability for effects arising from the use of the information.

Some of the Western Australian streamflow data used in this report were obtained from the Water Information System (WIN) and Hydstra database, managed by the Department of Water, Water Information Provision Section, Perth, Western Australia. Information supplied by the Department of Water is protected by the Copyright Act 1968. That copyright belongs to the State of Western Australia. Apart from any fair dealing for the purpose of private study, research, criticism or review, as permitted under the Copyright Act 1968, no part may be reproduced or reused for any purpose without the written permission of the Department of Water.

Streamflow data from NSW were sourced from version 9 and 9.2 of the Pinneena database (2006 and 2008). This data is compiled by the NSW Department of Water and Energy, Dumaresq-Barwon Border Rivers Commission, Manly Hydraulics Laboratory, Murray-Darling Basin Authority and State Water.

Streamflow data for the majority of gauge locations in Tasmania were primarily sourced from the Water Assessment Branch of the Department of Primary Industries, Parks, Water and Environment based in Hobart, Tasmania. Data captured by Hydro Tasmania were sourced directly from this organisation.

Streamflow data for sites in the Northern Territory was obtained from the Natural Resources Division of the Department of Natural Resources, Environment, The Arts and Sports in Darwin, and is used in accordance with the conditions of licences for the use of digital data and information (licences DNRM2006/0038.427A and B).

Thiess Hydrographic Services, as custodians of water data in Victoria, provided streamflow data for some Victorian sites.

The Department of Environment and Resource Management provided streamflow data for Queensland catchments as required.

Streamflow data from sites within the ACT were provided under an agreement with ACTEW.

11.   References

Australian Surveying and Land Information Group (2000) Nested Catchments and Sub-Catchments for the Australian Continent

Boughton, W. (1999) A Century if Water Resources in Australia 1900-1999, The Institution of Engineers, Australia

Bureau of Meteorology (2000) Climatic Atlas of Australia – Rainfall

Bureau of Meteorology (2001) Climatic Atlas of Australia – Evapotranspiration

Bureau of Rural Sciences (2000) Australian Groundwater Flow Systems - National Land and Water Resources Audit, January 2000

Cordery, I. (1998) The unit hydrograph method of flood estimation. Section 2 of Book V, Estimation of design flood hydrographs. Australian Rainfall and Runoff – A guide to flood estimation. The Institution of Engineers, Australia

CRC for Catchment Hydrology (2004) Soil Hydrological Properties of Australia. Prepared by Andrew Western and Neil McKenzie for the CRC for Catchment Hydrology.

Department of Climate Change (2008) Forest extent and change. Version 4, September 2008.

Geoscience Australia (2006) GEODATA TOPO 250K series 3 topographic data

Hutchinson, M.F., Stein, J.A. and Stein, J.L. (2007) GEODATA 9 second Digital Elevation Model (DEM-9S) Version 3, December 2007

Institution of Engineers, Australia (1987) Australian Rainfall and Runoff. (Ed: Pilgrim, D.H.) Institution of Engineers, Australia

Lowe, L (2009) Addressing uncertainties associated with water accounting. PhD Thesis. Department of Civil and Environmental Engineering, University of Melbourne, March 2009.

Lyne, V. And Hollick, M. (1979) Stochastic time-variable rainfall-runoff modelling. Institute of Engineers Australia National Conference. Publ. 79/10, 89-93

Morgan, J.N. and Sonquist, J.A. (1963) Problems in the analysis of survey data, and a proposal. Journal of the American Statistical Association, 58, 415-434

Neal, B.P., Nathan, R.J. and Evans, R. (2004) Survey of baseflows in unregulated streams of the Murray-Darling Basin. 9th Murray-Darling Basin Groundwater Workshop, 17-19 February 2004. Bendigo, Victoria.

Peel, M.C., Chiew, F.H.S., Western, A.W. ad McMahon, T.A. (2000) Estimation of unimpaired streamflow data and regionalisation of parameter values to estimate streamflow in ungauged catchments. Report prepared for the National Land and Water Resources Audit. Theme 1 – Water availability. July 2000

Pelleiter, P.M. (1988) Uncertainties in the single determination of river discharge: a literature review. Canadian Journal of Civil Engineeing. Volume 15, pages 834-850.

Raymond, O.L., Liu, S.F., Kilgour, P. (2007) Surface geology of Australia 1:1,000,000 scale, Victoria 3rd edition [Digital Dataset] Canberra: The Commonwealth of Australia, Geoscience Australia. http://www.ga.gov.au

Raymond, O.L., Liu, S.F., Kilgour, P., Retter, A.J., Connolly, D.P. (2007) Surface geology of Australia 1:1,000,000 scale, Victoria - 3rd edition [Digital Dataset] Canberra: The Commonwealth of Australia, Geoscience Australia. http://www.ga.gov.au

Raymond, O.L., Liu, S.F., Kilgour, P., Retter, A.J., Stewart, A.J., Stewart, G. (2007) Surface geology of Australia 1:1,000,000 scale, New South Wales 2nd edition [Digital Dataset] Canberra: The Commonwealth of Australia, Geoscience Australia. http://www.ga.gov.au

Scanlon, P. (2007) Using AS3778.2.3 to Estimate the uncertainty in flow estimates at flow monitoring sites in Gippsland. University of Melbourne.

Sheather, S.J. (2009) A Modern Approach to Regression with R, Springer Science+Business Media, New York

SKM (2003) Sustainable Diversion Limit Project: Estimation of Sustainable Diversion Limit parameters over winterfill periods in Victorian catchments. Prepared for the Department of Sustainability and Environment

SKM (2009) Project 7: Baseflow for catchment simulation. Phase 1 – Selection of baseflow separation approach. Report prepared for Engineers Australia.

SKM (2010) Project 7: Baseflow for catchment simulation. Data collection and catchment characteristics. Report prepared for Engineers Australia

Stewart, A.J., Sweet, I.P., Needham, R.S., Raymond, O.L., Whitaker, A.J., Liu, S.F., Phillips, D., Retter, A.J., Connolly, D.P., Stewart, G. (2008) Surface geology of Australia 1:1,000,000 scale, Western Australia [Digital Dataset] Canberra: The Commonwealth of Australia, Geoscience Australia. http://www.ga.gov.au

Whitaker, A.J., Champion, D.C., Sweet, I.P., Kilgour, P., Connolly, D.P. (2007) Surface geology of Australia 1:1,000,000 scale, Queensland 2nd edition [Digital Dataset] Canberra: The Commonwealth of Australia, Geoscience Australia. http:/www.ga.gov.au

Whitaker, A.J., Glanville, D.H., English, P.M., Stewart, A.J., Retter, A.J., Connolly, D.P., Stewart, G.A., Fisher, C.L. (2008) Surface geology of Australia 1:1,000,000 scale, South Australia [Digital Dataset] Canberra: The Commonwealth of Australia, Geoscience Australia. http://www.ga.gov.au

Appendix A   Subjective judgement involved in identifying flood events

Given the large number of catchments and events to be analysed, it was necessary to involve a number of operators in the event identification and definition tasks. It was considered essential that all operators completed the task consistently, without the introduction of bias into the outcomes. To understand any effects on the results that may have been introduced by the use of multiple operators, a set of common sites was analysed by each person involved in the assessment. The findings of this common analysis are discussed in this appendix.

Ten sites from different locations across Australia were analysed by all operators. One site was subsequently identified as having poor quality data and excluded from Phase 2 of this ARR project. The similarity of the events identified by each of the operators across the remaining nine sites were compared on the basis of the three baseflow parameters extracted for the broader study (described in Section 3.6 in more detail):

1) **Baseflow Peak Ratio**: Ratio of the peak baseflow (C) to the peak streamflow (A), given by C/A.

2) **Baseflow Under Peak Ratio**: Ratio of the baseflow at the time of the streamflow peak (B) to the peak streamflow (A), given by B/A.

3) **Baseflow Volume Ratio**: The event baseflow index (BFI), which is given by the total baseflow volume for the duration of the event divided by the total streamflow volume. This is the ratio of the shaded areas in the example hydrograph.

A fourth statistic, the event duration was also extracted for this comparison. This was calculated as the time between the start and the end of the event, as determined in the event analysis.

Statistical analysis of these values was undertaken to determine the significance of any variability in results between operators. At each site, the average parameter value for each operator for each of the above statistics was calculated. A two-factor Analysis of Variance (ANOVA) test was applied without replication for each of the parameters.

No significant difference in the average values of either the Baseflow Peak Ratio (Table 10) or Baseflow Under Peak Ratio (Table 11) was identified between the operators. This outcome is important as it indicates that the selection of the start and/or end of the event (through operator input) does not influence the values calculated for the baseflow contribution to the flood peak (Baseflow Peak Ratio and Baseflow Under Peak Ratio). This analysis shows that any difference in operator analysis should not significantly affect the parameter values characterising the magnitude of the peak baseflow that have been extracted for use in the study.

Table 10 Analysis of Variance for Baseflow Peak Ratio between nine sample sites and six operators

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Sites	0.766269	8	0.095784	1531.058	4.35E-39	2.244396
Between Operators	0.000227	4	5.68E-05	0.908396	0.470822	2.668437
Error	0.002002	32	6.26E-05			
Total	0.768498	44				

Table 11 Analysis of Variance for Baseflow Under Peak Ratio between nine sample sites and six operators

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Sites	0.529029	8	0.066129	2586.56	1.01E-42	2.244396
Between Operators	7.75E-05	4	1.94E-05	0.757669	0.560497	2.668437
Error	0.000818	32	2.56E-05			
Total	0.529925	44				

The ANOVA results demonstrated a significant difference between the operators in the values of Baseflow Volume Ratio (Table 12) and event duration (Table 13). This is not surprising, given the direct relationship between event length and the total volume of streamflow and baseflow generated during the event (reported in Baseflow Volume Ratio).

Table 12 Analysis of Variance for Baseflow Volume Ratio between nine sample sites and six operators

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Sites	1.008866	8	0.126108	882.9113	2.8E-35	2.244396
Between Operators	0.004992	4	0.001248	8.737327	6.94E-05	2.668437
Error	0.004571	32	0.000143			
Total	1.018428	44				

Table 13 Analysis of Variance for event duration between nine sample sites and six operators

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Sites	128.9418	8	16.11772	39.90103	1.67E-14	2.244396
Between Operators	55.56982	4	13.89245	34.39216	3.62E-11	2.668437
Error	12.92616	32	0.403942			
Total	197.4377	44				

A reduced analysis was done on three sites that had been initially completed by each operator and were then reviewed by a single person. Importantly, the ANOVA results for Baseflow Volume Ratio (Table 14) for this set of sites was not significant. This suggests that review by a single operator is sufficient to eliminate the effect of having used multiple operators to identify events.

Table 14 Analysis of Variance for Baseflow Volume Ratio between three sample sites and six operators after review by a single operator

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Sites	0.07361	2	0.036805	105.9114	1.75E-06	4.45897
Between Operators	0.003534	4	0.000884	2.542422	0.121719	3.837853
Error	0.00278	8	0.000348			
Total	0.079924	14				

Appendix B   Flood frequency distributions for each catchment

The following charts present the flood frequency distributions for each catchment considered as a part of this study. These plots present the data as obtained using the analysis approach as outlined in Section 3.5, which combines three different functions across the range of event sizes.

Appendix C   Variation in Baseflow Peak Factor with ARI of total flow peak for each catchment

The following charts present the variation in Baseflow Peak Factor with ARI for each catchment analysed as a part of this study. These plots present the data as obtained using the extraction method as outlined in Section 3. The fitted trend lines generally relate to those presented in Section 7.2, with the exception of the catchments that display flat or positive gradient behaviour with increasing event magnitude. As discussed in previous sections, it was assumed that the Baseflow Peak Factor for these catchments was actually invariant with ARI. All analysis undertaken as a part of this study has utilised the data described in the main text. The raw outputs are presented here for contextual information.

Appendix D   Variation in Baseflow Volume Factor with ARI of total flow peak for each catchment

The following charts present the variation in Baseflow Volume Factor with ARI for each catchment considered as a part of this study. These plots present the data as obtained using the extraction method as outlined in Section 3. The fitted trend lines generally relate to those presented in Section 7.3, with the exception of the catchments that display flat or positive gradient behaviour with increasing event magnitude. As discussed in previous sections, it was assumed that the Baseflow Volume Factor for these catchments was actually invariant with ARI. All analysis undertaken as a part of this study has utilised the data described in the main text. The raw outputs are presented here for contextual information.

Appendix E   Regression Statistics

The following figures show the diagnostics for each of the regressions developed in section 5.
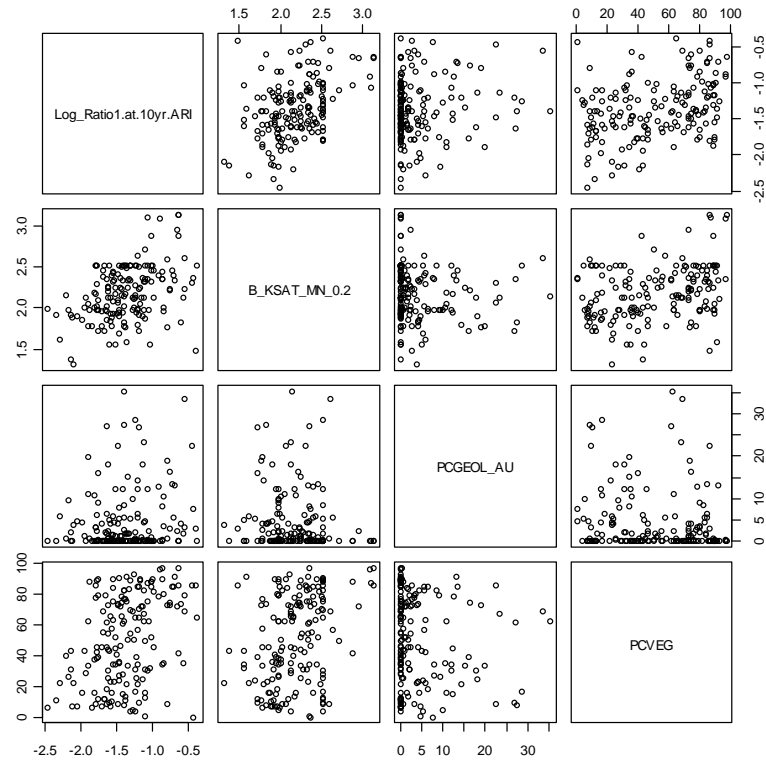
Figure 62 Scatter plot matrix of continuous predictor variables for cluster 1 Baseflow Peak Ratio regression
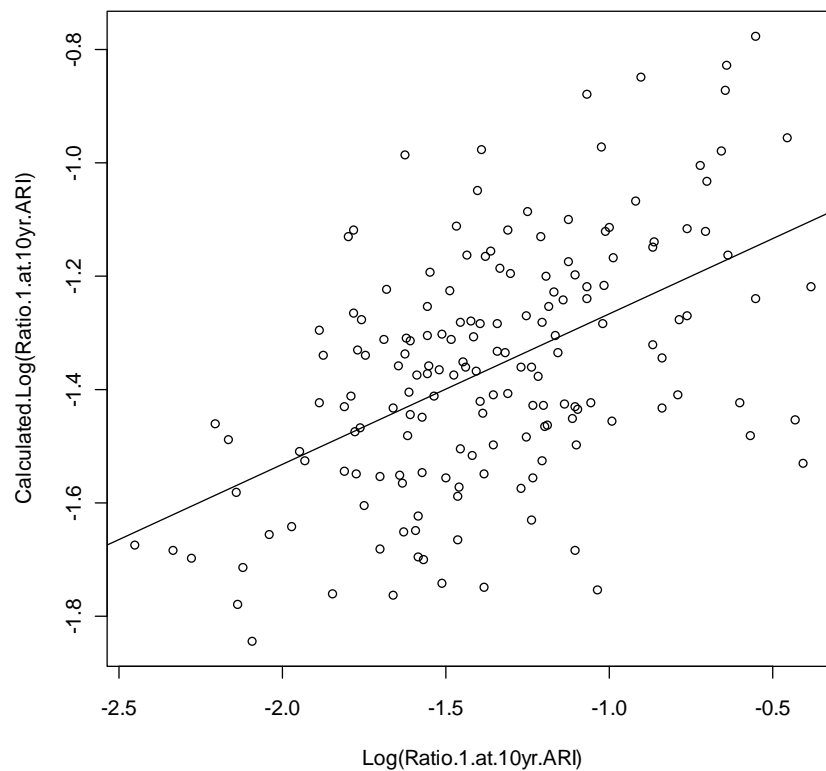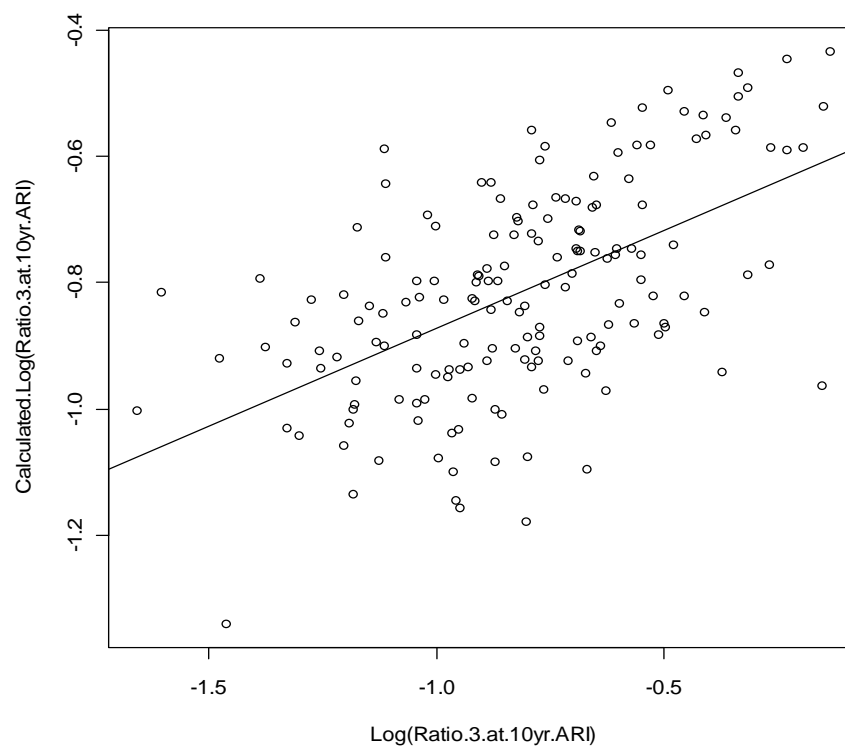

Figure 63 Fitted values against observed values for cluster 1 Baseflow Peak Ratio regression
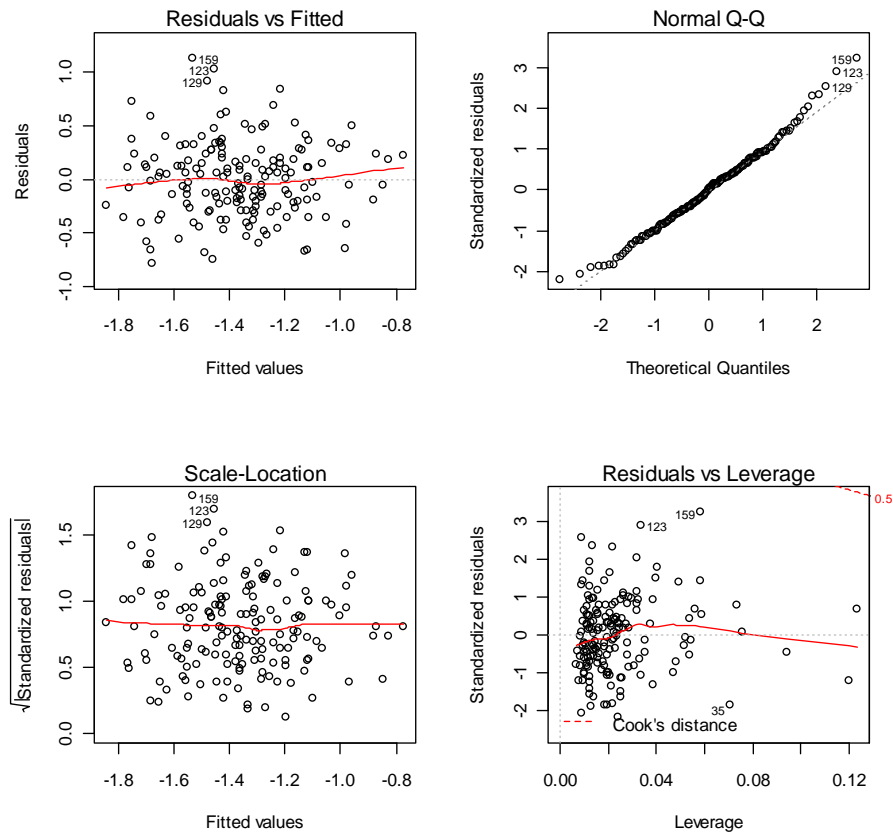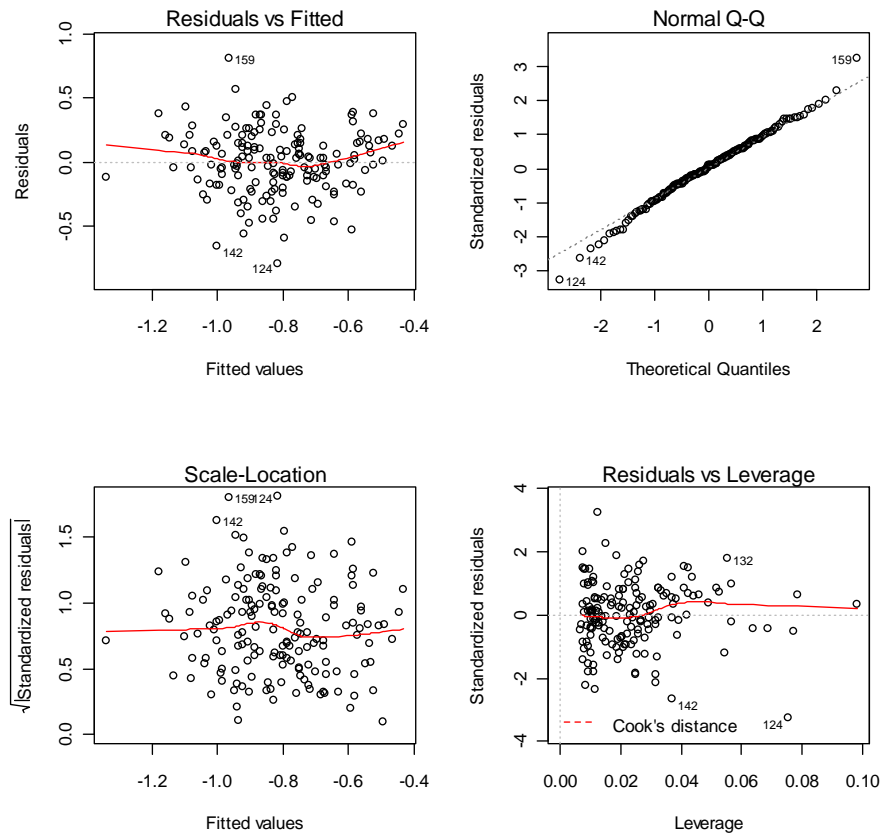
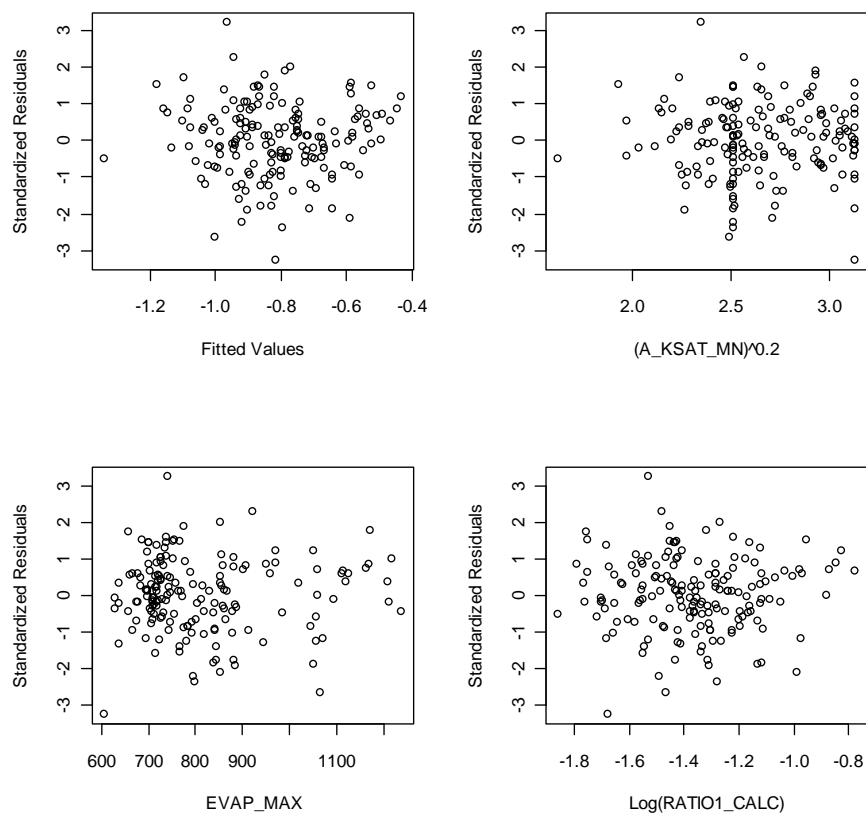**Figure 64 Diagnostic plots for cluster 1 Baseflow Peak Ratio regression**
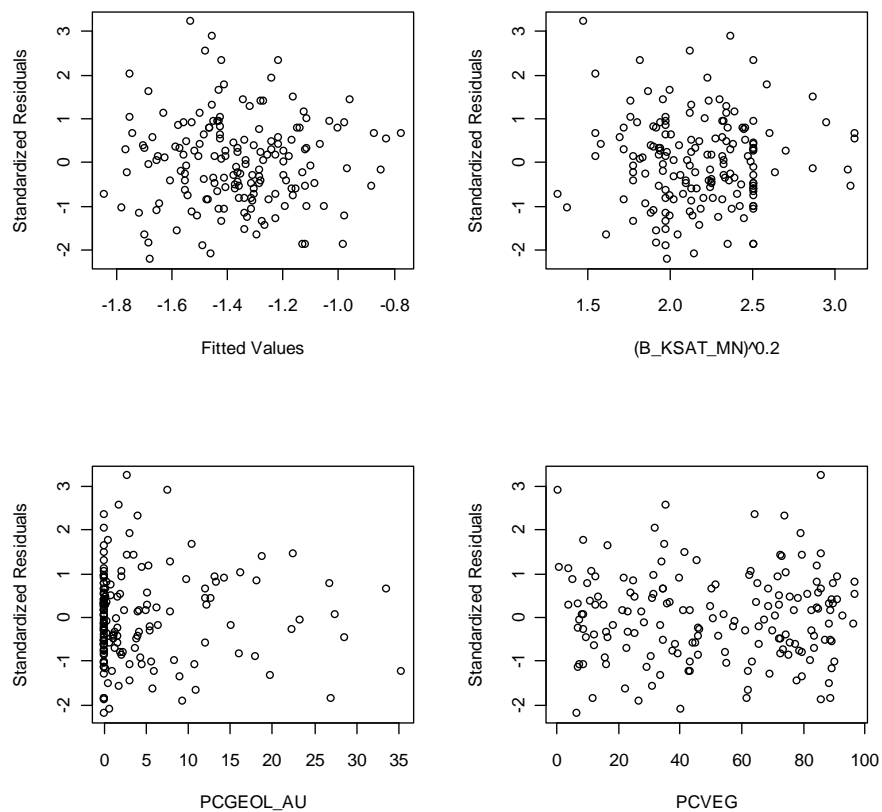
Figure 65 Standardised residuals for cluster 1 Baseflow Peak Ratio regression
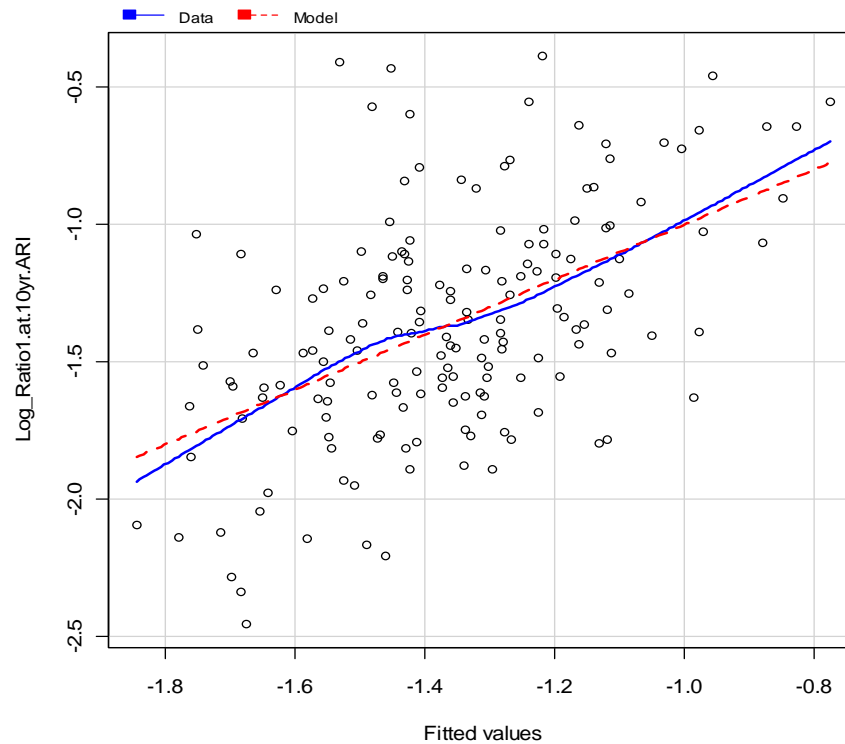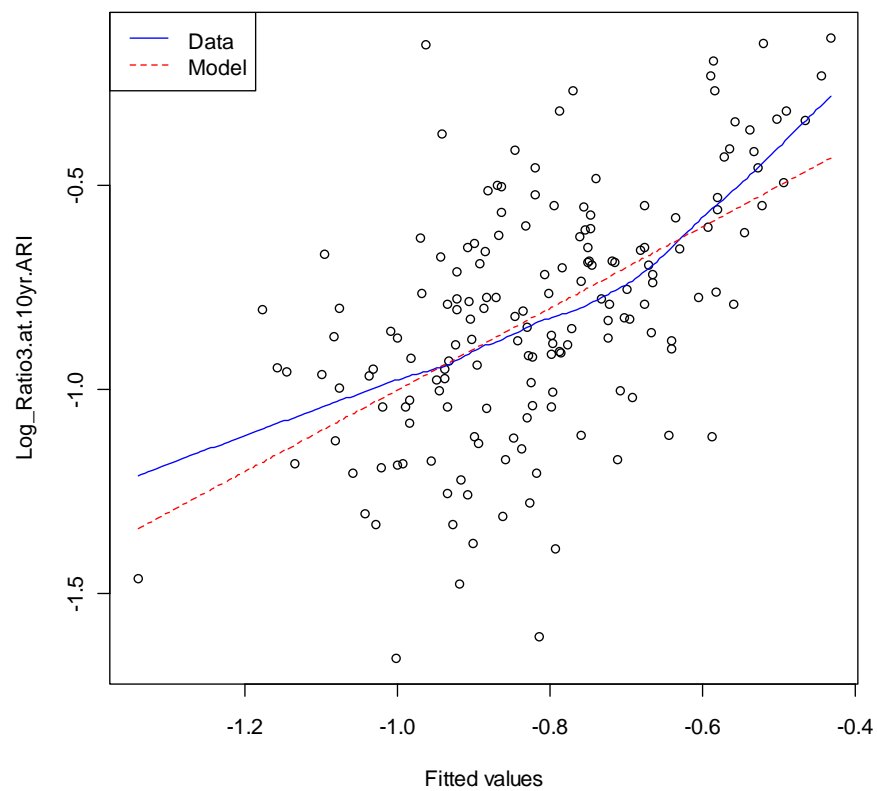
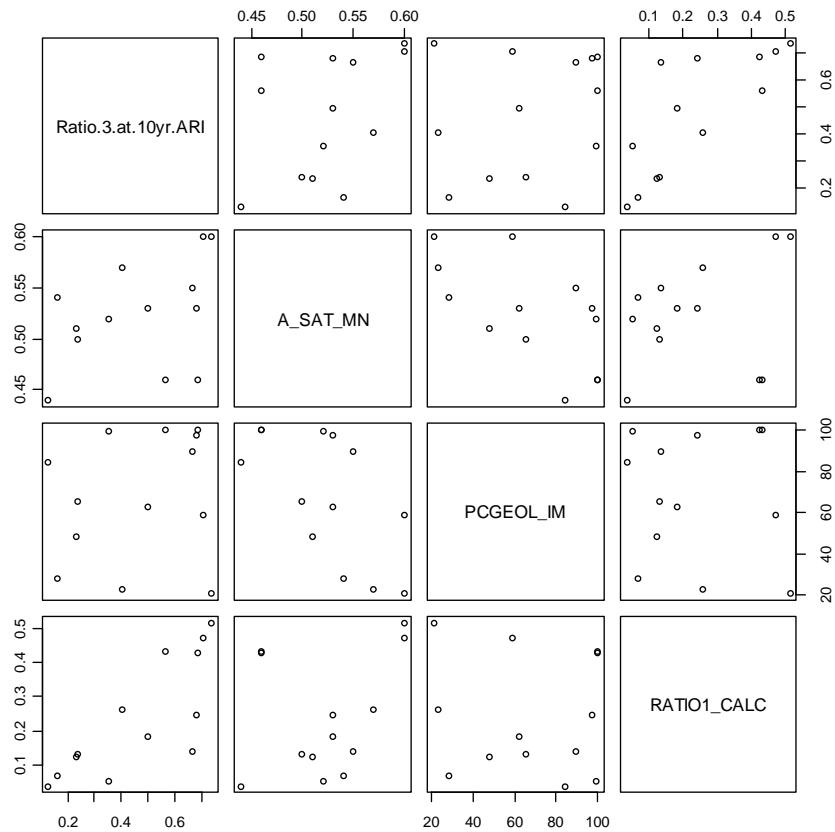**Figure 66 Marginal model plot for cluster 1 Baseflow Peak Ratio regression**
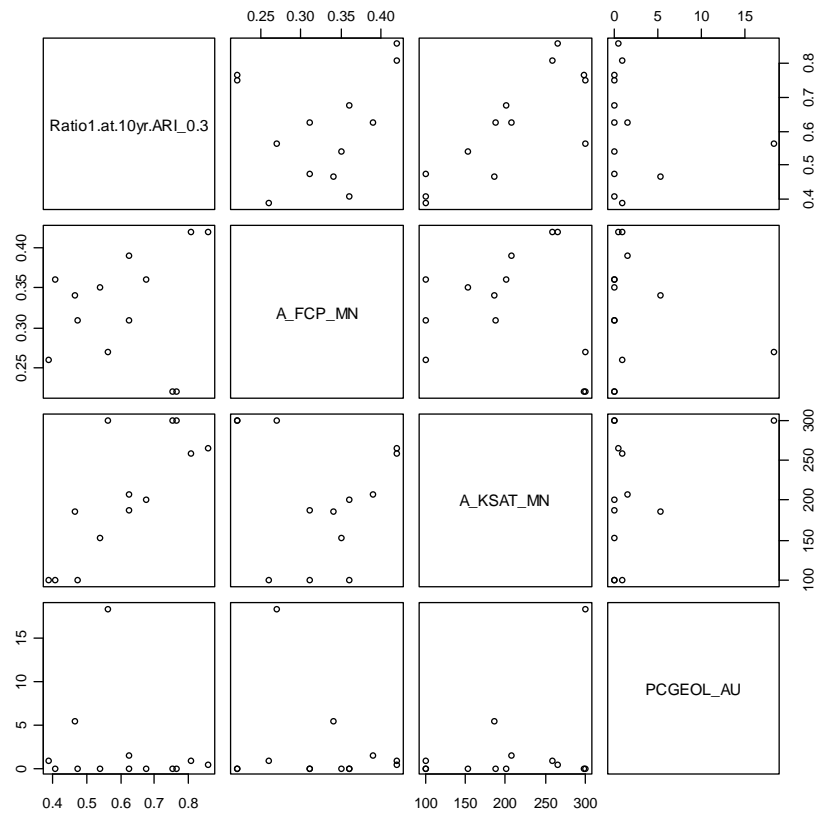
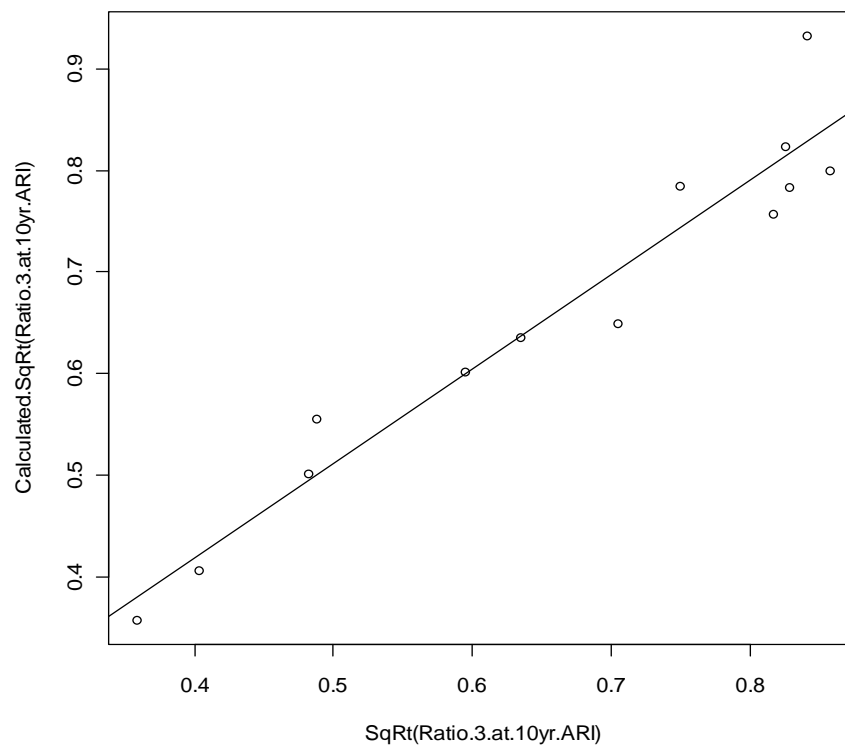**Figure 67 Scatter plot matrix of continuous predictor variables for cluster 2 Ratio1 regression**
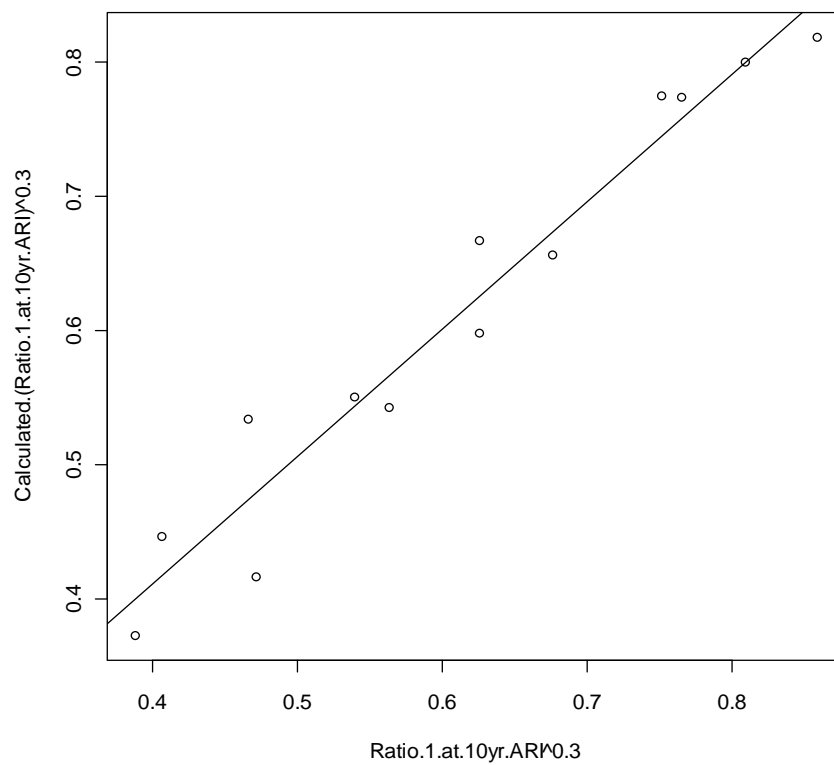
Figure 68 Fitted values against observed values for cluster 2 Baseflow Peak Ratio regression
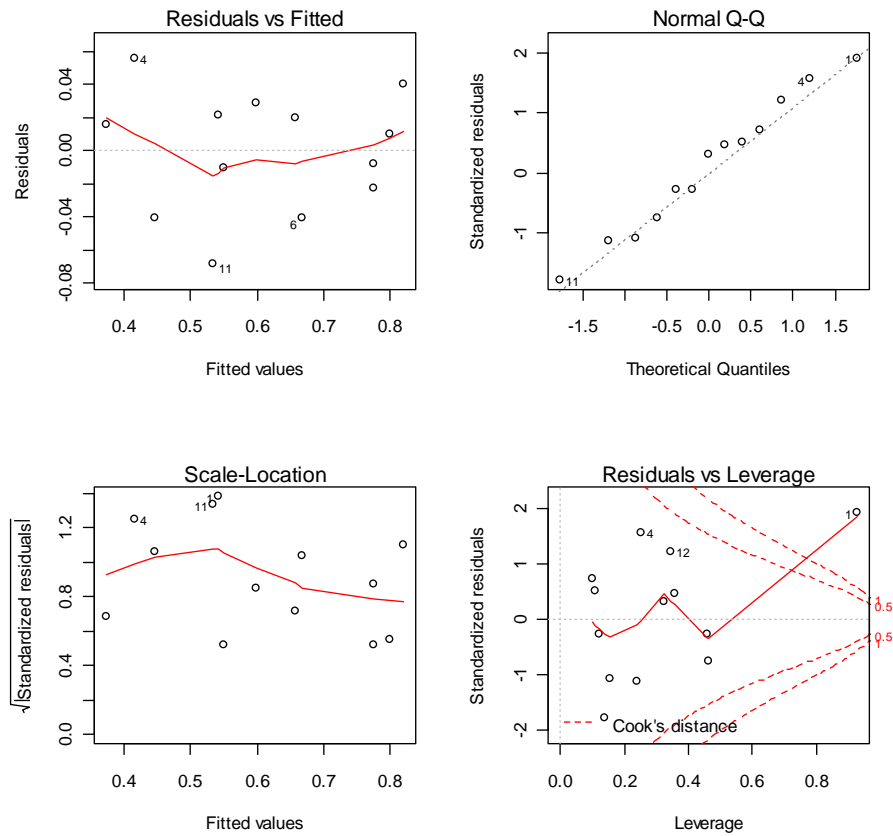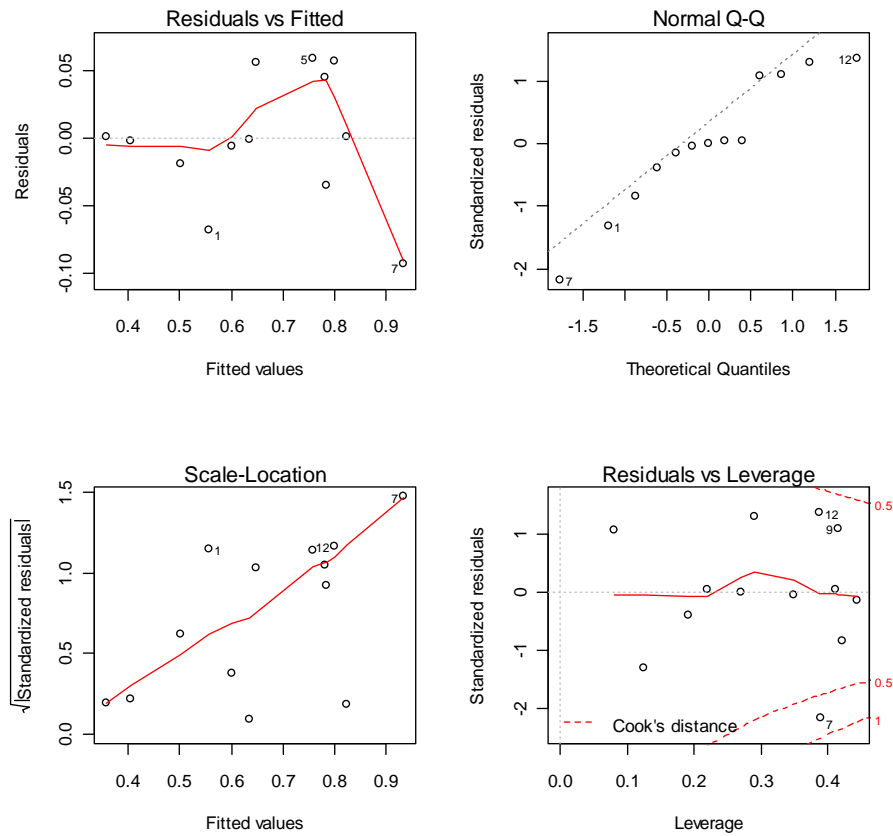
Figure 69 Diagnostic plots for cluster 2 Baseflow Peak Ratio regression
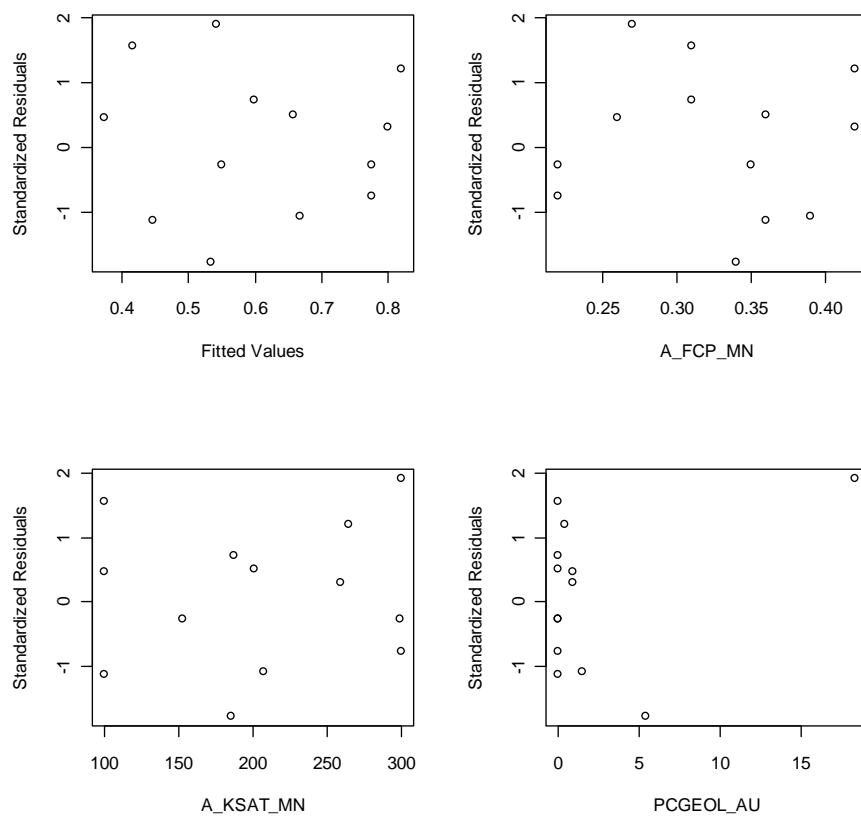
Figure 70 Standardised residuals for cluster 2 Baseflow Peak Ratio regression
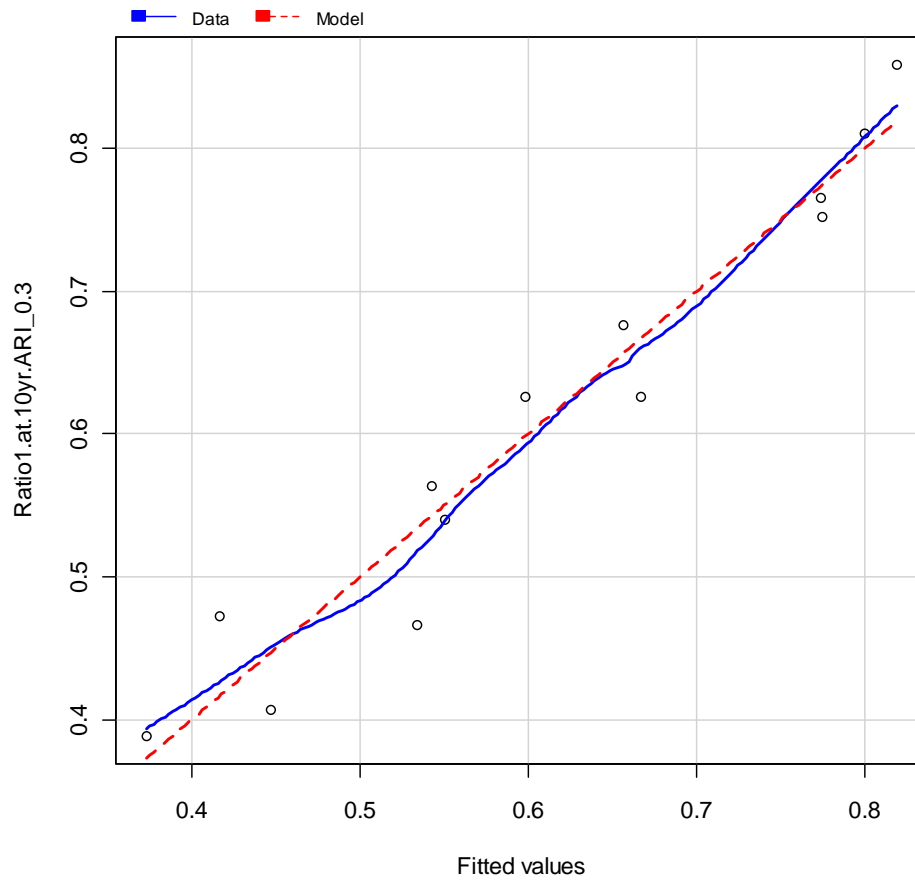
Figure 71 Marginal model plot for cluster 2 Baseflow Peak Ratio regression
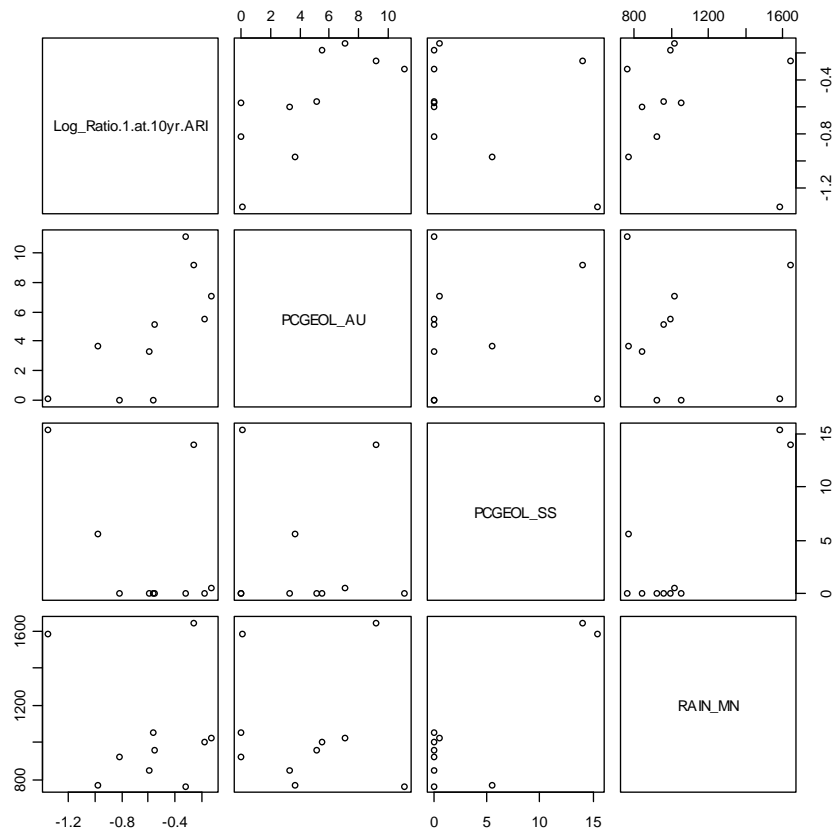
Figure 72 Scatter plot matrix of continuous predictor variables for cluster 3 Ratio1 regression
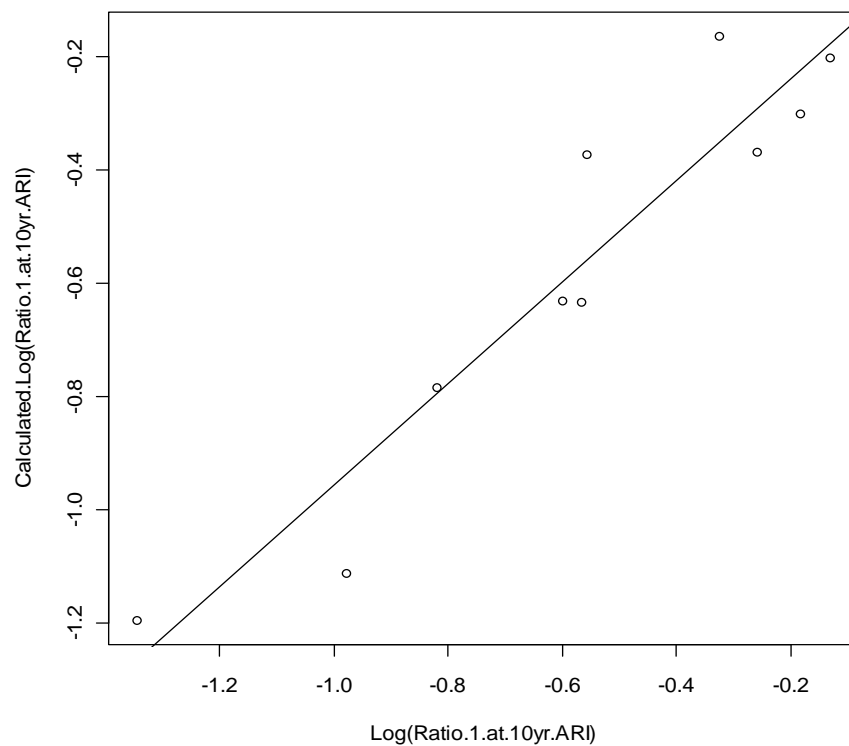
Figure 73 Fitted values against observed values for cluster 3 Baseflow Peak Ratio regression
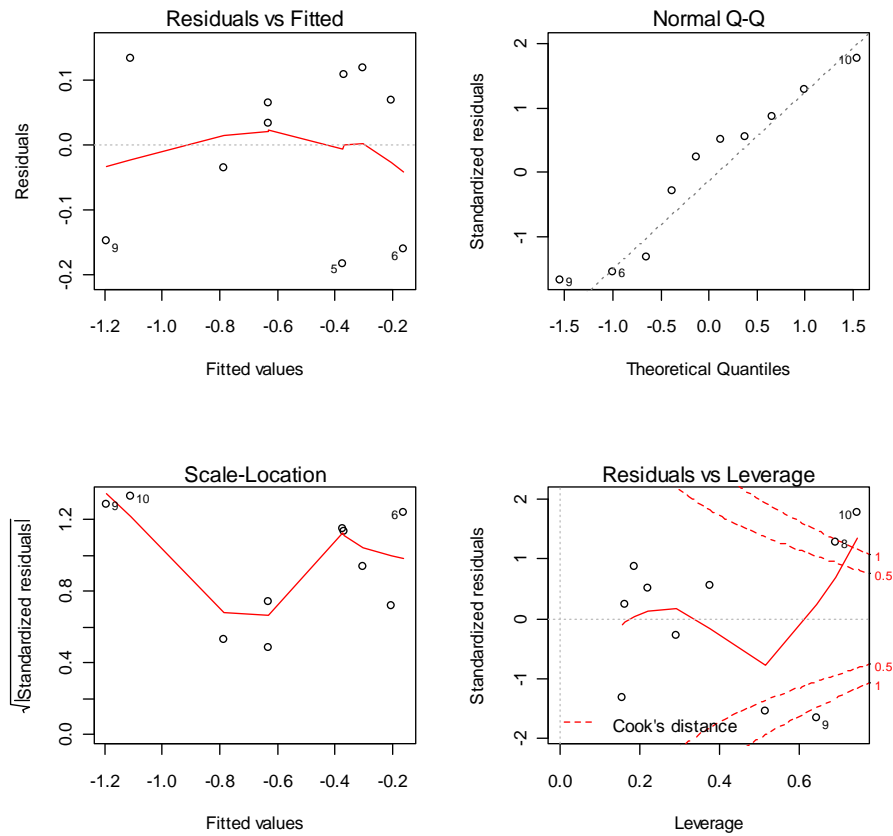
Figure 74 Diagnostic plots for cluster 3 Baseflow Peak Ratio regression
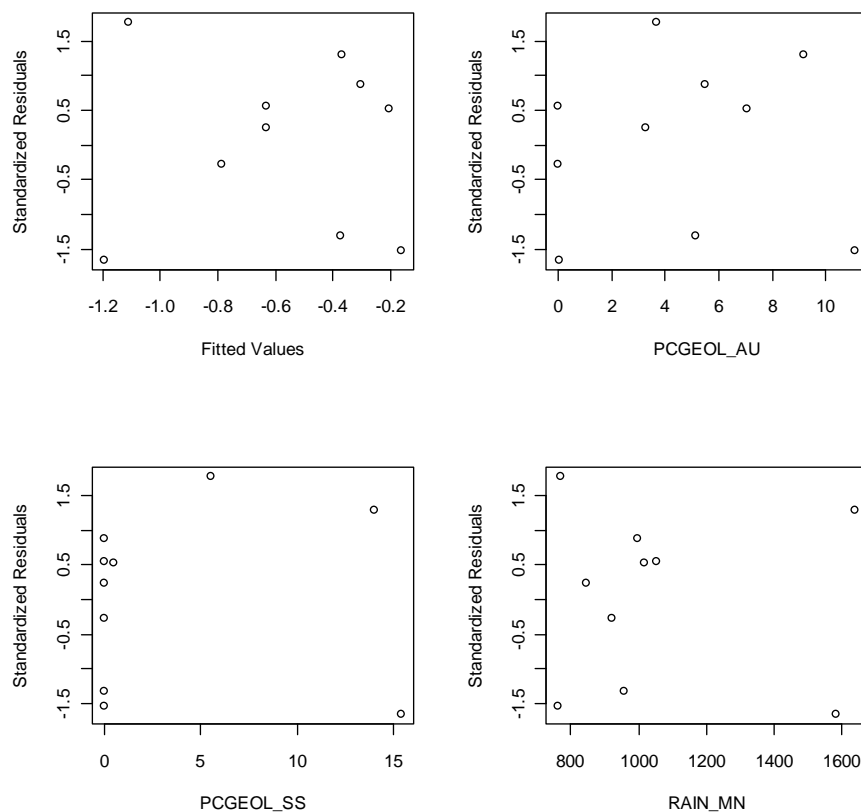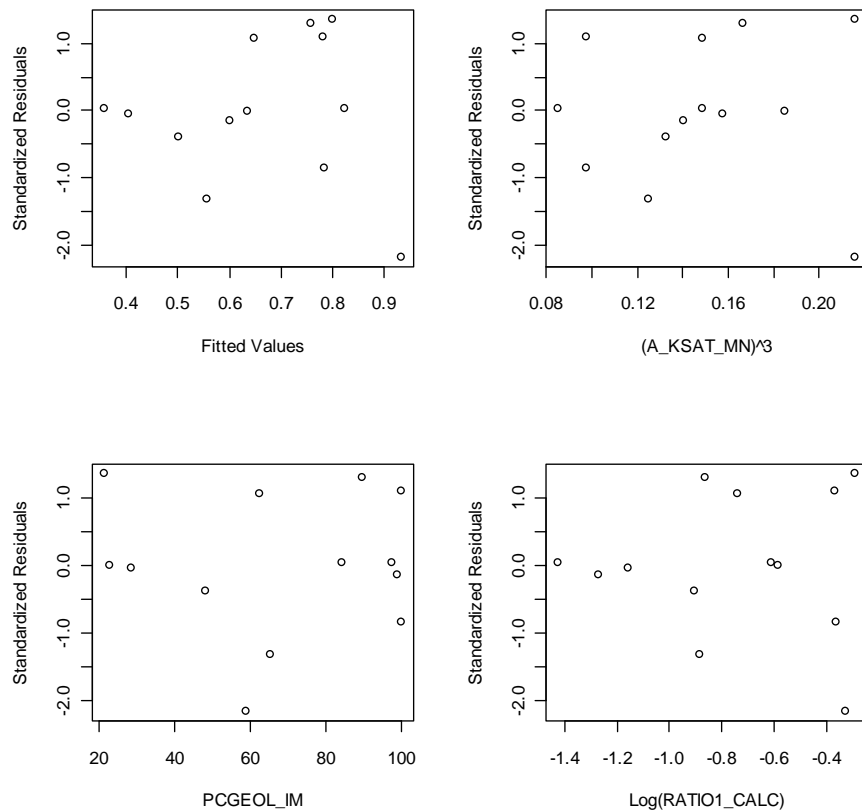

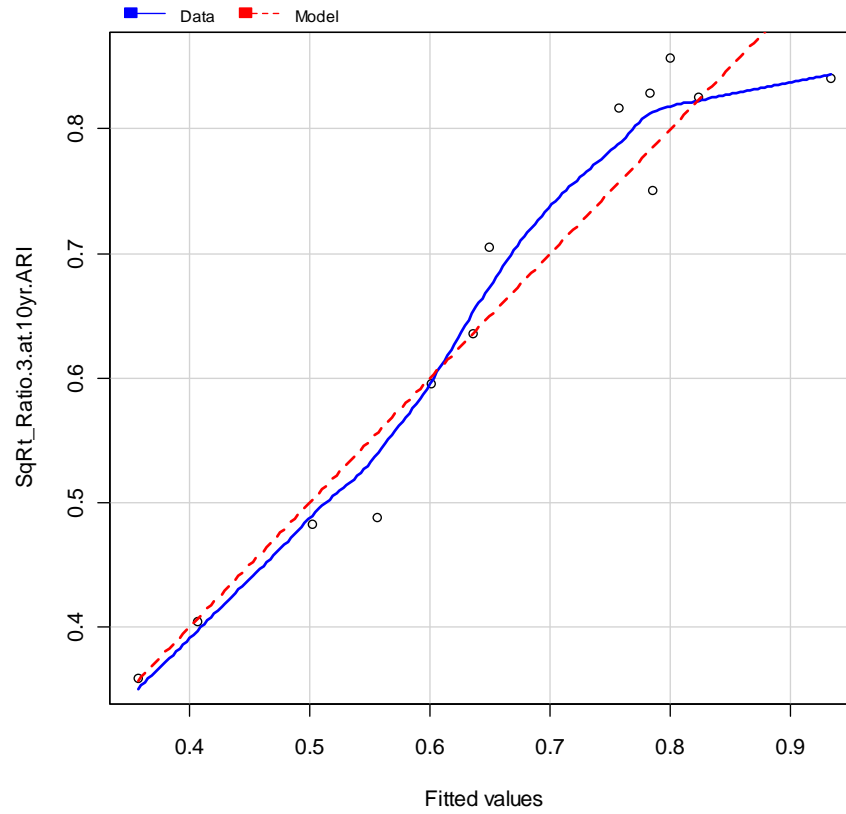**Figure 75 Standardised residuals for cluster 3 Baseflow Peak Ratio regression**
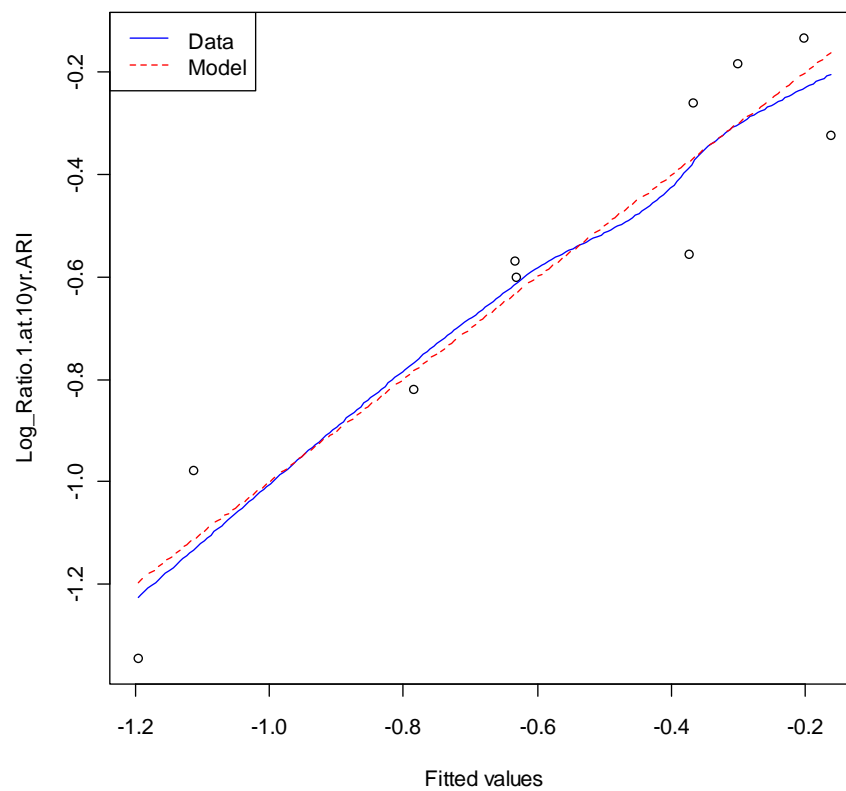
Figure 76 Marginal model plot for cluster 3 Baseflow Peak Ratio regression
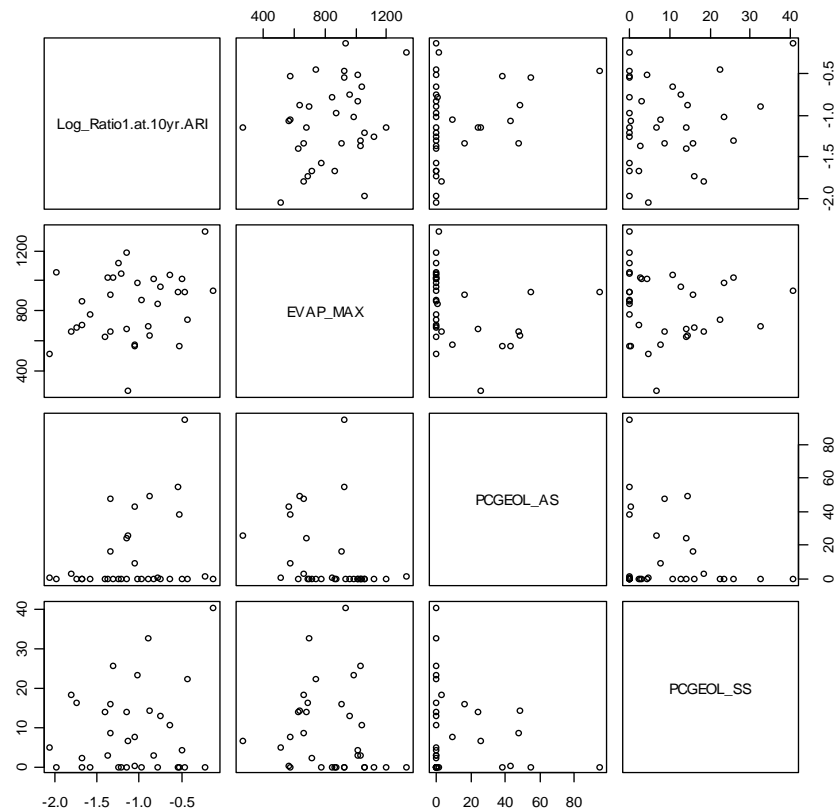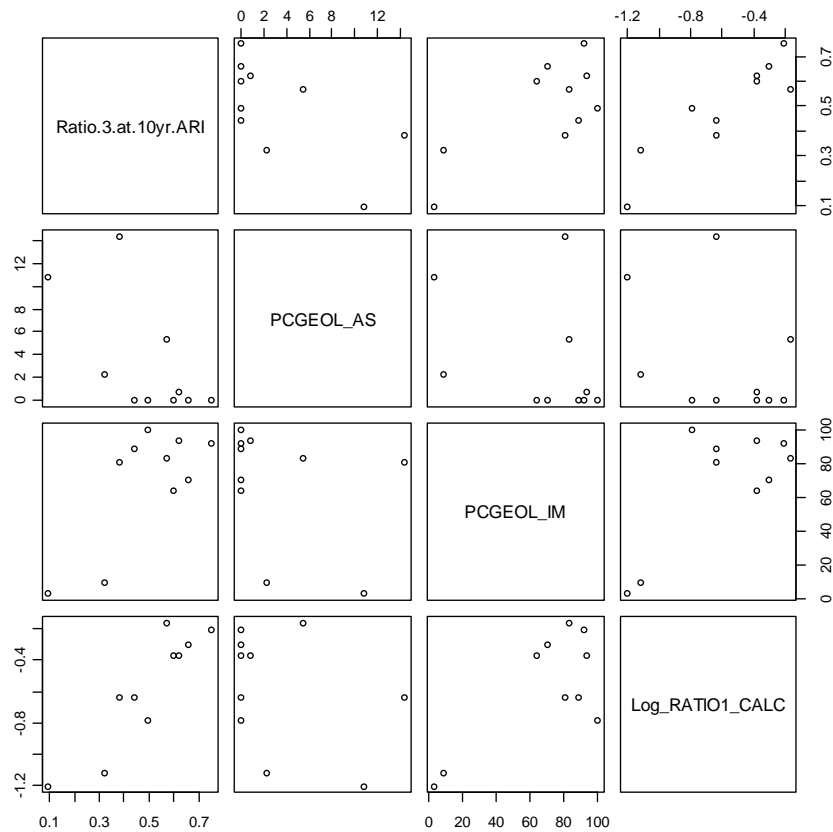
Figure 77 Scatter plot of continuous predictor variables for cluster 4 Baseflow Peak Ratio regression
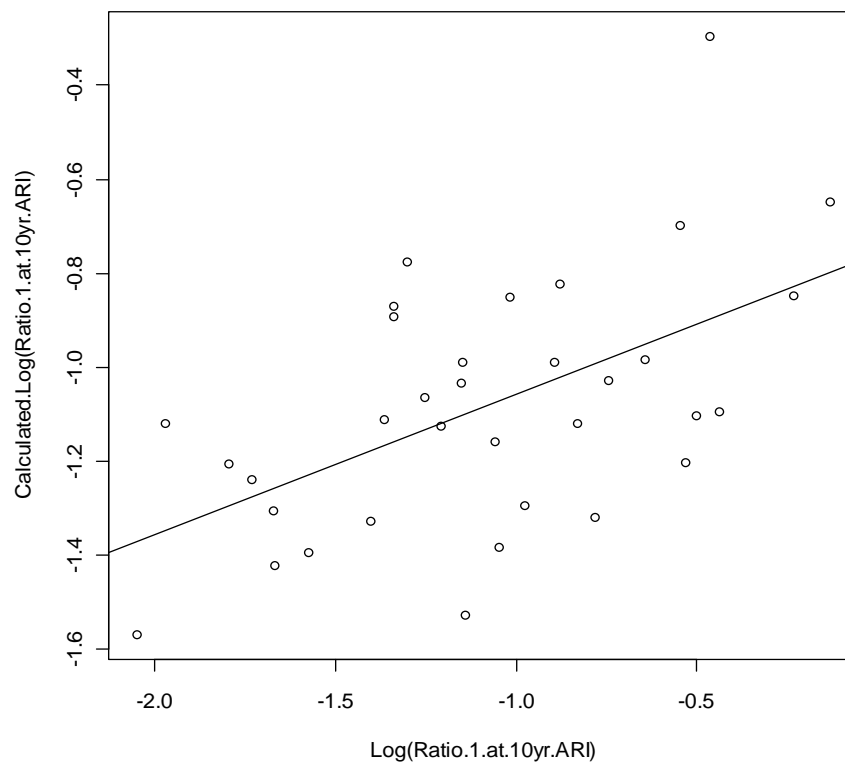
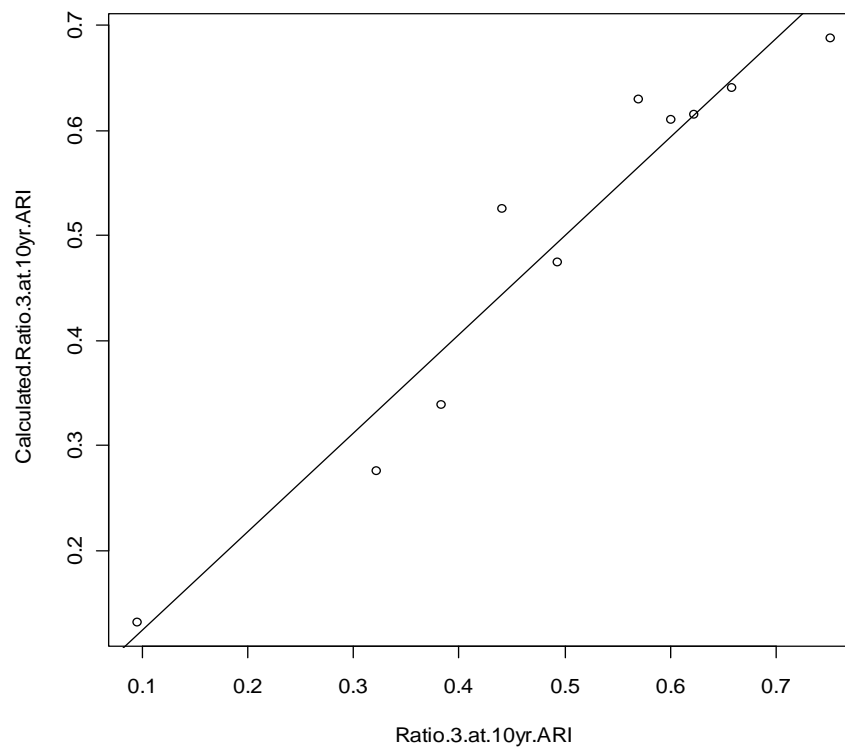**Figure 78 Fitted values against observed values for cluster 4 Baseflow Peak Ratio regression**
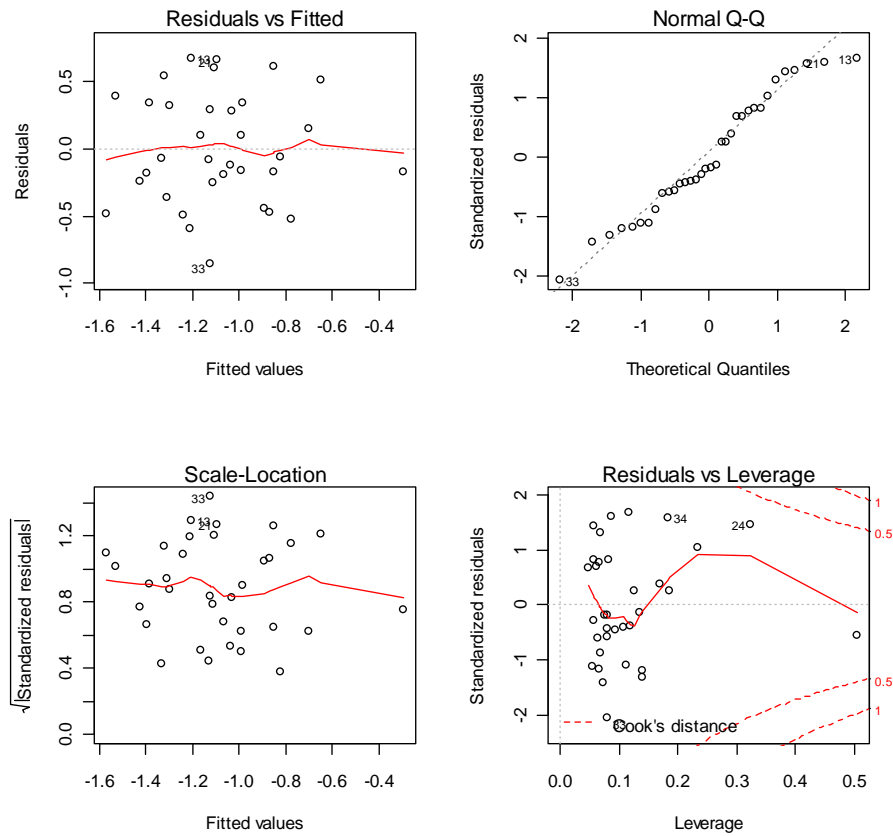
Figure 79 Diagnostic plots for cluster 4 Baseflow Peak Ratio regression
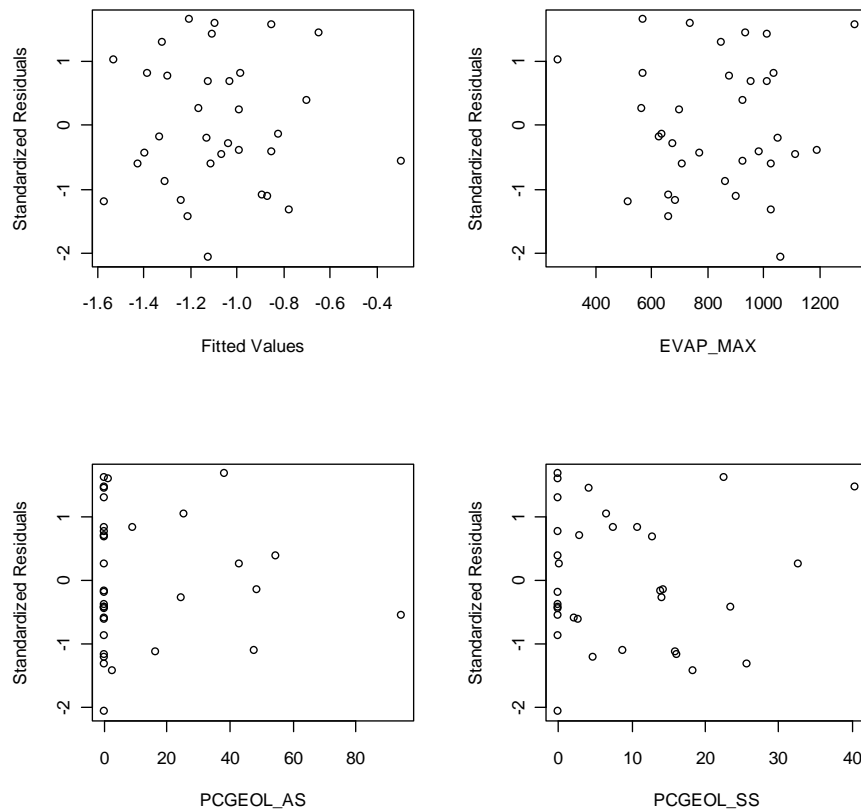
Figure 80 Standardised residuals for cluster 4 Baseflow Peak Ratio regression
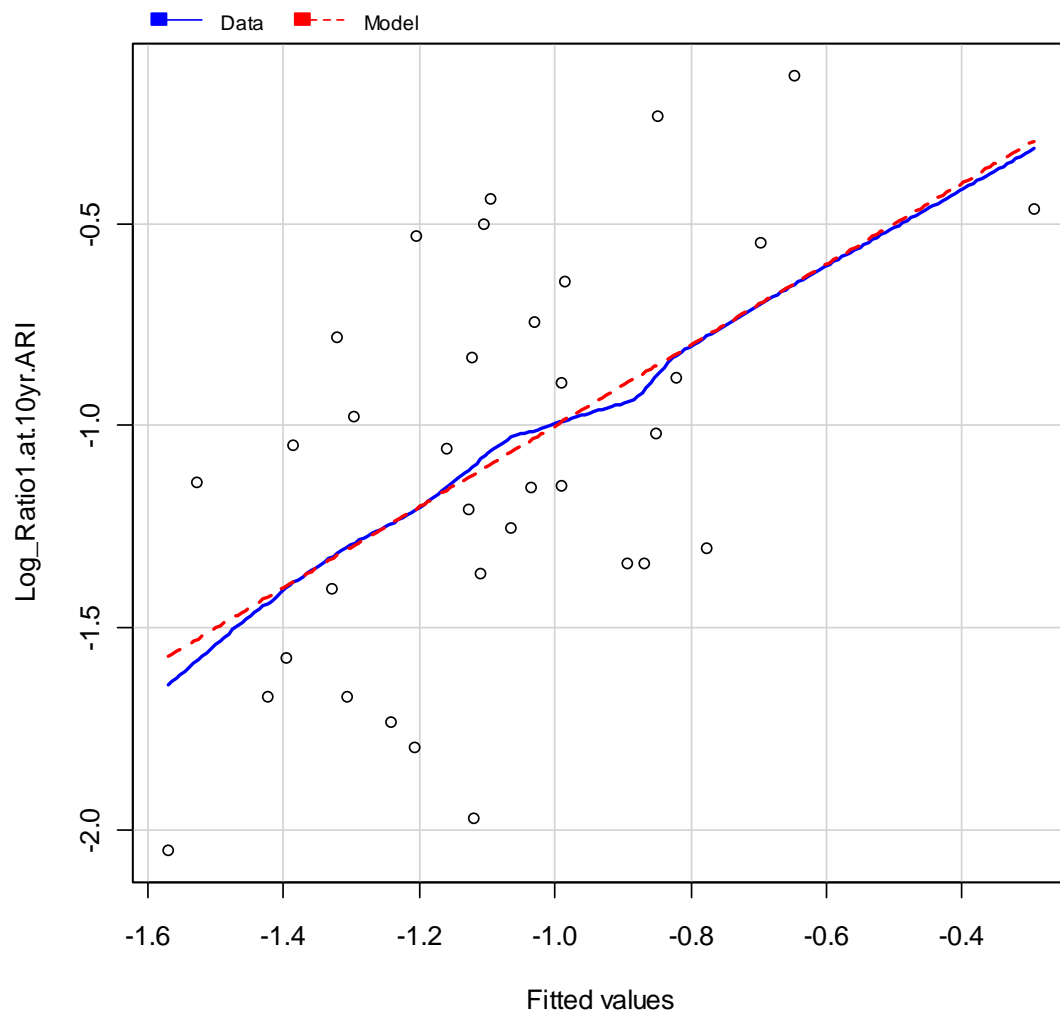
Figure 81 Marginal model plot for cluster 4 Baseflow Peak Ratio regression
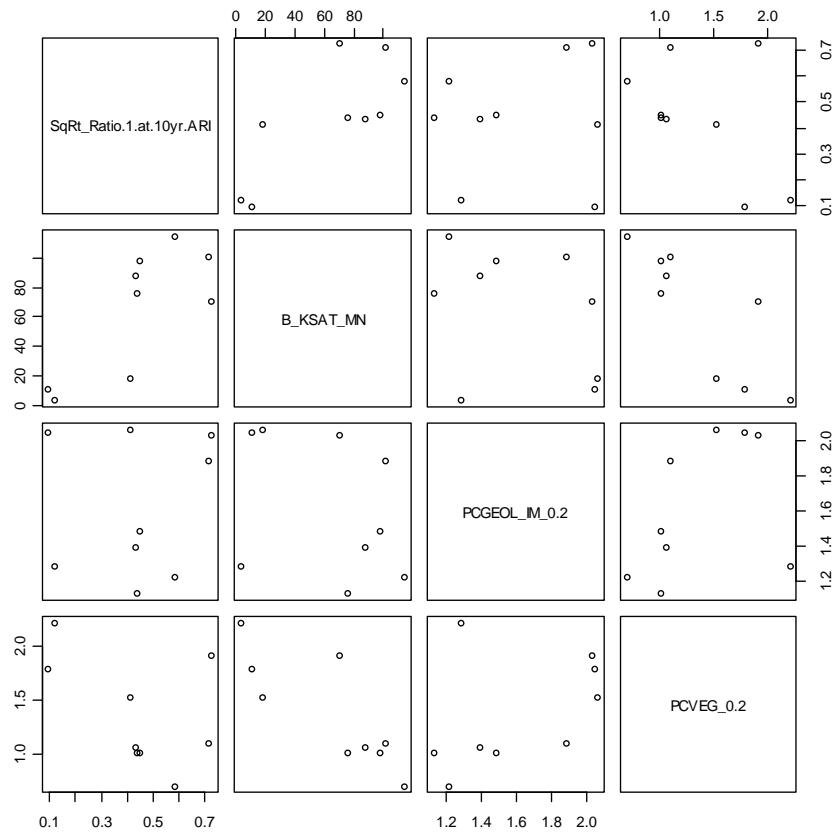
Figure 82 Scatter plot of continuous predictor variables for cluster 5 Baseflow Peak Ratio regression
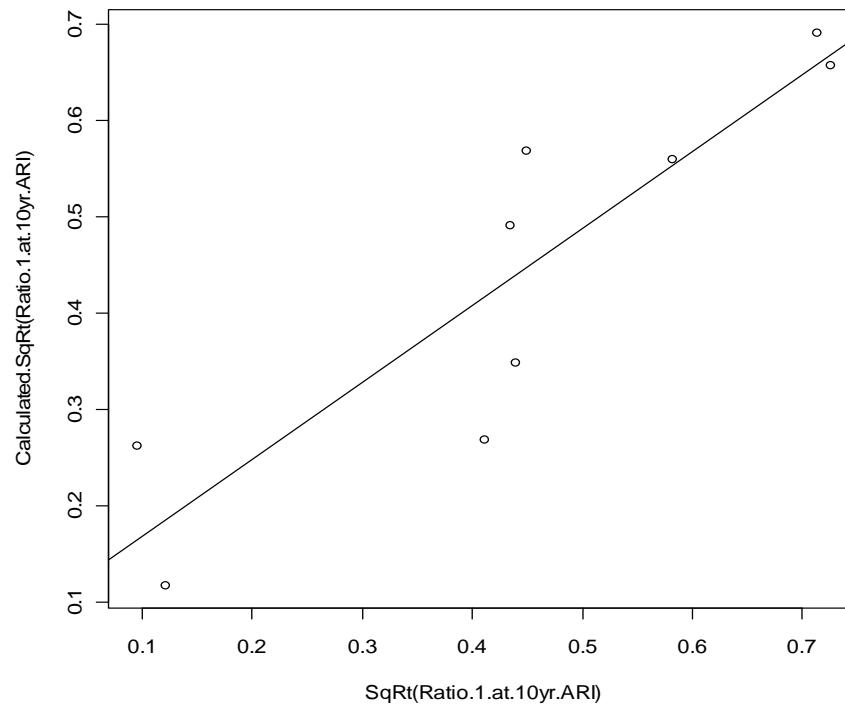
Figure 83 Fitted values against observed values for cluster 5 Baseflow Peak Ratio regression
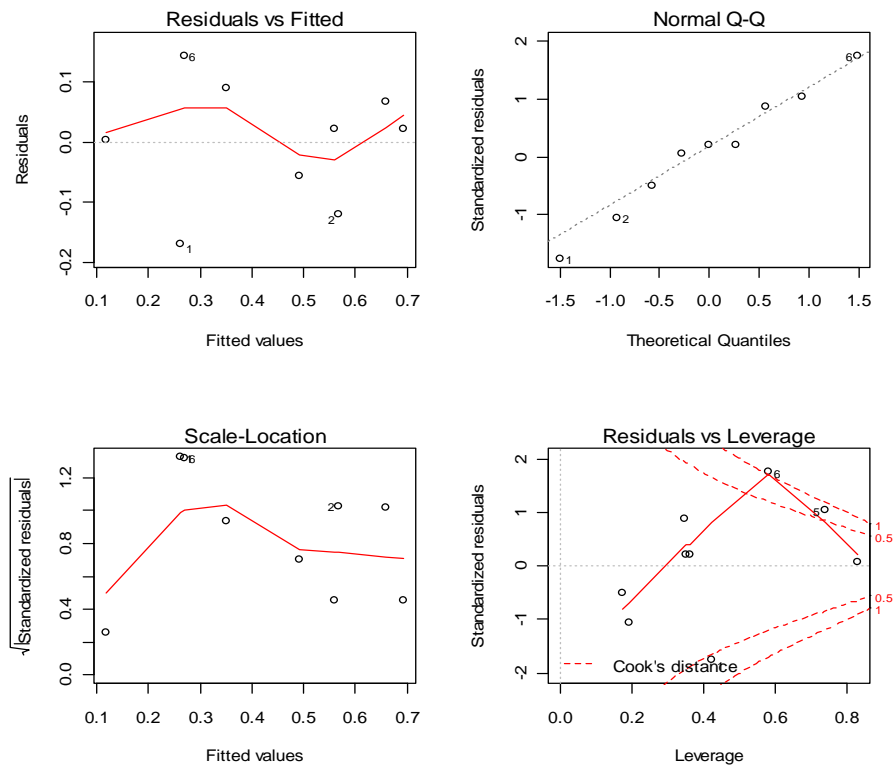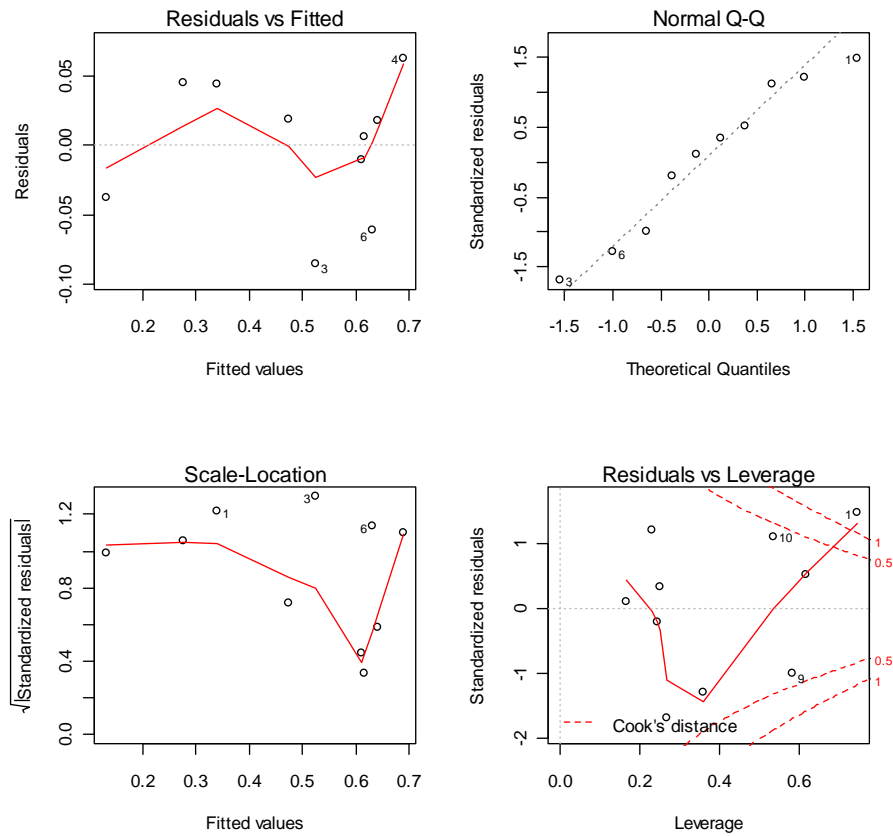
Figure 84 Diagnostic plots for cluster 5 Baseflow Peak Ratio regression
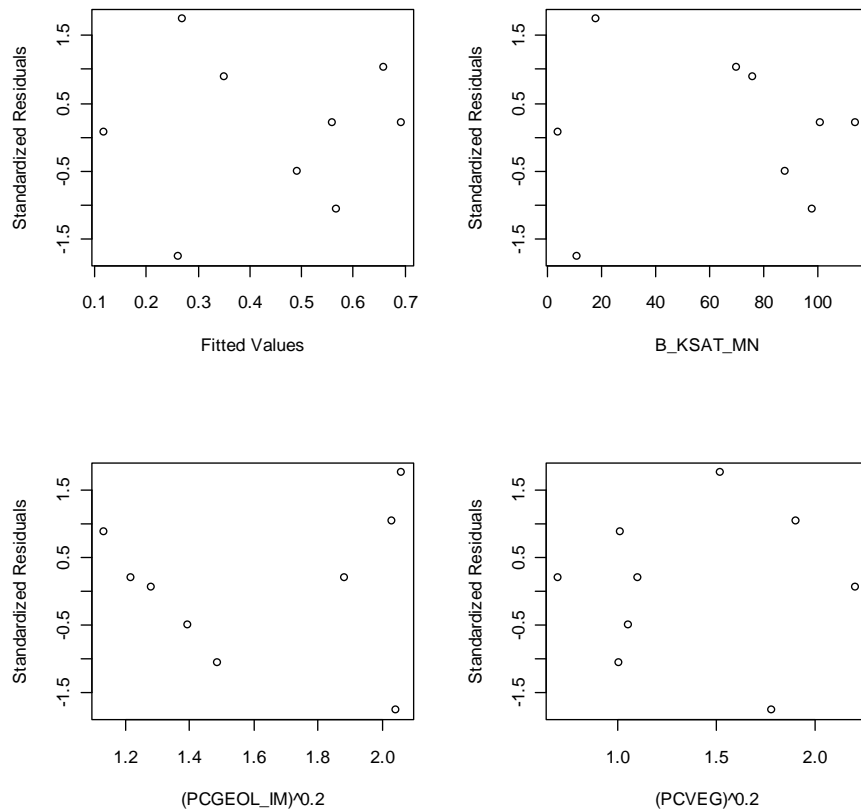
Figure 85 Standardised residuals for cluster 5 Baseflow Peak Ratio regression
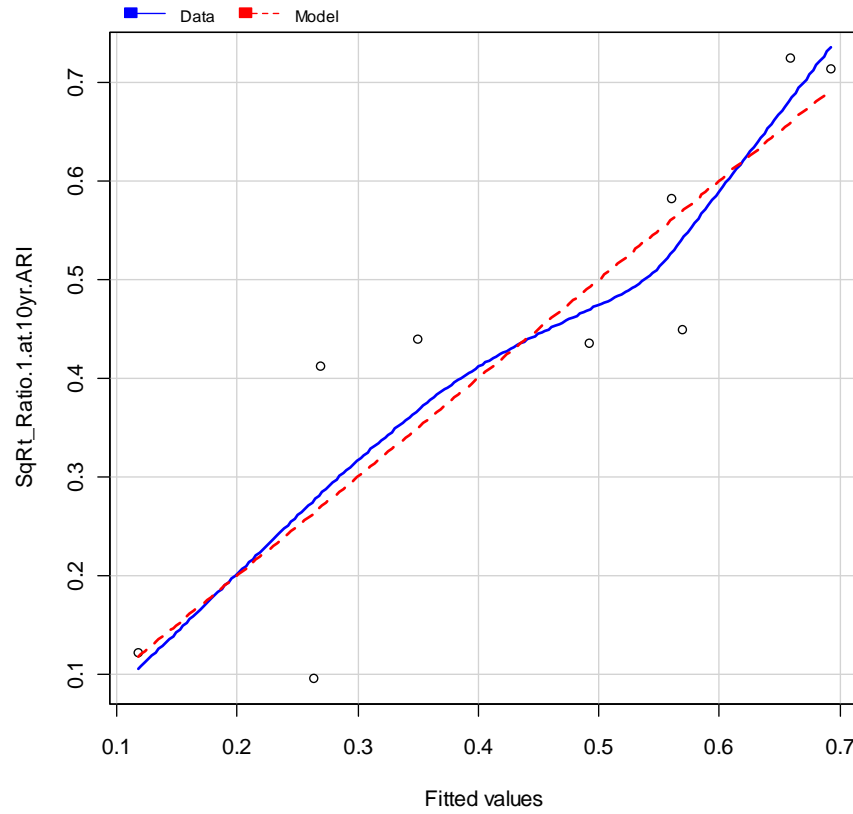
Figure 86 Marginal model plot for cluster 5 Baseflow Peak Ratio regression
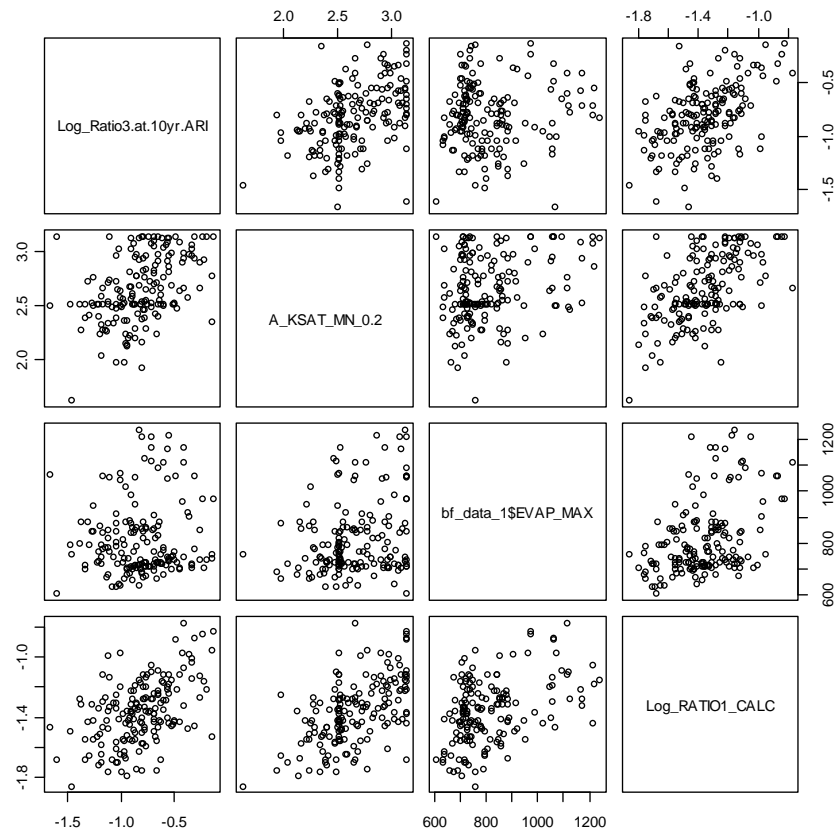
Figure 87 Scatter plot of continuous predictor variables for cluster 1 Baseflow Volume Ratio regression

Figure 88 Fitted values against observed values for cluster 1 Baseflow Volume Ratio regression

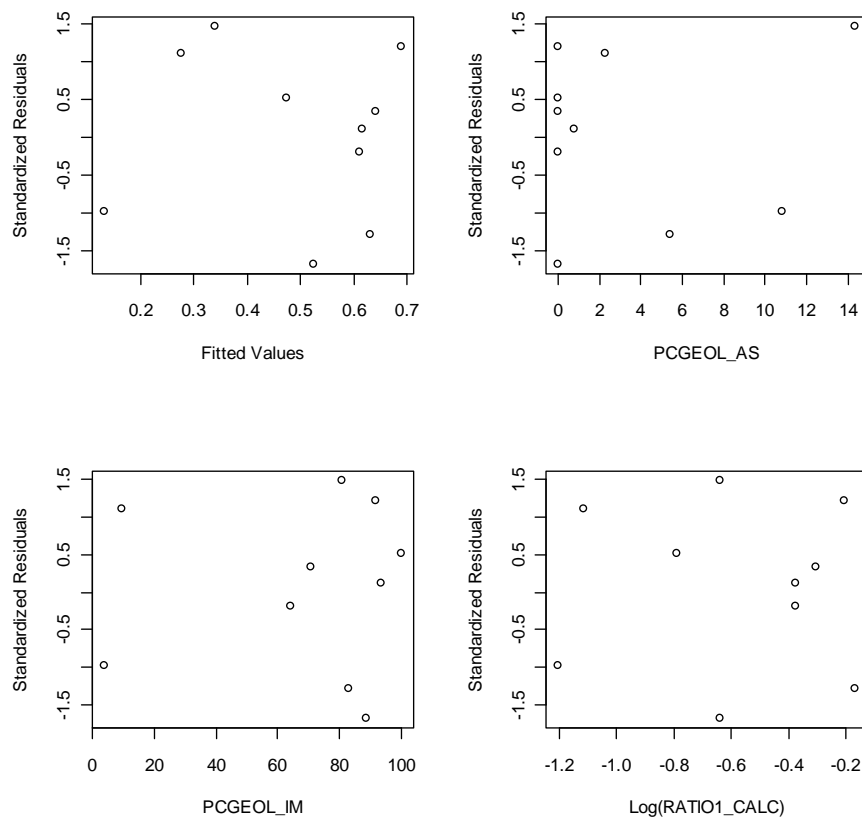**Figure 89 Diagnostic plots for cluster 1 Baseflow Volume Ratio regression**

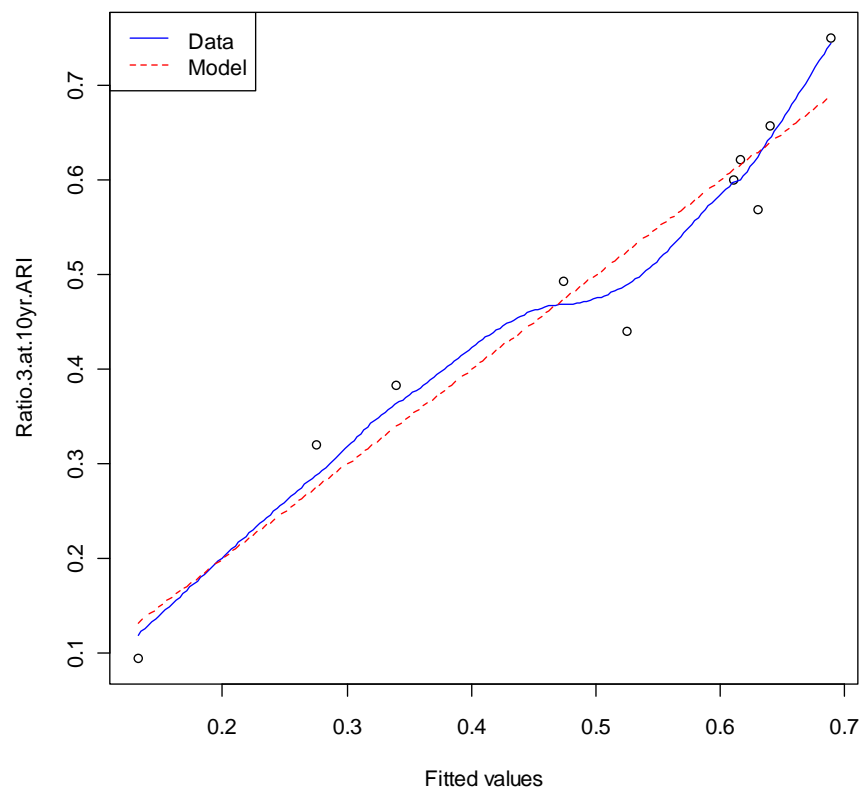**Figure 90 Standardised residuals for cluster 1 Baseflow Volume Ratio regression**

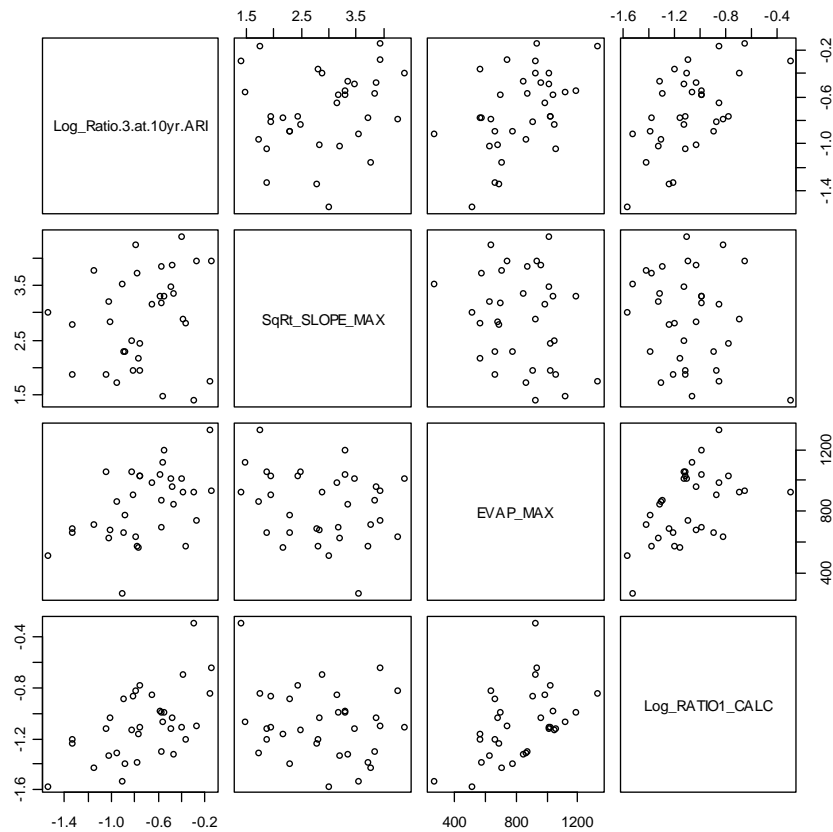**Figure 91 Marginal model plot for cluster 1 Baseflow Volume Ratio regression**

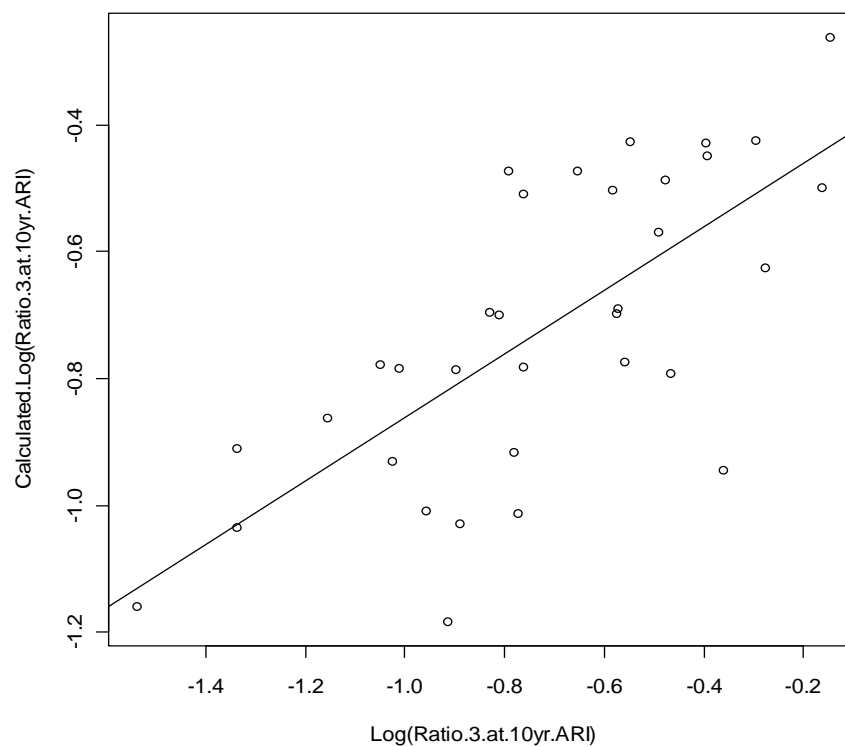**Figure 92 Scatter plot of continuous predictor variables for cluster 2 Baseflow Volume Ratio regression**

Figure 93 Fitted values against observed values for cluster 2 Baseflow Volume Ratio regression
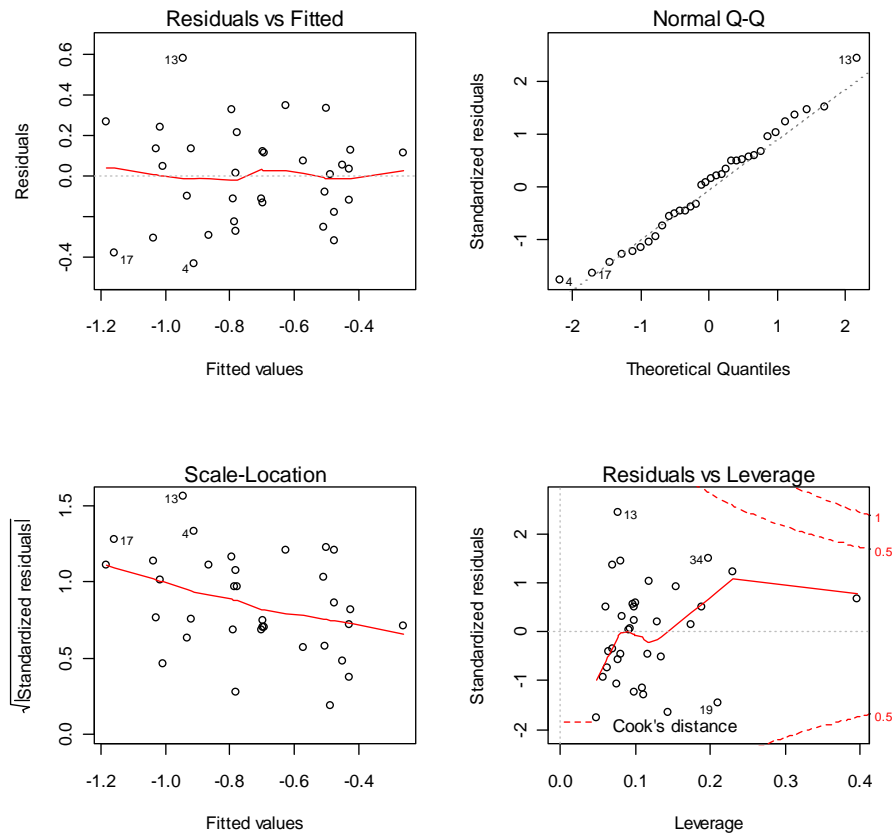
Figure 94 Diagnostic plots for cluster 2 Baseflow Volume Ratio regression

Figure 95 Standardised residuals for cluster 2 Baseflow Volume Ratio regression

Figure 96 Marginal model plot for cluster 2 Baseflow Volume Ratio regression

Figure 97 Scatter plot of continuous predictor variables for cluster 3 Baseflow Volume Ratio regression

Figure 98 Fitted values against observed values for cluster 3 Baseflow Volume Ratio regression

Figure 99 Diagnostic plots for cluster 3 Baseflow Volume Ratio regression
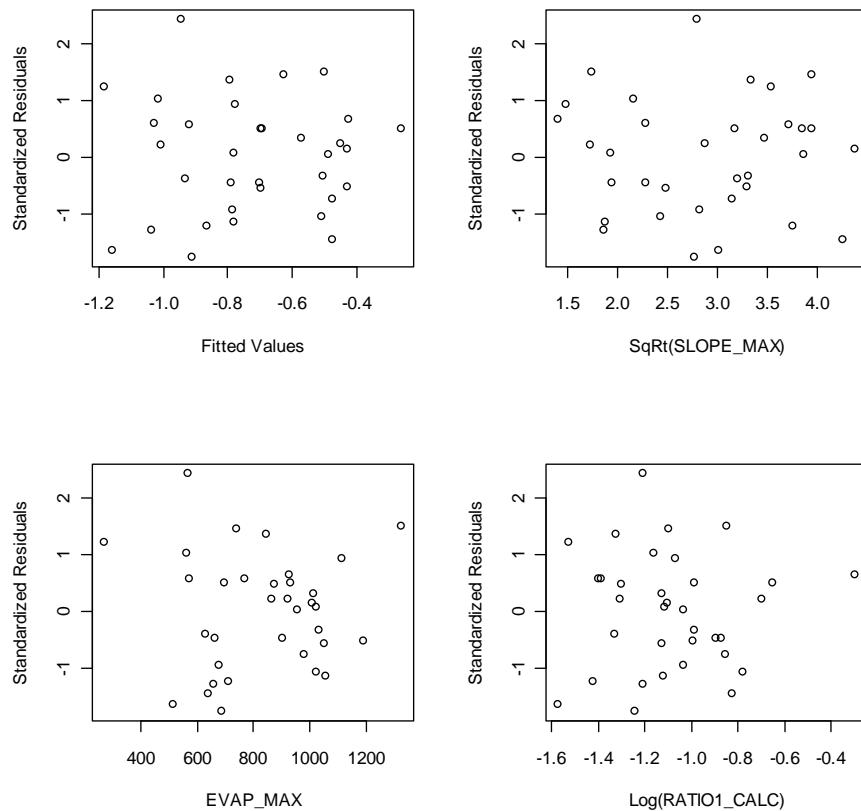

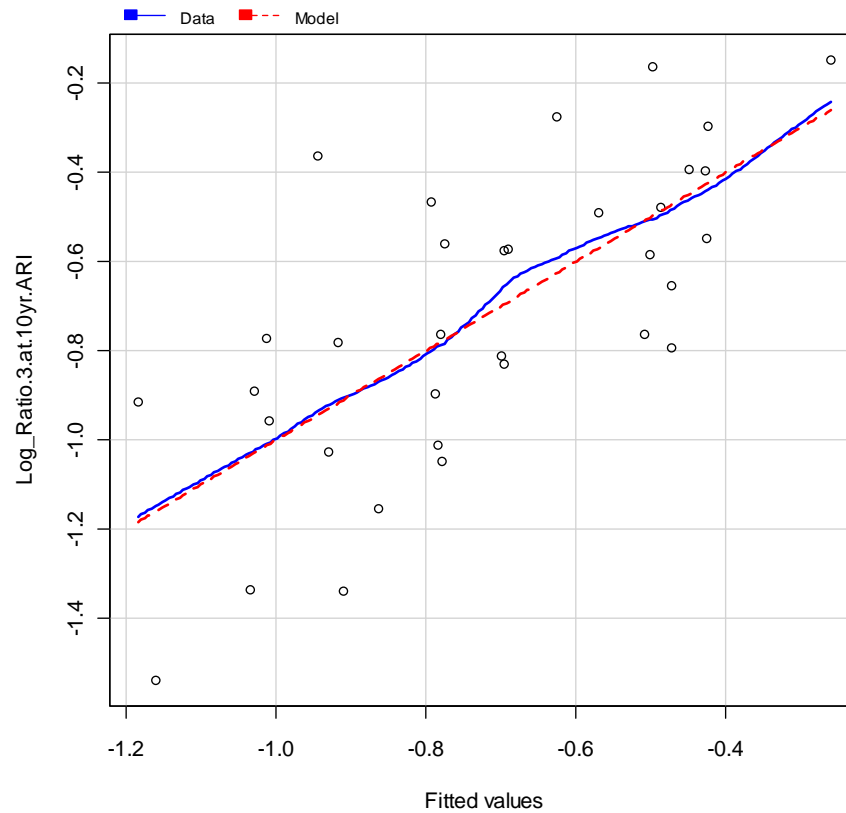**Figure 100 Standardised residuals for cluster 3 Baseflow Volume Ratio regression**

Figure 101 Marginal model plot for cluster 3 Baseflow Volume Ratio regression

Figure 102 Scatter plot of continuous predictor variables for cluster 4 Baseflow Volume Ratio regression

Figure 103 Fitted values against observed values for cluster 4 Baseflow Volume Ratio regression

Figure 104 Diagnostic plots for cluster 4 Baseflow Volume Ratio regression

Figure 105 Standardised residuals for cluster 4 Baseflow Volume Ratio regression

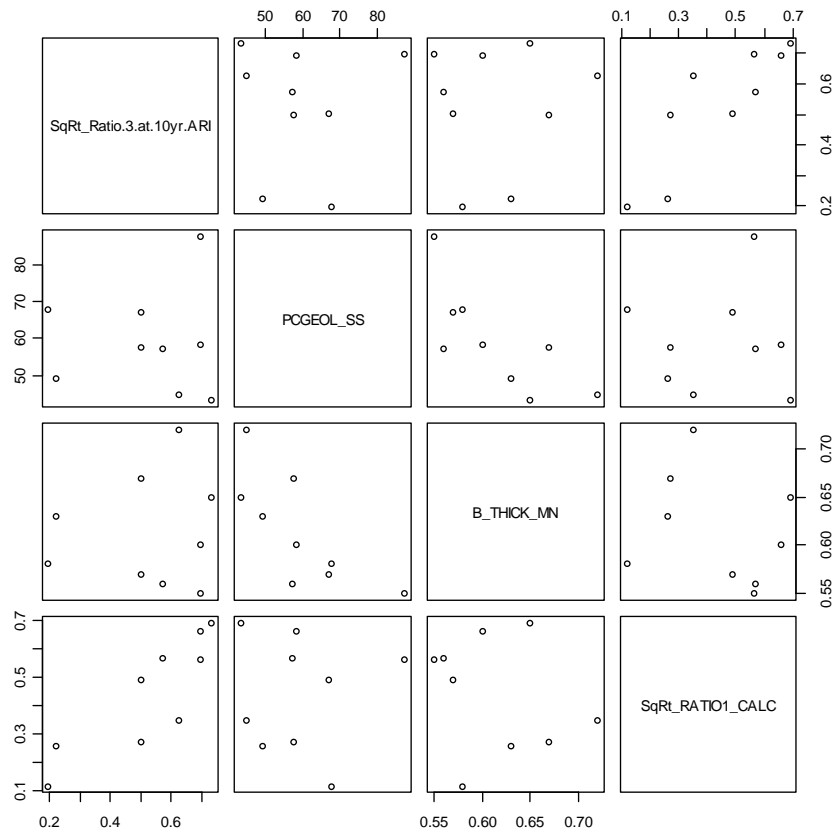**Figure 106 Marginal model plot for cluster 4 Baseflow Volume Ratio regression**

Figure 107 plot of continuous predictor variables for cluster 5 Baseflow Volume Ratio regression
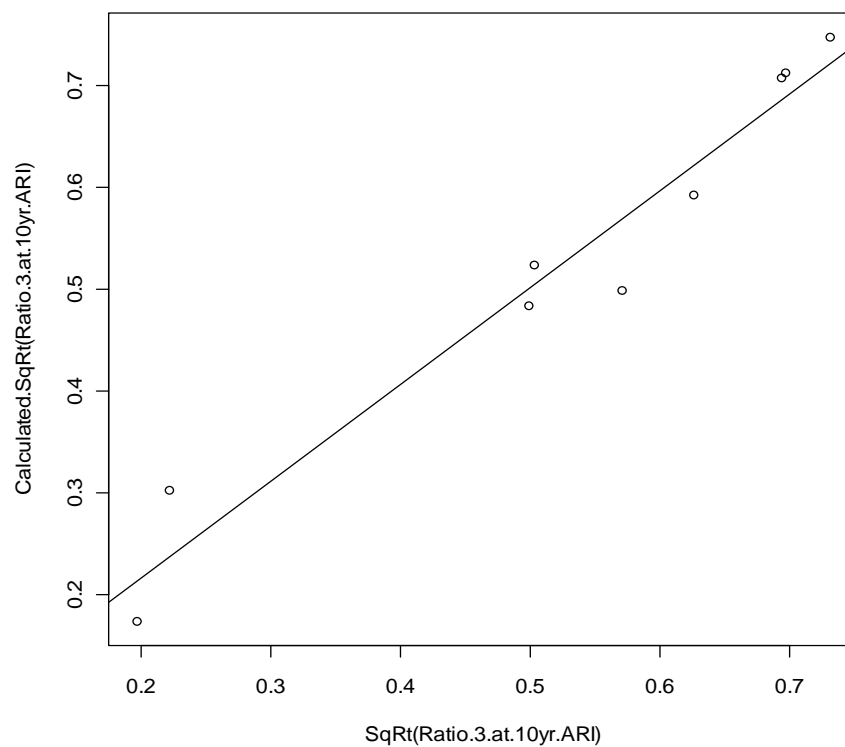

Figure 108 Fitted values against observed values for cluster 5 Baseflow Volume Ratio regression
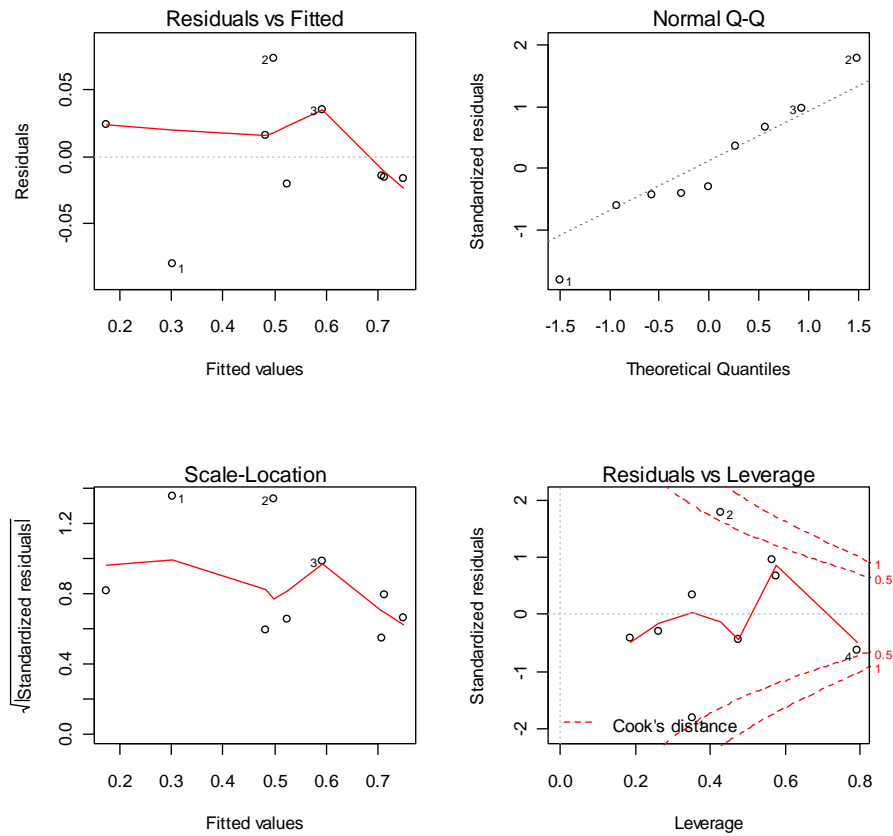
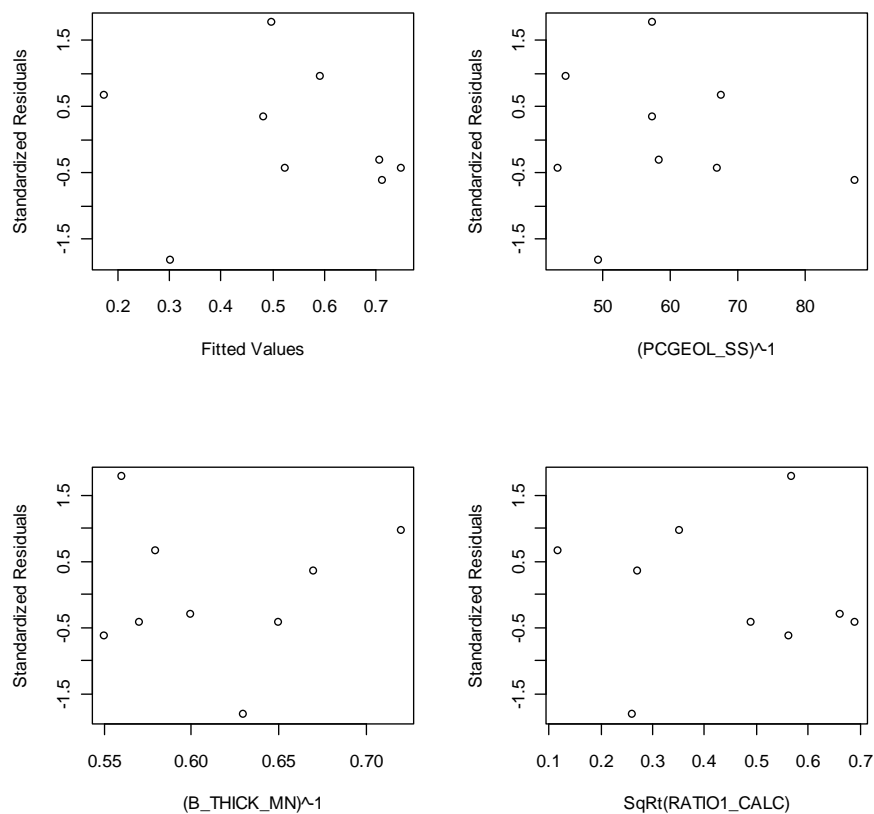**Figure 109 Diagnostic plots for cluster 5 Baseflow Volume Ratio regression**

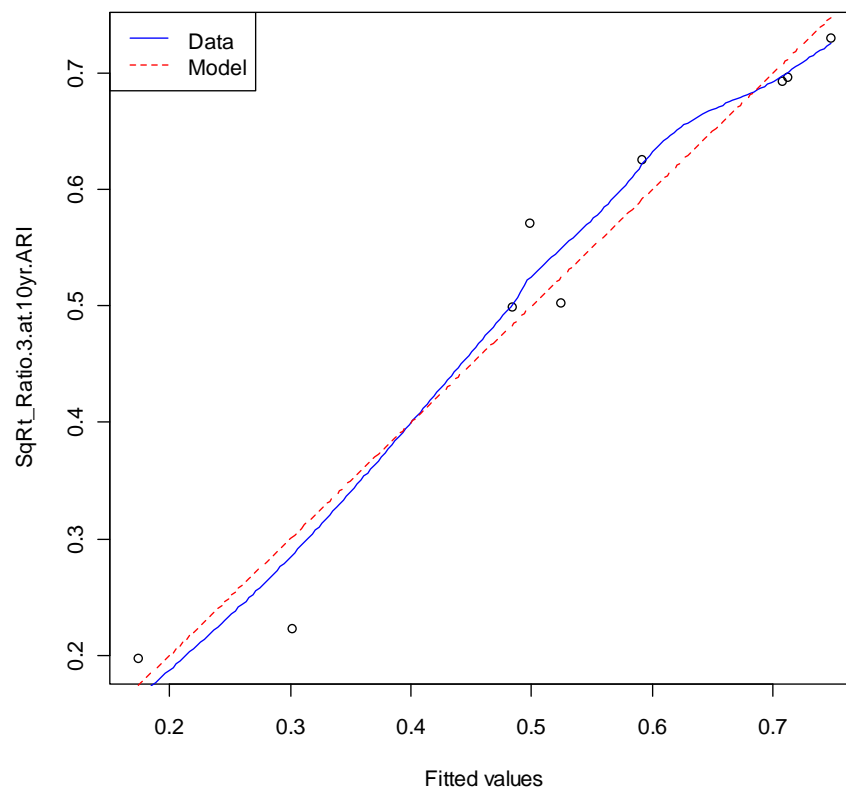**Figure 110 Standardised residuals for cluster 5 Baseflow Volume Ratio regression**

Figure 111 Marginal model plot for cluster 5 Baseflow Volume Ratio regression

Appendix E.1 Testing colinearity of variables

Colinearity of the variables can be tested by considering the variation inflation values. Values greater than 5 indicate that variables within the regression relationship are similar. The variation inflation factors relevant to the regressions developed in this study are presented in the following tables.

Table 15 Variation inflation factors calculated for Baseflow Peak Ratio regressions

	Variation Inflation Factor
Cluster 1	
$B\_KSAT\_MN^{0.2}$	1.12
PCGEOL_AU	1.02
PCVEG	1.12
Cluster 2	
A_FCP_MN	1.05
A_KSAT_MN	1.16
PCGEOL_AU	1.16
Cluster 3	
PCGEOL_AU	1.01
PCGEOL_SS	3.89
RAIN_MN	3.90
Cluster 4	
EVAP_MAX	1.13
PCGEOL_AS	1.16
PCGEOL_SS	1.09
Cluster 5	
B_KSAT_MN	3.32
$PCGEOL\_IM^{0.2}$	1.22
$PCVEG^{0.2}$	3.56

Table 16 Variation inflation factors calculated for Baseflow Volume Ratio regressions

	Variation Inflation Factor
Cluster 1	
A_KSAT_MN$^{0.2}$	1.48
EVAP_MAX	1.22
Log(RATIO1_CALC)	1.69
Cluster 2	
A_SAT_MN^{3}	2.02
PCGEOL_IM	1.73
Log(RATIO1_CALC)	1.23
Cluster 3	
PCGEOL_AS	1.18
PCGEOL_IM	2.22
Log(RATIO1_CALC)	2.26
Cluster 4	
SqRt(SLOPE_MAX)	1.07
EVAP_MAX	1.36
Log(RATIO1_CALC)	1.33
Cluster 5	
PCGEOL_SS	2.46
B_THICK_MN	2.61
SqRt(RATIO1_CALC)	1.14