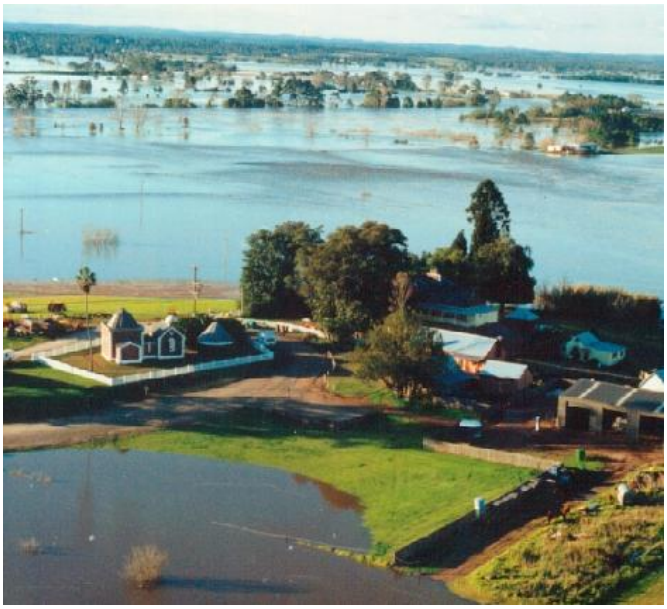


Australian Rainfall & Runoff

Revision Projects

PROJECT 5

Regional Flood Methods

STAGE 2 REPORT

P5/S2/015

June 2012

ENGINEERS AUSTRALIA
Water Engineering

ENGINEERS
AUSTRALIA
Water Engineering

Engineers Australia
Engineering House
11 National Circuit
Barton ACT 2600

Tel: (02) 6270 6528
Fax: (02) 6273 2358
Email:arr@engineersaustralia.org.au
Web: www.engineersaustralia.org.au

AUSTRALIAN RAINFALL AND RUNOFF
REVISION PROJECT 5: REGIONAL FLOOD METHODS

STAGE 2 REPORT

JUNE, 2012

Project Project 5: Regional Flood Methods	**AR&R Report Number** P5/S2/015
Date 21 June 2012	**ISBN** 978-0-85825-908-9
Contractor University of Western Sydney	**Contractor Reference Number** 20731.64125
Authors Ataur Rahman Khaled Haddad Mohammad Zaman Elias Ishak George Kuczera Erwin Weinmann	**Verified by** 

ACKNOWLEDGEMENTS

This project was made possible by funding from the Federal Government through the Department of Climate Change. This report and the associated project are the result of a significant amount of in kind hours provided by Engineers Australia Members.

ENGINEERS
AUSTRALIA
Water Engineering

Contractor Details

The University of Western Sydney
School of Engineering, Building XB, Kingswood
Locked Bag 1797, Penrith South DC, NSW 1797, Australia

Tel: (02) 4736 0145
Fax: (02) 4736 0833
Email: a.rahman@uws.edu.au
Web: www.uws.edu.au

University of
Western Sydney

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 AR&R 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 5: Regional Flood Methods

The most commonly encountered hydrological problem associated with estimating flood flows is that of estimating the flood flow of a given Annual Exceedence Probability (AEP) at a location where the historical monitored information is inadequate for frequency analysis. These locations are referred to as ungauged catchments. Numerous alternative techniques have been developed historically in the different regions of Australia to provide the necessary design flow predictions in ungauged catchments. The current diversity of approaches has resulted in predicted flows varying significantly at the interfaces between regions. It was recognised that there was a need to develop generic techniques that can be applied across the whole country, to test these techniques, and to develop appropriate guidance in their usage.

The aim of Stage 2 of Project 5 was to test the suitability of alternative national approaches to the estimation of design peak flow predictions for ungauged catchments.

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:

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

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, Chair NCWE, 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

Steve Finlay, Engineers Australia

Related Appointments:

ARR Project Engineer: Monique Retallick, WMAwater

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

PROJECT TEAM AND CONTRIBUTORS INCLUDING STATE TEAMS

Project Team Members:

- Ataur Rahman # (Research Project Leader)
- Khaled Haddad
- George Kuczera # (EA Project Manager)
- Erwin Weinmann #
- Ashish Sharma #

Contributors:

- James Ball #
- Mark Babister #
- Monique Retallick #
- Mohammad Zaman
- Elias Ishak
- William Weeks #
- Tom Micevski #
- Andre Hackelbusch #
- Luke Palmen #
- Guna Hewa #
- Trevor Daniell #
- David Kemp #
- Sithara Walpita Gamage
- Subhashini Wella Hewage #
- Syed Quddusi
- Md Jajal Uddin
- Fiona Ling #
- Crispin Smythe#
- Chris MacGeorge#
- Bryce Graham#
- James Pirozzi #
- Gavin McPherson #
- Chris Randall #
- Robert French #
- Wilfredo Caballero #

- Khaled Rima #
- Tarik Ahmed
- Lakshman Rajaratnam#
- Jerome Goh#
- Patrick Thompson#
- Neil Coles #
- Leanne Pearce #
- Mark Pearcey #
- John Ruprecht #

(# indicates unpaid team members)

Independent Review Team:

- Asaad Shamseldin, University of Auckland

LIST OF ORGANISATIONS PROVIDING DATA AND ASSISTANCE

- Australian Bureau of Meteorology
- Department of Sustainability and Environment (Victoria)
- Thiess Services Victoria
- Department of Transport and Main Roads (Qld)
- Department of Environment and Resource Management (Qld)
- ENTURA (formerly Hydro Tasmania)
- Department of Primary Industries, Parks, Water and Environment (TAS)
- Department of Water, Land and Biodiversity Conservation (SA)
- Department of Natural Resources, Environment, the Arts and Sport (NRETAS) (NT)
- University of Western Sydney
- University of Newcastle
- University of South Australia
- University of New South Wales
- Department of Environment, Climate Change and Water (NSW)
- Department of Water (Western Australia)
- WMAwater (NSW)

EXECUTIVE SUMMARY

In Australia, there are many streams where there is little/no recorded streamflow data. In these ungauged and poorly gauged catchments, there is insufficient information/data to obtain design flood estimates which are needed to size hydraulic structures, plan and design other water infrastructure and undertake various environmental and ecological studies. Regional flood frequency analysis (RFFA) is the most commonly adopted technique to derive design flood estimates on the ungauged catchments. A RFFA method attempts to transfer flood characteristics information from a group of gauged catchments to an ungauged catchment of interest. The RFFA methods recommended in the Australian Rainfall and Runoff (ARR) in 1987 need updating to reflect the advancements in RFFA methods and new additional streamflow data. To update the RFFA methods in the ARR, a project team was formed in 2008 and since then the team has been carrying out research and investigations, which have now formed part of Project 5 'Regional Flood Methods in Australia' in the ARR revision projects.

So far, Stage I and Stage II of Project 5 have been completed. The major outcomes of Stage I project are as follows.

Formation of Project 5 team and establishment of contacts and cooperations with various state agencies to obtain necessary streamflow data and relevant information. About 31 researchers/engineers from over 14 organisations of various Australian states directly contributed to Project 5 Stage I.

Preparation of initial version of national database which involved examination of a large number of potential stations from each state, short-listing of the stations, infilling the gaps in annual maximum flood series, test for outliers, test for trends and test for rating curve extrapolation error. In Stage I, databases for Victoria, NSW, Qld, Tasmania and SA were prepared.

Based on detailed literature review, consultation with Project 5 team and various state representatives and ARR Technical Committee, a number of RFFA methods were selected for detailed investigation which included the Probabilistic Rational Method, Quantile Regression Technique and Parameter Regression Technique. For the regression-based methods, both ordinary least squares and generalised least squares methods were considered. For the formation of regions, fixed state-based regions and region-of-influence (based on geographical proximity) were considered.

From initial trend analysis, a good number of stations showed trends in the annual maximum flood series data; these stations were not included in the development and testing of the RFFA methods. However, it was decided to conduct further investigation e.g. impact of serial and cross-correlation on the trends, and relationship between the identified trends and catchment and climate change/variability indices and impacts of the identified trends on regional flood estimates with respect to locations and ARIs of the flood estimates.

The major findings from Stage II project, presented in this report, are as follows.

A quality controlled national database consisting of 727 stations has been prepared.

It has been found that regression-based RFFA methods (such as the quantile regression technique (QRT) or parameter regression technique (PRT)) are preferable to the Probabilistic Rational Method.

It has also been found that Bayesian QRT and Bayesian PRT methods perform very similarly for various Australian states. Since the PRT method offers several additional advantages over the QRT (namely, in the PRT flood quantiles increase smoothly with increasing ARIs and from the regional LP3 distribution, flood quantiles of any ARI (in the range of 2 to 100 years) can be estimated), this has been recommended for general application in Australia.

From the comparison of fixed regions and region-of-influence (ROI) approaches, it has been found that, where a region contains a sufficient number of sites, the ROI approach outperforms the fixed regions. The mean annual flood model generally has the highest model error as compared to the SD and skew models. However, the SD and skew estimates are suffered greatly by sampling errors.

The developed RFFA methods in Stage II require data of two or three climatic and physical catchment characteristics (i.e. catchment area, representative design rainfall intensity and mean annual rainfall), which are easy to obtain. This would make the application of the recommended RFFA methods easy and simple.

It has been found that the recommended RFFA methods i.e. GLS-PRT-ROI and GLS-PRT-fixed region perform quite well for the smaller catchments in the database where there is no evidence that smaller catchments perform poorly than the medium and larger catchments. The possibility of extending the RFFA method to very small catchments beyond the limit of

the current Project 5 database has been examined; however, further study is needed to develop an acceptable method.

The development of a simple Large Flood Regionalisation Model for regional flood estimation in the major flood range was investigated in Stage I of the project (see Stage I report), which however did not consider the impacts of inter-station correlation of the annual maximum flood series among different pairs of stations on final design flood estimates. Some preliminary investigations on inter-station correlation have been undertaken in this report, which however needs further investigation.

There is insufficient streamflow data availability at both temporal and spatial scales in the arid and semi-arid regions of Australia that can be used to develop statistically meaningful RFFA methods. A simplified index type RFFA is recommended for arid/semi-arid regions of Australia where four separate regions are recommended at this stage (this needs further development and testing before inclusion in the ARR).

In the preliminary investigation (see Stage I report), about 13% of the selected stations (for Project 5) showed a trend in the annual maximum flood series data. In this report, the impacts of serial and cross-correlation on trend analysis have been investigated. At the significance level of 10% and with the consideration of the cross-correlation among the sites in the network, the field significance of downward trends in the annual maximum flood series was detected over the whole country. However, the field significance of upward trends was discovered to be statistically non-significant at 10% significant level. The impacts of the identified trends on regional flood quantile estimates for ARIs in the range of 2 to 100 years will be investigated in Stage III of the project. This is expected to produce climate change adjustment factors as a function of ARIs and locations across Australia.

The testing of the recommended RFFA methods for Australia by various states/stakeholders in cooperation with the Project 5 team has been recommended. A set of future tasks has been identified. Also, the scope of developing an application tool/software has been indicated.

Stage II has developed a firm basis for recommendations on the RFFA methods to be included in the revised ARR Chapter (4^{th} edition). It has also identified future research and development work in Stage III of the Project, required to develop the Stage II findings into a final set of methods, design databases, user guidelines and application tools.

The results presented in this report are applicable to the rural catchments in the vicinity of the catchments selected in this study; this should not be applied to urban catchments.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

AIC	Akaike information criterion
ARI	Average recurrence interval
ARR	Australian Rainfall and Runoff
AVP	Average variance of prediction
BGLSR	Bayesian Generalised Least Squares Regression
BIC	Bayesian information criterion
CE	Coefficient of efficiency
EVR	Error variance ratio
FF	Frequency factor
GLSR	Generalised Least Squares Regression
LFRM	Large flood regionalisation model
LP3	Log Pearson Type 3 distribution
MEV	Model error variance
MPRE	Mean percent relative error
MSRE	Mean squared relative error
OLSR	Ordinary Least Squares Regression
PRM	Probabilistic Rational Method
PRT	Parameter Regression Technique
QRT	Quantile Regression Technique
RE	Relative error
RFFA	Regional flood frequency analysis
ROI	Region-of-influence approach
RMSE	Root mean squared error
SEP	Standard error of prediction

1. Introduction

1.1 Background

To upgrade the regional flood estimation methods in Australian Rainfall and Runoff (ARR), a project team has been working since early 2006 with researchers from various Australian states, as described in the Stage I report of Project 5 'Regional Flood Methods in Australia' (see Stage I report, Rahman et al., 2009).

ARR Project 5 has the following broad objectives:

- To collate a quality controlled national database covering all the Australian states for testing and developing new regional flood frequency analysis (RFFA) methods.

- To select potential RFFA methods for detailed investigation using the prepared national database.

- To develop and test the selected RFFA methods to form the scientific basis of recommending suitable RFFA methods for inclusion in the 4th edition of ARR.

- To further test the recommended RFFA methods by various state agencies/stakeholders and the Project 5 team to assess their applicability in practical situations under a wide range of catchment and hydrologic conditions and make the necessary updates/improvements before including them in the ARR chapter.

- To develop application tools that will facilitate the application of the recommended RFFA methods by practitioners.

- To identify areas where improvements to the database and further development of RFFA methods would be desirable.

As in Dec 2010, Stage I and Stage II of the project have been completed. The major outcomes of Stage I project were:

1) Formation of Project 5 team and establishment of contacts and cooperation with various state agencies to obtain necessary streamflow data and relevant information. More than

31 researchers/engineers from various Australian states directly contributed to Project 5 Stage I.

2) Preparation of the initial version of the national database, which involved examination of a large number of potential stations from each state, short-listing of the stations, infilling the gaps in annual maximum flood series, testing for outliers, trends and for rating curve extrapolation error. In Stage I, databases for Victoria, NSW, Qld, Tasmania and SA were prepared.

3) Based on a detailed literature review, consultation with the Project 5 team and various state representatives and ARR Technical Committee members, a number of RFFA methods were selected for detailed investigation, which included the Probabilistic Rational Method, the Quantile Regression Technique and the Parameter Regression Technique. For the regression-based methods, both ordinary least squares (OLS) and generalised least squares (GLS) methods were considered. For the formation of regions, fixed state-based regions and region-of-influence (ROI) approaches (based on geographical proximity) were considered. It was found that regression-based methods outperformed the Probabilistic Rational Method. Furthermore, both the Quantile Regression Technique and the Parameter Regression Technique with GLS regression demonstrated potential for inclusion in the 4th edition of ARR. The superiority of the ROI approach over the fixed region was established for regions with sufficient number of sites.

4) From the initial trend analysis, a substantial number of stations showed trends in the annual maximum flood series data; these stations were not included in the development and testing of the RFFA methods. However, it was decided to conduct further investigations, e.g. impact of serial and cross-correlation on the trends, relationship between the identified trends and catchment and climate change/variability indices, and impacts of the identified trends on regional flood estimates with respect to locations and ARIs of the flood estimates.

5) It was found that a simple Large Flood Regionalisation Model can be coupled with the GLS regression to develop an easy to apply RFFA method for application in the large flood range.

6) It was found necessary to test the applicability of the selected RFFA methods to very small catchments, e.g. down to 0.10 km^2, for which little or no recorded streamflow data are available.

7) Up to Dec 2010, a total of five peer-reviewed journal papers and five conference papers have been published based on Project 5 Stage I research outcomes.

Project 5 Stage II set the following deliverables:

- Detailed testing on national basis of selected RFFA methods based on Stage I outcomes.
- Development of a method for incorporating climate change signals.
- Selection of a national method or regions for different methods.
- International bench marking of the approach.
- Testing of methods that incorporate climate change adjustments.

To achieve the above deliverables, an extensive research program has been undertaken to assess the performance of different RFFA methods when applied in a consistent manner in different regions, and to identify the best method for forming regions, dealing with boundary issues, establishing limits of application and exploring methods for dealing with special situations (small catchments, arid areas, rare events, changing climate). This report presents the detailed outcomes of this research.

1.2 Scope of the report

The report presents the data updating and results in relation to Project 5 Stage II. This broadly covers the following aspects:

- An update of the national database that was initially prepared as a part of Project 5 Stage I.
- A comparison of the short-listed regional flood estimation methods (as an outcome of Project 5 Stage I) to make a final recommendation of the regional flood estimation method(s) to be adopted for Australia in the 4th edition of the ARR.
- Detailed results of the development and testing of the recommended regional flood estimation method(s).
- Initial results of regional flood estimation methods for arid and semi-arid regions of Australia.
- Results from on-going investigations on the identification of time trends in the Australian annual maximum flood series data.

- Results from on-going investigations on the development of a simplified regional flood estimation method for large floods.
- Identification of the scope for further development and testing of the recommended regional flood estimation methods by various state agencies/stake holders in association with Project 5 team.
- Identification of the scope for the development of an application tool for routine application of the recommended regional flood estimation method(s).

1.3 Outline of the report

There are 12 chapters in the report, as follows.

Chapter 1 provides the background and scope of the project.

Chapter 2 details the data updating and archiving for all the Australian states/regions.

Chapter 3 presents a brief review of various RFFA methods and identifies potential methods for application in Australia. These methods include the Probabilistic Rational Method (PRM) and regression based methods such as the Quantile Regression Technique (QRT) and the Parameter Regression Technique (PRT). In the PRT, the parameters of the log-Pearson type 3 (LP3) distribution are regionalised. A detailed comparison of the PRM and QRT is presented in this chapter as well.

Chapter 4 presents results in relation to the development and comparison of the QRT and PRT. This chapter also compares fixed region and region-of-influence approaches for the formation of regions. Results of the application of the QRT and PRT methods are presented for all the Australian states and territories.

Chapter 5 presents the investigations on the applicability of the selected regional flood estimation method(s) to smaller catchments.

Chapter 6 presents the results of on-going investigations into the development of a large flood regionalisation model.

Chapter 7 presents the current state of development of regional flood estimation methods for selected arid and semi-arid regions of Australia.

Chapter 8 presents the investigations on the detection of time trends in Australian annual maximum flood series data.

Chapter 9 summarises the investigations made in this report.

Chapter 10 recommends the regional flood estimation method(s) to be included in the 4th edition of the ARR. It also presents the scope of the further development and testing of the recommended regional flood estimation methods by various state agencies/stakeholders in association with Project 5 team.

Chapter 11 presents the draft specification for developing an application tool for routine application of the recommended regional flood estimation methods.

Chapter 12 presents major conclusions derived from this study.

Finally, references and appendices are provided at the end.

Stage II has developed a firm basis for recommendations on the RFFA methods to be included in the revised ARR Chapter (4th edition). It also identified future research and development work in Stage III of the Project, required to develop the Stage II findings into a final set of methods, design databases, user guidelines and application tools.

2. Data updating and archiving

2.1 General

A total of 96 catchments from NSW & ACT, 131 catchments from Victoria, and 30 catchments from SA were selected as described in Project 5 Stage I report (Rahman et al., 2009). These selected catchments are listed in Appendix A of this report.

The initially selected catchments from Tasmania, Queensland and NT as provided in Stage I report have been updated as described below. Also, catchments from Western Australia have been selected here, which were not covered in the Stage I report. Similarly, data on arid/semi-arid regions have been added. The criteria of catchment selection is presented in Stage I report (Rahman et al., 2009).

2.2 Data for New South Wales and ACT

A total of 96 catchments have been selected from New South Wales and ACT (listed in Appendix Table A1).

The record lengths of annual maximum flood series of these 96 stations range from 25 to 75 years (mean: 37 years, median: 34 years and standard deviation: 11.4 years). The distribution of record lengths is shown in Figure 2.1. The record lengths and the number of stations constitute a dataset which is suitable for regional flood frequency analysis.

The catchment areas of the selected 96 catchments range from 8 km^2 to 1010 km^2 (mean: 353 km^2 and median: 267 km^2). The geographical distribution of the selected 96 catchments is shown in Figure 2.2. The distribution of catchment areas of these stations is shown in Figure 2.3.

This data set does not include 4 stations in the arid/semi-arid part of New South Wales (see Section 2.9).

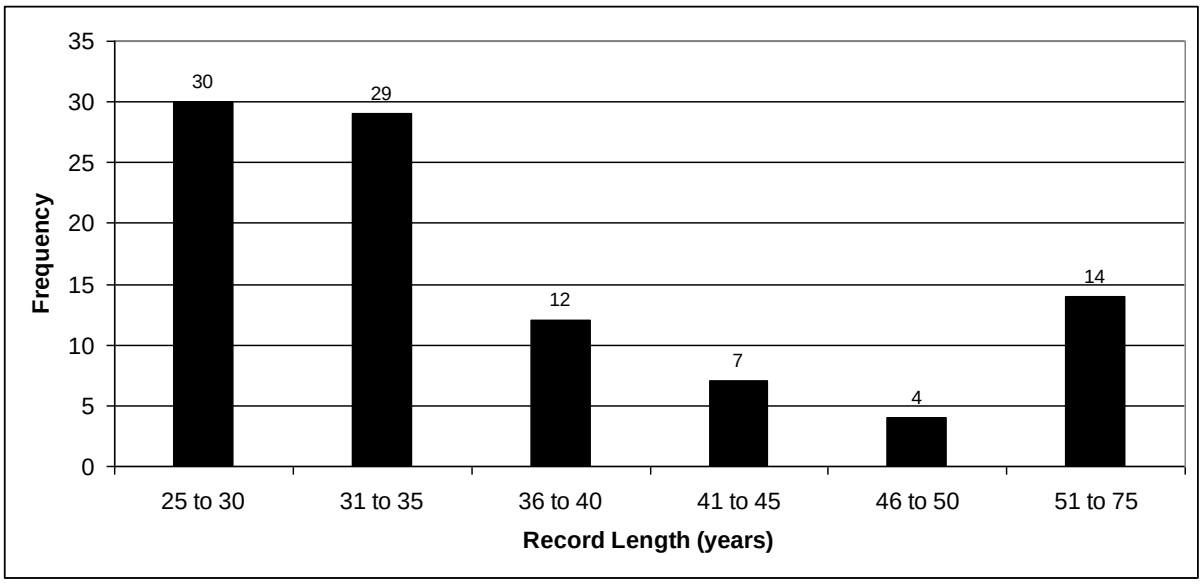


Figure 2.1 Distribution of streamflow record lengths of 96 stations from New South Wales.

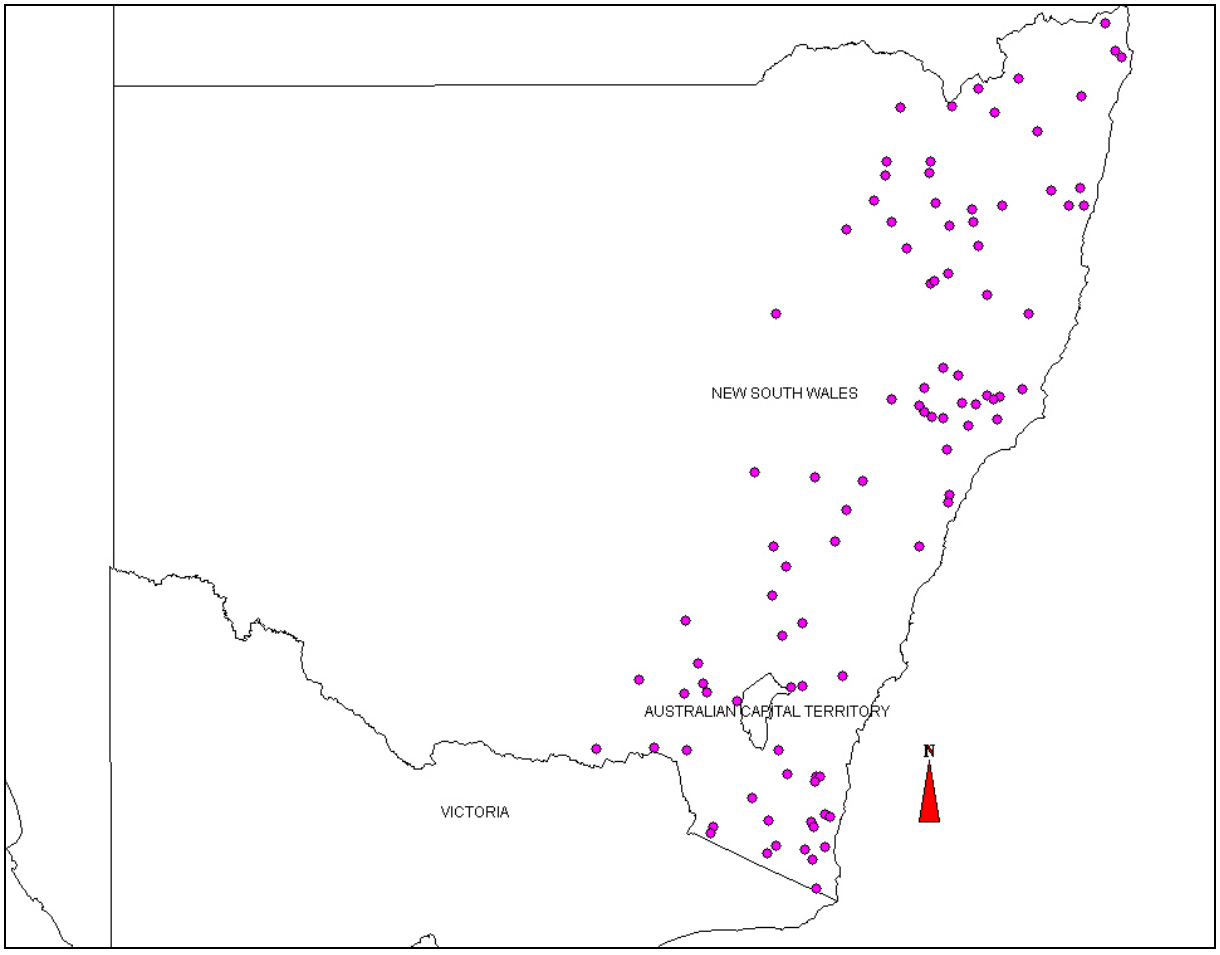


Figure 2.2 Geographical distributions of the selected 96 stations from NSW and ACT

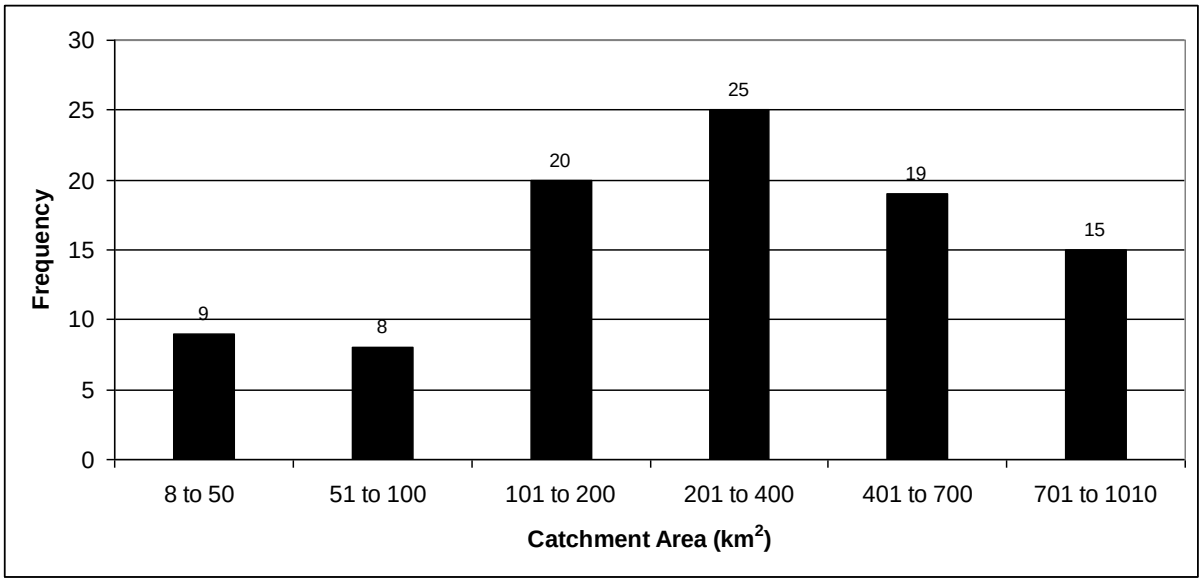


Figure 2.3 Distribution of catchment areas of 96 stations from NSW and ACT

2.3 Data for Victoria

A total of 131 catchments have been selected from Victoria (listed in Appendix Table A2).

The record lengths of annual maximum flood series of these 131 stations range from 26 to 52 years (mean: 33 years, median: 33 years and standard deviation: 4.6 years). The distribution of record lengths is shown in Figure 2.4.

The catchment areas of the selected 131 catchments range from 3 km^2 to 997 km^2 (mean: 321 km^2 and median: 289 km^2). The geographical distribution of the selected 131 catchments is shown in Figure 2.5. The distribution of catchment areas of these stations is shown in Figure 2.6.

This data set does not include 5 stations in the semi-arid part of Victoria (see Section 2.9).

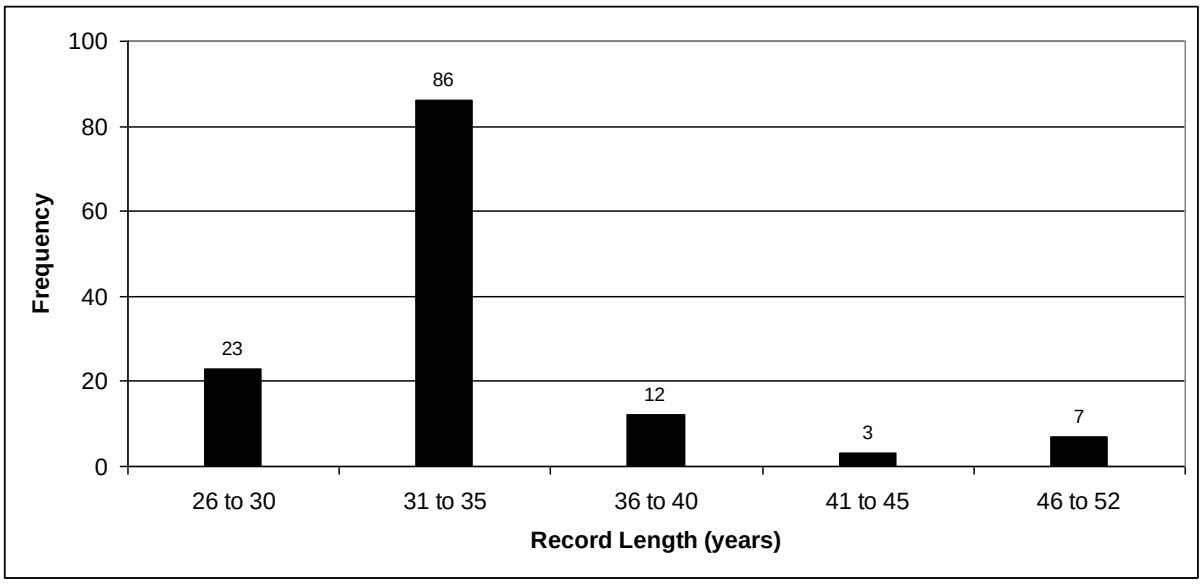


Figure 2.4 Distribution of streamflow record lengths of 131 stations from Victoria

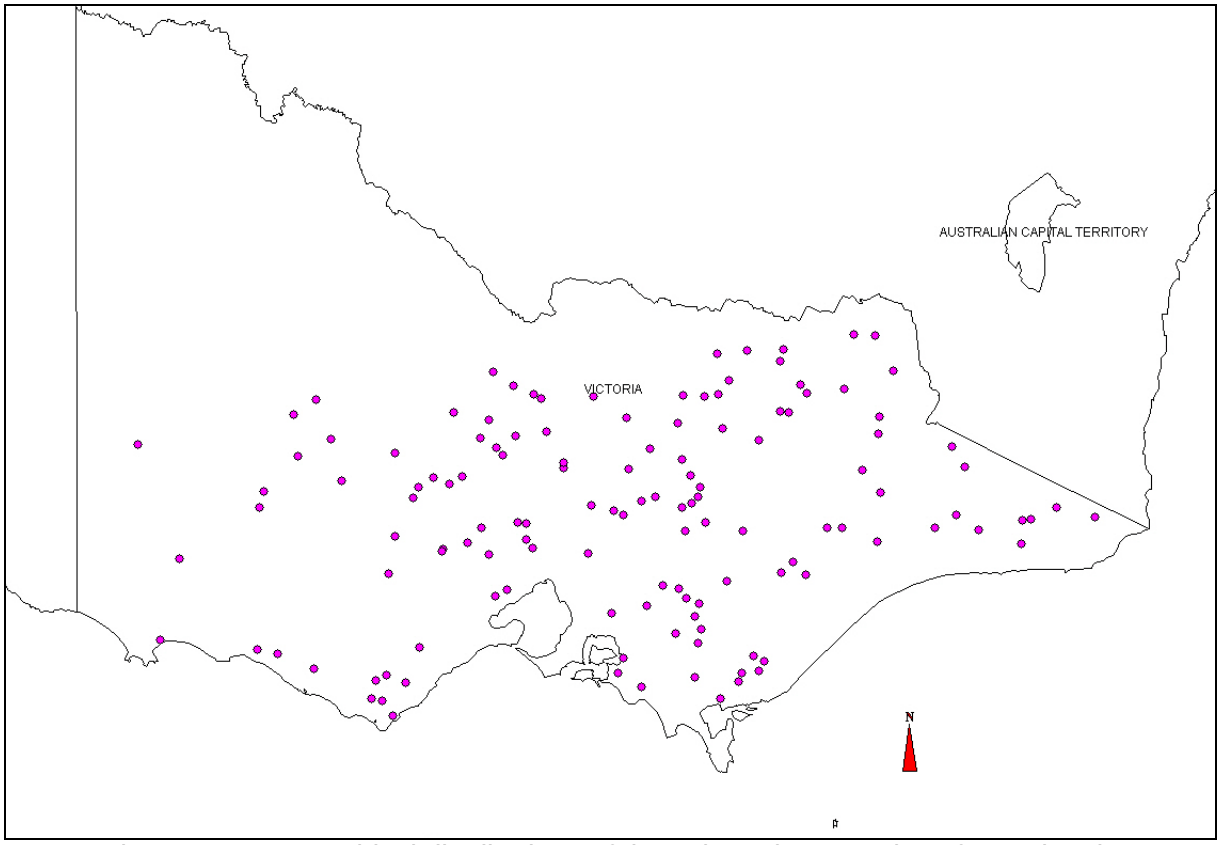


Figure 2.5 Geographical distributions of the selected 131 stations from Victoria

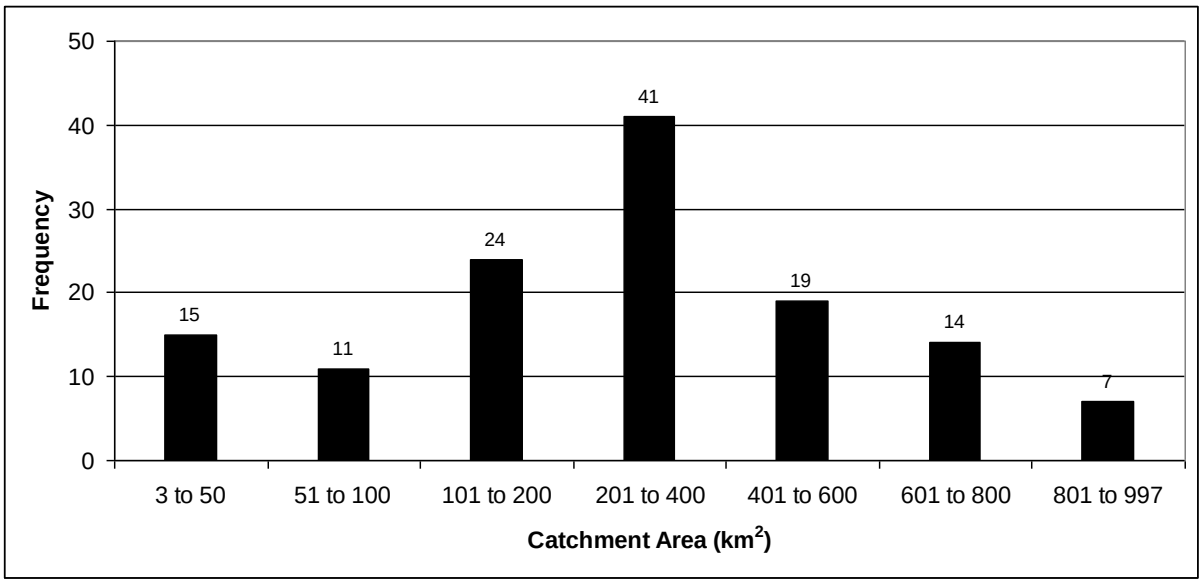


Figure 2.6 Distribution of catchment areas of 131 stations from Victoria

2.4 Data for South Australia

A total of 29 catchments have been selected from South Australia (listed in Appendix Table A3).

The record lengths of annual maximum flood series of these 29 stations range from 18 to 67 years (mean: 36 years, median: 34 years and standard deviation: 11.2 years). The distribution of record lengths is shown in Figure 2.7.

The catchment areas of the selected 30 catchments range from 0.6 km^2 to 708 km^2 (mean: 170 km^2 and median: 76.5 km^2). The geographical distribution of the selected 29 catchments is shown in Figure 2.8. The distribution of catchment areas of these stations is shown in Figure 2.9.

This data set does not include 6 stations in the arid/semi-arid part of South Australia (see Section 2.9).

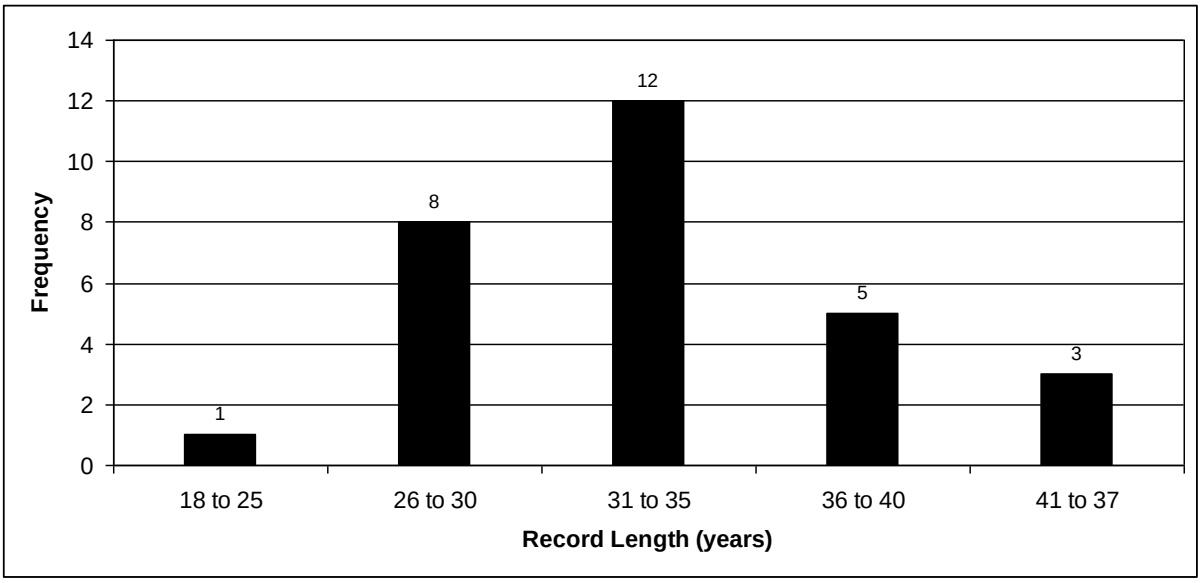


Figure 2.7 Distribution of streamflow record lengths of 29 stations from South Australia

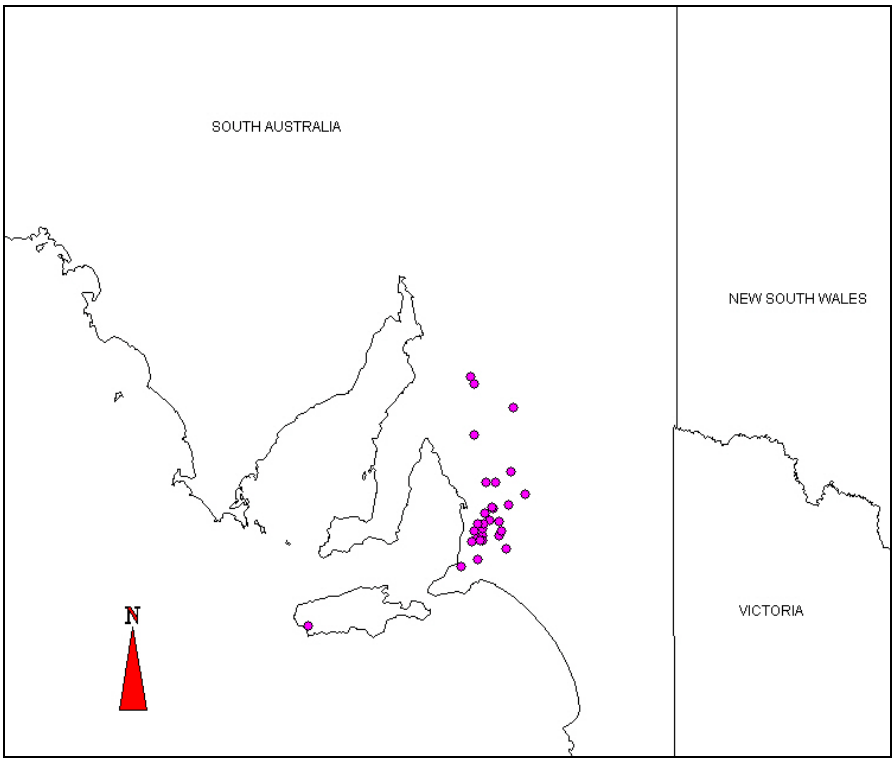


Figure 2.8 Geographical distributions of the selected 29 stations from South Australia

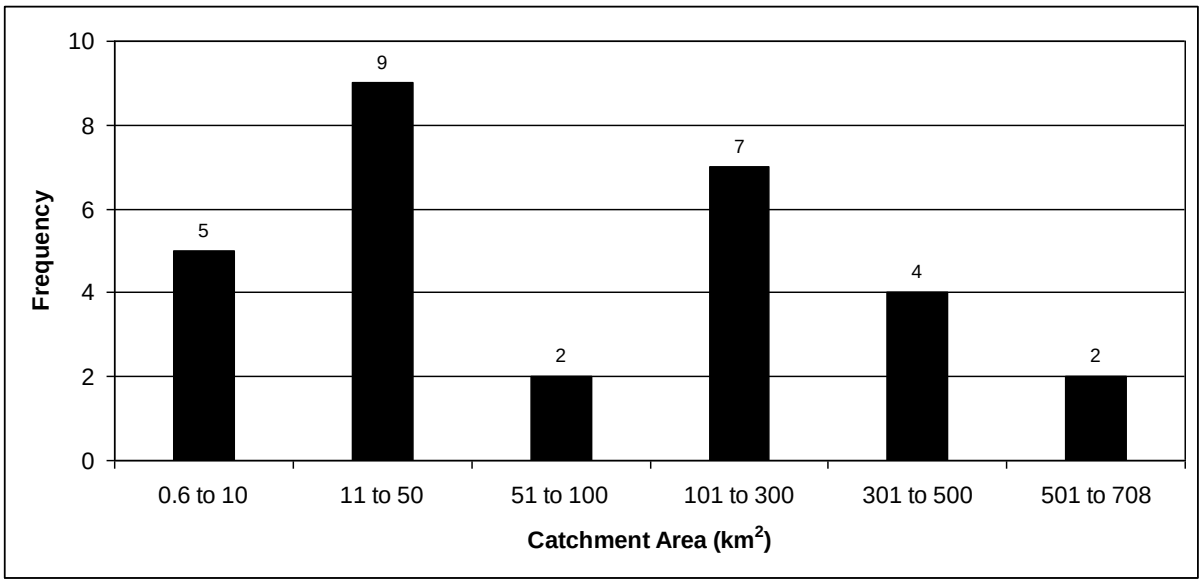


Figure 2.9 Distribution of catchment areas of 29 stations from South Australia

2.5 Data for Tasmania

A total of 53 catchments have been selected from Tasmania (listed in Appendix Table A4).

The record lengths of annual maximum flood series of these 53 stations range from 19 to 74 years (mean: 30 years, median: 28 years and standard deviation: 10.43 years). The distribution of record lengths is shown in Figure 2.10.

The catchment areas of the selected 53 catchments range from 1.3 km^2 to 1900 km^2 (mean: 323 km^2 and median: 158 km^2). The geographical distribution of the selected 53 catchments is shown in Figure 2.11. The distribution of catchment areas of these stations is shown in Figure 2.12.

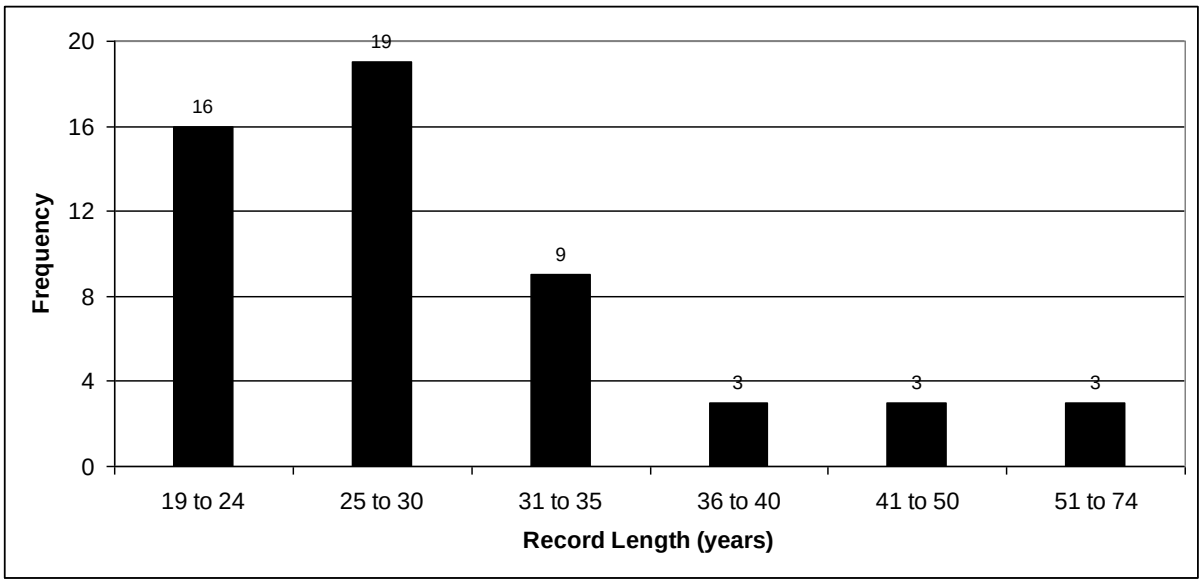


Figure 2.10 Distribution of streamflow record lengths of 53 stations from Tasmania

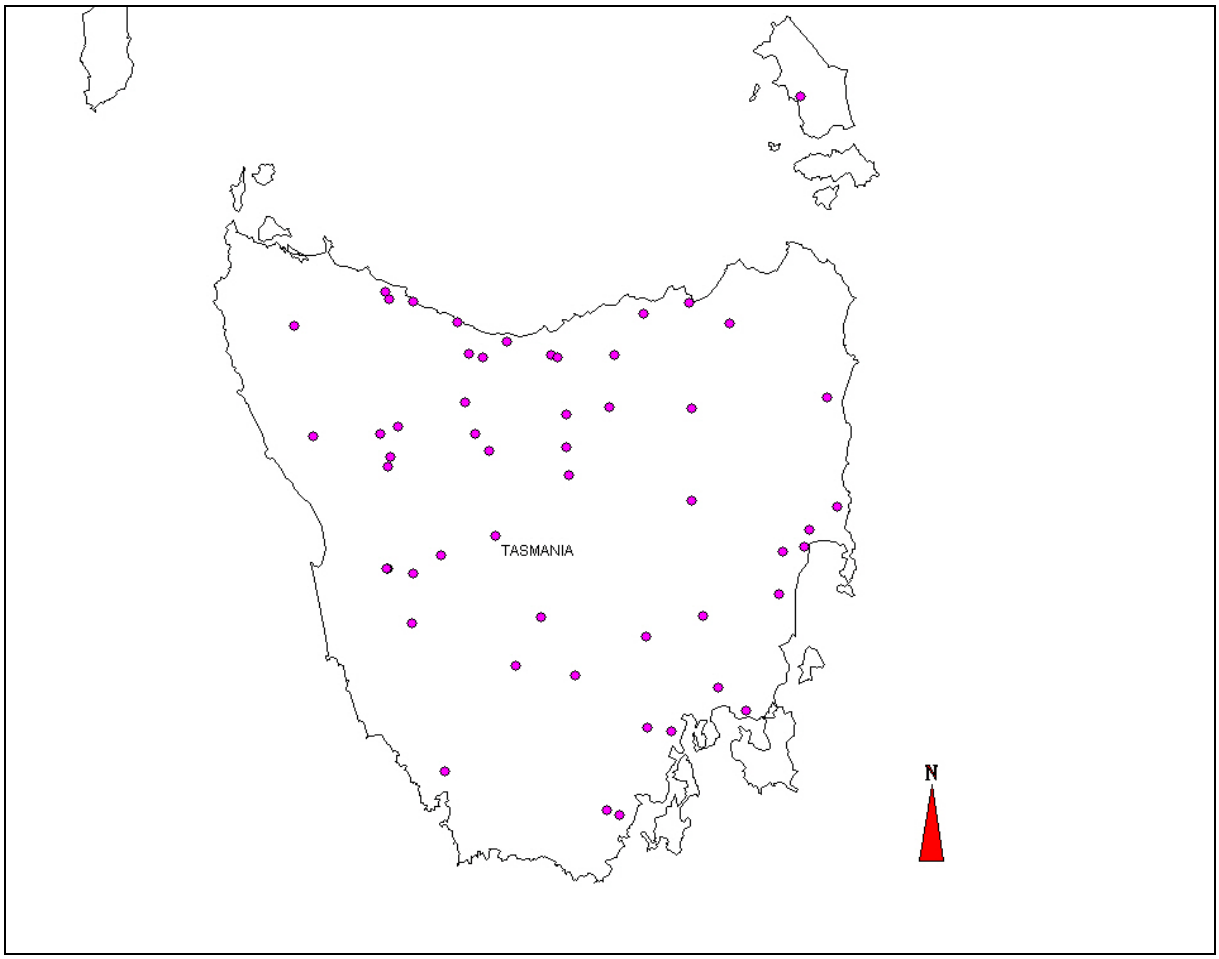


Figure 2.11 Geographical distributions of the selected 53 stations from Tasmania

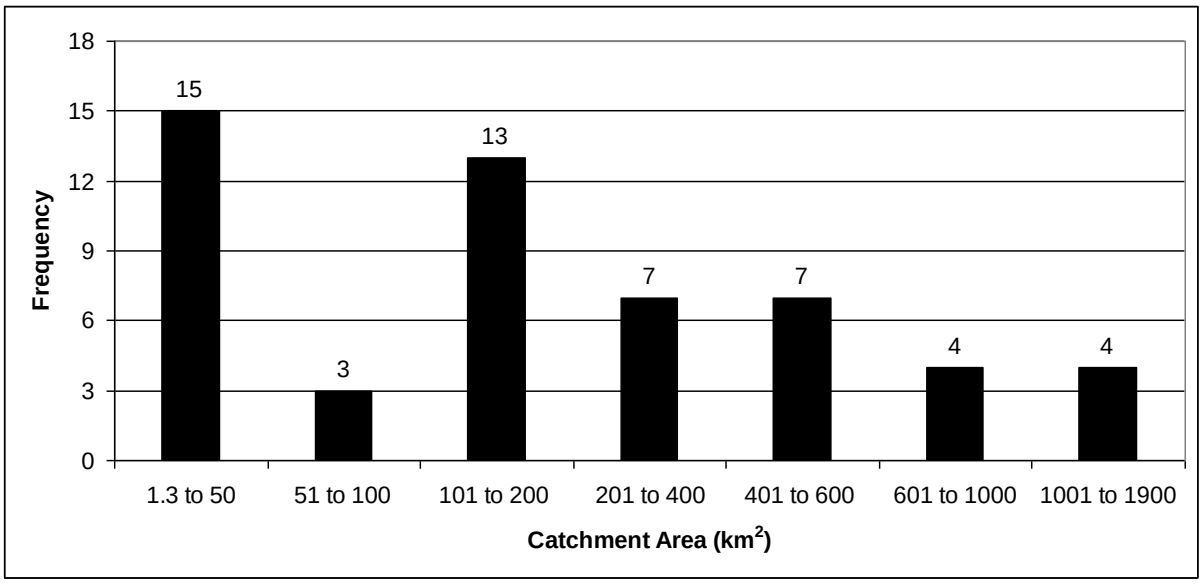


Figure 2.12 Distribution of catchment areas of 53 stations from Tasmania

2.6 Data for Queensland

A total of 172 catchments have been selected from Queensland (listed in Appendix Table A5).

The record lengths of annual maximum flood series of these 172 stations range from 25 to 97 years (mean: 41 years, median: 36 years and standard deviation: 15.2 years). The distribution of record lengths is shown in Figure 2.13.

The catchment areas of the selected 172 catchments range from 7 km^2 to 963 km^2 (mean: 325 km^2, median: 254 km^2). The geographical distribution of the selected 172 catchments is shown in Figure 2.14. The distribution of catchment areas of these stations is shown in Figure 2.15.

This data set does not include 16 stations in the arid/semi-arid part of Queensland (see Section 2.9).

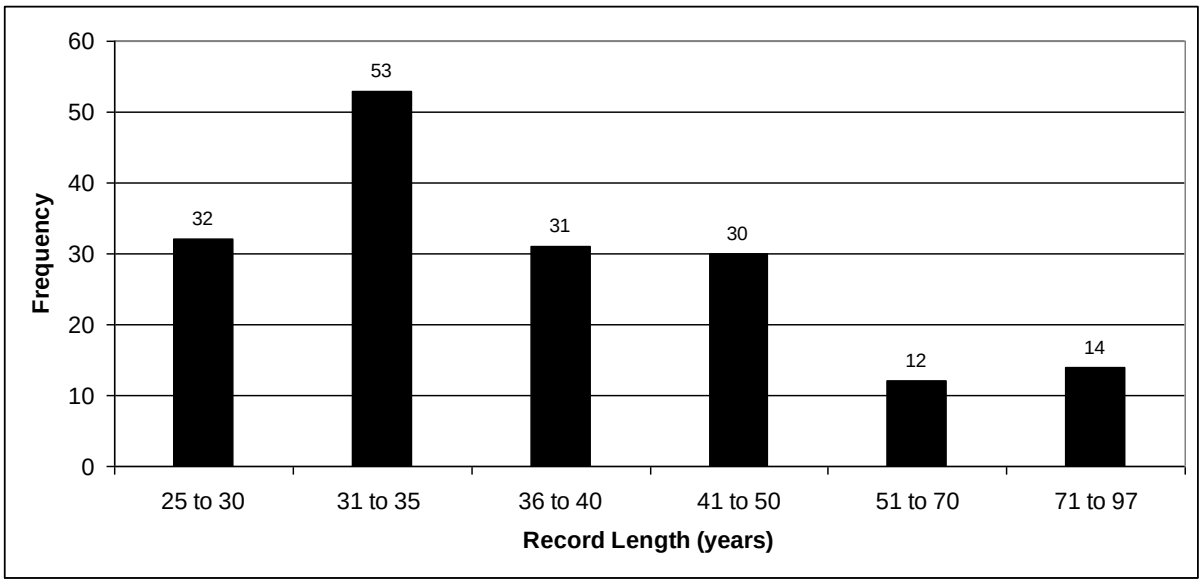


Figure 2.13 Distribution of streamflow record lengths of 172 stations from Queensland

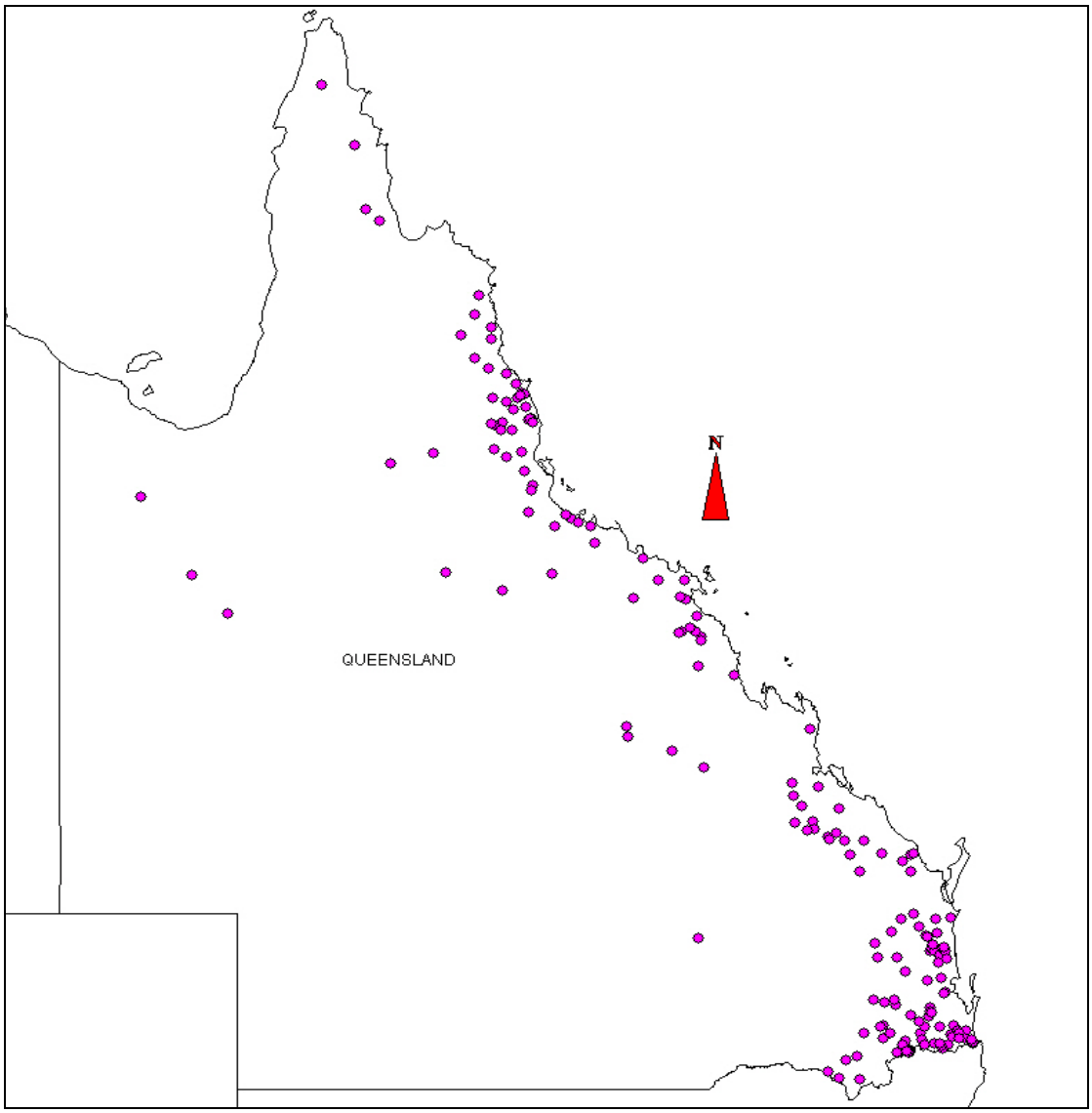


Figure 2.14 Geographical distributions of the selected 172 stations from Queensland

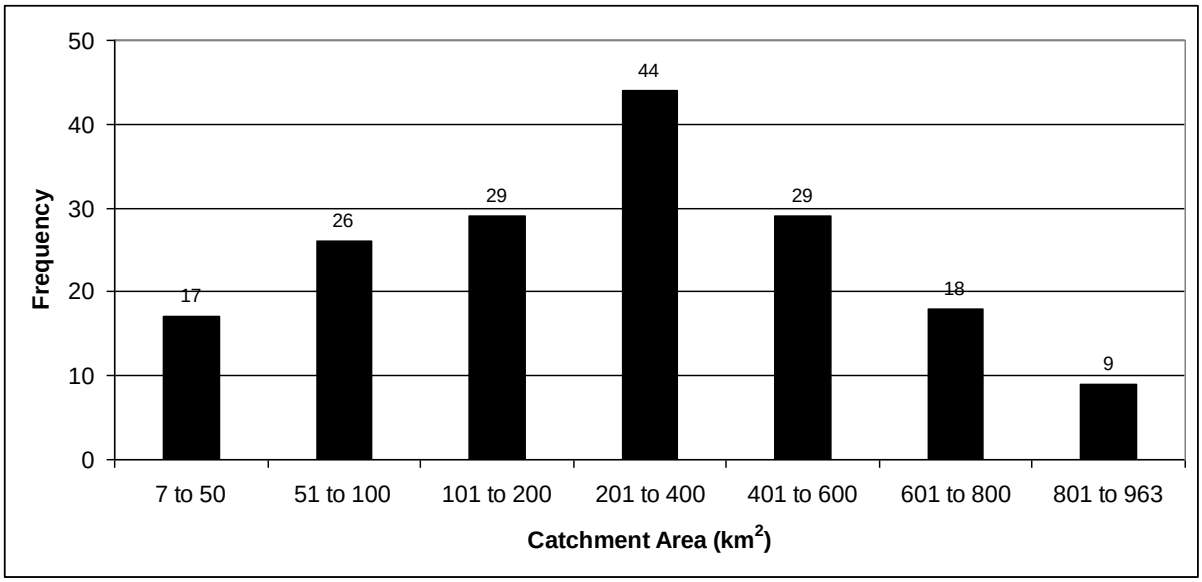


Figure 2.15 Distribution of catchment areas of 172 stations from Queensland

2.7 Data for Western Australia

A total of 146 catchments have been selected from Western Australia (listed in Appendix Table A6).

The record lengths of annual maximum flood series of these 146 stations range from 20 to 57 years (mean: 31 years, median: 30 years and standard deviation: 8.02 years). The distribution of record lengths is shown in Figure 2.16.

The catchment areas of the selected 146 catchments range from 0.1 km^2 to 7405.7 km^2 (mean: 323 km^2 and median: 60 km^2). The geographical distribution of the selected 146 catchments is shown in Figure 2.17. The distribution of catchment areas of these stations is shown in Figure 2.18.

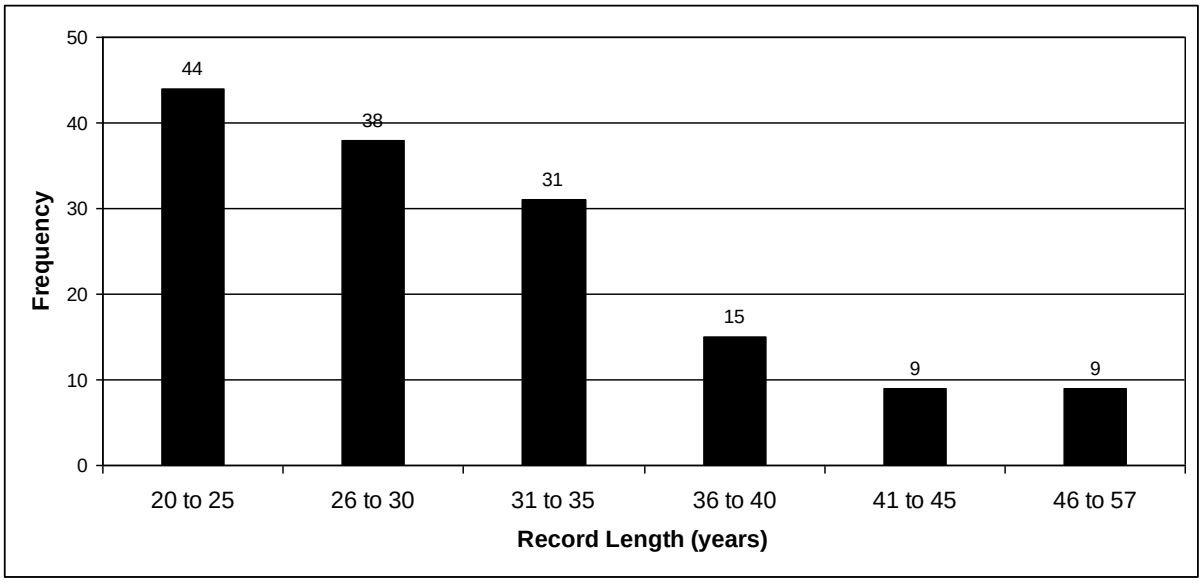


Figure 2.16 Distribution of streamflow record lengths of 146 stations from Western Australia

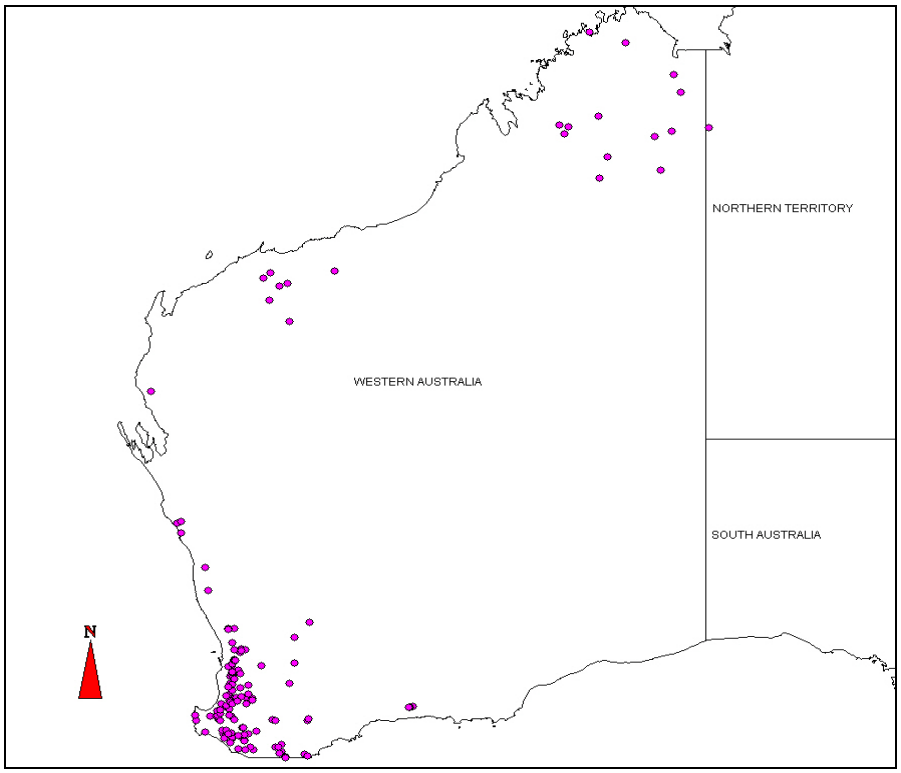


Figure 2.17 Geographical distributions of the selected 146 stations from Western Australia

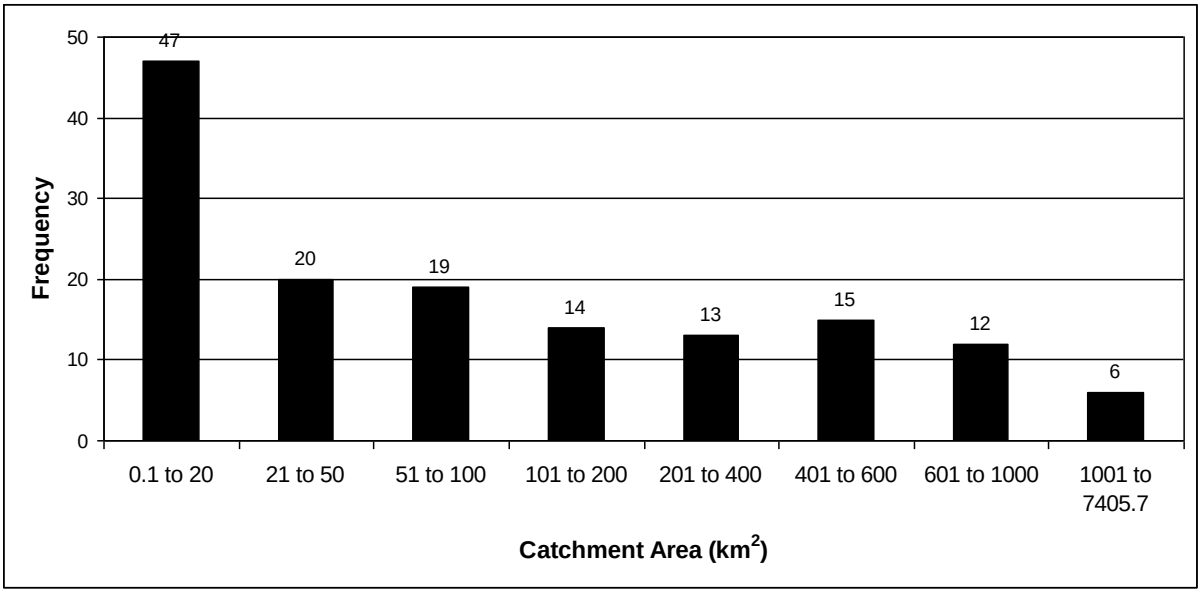


Figure 2.18 Distribution of catchment areas of 146 stations from Western Australia

2.8 Data for Northern Territory

A total of 55 catchments have been selected from Northern Territory (listed in Appendix Table A7).

The record lengths of annual maximum flood series of these 55 stations range from 19 to 54 years (mean: 35 years, median: 33 years and standard deviation: 11.30 years). The distribution of record lengths is shown in Figure 2.19.

The catchment areas of the selected 55 catchments range from 1.4 km^2 to 4325 km^2 (mean: 682 km^2 and median: 360 km^2). The geographical distribution of the selected 55 catchments is shown in Figure 2.20. The distribution of catchment areas of these stations is shown in Figure 2.21.

This data set does not include 14 stations in the arid/semi-arid part of the Northern Territory (see Section 2.9).

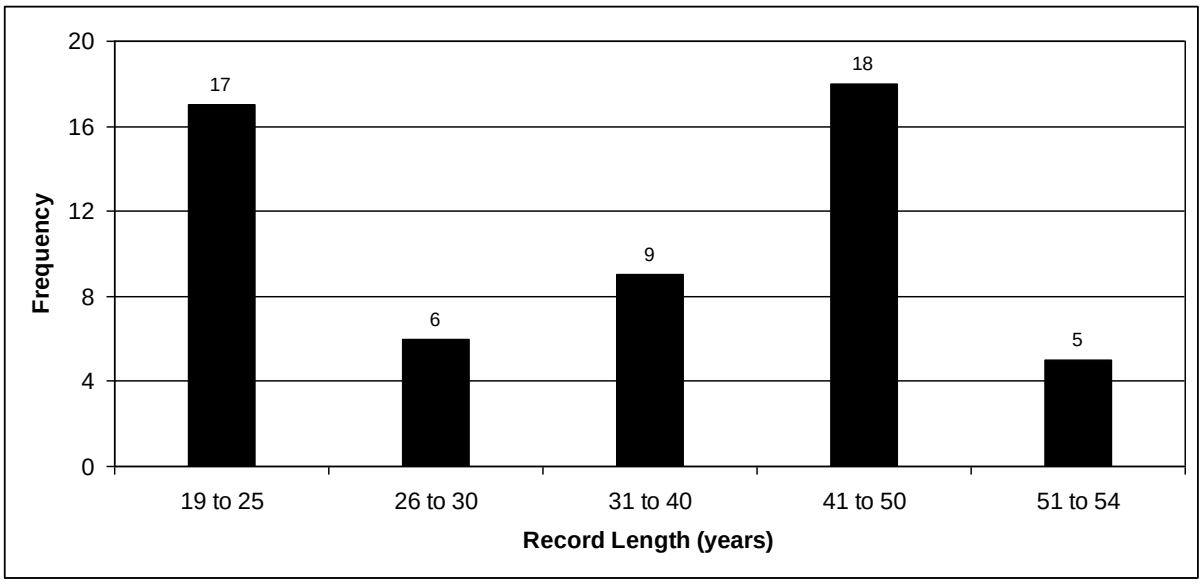


Figure 2.19 Distribution of streamflow record lengths of 55 stations from Northern Territory

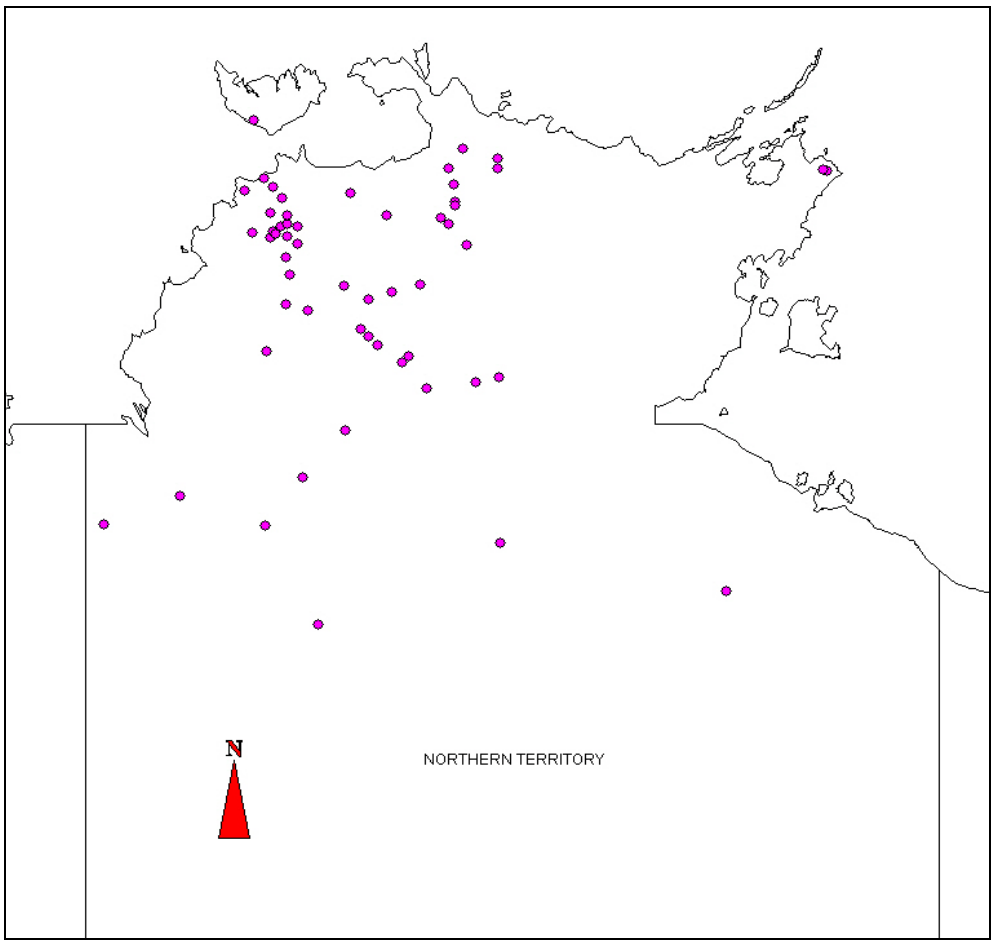


Figure 2.20 Geographical distributions of the selected 55 stations from Northern Territory

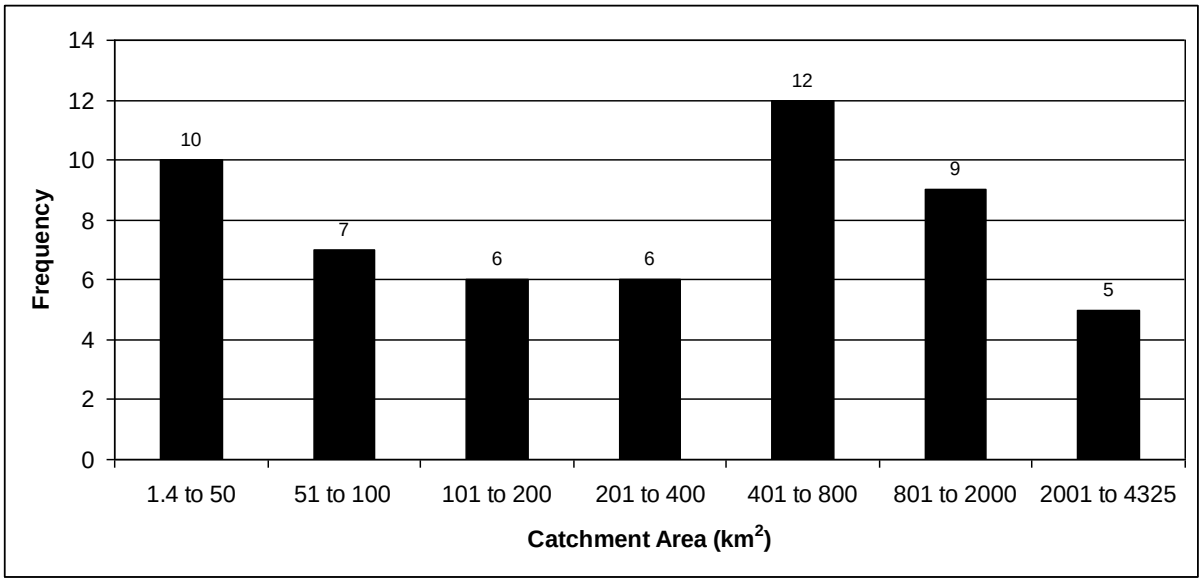


Figure 2.21 Distribution of catchment areas of 55 stations from Northern Territory

2.9 Data for arid and semi-arid regions

A total of 45 catchments have been selected from all over Australia for arid and semi-arid region (listed in Appendix Table A8).

The record lengths of annual maximum flood series of these 45 stations range from 10 to 46 years (mean: 25 years, median: 22 years and standard deviation: 10.0 years). The distribution of record lengths is shown in Figure 2.22.

The catchment areas of the selected 45 catchments range from 3.8 km^2 to 5975 km^2 (mean: 1152 km^2, median: 360 km^2). The geographical distribution of the selected 45 catchments is shown in Figure 2.23. The distribution of catchment areas of these stations is shown in Figure 2.24.

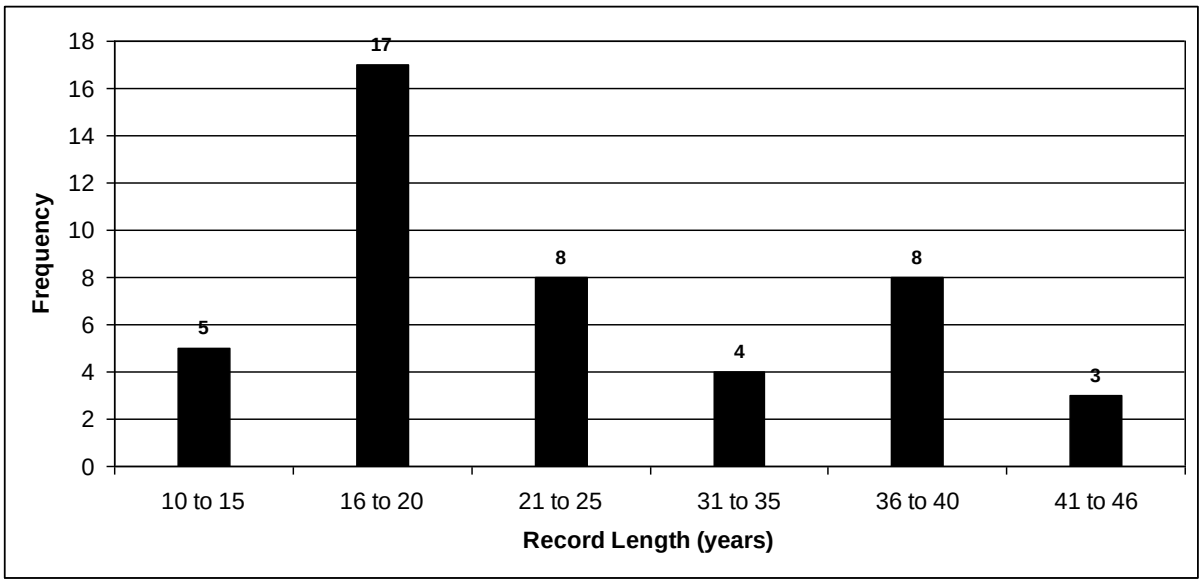


Figure 2.22 Distribution of streamflow record lengths of 45 stations from all over Australia for arid semi-arid regions

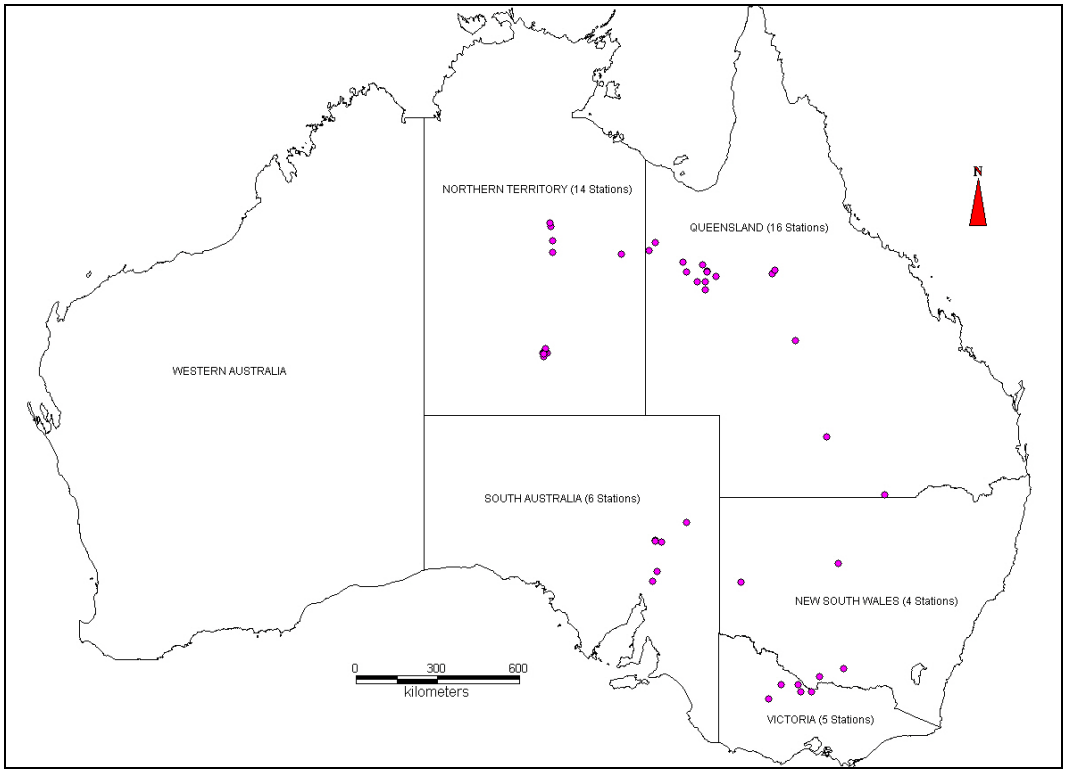


Figure 2.23 Geographical distributions of the selected 45 stations from all over Australia for arid semi-arid regions

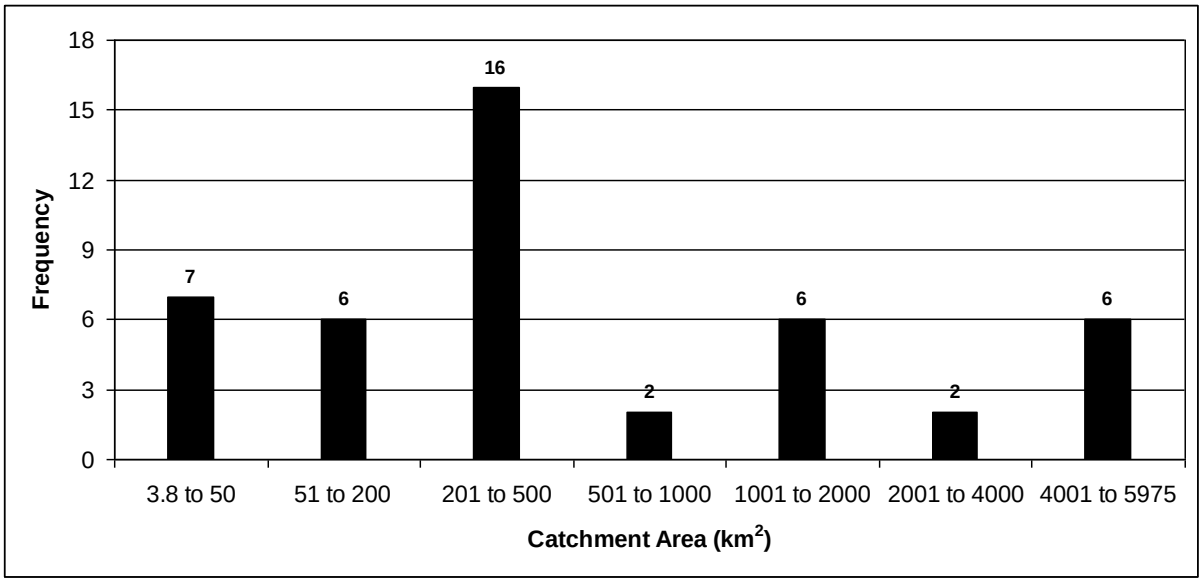


Figure 2.24 Distribution of catchment areas of 45 stations from all over Australia for arid semi-arid regions

2.10 All Australia (without arid and semi-arid database)

A total of 682 catchments have been selected from all over Australia.

The record lengths of the annual maximum flood series of these 682 stations range from 18 to 97 years (mean: 35 years, median: 33 years and standard deviation: 11.5 years). The distribution of record lengths is shown in Figure 2.25.

The catchment areas of the selected 682 catchments range from 0.1 km^2 to 7405.7 km^2 (mean: 350 km^2, median: 214 km^2). The geographical distribution of the selected 682 catchments is shown in Figure 2.26. The distribution of catchment areas of these stations is shown in Figure 2.27.

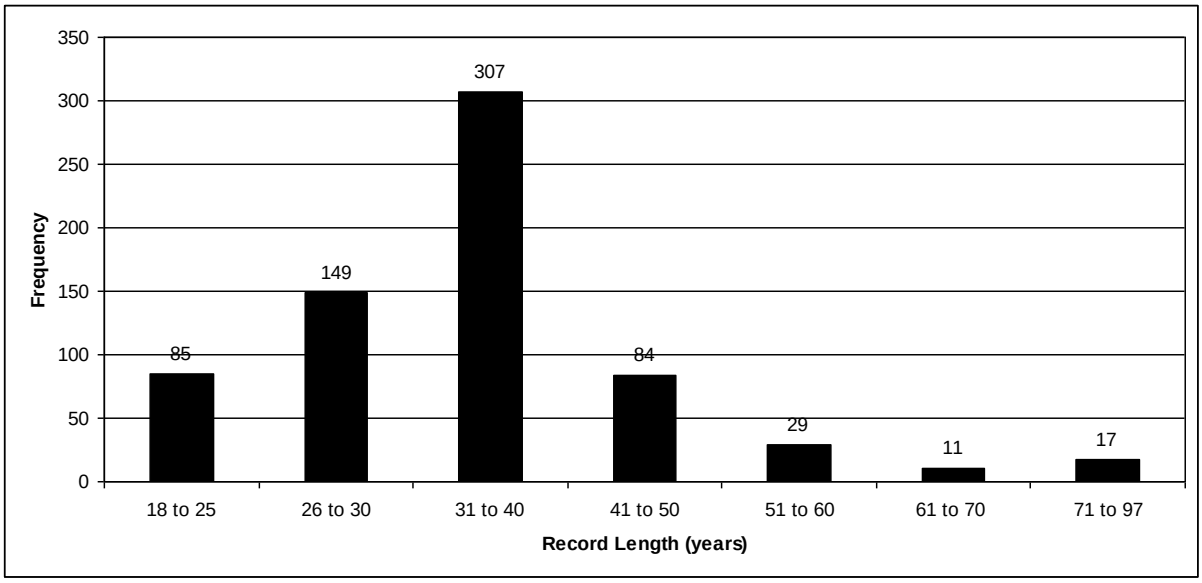


Figure 2.25 Distribution of streamflow record lengths of 682 stations from all over Australia

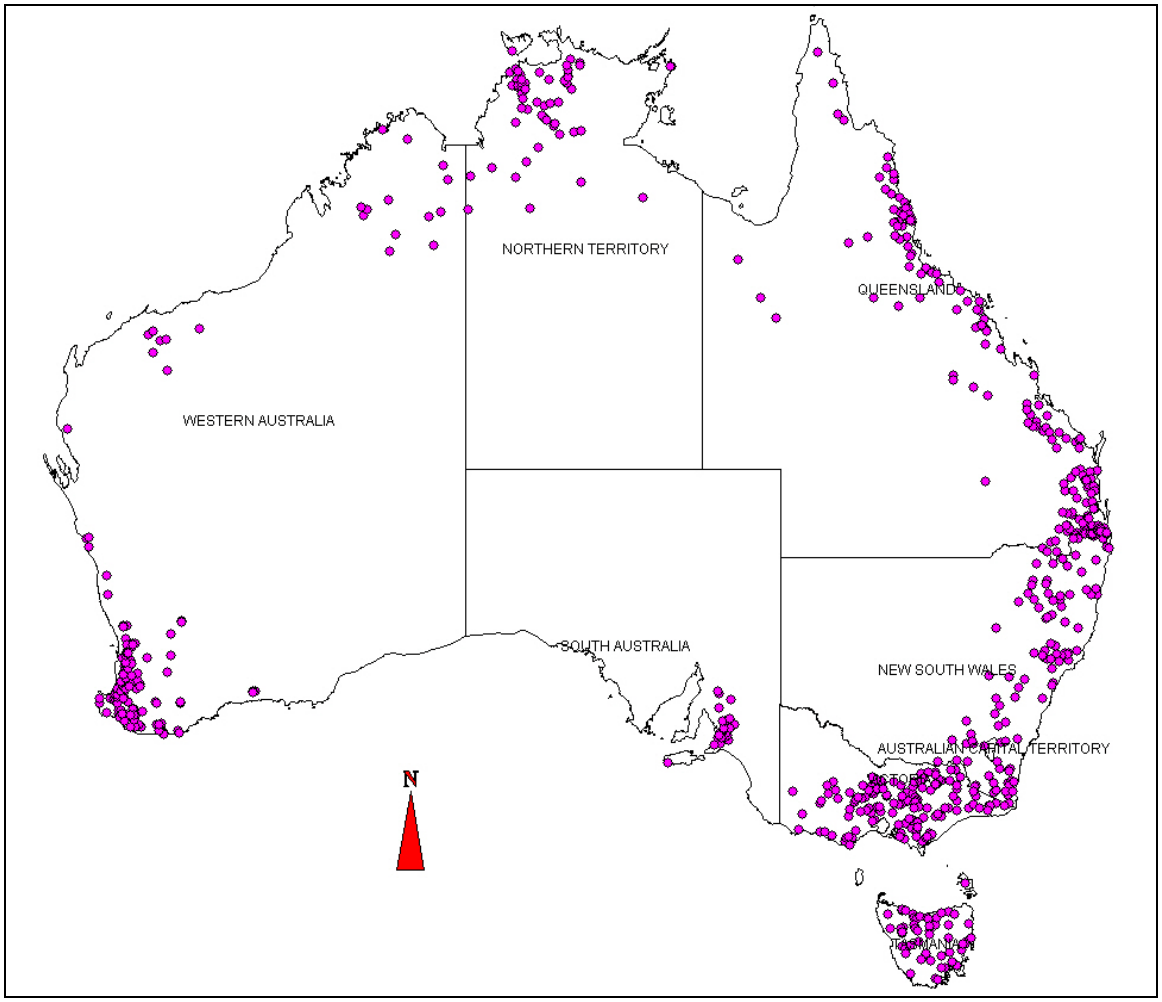


Figure 2.26 Geographical distributions of the selected 682 stations from all over Australia

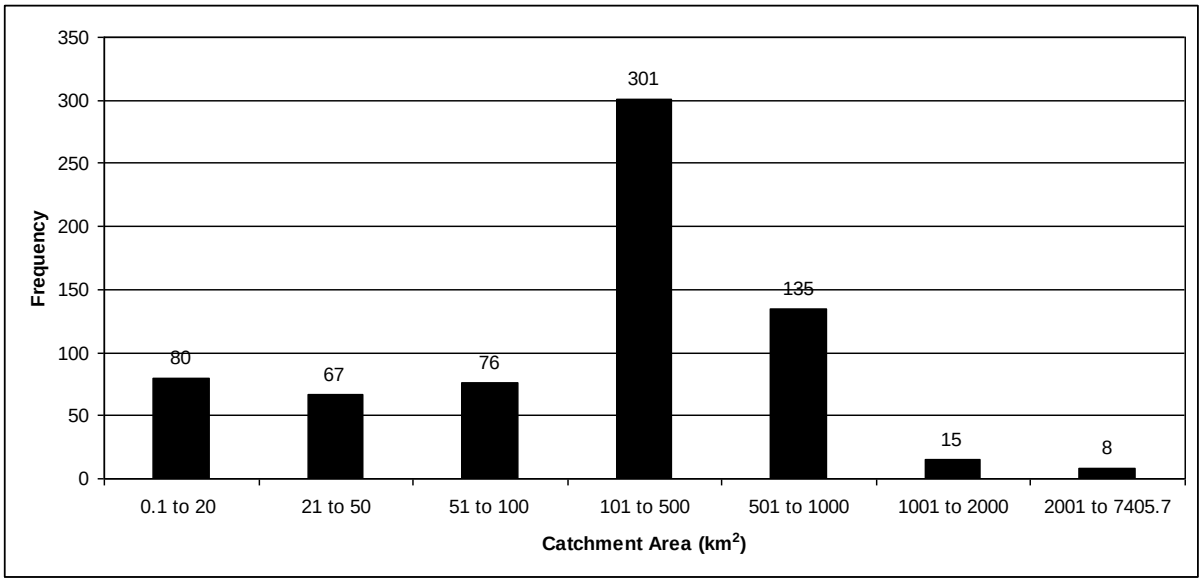


Figure 2.27 Distribution of catchment areas of 682 stations from all over Australia

2.11 Summary of all Australian data

The summary of all the Australian data prepared as a part of Project 5 is provided in Table 2.1

Table 2.1 Summary of selected stations Australia wide

State	No. of stations	Median streamflow record length (years)	Median catchment size (km^2)
NSW & ACT	96	34	267
Victoria	131	33	289
South Australia	29	34	76.5
Tasmania	53	28	158
Queensland	172	36	254
Western Australia	146	30	60
Northern Territory	55	33	360
Sub Total	**682**	-	-
Arid semi-arid region	45	22	360
TOTAL	**727**	-	-

2.12 Archiving of the data

The list of selected catchments, annual maximum flood series data, estimated flood quantiles and abstracted catchment characteristics data of all the states have been saved in a CD and archived. The selected catchment characteristics are provided in Stage I report (Rahman et al., 2009)

3. Overview and comparison of regional flood frequency analysis (RFFA) methods

3.1 General

Estimation of peak flows on small to medium sized rural catchments is required for the design of culverts, small to medium sized bridges, causeways and soil conservation works and many other water resources management tasks (Pilgrim, 1987). Typically, these catchments are ungauged. In such cases, peak flow estimates are obtained using a regional flood frequency analysis (RFFA) which transfers information from gauged catchments to the catchment under consideration. RFFA techniques are preferred in situations where catchment rainfall-runoff modelling is unwarranted due to a full streamflow hydrograph not being required for resolution of the design flood problem. A RFFA technique is expected to be simple, requiring readily obtainable input data to obtain design flood estimates relatively quickly.

This chapter provides a brief description of various RFFA methods and results of the comparison of some of the most commonly adopted RFFA methods.

3.2 Classification of RFFA methods

There are many RFFA methods in the literatures ranging from simple Rational Method (Mulvany, 1851) to non-linear models like Artificial Neural Network (ANN). The selection of a RFFA method for general application depends on factors such as:

1) Quantity and quality of temporal and spatial data availability: A RFFA method is ‘as good as’ the quantity and quality of the observed streamflow and other data. There is little merit in developing a highly complex RFFA model when there is limited data availability.
2) Ease of application: The recommended RFFA methods should be ‘easy to apply’ by the end-users. To facilitate this, simple maps, graphs, equations and/or computer models should be made available.
3) Simplicity of the model: Recommended RFFA methods/prediction equations should contain ‘easy to obtain’ predictor variables. The number of predictor variables should be as few as possible, and these variables should have plausible physical significance in terms of regional floods.
4) Robustness: The recommended RFFA methods should be unbiased and robust.

5) Measure of uncertainty: the recommended RFFA methods should provide a measure of uncertainty with the estimated flood quantiles.

All RFFA methods use the results of at-site FFA as basic data. A RFFA method then essentially consists of two principal steps: (i) Formation of regions: This involves formation of regions from the available streamflow gauging stations. (ii) Development of regional estimation models: This involves development of prediction equations to estimate flood quantiles, based on the results of at-site FFA within the region. Various RFFA methods are briefly described below.

3.3 Formation of regions

In RFFA, formation of regions can be based on proximity in geographic or catchment attributes space. A region can be fixed, having a definite boundary or it can be formed in geographic or catchment attributes space with respect to the ungauged catchment of interest (i.e. where flood quantile estimation is desired). Various methods of the formation of regions in RFFA are illustrated in Figure 3.3.1. The allocation of an ungauged catchment to regions formed in catchment attributes space is often problematic. Acreman and Wiltshire (1987) proposed regions without fixed boundaries. Subsequently, Burn (1990a, 1990b) and Zrinji and Burn (1994) proposed the region-of-influence (ROI) approach where each site of interest (i.e. catchment where flood quantiles are to be estimated) can form its own region.

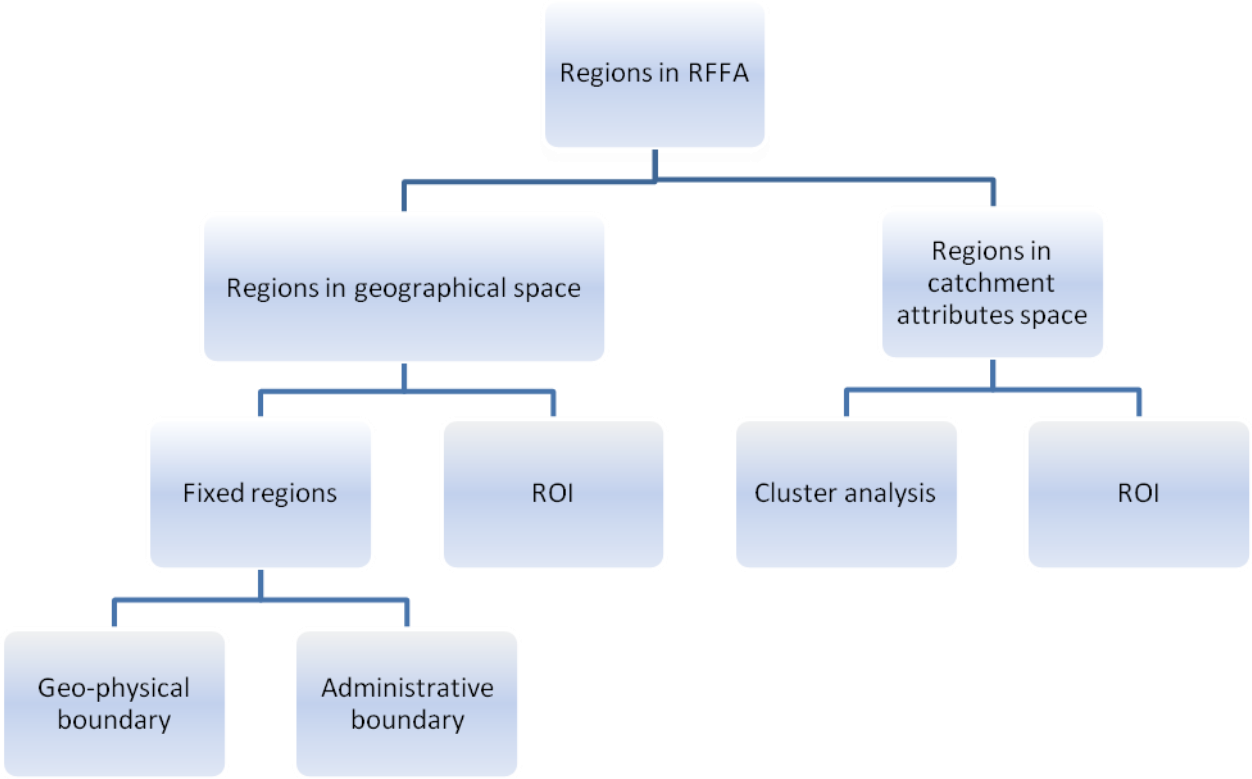


Figure 3.3.1 Methods of formation of regions in RFFA

3.4 Development of regional estimation models

3.4.1 Probabilistic Rational Method (PRM)

The Rational Method, introduced by Mulvany (1851), has been widely regarded as a deterministic method for estimating the peak discharge from an individual storm. However in ARR1987 (IEAust, 1987), it was presented as a probabilistic method (referred to as Probabilistic Rational Method (PRM)). This means that the runoff coefficient becomes a simple transfer function which converts the design rainfall of given ARI to a design peak flow of the same ARI.

The Rational Method has often been recommended for application to only small catchments below some arbitrary limit such as 25 km^2. This limited range of applicability reflects the inadequate manner in which the method considers physical factors such as the effects of temporary storage on the catchment, and temporal and spatial variations of rainfall intensity. These physical considerations have little relevance to the probabilistic interpretation of the PRM, where their effects are incorporated in the recorded floods, and hence in the flood frequency statistics and the derived values of the runoff coefficient C_Y. As mentioned in ARR1987, the PRM derived from observed data should be valid for catchment areas and ARIs up to and somewhat beyond the maximum areas and record lengths used in their derivation (I. E. Aust., 1987).

In ARR1987, the PRM is represented by:

$$Q_Y = 0.278 C_Y A I_{tc,Y} \qquad (3.4.1)$$

where Q_Y is the peak flow rate (m^3/s) for an ARI of *Y* years; C_Y is the runoff coefficient (dimensionless) for ARI of *Y* years; $I_{tc,Y}$ is the average rainfall intensity (mm/h) for a time of concentration t_c (hours) and ARI of *Y* years; and *A* is the catchment area (km^2).

From Equation 3.4.1, the value of the runoff coefficient is given by:

$$C_Y = \frac{Q_Y}{0.278 I_{tc,Y} A} \qquad (3.4.2)$$

The values of Q_Y for a station can be obtained from at-site flood frequency analysis, subject to the availability of reasonably long streamflow records. Values for $I_{tc,Y}$ at a given location can be found from Book II Section 1 of ARR. The catchment and rainfall characteristics and conditions affecting the relation between Q_Y, *A* and I_Y are incorporated in C_Y, but not necessarily in a physically realistic fashion.

In the deterministic interpretation of the Rational Method, the critical rainfall duration is t_c, which is considered to be the travel time from the most remote point on the catchment to the outlet, or the time taken from the start of rainfall until all of the catchment is simultaneously contributing flow to the outlet. For the probabilistic interpretation of the Rational Method, as in the PRM, these physical measures are not really relevant. However, Equation 3.4.2 shows that the value of C_Y depends on the duration of rainfall, and some design duration related to catchment characteristics must be specified as part of the overall procedure. A typical response time of flood runoff is appropriate, and the 'time of concentration' is a convenient measure. In this context, its accuracy regarding travel time is much less important than the consistency and reproducibility of derived C_Y values. Also, values of C_Y cannot be compared unless consistent estimates of t_c are used in their derivation. Pegram (2002) and French (2002) discussed various methods of estimating t_c as well as the strengths and weaknesses of the PRM.

One commonly adopted equation to estimate t_c is:

$$t_c = 0.76A^{0.38} \qquad (3.4.3)$$

where t_c is the time of concentration (hours) and *A* is the catchment area (km^2).

An alternative approach is to use the Bransby William formula, as given below:

$$t_c = \frac{58L}{A^{0.1}S_e^{0.2}} \qquad (3.4.4)$$

where t_c is in minutes; *L* is the mainstream length measured to the catchment divide (km); *A* is the catchment area (km^2) and S_e is the equal area slope of the main stream projected to the catchment divide (m/km). This is the slope of a straight line drawn on a profile of a stream such that the line passes through the outlet and has the same area below and above the stream profile.

Equation 3.4.3 was adopted with the PRM for eastern NSW and Victoria in ARR1987. This was also adopted by Weeks (1991) in an attempt to develop the PRM for Queensland. In this current study, Equation 3.3 is used as it is easier to apply. In the development of the PRM for a region, the C_{10} value for each individual catchment is estimated using Equation 3.2, and these 'site' values are then 'regionalised' by plotting a C_{10} contour map. A frequency factor (FF_Y) is used to convert Q_{10} to Q_Y. The value of FF_Y is estimated for each of the model catchments using Equations 3.2 and 3.5; the average or median FF_Y value is then used in the design.

$$FF_Y = \frac{C_Y}{C_{10}} \qquad (3.4.5)$$

3.4.2 Quantile Regression Technique (QRT)

United States Geological Survey (USGS) proposed a quantile regression technique (QRT) where a large number of gauged catchments are selected from a region and flood quantiles are estimated from recorded streamflow data, which are then regressed against catchment characteristics variables that are most likely to govern the flood generation process. Studies by Benson (1962) suggested that *Y*-year flood peak discharges could be estimated directly using catchment characteristics data by multiple regression analysis. The quantile regression technique can be expressed as follows:

$$Q_Y = aB^bC^cD^d \ldots \qquad (3.4.6)$$

where *B, C, D, …* are catchment characteristics variables and Q_Y is the flood magnitude with *Y*-year ARI (flood quantile), and *a, b, c, …* are regression coefficients.

There have been various techniques and many applications of regression models that have been adopted for hydrological regression. The USGS has been applying the QRT for several decades. A well known study using the QRT with an Ordinary Least Squares (OLS)

procedure was carried out by Thomas and Benson (1970). The study tested four regions in the United States for design flood estimation using multiple regression techniques that related streamflow characteristics to catchment characteristics. This study found that the QRT was able to predict quantile estimates more accurately as compared to previous methods adopted by the USGS. However, the study noted that the equations were not statistically sound.

The OLS estimator has traditionally been used by hydrologists to estimate the regression coefficients in Equation 3.4.6. But in order for the OLS model to be statistically efficient and robust, the annual maximum flood series in the region must be uncorrelated, all the sites in the region should have equal record length and all estimates of *Y*-year events have equal variance. Since the annual maximum flow data in a region do not generally satisfy these criteria, the assumption that the model residual errors in OLS are homoscedastic is violated and the OLS approach can provide distorted estimates of the model's predictive precision (model error) and the precision with which the regression coefficients are being estimated (Stedinger and Tasker, 1985).

To overcome the above problems in OLS, Stedinger and Tasker (1985) proposed the Generalised Least Squares (GLS) procedure which can result in remarkable improvements in the precision with which the parameters of regional hydrologic regression models can be estimated, in particular when the record length varies widely from site to site. In the GLS model, the assumptions of equal variance of the *Y*-year events and zero cross-correlation for quantiles are relaxed. The Bayesian GLS (BGLS) regression offers additional advantages by providing a realistic description of the possible values of the model error variance, especially in the case where sampling error tends to dominate the model errors in the regional analysis (Reis et al., 2005).

3.4.3 Parameter Regression Technique (PRT)

In the parameter regression technique (PRT), the parameters of a particular probability distribution are regressed against the catchment characteristics similar to QRT. Here, both the OLS and GLS methods (including BGLS) can be used to develop the prediction equations for the mean, standard deviation and skewness of the annual maximum flood series. These equations are then used to predict the mean, standard deviation and skewness of the annual maximum flood series for an ungauged catchment to fit a particular probability

distribution. This fitted probability distribution is then used to estimate the flood quantiles for the ungauged catchment.

3.4.4 Index Flood Method

The key assumption in the index flood method is that the distribution of floods at different sites within a homogeneous region is the same except for a site-specific scale or index flood factor. Homogeneity with regard to the index flood relies on the concept that the standardised flood peaks from individual sites in the region follow a common probability distribution with identical parameter values. From all the method examined in this project, the Index Flood Method involves the strongest assumptions on homogeneity. The method is used in many countries.

ARR1987 (I.E Aust., 1987; 2001) did not favour the index flood method as a design flood estimation technique. The index flood method had been criticised on the grounds that the coefficient of variation of the flood series C_v may vary approximately inversely with catchment area, thus resulting in flatter flood frequency curves for larger catchments. This had particularly been noticed in the case of humid catchments that differed greatly in size (Dawdy, 1961; Benson, 1962; Riggs, 1973; Smith, 1992).

There have been recent studies carried out by Bates et al. (1998) and Rahman et al. (1999) where the development of an application for design flood estimation in ungauged catchments in south-east Australia was tested using the index flood method. The method involved the assignment of ungauged catchments to a particular homogenous group identified (through the use of *L*-moments) on the basis of catchment characteristics as opposed to geographical proximity. The relationships sought were developed by statistical procedures such as canonical correlation analysis, tree based modelling and other multivariate statistical techniques. This allowed for the development of a RFFA method using up to 12 independent catchment characteristics variables.

Although the results of this method showed promise when compared to the PRM, its limitations were already evident in that it needed a large number of independent variables which are very time consuming to obtain. The results of this method also depend upon the correct assignment of an ungauged catchment to a homogenous group, thus any wrong assignment would greatly increase error in quantile estimation.

3.4.5 Probabilistic Model (PM)/ Large Flood Regionalisation Model (LFRM)

The Probabilistic Model presented by Majone and Tomirotti (2004) assumes that the maximum observed floods Q_{max} from the annual flood series of each of the sites in a region (after standardisation by the at-site average flood and a function of the coefficient of variation of annual flood series) can be pooled (similar to the station-year approach) and assumed to follow a single probability distribution. That is, the standardised Q_{max} across various sites form a homogeneous region. This is similar to the assumption of the index flood method but, by allowing for differences in the standard deviation of annual floods, it overcomes a major weakness of the index flood method. The Probabilistic Model is also referred to as Large Flood Regionalisation Model (LFRM) in this report.

The main focus of the PM/ LFRM is the prediction of flood quantiles of higher ARIs. To apply the PM/ LFRM to ungauged catchments, one needs to develop prediction equations for the mean and coefficient of variation of the annual flood series. Majone et al. (2007) applied the PM to flood data from 8500 gauging stations across the world and found that the method can provide quite reasonable design flood estimates for higher ARIs.

The method has been applied for the NSW and Victoria data set and has shown promising results (Haddad, Rahman and Weinmann, 2011).

3.4.6 Summary of the classification of RFFA methods

Based on the discussion presented in this chapter, various methods of the formation of regions and development of the regional estimation models are illustrated in Figure 3.4.1.

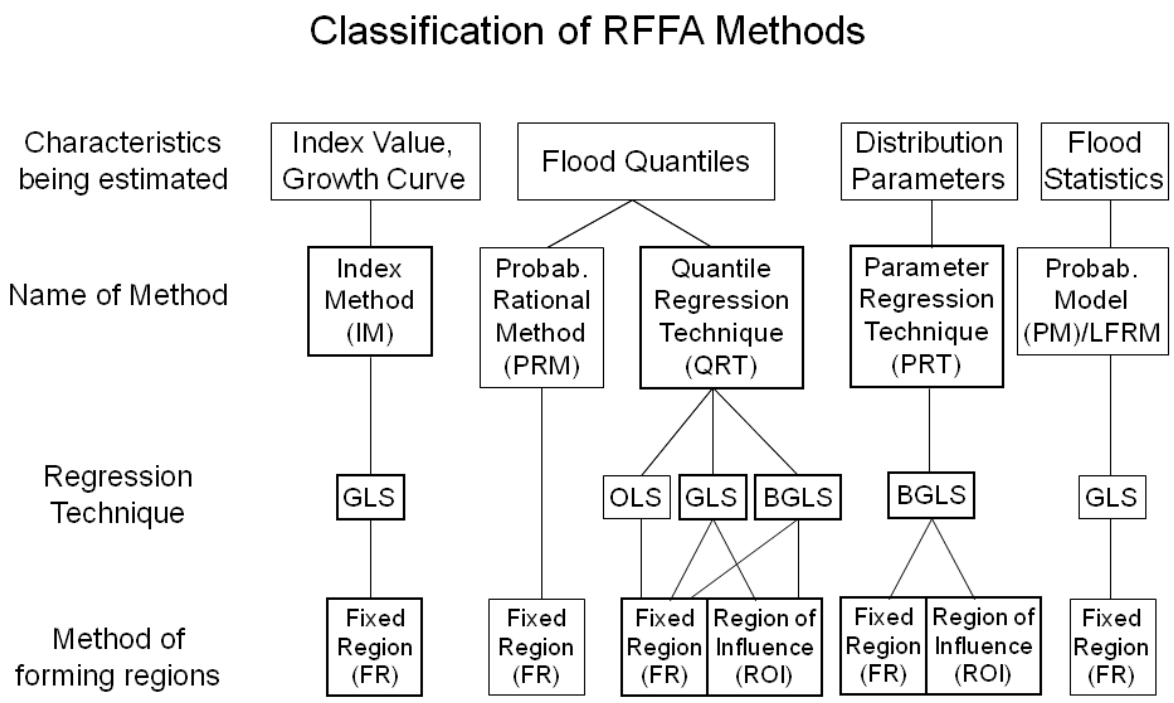


Figure 3.4.1 Various RFFA methods including formation of regions and development of estimation models

3.5 Comparison of the commonly adopted RFFA methods

Various RFFA methods were compared in Stage I of Project 5 (Rahman et al., 2009). These included a comparison of the PRM and QRT, OLS and GLS regression and fixed region and ROI and application of PM/LFRM. The results of these have been summarised in a number of refereed papers (e.g. Haddad et al., 2006, 2008, 2009a,b, 2010a,b, 2011; Haddad and Rahman, 2008, 2010, 2011a, b, c; Rahman et al., 2008, 2010, 2011; Hackelbusch et al., 2009; Pirozzi et al., 2009; Pirozzi and Rahman, 2010; Zaman et al., 2010; Ishak et al., 2009, 2011; Aziz et al., 2010).

A further comparison of the PRM and QRT is presented the Section 3.5.1. The development and testing of the BGLS-QRT and BGLS-PRT are presented in Chapter 4.

3.5.1 Comparison of PRM and QRT

Differences and similarity between PRM and QRT

The PRM and QRT have been described in Sections 3.4.1 and 3.4.2. To allow a more direct comparison of the performance of the PRM and QRT, a special form of the QRT has been adopted in this chapter, using the same explanatory variables as the PRM, as expressed by the following equation:

$$Q_Y = aA^b I_{tc,Y}{}^c \qquad (3.5.1)$$

In this study, regression coefficients *a, b* and *c* in Equation 3.5.1 are estimated using GLS regression.

Equation 3.5.1 reduces to Equation 3.4.1 for Y = 10 years under the following conditions: (i) there are two independent variables in both the equations: A and I_{tc_10}; ii) b = 1 and c = 1; and (iii) $a = 0.278 * C_{10}$. The PRM attempts to lump coefficients *a, b* and *c* in C_{10}, which allows mapping of C_{10} in one single contour map. The QRT does not require maps of coefficients as the set of coefficients determined from regression applies over the whole region under consideration.

There may be advantage in lumping the effects of *a*, *b* and *c* in a single coefficient like C_{10} but issues arise in how to regionalise the C_{10} values to allow the determination of the C_{10}

value for a given ungauged catchment. Mapping and simple interpolation are based on the assumption that the flood producing characteristics vary in a continuous fashion over the region, which appears to be too simplistic, as it neglects the obvious discontinuities at catchment boundaries. The use of simple interpolation routines to generalise the C_{10} values from gauged catchments to the whole region also assumes that the C_{10} value obtained for each gauged catchment is error free. This limitation could be overcome by using a surface fitting technique for the mapping of C_{10} values, with allowance for estimated errors in the C_{10} site values.

It may be argued that QRT is superior to PRM as it develops prediction equation for each ARI of interest, thus the use of FF_Y is not required. In principle, use of FF_Y in the PRM could be avoided by preparing separate maps of runoff coefficients for various ARIs e.g. C_2, C_5, C_{10}, C_{20}, C_{50} and C_{100}. The FF_Y is similar to the regional growth factors used in the index flood approach, i.e. the same frequency factor is applicable for all the stations in the region. However, this assumption of regional homogeneity with FF_Y has never been tested.

In this special application of the QRT (Equation 3.5.1), all the variation in flood quantile estimates for different catchments in the region is explained by differences in A and I_{tc_Y}. However, a significant advantage of the QRT is that it can include additional catchment variables (other than A and I_{tc_Y}) in the regression equation without much difficulty. The influence of other flood producing factors (which may be reflected in the runoff coefficient of the PRM) can be allowed for without any assumption of geographic contiguity of such influences.

Uncertainty analysis for the PRM is quite difficult to undertake as the errors in the interpolation between the two nearest points on the C_{10} curves and in the curves themselves are difficult to quantify. In contrast, the coefficients *a, b* and *c* in QRT can be estimated using methods such as GLS regression, which accounts for variation in record lengths from site to site and inter-station correlation, and a rigorous uncertainty analysis can be undertaken for the sampling and model error. The PRM needs a greater spatial coverage of the available stations to increase the density of contour curves; for QRT it would be enough to have a reasonable number of stations covering the expected variability in the independent variables and a sample size which is large enough (say about 40 catchments) to estimate regression coefficients *a*, *b* and *c* (and possibly coefficients associated with additional explanatory variables) with sufficient accuracy. Furthermore, by integrating the QRT with the region-of-influence approach, an appropriate region size for the ungauged catchment of interest can be established on the criterion of minimum model error variance. This is likely to result in a

more accurate set of estimated regression coefficients for the ungauged catchment of interest.

Description of data used in the comparison of PRM and QRT

The study uses data from 107 catchments from NSW (Figure 3.5.1). These catchments are not affected by any major regulation and have not undergone major land use changes during the period of streamflow data availability. The initial data set consisted of stations with at least 25 years of data. This however did not provide sufficient spatial coverage at a few locations in the study area to be able to develop a C_{10} contour map with reasonable resolution; some additional stations were therefore selected with streamflow record lengths slightly smaller than 25 years to fill gaps in spatial coverage. The overall streamflow record lengths of the selected stations range from 18 to 74 years with mean and median of 32 and 30 years, respectively. The distribution of annual maximum flood record lengths of the selected 107 stations is shown in Figure 3.5.2.

The catchment areas of the selected catchments range from 8 to 1010 km^2 with mean and median values of 325 km^2 and 236 km^2, respectively. The distribution of catchment areas is shown in Figure 3.5.3. It should be noted that there are only 2 and 6 catchments smaller than 10 km^2 and 20 km^2, respectively. Thus, the application of the developed PRM and QRT to catchments smaller than 8 km^2 needs special consideration, which is discussed in Chapter 5.

The streamflow data for these catchments were prepared by following the stringent criteria detailed in Rahman et al. (2009) and Haddad et al. (2010). As far as possible, gaps in the annual maximum flood series were infilled and outliers in the data series were detected. To limit the error in flood frequency analysis arising from rating curve extrapolation, an empirical procedure was adopted, as discussed in Haddad et al. (2010). This method is based on the ratio of the estimated flow and the maximum measured flow (rating ratio) at a gauging station. The stations with rating ratio values greater than 20 were excluded from the database. Also, the effects of rating curve error were accounted for in the at-site flood frequency analyses using the in-built facility in the FLIKE software (Kuczera, 1999). Stations with significant time trends in the annual maximum flood series were excluded from the analysis.

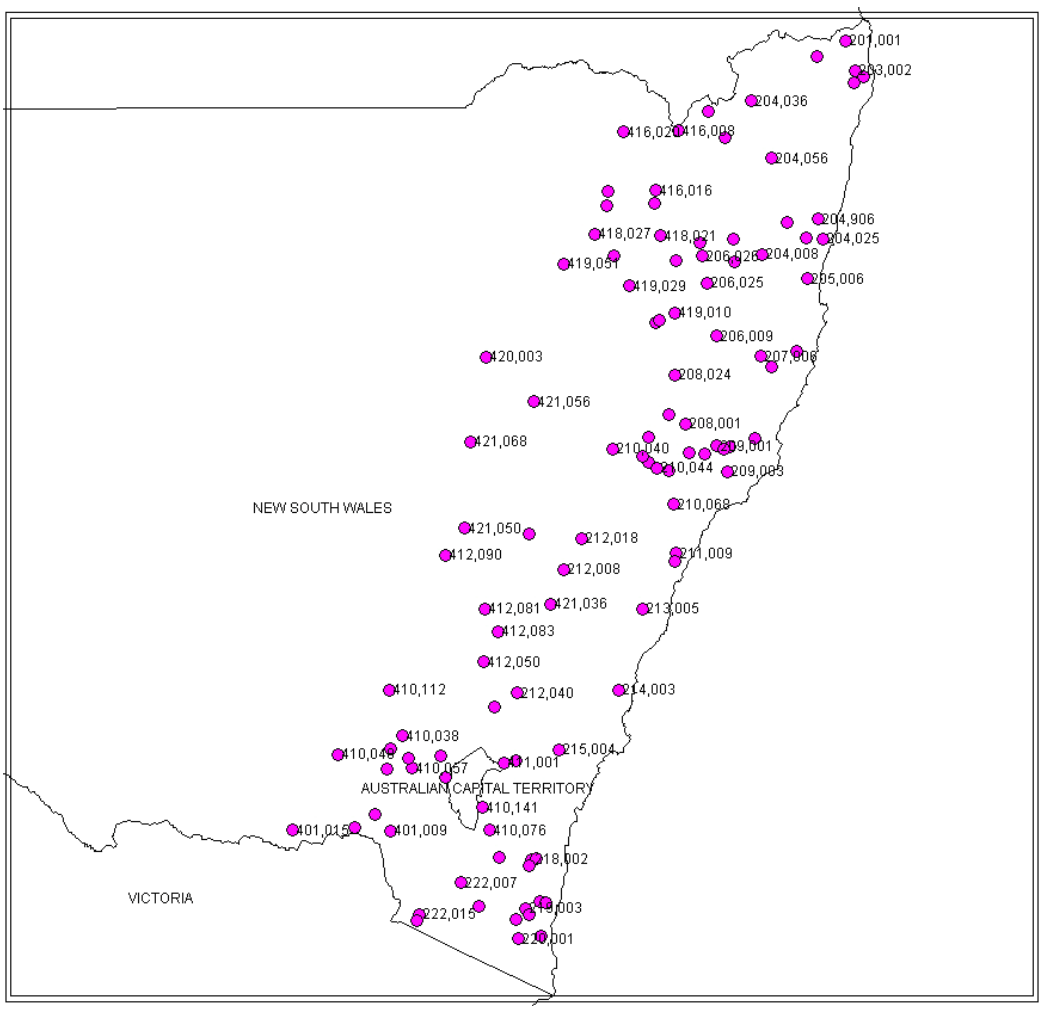


Figure 3.5.1 Locations of the selected 107 catchments from NSW

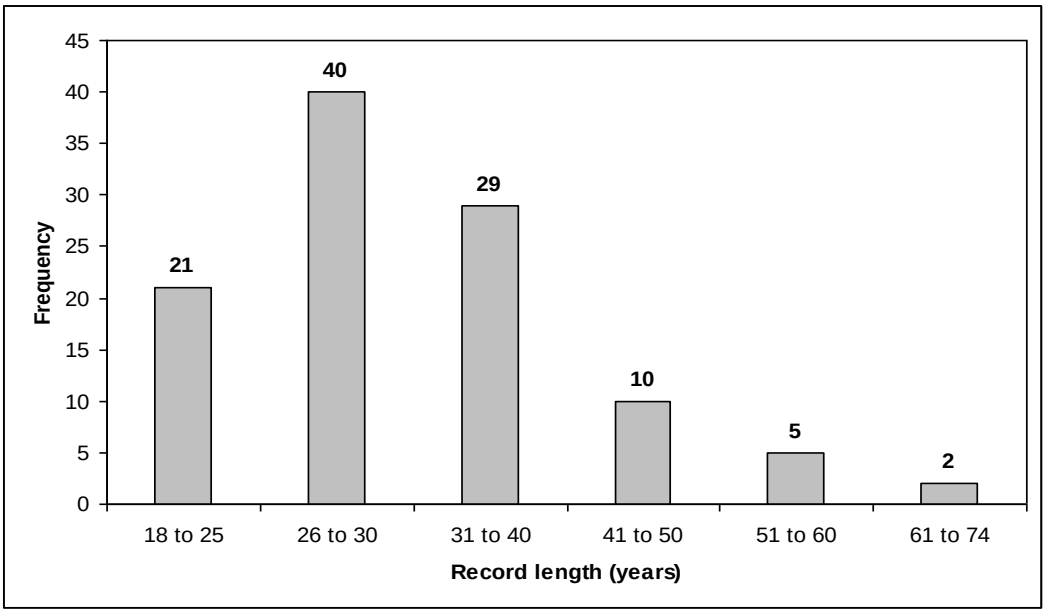


Figure 3.5.2 Distribution of streamflow record lengths

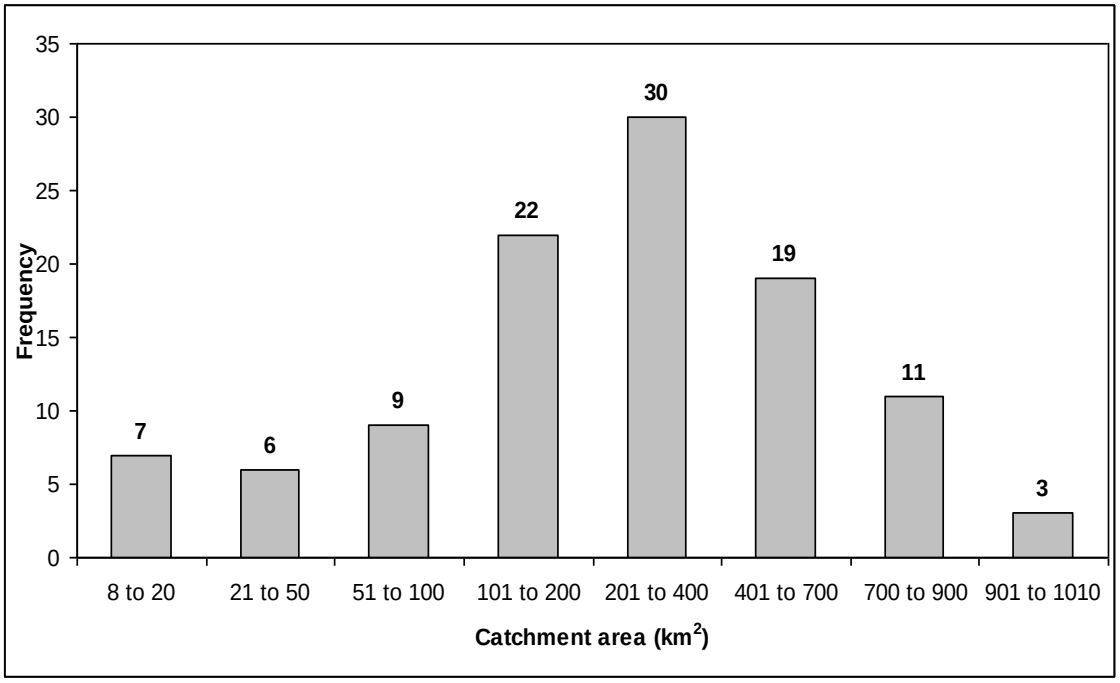


Figure 3.5.3 Distribution of catchment sizes.

Approach adopted in the comparison between PRM and QRT

At-site flood frequency estimates for ARIs of 2, 5, 10, 20, 50 and 100 years for each of the selected 107 stations were obtained by fitting a Log Pearson Type 3 (LP3) distribution using a Bayesian parameter estimation procedure as implemented in the FLIKE software (Kuczera, 1999). The LP3 distribution (with a Bayesian parameter estimation procedure) generally provided better fitting to the observed annual maximum flood series data, and hence adopted in this study.

A one-at-a-time validation approach was used for both the PRM and QRT, i.e. in an individual run, one catchment (from the 107) was left out for independent testing, leaving 106 model catchments. The procedure was repeated 107 times so that each of the catchments was selected once for independent testing. In this process, a total of 107 contour maps of C10 were produced; similarly 107 sets of prediction equations for Q_Y (Y = 2, 5, 10, 20, 50 and 100 years) were developed using GLS regression. In both the PRM and QRT, all the 106 model catchments in an individual run were assumed to form one region i.e. the region-of-influence approach was not used to form a separate region for each catchment with a reduced number of neighbouring stations.

In addition to the one-at-a-time validation approach, as discussed above, an alternative split-sample validation was conducted where 20% of the catchments (i.e. 21 catchments) were selected randomly as test catchments. The QRT and PRM were then developed based on the remaining 86 catchments (model catchments) and subsequently applied to the 21 test catchments.

For the PRM, C_{10} values were estimated for the model data set and a map of C_{10} was produced. The GIS program Mapinfo's Vertical Mapper add-on was then used to develop the C_{10} contour map. A spreadsheet containing the latitude, longitude and C_{10} values for each model catchment was produced and entered into the mapping program with the C_{10} values represented in the z axis. The program used a kriging method (Kottegoda and Rosso, 1997) to create a model of the C_{10} surface, from which isopleths were developed. The isopleths were labelled and the test catchment was located on the map. Linear interpolation was then used to estimate the C_{10} value for the test catchment from the contour map.

For the QRT, prediction equations were developed using the model data set for Q_Y (Y = 2, 5, 10, 20, 50 and 100 years) using GLS regression as described in Stedinger and Tasker (1985), Tasker and Stedinger (1989) and Haddad et al. (2006, 2008 and 2009).

To assess the relative performance of the PRM and QRT, a number of evaluation statistics were employed: root mean squared error (RMSE), relative error (RE), mean squared relative error (MSRE), mean percent relative error (MPRE), coefficient of efficiency (CE) and the ratio of predicted and observed values (ratio). These statistics are defined below:

$$RMSE = \sqrt{\frac{\sum\left(Q_{pred} - Q_{obs}\right)^2}{n}} \tag{3.5.2}$$

$$RE = \frac{Q_{pred} - Q_{obs}}{Q_{obs}} \times 100 \tag{3.5.3}$$

$$MSRE = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Q_{pred} - Q_{obs}}{Q_{obs}}\right)^2 \tag{3.5.4}$$

$$MPRE = \frac{100}{n}\sum_{i=1}^{n} abs\left(\frac{Q_{pred} - Q_{obs}}{Q_{obs}}\right) \tag{3.5.5}$$

$$CE = 1 - \frac{\sum_{i=1}^{n}\left(Q_{pred} - Q_{obs}\right)^2}{\sum_{i=1}^{n}\left(Q_{obs} - \overline{Q}\right)^2} \quad (3.5.6)$$

$$ratio = \frac{Q_{pred}}{Q_{obs}} \quad (3.5.7)$$

where Q_{obs} is the observed flood quantile obtained from at-site flood frequency analysis using FLIKE (Kuczera, 1999), Q_{pred} is the predicted flood quantile obtained by applying the PRM or QRT based on one-at-a-time or split-sample validation approach, $\overline{Q}$ is the mean of the Q_{obs} values for a given ARI and *n* is the number of catchments.

The RMSE, MSRE and MPRE provide an indication of the overall accuracy of a model. The CE provides an indication of how good a model is at predicting values away from the mean. The CE ranges from - ∞ in the worst case to +1 for a perfect model. The Q_{pred}/Q_{obs} *ratio* gives an indication of the degree of bias (i.e. systematic over- or under estimation), where a value of 1 indicates perfect agreement between the Q_{pred} and Q_{obs}. The Q_{pred}/Q_{obs} *ratio* values were counted based on a number of thresholds, e.g. 0.7 to 1.4 and 0.5 to 2. Here, *ratio* values smaller than 0.5 and greater than 2 may be used to identify cases showing 'gross under-estimation', and 'gross over-estimation', respectively. It should be mentioned here that these are only arbitrary limits and could be expected to provide a reasonable guide about the relative accuracy of the method as far as the practical application of the method is concerned.

Results of comparison between PRM and QRT

A typical C_{10} contour map is shown in Figure 3.5.4. The value of the runoff coefficients tends to decrease from east to west (similar to the C_{10} contour map in ARR1987). At many locations, a higher C_{10} value is surrounded by relatively much smaller values. These C_{10} estimates at individual locations are affected by the errors in Q_{10} and $I_{tc,10}$ estimates. Given the streamflow record lengths (18 to 74 years with a mean of 32 years) and the method of at-site flood frequency analysis employed in this study, the magnitude of error for Q_{10} is likely to be smaller than the error in the corresponding ARR1987 estimates, where stations with a minimum of 10 years of streamflow data were included). The error in $I_{tc,10}$ value can be further reduced in the near future by using the new design rainfall estimates; these will be

based on longer record lengths and superior statistical techniques and will be available in the near future as a part of the updated ARR. The frequency factors for each of the 107 catchments for a given ARI were obtained using Equation 3.4.5. The median frequency factor value for a given ARI was then computed and is shown in Table 3.5.1.

Table 3.5.1 Median of Frequency factors of different ARIs for PRM

ARI (years)	Frequency factor
2	0.37
5	0.73
10	1.00
20	1.20
50	1.45
100	1.58

The GLS-based regression was adopted for developing the prediction equations for ARIs of 2, 5, 10, 20, 50 and 100 years based on 106 model catchments in an individual run. There were altogether 107 sets of these equations. The regression coefficients for a given ARI were very similar over the 107 sets of these equations, which is as expected as only one catchment was different between any two runs. A typical set of equations is shown below (Equations 3.14-3.19).

$$\log_{10}(Q_2) = -3.46 + 1.25\log_{10}(A) + 2.40\log_{10}(I_{tc,2}) \quad (3.5.8)$$

$$\log_{10}(Q_5) = -2.73 + 1.15\log_{10}(A) + 2.10\log_{10}(I_{tc,5}) \quad (3.5.9)$$

$$\log_{10}(Q_{10}) = -2.33 + 1.09\log_{10}(A) + 1.94\log_{10}(I_{tc,10}) \quad (3.5.10)$$

$$\log_{10}(Q_{20}) = -1.99 + 1.05\log_{10}(A) + 1.78\log_{10}(I_{tc,20}) \quad (3.5.11)$$

$$\log_{10}(Q_{50}) = -1.58 + 0.99\log_{10}(A) + 1.59\log_{10}(I_{tc,50}) \quad (3.5.12)$$

$$\log_{10}(Q_{100}) = -1.30 + 0.94\log_{10}(A) + 1.48\log_{10}(I_{tc,100}) \quad (3.5.13)$$

It is reassuring to observe that the regression coefficients in this set of equations vary in a regular fashion with increasing ARI.

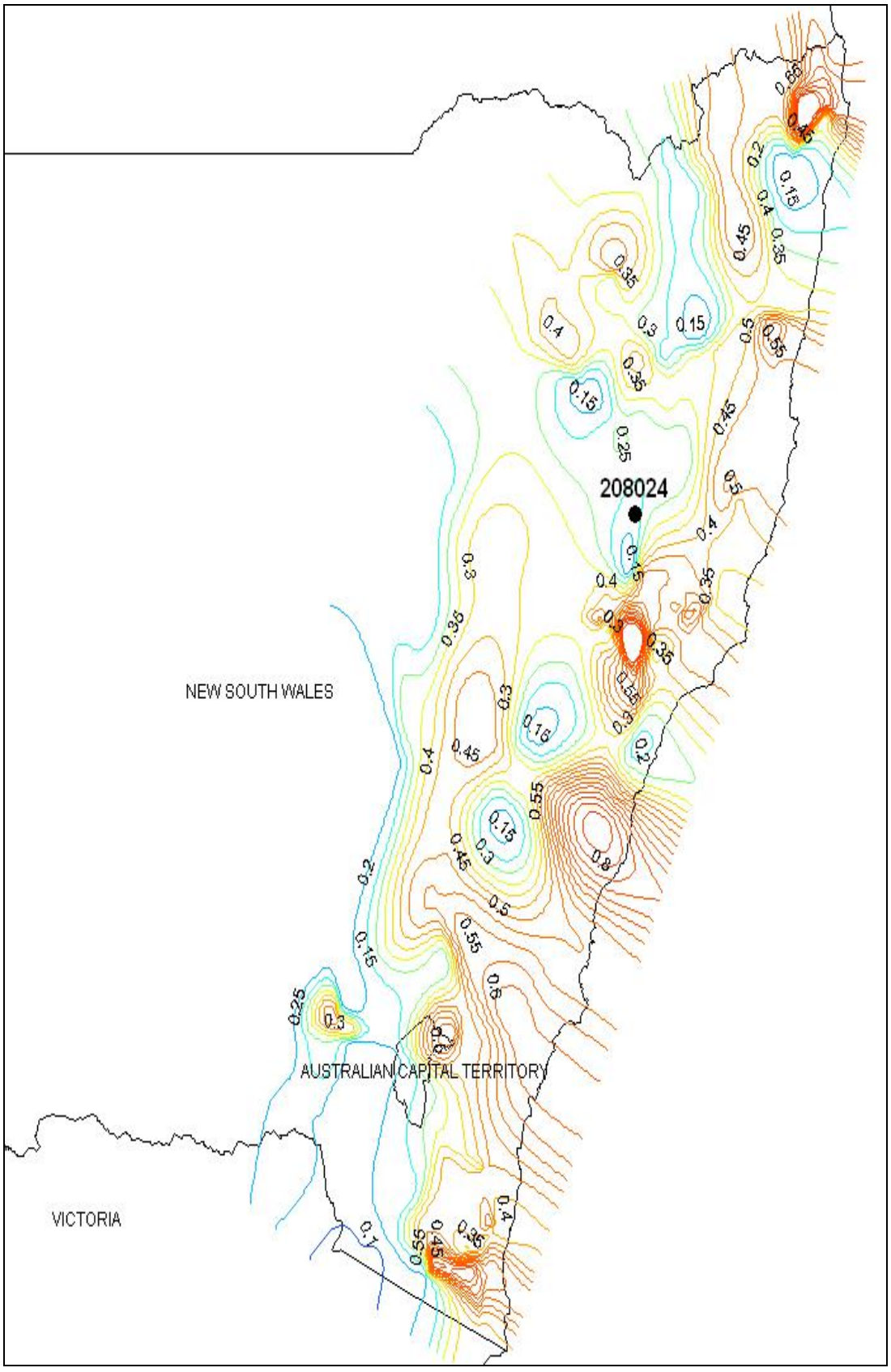


Figure 3.5.4 Typical C10 contour map

Various diagnostic plots related to the prediction equations from the GLS regression were examined. The plots of standardised regression residuals vs. predicted flood quantiles (Figure 3.5.5) showed some trends; however, based on the Kolmogorov-Smirnov and Anderson-Darling tests, the hypothesis that the standardised residuals are normally distributed could not be rejected at the 10% significance level. The QQ-plots of the standardised residuals (Figure 3.6 for ARI = 20 years) show that the assumption of normality for the standardised residuals is well satisfied with all the points closely following a straight line. If the standardised residuals were indeed normally and independently distributed with mean 0 and variance 1, then the slope of the best fit line in the QQ-plot, which can be interpreted as the standard deviation of the sample, should approach 1 and the intercept, which is the mean of the sample, should approach 0 as the number of sites increases. Figure 3.5.6 indeed shows that the fitted line passes through the origin (0, 0) and it has a slope approximately equal to 1. These results indicate that the developed prediction equations satisfy the homogeneity and normality of the residual assumption quite well.

Tables 3.5.2 and 3.5.3 summarise various error statistics for the PRM and QRT. These values are based on independent testing of the PRM and QRT i.e. based on one-at-a-time validation as explained previously. The RMSE values for the PRM are 3 to 12% higher than those of the QRT. The MPRE values for the PRM are 2 to 17% higher than those of QRT. With respect to MSRE, QRT performs better by 1 to 47% for ARIs of 2 to 20 years. However, PRM performs better for ARIs of 50 and 100 years by 4 to 12%. In relation to CE, the higher the values, the better the performance; here QRT performs better by 2 to 12%. In terms of median relative error values (Table 3.5.4), QRT performs better by 11 to 27% for ARIs of 10 to 100 years. For 2 and 5 years ARIs, both the methods perform very similarly.

Table 3.5.2 Comparison of RMSE, MPRE & MSRE for the PRM and QRT.

ARI(years)	RMSE (m^3/s)		MPRE (%)		MSRE	
	PRM	QRT	PRM	QRT	PRM	QRT
2	58.17	56.44	64.21	54.61	0.89	0.60
5	121.79	109.27	48.67	45.01	0.48	0.41
10	193.73	171.8	48.45	45.29	0.47	0.43
20	296.88	265.63	50.12	48.03	0.51	0.50
50	507.65	452.77	56.27	53.32	0.64	0.67
100	732.13	662.62	60.54	59.10	0.75	0.86

Table 3.5.3 Comparison of CE for the PRM and QRT.

ARI(years)	CE	
	PRM	QRT
2	0.64	0.66
5	0.74	0.79
10	0.71	0.77
20	0.65	0.71
50	0.50	0.60
100	0.37	0.48

Table 3.5.4 Median relative error values (%) for the PRM and QRT. The absolute values of the relative errors are considered in obtaining the median value.

Method	Median relative error (%)					
	Q_2	Q_5	Q_{10}	Q_{20}	Q_{50}	Q_{100}
PRM	43	33	36	36	40	44
QRT	44	34	32	30	31	32

Examples of the Q_{pred}/Q_{obs} ratio values for all the six ARIs and 107 test catchments (based on one-at-a-time validation) are presented in Tables 3.5.5 and 3.5.6. Out of the 642 cases (6 ARIs and 107 test catchments), PRM and QRT produce 283 and 305 cases respectively in the range of $0.7 \leq ratio \leq 1.4$, which are equivalent to 44% and 48% of the cases respectively. PRM and QRT respectively produce 24% and 25% cases with $ratio < 0.7$ and 32% and 28% cases with $ratio > 1.4$. There are 77% and 79% of cases in the range of $0.5 \leq ratio \leq 2$ for the PRM and QRT respectively. For the PRM, there are 10% and 13% cases respectively with $ratio < 0.5$ and $ratio > 2$. For the QRT, there are 11% and 10% cases respectively with $ratio < 0.5$ and $ratio > 2$. These results demonstrate that in relation to Q_{pred}/Q_{obs} *ratio*, QRT performs slightly better than the PRM. The results also show that with both PRM and QRT, there are about 10% of cases with gross under-estimation, and another 10% of cases with gross over-estimation, which are expected for these types of approximate RFFA methods.

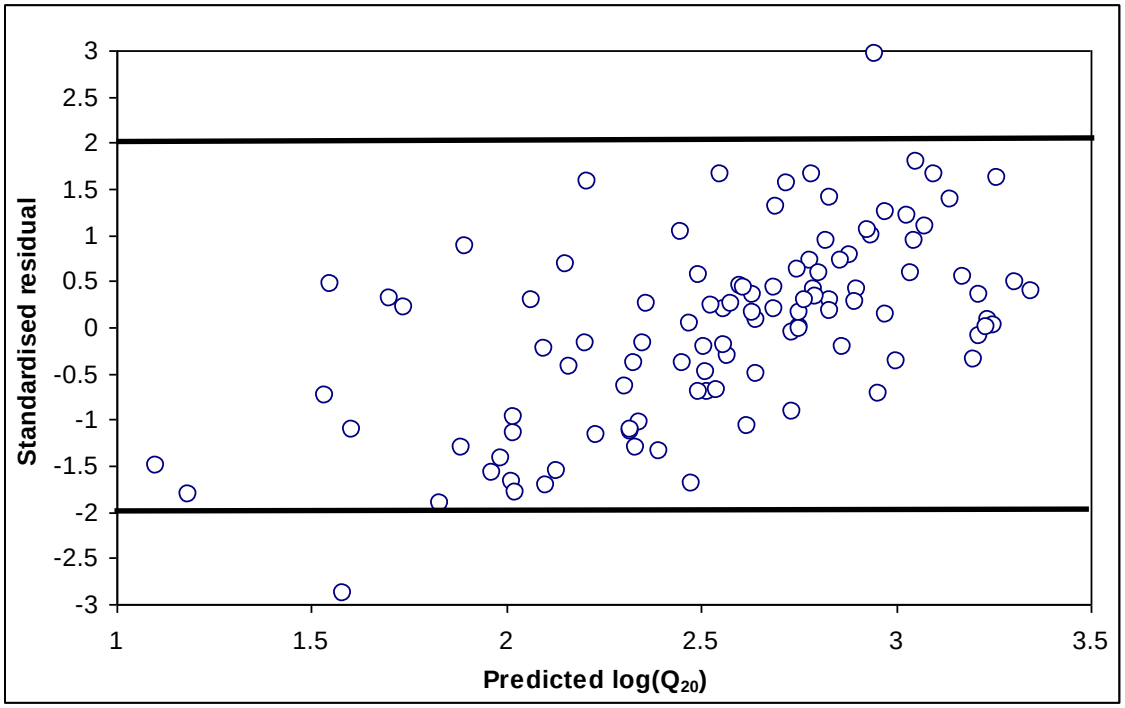


Figure 3.5.5 Standardised residuals vs. predicted quantiles for ARI = 20 years (the heavy lines show the bound of ± 2.0×standardised residual)

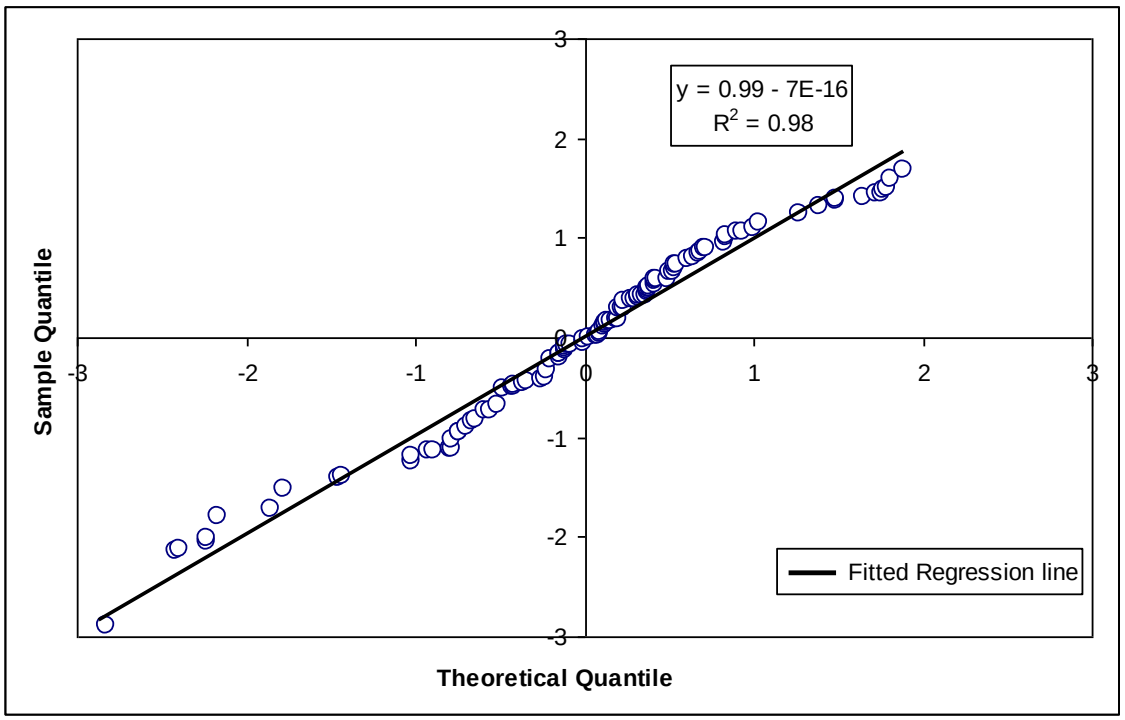


Figure 3.5.6 QQ-plot of the standardised residuals (ARI = 20 years)

Figures 3.5.7 and 3.5.8 present Q_{pred}/Q_{obs} *ratio* values vs. catchment size for the PRM and QRT, respectively. These plots show that the performance of both the PRM and QRT is similar over all catchment sizes, i.e. there is no evidence that the methods perform poorly for

smaller catchments. The box plots of Q_{pred}/Q_{obs} *ratio* values for various ARIs were examined and it was found that generally QRT shows a narrower band of *ratios*. Figure 3.5.9 shows such a plot for the 20 years ARI, which highlights a few outliers for both the PRM and QRT. These outlier catchments were examined but no unusual catchment characteristics were found when compared to the other catchments in the data set. This result indicates that there are likely to be a few cases when PRM and QRT may perform poorly.

The overall results from the split-sample validation test (where 20% catchments were randomly selected as test catchments as explained previously) favoured the QRT over the PRM, where the median relative error values for the QRT were smaller by 5%, 10%, 6% and 12% than those of the PRM for ARIs of 2, 20, 50 and 100 years, respectively. However, for ARIs of 5 and 10 years, PRM shows smaller median relative error values than the QRT by 9% and 2%, respectively. In terms of the count of the Q_{pred}/Q_{obs} *ratio* values, QRT shows 55% ($0.7 \leq$ *ratio* ≤ 1.4), 28% (*ratio* < 0.7) and 17% (*ratio* > 1.4) as compared to 51% ($0.7 \leq$ *ratio* ≤ 1.4), 32% (*ratio* < 0.7) and 17% (*ratio* > 1.4) for the PRM; these results overall favour the QRT.

Table 3.5.5 Summary of counts based on Q_{pred}/Q_{obs} *ratio* values for PRM.

	Count			Percentage		
ARI (years)	*ratio* < 0.7	0.7 ≤ *ratio* ≤ 1.4	*ratio* > 1.4	*ratio* < 0.7	0.7 ≤ *ratio* ≤ 1.4	*ratio* > 1.4
2	24	42	41	22%	39%	38%
5	19	57	31	18%	53%	29%
10	22	53	32	21%	50%	30%
20	27	47	33	25%	44%	31%
50	30	44	33	28%	41%	31%
100	33	40	34	31%	37%	32%
Sum/average	**155**	**283**	**204**	**24%**	**44%**	**32%**

Table 3.5.6 Summary of counts based on Q_{pred}/Q_{obs} *ratio* values for QRT.

	Count			Percentage		
ARI (years)	*ratio* < 0.7	0.7 ≤ *ratio* ≤ 1.4	*ratio* > 1.4	*ratio* < 0.7	0.7 ≤ *ratio* ≤ 1.4	*ratio* > 1.4
2	28	40	39	26%	37%	36%
5	26	51	30	24%	48%	28%
10	25	55	27	23%	51%	25%
20	24	54	29	22%	50%	27%
50	26	53	28	24%	50%	26%
100	29	52	26	27%	49%	24%
Sum/average	**158**	**305**	**179**	**25%**	**48%**	**28%**

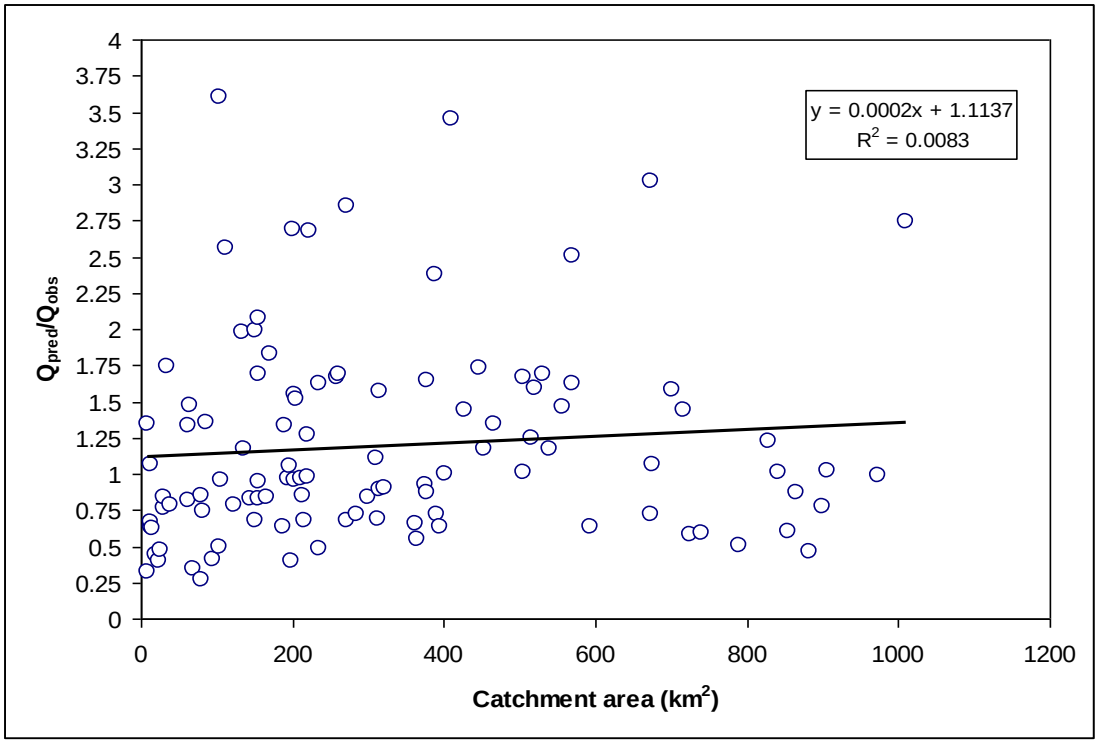


Figure 3.5.7 Q_{pred}/Q_{obs} *ratio* values vs. catchment area for ARI = 20 years for PRM

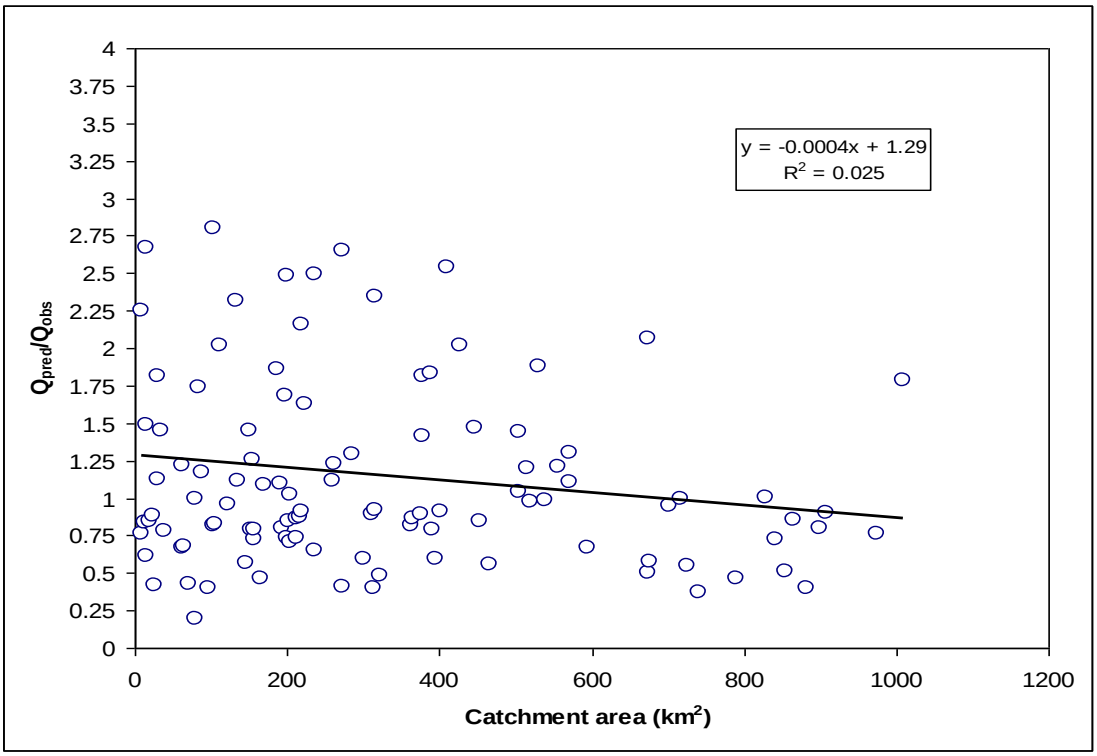


Figure 3.5.8 Q_{pred}/Q_{obs} *ratio* values vs. catchment area for ARI = 20 years for QRT

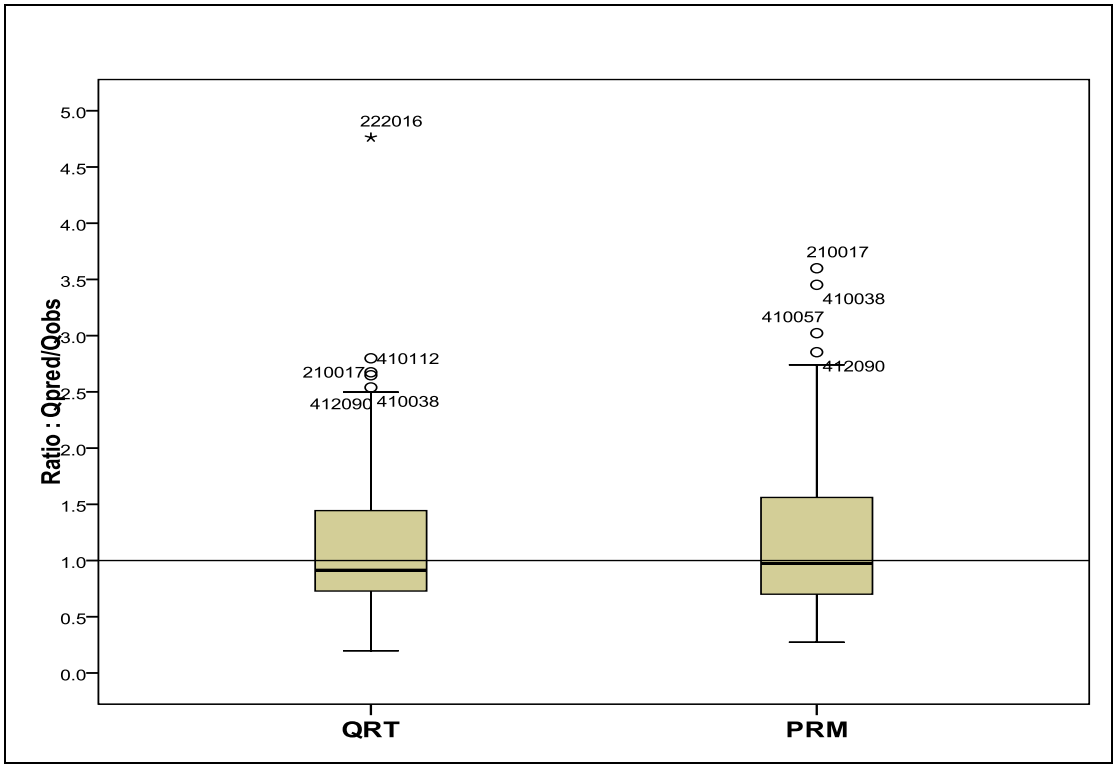


Figure 3.5.9 Box plot showing distribution of Q_{pred}/Q_{obs} *ratio* values for PRM and QRT (ARI = 20 years)

Findings from the comparison of PRM and QRT

This section has compared two commonly used regional flood frequency analysis methods using data from 107 catchments in NSW, the Probabilistic Rational Method (PRM) and the Generalised Least Squares (GLS) based Quantile Regression Technique (QRT). To make a valid comparison, the same predictor variables and data set have been used with both the methods. The comparison examines the specific features of each method and assesses its performance using a one-at-a-time validation method, where each of the 107 study catchments is tested independently, as well as a split sample approach, leaving a randomly selected 20% of catchments for independent testing. The following conclusions can be drawn from this study:

- Based on a range of evaluation statistics (such as root mean squared error, median relative error, mean squared relative error, standard error, coefficient of efficiency and ratio of predicted and observed flood quantiles), the QRT has been found to outperform the PRM.
- There is no evidence that PRM and QRT perform poorly for smaller catchments as far as the range of smaller catchments used in this study is concerned. The applicability

of these methods for catchments smaller than 8 km^2 could not be tested, due to limitations of streamflow data for these catchments.

- The particular advantage of the QRT is that it does not require a contour map of the runoff coefficient as with the PRM. The GLS-based QRT also offers rigorous uncertainty analysis of the estimated flood quantiles by differentiating the sampling and model error. The QRT can also be integrated with the region-of-influence approach where a region can be formed around an ungauged catchment by selecting an 'appropriate number' of neighbouring gauged catchments based on the criterion of minimum model error variance. Hence, QRT offers much greater flexibility and potential in terms of error analysis and further development.
- In the application of the PRM and QRT, the users should note that these are approximate methods and are likely to provide reasonably accurate results in most cases. However, there are likely to be a few cases where relative errors in design flood estimates may be quite high. Thus, the users should consider checking of the results with alternative methods.

4. Development of Quantile Regression Technique (QRT) and Parameter Regression Technique (PRT) using fixed region and ROI

4.1 Introduction

Regression based methods are widely used in RFFA which is based on the concept that spatial variations in flood flow statistics are closely related with variations in regional catchment and climatic characteristics (Pandey and Nguyen, 1999). The most common form of the regression approach is to develop a regression equation for a flood quantile of interest, known as the quantile regression technique (QRT) (Benson, 1962; Thomas and Benson, 1970). The USGS has adopted the QRT as the standard RFFA method since the 1960s (Gupta et al., 1994).

As an alternative to the QRT, the parameters of a probability distribution can be regressed against the explanatory variables; for example, in the case of the log Pearson Type 3 (LP3) distribution, regression equations can be developed for the first three moments i.e. the mean, standard deviation and skewness of the logarithms of annual maximum flood series. For an ungauged catchment, these equations can then be used to predict the mean, standard deviation and skewness to fit an LP3 distribution. This method here is referred to as 'parameter regression technique' (PRT).

The PRT offers three significant advantages over regionalizing quantiles:

- It ensures flood quantiles increase smoothly with increasing average recurrence interval (ARI), an outcome that may not always be achieved with quantile regression;
- It is straightforward to combine any at-site flood information with regional estimates using the approach described by Micevski and Kuczera (2009) to produce more accurate quantile estimates; and
- It permits quantiles to be estimated for any ARI in the range of interest.

There has been little research on the applicability of the PRT as compared to the QRT in RFFA. This section compares the performances of the QRT and PRT for different Australian states.

4.2 Methods

4.2.1 Fixed regions vs. region-of-influence (ROI) approaches

In RFFA, formation of regions can be based on geographic and administrative boundaries or in catchment characteristics data space. The allocation of an ungauged catchment to regions formed in catchment characteristics data space is often problematic. Acreman and Wiltshire (1987) proposed regions without fixed boundaries. Subsequently, Burn (1990a, 1990b) and Zrinji and Burn (1994) proposed the region-of-influence (ROI) approach where each site of interest (i.e. catchment where flood quantiles are to be estimated) can form its own region. Tasker et al. (1996) compared five methods of developing regression models for ungauged catchments using data from 204 gauging stations in Arkansas. The formation of regions in these methods was based on proximity in geographical space or catchment attributes space. They found the ROI approach was the best among the five methods considered. A key advantage of the ROI approach is that it can overcome the inconsistency in flood quantile estimates at the boundary of two neighbouring administrative regions (e.g. state borders). A recent study by Eng et al. (2005) compared the performance of ROI approaches based on predictor-variable similarity or geographical proximity for estimating the 50-year peak discharge, using an ordinary least squares approach with 1091 sites in southeastern USA. They found that using geographical proximity produced the smallest predictive errors over the study region. Similar results demonstrating the superiority of geographical proximity over predictor-variable similarity have been shown by others (e.g. Merz and Blöschl, 2005, Kjeldsen and Jones, 2006).

In this chapter, the performances of the fixed regions (based on state boundary) and ROI approaches are compared. The ROI approach uses the physical distance between sites as the distance metric (i.e. geographic proximity). In applying the ROI approach, in the first iteration, the 15 nearest sites to the site of interest are selected and the regional BGLS regression is performed and the predictive variance is noted. The second iteration proceeds with the next five closest stations being added to the previous ROI and repeating the BGLS regression. This procedure terminates when all the eligible sites have been included in the ROIs. The ROI for the site of interest is then selected as the one exhibiting the lowest predictive variance. Here, in implementing the ROI approach, the states of NSW, Victoria, Queensland, Tasmania and south-east WA are treated separately so that a comparison between the ROI and fixed regions can be made. In future testing of the ROI, the database

for NSW, Victoria and Queensland will be combined as they form a continuous spatial distribution.

The ROI approach adopted here is fundamentally different to that of Tasker et al. (1996) in that it seeks to minimise

- It seeks to minimise regression model's predictive error variance rather than selecting or assuming a fixed number of sites that minimise a distance metric in catchment characteristic space;
- the ROI criterion of Tasker et al. (1996) cannot guarantee minimum predictive variance; and
- moreover, the selection of sites that are minimally different in catchment characteristic space may result in greater uncertainty in the estimated regression coefficients.

4.2.2 Bayesian generalised least squares regression

Hydrologists commonly use ordinary least squares (OLS) estimators that are appropriate and statistically efficient if the flow records are of equal length and if concurrent flows between any pair of stations are uncorrelated (Tasker et al., 1986). These assumptions are often violated with regional annual maximum flood series data. To overcome the problems with the GLS regression, Stedinger and Tasker (1985, 1986) developed a GLS model that accounts for the differences in at-site record lengths and inter-site correlation among at-site estimators. Stedinger and Tasker (1985, 1986) showed in a Monte Carlo simulation that the GLS estimators provide model regression parameters with smaller mean–squared errors than the competing OLS estimators, provide relatively unbiased estimates of the variance of the regression parameters and produce more accurate estimates of the regression model error. GLS regression has been widely adopted in hydrology (e.g., Tasker and Stedinger, 1989; Madsen et al., 1995; Madsen et al., 1997; Kroll and Stedinger, 1998; Reis et al., 2005; Eng et al., 2005; Griffis and Stedinger, 2007; Gruber and Stedinger, 2008; Hackelbusch et al., 2009; Micevski and Kuczera, 2009).

Reis et al. (2005) and Gruber et al. (2007) introduced a Bayesian analysis of the GLS model which provides more accurate measure of the model error variance and a more realistic description of the possible values of the model error variance in cases where the method of

moments estimator of the model error variance as described by Stedinger and Tasker (1985) may be zero or close to it; this occurs when sampling errors dominate the regional analysis.

GLS regression

The GLS regression assumes that the hydrological variable of interest (e.g. a flood quantile or a parameter of the LP3 distribution) denoted by y_i for a given site *i* can be described by a function of catchment characteristics (explanatory variables) with an additive error:

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_j X_{ij} + \delta_i \quad i = 1, 2, ..., n \qquad (4.2.1)$$

where X_{ij} (*j* = 1,…, *k*) are explanatory variables, β_j are the regression coefficients, δ_i is the model error which is assumed to be normally and independently distributed with model error variance σ_δ^2 and *n* is the number of sites in the region. In all cases only an at-site estimate of y_i denoted as $\hat{y}_i$ is available. To account for the error in this data, a sampling error η_i must be introduced into the model so that:

$$\hat{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta} + \boldsymbol{\delta} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \text{ where } \hat{y}_i = y_i + \eta_i;\ i = 1, 2, ..., n \qquad (4.2.2)$$

Thus the observed regression model errors $\boldsymbol{\varepsilon}_i$ are the sum of the model errors δ_i and the sampling errors η_i. The total error vector $\Lambda(\sigma_\delta^2)$ has mean zero and a covariance matrix:

$$E\left[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\right] = \boldsymbol{\Lambda}(\sigma_\delta^2) = \sigma_\delta^2 \mathbf{I} + \sum(\hat{\mathbf{y}}) \qquad (4.2.3)$$

where $\sum(\hat{\mathbf{y}})$ is the covariance matrix of the sampling errors in the sample estimators of the flood quantiles or the parameters of the LP3 distribution. The variance of η_i depends on the record length available at each site and the cross-correlation of the sites flood data. Therefore the observed regression model errors are a combination of time-sampling error η_i and an underlying model error δ_i.

In this regional framework, σ_δ^2 can be viewed as a heterogeneity measure. Madsen et al. (2002; 1997) showed that the regional average GLS estimator is a general extension of the

record-length-weighted average commonly applied in the index-flood procedure; however the record-length-weighted average estimator neglects inter-site correlation and regional heterogeneity (Stedinger et al., 1992 and Stedinger and Lu, 1995).

The GLS estimator of $\boldsymbol{\beta}$ and its respective covariance matrices for known σ_δ^2 are given by:

$$\hat{\beta}_{GLS} = \left[\mathbf{X}^T \Lambda(\sigma_\delta^2)^{-1} \mathbf{X}\right]^{-1} \mathbf{X}^T \Lambda(\sigma_\delta^2)^{-1} \hat{\mathbf{y}} \qquad (4.2.4)$$

$$\Sigma[\hat{\beta}_{GLS}] = \left[\mathbf{X}^T \Lambda(\sigma_\delta^2)^{-1} \mathbf{X}\right]^{-1} \qquad (4.2.5)$$

The model error variance σ_δ^2 can be estimated by either generalised method of moments (MOM) or maximum likelihood (ML) estimators as described by Stedinger and Tasker (1986). The MOM estimator is determined by iteratively solving equation 4.2.6 along with the generalised residual mean square error equation:

$$(\hat{\mathbf{y}} - \mathbf{X}\hat{\boldsymbol{\beta}}_{GLS})^T [\hat{\sigma}_\delta^2 \mathbf{I} + \sum(\hat{\mathbf{y}})]^{-1} (\hat{\mathbf{y}} - \mathbf{X}\hat{\boldsymbol{\beta}}_{GLS}) = n - (k+1) \qquad (4.2.6)$$

In some situations, the sampling covariance matrix explains all the variability observed in the data, which means the left-hand side of equation 4.2.6 will be less than n – (k+1) even if $\hat{\sigma}_\delta^2$ is zero. In these circumstances, the MOM estimator of the model error variance is generally taken to be zero (Stedinger and Tasker, 1985; 1986).

Bayesian GLS regression

Bayesian inference is an alternative to the classical statistical approach. In a Bayesian framework, the parameters of the model are considered to be random variables, whose probability density function should be estimated. Reis et al. (2005) developed a Bayesian approach to estimate the regional model parameters and showed that the Bayesian approach can provide a realistic description of the possible values of the model error variance, especially in the case where sampling error tends to dominate the model errors in the regional analysis.

With the Bayesian approach, it is assumed that there is no prior information on any of the β parameters thus a multivariate normal distribution with mean zero and a large variance (e.g. greater than 100) is used as a prior for the regression coefficient parameters as suggested by Reis et al. (2005). This prior is considered virtually non-informative, which produces a probability distribution function that is generally flat in the region of interest. The prior information for the model error variance σ_δ^2 is represented by an informative one-parameter exponential distribution, which represents the reciprocal of the prior mean of the model error variance. Reis et al. (2005) made a detail discussion on the derivation of the prior for the model error variance for regionalising the skew. A variance of 6 is adopted in our study for the regionalisation of skew following Reis et al. (2005).

To derive the prior distribution for the standard deviation, mean flood and quantiles of the LP3 distribution we used an informative one-parameter exponential distribution where the reciprocal of the residual error variance estimates from the OLS regression is taken as the prior mean of the model error variance. For the mean flood and flood quantiles, the model error variance tends to dominate the regional analysis. In this case a zero or negative value for the model error variance is highly unlikely.

A negative model error variance is unrealistic as noted by Reis et al. (2005). In this situation equation 4.2.6 may introduce further uncertainty into the regional model. A Bayesian estimator of the model error variance as discussed above may be used to safeguard against this happening. Further details on this can be found in Reis et al. (2005) and Micevski and Kuczera (2009). In summary, the Bayesian estimator offers a better way of dealing with the model error and quantifying uncertainty associated with this.

Selection of predictor variables

In the OLS regression, several statistics are used to justify the model selection such as the traditional coefficient of determination (R^2), *F* statistics, Durbin Watson Statistics, Akaike information criterion (AIC) and Bayesian information criteria (BIC) (Gelman et al., 2004). Among these statistics, the AIC and BIC penalise for the extra complexity in the model, which means that an extra predictor variable must improve the model significantly to justify its inclusion. Provided below is a brief discussion on the Bayesian GLS regression statistics that guided our model selection procedure.

Average Variance of Prediction

In RFFA, the objective is to make prediction at both gauged and ungauged sites, hence a statistic appropriate for evaluation of model selection is the variance of prediction, which in many cases depends on the explanatory variables at both gauged and ungauged sites. Hence, Tasker and Stedinger (1989) suggested the use of the average variance prediction (AVP) computed with the sites in the regression. The assumption here is that these sites are representatives of all the sites in the region, or at least the sites at which predictions are needed.

By using a GLS regression model one can predict a hydrological statistic on average over a new region. Thus this becomes the average variance of prediction AVP_{new} for a new site which is made up of the uncertainty in estimating the coefficients of the regression model and the average model error (Tasker and Stedinger, 1986).

Consider a new site not used in the derivation of the regional model. Let $\mathbf{x}_o$ be the vector of characteristics at the new site. The expected value of the variable of interest y_o is $\mathbf{x}_o^T\hat{\boldsymbol{\beta}}$ where $\hat{\boldsymbol{\beta}}$ is the expected GLS value of $\boldsymbol{\beta}$. The predictive variance of y_o is (Reis et al. 2005):

$$Var(y_o) = \sigma_\delta^2 + \mathbf{x}_i(\mathbf{X^T}\boldsymbol{\Lambda}^{-1}\mathbf{X})^{-1}\mathbf{x}_i^T \qquad (4.2.7a)$$

The second term is the contribution of uncertainty in $\boldsymbol{\beta}$ to y_o. For Bayesian GLS analysis according to Gruber et al. (2007):

$$AVP_{new} = E[\sigma_\delta^2] + \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i Var[\boldsymbol{\beta}|\hat{\mathbf{y}}]\mathbf{x}_i^T \qquad (4.2.7b)$$

Also, if the prediction is for a site that was used in the estimation of the regional regression model, the measure of prediction AVP_{old} requires an additional term:

$$AVP_{old} = E[\sigma_\delta^2] + \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i Var[\boldsymbol{\beta}|\hat{\mathbf{y}}]\mathbf{x}_i^T - 2\sigma_\delta^2\mathbf{x}_i(\mathbf{X^T}\boldsymbol{\Lambda}^{-1}\mathbf{X})^{-1}\mathbf{X^T}\boldsymbol{\Lambda}^{-1}\mathbf{e}_i \qquad (4.2.8)$$

where $\mathbf{e}_i$ is a unit column vector with 1 at the *i*th row and 0 otherwise.

Bayesian and Akaike Information Criteria

The Akaike information criterion (AIC) is given by equation 4.2.9, where $\Pi(Y)$ is the log-likelihood maximised function, *n* is the number of sites in the region and *k* is the number of predictor variables in the fitted regression model. The first term on the right hand side of equation 4.2.9 measures essentially the true lack of fit while the second term measures model complexity which is related to the number of predictor variables.

$$\mathrm{AIC} = -2\Pi(\mathrm{Y}) + 2k \qquad (4.2.9)$$

In practice, after the computation of the AIC for all of the competing models, one selects the model with the minimum AIC value, $\mathrm{AIC}_{\mathrm{min.}}$ The Bayesian information criterion (BIC) is very similar to AIC, but is developed in a Bayesian framework:

$$\mathrm{BIC} = -2\Pi(\mathrm{Y}) + \ln(n)k \qquad (4.2.10)$$

The BIC penalises more heavily models with higher values of *k* than does AIC. Since $\Pi(Y)$ depends on the sample size (*n*), the competing models can be compared using AIC and BIC only if fitted using the same sample, as done in this study.

Regression Diagnostics

The assessment of the regional regression model is made by using a number of statistical diagnostics such as a pseudo–coefficient of determination and standard error of prediction. An analysis of variance for the Bayesian GLS models is undertaken to examine which proportion of the sampling and model errors dominates the regional analysis. The standardised residuals are used to identify outlier sites; absence of outlier in regression diagnostics indicates the overall adequacy of the regional model. These statistics are described below.

Co-efficient of Determination (R^2) and Analysis of Variance

The traditional coefficient of determination (R^2) measures the degree to which a model explains the variability in the dependent variable. It uses the partitioning of the sum of squared deviations and associated degrees of freedom to describe the variance of the signal versus the model error. Traditionally for OLS regression, the Total-Sum-of-Squared deviations about the mean (SST) is divided into two separate terms, the Sum-of-Squared Errors explained by the regression model (SSR) and the residual Sum-of-Squared Errors (SSE), where SST = SSR + SSE.

Reis et al. (2005) proposed a pseudo co-efficient of determination ($\overline{R}^2_{GLS}$) appropriate for use with the GLS regression. For traditional R^2, both the SSE and SST include sampling and model error variances, and therefore this statistic can grossly misrepresent the true power of the GLS model to explain the actual variation in the y_i. Hence, for the GLS regression a more appropriate pseudo co-efficient of determination is defined by:

$$\overline{R}^2_{GLS} = \frac{n[\hat{\sigma}^2_\delta(0) - \hat{\sigma}^2_\delta(k)]}{n\hat{\sigma}^2_\delta(0)} = 1 - \frac{\hat{\sigma}^2_\delta(k)}{\hat{\sigma}^2_\delta(0)} \qquad (4.2.11)$$

where $\hat{\sigma}^2_\delta(k)$ and $\hat{\sigma}^2_\delta(0)$ are the model error variances when k and no explanatory variables are used, respectively. Here, $\overline{R}^2_{GLS}$ measures the improvement of a GLS regression model with k explanatory variables against the estimated error variance for a model without any explanatory variable. If $\hat{\sigma}^2_\delta(k)$ = 0, $\overline{R}^2_{GLS}$ = 1 as it should, even though the model is not perfect because var[$\eta_i + \delta_i$] is still not zero because var[η_i] > 0. A pseudo ANOVA table is used in GLS regression for the error variance analysis as presented by Reis et al. (2005) and Griffis and Stedinger (2007).

Standard Error of Prediction

If the residuals have a nearly normal distribution, the standard error of prediction in percent (SEP) for the true flood quantiles/flood statistics is described by:

$$SEP(\%) = 100 \times [\exp(AVP_{new}) - 1]^{0.5} \qquad (4.2.12)$$

where the regression models independent and dependant variables have been transformed by $\log_e$.

Analysis of Residuals and Z-score

Analysis of residuals provides a means of assessing the model fit and identifying possible outliers. In this study, the standardised residual (r_{si}) (Tasker et al., 1996) is used, which is the residual r_i divided by the square root of its variance. This is calculated as:

$$r_{si} = \frac{r_i}{[\lambda_i + \mathbf{x}_i(\mathbf{X}^{\mathbf{T}}\mathbf{\Lambda}^{-\mathbf{1}}\mathbf{X})^{-1}\mathbf{x}_i^T]^{0.5}} \quad where\ \lambda_i\ is\ the\ diagonal of\ \mathbf{\Lambda} \tag{4.2.13}$$

To assess the adequacy of the model in estimating flood quantiles, a Z score is used, which is defined as:

$$Z_{score} = \frac{LNQ_{ARI,i} - LN\hat{Q}_{ARI,i}}{\sqrt{\sigma^2_{ARI,i} + \hat{\sigma}^2_{ARI,i}}} \tag{4.2.14}$$

Here the numerator is the difference between the at-site flood quantile and regional flood quantile (estimated from the developed regression equation) and the denominator is the square root of the sum of the variances of the at-site ($\sigma^2_{ARI,i}$) and regional ($\hat{\sigma}^2_{ARI,i}$) flood quantiles in natural logarithm space. It is reasonable to assume that the errors in the two estimators are independent because $Q_{ARI,i}$ is an unbiased estimator of the true quantile estimators based upon the at-site data, whereas the error in $\hat{Q}_{ARI,i}$ is mostly due to the failure of the best regional model to estimate accurately the true at-site flood quantile. The use of log space makes the difference approximately normally distributed and hence enables the use of standard statistical tests.

4.2.3 At-site Flood Frequency Analysis and Quantile and Parameter Regression Technique

At-site flood quantiles for ARIs of 2, 5, 10, 20, 50 and 100 years were estimated by FLIKE (at-site flood frequency analysis software) using an LP3 distribution with Bayesian parameter estimation procedure as described in Kuczera (1999). No prior information was used in fitting the LP3 distribution. The parameters of the LP3 distribution were also extracted from the FLIKE software.

To regionalise the flood quantiles the sampling covariance matrix (Σ) of the LP3 distribution is required. Tasker and Stedinger (1989) and Griffis and Stedinger (2007) provide the approximate estimator of the components of Σ matrix of the LP3 distribution. The skew and standard deviation in the Σ matrix are subject to estimation uncertainty. In this study to avoid correlation between the residuals and the fitted quantiles, (i) the inter site correlation between the concurrent annual maximum flood series (ρ_{ij}) is estimated as a function of the distance

between sites *i* and *j,* (ii) the standard deviations σ_i and σ_j are estimated using a separate OLS/GLS regression using the explanatory variables used in the study, and (iii) the regional skew is used in place of the population skew γ as suggested by Tasker and Stedinger (1989). This analysis above used the regional estimates of the standard deviation and skew obtained from Bayesian GLS regression. The detailed information on the covariance matrices associated with the standard deviation and skew can be found in Reis et al. (2005) and Griffis and Stedinger (2007).

For the parameter regression technique (PRT), we adopted the GLS regression (Tasker and Stedinger, 1989 and Griffis and Stedinger, 2007) using a Bayesian framework (Reis et al., 2005) to develop regression equations for the parameters of the LP3 distribution (i.e. mean μ, standard deviation σ, and skew coefficient γ of the logarithms of the annual maximum flood series). The regional values of standard deviation and skew were taken from the Σ matrix of the flood quantile modelling as mentioned above. The covariance matrix for the mean flood was obtained following Stedinger and Tasker (1986)

4.2.4 Evaluation Statistics

We evaluate the overall performance of the Bayesian GLS regression method by using one-at-a-time cross validation. The site of interest is left out in building the model so it is in effect being treated as an ungauged site. This is repeated for all the sites considered in the study. The advantage of the one-at-a-time cross-validation procedure is that it generates quantile or moment estimates for the site of interest which are independent from the site itself. To compare model adequacy we adopt a number of evaluation statistics (equations 4.2.15 to 4.2.17)) being the relative error (RE), relative root mean square error (RMSE) and the mean ratio of the predicted flow to observed flow (ratio).

$$RE(\%) = \frac{Q_{pred} - Q_{obs}}{Q_{obs}} \times 100 \qquad (4.2.15)$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{Q_{pred} - Q_{obs}}{Q_{obs}}\right)^2} \qquad (4.2.16)$$

$$ratio = \frac{Q_{pred}}{Q_{obs}} \qquad (4.2.17)$$

where Q_{obs} is the observed flood quantile obtained from at-site flood frequency analysis, Q_{pred} is the predicted flood quantile obtained from the Bayesian GLS-QRT or Bayesian GLS-PRT based on the one-at-a-time cross validation approach and *n* is the number of sites in the region.

The RMSE (%) and RE provide an indication of the overall accuracy of a model. The average value of the Q_{pred}/Q_{obs} ratio gives an indication of the degree of bias (i.e. systematic over- or under estimation), where a value of 1 indicates a good average agreement between the Q_{pred} and Q_{obs}. A Q_{pred}/Q_{obs} ratio value in the range of 0.5 to 2 may be regarded as 'desirable (D)', a value smaller than 0.5 may be regarded as 'gross underestimation (U)', and a value greater than 2 may be regarded as 'gross overestimation (O)'. It should be mentioned here that these are only arbitrary limits and would provide a reasonable guide about the relative accuracy of the methods as far as the practical application of the methods is concerned. In applying these evaluation statistics to compare the alternative models, factors such as data error (e.g. measurement error and error due to rating curve extrapolation) and the error associated with the at-site flood frequency analysis have not been considered.

4.3 Results for NSW and ACT

4.3.1 QRT and PRT – fixed and ROI approaches

A total of 96 catchments were used from NSW and ACT for the analyses presented here. These catchments are listed in Table A1. The record lengths of annual maximum flood series of these 96 stations range from 25 to 75 years (mean: 37 years, median: 34 years and standard deviation: 11.4 years). The catchment areas of the selected 96 catchments range from 8 km^2 to 1010 km^2 (mean: 353 km^2 and median: 267 km^2). The geographical distribution of the selected 96 catchments is shown in Figure 2.2. The distribution of the catchment areas of these stations is shown in Figure 2.3.

In the fixed region approach, all the 96 catchments were considered to have formed one region, however, one catchment was left out for cross-validation and the procedure was repeated 96 times to implement the one-at-a-time cross validation. The ROI approach in this study was applied to the parameters (i.e. mean, standard deviation and skew) and flood quantiles of the LP3 distribution to further reduce the heterogeneity unaccounted for by the fixed-region BGLS model.

The ROI approach in this paper uses the physical distance between sites as the distance metric (i.e. geographic proximity). In the first iteration, the 15 nearest stations to the site of interest are selected and a regional BGLS regression is performed and the predictive variance is noted. The second iteration proceeds with the next five closest stations being added to the ROI and repeating the regression. This procedure terminates when all 96 sites have been included in the ROI. The ROI for the site of interest is then selected as the one which shows the lowest predictive variance.

Table 4.3.1 shows different combinations of predictor variables for Q_{10} QRT model and the first three moments of the LP3 distribution. Figures 4.3.1 and 4.3.2 show example plots of the statistics used in selecting the best set of predictor variables for Q_{10} and the skew models. According to the model error variance, combinations 19, 18, 20, 23, 16, 6, 4, 25 and 10 were potential sets of predictor variables for the Q_{10} model. Combinations 18, 19, 20 and 23 contained 3 to 4 predictor variables while combinations 16, 6, 4, 25 and 10 contained 2 predictor variables with similar model error variances and R^2_{GLS}. The average variance of prediction for an old and new site (AVPO) and (AVPN) and the Akaike information and Bayesian information criteria (AIC) and (BIC) values favour combination 10, and hence this

was finally selected as the best set of predictor variables for the Q_{10} model which includes area and design rainfall intensity $I_{tc,10.}$

For the skew model, combination 9 showed a slightly higher model error variance than combination 1 and the highest R^2_{GLS} (see Figure 4.3.2) as well as the lowest AIC and BIC. Combination 1 without any predictor variables however showed a lower AVPO and AVPN as compared to combination 9. Both combinations 1 and 9 were trialled in this study.

A similar procedure was adopted in selecting the best set of predictor variables for the other models with the QRT and PRT. The set of predictor variables selected as above was used in the one-at-a-time cross validation (with fixed regions) and region-of-influence (ROI) approach.

The significance of the estimated regression coefficient values shown in Equations 4.3.1 to 4.3.9 was evaluated using the Bayesian plausibility value (BPV) as described by Reis et al. (2005) and Gruber et al. (2007). The BPV allows one to perform the equivalent of a classical hypothesis *p*-value test within a Bayesian framework. The BPV was carried out at the 5% significance level. The advantage of the BPV is that it uses the posterior distribution of each β-parameter. The BPVs for the regression coefficients associated with variable area and design rainfall intensity $I_{tc,ARI}$ for the QRT over all the ARIs were smaller than 0%. The BPVs for the skew model were 6% and 7% for area and forest, respectively indicating that these are reasonably good predictors for skew in this application. The BPVs for the mean flood model were 0% for both the predictor variables (area and $^2I_{12}$). For the standard deviation model, the BPVs for the predictor variables rain and S1085 were 2%.

Regression equations developed for the QRT and PRT for the fixed region are given by Equations 4.3.1 to 4.3.9.

Table 4.3.1 Different combinations of predictor variables considered for the QRT models and the parameters of the LP3 distribution (QRT and PRT fixed region NSW)

Combination	Combinations for mean, standard deviation & skew models	Combinations for flood quantile model
1	Const	Const
2	Const, area	Const, area
3	Const, area, $^{2}I_{1}$	Const, area, $^{2}I_{1}$
4	Const, area, $^{50}I_{1}$	Const, area, $^{2}I_{12}$
5	Const, area, $^{50}I_{12}$	Const, area, $^{50}I_{1}$
6	Const, area, $^{2}I_{12}$	Const, area, $^{50}I_{12}$
7	Const, area, S1085	Const, area, rain
8	Const, area, sden	Const, area, for
9	Const, area, forest	Const, area, forest, evap
10	Const, area, evap	Const, area, $I_{tc,ARI}$
11	Const, area, rain	Const, area, evap
12	Const, rain, S1085	Const, area, S1085
13	Const, sden, S1085	Const, area, sden
14	Const, evap, sden	Const, sden, rain
15	Const, forest	Const, for, rain
16	Const, S1085, forest	Const, area, $^{50}I_{12}$, rain
17	-	Const, area, $^{50}I_{12}$, sden
18	-	Const, area, $^{50}I_{12}$, rain, evap
19	-	Const, area, $^{50}I_{12}$, $I_{tc,ARI}$, evap
20	-	Const, area, $^{50}I_{12}$, $I_{tc,ARI}$, rain, evap
21	-	Const, area, $^{50}I_{12}$, $I_{tc,ARI}$, sden
22	-	Const, area, $^{50}I_{12}$, $I_{tc,ARI}$, S1085
23	-	Const, area, $I_{tc,ARI}$, evap
24	-	Const, area, $I_{tc,ARI}$, rain
25	-	Const, area, $^{2}I_{1}$, $I_{tc,ARI}$

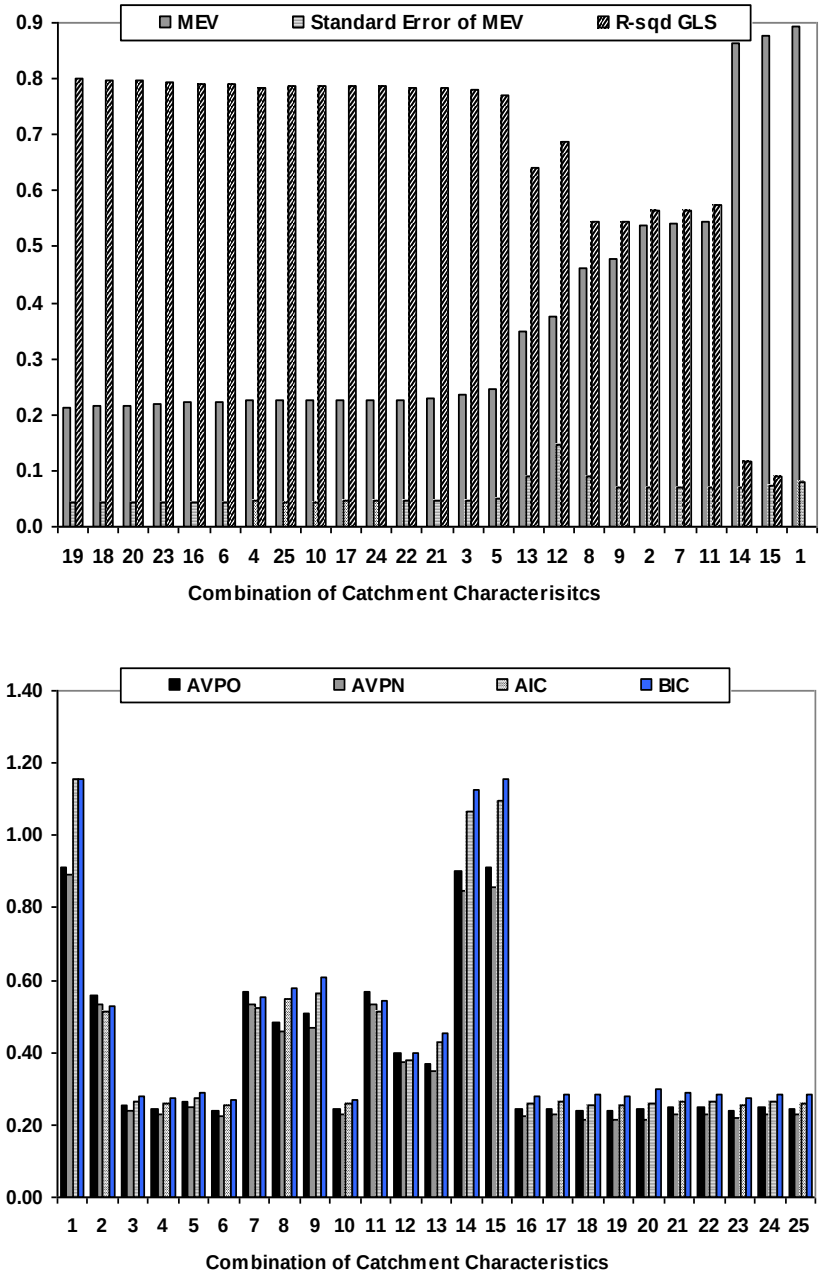


Figure 4.3.1 Selection of predictor variables for the BGLS regression model for Q_{10} model (QRT, fixed region NSW), MEV = model error variance, AVPO = average variance of prediction (old), AVPN = average variance of prediction (new) AIC = Akaike information criterion, BIC = Bayesian information criterion

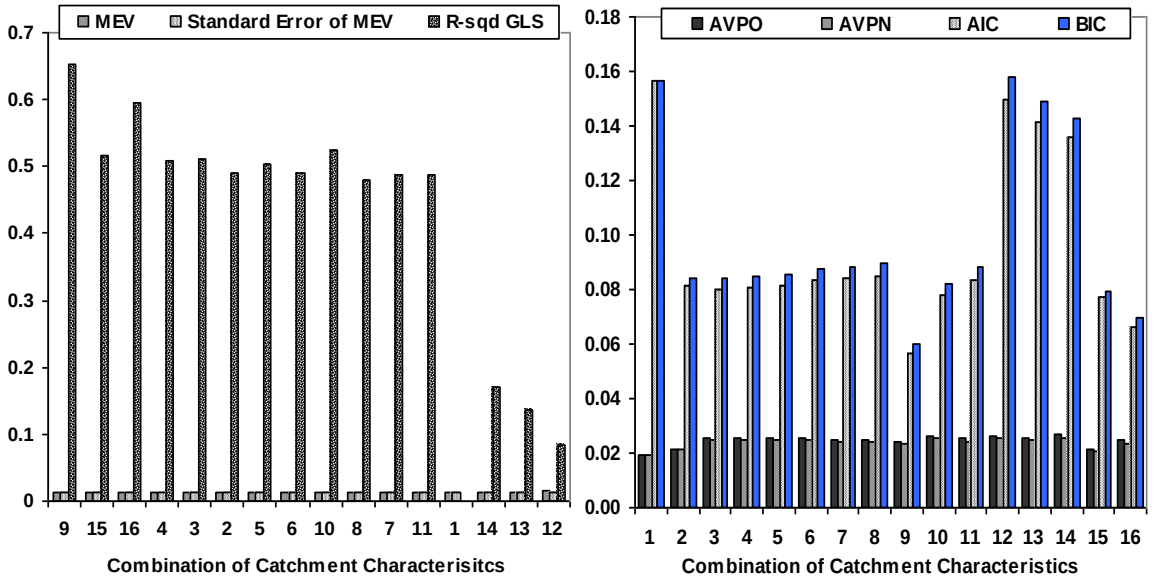


Figure 4.3.2 Selection of predictor variables for the BGLS regression model for skew

$\ln(Q_2) = 4.06 + 1.26(\boldsymbol{z}area) + 2.42(\boldsymbol{z}I_{tc,2})$ (4.3.1)

$\ln(Q_5) = 5.11 + 1.19(\boldsymbol{z}area) + 2.08(\boldsymbol{z}I_{tc,5})$ (4.3.2)

$\ln(Q_{10}) = 5.56 + 1.14(\boldsymbol{z}area) + 1.93(\boldsymbol{z}I_{tc,10})$ (4.3.3)

$\ln(Q_{20}) = 5.91 + 1.09(\boldsymbol{z}area) + 1.79(\boldsymbol{z}I_{tc,20})$ (4.3.4)

$\ln(Q_{50}) = 6.55 + 1.01(\boldsymbol{z}area) + 1.73(\boldsymbol{z}I_{tc,50})$ (4.3.5)

$\ln(Q_{100}) = 6.47 + 0.97(\boldsymbol{z}area) + 1.50(\boldsymbol{z}I_{tc,100})$ (4.3.6)

$M = 4.09 + 0.67(\boldsymbol{z}area) + 2.31(\boldsymbol{z}I_{12,2})$ (4.3.7)

$stdev = 1.22 - 0.59(\boldsymbol{z}rain) - 0.13(\boldsymbol{z}S1085)$ (4.3.8)

$skew = -0.42 - 0.10(\boldsymbol{z}area) - 0.10(\boldsymbol{z}forest)$ (4.3.9)

where

$$z(x_i) = \ln(x_i) - \frac{\sum_{i=1}^{n} \ln(x_i)}{n} \quad (4.3.10)$$

where, $\ln(x_i)$ is the logarithm of the catchment variable at a site; and

$\frac{\sum_{i=1}^{n} \ln(x_i)}{n}$ is the arithmetic average of the same log transformed catchment variable over all the sites in the region. These average values for the required predictor variables for all the states/regions are provided in Section 4.10 (Table 4.10.1). Here, M = average of $\ln(Q)$, where

Q is annual maximum flood series and stdev is the standard deviation of ln(*Q*) and skew is the skewness of ln(*Q*) data.

The Pseudo Analysis of Variance (ANOVA) tables for the Q_{20} and Q_{100} models and the parameters of the LP3 distribution are presented in Tables 4.3.2 – 4.3.6 for the fixed regions and ROI. A Pseudo ANOVA table describes how the total variation among the $\hat{y}_i$ values can be apportioned between that explained by the model error and sampling error. This is an extension of the ANOVA in the OLS regression which does not recognize and correct for the expected sampling variance (Reis et al., 2005). An error variance ratio (EVR) is used in Pseudo ANOVA, which is the ratio of the sampling error variance to model error variance. An EVR greater than 0.20 indicates that the sampling variance is not negligible when compared to the model error variance, which suggests the need for a GLS regression analysis (Gruber et al., 2007).

For the LP3 parameters, the sampling error (i.e. EVR) increases as the order of moment increases. The ROI shows a reduced model error variance (i.e. a reduced heterogeneity) as compared to the fixed regions. The model error dominates the regional analysis for the mean flood and the standard deviation models for both the fixed regions and ROI. However, ROI shows a higher EVR than the fixed regions; e.g. for the mean flood model the EVR is 0.30 for the ROI and 0.17 for the fixed region (Table 4.3.4). For the standard deviation model the EVR is 0.77 for the ROI and 0.35 for the fixed region (Table 4.3.5)

The EVR values for the skew model are 19 and 18 for the fixed regions and ROI respectively (Tables 4.3.6), which are much higher than the recommended limit of 0.20. This clearly indicates that the GLS regression is the preferred modeling choice over OLS for the skew model. An OLS model for the skew would have clearly given misleading results as it does not distinguish between the model error and sampling error. Importantly, what is clear is that if a method of moments estimator was used to estimate the model error variance σ_δ^2 for the skew model, the model error variance would have been grossly underestimated as the sampling error has heavily dominated the regional analysis. A more reasonable estimate of the model error variance has been achieved with the Bayesian procedure as it represents the values of σ_δ^2 by computing expectations over the entire posterior distribution. As far as the ROI is concerned there is little change in the EVR as compared to the fixed region (as the skew model tends to include more stations in the regional analysis).

For the fixed regions, the mean flood model has the model error variance of 27.7, which is

much higher than 5.6 (for the standard deviation model) and 1.22 (for the skew model) (Tables 4.3.4, 4.3.5 and 4.3.6). For the ROI, the mean flood model also shows a much higher model error variance than those of the standard deviation and skew models. These results indicate that the mean flood has the greater level of heterogeneity associated with it as compared to the standard deviation and skew.

The pseudo ANOVA tables were also prepared for the flood quantile models. Tables 4.3.2 and 4.3.3 show the results for Q_{20} and Q_{100} models, respectively. Here the ROI shows a higher EVR than the fixed region. It can also be clearly seen that the model error (δ) terms for ROI of tables 4.3.2 and 4.3.3 are smaller than that of the fixed region. This is due to the fact that ROI has found an optimum number of sites based on the minimum model error variance which naturally uses a smaller number of sites than that of the fixed region. This suggests that sub regions may exist in larger regions and that the BGLS regression should be used with ROI in developing the flood quantile models.

`

Table 4.3.2 Pseudo ANOVA table for Q_{20} model (QRT, fixed region and ROI NSW)

Source	**Degrees of Freedom**		**Sum of Squares**		
	Fixed region	ROI	*Equations*	Fixed region	ROI
Model	*k=3*	*k=3*	$n(\sigma^2_{\delta 0} - \sigma^2_{\delta}) =$	61.1	61.1
Model error δ	*n-k-1=92*	*n-k-1=48*	$n(\sigma^2_{\delta}) =$	23.5	17.3
Sampling error η	*N = 96*	*N = 52*	$tr[\Sigma(\hat{y})] =$	7.6	7.0
Total	*2n-1 = 191*	*2n-1 = 103*	**Sum of the above =**	**92**	**86**
			EVR	**0.32**	**0.43**

Table 4.3.3 Pseudo ANOVA table for Q_{100} (QRT, fixed region and ROI NSW)

Source	**Degrees of Freedom**		**Sum of Squares**		
	Fixed region	ROI		Fixed region	ROI
Model	*k=3*	*k=3*		50.0	50.0
Model error δ	*n-k-1=92*	*n-k-1=51*		33.6	26.1
Sampling error η	*N = 96*	*N = 55*		10.9	10.0
Total	*2n-1 = 191*	*2n-1 = 109*	**Sum of the above =**	**95**	**87**
			EVR	**0.32**	**0.42**

Table 4.3.4 Pseudo ANOVA table for the mean flood model (PRT, fixed region and ROI NSW

Source	**Degrees of Freedom**		**Sum of Squares**		
	Fixed region	ROI		Fixed region	ROI
Model	*k=3*	*k=3*		61.5	61.2
Model error δ	*n-k-1=92*	*n-k-1=32*		27.7	16.5
Sampling error η	*N = 96*	*N = 36*		5	4.5
Total	*2n-1 = 191*	*2n-1 = 71*	**Sum of the above =**	**94**	**83**
			EVR	**0.17**	**0.3**

Table 4.3.5 Pseudo ANOVA table for the standard deviation model (PRT, fixed region and ROI NSW)

Source	**Degrees of Freedom**		**Sum of Squares**		
	Fixed region	ROI		Fixed region	ROI
Model	*k=3*	*k=3*		3.1	3.1
Model error δ	*n-k-1=92*	*n-k-1=43*		5.6	4.4
Sampling error η	*N = 96*	*N = 47*		3.6	3.4
Total	*2n-1 = 191*	*2n-1 = 93*	**Sum of the above =**	**12**	**11**
			EVR	**0.35**	**0.77**

Table 4.3.6 Pseudo ANOVA table for the skew model (PRT, fixed region and ROI NSW)

Source	**Degrees of Freedom**		**Sum of Squares**		
	Fixed region	ROI		Fixed region	ROI
Model	*k=3*	*k=3*		0.1	0.1
Model error δ	*n-k-1=92*	*n-k-1=91*		1.22	1.21
Sampling error η	*N = 96*	*N = 95*		24	23
Total	*2n-1 = 191*	*2n-1 = 189*	**Sum of the above =**	**25**	**23**
			EVR	**19**	**18**

To assess the underlying model assumptions (i.e. the normality of residuals), the plots of the standardised residuals vs. predicted values were examined. The predicted values were obtained from one-at-a-time cross validation. Figures 4.3.3 to 4.3.5 show the plots for the Q_{20} and the mean flood models with the fixed region and ROI. If the underlying model

assumption is satisfied to a large extent the standardised residual values should not exceed ± 2 limits; in practice, 95% of the standardised residuals should fall between ± 2. The results in Figures 4.3.3 to 4.3.5 reveal that the developed equations satisfy the normality of residual assumption quite satisfactorily. Also no specific trend (heteroscedasicity) can be identified with the standardised values being almost equally distributed below and above zero. Similar results were obtained for the skew, standard deviation and other flood quantile models.

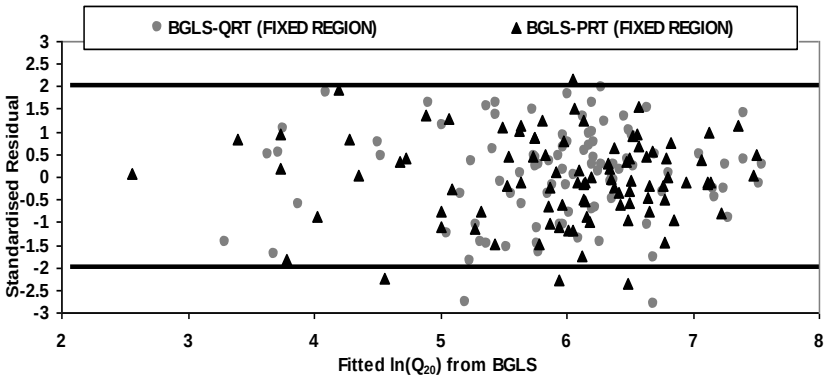


Figure 4.3.3 plots of standardised residuals vs. predicted values for ARI of 20 years (QRT and PRT, fixed region, NSW)

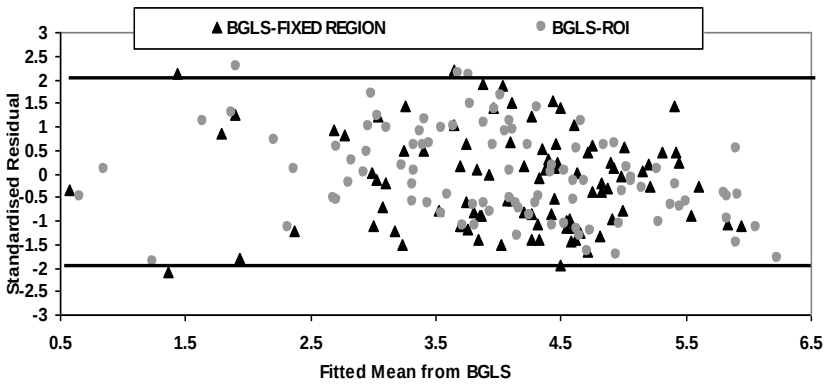


Figure 4.3.4 plot of standardised residuals vs. predicted values for the mean flood (PRT, fixed region, ROI, NSW)

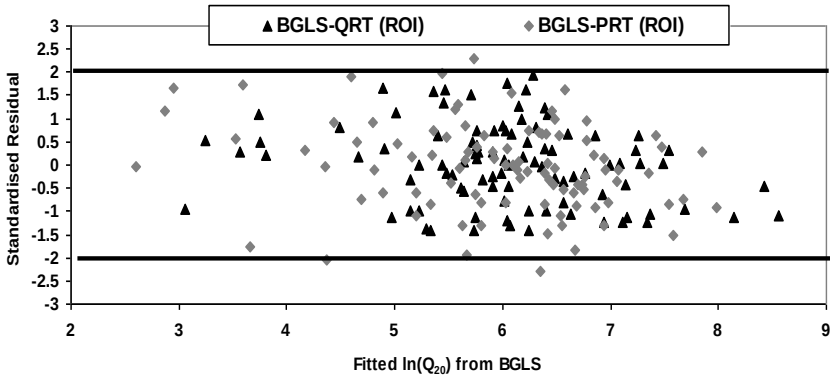


Figure 4.3.5 plots of standardised residuals vs. predicted values for ARI of 20 years (QRT and PRT, ROI, NSW)

The QQ-plots of the standardised residuals (Equation 4.2.13) vs. normal score (Equation 4.2.14) for the fixed region (based on one-at-a-time cross validation) and ROI were examined. Figures 4.3.6 to 4.3.8 present results for the Q_{20} (fixed region and ROI) and skew (ROI) models, which show that all the points closely follow a straight line. This indicates that the assumption of normality and the homogeneity of variance of the standardised residuals have largely been satisfied. If the standardised residuals are indeed normally and independently distributed N(0,1) with mean 0 and variance 1 then the slope of the best fit line in the QQ-plot, which can be interpreted as the standard deviation of the normal score (Z score) of the quantile, should approach 1 and the intercept, which is the mean of the normal score of the quantile should approach 0 as the number of sites increases. Figures 4.3.6 to 4.3.8 indeed show that the fitted lines for the developed models pass through the origin (0, 0) and it has a slope approximately equal to one. The ROI approach approximates the normality of the residuals slightly better (i.e. a better match with the fitted line) than the fixed region approach. Similar results were also found for the mean, standard deviation and other flood quantile models.

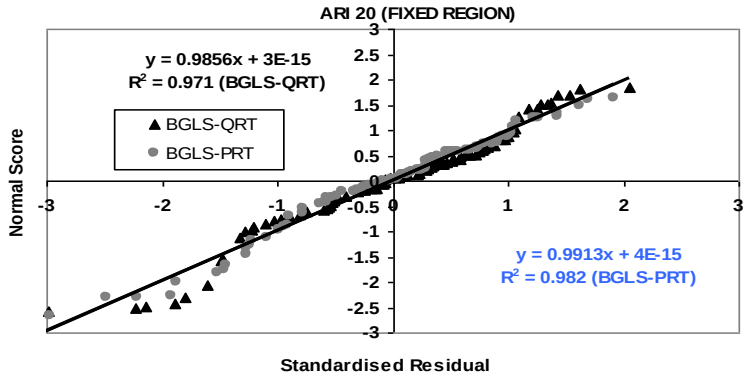


Figure 4.3.6 QQ-plot of the standardised residuals vs. Z score for ARI of 20 years (QRT and PRT, fixed region, NSW)

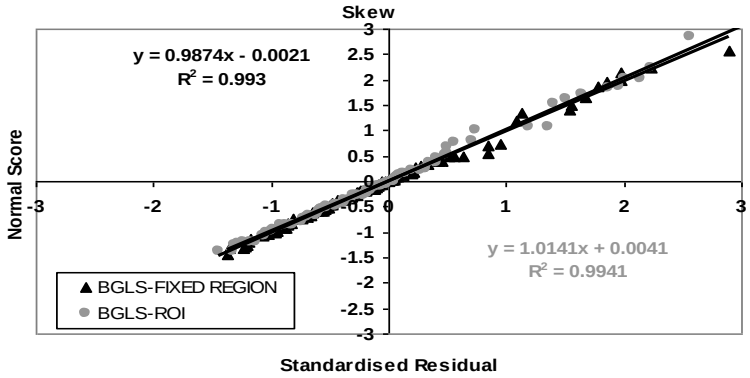


Figure 4.3.7 QQ-plot of the standardised residuals vs. Z score for the skew model (PRT, fixed region, ROI, NSW)

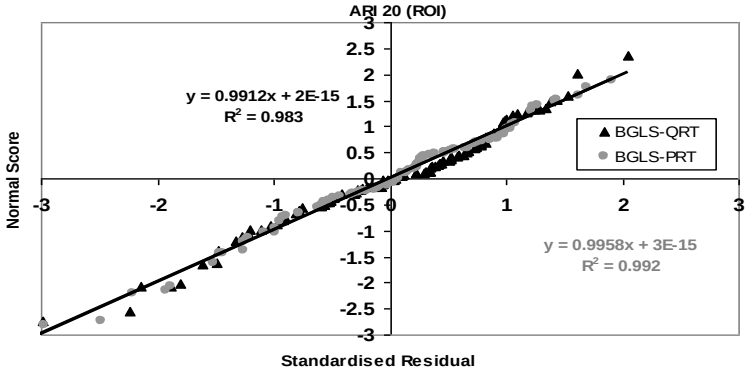


Figure 4.3.8 QQ-plot of the standardised residuals vs. Z score for ARI of 20 years (QRT and PRT, ROI, NSW)

The summary of various regression diagnostics (the relevant equations are described in Section 4.2.2) is provided in Table 4.3.7. This shows that for the mean flood model, the model error variance (MEV) and average standard error of prediction (SEP) are much higher than those of the standard deviation and skew models. This indicates that the mean flood model exhibits a higher degree of heterogeneity than the standard deviation and skew models. Indeed the issue here is that sampling error becomes larger as the order of the moment increases, therefore in case of the skew the spatial variation is a second order effect that is not really detectable. The R^2_{GLS} value for the mean flood model with the ROI is 8% higher than the fixed region. These indicate that the ROI should be preferred over the fixed region for developing the mean flood model. For the standard deviation model, ROI also shows 2% smaller SEP and 9% higher R^2_{GLS} values. This indicates that the ROI is preferable over the fixed region for the standard deviation model. The SEP and R^2_{GLS} values for the skew model are very similar for the fixed region and ROI as the number of sites in the fixed region and ROI is very similar.

Interestingly one can see from Table 4.3.7 that the SEP values for all the flood quanitle models are 5% to 11% smaller for the ROI cases than the fixed region. Also, the R^2_{GLS} values for ROI cases are 4% to 7% higher than the fixed region. These show that the ROI approach performs better overall than the fixed region approach.

Table 4.3.7 Regression diagnostics for fixed region and ROI for NSW

Model	Fixed region				ROI			
	MEV	AVP	SEP (%)	R^2_{GLS} (%)	MEV	AVP	SEP (%)	R^2_{GLS} (%)
Mean	0.29	0.31	60	76	0.19	0.23	51	84
Stdev	0.058	0.062	25	37	0.046	0.054	23	46
Skew	0.013	0.024	16	65	0.013	0.023	16	65
Q_2	0.31	0.33	63	77	0.20	0.24	52	84
Q_5	0.23	0.24	52	79	0.16	0.20	47	85
Q_{10}	0.23	0.24	52	79	0.16	0.20	46	85
Q_{20}	0.25	0.27	55	76	0.18	0.22	49	83
Q_{50}	0.35	0.37	66	70	0.25	0.28	56	74
Q_{100}	0.35	0.38	68	65	0.29	0.34	63	70

Table 4.3.8 shows number of sites in a region and associated model error variances for the ROI and fixed region models. This shows that the ROI mean flood model has fewer sites on average (36 out of 96 i.e. 37% of the available sites) than the standard deviation and skew models. The ROI skew model has the highest number of sites which includes nearly all the sites in NSW. The model error variance for the fixed region mean flood model is 34% higher than the corresponding ROI model. The model error variances for all the ROI models (except the skew model) are smaller than the fixed region models. This shows that the fixed region models experience a greater heterogeneity than the ROI. If the fixed region models are made too big, the model error will be inflated by heterogeneity unaccounted for by the catchment characteristics. Figure 4.3.9 shows the resulting sub-regions in NSW (with minimum model error variances) for the ROI mean flood and skew models. For the mean flood model, there are distinct sub-regions while the sub-region for the skew model captures the entire study area. Similar results were found by Hackelbusch et al. (2009). The significance of this finding is that if sub-regions do exist they are most likely to be captured by the ROI.

Table 4.3.8 Model error variances associated with fixed region and ROI for NSW (*n* = number of sites of the parameters and flood quantiles

Parameter/ Quantiles	Mean	Stdev	Skew	Q_2	Q_5	Q_{10}	Q_{20}	Q_{50}	Q_{100}
ROI (*n*)	36	47	95	31	42	48	52	53	55
$\hat{\sigma}_{\delta}^2$	0.19	0.046	0.013	0.20	0.16	0.16	0.18	0.25	0.29
Fixed region (*n*)	96	96	96	96	96	96	96	96	96
$\hat{\sigma}_{\delta}^2$	0.29	0.058	0.013	0.21	0.23	0.23	0.25	0.35	0.35

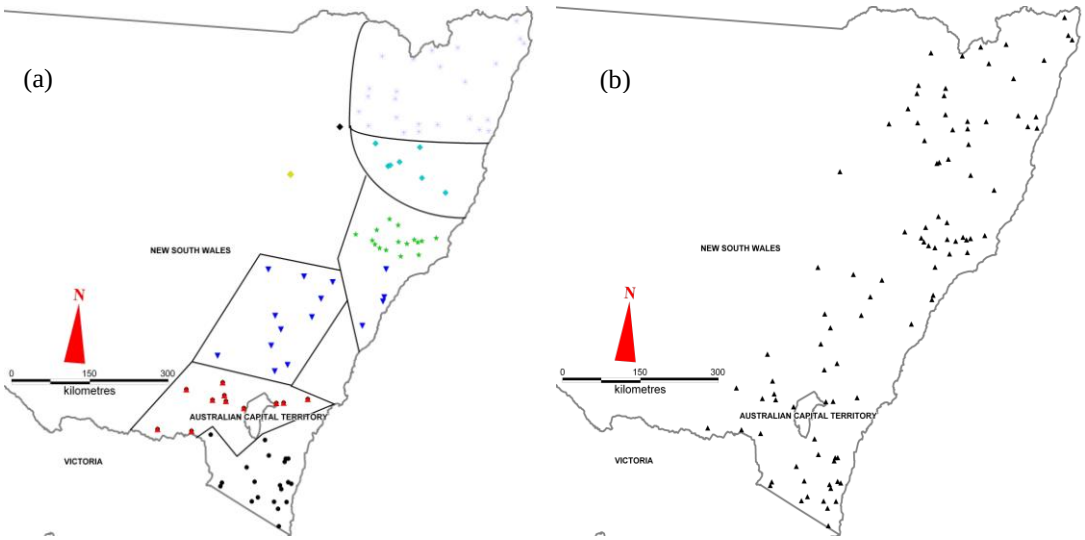


Figure 4.3.9 Binned minimum model error variance for (a) mean flood model and (b) skew model for NSW

Table 4.3.9 presents the relative root mean square error (RMSE) (Equation 4.2.16) and the relative error (RE) (Equation 4.2.15) values for the PRT and QRT models with both the fixed region and ROI. In terms of RMSE, ROI gives smaller values than the fixed regions for all the ARIs. The QRT-ROI shows smaller RMSE values than the PRT-ROI for all the ARIs except for the 5 years. In terms of RE, ROI gives 1 to 8% smaller values than the fixed regions. The PRT-ROI gives smaller values of RE (by 2% to 7%) for ARIs of 2 to 10 years than the QRT-ROI. However, the QRT-ROI gives smaller RE values (by 1% to 3%) for ARIs of 20 to 100 years than the QRT-ROI. These statistics reveal that there are only modest differences between the performances of QRT and PRT.

Tables 4.3.10 and 4.3.11 show results of counting the Q_{pred}/Q_{obs} ratios for the QRT and PRT. The use of this ratio has been discussed in Section 4.2.4. It was found that ROI provided relatively better results on average overall the ARIs than the fixed regions. For QRT, 82% compared to 74% were in the desirable range (ROI and fixed, respectively). For PRT, 78% of cases for ROI compared to 77% for fixed were in the 'desirable range' of estimates. Indeed it can be seen that the PRT-ROI methods are very similar, as would be expected, with only an average of 4% difference between the methods. The ROI-PRT showed 12% underestimation as compared to 9% for the QRT. The cases for overestimation were very similar: 10% for PRT-ROI compared to 9% for QRT-ROI. The evaluation statistics show that the PRT does not perform worse that the QRT and that the PRT is a viable option for design flood estimation for NSW. These results are in agreement with the results in Tables 4.3.9.

Table 4.3.9 Evaluation statistics (RMSE and RE) from one-at-a-time cross validation for NSW

Model	RMSE (%)				RE (%)			
	PRT		QRT		PRT		QRT	
	Fixed region	ROI	Fixed region	ROI	Fixed region	ROI	Fixed region	ROI
Q_2	7.3	6.2	6.8	5.9	46	38	44	40
Q_5	6.5	5.4	7.0	5.9	37	30	38	36
Q_{10}	6.7	5.6	7.4	5.5	37	29	37	36
Q_{20}	7.2	5.7	8.3	5.3	36	34	35	31
Q_{50}	8.1	7.0	10.0	6.7	38	34	36	32
Q_{100}	9.0	7.5	10.0	7.2	40	36	38	35

Table 4.3.10 Summary of counts based on Q_{pred}/Q_{obs} ratio values for QRT and PRT for NSW (fixed region). “U” = gross underestimation, “D” = desirable and “O” = gross overestimation

ARI (years)	Count (QRT)			Percent (QRT)			Count (PRT)			Percent (PRT)		
	U	D	O	U	D	O	U	D	O	U	D	O
2	16	70	10	17	73	10	10	69	17	10	72	18
5	8	77	11	8	80	11	8	77	11	8	80	11
10	12	72	12	13	75	13	10	76	10	10	79	10
20	12	72	12	13	75	13	11	77	8	11	80	8
50	5	72	19	5	75	20	12	75	9	13	78	9
100	17	65	14	18	68	15	15	72	9	16	75	9
Sum / average	**70**	**428**	**78**	**12**	**74**	**14**	**66**	**446**	**64**	**11**	**77**	**11**

Table 4.3.11 Summary of counts based on Q_{pred}/Q_{obs} ratio values for QRT and PRT for NSW (ROI). "U" = gross underestimation, "A" = acceptable and "O" = gross overestimation

ARI (years)	Count (QRT)			Percent (QRT)			Count (PRT)			Percent (PRT)		
	U	D	O	U	D	O	U	D	O	U	D	O
2	10	77	9	10	80	9	10	76	10	10	79	10
5	8	80	8	8	83	8	10	78	8	10	81	8
10	11	79	6	11	82	6	10	78	8	10	81	8
20	10	80	6	10	83	6	12	77	7	13	80	7
50	5	81	10	5	84	10	13	71	12	14	74	13
100	6	76	14	6	79	15	13	71	12	14	74	13
Sum / average	**50**	**473**	**53**	**9**	**82**	**9**	**68**	**451**	**57**	**12**	**78**	**10**

4.3.2 PRT-ROI with constant SD and skew models

Below we present the results of the region-of-influence (ROI) analysis for the parameter regression technique (PRT) using constant standard deviation and skew (i.e. no predictor variables in the regression equation of the standard deviation and skew). The main aspect of this analysis is to determine if there is any loss in accuracy and efficiency, especially in the mid to higher ARIs (i.e 20 to 100 years), when using a constant standard deviation and skew as compared to models with explanatory variables.

Firstly, in Figure 4.3.10, we present the standardised residual vs. the fitted quantile plot of the 20 year ARI, superimposing the estimate made by the PRT-ROI with constant standard deviation and skew on the estimates by the previous QRT-ROI and PRT-ROI models.. Indeed one can observe that the PRT-ROI estimate of the 20 year ARI with constant standard deviation and skew performs equally well as the competing models. Nearly all the standardised residuals fall within the $\pm$ 2 limits, suggesting that the use of explanatory variables does not really add any more meaningful information to the analysis. Secondly we show the QQ-plot (Figure 4.3.11) of the competing models which shows that the use of a constant standard deviation and skew does not result in any gross errors. The residual analysis also reveals that the major assumptions of the regression have been largely satisfied (i.e. normality of the residuals).

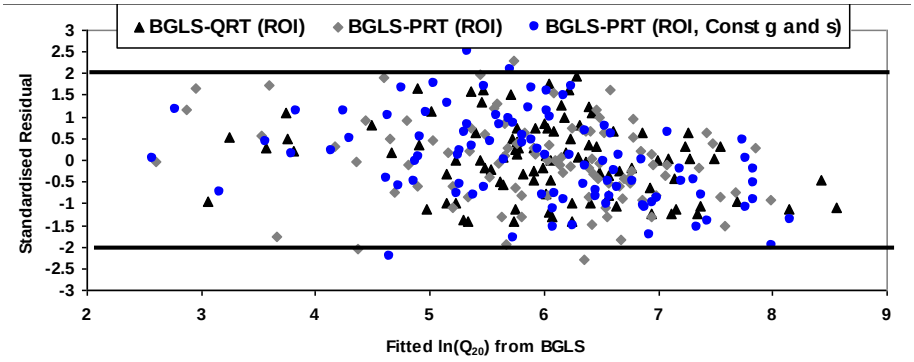


Figure 4.3.10 Plots of standardised residuals vs. predicted values for ARI of 20 years (QRT and PRT, ROI and PRT-ROI with constant standard deviation and skew, NSW)

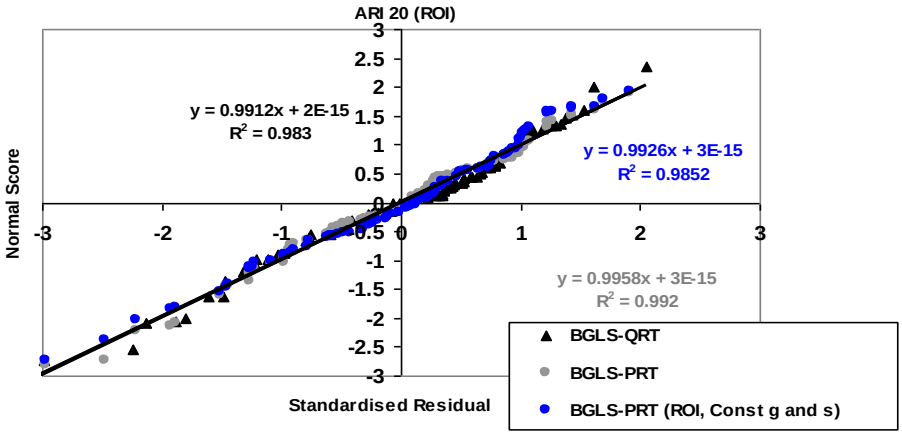


Figure 4.3.11 QQ-plot of the standardised residuals vs. Z score for ARI of 20 years (QRT and PRT, ROI, and PRT-ROI with constant standard deviation and skew, NSW)

We now present the results based on the evaluation statistics (i.e. Equation 4.2.16 and 4.2.15) to compare the flood quantiles from PRT-ROI using a constant standard deviation and skew to PRT using a standard deviation and skew as a function of catchment variables. The evaluation statistics (see Table 4.3.12) from the validation reveal that there is no real loss of accuracy (as compared to at-site flood quantiles) if a constant standard deviation and skew model is adopted to estimate the flood quantiles up to the 20 years ARI. The results at the higher ARIs (50 and 100 years) show that using a constant standard deviation and skew may affect the results slightly. The larger ARI estimation may require further information

which may be provided by the explanatory variables for the estimation of standard deviation and skew. This needs further investigation, which is left for future research.

Table 4.3.12 Evaluation statistics (RMSE and RE) from one-at-a-time cross validation for NSW. Blue colour indicates the results where the quantiles are estimated using constant standard deviation and skew (ROI(C) indicates region-of-influence model with regional average constant SD and skew, i.e. the SD and skew models do not have any predictor variables)

Model	RMSE (%)			RE (%)		
		PRT			PRT	
	Fixed region	ROI	**ROI (C)**	Fixed region	ROI	**ROI (C)**
Q_2	7.3	6.2	**6.3**	46	38	**37**
Q_5	6.5	5.4	**5.9**	37	30	**32**
Q_{10}	6.7	5.6	**6.0**	37	29	**33**
Q_{20}	7.2	5.7	**6.3**	36	34	**34**
Q_{50}	8.1	7.0	**7.7**	38	34	**35**
Q_{100}	9.0	7.5	**8.5**	40	36	**39**

Finally we consider the results of the counting of the Q_{pred}/Q_{obs} ratios. This reveals that there is no notable difference between the PRT-ROI (that utilises standard deviation and skew predicted from regression models) and PRT-ROI that utilises a constant standard deviation and skew. The difference between the two methods in the number of estimates in the desirable range is only 1% on average (i.e. 78% and 77%). The numbers in the gross overestimation and gross underestimation range show little difference on average. There seems to be little loss in accuracy in using a constant standard deviation and skew; however, further investigation using data from the other states is needed before recommending this type of model.

4.4 Results for Victoria

A total of 131 catchments were used from Victoria for the analyses presented here. These catchments are listed in Appendix (Table A2). The record lengths of the annual maximum flood series of these 131 stations range from 26 to 52 years (mean: 33 years, median: 33 years and standard deviation: 4.6 years). The distribution of record lengths is shown in Figure 2.4. The catchment areas of the selected 131 catchments range from 3 km^2 to 997 km^2 (mean: 321 km^2 and median: 289 km^2). The geographical distribution of the selected 131 catchments is shown in Figure 2.5. The distribution of catchment areas of these stations is shown in Figure 2.6.

In the fixed region approach, all the 131 catchments were considered to have formed one region, however, one catchment was left out for cross-validation and the procedure was repeated 131 times to implement one-at-a-time cross validation. In the region-of-influence (ROI) approach, an optimum region was formed for each of the 131 catchments by starting with 15 stations and then consecutively adding 5 stations at each iteration.

Table 4.4.1 shows different combinations of predictor variables for the Q_{10} QRT model and the models for the first three moments of the LP3 distribution. Figures 4.4.1 and 4.4.2 show example plots of the statistics used in selecting the best set of predictor variables for Q_{10} and the skew models. According to the model error variance, combinations 8, 11, 9, 3, 21, 13, 7, 12, 4, 21, 5, 23 and 10 were potential sets of predictor variables for the Q_{10} model. Combinations 21, 23, 25 and 23 contained 3 to 4 predictor variables while the rest of the combinations contained 2 predictor variables with very similar model error variances and R^2_{GLS}. The AVPO, AVPN, AIC and BIC values favoured combinations 2, 3 and 7. However combination 10 which included area and design rainfall intensity $I_{tc,10.}$ was finally selected as the best set of predictor variables as it has regression coefficients showing 4 times the posterior standard deviation away from zero as compared to combinations 2, 3, and 7. These set of predictor variables were also found significant for NSW.

For the skew model, combination 13 showed the lowest model error variance and the highest R^2_{GLS} (see Figure 4.4.2) as well as the lowest AIC and BIC values. The comparison with combination 1 (having no explanatory variables i.e. regional mean model) showed a higher AVPO and AVPN and a higher standard error in the model error variance estimate for combination 1 as compared to combination 13. Combination 13 certainly shows an improvement over all the combinations and hence was selected. A similar procedure was

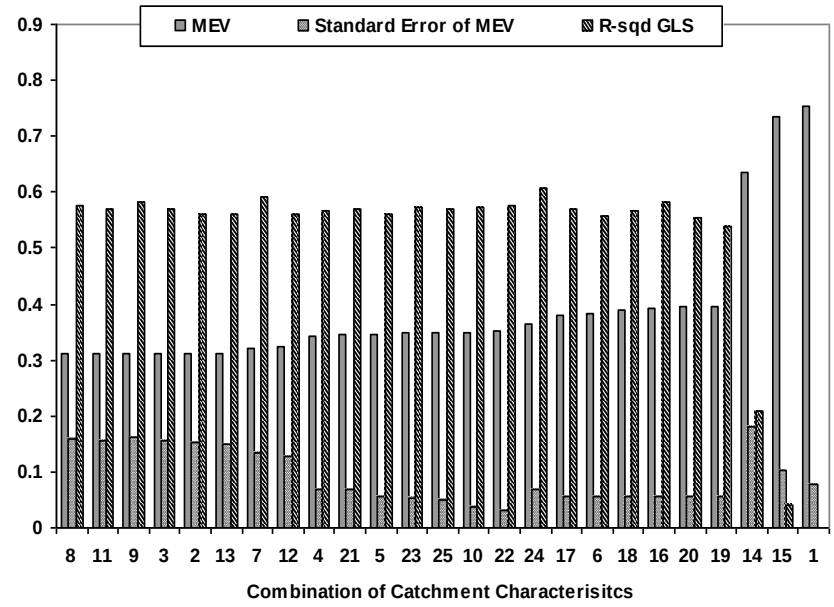
adopted in selecting the best set of predictor variables for other models with the QRT and PRT. The sets of predictor variables selected as above were used in the one-at-a-time cross validation with the fixed regions and region-of-influence (ROI) approaches.

The significance of the estimated regression coefficient values shown in Equations 4.4.1 to 4.4.9 was evaluated using the Bayesian plausibility value (BPV) as described by Reis et al. (2005) and Gruber et al. (2007). The BPVs for the regression coefficients associated with variable area and design rainfall intensity $I_{tc,ARI}$ for the QRT for the ARIs of 2 – 20 years were smaller than 0%, while for the ARIs of 50 and 100 years the BPVs were less than 0% for area and less than 0.05% for design rainfall intensity $I_{tc,ARI}$. The BPVs for the skew model were 0% and 1% for rain and evap, respectively indicating that these are reasonably good predictors for skew in Victoria. The BPVs for the mean flood model were 0% for both the predictor variables (area and $^2I_{12}$). For the standard deviation model, the BPVs for the predictor variables rain and evap were 0.19% and 0.86% respectively.

Regression models developed for the QRT and PRT for the fixed region are given by Equations 4.4.1 to 4.4.9.

Table 4.4.1 Different combinations of predictor variables considered for the QRT models and the parameters of the LP3 distribution (QRT and PRT fixed region Victoria)

Combination	Combinations for mean, standard deviation & skew models	Combinations for flood quantile model
1	Const	Const
2	Const, area	Const, area
3	Const, area, $^{2}I_{1}$	Const, area, $^{2}I_{1}$
4	Const, area, $^{50}I_{12}$	Const, area, $^{2}I_{12}$
5	Const, area, sden	Const, area, $^{50}I_{12}$
6	Const, area, $^{2}I_{12}$	Const, area, sden
7	Const, area, evap	Const, area, evap
8	Const, area, rain	Const, area, rain
9	Const, area, QSA	Const, area, rain, QSA
10	Const, area, forest	Const, area, $I_{tc,ARI}$
11	Const, area, S1085	Const, area, QSA
12	Const, S1085, evap	Const, area, forest
13	Const, rain, evap	Const, area, S1085
14	Const, forest, rain	Const, S1085, evap
15	Const, QSA	Const, rain, evap
16	Const, evap, QSA	Const, area, sden, evap
17	-	Const, area, sden, S1085
18	-	Const, area, sden, evap, QSA
19	-	Const, area, sden , $I_{tc,ARI}$, QSA
20	-	Const, area, sden , $I_{tc,ARI}$, evap, QSA
21	-	Const, area, $I_{tc,ARI}$, S1085
22	-	Const, area, $I_{tc,ARI}$, forest
23	-	Const, area, $I_{tc,ARI}$, QSA
24	-	Const, area, $I_{tc,ARI}$, evap
25	-	Const, area, $^{2}I_{1}$, $I_{tc,ARI}$



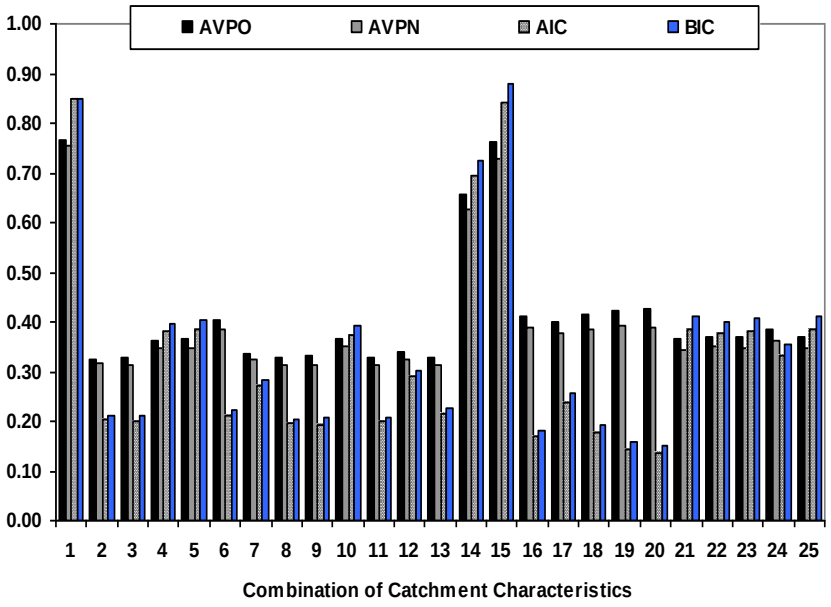


Figure 4.4.1 Selection of predictor variables for the BGLS regression model for Q_{10} (QRT, fixed region Victoria), MEV = model error variance, AVPO = average variance of prediction (old), AVPN = average variance of prediction (new) AIC = Akaike information criterion, BIC = Bayesian information criterion

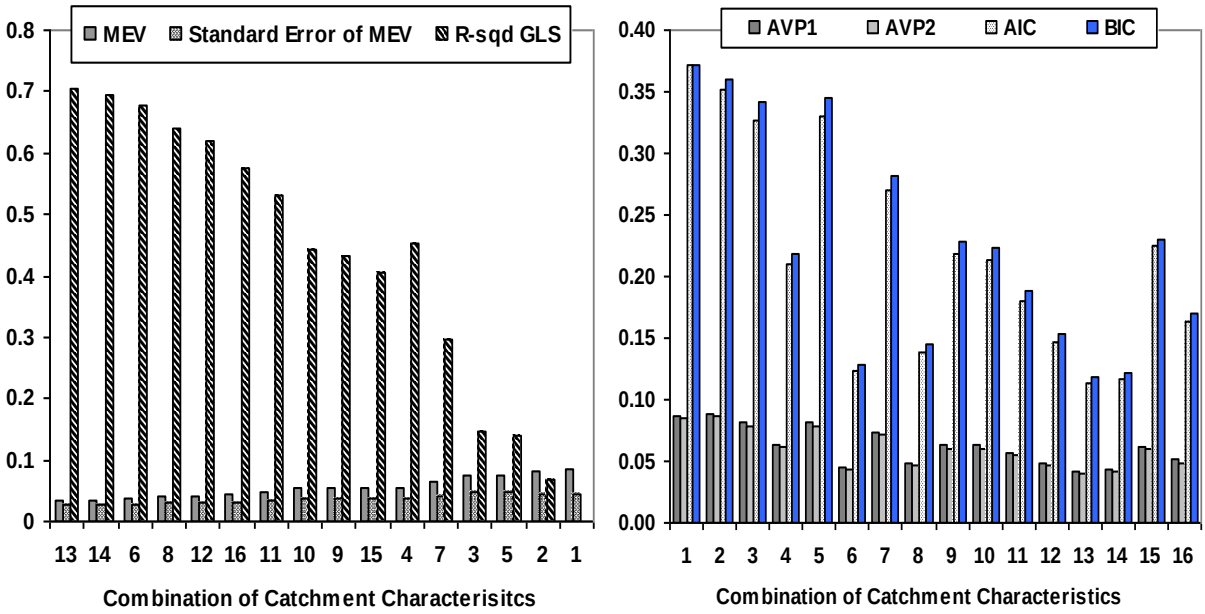


Figure 4.4.2 Selection of predictor variables for the BGLS regression model for skew

$$\ln(Q_2) = 3.38 + 0.90(\boldsymbol{z}area) + 1.35(\boldsymbol{z}I_{tc,2}) \quad (4.4.1)$$

$$\ln(Q_5) = 4.17 + 0.92(\boldsymbol{z}area) + 1.32(\boldsymbol{z}I_{tc,5}) \quad (4.4.2)$$

$$\ln(Q_{10}) = 4.55 + 0.94(\boldsymbol{z}area) + 1.42(\boldsymbol{z}I_{tc,10}) \quad (4.4.3)$$

$$\ln(Q_{20}) = 4.82 + 0.97(\boldsymbol{z}area) + 1.50(\boldsymbol{z}I_{tc,20}) \quad (4.4.4)$$

$$\ln(Q_{50}) = 5.17 + 0.99(\boldsymbol{z}area) + 1.62(\boldsymbol{z}I_{tc,50}) \quad (4.4.5)$$

$$\ln(Q_{100}) = 5.24 + 0.99(\boldsymbol{z}area) + 1.63(\boldsymbol{z}I_{tc,100}) \quad (4.4.6)$$

$$M = 3.22 + 0.61(\boldsymbol{z}area) + 1.50(\boldsymbol{z}I_{12,2}) \quad (4.4.7)$$

$$stdev = 1.16 - 0.83(\boldsymbol{z}rain) + 1.49(\boldsymbol{z}evap) \quad (4.4.8)$$

$$skew = -0.65 + 0.74(\boldsymbol{z}rain) - 3.25(\boldsymbol{z}evap) \quad (4.4.9)$$

where *z()* is explained by Equation 4.3.10.

It is reassuring to observe that the regression coefficients in the QRT set of equations vary in a regular fashion with increasing ARI.

The Pseudo Analysis of Variance (ANOVA) tables for the Q_{20} and Q_{100} models and the parameters of the LP3 distribution are presented in Tables 4.4.2 – 4.4.6 for the fixed regions and ROI approaches.

For the LP3 parameters, the sampling error increases as the order of moment increases i.e. the EVR increases with the order of the moment. The ROI shows a reduced model error variance (i.e. a reduced heterogeneity) as compared to the fixed regions. The model error

dominates the regional analysis for the mean flood in both the fixed region and ROI approach. The standard deviation models for both the fixed regions and ROI are dominated by the sampling error (i.e. sampling error variance 1.25 and 1.3 times the model error variance, see Table 4.4.5). In all cases ROI shows a higher EVR than the fixed regions e.g. for the mean flood model the EVR is 0.20 for the ROI and 0.16 for the fixed region (Table 4.4.4).

The EVR values for the skew model are 8.4 and 9.5 for the fixed regions and ROI respectively (Tables 4.4.6), which are much higher than the recommended limit of 0.20. This clearly indicates that the GLS regression is the preferred form of modeling, especially in the case of the skew and standard deviation models. The sampling error estimate has proved to be important, thus OLS and even a method of moment's GLS estimator would have certainly provided an unstable estimate of the model error variance σ_{δ}^{2}. Indeed the standard deviation and skew models would have been grossly underestimated the model error variance, as the sampling error has dominated the regional analysis. The BGLS analysis has proved to be superior in the handling of the uncertainty of the model error variance. Furthermore for the skew model, the ROI has included more sites than the mean and standard deviation models.

Pseudo ANOVA tables were also prepared for the flood quantile models. Tables 4.4.2 and 4.4.3 show the results for the Q_{20} and Q_{100} models, respectively. Here the ROI shows a higher EVR (nearly double) than the fixed region. This suggests that the BGLS regression should be the preferred modeling choice in combination with ROI in developing the flood quantile models.

Table 4.4.2 Pseudo ANOVA table for Q_{20} model for Victoria (QRT, fixed region and ROI)

Source	**Degrees of Freedom**		**Sum of Squares**		
	Fixed region	ROI	*Equations*	Fixed region	ROI
Model	*k=3*	*k=3*	$n(\sigma^2_{\delta 0} - \sigma^2_{\delta}) =$	45.2	45.2
Model error δ	*n-k-1=127*	*n-k-1=48*	$n(\sigma^2_{\delta}) =$	55.2	24.4
Sampling error η	*N = 131*	*N = 52*	$tr[\Sigma(\hat{y})] =$	7.4	7.2
Total	*2n-1 = 261*	*2n-1 = 103*	**Sum of the above =**	**108**	**77**
			EVR	**0.13**	**0.30**

Table 4.4.3 Pseudo ANOVA table for Q_{100} model for Victoria (QRT, fixed region and ROI)

Source	**Degrees of Freedom**		**Sum of Squares**		
	Fixed region	ROI		Fixed region	ROI
Model	*k=3*	*k=3*		29	29
Model error δ	*n-k-1=127*	*n-k-1=53*		77	40
Sampling error η	*N = 131*	*N = 57*		11	10
Total	*2n-1 = 261*	*2n-1 = 113*	**Sum of the above =**	**117**	**79**
			EVR	**0.14**	**0.25**

Table 4.4.4 Pseudo ANOVA table for the mean flood model for Victoria (PRT, fixed region and ROI)

Source	**Degrees of Freedom**		**Sum of Squares**		
	Fixed region	ROI		Fixed region	ROI
Model	*k=3*	*k=3*		46	45
Model error δ	*n-k-1=127*	*n-k-1=39*		37.5	28
Sampling error η	*N = 131*	*N = 43*		6.1	6
Total	*2n-1 = 261*	*2n-1 = 85*	**Sum of the above =**	**90**	**79**
			EVR	**0.16**	**0.2**

Table 4.4.5 Pseudo ANOVA table for the standard deviation model for Victoria (PRT, fixed region and ROI)

Source	**Degrees of Freedom**		**Sum of Squares**		
	Fixed region	ROI		Fixed region	ROI
Model	*k=3*	*k=3*		7.6	7.6
Model error δ	*n-k-1=127*	*n-k-1=76*		5.7	5.4
Sampling error η	*N = 131*	*N = 80*		7.1	6.8
Total	*2n-1 = 261*	*2n-1 = 159*	**Sum of the above =**	**20.3**	**20**
			EVR	**1.25**	**1.3**

Table 4.4.6 Pseudo ANOVA table for the skew model for Victoria (PRT, fixed region and ROI)

Source	**Degrees of Freedom**		**Sum of Squares**		
	Fixed region	ROI		Fixed region	ROI
Model	*k=3*	*k=3*		6.5	7.3
Model error δ	*n-k-1=127*	*n-k-1=113*		4.5	3.7
Sampling error η	*N = 131*	*N = 117*		38	35
Total	*2n-1 = 261*	*2n-1 = 233*	**Sum of the above =**	**49**	**48**
			EVR	**8.4**	**9.5**

We assessed the underlying model assumptions (i.e. the normality of residuals), by examining the plots of the standardised residuals vs. predicted values. The predicted values were obtained from one-at-a-time cross validation. Figures 4.4.3 to 4.4.5 show the plots for the Q_{20} and the skew models with the fixed region and ROI. It can be seen that most of the standardised residuals fall between the ± 2 limits, thus satisfying the normality of residuals to a large extent. What is noteworthy is that when ROI is used, both the QRT and PRT (for quantiles and skew estimates) provide a better approximation, with a greater proportion of the standardised residuals falling within the ± 2 limits. The results in Figures 4.4.3 to 4.4.5 reveal that the developed equations satisfy the normality of residual assumption quite satisfactorily. Also no specific pattern (heteroscedasicity) can be identified with the standardised values being almost equally distributed below and above zero; this is especially seen for the ROI case (see Figure 4.4.4 and 4.4.5). Similar results were obtained for the mean flood, standard deviation and other flood quantile models.

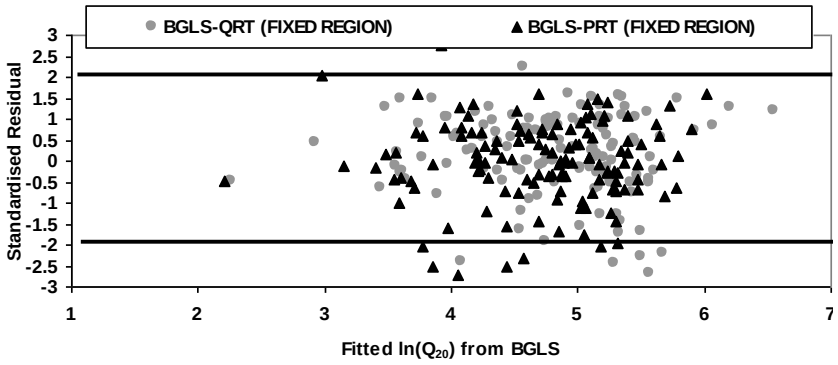


Figure 4.4.3 plots of standardised residuals vs. predicted values for ARI of 20 years (QRT and PRT, fixed region, Victoria)

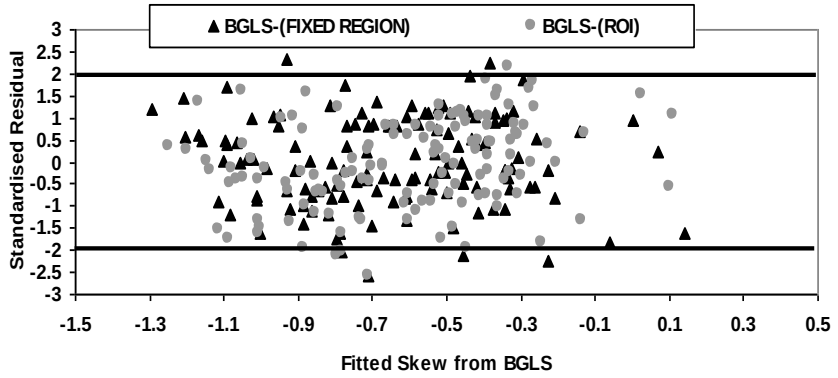


Figure 4.4.4 plot of standardised residuals vs. predicted values for the skew model (PRT, fixed region, ROI, Victoria)

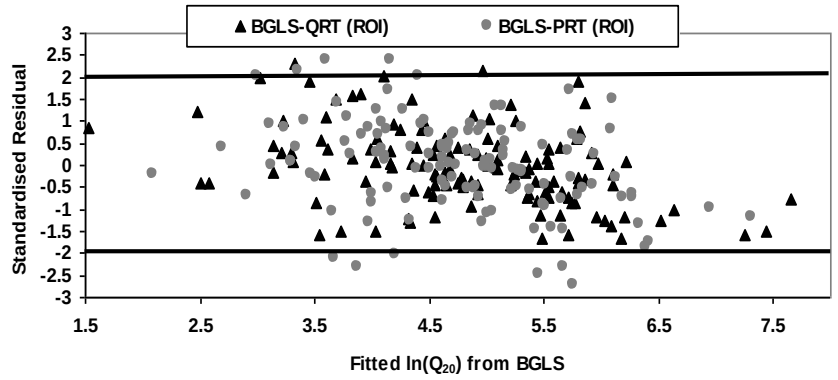


Figure 4.4.5 plots of standardised residuals vs. predicted values for ARI of 20 years (QRT and PRT, ROI, Victoria)

The QQ-plots of the standardised residuals (Equation 4.2.13) vs. normal score (Equation 4.2.14) for the fixed region (based on one-at-a-time cross validation) and ROI were

examined. Figures 4.4.6 to 4.4.8 present the results for the Q_{20} and skew models, which show that all the points closely follow a straight line. This indicates that the assumptions of normality and the homogeneity of variance of the standardised residuals have largely been satisfied. If the standardised residuals are indeed normally and independently distributed N(0,1) with mean 0 and variance 1 then the slope of the best fit line in the QQ-plot, which can be interpreted as the standard deviation of the normal score (Z score) of the quantile, should approach 1 and the intercept, which is the mean of the normal score of the quantile should approach 0 as the number of sites increases. Figures 4.4.6 to 4.4.8 indeed show that the fitted lines for the developed models pass through the origin (0, 0) and have a slope approximately equal to one. The ROI approach approximates the normality of the residuals slightly better (i.e. a better match with the fitted line) than the fixed region approach for both Q_{20} and skew models. Similar results were also found for the mean, standard deviation and other flood quantile models.

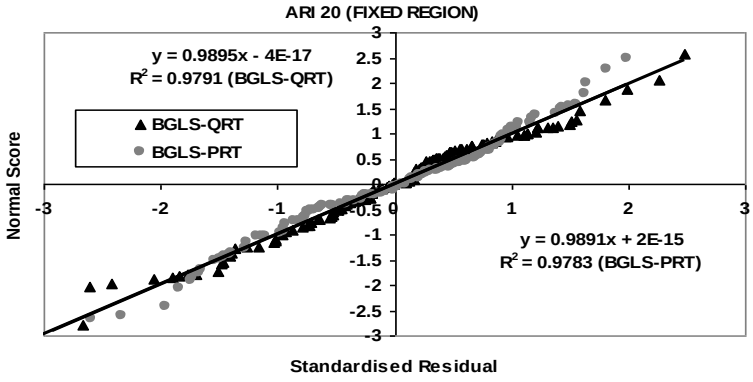


Figure 4.4.6 QQ-plot of the standardised residuals vs. Z score for ARI of 20 years (QRT and PRT, fixed region, Victoria)

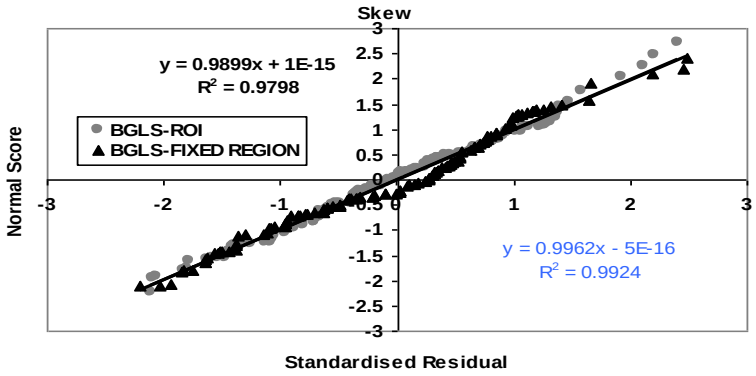


Figure 4.3.7 QQ-plot of the standardised residuals vs. Z score for the skew model (PRT, fixed region, ROI, Victoria)

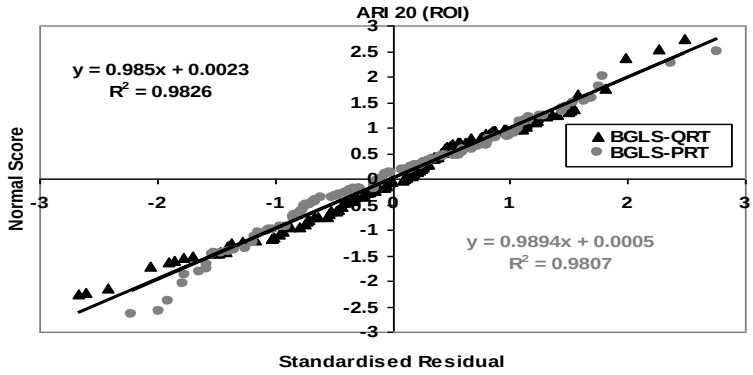


Figure 4.4.8 QQ-plot of the standardised residuals vs. Z score for ARI of 20 years (QRT and PRT, ROI, Victoria)

The summary of various regression diagnostics (the relevant equations are described in Section 4.2.2) is provided in Table 4.4.7. This shows that for the mean flood model, the model error variance (MEV) and average standard error of prediction (SEP) are much higher than those of the standard deviation and skew models. This indicates that the mean flood model exhibits a higher degree of heterogeneity than the standard deviation and skew models. Indeed the issue here is that sampling error becomes larger as the order of the moment increases, therefore, in case of the skew, the spatial variation is a second order effect that is not really detectable. The R^2_{GLS} value for the mean flood model with the ROI is only 1% higher than that for the fixed region, which is a negligible increase. These results indicate that the ROI should be preferred over the fixed region for developing the mean flood model, as this reduces the level of heterogeneity in the region. For the standard deviation model, ROI also shows 1% smaller SEP and similar R^2_{GLS} values. In this case the reduction in SEP% is negligible, however, even with this small improvement, ROI is still preferable over the fixed region for the standard deviation model. The SEP and R^2_{GLS} values for the skew model are very similar for the fixed region and ROI, as the number of sites in the fixed region and ROI is very similar.

Interestingly one can see from Table 4.4.7 that the SEP values for all the flood quanitle models are 6% to 27% smaller for the ROI cases than the fixed region. Also, the R^2_{GLS} values for ROI cases are 2% to 12% higher than the fixed region. These results demonstrate the superiority of the ROI approach over the fixed region analysis.

Table 4.4.7 Regression diagnostics for fixed region and ROI for Victoria

Model	Fixed region				ROI			
	MEV	AVP	SEP (%)	R^2_{GLS} (%)	MEV	AVP	SEP (%)	R^2_{GLS} (%)
Mean	0.29	0.31	60	62	0.21	0.23	46	63
Stdev	0.044	0.049	22	65	0.041	0.050	21	65
Skew	0.034	0.040	20	70	0.028	0.037	19	73
Q_2	0.27	0.28	57	63	0.20	0.23	51	65
Q_5	0.29	0.31	60	61	0.20	0.23	50	64
Q_{10}	0.35	0.37	67	57	0.23	0.26	54	61
Q_{20}	0.35	0.37	67	57	0.19	0.22	48	66
Q_{50}	0.47	0.49	80	49	0.27	0.32	61	61
Q_{100}	0.59	0.60	91	45	0.29	0.35	64	54

Table 4.4.8 shows number of sites in a region and associated model error variances for the ROI and fixed region models. This shows that the ROI mean flood model has fewer sites on average (43 out of 131 i.e. 33% of the available sites) than the standard deviation and skew models. The ROI skew model has the highest number of sites which includes nearly all the sites in Victoria (i.e. 117/131 = 90%). The model error variance for the fixed region mean flood model is 28% higher than the corresponding ROI model. The model error variances for all the ROI models in this case are smaller than the fixed region models. This shows that the fixed region models experience a greater heterogeneity than the ROI. If the fixed regions are made too large, the model error will be inflated by heterogeneity unaccounted for by the catchment characteristics. Figure 4.4.9 shows the resulting sub-regions in Victoria (with minimum model error variances) for the ROI Q_{20} and skew models. For the Q_{20} flood quantile and skew models, there is evidence of distinct sub-regions. For the Q_{20} flood quantile, three distinctive regions are formed from north to south Victoria (i.e. north, south and on the Great Dividing Range), while the sub-region for the skew model captures the east, middle and west regions with some overlapping stations, as expected. The significance of this finding is that if sub-regions do exist in a state they are most likely to be captured by the ROI.

Table 4.4.8 Model error variances of the parameters and flood quantiles associated with fixed region and ROI for Victoria (*n* = number of sites)

Parameter/ Quantiles	Mean	Stdev	Skew	Q_2	Q_5	Q_{10}	Q_{20}	Q_{50}	Q_{100}
ROI (*n*)	43	83	117	41	45	52	52	57	57
$\hat{\sigma}^2_\delta$	0.21	0.041	0.028	0.20	0.20	0.23	0.19	0.27	0.29
Fixed region (*n*)	131	131	131	131	131	131	131	131	131
$\hat{\sigma}^2_\delta$	0.29	0.044	0.034	0.27	0.29	0.35	0.35	0.47	0.59

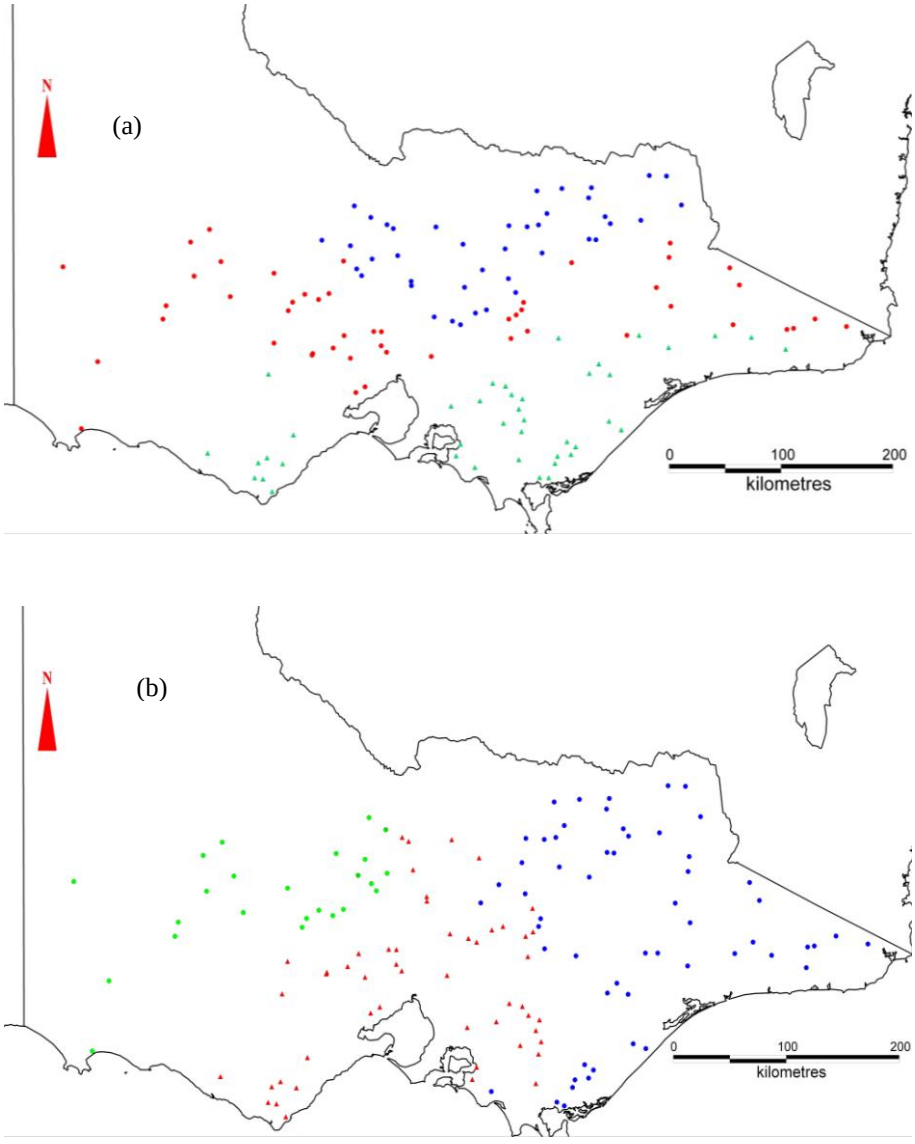


Figure 4.4.9 Binned minimum model error variance for Victoria (a) ARI = 20 flood quantile model and (b) skew model

Table 4.4.9 presents the root mean square error (RMSE) (Equation 4.2.16) and relative error (RE) (Equation 4.2.15) values for the PRT and QRT models with both the fixed region and ROI. In terms of RMSE, ROI gives smaller values than the fixed regions for all the ARIs. The QRT-ROI shows smaller RMSE values than the PRT-ROI for all the ARIs except for the 2 and 5 years. In terms of RE, ROI gives 1 to 5% smaller values than the fixed regions. The PRT-ROI gives smaller values of RE (by 1% to 2%) for ARIs of 2 to 10 years than the QRT-ROI. However, the QRT-ROI gives smaller RE values (by 1% to 5%) for ARIs of 20 to 100 years than the PRT-ROI. As with NSW, these results reveal only modest differences between the performances of the QRT and PRT.

Tables 4.4.10 and 4.4.11 show results of counting the Q_{pred}/Q_{obs} ratios for the QRT and PRT with both the ROI and fixed regions. For the QRT, it was found that the ROI and fixed region provided very similar results, with 76% and 75% of the ratios being in the desirable range on average. For the PRT-ROI, 73% of cases were in the desirable range, compared to 72% for the PRT-fixed region. The PRT-ROI showed 15% underestimation on average as compared to 12% for the QRT-ROI. The cases for overestimation were the same (12%) for both the methods. These results are in agreement with the results provided in Table 4.4.9.

Table 4.4.9 Evaluation statistics (RMSE and RE) from one-at-a-time cross validation for Victoria

Model	RMSE (%)				RE (%)			
	PRT		QRT		PRT		QRT	
	Fixed region	ROI	Fixed region	ROI	Fixed region	ROI	Fixed region	ROI
Q_2	5.6	5.5	7.7	6.8	38	37	37	37
Q_5	6.9	6.8	8.7	6.8	38	36	35	35
Q_{10}	8.2	8.0	10.7	6.9	37	37	36	35
Q_{20}	9.6	9.2	11.2	7.4	41	40	38	33
Q_{50}	11.5	11	11.3	9.5	41	40	41	40
Q_{100}	13	12.7	14	12	46	45	44	44

Table 4.4.10 Summary of counts/percentages based on Q_{pred}/Q_{obs} ratio values for QRT and PRT for Victoria (fixed region). “U” = gross underestimation, “D” = desirable range and “O” = gross overestimation

Model	Count (QRT)			Percent (QRT)			Count (PRT)			Percent (PRT)		
	U	D	O	U	D	O	U	D	O	U	D	O
Q_2	10	109	12	8	83	9	31	93	7	24	71	5
Q_5	12	104	15	9	79	11	22	99	10	17	76	8
Q_{10}	11	102	18	8	78	14	20	98	13	15	75	10
Q_{20}	16	96	19	12	73	15	23	94	14	18	72	11
Q_{50}	20	89	22	15	68	17	25	91	15	19	69	11
Q_{100}	23	89	19	18	68	15	24	88	19	18	67	15
Sum / average	**92**	**589**	**105**	**12**	**75**	**13**	**145**	**563**	**78**	**18**	**72**	**10**

Table 4.4.11 Summary of counts/percentages based on Q_{pred}/Q_{obs} ratio values for QRT and PRT for Victoria (ROI). “U” = gross underestimation, “D” = desirable range and “O” = gross overestimation

ARI (years)	Count (QRT)			Percent (QRT)			Count (PRT)			Percent (PRT)		
	U	D	O	U	D	O	U	A	O	U	D	O
2	15	103	13	11	79	10	18	96	17	14	73	13
5	10	106	15	8	81	11	15	102	14	11	78	11
10	14	106	11	11	81	8	15	100	16	11	76	12
20	16	100	15	12	76	11	19	96	16	15	73	12
50	20	93	18	15	71	14	25	90	16	19	69	12
100	20	91	20	15	69	15	26	90	15	20	69	11
Sum / average	**95**	**599**	**92**	**12**	**76**	**12**	**118**	**574**	**94**	**15**	**73**	**12**

4.5 Results for Tasmania

4.5.1 QRT and PRT – fixed and ROI approaches (Tasmania considered as a single region)

A total of 53 catchments were used from Tasmania for the analyses presented here. These catchments are listed in the appendix. The locations of these catchments are shown in Figure 2.11. The annual maximum flood series record lengths of these 53 stations range from 19 to 74 years (mean 30 years, median 28 years and standard deviation 10 years). The catchment areas of these 53 stations range from 1.3 to 1900 km^2 (mean 323 km^2, median 158 km^2 and standard deviation of 417 km^2).

In the fixed region approach, all the 53 catchments were considered to have formed one region, however, one catchment was left out for cross-validation and the procedure was repeated 53 times to implement one-at-a-time cross validation. In the region-of-influence (ROI) approach, an optimum region was formed for each of the 53 catchments by starting with 15 stations and then consecutively adding 1 station at each iteration.

Table 4.5.1 shows different combinations of predictor variables for the Q_{10} QRT model and the models for the first three moments of the LP3 distribution. Figure 4.5.1 and 4.5.2 show example plots of the statistics used in selecting the best set of predictor variables for the Q_{10} and skew models. According to the model error variance, combinations 6, 16, 18, 20, 17, 19 and 4 were potential sets of predictor variables for the Q_{10} model. Combinations 16, 18, 20, 17, 19 and 4 contained 3 to 4 predictor variables, while combinations 6 and 4 contained 2 predictor variables. Indeed combination 6 with the 2 predictor variables (area and design rainfall intensity ${}^{50}I_{12}$) showed the lowest model error variance and the highest R^2_{GLS}. The AVPO, AVPN, AIC and BIC values favour combination 6 as well. We also compared combination 6 to combination 10 (the latter also contains 2 predictor variables, area and design rainfall intensity $I_{tc,10}$). Combination 6 had a smaller model error variance while also showing the regression coefficient for variable ${}^{50}I_{12}$ to be 5.5 times the posterior standard deviation away from zero, as compared to 4 times for $I_{tc,10}$. Hence, combination 6 was finally selected as the best set of predictor variables for the Q_{10} model.

For the skew model, combination 4 showed the lowest model error variance (0.034) and the highest R^2_{GLS} (52%) (see Figure 4.5.2), as well as the lowest AIC and BIC. Combination 1 without any explanatory variables ranked 13 out of the 16 possible combinations (model error

variance of 0.045); it also showed higher AVPO and AVPN as compared to combination 4, hence combination 4 was finally selected.

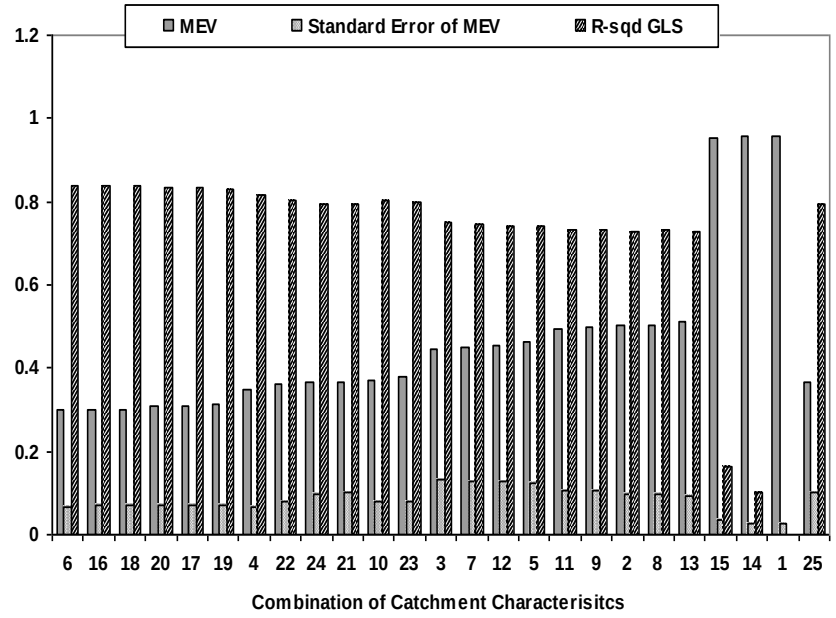
A similar procedure was adopted in selecting the best set of predictor values for other models with the QRT and PRT. The sets of predictor variables selected as above were used in the one-at-a-time cross validation with fixed regions and region-of-influence (ROI) approaches.

The BPV values for the regression coefficients associated with the QRT over all the ARIs were between 2% and 8% for the variable area and 0.000% for design rainfall intensity $^{50}I_{12.}$ This justifies the inclusion of predictor variables area and $^{50}I_{12}$ in the prediction equations for QRT. The BPVs for the skew model were 23% and 11% for area and $^{50}I_1$, respectively indicating these variables are not very good predictors for skew. The BPVs for the mean model were close to 1% for both the predictor variables. For the standard deviation model, the BPV for the predictor variable rain was 1%.

Regression equations developed for the QRT and PRT for the fixed region are given by Equations 4.5.1 to 4.5.9.

Table 4.5.1 Different combinations of predictor variables considered for the QRT models and the parameters of the LP3 distribution (QRT and PRT fixed region Tasmania)

Combination	Combinations for mean, standard deviation & skew models	Combinations for flood quantile model
1	Const	Const
2	Const, area	Const, area
3	Const, area, ($^{2}I_{1}$)	Const, area, $^{2}I_{1}$
4	Const, area, ($^{50}I_{1}$)	Const, area, $^{2}I_{12}$
5	Const, area, ($^{2}I_{12}$)	Const, area, $^{50}I_{1}$
6	Const, area, ($^{50}I_{12}$)	Const, area, $^{50}I_{12}$
7	Const, area, rain	Const, area, rain
8	Const, area, for	Const, area, for
9	Const, area, evap	Const, area, forest, evap
10	Const, area, S1085	Const, area, $I_{tc,ARI}$
11	Const, area, sden	Const, area, evap
12	Const, sden, rain	Const, area, S1085
13	Const, for, rain	Const, area, sden
14	Const, S1085, for	Const, sden, rain
15	Const, evap	Const, for, rain
16	Const, rain, evap	Const, area, $^{50}I_{12}$, rain
17	Const, rain	Const, area, $^{50}I_{12}$, sden
18	-	Const, area, $^{50}I_{12}$, rain, evap
19	-	Const, area, $^{50}I_{12}$, $I_{tc,ARI}$, evap
20	-	Const, area, $^{50}I_{12}$, $I_{tc,ARI}$, rain, evap
21	-	Const, area, $^{50}I_{12}$, $I_{tc,ARI}$, sden
22	-	Const, area, $^{50}I_{12}$, $I_{tc,ARI}$, S1085
23	-	Const, area, $I_{tc,ARI}$, evap
24	-	Const, area, $I_{tc,ARI}$, rain
25	-	Const, area, $^{2}I_{1}$, $I_{tc,ARI}$



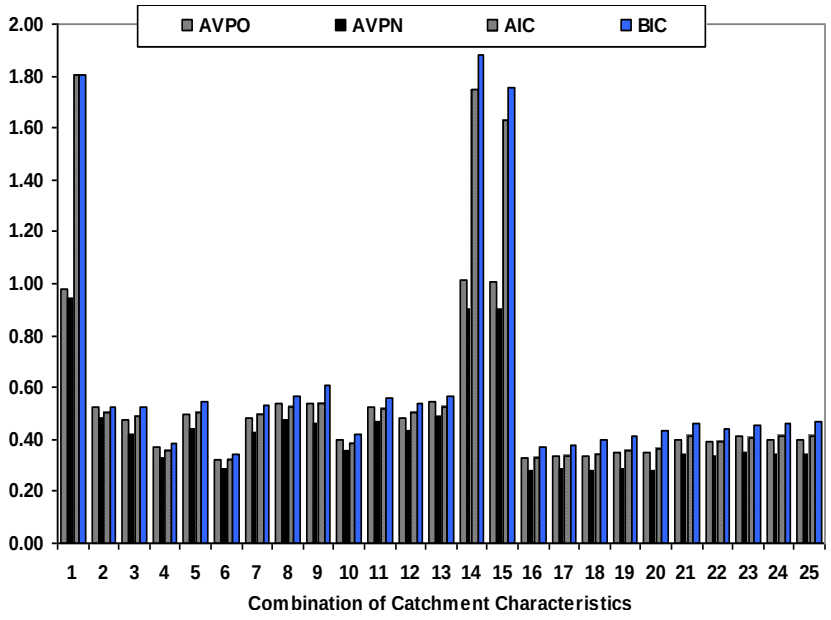


Figure 4.5.1 Selection of predictor variables for the BGLS regression model for Q_{10} (QRT, fixed region Tasmania), MEV = model error variance, AVPO = average variance of prediction (old), AVPN = average variance of prediction (new) AIC = Akaike information criterion, BIC = Bayesian information criterion

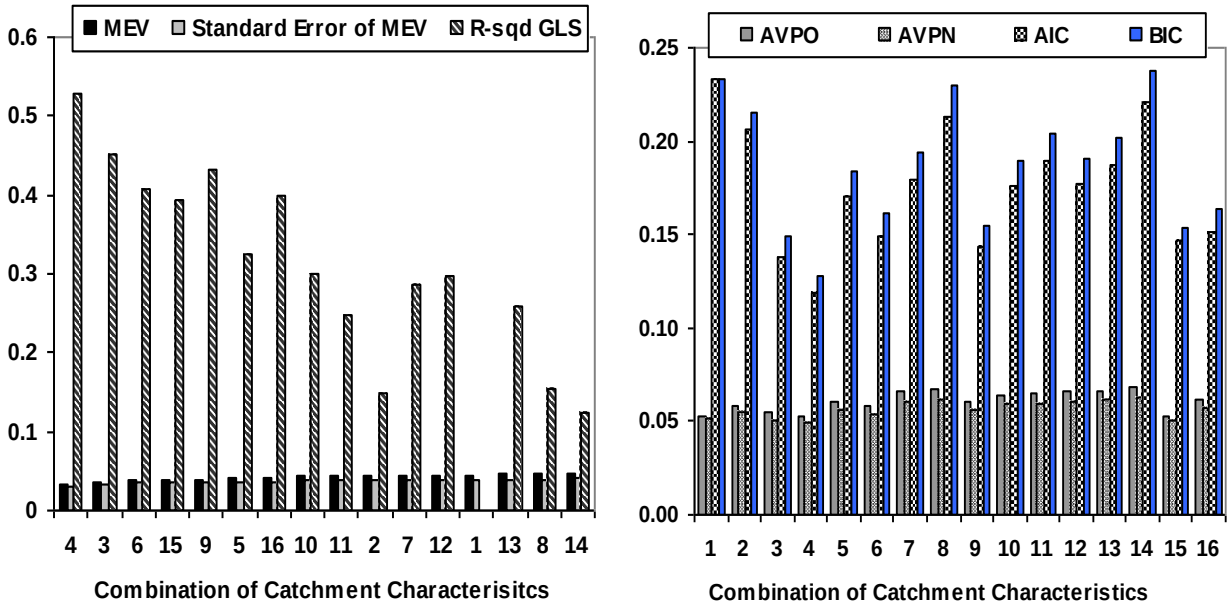


Figure 4.5.2 Selection of predictor variables for the BGLS regression model for skew

$$\ln(Q_2) = 4.18 + 0.91(\boldsymbol{z}area) + 3{,}35(\boldsymbol{z}I_{12,50}) \quad (4.5.1)$$

$$\ln(Q_5) = 4.59 + 0.89(\boldsymbol{z}area) + 2.80(\boldsymbol{z}I_{12,50}) \quad (4.5.2)$$

$$\ln(Q_{10}) = 4.87 + 0.85(\boldsymbol{z}area) + 2.57(\boldsymbol{z}I_{12,50}) \quad (4.5.3)$$

$$\ln(Q_{20}) = 5.09 + 0.84(\boldsymbol{z}area) + 2.39(\boldsymbol{z}I_{12,50}) \quad (4.5.4)$$

$$\ln(Q_{50}) = 5.45 + 0.84(\boldsymbol{z}area) + 2.23(\boldsymbol{z}I_{12,50}) \quad (4.5.5)$$

$$\ln(Q_{100}) = 5.48 + 0.82(\boldsymbol{z}area) + 2.02(\boldsymbol{z}I_{12,50}) \quad (4.5.6)$$

$$M = 4.00 + 0.90(\boldsymbol{z}area) + 3.85(\boldsymbol{z}I_{12,2}) \quad (4.5.7)$$

$$stdev = 0.64 + 0.55(\boldsymbol{z}rain) \quad (4.5.8)$$

$$skew = -0.05 + 0.07(\boldsymbol{z}area) + 1.20(\boldsymbol{z}I_{1,50}) \quad (4.5.9)$$

where $z()$ is explained by Equation 4.3.10.

It is reassuring to observe that the regression coefficients in the QRT set of equations vary in a regular fashion with increasing ARI.

The Pseudo Analysis of Variance (ANOVA) tables for the Q_{20} and Q_{100} models and the parameters of the LP3 distribution are presented in Tables 4.5.2 – 4.5.6 for the fixed regions and ROI. This is an extension of the ANOVA in the OLS regression which does not recognize and correct for the expected sampling variance (Reis et al., 2005).

For the LP3 parameters, the sampling error increases as the order of moment increases i.e. the EVR increases with the order of the moments. The ROI shows a reduced model error variance (i.e. a reduced heterogeneity) as compared to the fixed regions, as fewer sites have been used. The model error dominates the regional analysis for the mean flood and the standard deviation models for both the fixed regions and ROI. However, ROI shows a higher EVR than the fixed regions, e.g. for the mean flood model the EVR is 0.20 for the ROI and 0.06 for the fixed region (Table 4.5.4). For the standard deviation model the EVR is 0.66 for the ROI and 0.54 for the fixed region, which is a 12% increase in EVR (Table 4.5.5). This shows that ROI indeed deals better with heterogeneity, even if only slightly.

The EVR values for the skew model are 9 and 9.3 for the fixed regions and ROI respectively (Tables 4.4.6), which are much higher than the recommended limit of 0.20. Again GLS regression should be the preferred modeling choice over the OLS. Given that the skew model has a high sampling error component, an OLS model would give misleading results. The advantage of GLS is that it can distinguish between the variance due to model error and sampling error. Importantly, the Bayesian procedure adds another dimension to the analysis, by computing expectations over the entire posterior distribution. It has provided a more reasonable estimate of the model error variance where the method of moment's estimator would have been grossly underestimated the model error variance, as the sampling error has overwhelmed the analysis. As far as the ROI is concerned, there is little change in the EVR as compared to the fixed region, as the skew model tends to include more stations in the regional analysis.

For the fixed regions, the mean flood model has a model error variance of 17.8, which is much higher than 3.6 (for the standard deviation model) and 1.74 (for the skew model) (Tables 4.5.4, 4.5.5 and 4.5.6). For the ROI, the mean flood model also shows much higher model error variance than those of the standard deviation and skew models. These results indicate that the mean flood has the greater level of heterogeneity associated with it as compared to the standard deviation and skew models.

Pseudo ANOVA tables were also prepared for the flood quantile models. Tables 4.5.2 and 4.5.3 show the results for the Q_{20} and Q_{100} models, respectively. Here the ROI shows a higher EVR than the fixed region. This suggests that the BGLS regression should be used with ROI in developing the flood quantile models, especially as the ARI increases.

Table 4.5.2 Pseudo ANOVA table for Q_{20} model for Tasmania (QRT, fixed region and ROI)

Source	**Degrees of Freedom**		**Sum of Squares**		
	Fixed region	ROI	*Equations*	Fixed region	ROI
Model	*k=3*	*k=3*	$n(\sigma^2_{\delta 0} - \sigma^2_{\delta}) =$	34.3	37.5
Model error δ	*n-k-1=48*	*n-k-1=30*	$n(\sigma^2_{\delta}) =$	15.5	12.2
Sampling error η	*N = 52*	*N = 34*	$tr[\Sigma(\hat{y})]$ =	2.08	1.99
Total	*2n-1 = 103*	*2n-1 = 67*	**Sum of the above =**	**51.9**	**51.7**
			EVR	**0.13**	**0.16**

Table 4.5.3 Pseudo ANOVA table for Q_{100} model for Tasmania (QRT, fixed region and ROI)

Source	**Degrees of Freedom**		**Sum of Squares**		
	Fixed region	ROI		Fixed region	ROI
Model	*k=3*	*k=3*		30.7	34.1
Model error δ	*n-k-1=48*	*n-k-1=20*		19.0	15.7
Sampling error η	*N = 52*	*N = 52*		3.3	3.13
Total	*2n-1 = 103*	*2n-1 = 103*	**Sum of the above =**	**53.0**	**52.9**
			EVR	**0.17**	**0.2**

Table 4.5.4 Pseudo ANOVA table for the mean flood model for Tasmania (PRT, fixed region and ROI)

Source	**Degrees of Freedom**		**Sum of Squares**		
	Fixed region	ROI		Fixed region	ROI
Model	*k=3*	*k=3*	$n(\sigma^2_{\delta 0} - \sigma^2_{\delta}) =$	30.5	54.6
Model error δ	*n-k-1=48*	*n-k-1=24*	$n(\sigma^2_{\delta}) =$	17.8	7.1
Sampling error η	*N = 52*	*N = 28*	$tr[\Sigma(\hat{y})]$ =	1.13	1.02
Total	*2n-1 = 103*	*2n-1 = 55*	**Sum of the above =**	**49.4**	**63**
			EVR	**0.06**	**0.2**

Table 4.5.5 Pseudo ANOVA table for the standard deviation model for Tasmania (PRT, fixed region and ROI)

Source	**Degrees of Freedom**		**Sum of Squares**		
	Fixed region	ROI		Fixed region	ROI
Model	*k=2*	*k=2*		3.6	3.5
Model error δ	*n-k-1=49*	*n-k-1=33*		3.6	3.3
Sampling error η	*N = 52*	*N = 52*		1.9	2.2
Total	*2n-1 = 103*	*2n-1 = 103*	**Sum of the above =**	**9.1**	**9.0**
			EVR	**0.54**	**0.66**

Table 4.5.6 Pseudo ANOVA table for the skew model for Tasmania (PRT, fixed region and ROI)

Source	**Degrees of Freedom**		**Sum of Squares**		
	Fixed region	ROI		Fixed region	ROI
Model	*k=3*	*k=3*		0.62	1.80
Model error δ	*n-k-1=48*	*n-k-1=46*		1.74	1.54
Sampling error η	*N = 52*	*N = 50*		15.5	14.4
Total	*2n-1 = 103*	*2n-1 = 99*	**Sum of the above =**	**17.8**	**17.7**
			EVR	**9.0**	**9.3**

To assess the underlying model assumptions (i.e. the normality of residuals), the plots of the standardised residuals vs. predicted values were examined. The predicted values were obtained from one-at-a-time cross validation. Figures 4.5.3 to 4.5.5 show the plots for the Q_{20} and the mean flood models with the fixed region and ROI. The underlying model assumptions are satisfied to a large extent, as 95% of the standardised residuals values fall between the limits of ± 2. The ROI shows standardised residuals closer to the ± 2 limits. The results in Figures 4.5.3 to 4.5.5 reveal that the developed equations satisfy the normality of residual assumption quite satisfactorily. Also no specific pattern (heteroscedasicity) can be identified, with the standardised values being almost equally distributed below and above zero. Similar results were obtained for the skew, standard deviation and other flood quantile models.

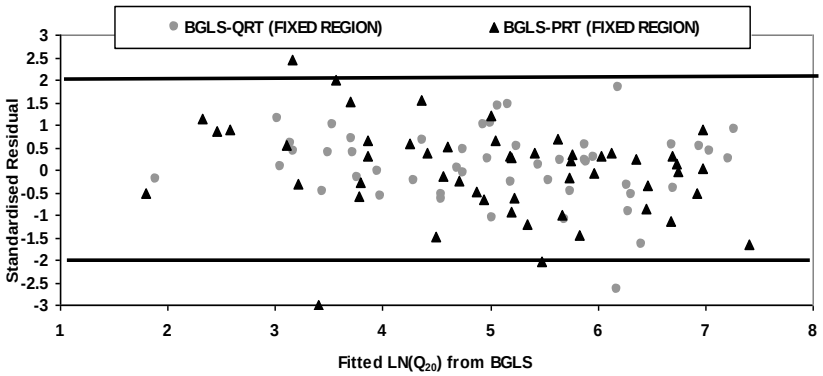


Figure 4.5.3 plots of standardised residuals vs. predicted values for ARI of 20 years (QRT and PRT, fixed region, Tasmania)

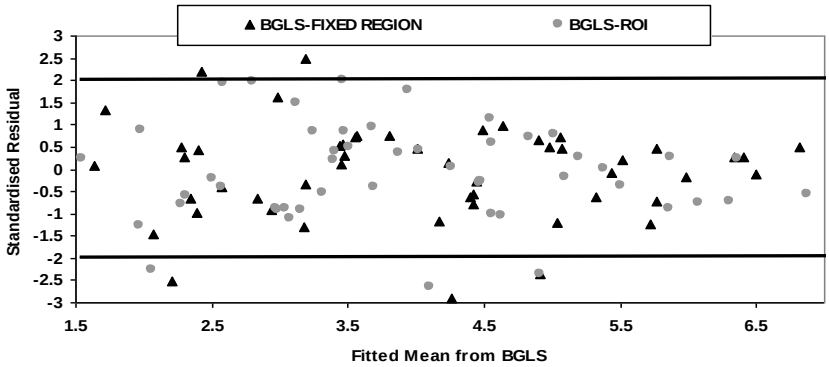


Figure 4.5.4 plot of standardised residuals vs. predicted values for the mean flood (PRT, fixed region, ROI, Tasmania)

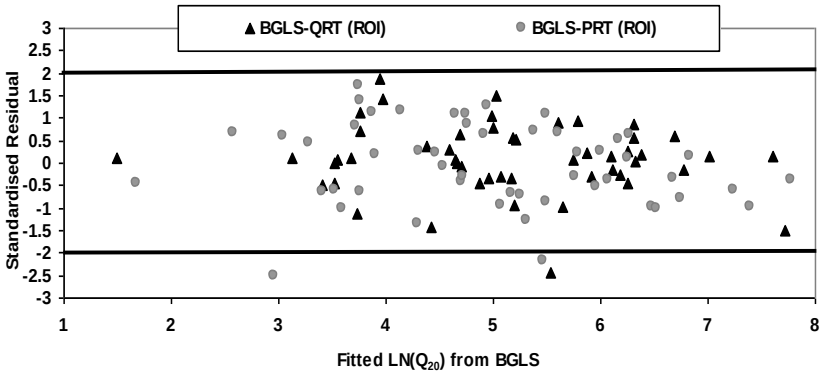


Figure 4.5.5 plots of standardised residuals vs. predicted values for ARI of 20 years (QRT and PRT, ROI, Tasmania)

The QQ-plots of the standardised residuals (Equation 4.2.13) vs. normal score (Equation 4.2.14) for the fixed region (based on one-at-a-time cross validation) and ROI were examined. Figures 4.5.6 to 4.5.8 present results for the Q_{20} and the skew models, which show that all the points closely follow a straight line. This indicates that the assumption of normality and the homogeneity of variance of the standardised residuals have largely been satisfied. The standardised residuals are indeed normally and independently distributed N(0,1) with mean 0 and variance 1 as the slope of the best fit line in the QQ-plot, which can be interpreted as the standard deviation of the normal score (Z score) of the quantile, should approach 1 and the intercept, which is the mean of the normal score of the quantile should approach 0 as the number of sites increases. It can be observed from Figures 4.5.6 to 4.5.8 that the fitted lines for the developed models pass through the origin (0, 0) and have a slope approximately equal to one. The ROI approach approximates the normality of the residuals slightly better (i.e. a better match with the fitted line) than the fixed region approach. Similar results were also found for the mean, standard deviation and other flood quantile models.

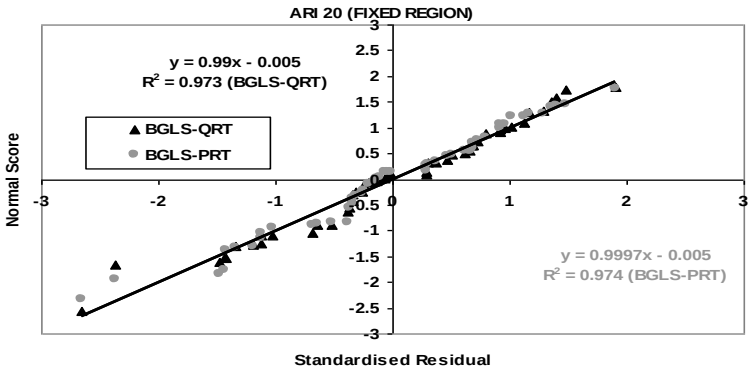


Figure 4.5.6 QQ-plot of the standardised residuals vs. Z score for ARI of 20 years (QRT and PRT, fixed region, Tasmania)

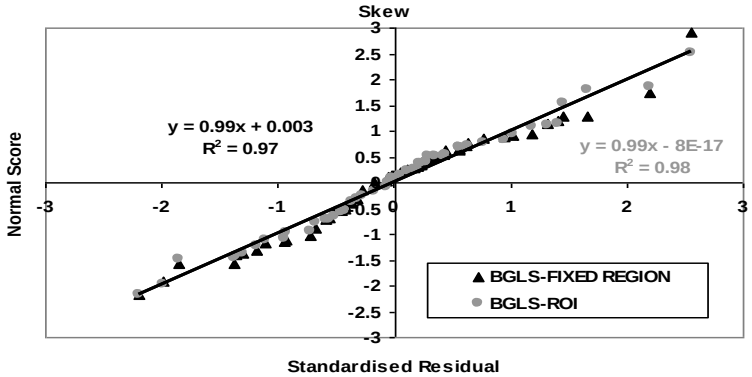


Figure 4.5.7 QQ-plot of the standardised residuals vs. Z score for the skew model (PRT, fixed region, ROI, Tasmania)

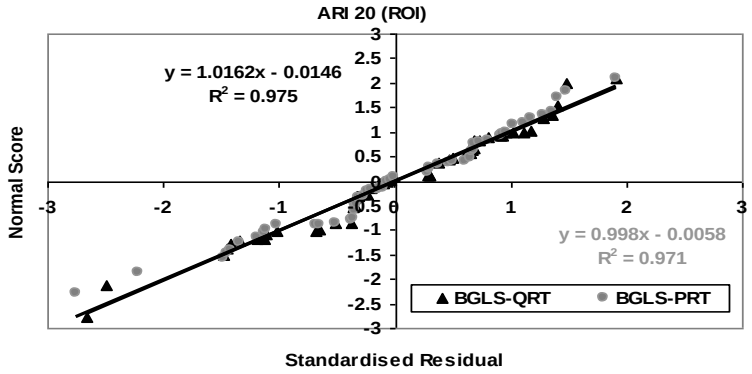


Figure 4.5.8 QQ-plot of the standardised residuals vs. Z score for ARI of 20 years (QRT and PRT, ROI, Tasmania)

The summary of various regression diagnostics (the relevant equations are described in Section 4.2.2) is provided in Table 4.5.7. This shows that for the mean flood model, the model error variance (MEV) and average standard error of prediction (SEP) are much higher than those of the standard deviation and skew models. This indicates that the mean flood model exhibits a higher degree of heterogeneity than the standard deviation and skew models, this result is also supported by the ANOVA analysis. Indeed the issue here is that sampling error becomes larger as the order of the moment increases, therefore, in case of the skew, the spatial variation is a second order effect that is not really detectable. For the mean flood model, the ROI shows a model error variance which is 11% smaller than for the fixed region. Also, the R^2_{GLS} value for the mean flood model with the ROI is 2% higher than for the fixed region. The reasonable reduction in MEV alone indicates that the ROI should be preferred over the fixed region analysis for developing the mean flood model. For the standard deviation model, ROI also shows 8% smaller SEP and 5% higher R^2_{GLS} values. This indicates that the ROI is preferable to the fixed region for the standard deviation model. What is also noteworthy (as seen from Table 4.5.7) is that the SEP% for the skew model is slightly larger for the ROI than the fixed region analysis. This may be due to the fact that, if the number of sites are reduced (smaller ROI), the predictive variance may be slightly inflated in the skew region. The R^2_{GLS} values for the skew models are similar for the fixed region and ROI, with the latter providing only a 2% increase.

One can see from Table 4.5.7 that the SEP values for all the flood quanitle models are 2% to 11% smaller for the ROI cases than the fixed region; the best result is obtained for ARI = 2

years. Also, the R^2_{GLS} values for ROI cases are 3% to 6% higher than the fixed region. These results show that the ROI generally outperforms the fixed region approach.

Table 4.5.7 Regression diagnostics for fixed region and ROI for Tasmania

Model	Fixed region				ROI			
	MEV	AVP	SEP (%)	R^2_{GLS} (%)	MEV	AVP	SEP (%)	R^2_{GLS} (%)
Mean	0.35	0.37	67	86	0.24	0.27	56	88
Stdev	0.071	0.076	28	51	0.042	0.046	20	56
Skew	0.034	0.050	22	52	0.031	0.050	23	54
Q_2	0.55	0.59	83	76	0.38	0.419	72	79
Q_5	0.33	0.36	61	82	0.25	0.28	57	86
Q_{10}	0.30	0.32	58	84	0.23	0.26	54	87
Q_{20}	0.30	0.33	58	83	0.23	0.26	55	87
Q_{50}	0.34	0.37	62	82	0.27	0.30	60	86
Q_{100}	0.37	0.40	66	79	0.30	0.34	64	85

Table 4.5.8 shows number of sites and associated model error variances for the ROI and fixed region models. This shows that the ROI mean flood model has fewer sites on average (28 out of 52 i.e. 54%) than the standard deviation and skew models. The ROI skew model has the highest number of sites which includes nearly all the sites in Tasmania (50 out of 52 i.e. 96%). The model error variances for all the ROI models (except the skew model) are smaller than the fixed region models. This shows that the fixed region models experience a greater heterogeneity than the ROI. If the fixed regions are made too large, the model error will be inflated by heterogeneity that will go unaccounted for by the catchment characteristics. Figure 4.5.9 shows the resulting sub-regions in Tasmania (with minimum model error variances) for the ROI mean flood and skew models. For the mean flood and skew models, there are two distinct sub-regions. The regions can be classified as east and west Tasmania for which there are two distinct types of rainfall regimes and districts.

Table 4.5.8 Model error variances associated with fixed region and ROI for Tasmania (n = number of sites in the region)

Parameter/ Quantiles	Mean	Stdev	Skew	Q_2	Q_5	Q_{10}	Q_{20}	Q_{50}	Q_{100}
ROI (n)	28	36	50	30	35	35	34	33	33
$\hat{\sigma}^2_\delta$	0.24	0.042	0.031	0.38	0.25	0.23	0.23	0.27	0.30
Fixed region (n)	52	52	52	52	52	52	52	52	52
$\hat{\sigma}^2_\delta$	0.35	0.067	0.034	0.55	0.33	0.30	0.30	0.34	0.37

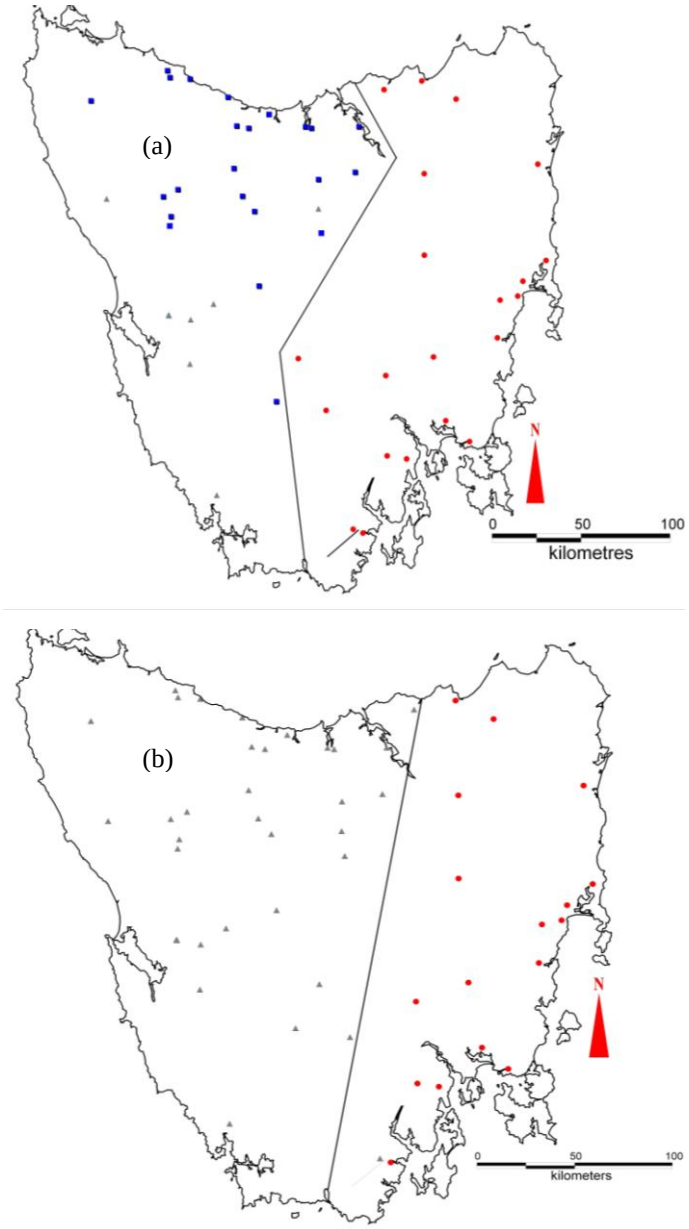


Figure 4.5.9 Binned minimum model error variance for Tasmania (a) mean flood model and (b) skew model

Table 4.5.9 presents the root mean square error (RMSE) (Equation 4.2.16) and relative error (RE) (Equation 4.2.15) values for the PRT and QRT models with both the fixed region and ROI. In terms of RMSE, ROI gives smaller values than the fixed regions for all the ARIs. The PRT-ROI shows smaller RMSE values than the ROI for all the ARIs, however for ARIs of 5, 10 and 20 years the increase is negligible (i.e. 1 to 2 %). In terms of RE, ROI gives 0 to 9% smaller values than the fixed regions. The PRT-ROI gives larger values of RE (by 13%) for

both the 50 and 100 years ARIs. For ARIs of 2 to 20 years, the QRT-ROI gives smaller RE values (by 1% to 13%) than the PRT-ROI.

Finally the results of counting the Q_{pred}/Q_{obs} ratios for the QRT and PRT for the ROI and fixed regions are provided in Tables 4.5.10 and 4.5.11. The QRT-ROI has 85% ratio values in the desirable range, compared to 81% for the QRT-fixed region. The PRT-ROI has 78% ratio values in the desirable range, compared to 74% for the PRT-fixed region. These results show that ROI performs better than the fixed regions with both the QRT and PRT. The PRT-ROI shows 16% underestimation as compared to 8% for the QRT-ROI. The cases for overestimation were very similar for both the methods.

Table 4.5.9 Evaluation statistics (RMSE and RE) from one-at-a-time cross validation for Tasmania

Model	RMSE (%)				RE (%)			
	PRT		QRT		PRT		QRT	
	Fixed region	ROI	Fixed region	ROI	Fixed region	ROI	Fixed region	ROI
Q_2	11	10	16	12	33	31	38	30
Q_5	9	7	11	8	35	30	34	25
Q_{10}	10	7	11	8	34	37	30	24
Q_{20}	10	7	13	9	36	37	27	27
Q_{50}	11	7	13	10	39	41	29	28
Q_{100}	12	7	13	10	49	42	33	29

Table 4.5.10 Summary of counts/percentages based on Q_{pred}/Q_{obs} ratio values for QRT and PRT for Tasmania (fixed region). “U” = gross underestimation, “D” = desirable range and “O” = gross overestimation

Model	Count (QRT)			Percent (QRT)			Count (PRT)			Percent (PRT)		
	U	D	O	U	D	O	U	D	O	U	D	O
Q_2	2	41	9	4	79	17	5	41	6	10	79	12
Q_5	2	44	6	4	85	12	6	41	5	12	79	10
Q_{10}	3	46	3	6	88	6	6	41	5	12	79	10
Q_{20}	4	45	3	8	87	6	9	37	6	17	71	12
Q_{50}	6	40	6	12	77	12	10	36	6	19	69	12
Q_{100}	9	38	5	17	73	10	10	36	6	19	69	12
Sum / average	**26**	**254**	**32**	**8**	**81**	**10**	**46**	**232**	**34**	**15**	**74**	**11**

Table 4.5.11 Summary of counts/percentages based on Q_{pred}/Q_{obs} ratio values for QRT and PRT for Tasmania (ROI). "U" = gross underestimation, "D" = desirable range and "O" = gross overestimation

ARI (years)	Count (QRT)			Percent (QRT)			Count (PRT)			Percent (PRT)		
	U	D	O	U	D	O	U	D	O	U	D	O
2	3	45	4	6	87	8	6	43	3	12	83	6
5	2	45	5	4	87	10	7	42	3	13	81	6
10	3	45	4	6	87	8	9	41	2	17	79	4
20	4	45	3	8	87	6	9	40	3	17	77	6
50	6	42	4	12	81	8	9	39	4	17	75	8
100	6	42	4	12	81	8	9	39	4	17	75	8
Sum / average	**24**	**264**	**24**	**8**	**85**	**8**	**49**	**244**	**19**	**16**	**78**	**6**

4.5.2 PRT – fixed region (Tasmania considered having two regions)

Based on the region-of-influence (ROI) approach, prominent spatial variations were found in the model error variance for the mean flood model. This can be seen in Figure 4.5.9a. The much greater spatial variability of the mean is dominated by local factors (as compared to the higher moments).

The results of this analysis concur with previous studies (McConachy et. al., 2003, Gamble et. al., 1998, Xuereb et. al, 2001) which showed that large rainfalls over Tasmania are not meteorologically homogeneous. In the east of the state, the largest rainfall events occur in the warmer spring and summer months when low pressure systems in the Tasman Sea can direct an easterly onshore flow over Tasmania. The heaviest rainfalls in the west of the state are due to the passage of fronts, sometimes associated with an intense extratropical cyclone with a westerly or southwesterly airstream (Xuereb et.al., 2001).

Based on this finding and on past results, as described above, it was decided to treat Tasmania as two different regions (i.e. east and west) and apply the Parameter Regression Technique (PRT) to both these regions. This analysis is discussed below.

A total of 32 catchments were located in west Tasmania and 21 stations in east Tasmania. The locations of these catchments are shown in Figure 4.5.9a. The annual maximum flood series record lengths for west Tasmania range from 20 to 58 years (mean 31 years, median 28 years and standard deviation 9.75 years). The catchment areas of these 32 stations

range from 4.5 to 1600 km^2 (mean 353 km^2, median 184 km^2 and standard deviation of 418 km^2). For east Tasmania the record lengths range from 19 to 74 years (mean 28 years, median 27 years and a standard deviation 11 years). The catchment areas of these 21 stations range from 1.3 to 1900 km^2 (mean 276 km^2, median 136 km^2 and standard deviation of 420 km^2)

In the fixed region approach, all the 32 catchments in west Tasmania were considered to have formed one region; however, one catchment was left out for cross-validation and the procedure was repeated 32 times to implement one-at-a-time cross validation. A similar approach was adopted for east Tasmania. For this analysis no predictor variable was used for the standard deviation and skew models (e.g. a regional average value was used, which is the skew values weighted by the error covariance matrix). For the mean flood model, predictor variables were selected as explained below.

Table 4.5.1 shows different combinations of predictor variables used in the mean flood model. Figures 4.5.10 and 4.5.11 show example plots of the statistics used in selecting the best set of predictor variables for the mean flood models of west Tasmania and east Tasmania. According to the model error variance, combinations 5 and 6 were potentially the best sets of predictor variables for mean flood models for both west Tasmania and east Tasmania. Indeed combination 5 with two predictor variables (area and design rainfall intensity ${}^2I_{12}$) showed the lowest model error variance and the highest ${R^2}_{GLS}$. The AVPO, AVPN, AIC and BIC values favour combination 5 as well. We also compared combination 5 to combination 6 (also two predictor variables, area and design rainfall intensity ${}^{50}I_{12}$). The regression coefficient for the variable ${}^{50}I_{12}$ was 6 times the posterior standard deviation away from zero as compared to 7.2 times for ${}^2I_{12}$. Hence, combination 5 was finally selected as the best set of predictor variables for the mean flood model. The BPVs for the predictor variables for the mean model were smaller than 0.5% for both the predictor variables (area and ${}^2I_{12}$) for both west Tasmania and east Tasmania.

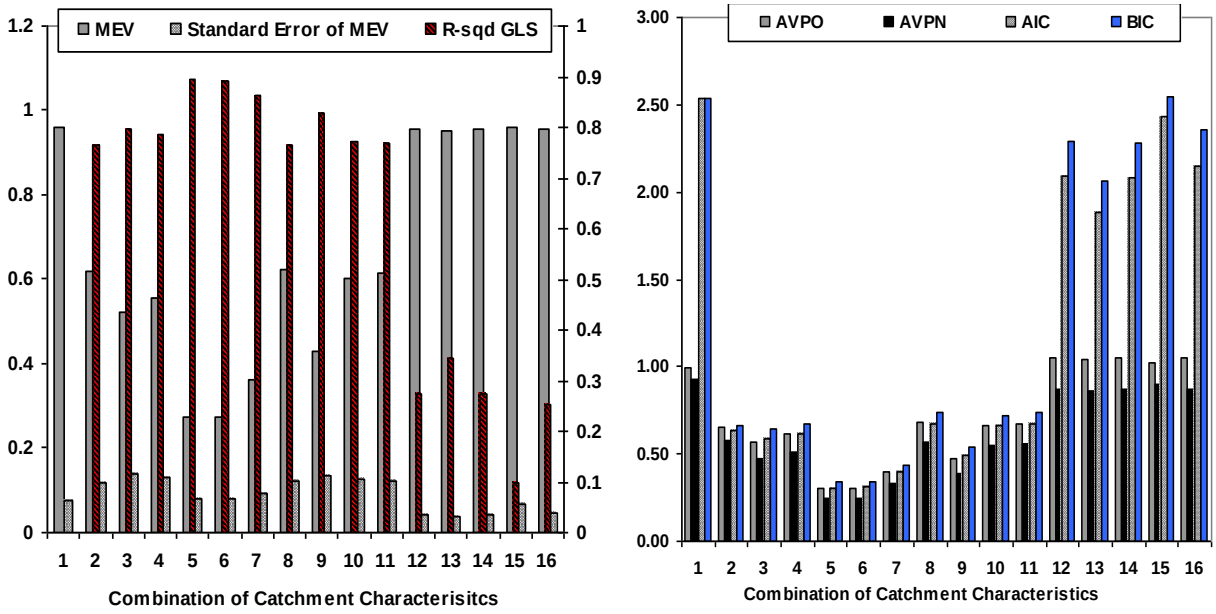


Figure 4.5.10 Selection of predictor variables for the BGLS regression model for the mean flood in west Tasmania – (Note R^2 GLS uses the right hand axis)

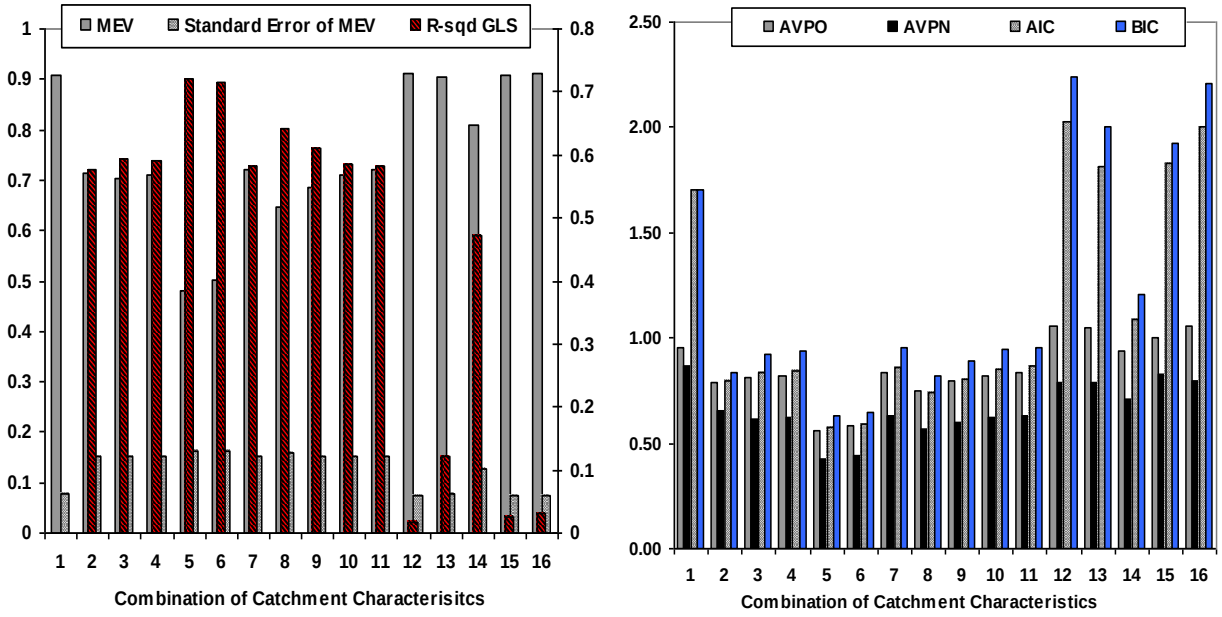


Figure 4.5.11 Selection of predictor variables for the BGLS regression model for the mean flood in east Tasmania – (Note R^2 GLS uses the right hand axis)

Regression equations developed for the mean flood for the two fixed regions in Tasmania are given by Equations 4.5.10 to 4.5.12 for west Tasmania and 4.5.13 to 4.5.15 for east Tasmania

west Tasmania:

$M = 4.32 + 0.95[\ln(area) - 5.00] + 3.47[\ln(I_{12,2}) - 1.55]$ (4.5.10)

stdev = 0.42 (4.5.11)

skew = 0.022 (4.5.12)

east Tasmania:

$M = 3.44 + 0.77[\ln(area) - 4.74] + 4.03[\ln(I_{12,2}) - 1.40]$ (4.5.13)

stdev = 0.98 (4.5.14)

skew = - 0.25 (4.5.15)

The Pseudo Analysis of Variance (ANOVA) tables for west Tasmania and east Tasmania for the parameters of the LP3 distribution are presented in Tables 4.5.12 – 4.5.14.

For the LP3 parameters, the sampling error increases as the order of moment increases i.e. the EVR increases with the order of moments. The model error dominates the regional analysis for the mean flood and the standard deviation models for both west Tasmania and east Tasmania. For the mean flood model the EVR is 0.03 for west Tasmania and 0.09 for east Tasmania (Table 4.5.12). For the standard deviation model the EVR is 0.16 for west Tasmania and 0.11 for east Tasmania.

The EVR values for the skew models are 3.8 and 4.7 for west Tasmania and east Tasmania respectively (Table 4.5.14), which are much higher than the recommended limit of 0.20. Given that the skew has a high sampling error component, an OLS model would give misleading results. These results are consistent with the other states in that the mean flood model always shows a much higher model error variance than those of the standard deviation and skew models. These results indicate that the mean flood has the greater level of heterogeneity associated with it compared to the standard deviation and skew.

Table 4.5.12 Pseudo ANOVA table for the mean flood model (PRT, fixed region for west Tasmania and east Tasmania)

Source	**Degrees of Freedom**		**Sum of Squares**		
	West Tasmania	East Tasmania		West Tasmania	East Tasmania
Model	*k=3*	*k=3*	$n(\sigma^2_{\delta 0} - \sigma^2_{\delta}) =$	21	8.5
Model error δ	*n-k-1=28*	*n-k-1=17*	$n(\sigma^2_{\delta}) =$	8.5	10
Sampling error η	*N = 32*	*N = 21*	$tr[\Sigma(\hat{y})] =$	0.27	0.91
Total	*2n-1 = 63*	*2n-1 = 41*	**Sum of the above =**	**30**	**19.45**
			EVR	**0.03**	**0.09**

Table 4.5.13 Pseudo ANOVA table for the standard deviation model (PRT, fixed region for west Tasmania and east Tasmania)

Source	**Degrees of Freedom**		**Sum of Squares**		
	West Tasmania	East Tasmania		West Tasmania	East Tasmania
Model	*k=1*	*k=1*		0.0	0.0
Model error δ	*n-k-1=30*	*n-k-1=19*		4.8	3.9
Sampling error η	*N = 32*	*N = 21*		0.75	0.42
Total	*2n-1 = 63*	*2n-1 = 41*	**Sum of the above =**	**5.54**	**4.35**
			EVR	**0.16**	**0.11**

Table 4.5.14 Pseudo ANOVA table for the skew model (PRT, fixed region for west Tasmania and east Tasmania)

Source	**Degrees of Freedom**		**Sum of Squares**		
	West Tasmania	East Tasmania		West Tasmania	East Tasmania
Model	*k=1*	*k=1*		0.0	0.0
Model error δ	*n-k-1=30*	*n-k-1=19*		2.37	1.45
Sampling error η	*N = 32*	*N = 21*		9.0	6.9
Total	*2n-1 = 63*	*2n-1 = 41*	**Sum of the above =**	**11.3**	**8.3**
			EVR	**3.8**	**4.7**

To assess the underlying model assumptions (i.e. the normality of residuals), the plots of standardised residuals vs. predicted values were examined for west Tasmania and east Tasmania. The predicted values were obtained from one-at-a-time cross validation. Figures

4.5.12 and 4.5.13 show the plots for Q_{20} predicted from PRT and the mean flood models for west Tasmania and east Tasmania. The underlying model assumptions are satisfied to a large extent as 95% of the standardised residual values fall between the magnitudes of ± 2. The results in Figures 4.5.12 to 4.5.13 reveal that the developed equations satisfy the normality of residual assumption quite satisfactorily. Also no specific pattern (heteroscedasicity) can be identified with the standardised values being almost equally distributed below and above zero. Similar results were obtained for the skew, standard deviation and other flood quantiles estimated from PRT. This result indicates that there is no major issue with treating Tasmania as two different regions and that no true outlier sites occur with this subdivision.

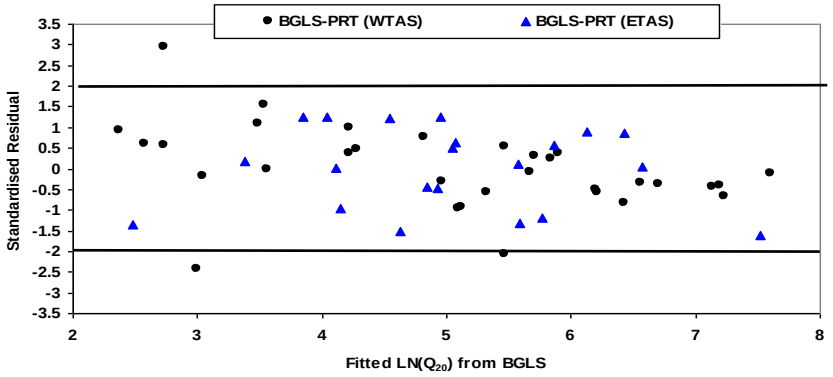


Figure 4.5.12 Plot of standardised residuals vs. predicted values for ARI of 20 years (PRT, west Tasmania and east Tasmania)

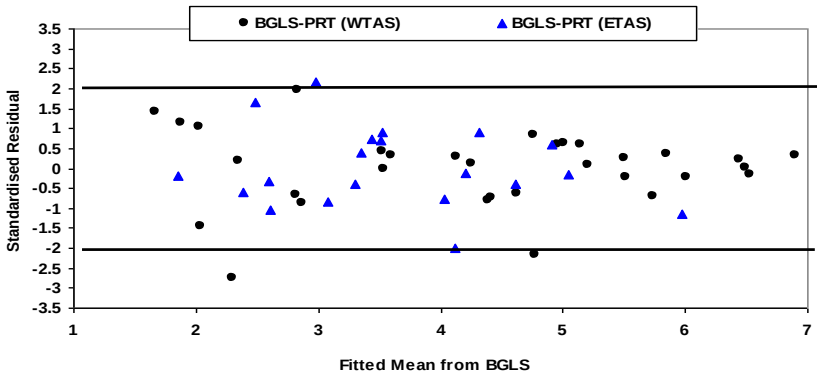


Figure 4.5.13 Plot of standardised residuals vs. predicted values for the mean flood (PRT, west Tasmania and east Tasmania)

The QQ-plots of the standardised residuals (Equation 4.2.13) vs. normal score (Equation 4.2.14) for west Tasmania and east Tasmania (based on one-at-a-time cross validation)

were examined. Figure 4.5.14 presents the result for the Q_{20} model, which shows that all the points closely follow a straight line. This indicates that the assumption of normality and the homogeneity of variance of the standardised residuals have largely been satisfied. The standardised residuals are indeed normally and independently distributed N(0,1) with mean 0 and variance 1, as the slope of the best fit line in the QQ-plot, which can be interpreted as the standard deviation of the normal score (Z score) of the quantile, should approach 1 and the intercept, which is the mean of the normal score of the quantile should approach 0 as the number of sites increases. It can be observed from Figure 4.5.14 that the fitted lines for the developed models pass through the origin (0, 0) and it has a slope approximately equal to one. Similar results were also found for the mean, standard deviation and other flood quantile models. This result indicates that there is no major issue with treating Tasmania as two different regions and that no true outlier sites occur with this subdivision, as found with the standardised residuals vs. predicted values plot.

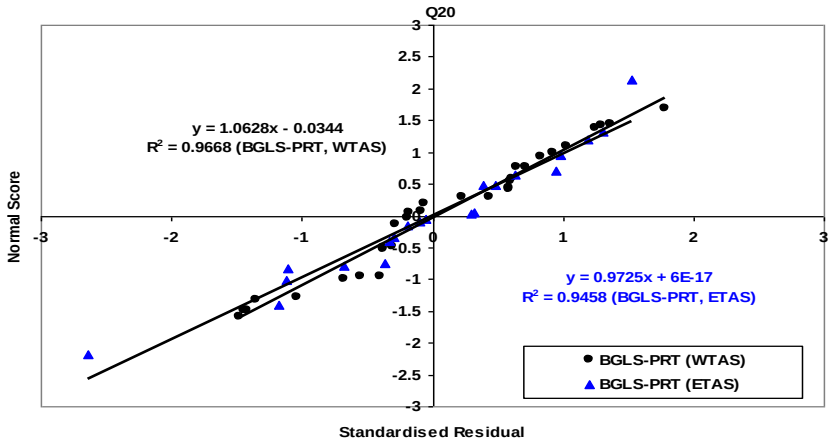


Figure 4.5.14 QQ-plot of the standardised residuals vs. Z score for ARI of 20 years (PRT, west Tasmania and east Tasmania)

The summary of various regression diagnostics for west Tasmania and east Tasmania (the relevant equations are described in Section 4.2.2) is provided in Table 4.5.15. This shows that for the mean flood model (for both west Tasmania and east Tasmania) the model error variance (MEV) and average standard error of prediction (SEP) are much higher than those of the standard deviation and skew models. This indicates that the mean flood model exhibits a higher degree of heterogeneity than the standard deviation and skew models, this result also supports the ANOVA analysis. Indeed the issue here is that sampling error becomes larger as the order of the moment increases, therefore, in case of the skew, the spatial variation is a second order effect that is not really detectable. For the mean flood model, west Tasmania shows a model error variance which is 29% smaller than east Tasmania. Also, the

R^2_{GLS} value for west Tasmania mean flood model is 18% higher than east Tasmania. The notable difference in MEV for the mean flood model indicates that it is justified to treat west Tasmania and east Tasmania as two separate regions. For the standard deviation and skew models, west Tasmania has very similar SEP values to east Tasmania, i.e. 41% and 48% respectively for the standard deviation model, and 31% and 30% respectively for the skew model. The R^2_{GLS} values for the standard deviation and skew models for west Tasmania and east Tasmania are zero as no explanatory variables are used.

Table 4.5.15 Regression diagnostics for west Tasmania and east Tasmania

Model	West Tasmania				East Tasmania			
	MEV	AVP	SEP (%)	R^2_{GLS} (%)	MEV	AVP	SEP (%)	R^2_{GLS} (%)
Mean	0.28	0.30	60	89	0.50	0.59	89	71
Stdev	0.14	0.15	41	0	0.20	0.22	48	0
Skew	0.076	0.091	31	0	0.073	0.090	30	0

Table 4.5.16 presents the root mean square error (RMSE) (Equation 4.2.16) and relative error (RE) (Equation 4.2.15) values for the flood quantiles estimated by PRT with both west Tasmania and east Tasmania. In terms of RMSE, west Tasmania gives smaller values than east Tasmania for most of the ARIs. In terms of RE, west Tasmania provides smaller values than east Tasmania over all the ARIs by 17 to 27%.

Finally the results of counting the Q_{pred}/Q_{obs} ratios for the PRT (for west Tasmania and east Tasmania) are provided in Table 4.5.17. It was found that west Tasmania provided relatively better results with 81% of the ratio values in the desirable range, which is 62% for east Tasmania.

Table 4.5.16 Evaluation statistics (RMSE and RE) from one-at-a-time cross validation for west Tasmania and east Tasmania

Model	RMSE (%) PRT		RE (%) PRT	
	West Tasmania	East Tasmania	West Tasmania	East Tasmania
Q_2	8.8	12.3	29	49
Q_5	8.6	8.2	28	55
Q_{10}	8.5	8.4	31	54
Q_{20}	8.4	9.0	35	52
Q_{50}	8.3	9.7	40	59
Q_{100}	8.2	10.2	40	63

Table 4.5.17 Summary of counts based on Q_{pred}/Q_{obs} ratio values for PRT for west Tasmania and east Tasmania. "U" = gross underestimation, "D" = desirable and "O" = gross overestimation

Model	Count (West Tasmania)			Percent (West Tasmania)			Count (East Tasmania)			Percent (East Tasmania)		
	U	D	O	U	D	O	U	D	O	U	D	O
Q_2	3	27	2	9	84	6	3	15	3	14	71	14
Q_5	3	27	2	9	84	6	3	15	3	14	71	14
Q_{10}	3	27	2	9	84	6	4	13	4	19	62	19
Q_{20}	4	25	3	13	78	9	4	12	5	19	57	24
Q_{50}	4	25	3	13	78	9	4	12	5	19	57	24
Q_{100}	5	24	3	16	75	9	5	11	5	24	52	24
Sum / average	**22**	**155**	**15**	**11**	**81**	**8**	**23**	**78**	**25**	**18**	**62**	**20**

4.6 Results for Queensland

A total of 172 catchments were used from Queensland for the analyses presented here. These catchments are listed in the appendix. The locations of these catchments are shown in Figure 2.14. The annual maximum flood series record lengths of these 172 stations range from 25 to 94 years (mean 40 years, median 36 years and a standard deviation 14.5 years). The catchment areas of these 172 stations range from 7 to 963 km^2 (mean 325 km^2, median 254 km^2 and standard deviation of 252 km^2).

In the fixed region approach, all the 172 catchments were considered to have formed one region, however, one catchment was left out for cross-validation and the procedure was repeated 172 times to implement one-at-a-time cross validation. In the region-of-influence (ROI) approach, an optimum region was formed for each of the 172 catchments by starting with 15 stations and then consecutively adding 5 stations at each iteration.

Table 4.6.1 shows different combinations of predictor variables for the Q_{10} QRT model and the models for the first three moments of the LP3 distribution. Figure 4.6.1 and 4.6.2 show example plots of the statistics used in selecting the best set of predictor variables for Q_{10} and the skew models. According to the model error variance, combinations 8, 9 and 18 were the top 3 potential sets of predictor variables for the Q_{10} model. Combination 18 contained 5 predictor variables while combinations 8 and 9 each contained 2 predictor variables. Indeed combination 8 with the 2 predictor variables (area and design rainfall intensity ${}^{50}I_{72}$) showed the lowest model error variance (and the lowest standard error of MEV) and the highest R^2_{GLS}. The AVPO, AVPN, AIC and BIC values favour combination 8 as well. We also compared combination 8 to combination 10 (the latter contained 2 predictor variables, area and design rainfall intensity $I_{tc,10}$). Combination 8 had a smaller model error variance, while both the variables (${}^{50}I_{72}$ and $I_{tc,10}$) were at least 10 times the posterior standard deviation away from zero. Overall, given that there was not an overly large difference in R^2_{GLS} and model error variance and for consistency with the other eastern states (NSW and Victoria), combination 10 was finally selected as the best set of predictor variables for the Q_{10} model.

For the skew model, combination 13 showed the lowest model error variance (0.0152) and the highest R^2_{GLS} (46%) (see Figure 4.6.2) as well as the lowest AIC and BIC. Combination 1 without any explanatory variables ranked 4 out of the 16 possible combinations (model error variance of 0.0159); the AVPO and AVPN as compared to combination 13 were quite

comparable. In this case, given the relatively small difference in model error variance, it may be argued that a regional average skew may be applicable for Queensland.
A similar procedure was adopted in selecting the best set of predictor values for other models with the QRT and PRT. The sets of predictor variables selected as above were used in the one-at-a-time cross validation with the fixed regions and region-of-influence (ROI) approaches.

To see how statistically significant the variables were for the regression equations, the BPV values for the regression coefficients were calculated. For the QRT over all the ARIs, the BPV values were 0% for both the variables area and design rainfall intensity $I_{tc,ARI}$. This justifies the inclusion of predictor variables area and $I_{tc,ARI}$ in the prediction equations for QRT. The BPVs for the skew model were 8% and 4% for $^{50}I_{72}$ and rain, respectively indicating that rain is a good predictor (5% level of significance) for skew in Queensland. The BPVs for the mean flood model were 0% for both the predictor variables $^{50}I_{72}$ and rain. For the standard deviation model, the BPVs for the predictor variables area and $^{2}I_{1}$ were 42% and 2% respectively which indicates that area is not a good predictor for the standard deviation model in Queensland.

Regression equations developed for the QRT and PRT for the fixed region are given by Equations 4.6.1 to 4.6.9.

Table 4.6.1 Different combinations of predictor variables considered for the QRT models and the parameters of the LP3 distribution for Queensland (QRT and PRT fixed region)

Combination	Combinations for mean, standard deviation & skew models	Combinations for flood quantile model
1	Const	Const
2	Const, area	Const, area
3	Const, area, $^{2}I_{1}$	Const, area, $^{2}I_{1}$
4	Const, area, $^{50}I_{1}$	Const, area, $^{2}I_{12}$
5	Const, area, $^{50}I_{12}$	Const, area, $^{50}I_{1}$
6	Const, area, $^{2}I_{12}$	Const, area, $^{50}I_{12}$
7	Const, area, rain	Const, area, rain
8	Const, area, $^{50}I_{72}$	Const, area, $^{50}I_{72}$
9	Const, area, evap	Const, area, $^{50}I_{72}$, evap
10	Const, area, S1085	Const, area, $I_{tc,ARI}$
11	Const, area, SL*	Const, area, evap
12	Const, SL, rain	Const, area, S1085
13	Const, $^{50}I_{72}$, rain	Const, area, SL
14	Const, S1085, $^{50}I_{72}$	Const, SL, rain
15	Const, evap	Const, $^{50}I_{72}$, rain
16	Const, rain, evap	Const, area, $^{50}I_{12}$, rain
17	-	Const, area, $^{50}I_{12}$, SL
18	-	Const, area, $^{50}I_{12}$, rain, evap
19	-	Const, area, $^{50}I_{12}$, $I_{tc,ARI}$, evap
20	-	Const, area, $^{50}I_{12}$, $I_{tc,ARI}$, rain, evap
21	-	Const, area, $^{50}I_{12}$, $I_{tc,ARI}$, SL
22	-	Const, area, $I_{tc,ARI}$, S1085
23	-	Const, area, $I_{tc,ARI}$, evap
24	-	Const, area, $I_{tc,ARI}$, rain
25	-	Const, area, $^{2}I_{1}$, $I_{tc,ARI}$

* SL = Stream Length

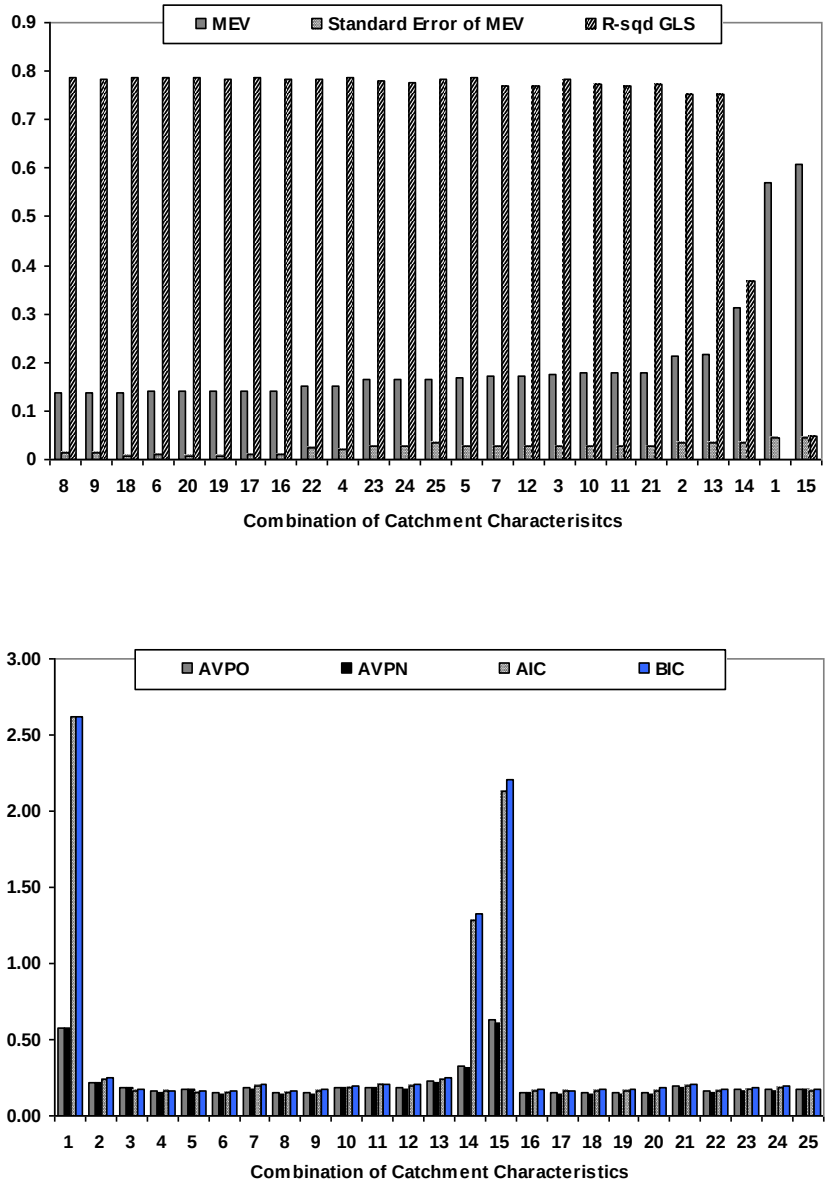


Figure 4.6.1 Selection of predictor variables for the BGLS regression model for Q_{10} (QRT, fixed region Queensland), MEV = model error variance, AVPO = average variance of prediction (old), AVPN = average variance of prediction (new) AIC = Akaike information criterion, BIC = Bayesian information criterion

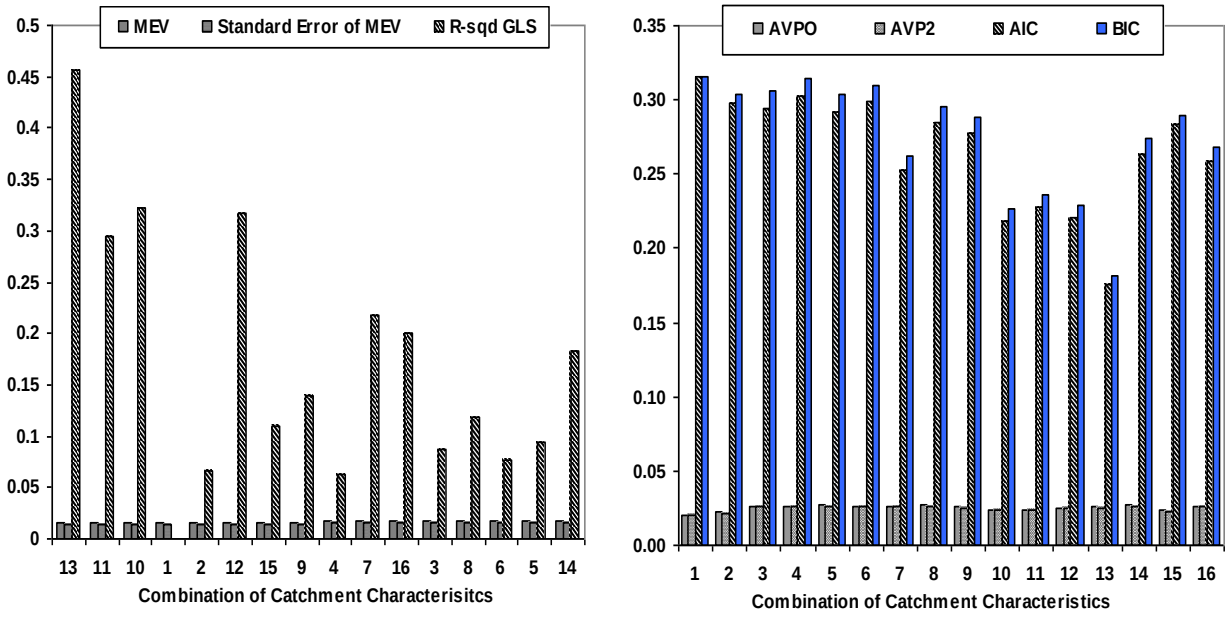


Figure 4.6.2 Selection of predictor variables for the BGLS regression model for skew

$$\ln(Q_2) = 4.80 + 1.35(\boldsymbol{z}area) + 2.57(\boldsymbol{zI}_{tc,ARI}) \quad (4.6.1)$$

$$\ln(Q_5) = 5.77 + 1.16(\boldsymbol{z}area) + 1.95(\boldsymbol{zI}_{tc,ARI}) \quad (4.6.2)$$

$$\ln(Q_{10}) = 6.25 + 1.00(\boldsymbol{z}area) + 1.67(\boldsymbol{zI}_{tc,ARI}) \quad (4.6.3)$$

$$\ln(Q_{20}) = 6.59 + 0.98(\boldsymbol{z}area) + 1.42(\boldsymbol{zI}_{tc,ARI}) \quad (4.6.4)$$

$$\ln(Q_{50}) = 6.97 + 0.91(\boldsymbol{z}area) + 1.19(\boldsymbol{zI}_{tc,ARI}) \quad (4.6.5)$$

$$\ln(Q_{100}) = 7.23 + 0.86(\boldsymbol{z}area) + 1.01(\boldsymbol{zI}_{tc,ARI}) \quad (4.6.6)$$

$$M = 4.71 + 0.74(\boldsymbol{z}area) + 1.97(\boldsymbol{zI}_{12,50}) \quad (4.6.7)$$

$$stdev = 1.37 - 0.03(\boldsymbol{z}area) - 1.41(\boldsymbol{zI}_{1,2}) \quad (4.6.8)$$

$$skew = -0.63 - 0.32(\boldsymbol{zI}_{72,50}) + 0.36(\boldsymbol{z}rain) \quad (4.6.9)$$

where *z()* is explained by Equation 4.3.10.

It is reassuring to observe that the regression coefficients in the QRT set of equations vary in a regular fashion with increasing ARI.

The Pseudo Analysis of Variance (ANOVA) tables for Q_{20} and Q_{100} models and the parameters of the LP3 distribution are presented in Tables 4.6.2 – 4.6.6 for the fixed regions and ROI.

For the LP3 parameters, the sampling error increases as the order of moment increases i.e. the EVR increases with the order of moments. The ROI in each case of the LP3 parameters shows a reduced model error variance (i.e. a reduced heterogeneity) as compared to the fixed regions as fewer sites have been used. The model error dominates the regional analysis for the mean flood and the standard deviation models for both the fixed regions and ROI. The ROI for the mean flood shows a higher EVR than the fixed regions e.g. the EVR is 0.40 for the ROI and 0.26 for the fixed region (Table 4.6.4) which is a 14% increase. For the standard deviation model, the EVR values are 0.36 and 0.73 for the ROI and fixed region (Table 4.6.5). This indicates that EVR is almost double for the ROI as compared to the fixed region, which means that the ROI procedure deals better with the heterogeneity than the fixed region.

The EVR values for the skew model are 17 and 19 for the fixed regions and ROI respectively (Tables 4.6.6), which are much higher than those of the mean flood and standard deviation models. Hence, GLS regression should be the preferred modeling choice for all the three parameters of the LP3 distribution, especially for the skew where the EVR is very high. The ROI combined with GLS is also advantageous as there is certainly a reduction in the model error variance. The EVR for ROI as compared to the fixed region for the skew model is slightly higher, as the reduction in the number sites (172 to 150) has slightly decreased the overall heterogeneity in the model. In any case the skew model tends to include more stations in the regional analysis, due to the fact that most of the error in the regional model is sampling error. These results indicate, as found for NSW, Victoria and Tasmania, that the mean flood shows greater levels of heterogeneity when compared to the standard deviation and skew.

Pseudo ANOVA tables were also prepared for the flood quantile models. Tables 4.6.2 and 4.6.3 show the results for Q_{20} and Q_{100} models, respectively. Here the ROI shows a higher EVR (nearly double) than the fixed region. This suggests that the BGLS regression and ROI should be the preferred modeling approach in developing the flood quantile models, especially as the ARI increases.

Table 4.6.2 Pseudo ANOVA table for Q_{20} model of Queensland (QRT, fixed region and ROI)

Source	**Degrees of Freedom**		**Sum of Squares**		
	Fixed region	ROI	*Equations*	Fixed region	ROI
Model	*k=3*	*k=3*	$n(\sigma^2_{\delta 0} - \sigma^2_{\delta}) =$	46	59
Model error δ	*n-k-1=168*	*n-k-1=77*	$n(\sigma^2_{\delta}) =$	25	12
Sampling error η	*N = 172*	*N = 81*	$tr[\Sigma(\hat{y})] =$	13	12
Total	*2n-1 = 343*	*2n-1 = 161*	**Sum of the above =**	**84**	**83**
			EVR	**0.53**	**0.97**

Table 4.6.3 Pseudo ANOVA table for Q_{100} model of Queensland (QRT, fixed region and ROI)

Source	**Degrees of Freedom**		**Sum of Squares**		
	Fixed region	ROI		Fixed region	ROI
Model	*k=3*	*k=3*		44	59
Model error δ	*n-k-1=168*	*n-k-1=92*		35	20
Sampling error η	*N = 172*	*N = 96*		19	17
Total	*2n-1 = 343*	*2n-1 = 191*	**Sum of the above =**	**98**	**97**
			EVR	**0.54**	**0.86**

Table 4.6.4 Pseudo ANOVA table for the mean flood model of Queensland (PRT, fixed region and ROI)

Source	**Degrees of Freedom**		**Sum of Squares**		
	Fixed region	ROI		Fixed region	ROI
Model	*k=3*	*k=3*	$n(\sigma^2_{\delta 0} - \sigma^2_{\delta}) =$	105	122
Model error δ	*n-k-1=168*	*n-k-1=34*	$n(\sigma^2_{\delta}) =$	39	22
Sampling error η	*N = 172*	*N = 38*	$tr[\Sigma(\hat{y})] =$	10.2	9.0
Total	*2n-1 = 343*	*2n-1 = 75*	**Sum of the above =**	**155**	**154**
			EVR	**0.26**	**0.40**

Table 4.6.5 Pseudo ANOVA table for the standard deviation model of Queensland (PRT, fixed region and ROI)

Source	**Degrees of Freedom**		**Sum of Squares**		
	Fixed region	ROI		Fixed region	ROI
Model	*k=3*	*k=3*		9	21
Model error δ	*n-k-1=168*	*n-k-1=47*		22	9.7
Sampling error η	*N = 172*	*N = 51*		7.9	7.1
Total	*2n-1 = 343*	*2n-1 = 93*	**Sum of the above =**	**39**	**38**
			EVR	**0.36**	**0.73**

Table 4.6.6 Pseudo ANOVA table for skew model of Queensland (PRT, fixed region & ROI)

Source	**Degrees of Freedom**		**Sum of Squares**		
	Fixed region	ROI		Fixed region	ROI
Model	*k=3*	*k=3*		0.11	0.65
Model error δ	*n-k-1=168*	*n-k-1=146*		2.6	2.1
Sampling error η	*N = 172*	*N = 150*		45	40
Total	*2n-1 = 343*	*2n-1 = 299*	**Sum of the above =**	**48**	**43**
			EVR	**17**	**19**

The underlying model assumptions are examined (i.e. the normality of residuals), using the plots of the standardised residuals vs. predicted values. The predicted values were obtained from one-at-a-time cross validation. Figures 4.6.3 to 4.6.5 show the plots for the Q_{20} and the mean flood models with the fixed region and ROI. It can be seen that most of the standardised residuals fall between the magnitudes of ± 2, hence the underlying model assumptions are satisfied quite well. The ROI in each case shows standardised residuals closer to the ± 2 limits indicating that ROI deals better with the sites that may be outliers. The results in Figures 4.6.3 to 4.6.5 reveal that the developed equations satisfy the normality of residual assumption quite satisfactorily. Also no specific pattern (heteroscedasicity) can be identified with the standardised values being almost equally distributed below and above zero. Similar results were obtained for the skew, standard deviation and other flood quantile models.

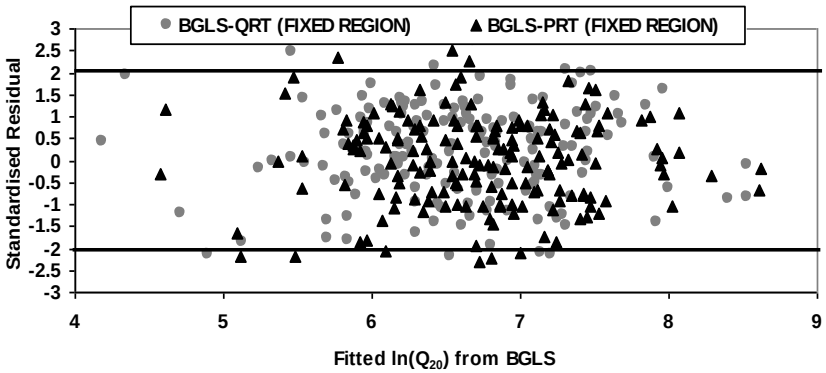


Figure 4.6.3 Plots of standardised residuals vs. predicted values for ARI of 20 years (QRT and PRT, fixed region, Queensland)

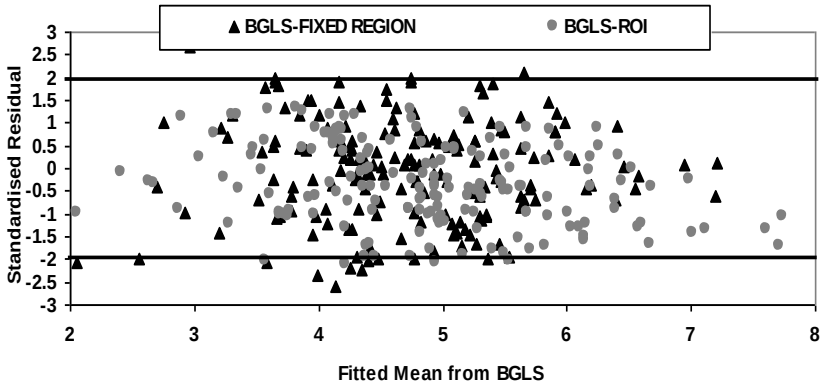


Figure 4.6.4 Plot of standardised residuals vs. predicted values for the mean flood (PRT, fixed region, ROI, Queensland)

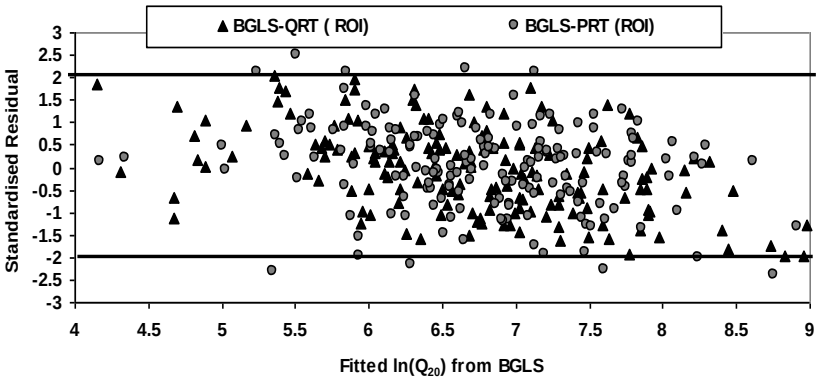


Figure 4.6.5 Plots of standardised residuals vs. predicted values for ARI of 20 years (QRT and PRT, ROI, Queensland)

The QQ-plots of the standardised residuals (Equation 4.2.13) vs. normal score (Equation 4.2.14) for the fixed region (based on one-at-a-time cross validation) and ROI were examined. Figures 4.6.6 to 4.6.8 present the results for the Q_{20} and skew models, which show that all the points closely follow a straight line. This indicates that the assumption of normality and the homogeneity of variance of the standardised residuals have largely been satisfied. The standardised residuals are indeed normally and independently distributed N(0,1), with mean 0 and variance 1, as the slope of the best fit line in the QQ-plot, which can be interpreted as the standard deviation of the normal score (Z score) of the quantile, should approach 1 and the intercept, which is the mean of the normal score of the quantile should approach 0 as the number of sites increases. It can be observed from Figures 4.6.6 to 4.6.8, that the fitted lines for the developed models pass through the origin (0, 0) and have a slope approximately equal to one. The ROI approach approximates the normality of the residuals slightly better (i.e. a better match with the fitted line) than the fixed region approach. Similar results were also found for the mean, standard deviation and other flood quantile models.

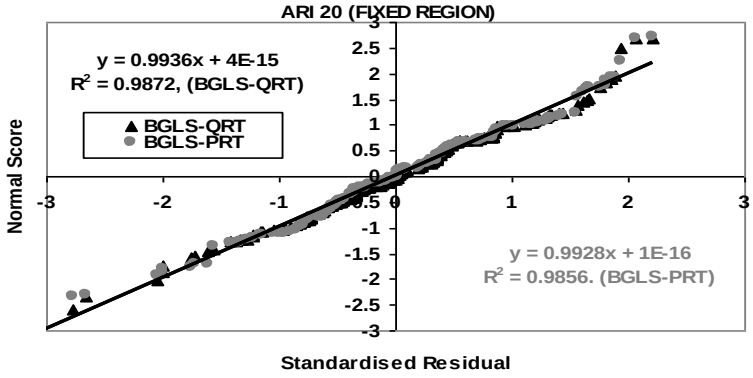


Figure 4.6.6 QQ plot of the standardised residuals vs. Z score for ARI of 20 years (QRT and PRT, fixed region, Queensland)

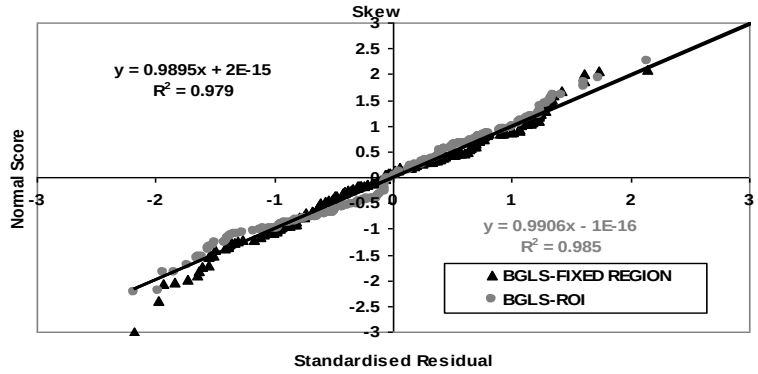


Figure 4.6.7 QQ plot of the standardised residuals vs. Z score for the skew model (PRT, fixed region, ROI, Queensland)

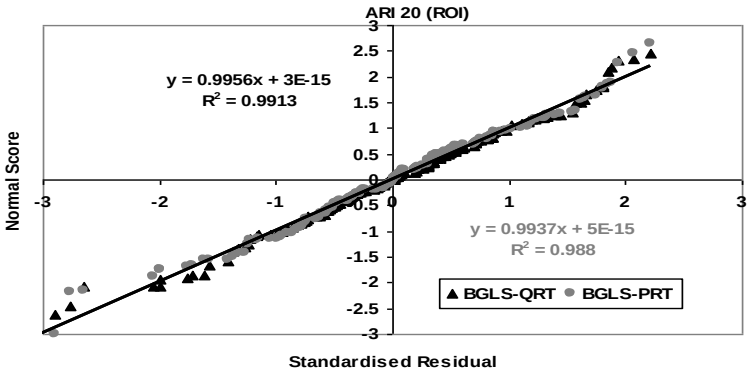


Figure 4.6.8 QQ plot of the standardised residuals vs. Z score for ARI of 20 years (QRT and PRT, ROI, Queensland)

The summary of various regression diagnostics (the relevant equations are described in Section 4.2.2) is provided in Table 4.6.7. This shows that for the mean flood model the model error variance (MEV) and average standard error of prediction (SEP) are much higher than those of the standard deviation and skew models. This indicates that the mean flood model exhibits a higher degree of heterogeneity than the standard deviation and skew models; this result is also supported by the ANOVA analysis. Indeed the issue here is that sampling error becomes larger as the order of the moment increases, therefore, in case of the skew model, the spatial variation is a second order effect that is not really detectable. For the mean flood model, the ROI shows a model error variance which is 35% smaller than for the fixed region. The R^2_{GLS} value for the mean flood model with the ROI is only 1% higher than that for the fixed region, which is negligible. Given that the model error dominates the regional analysis for the mean flood, it would be preferable that ROI be used over a fixed region analysis for developing the mean flood model for Queensland. For the standard deviation model, ROI also shows 14% smaller and 12% higher SEP and R^2_{GLS} values, respectively, again indicating the relative advantage of ROI. Again, ROI should be preferred over the fixed region analysis for the standard deviation model. From Table 4.6.7 the SEP% for the skew model is the same for ROI and the fixed region analysis. This is attributed to the fact that the skew model tends to include more sites due to the very low model error, for a big region like Queensland ROI is very close to including all the sites in region to capture the variability not accounted for by the catchment characteristics, hence in this case the model error variance is similar, leading to similar SEP values. The R^2_{GLS} values for the skew models are also the same for both the ROI and fixed region.

One can see from Table 4.6.7 that the SEP values for all the flood quanitle models are 5% to 13% smaller for the ROI cases than the fixed region; the best results are obtained for ARIs of 2 and 10 years (i.e. 13% and 12% reduction respectively). Also, the R^2_{GLS} values for ROI cases are 1% to 5% higher than the fixed region. These results support the use of the ROI approach over a fixed region.

Table 4.6.7 Regression diagnostics for fixed region and ROI for Queensland

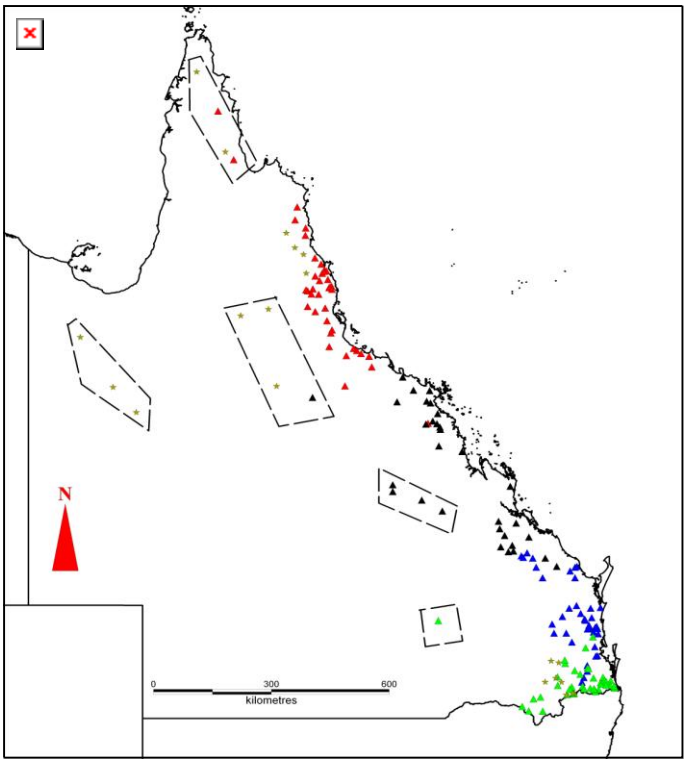
Model	Fixed region				ROI			
	MEV	AVP	SEP (%)	R^2_{GLS} (%)	MEV	AVP	SEP (%)	R^2_{GLS} (%)
Mean	0.23	0.24	52	77	0.14	0.15	40	78
Stdev	0.13	0.14	38	34	0.056	0.061	24	46
Skew	0.015	0.024	16	44	0.014	0.026	16	44
Q_2	0.26	0.27	56	75	0.15	0.18	43	79
Q_5	0.17	0.18	44	79	0.08	0.11	34	83
Q_{10}	0.18	0.19	45	74	0.07	0.11	33	79
Q_{20}	0.15	0.16	41	77	0.07	0.13	36	80
Q_{50}	0.17	0.19	45	72	0.10	0.14	39	77
Q_{100}	0.20	0.22	49	72	0.12	0.16	40	73

Table 4.6.8 shows number of sites and associated model error variances for the ROI and fixed region models. This shows that the ROI mean flood model has fewer sites (42 out of 172 i.e. 24%) than the standard deviation and skew models. The ROI skew model has the highest number of sites which includes nearly all the sites in Queensland (150 out of 172 i.e. 87%). The model error variances for all the ROI models (including the skew model) are smaller than the fixed region models. This shows that the fixed region models experience a greater heterogeneity than the ROI. If the fixed regions are made too large, the model error will be inflated by heterogeneity that will go unaccounted for by the catchment characteristics, this is especially the case for the flood quantile and mean flood models. Figure 4.6.9 shows the resulting sub-regions in Queensland (with minimum model error variances) for the ROI mean flood and standard deviation models. For the mean flood some of the stations have overlapping regions, these are the sites that showed very similar model error variances, therefore they are flexible sites and can be included in either region for estimating the mean flood. For the mean flood model there are sites that do not belong to any region as they had relatively different model error variances from the rest of the sites (these are the sites that are enclosed by a rectangle). For the standard deviation model there

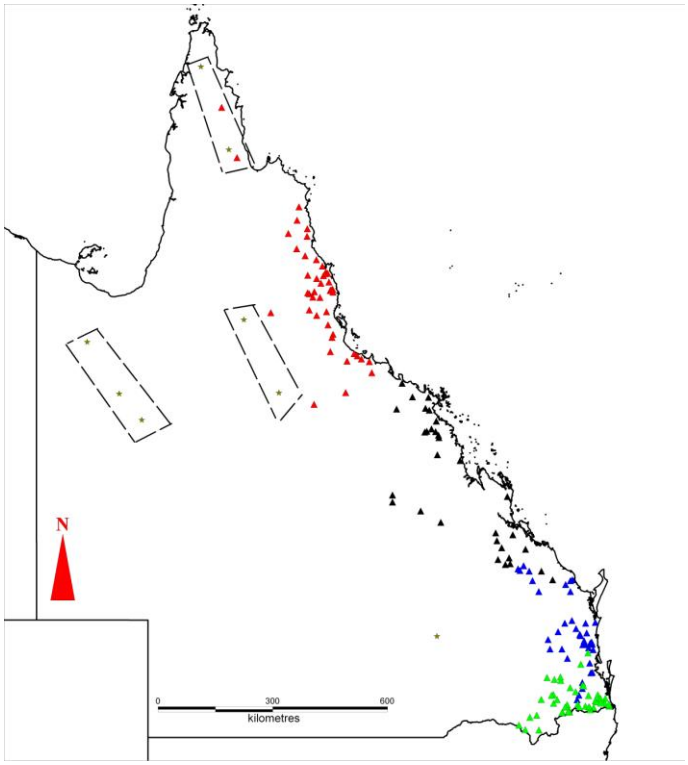
are some overlapping sites but fewer outlying sites, as the standard deviation model can include more sites in the region.

Table 4.6.8 Model error variances of the parameters and flood quantiles associated with fixed region and ROI for Queensland (n = number of sites)

Parameter/ Quantiles	Mean	Stdev	Skew	Q_2	Q_5	Q_{10}	Q_{20}	Q_{50}	Q_{100}
ROI (n)	42	65	150	60	65	74	80	88	90
$\hat{\sigma}_\delta^2$	0.15	0.056	0.014	0.14	0.08	0.07	0.07	0.10	0.12
Fixed region (n)	172	172	172	172	172	172	172	172	172
$\hat{\sigma}_\delta^2$	0.23	0.14	0.015	0.26	0.17	0.18	0.15	0.17	0.20



(a)



(b)

Figure 4.6.9 Binned minimum model error variance for (a) mean flood model and (b) standard deviation model for Queensland

Table 4.6.9 presents the root mean square error (RMSE) (Equation 4.2.16) and relative error (RE) (Equation 4.2.15) values for the PRT and QRT models with both the fixed region and ROI. In terms of RMSE, ROI gives smaller values than the fixed regions for all the ARIs. The QRT-ROI shows smaller RMSE values than the PRT-ROI for all the ARIs. The best result was found for the 2-year ARI for PRT-ROI and the 20-year ARI for the QRT-ROI. In terms of RE, ROI gives 0 to 5% smaller values than the fixed regions. The PRT-ROI gives larger values of RE (by 5 and 9%) for both the 50 and 100 years ARIs respectively. For the ARI of 2 years the PRT-ROI gives a smaller RE value (by 4%) than the QRT-ROI. These results suggest modest differences between the performances of the QRT and PRT for Queensland (similar to NSW & ACT and Victoria).

Finally the results of counting the Q_{pred}/Q_{obs} ratios for the QRT and PRT (for both the ROI and fixed regions) are provided in Tables 4.6.10 and 4.6.11. The QRT-ROI had 89% ratio values in the desirable range, compared to 84% for the QRT-fixed region. The PRT-ROI had 85% ratio values in the desirable range, compared to 79% for the PRT-fixed region. These results show that ROI performs better than the fixed regions with both the QRT and PRT. The PRT-ROI shows 4% underestimation as compared to 6% for the QRT-ROI. The cases for overestimation are 11% and 4% for PRT-ROI and QRT-ROI, respectively. These results favour the PRT-ROI, as slight over-estimation is preferable to underestimation.

Table 4.6.9 Evaluation statistics (RMSE and RE) from one-at-a-time cross validation for Queensland

Model	RMSE (%)				RE (%)			
	PRT		QRT		PRT		QRT	
	Fixed region	ROI	Fixed region	ROI	Fixed region	ROI	Fixed region	ROI
Q_2	8.2	6.9	6.1	5.6	39	35	39	39
Q_5	6.8	6.0	4.8	4.4	33	34	34	32
Q_{10}	6.9	6.0	5.2	4.7	34	30	32	31
Q_{20}	7.2	6.5	5.0	4.4	35	33	31	29
Q_{50}	7.8	6.8	5.3	4.9	37	36	32	31
Q_{100}	8.5	7.9	5.8	5.3	41	40	36	31

Table 4.6.10 Summary of counts based on Q_{pred}/Q_{obs} ratio values for QRT and PRT for Queensland (fixed region). "U" = gross underestimation, "D" = desirable range and "O" = gross overestimation

Model	Count (QRT)			Percent (QRT)			Count (PRT)			Percent (PRT)		
	U	D	O	U	D	O	U	D	O	U	D	O
Q_2	20	139	13	12	81	8	19	129	24	11	75	14
Q_5	9	154	9	5	90	5	9	141	22	5	82	13
Q_{10}	17	143	12	10	83	7	10	141	21	6	82	12
Q_{20}	13	149	10	8	87	6	11	143	18	6	83	10
Q_{50}	18	141	13	10	82	8	17	132	23	10	77	13
Q_{100}	20	138	14	12	80	8	24	125	23	14	73	13
Sum / average	**97**	**864**	**71**	**9**	**84**	**7**	**90**	**811**	**131**	**9**	**79**	**13**

Table 4.6.11 Summary of counts based on Q_{pred}/Q_{obs} ratio values for QRT and PRT for Queensland (ROI). "U" = gross underestimation, "A" = desirable range and "O" = gross overestimation

ARI (years)	Count (QRT)			Percent (QRT)			Count (PRT)			Percent (PRT)		
	U	D	O	U	D	O	U	D	O	U	D	O
2	21	140	11	12	81	6	1	144	27	1	84	16
5	11	157	4	6	91	2	3	149	20	2	87	12
10	11	155	6	6	90	3	5	149	18	3	87	10
20	9	157	6	5	91	3	6	147	19	3	85	11
50	6	156	10	3	91	6	11	145	16	6	84	9
100	5	158	9	3	92	5	11	145	16	6	84	9
Sum / average	**63**	**923**	**46**	**6**	**89**	**4**	**37**	**879**	**116**	**4**	**85**	**11**

4.7 Results for the NT

From the NT, a total of 55 catchments were selected; 51 of these catchments were located in Drainage Division VIII (Timor Sea Division), the remaining 4 catchments were from Drainage Division IX. Here, the 51 catchments are used to develop the RFFA model, which is applicable to north-western NT i.e. the part of Drainage Division VIII (Timor Sea Division) falling in the NT. The locations of these 51 stations are shown in Figure 4.7.1.

The annual maximum flood series record lengths of these 51 stations range from 19 to 54 years (mean 35 years, median 37 years and standard deviation 11.5 years). The catchment areas of these 51 stations range from 1.4 to 4325 km^2 (mean 581 km^2, median 352 km^2 and standard deviation of 782 km^2).

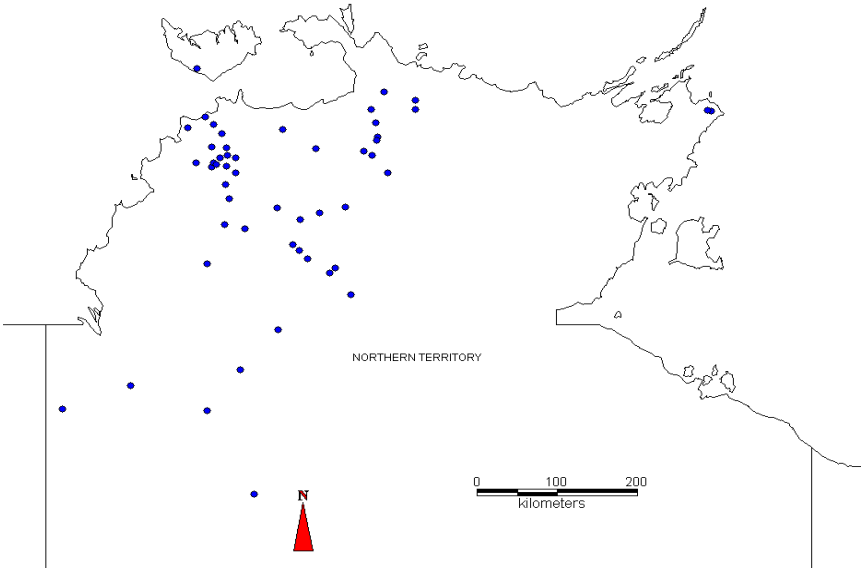


Figure 4.7.1 Geographical distributions of the selected 51 catchments from the NT (Drainage Division VIII -Timor Sea Division)

In the fixed region approach, a parameter regression technique is used, all the 51 catchments were considered to have formed one region, however, one catchment was left out for cross-validation and the procedure was repeated 51 times to implement the one-at-a-time cross validation. No region-of-influence (ROI) approach was used for the NT as the total number of stations is too small with a highly sparse desnsity for the ROI application.

Table 4.7.1 shows different combinations of predictor variables for first three moments of the LP3 distribution. Figures 4.7.2 and 4.7.3 show example plots of the statistics used in selecting the best set of predictor variables for the mean flood and the skew models.

According to the model error variance, combinations 2 to 11 were the top potential sets of predictor variables for the mean flood model. Combination 2 only contained 1 predictor variable area, while combinations 3, 4, 5, 6 and 7 contained 2 predictor variables: area and basic design rainfall intensity. Combinations 3, 4, 5, 6 and 7 showed no significant difference in $R^2_{GLS.}$ There are also negligible differences in the model error variances. Combinations 3, 4, 5 and 6 all had negative coefficients of design rainfall intensity, therefore it was decided to go for combination 7 (area and rain). The AVPO, AVPN, AIC and BIC for combination 7 was comparable to combinations 3, 4, 5 and 6. The regression coefficient for the variable rain was approximately 2 times the posterior standard deviation away from zero compared to 1 time for the design rainfall intensity.

This suggested that mean annual rainfall (rain) is preferable to design rainfall intensity, unlike other Australian states. It should also be noted here that design rainfall intensity does not show any relationship with the mean flood or other flood quantiles (see Rahman et al., 2011), which is somewhat unexpected, as for all other Australian states, rainfall intensity has appeared to be the 2nd most important predictor variable after area (Rahman *et* al., 2009; Haddad *et* al., 2009; Rahman 2005). The study by Weeks and Rajaratnam (2005) also found that design rainfall intensity does not appear in the prediction equations in the NT. It appears that there might be some problems with the ARR87 design rainfall data for the NT, it may be that only few stations were used to derive ARR87 design rainfall data for the NT. Once the updated design rainfall intensity data are available with the 4th edition of the ARR, the regional flood frequency analysis methods developed here for the NT need to be updated, which might include design rainfall intensity as a predictor variable, similar to the other Australian states.

For the skew model, combination 1 with no explanatory variables showed one of the lowest model error variances (0.0286) and the lowest AVPO and AVPN (see Figure 4.7.3). The next best combination was 15 with a slightly smaller model error variance; however the R^2_{GLS} was very poor (smaller than 24%). Therefore there was enough evidence to stay with combination 1, as the other models did not show major improvement in model error variance with the added explanatory variables. In this case, given the relatively small difference in model error variance, it may be argued that a regional average skew is applicable for the NT. A similar outcome was obtained for the standard deviation model.

The set of predictor variables selected as above was used in the fixed region regression with one-at-a-time cross validation approach. The BPVs for the mean flood model were 0% and 31% for both the predictor variables (area and rain respectively). This does suggest that rain

is not a particularly good descriptor in this case. For the design rainfall intensities the BPVs ranged from 31% to 59%, thus rain was the better alternative. Indeed there could be some problem with the design rainfall intensities data for the NT, as discussed before. Regression equations developed for the PRT for the fixed region are given by Equations 4.7.1 to 4.7.3.

Table 4.7.1 Different combinations of predictor variables considered for the the parameters of the LP3 distribution for the NT (PRT fixed region)

Combination	Combinations for mean, standard deviation & skew models
1	Const
2	Const, area
3	Const, area, $^{2}I_{1}$
4	Const, area, $^{50}I_{1}$
5	Const, area, $^{50}I_{12}$
6	Const, area, $^{2}I_{12}$
7	Const, area, rain
8	Const, area, S1085
9	Const, area, evap
10	Const, area, E85*
11	Const, area, SL*
12	Const, SL, rain
13	Const, S1085, rain
14	Const, E85, S1085
15	Const, evap
16	Const, rain, evap

* E85 = Slope S1085

* SL = Stream Length

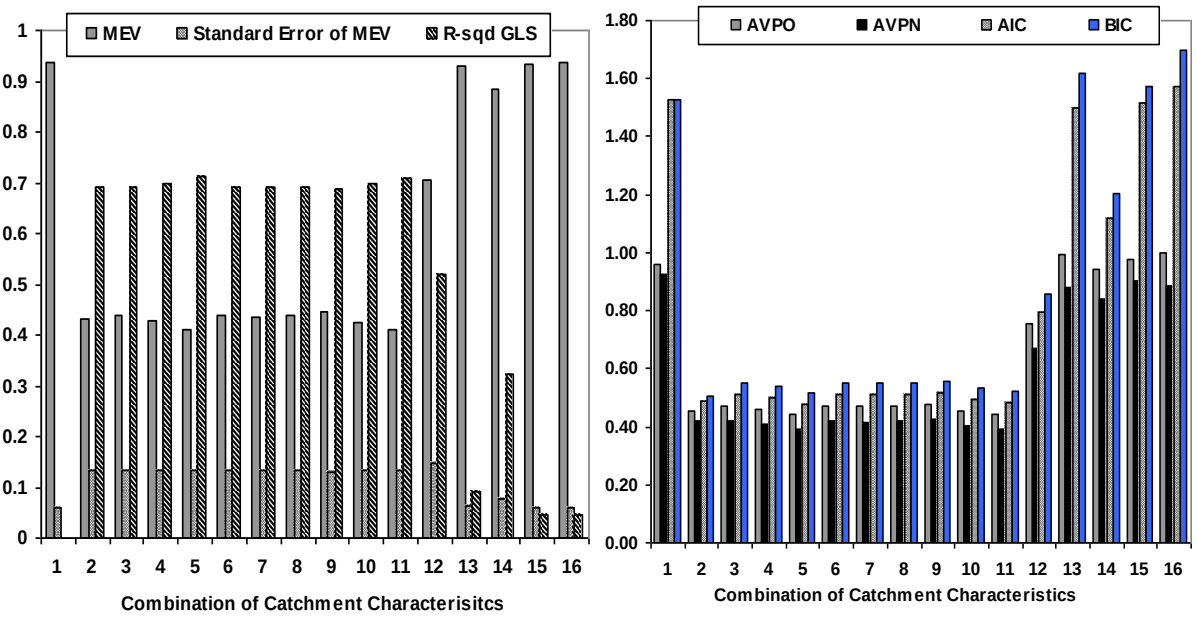


Figure 4.7.2 Selection of predictor variables for the BGLS regression model for the mean flood (PRT, fixed region NT), MEV = model error variance, AVPO = average variance of prediction (old), AVPN = average variance of prediction (new) AIC = Akaike information criterion, BIC = Bayesian information criterion

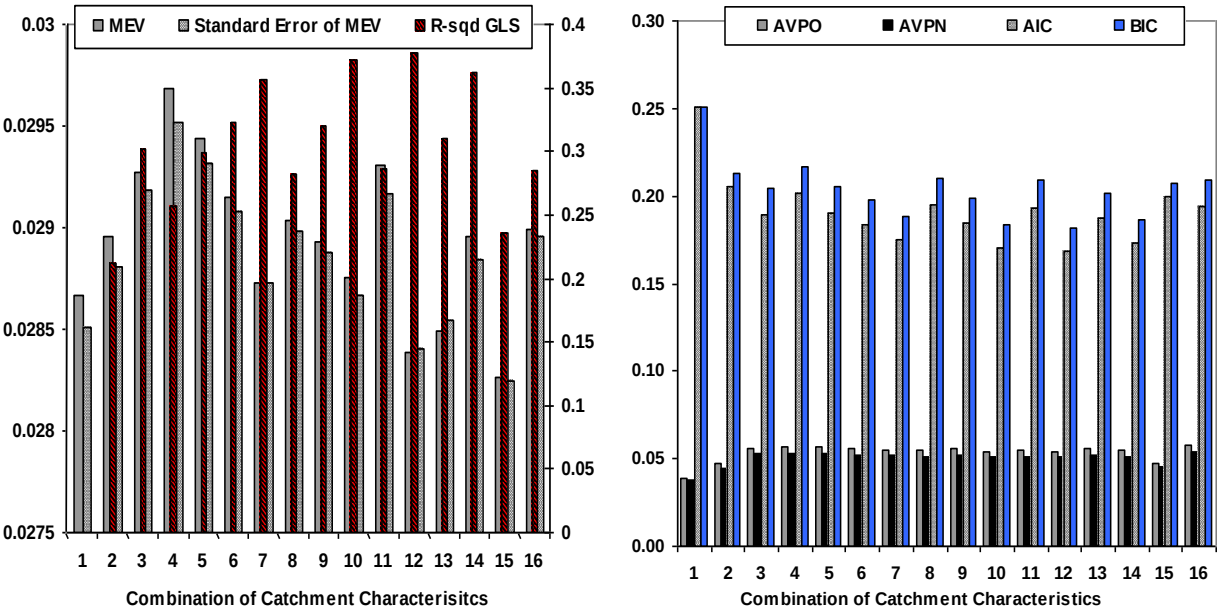


Figure 4.7.3 Selection of predictor variables for the BGLS regression model for skew (R-sqd GLS uses right hand axis)

M = 4.40 + 0.61[ln(*area*) - 5.41] + 0.43[ln(*rain*) - 7.17] (4.7.1)

stdev = 1.04 (4.7.2)

skew = - 0.90 (4.7.3)

The Pseudo Analysis of Variance (ANOVA) tables for the parameters of the LP3 distribution are presented in Tables 4.7.2 and 4.7.4. For the LP3 parameters, the sampling error increases as the order of moment increases, i.e. the EVR (sampling error to model error ratio) increases with the order of moments. The model error dominates the regional analysis for the mean flood and the standard deviation models; this is more pronounced for the mean flood model (0.09 compared to 0.28). The EVR value for the skew model is 12 (Table 4.7.4) which is much higher than that of the mean flood and standard deviation models. This indicates that the skew model is dominated by sampling error and in this case the GLS regression modeling framework should be the preferred. These results indicate, as found for NSW, Victoria, Tasmania and Queensland, that the mean flood shows greater levels of heterogeneity when compared to the standard deviation and skew models.

Table 4.7.2 Pseudo ANOVA table for the mean flood model of the NT (PRT)

Source	**Degrees of Freedom**	**Sum of Squares**
Model	*k=3*	26.0
Model error δ	*n-k-1=47*	22.0
Sampling error η	*N = 51*	2.05
Total	*2n-1 = 101*	**50.05**
	EVR	**0.09**

Table 4.7.3 Pseudo ANOVA table for the standard deviation model of the NT (PRT)

Source	**Degrees of Freedom**	**Sum of Squares**
Model	*k=1*	0.0
Model error δ	*n-k-1=49*	4.7
Sampling error η	*N = 51*	1.32
Total	*2n-1 = 101*	**5.98**
	EVR	**0.28**

Table 4.7.4 Pseudo ANOVA table for the skew model of the NT (PRT)

Source	Degrees of Freedom	Sum of Squares
Model	*k=1*	0.0
Model error δ	*n-k-1=49*	1.46
Sampling error η	*N = 51*	17.5
Total	*2n-1 = 101*	**18.99**
	EVR	**12.0**

The underlying model assumptions are examined (i.e. the normality of residuals), using the plots of the standardised residuals vs. predicted values. The predicted values were obtained from one-at-a-time cross validation. Figure 4.7.4 shows the plots for Q_{20} and Q_{50} estimated by PRT. It can be seen that most of the standardised residuals fall between the magnitudes of ± 2, hence the underlying model assumptions are satisfied satisfactorily. The result in Figure 4.7.4 reveals that the developed equations satisfy the normality of residual assumption quite satisfactorily. Also no specific pattern (heteroscedasicity) can be identified, with the standardised values being almost equally distributed below and above zero. Similar results were obtained for the mean, standard deviation, skew and other flood quantile models estimated by PRT.

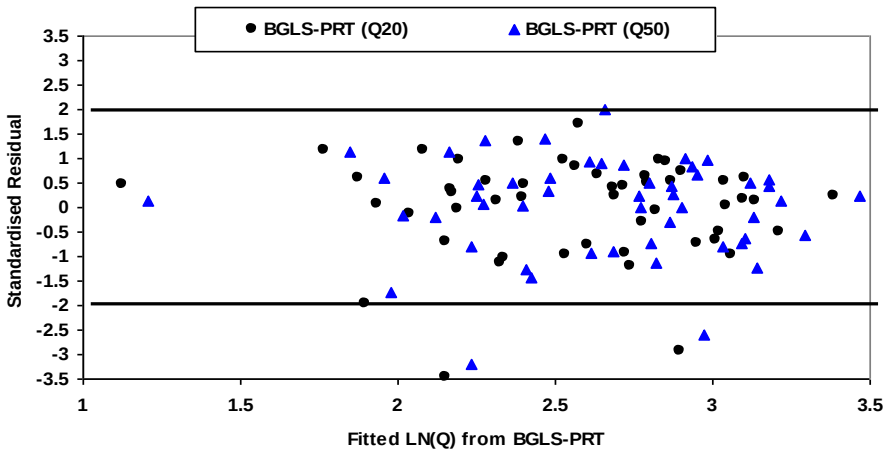


Figure 4.7.4 plots of standardised residuals vs. predicted values for ARI of 20 and 50 years (PRT, NT)

The QQ-plots of the standardised residuals (Equation 4.2.13) vs. normal score (Equation 4.2.14) for the one-at-a-time cross validation were examined. Figures 4.7.5 and 4.7.6 present the results for Q_{20}, Q_{50} and the mean flood models, which show that most of the points closely follow a straight line, while some points also fall away from the line. This indicates

that the assumption of normality and the homogeneity of variance of the standardised residuals have been satisfied to some degree. The standardised residuals are approximately normally and independently distributed N(0,1) (with mean 0 and variance 1) as the slope of the best fit line in the QQ-plot, which can be interpreted as the standard deviation of the normal score (Z score) of the quantile, should approach 1 and the intercept, which is the mean of the normal score of the quantile should approach 0 as the number of sites increases. It can be observed from Figures 4.7.5 and 4.7.6 that the fitted lines for the developed models pass through the origin (0, 0) and have a slope approximately equal to one. Similar results were also found for the standard deviation and skew models and for other flood quantiles estimated from the PRT.

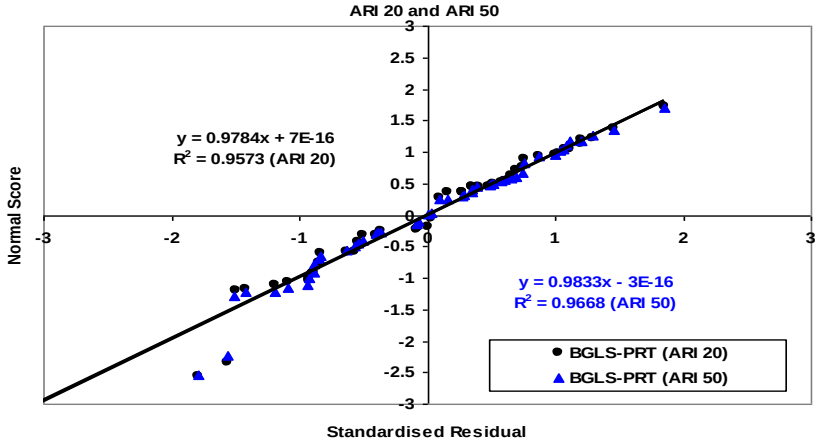


Figure 4.7.5 QQ plot of the standardised residuals vs. Z score (ARI of 20 and 50 years, PRT)

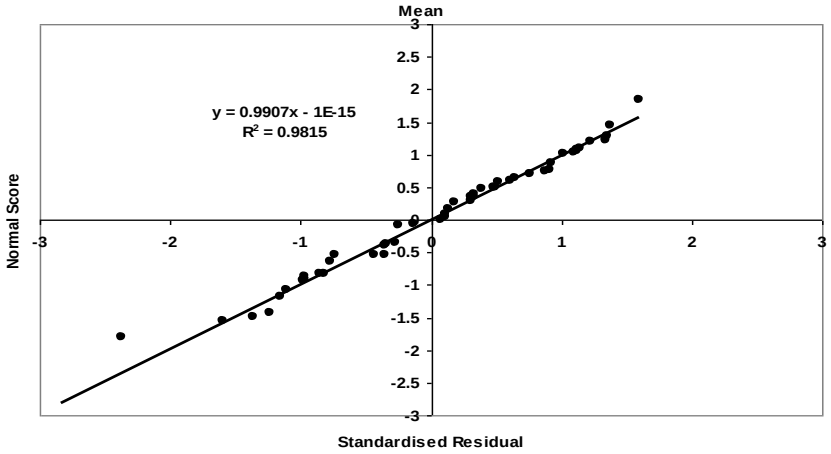


Figure 4.7.6 QQ plot of the standardised residuals vs. Z score for the mean flood model (PRT) for the NT

The summary of various regression diagnostics (the relevant equations are described in Section 4.2.2) is provided in Table 4.7.5. This shows that for the mean flood model the model

error variance (MEV) and average standard error of prediction (SEP) are much higher than those of the standard deviation and skew models. The large model error variance for the mean flood model indicates a higher degree of heterogeneity than the standard deviation and skew models; this result is in line with the ANOVA analysis and results for other Australian states. For the standard deviation and skew models, the SEP was 32% and 20% respectively, again indicating that both the standard deviation and skew may be regionalised more accurately than the mean flood. The R^2_{GLS} values for the standard deviation and skew models were 0%, as no variables were used, i.e. a regional average value is adopted. The regional average in this case is not the simple arithmetic average, but it is the regional average skew weighted by the error covariance matrix.

Table 4.7.5 Regression diagnostics for fixed region analysis for NT

Model	Fixed region			
	MEV	AVP	SEP (%)	R^2_{GLS} (%)
Mean	0.45	0.49	79	68
Stdev	0.093	0.097	32	0
Skew	0.029	0.040	20	0

Table 4.7.6 presents the values of root mean square error (RMSE) (Equation 4.2.16) and relative error (RE) (Equation 4.2.15) for the quantiles estimated by the PRT. In terms of RMSE, the 2 years ARI shows the highest value. The smallest value is found for the 100-year ARI. The RE values range from 36% to 54%, the smallest value being for the 5 years ARI. Again, the RE values are considered to be in a reasonable range for the ARIs considered in this study and compare reasonably well with the other Australian states.

Table 4.7.7 shows the results of counting the Q_{pred}/Q_{obs} ratios for the PRT method. There are on average 75% of cases in the desirable estimation range. The cases for overestimation on average are 7%, while the underestimation is 18% on average.

Table 4.7.6 Evaluation statistics (RMSE and RE) from one-at-a-time cross validation for NT

Model	RMSE (%)	RE (%)
	PRT	PRT
Q_2	20	42
Q_5	16	36
Q_{10}	13	39
Q_{20}	11	42
Q_{50}	9	48
Q_{100}	8	54

Table 4.7.7 Summary of counts based on Q_{pred}/Q_{obs} ratio values for PRT for NT (fixed region). "U" = gross underestimation, "D" = desirable range and "O" = gross overestimation

Model	Count (PRT)			Percent (PRT)		
	U	D	O	U	D	O
Q_2	7	39	5	14	76	10
Q_5	7	39	5	14	76	10
Q_{10}	7	41	3	14	80	6
Q_{20}	8	40	3	16	78	6
Q_{50}	11	37	3	22	73	6
Q_{100}	14	35	2	27	69	4
Sum / average	**54**	**231**	**21**	**18**	**75**	**7**

4.8 Results for Western Australia

The WA was divided into three distinct regions in consultation with the WA state team. This was done since WA is too large and there are concentrations of stream gauging stations in three parts of WA, which are separated by long distances. These three regions are:

Kimberley region: 14 stations (top part of WA from Drainage Division VIII)
Pilbara region: 12 stations (middle western part of WA, Drainage Division VII)
South-west WA: 120 stations (Drainage Division VI)

The locations of these stations can be seen in Figure 2.17 and their details can be found in Appendix. The RFFA methods are developed separately for each of these three regions, as detailed below.

4.8.1 Kimberley region

A total of 14 catchments were used from the Kimberley region. The annual maximum flood series record lengths of these 14 stations range from 23 to 42 years (mean 33 years, median 31 years and standard deviation 6 years). The catchment areas of these 14 stations range from 30.6 to 7405.70 km^2 (mean 1739 km^2, median 701 km^2 and standard deviation of 2343 km^2).

A fixed region approach was considered, due to the small number of stations in the region, i.e. all the 14 catchments are considered to have formed one region. One catchment at a time was left out for cross-validation; the procedure was repeated 14 times so that each of the 14 catchments is tested independently. No region-of-influence (ROI) approach was used for the Kimberley region.

Table 4.8.1 shows different combinations of predictor variables for the Q_{10} QRT model and for models of the first three moments of the LP3 distribution. Figures 4.8.1 and 4.8.2 show example plots of the statistics used in selecting the best set of predictor variables for the Q_{10} flood quantile and the skew. The model error variance shows that combinations 3, 4, 5, 6, 7 and 8 are the top potential sets of predictor variables for the Q_{10} model. All these combinations did not differ greatly in model error variance and R^2_{GLS} values. The AVPO, AVPN, AIC and BIC were also very similar. In this case combination 6 with variables area

and $^2I_{12}$ was adopted, similar to most other Australian states. It was also found that the variable $^2I_{12}$ was significant in the regression analysis, with the regression coefficient being greater than three times the posterior standard deviation away from zero.

For the skew model, combination 1 with no explanatory variable had the lowest model error variance of 0.0858. The AVPO and AVPN values were also lowest with combination 1 (see Figure 4.8.2). There is enough evidence to stay with combination 1 (it may be argued that a regional average skew maybe applicable), as the increase in R^2_{GLS} value for combination 7 was insignificant. Also the variables with combination 7 are not highly significant as the regression coefficients with both area and rain were less than two standard deviations away from zero. In this case, combination 1 was adopted. A similar outcome was obtained for the standard deviation model.

A similar procedure as discussed above was adopted in selecting the best set of predictor values for other models with the QRT and PRT. The set of predictor variables selected as above were used in the one-at-a-time cross validation approach.

To assess statistical significance of the variables, the BPV values for the regression coefficients were calculated. For the QRT (for all the ARIs) the BPV values were smaller than 0.000 and 1% for area and rainfall intensity ($^2I_{12}$) respectively. Hence, the inclusion of predictor variables area and rainfall intensity in the prediction equations for the QRT was justified. Regression equations developed for the QRT and PRT for the fixed region are given by Equations 4.8.1 to 4.8.9.

Table 4.8.1 Different combinations of predictor variables considered for the QRT models and the parameters of the LP3 distribution (QRT and PRT fixed region, Kimberley region WA)

Combination	Combinations for mean, standard deviation, skew and Flood Quantile Models
1	Const
2	Const, area
3	Const, area, ($^{2}I_{1}$)
4	Const, area, ($^{50}I_{1}$)
5	Const, area, ($^{50}I_{12}$)
6	Const, area, ($^{2}I_{12}$)
7	Const, area, rain
8	Const, area, evap
9	Const, rain
10	Const, rain, evap
11	Const, evap

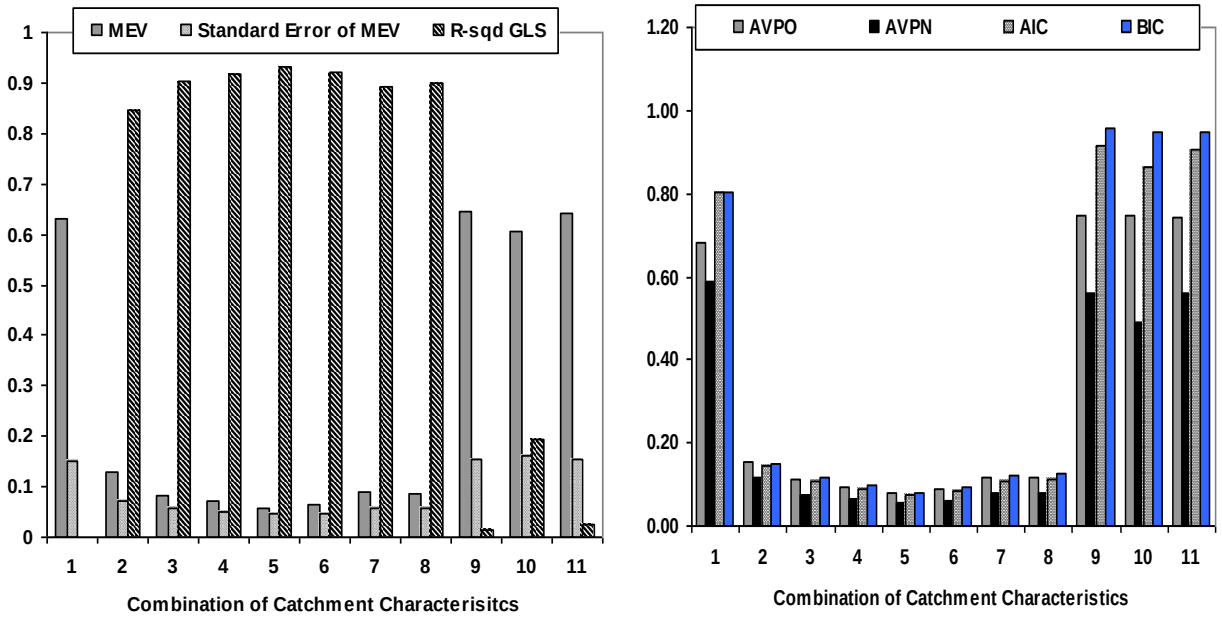


Figure 4.8.1 Selection of predictor variables for the BGLS regression model for Q_{10} (QRT, fixed region, Kimberley region WA), MEV = model error variance, AVPO = average variance of prediction (old), AVPN = average variance of prediction (new) AIC = Akaike information criterion, BIC = Bayesian information criterio

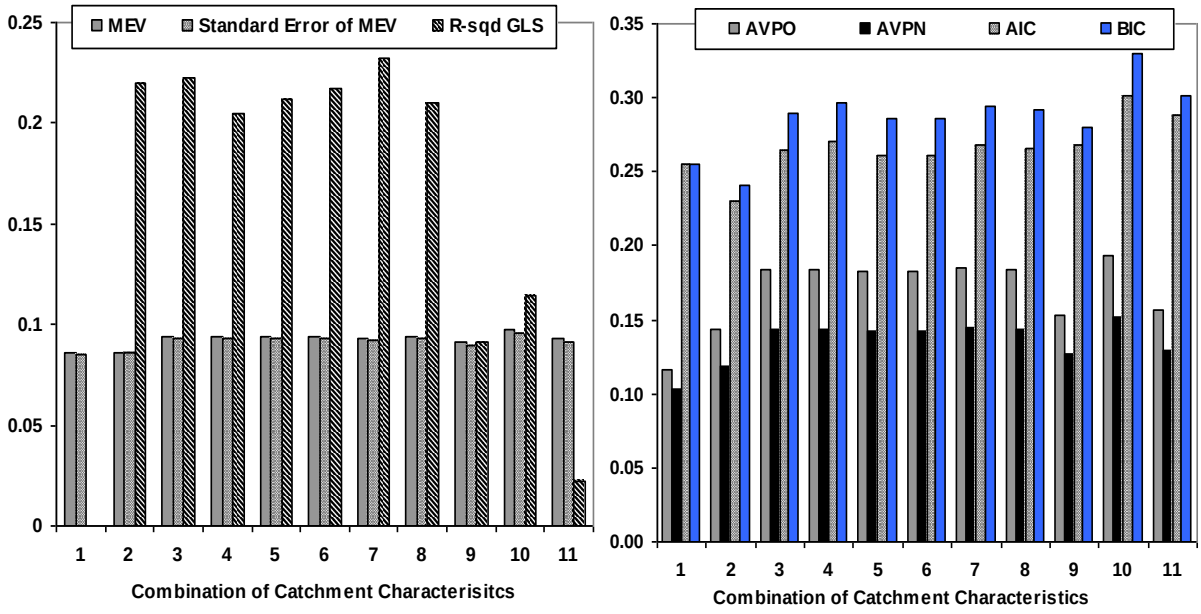


Figure 4.8.2 Selection of predictor variables for the BGLS regression model for skew

Equations for QRT:

$\ln(Q_2) = 5.93 + 0.48[\ln(area) - 6.46] + 0.93[\ln(I_{12,2}) - 1.87]$ (4.8.1)

$\ln(Q_5) = 6.71 + 0.53[\ln(area) - 6.46] + 1.44[\ln(I_{12,2}) - 1.87]$ (4.8.2)

$\ln(Q_{10}) = 7.09 + 0.56[\ln(area) - 6.46] + 1.66[\ln(I_{12,2}) - 1.87]$ (4.8.3)

$\ln(Q_{20}) = 7.39 + 0.57[\ln(area) - 6.46] + 1.83[\ln(I_{12,2}) - 1.87]$ (4.8.4)

$\ln(Q_{50}) = 7.72 + 0.59[\ln(area) - 6.46] + 1.98[\ln(I_{12,2}) - 1.87]$ (4.8.5)

$\ln(Q_{100}) = 7.94 + 0.60[\ln(area) - 6.46] + 2.08[\ln(I_{12,2}) - 1.87]$ (4.8.6)

Equations for PRT:

$M = 5.79 + 0.48[\ln(area) - 6.46] + 0.75[\ln(I_{12,2}) - 1.87]$ (4.8.7)

$stdev = 1.05$ (4.8.8)

$skew = - 0.88$ (4.8.9)

The Pseudo Analysis of Variance (ANOVA) tables for the Q_2, Q_{20} and Q_{100} models and the parameters of the LP3 distribution are presented in Tables 4.8.2 and 4.8.7.

For the LP3 parameters, the sampling error increases as the order of moment increases i.e. the EVR (sampling error to model error ratio) increases with the order of the moment. The model error dominates the regional analysis for the mean flood and the standard deviation models; this is more pronounced for the mean flood model (EVR = 0.46 compared to EVR =

0.95, mean and standard deviation repectively). However these results also indicate that the sampling error has had a slight effect on the analysis, as the EVR values are larger than 0.2. The EVR value for the skew model is 4.0 (Table 4.8.7) which is much higher than that of the mean flood and standard deviation models. The GLS regression modeling framework in this case has performed quite well in the estimation of the parameters of the LP3 distribution. It should be remembered, however, that there are only 14 stations in this region, thus the sampling error might be overestimated in this case.

Pseudo ANOVA tables were also prepared for the flood quantile models. Tables 4.8.2 to 4.8.4 show the results for the Q_2, Q_{20} and Q_{100} models, respectively. What is clear is that the model errors are relatively high. The Q_2 model has shown the highest heterogeneity in this case (EVR = 0.42). The Q_{100} model shows the highest EVR, as expected (EVR = 0.90).

Table 4.8.2 Pseudo ANOVA table for Q_2 model of Kimberley Region WA (QRT)

Source	**Degrees of Freedom**	**Sum of Squares**
Model	*k=3*	$n(\sigma^2_{\delta 0} - \sigma^2_{\delta}) = 6.4$
Model error δ	*n-k-1=10*	$n(\sigma^2_{\delta}) = 1.0$
Sampling error η	*N = 14*	$tr[\Sigma(\hat{y})] = 0.4$
Total	*2n-1 = 27*	**Sum of the above = 7.8**
	EVR	**0.42**

Table 4.8.3 Pseudo ANOVA table for Q_{20} model of Kimberley Region WA (QRT)

Source	**Degrees of Freedom**	**Sum of Squares**
Model	*k=3*	8.25
Model error δ	*n-k-1=10*	0.8
Sampling error η	*N = 14*	0.6
Total	*2n-1 = 27*	**9.7**
	EVR	**0.75**

Table 4.8.4 Pseudo ANOVA table for Q_{100} model of Kimberley Region WA (QRT)

Source	Degrees of Freedom	Sum of Squares
Model	*k=3*	8.6
Model error δ	*n-k-1=10*	0.9
Sampling error η	*N = 14*	0.8
Total	*2n-1 = 27*	**10.4**
	EVR	**0.90**

Table 4.8.5 Pseudo ANOVA table for mean flood model of Kimberley Region WA (PRT)

Source	Degrees of Freedom	Sum of Squares
Model	*k=3*	6.2
Model error δ	*n-k-1=10*	1.0
Sampling error η	*N = 14*	0.48
Total	*2n-1 = 27*	**7.7**
	EVR	**0.46**

Table 4.8.6 Pseudo ANOVA table for standard deviation model, Kimberley Region WA (PRT)

Source	Degrees of Freedom	Sum of Squares
Model	*k=1*	0.0
Model error δ	*n-k-1=12*	0.5
Sampling error η	*N = 14*	0.35
Total	*2n-1 = 27*	**0.89**
	EVR	**0.65**

Table 4.8.7 Pseudo ANOVA table for skew model of Kimberley Region WA (PRT)

Source	Degrees of Freedom	Sum of Squares
Model	*k=1*	0.0
Model error δ	*n-k-1=12*	1.16
Sampling error η	*N = 14*	4.6
Total	*2n-1 = 27*	**5.8**
	EVR	**4.0**

The underlying model assumptions are examined (i.e. the normality of residuals), using the plots of the standardised residuals vs. predicted values (see below). The predicted values were obtained from one-at-a-time cross validation. Figures 4.8.3 and 4.8.4 show the plots for Q_{20} and Q_{100}. It can be seen that most of the standardised residuals fall between the magnitudes of ± 2 for Q_{20} and Q_{100} respectively, hence the underlying model assumptions are satisfied satisfactorily for both QRT and PRT. Also no specific trend can be identified, with the standardised values being almost equally distributed below and above zero. Similar results were obtained for the mean, standard deviation, skew and other flood quantile models.

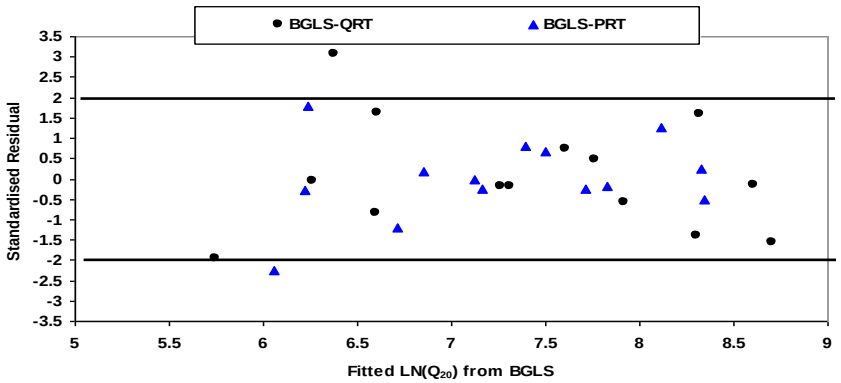


Figure 4.8.3 plots of standardised residuals vs. predicted values for ARI of 20 years (QRT and PRT, Kimberley Region, WA)

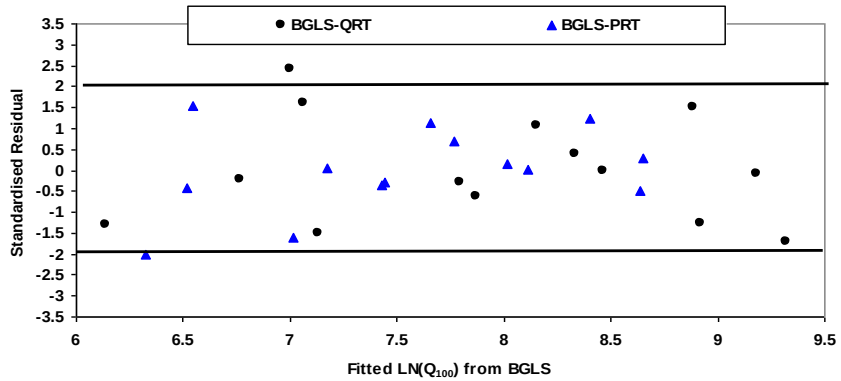


Figure 4.8.4 plots of standardised residuals vs. predicted values for ARI of 100 years (QRT and PRT, Kimberley Region, WA)

The summary of various regression diagnostics (the relevant equations are described in Section 4.2.2) is provided in Table 4.8.8. One can see from Table 4.8.8 that the SEP values for all the flood quanitle models ranged from 31% to 33%, which are quite reasonable as compared to the other Australian states/regions. While the sampling error has had a slight impact on the analysis (see ANOVA results), the moderately large model errors have still dominated the regional analysis; this has been reflected in the moderately high AVPs and SEPs (see Table 4.8.8). The lowest SEP was found for the ARIs of 5, 10 and 20 years. The highest R^2_{GLS} value was found for the 20 and 50 years ARIs (93%). Overall, the R^2_{GLS} values obtained (88% to 93%) seem to be relatively high when compared to the other states of Australia. In saying this, however, it must be kept in mind that the analysis has been undertaken on a small sample with only 14 stations; thus little confidence can be placed on the above statistics.

Table 4.8.8 Regression diagnostics for fixed region analysis for Kimberly (14 stations)

Model	Fixed region			
	MEV	AVP	SEP (%)	R^2_{GLS} (%)
Q_2	0.076	0.10	33	88
Q_5	0.069	0.093	31	91
Q_{10}	0.066	0.091	31	92
Q_{20}	0.063	0.091	31	93
Q_{50}	0.066	0.097	32	93
Q_{100}	0.071	0.11	33	92

Table 4.8.9 presents the root mean square error (RMSE) (Equation 4.2.16) and relative error (RE) (Equation 4.2.15) values for the PRT and QRT models. In terms of RMSE, PRT over all the ARIs provides negligibly higher values as compared to QRT (A difference of 0.1% to 0.9% over all the ARIs). In terms of RE, PRT gives 1 to 4% smaller values than the QRT, except for the 100-year ARI, where QRT is smaller by 3%. Overall it can be seen that QRT and PRT perform similarly over all the ARIs considered here.

Table 4.8.10 shows the results of counting the Q_{pred}/Q_{obs} ratios for the QRT and PRT. The use of this ratio has been discussed in Section 4.2.4. There are more desirable cases on average for the QRT than the PRT, i.e. 95% vs. 87%. The PRT and QRT on average show similar gross underestimation (i.e. 6% and 4%, PRT and QRT respectively). The PRT shows slightly more cases of overestimation on average as compared to QRT (7% and 1% for PRT

and QRT respectively). Overll, there are only modest differences between the performances of the QRT and PRT (similar to other states).

Table 4.8.9 Evaluation statistics (RMSE and RE) from one-at-a-time cross validation for the Kimberley Region, WA

Model	RMSE (%)		RE (%)	
	PRT Fixed region	QRT Fixed region	PRT Fixed region	QRT Fixed region
Q_2	4.3	4.2	22	23
Q_5	4.6	3.7	20	22
Q_{10}	4.4	3.5	18	21
Q_{20}	4.3	3.4	21	25
Q_{50}	4.3	3.6	30	33
Q_{100}	4.4	3.8	38	35

Table 4.8.10 Summary of counts based on Q_{pred}/Q_{obs} ratio values for QRT and PRT for Kimberley Region, WA (fixed region). "U" = gross underestimation, "D" = desirable range and "O" = gross overestimation

Model	Count (QRT)			Percent (QRT)			Count (PRT)			Percent (PRT)		
	U	D	O	U	D	O	U	D	O	U	D	O
Q_2	0	13	1	0	93	7	0	13	1	0	93	7
Q_5	1	13	0	7	93	0	1	12	1	7	86	7
Q_{10}	1	13	0	7	93	0	1	12	1	7	86	7
Q_{20}	1	13	0	7	93	0	1	12	1	7	86	7
Q_{50}	0	14	0	0	100	0	1	12	1	7	86	7
Q_{100}	0	14	0	0	100	0	1	12	1	7	86	7
Sum / average	**3**	**80**	**1**	**4**	**95**	**1**	**5**	**73**	**6**	**6**	**87**	**7**

4.8.2 Pilbara region

A total of 12 catchments were used from the Pilbara region (details can be found in the Appendix). The annual maximum flood series record lengths of these 12 stations range from 20 to 34 years (mean 28 years, median 28 years and standard deviation 4.5 years). The catchment areas of these 12 stations range from 0.1 to 1000 km^2 (mean 347 km^2, median 205 km^2 and standard deviation of 366 km^2).

A fixed region approach was considered due to the small number of stations in the region, i.e. all the 12 catchments are considered to have formed one region. One catchment at a time was left out for cross-validation; the procedure was repeated for 12 times so that each of the 12 catchments is tested independently. No region-of-influence (ROI) approach was used for the Pilbara region.

Table 4.8.11 shows different combinations of predictor variables for the Q_{10} QRT model and the models of the first three moments of the LP3 distribution. Figures 4.8.5 and 4.8.6 show example plots of the statistics used in selecting the best set of predictor variables for the Q_{10} flood quantile and the skew. The model error variance shows that combinations 11, 3, 4 and 25 are the top 4 potential sets of predictor variables for the Q_{10} model. Combination 25 contained 3 predictor variables, two of them being rainfall intensity. Combinations 11 (area and evap), 3 (area and rainfall intensity ${}^2I_1$) and 4 (area and ${}^2I_{12}$) were compared to combination 10 (area and $I_{tc,10}$). Combinations 11, 4 and 3 clearly had smaller model error variances than combination 10; however the standard error of the model error variances did not differ greatly. Also there was no significant difference in the R^2_{GLS} values over the three combinations. The AVPO, AVPN, AIC and BIC were the lowest with combination 11. In any case combination 4 with variables area and ${}^2I_{12}$ was adopted as it contained the design rainfall intensity which is the most accurately estimated of all the basic design durations.

For the skew model, combination 11 with two explanatory variables (area and sden) showed the lowest model error variance (0.080) and an R^2_{GLS} value of 57%, while showing the lowest AIC and BIC. (see Figure 4.8.6). The next best combination was 12 with a slightly higher model error variance, with a R^2_{GLS} value of 50%. Combination 1 with no explanatory variables had a model error variance of 0.085 which is comparable to combination 11. The AVPO and AVPN values were also smaller with combination 1 (see Figure 4.8.6). While there is enough evidence to stay with combination 1 (it may be argued that a regional average skew is applicable), the increase in R^2_{GLS} value for combination 11 was significant. In this case

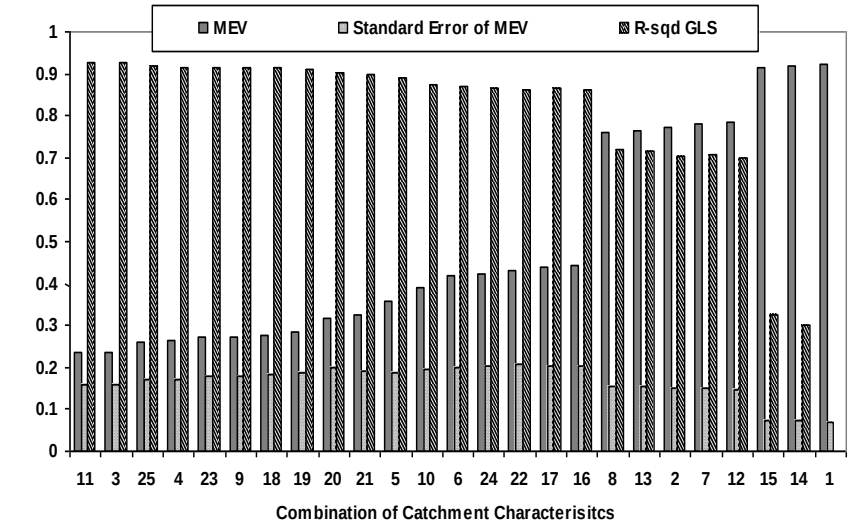
combination 11 was adopted. A similar outcome was obtained for the standard deviation model.

A similar procedure, as discussed above, was adopted in selecting the best set of predictor values for other models with the QRT and PRT. The set of predictor variables selected as above were used in the one-at-a-time cross validation approach.

To assess the statistical significance of the predictor variables, BPV values for the regression coefficients were calculated. For the QRT (for all the ARIs), the BPV values were 0% for area and rainfall intensity ($^2I_{12}$). Hence, the inclusion of these predictor variables is justified. The BPVs for the skew model were 30% and 8% for area and sden, respectively, indicating that area and sden may not be good predictors for skew in this case. The BPVs for the mean flood model were 0% for both the predictor variables (area and $^2I_{12}$). For the standard deviation model, the BPVs for the predictor variables area and forest were 10%, indicating the potential of adopting a constant standard deviation, similar to other states. Regression equations developed for the QRT and PRT for the fixed region are given by Equations 4.8.10 to 4.8.18.

Table 4.8.11 Different combinations of predictor variables considered for the QRT models and the parameters of the LP3 distribution (QRT and PRT fixed region: Pilbara region, WA)

Combination	Combinations for mean, standard deviation & skew models	Combinations for flood quantile model
1	Const	Const
2	Const, area	Const, area
3	Const, area, ($^{2}I_{1}$)	Const, area, $^{2}I_{1}$
4	Const, area, ($^{50}I_{1}$)	Const, area, $^{2}I_{12}$
5	Const, area, ($^{2}I_{12}$)	Const, area, $^{50}I_{1}$
6	Const, area, ($^{50}I_{12}$)	Const, area, $^{50}I_{12}$
7	Const, area, rain	Const, area, rain
8	Const, area, for	Const, area, for
9	Const, area, evap	Const, area, forest, evap
10	Const, area, S1085	Const, area, $I_{tc,ARI}$
11	Const, area, sden	Const, area, evap
12	Const, sden, rain	Const, area, S1085
13	Const, for, rain	Const, area, sden
14	Const, S1085, for	Const, sden, rain
15	Const, evap	Const, for, rain
16	Const, rain, evap	Const, area, $^{50}I_{12}$, rain
17	-	Const, area, $^{50}I_{12}$, sden
18	-	Const, area, $^{50}I_{12}$, rain, evap
19	-	Const, area, $^{50}I_{12}$, $I_{tc,ARI}$, evap
20	-	Const, area, $^{50}I_{12}$, $I_{tc,ARI}$, rain, evap
21	-	Const, area, $^{50}I_{12}$, $I_{tc,ARI}$, sden
22	-	Const, area, $^{50}I_{12}$, $I_{tc,ARI}$, S1085
23	-	Const, area, $I_{tc,ARI}$, evap
24	-	Const, area, $I_{tc,ARI}$, rain
25	-	Const, area, $^{2}I_{1}$, $I_{tc,ARI}$



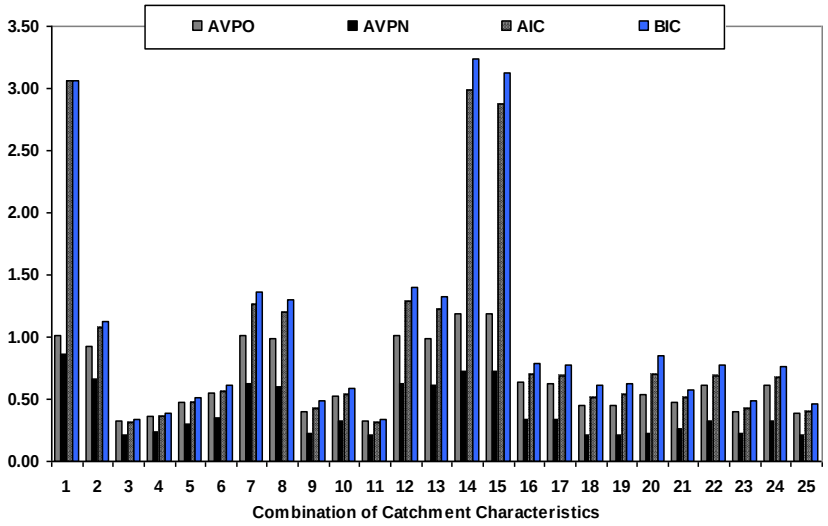


Figure 4.8.5 Selection of predictor variables for the BGLS regression model for Q_{10} of Pilbara Region, WA (QRT, fixed region), MEV = model error variance, AVPO = average variance of prediction (old), AVPN = average variance of prediction (new) AIC = Akaike information criterion, BIC = Bayesian information criterion

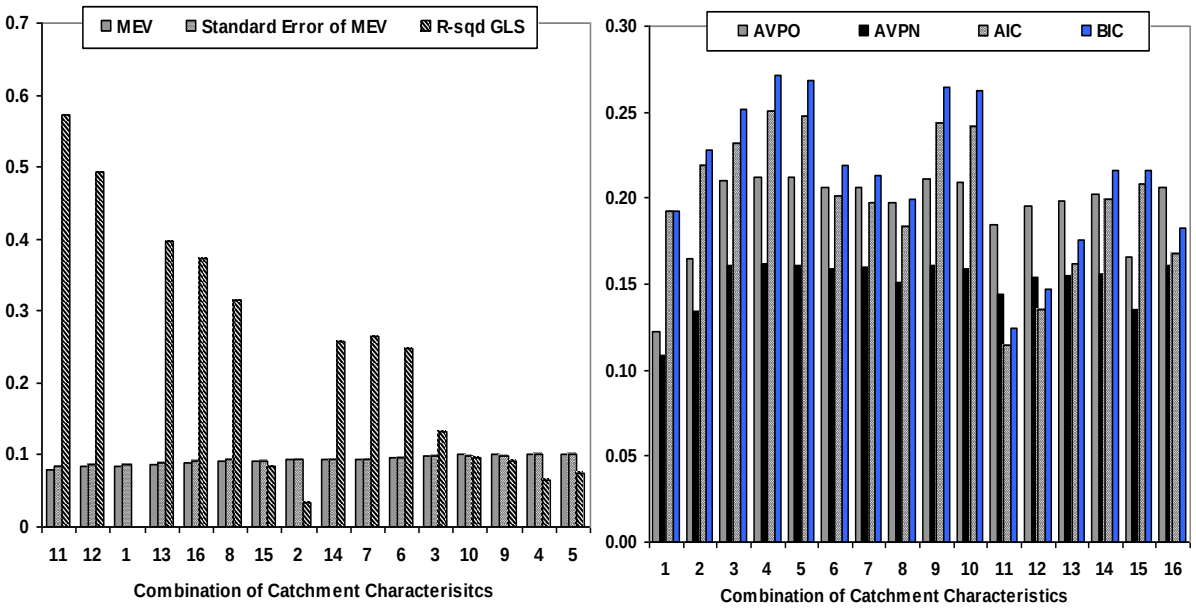


Figure 4.8.6 Selection of predictor variables for the BGLS regression model for skew

Prediction equations for QRT:

$\ln(Q_2) = 2.66 + 0.51[\ln(area) - 4.71] + 8.08[\ln(I_{12,2}) - 1.47]$ (4.8.10)

$\ln(Q_5) = 3.90 + 0.48[\ln(area) - 4.71] + 7.20[\ln(I_{12,2}) - 1.47]$ (4.8.11)

$\ln(Q_{10}) = 4.51 + 0.45[\ln(area) - 4.71] + 6.74[\ln(I_{12,2}) - 1.47]$ (4.8.12)

$\ln(Q_{20}) = 5.01 + 0.44[\ln(area) - 4.71] + 6.19[\ln(I_{12,2}) - 1.47]$ (4.8.13)

$\ln(Q_{50}) = 5.59 + 0.41[\ln(area) - 4.71] + 5.66[\ln(I_{12,2}) - 1.47]$ (4.8.14)

$\ln(Q_{100}) = 5.87 + 0.39[\ln(area) - 4.71] + 5.34[\ln(I_{12,2}) - 1.47]$ (4.8.15)

Prediction equations for PRT:

$M = 2.54 + 0.52[\ln(area) - 4.71] + 8.08[\ln(I_{12,2}) - 1.47]$ (4.8.16)

$stdev = 1.45 + 0.10(\boldsymbol{z}area) + 0.07(\boldsymbol{z}forest)$ (4.8.17)

$skew = - 0.49 – 0.08(\boldsymbol{z}area) – 0.64(\boldsymbol{z}sden)$ (4.8.18)

where *z ()* is explained by Equation 4.3.10.

It is reassuring to observe that the regression coefficients in the QRT set of equations vary in a regular fashion with increasing ARI.

The Pseudo Analysis of Variance (ANOVA) tables for Q_2, Q_{20} and Q_{100} models and the parameters of the LP3 distribution are presented in Tables 4.8.12 and 4.8.17.

For the LP3 parameters, the sampling error increases as the order of moment increases i.e. the EVR (sampling error to model error ratio) increases with the order of the moment. The model error dominates the regional analysis for the mean flood and the standard deviation models; this is more pronounced for the mean flood model (0.47 compared 0.68). However these results also indicate that the sampling error has had a slight effect on the analysis as the EVR values are larger than 0.2. The EVR value for the skew model is 4.2 (Table 4.8.17) which is much higher than that of the mean flood and standard deviation models. The GLS regression modeling framework in this case has performed quite well in the estimation of the parameters of the LP3 distribution. It should be remembered however that there is only 12 stations in this region, thus the sampling error might be over estimated in this case.

The pseudo ANOVA tables were also prepared for the flood quantile models. Tables 4.8.12 and 4.8.14 show the results for Q_2, Q_{20} and Q_{100} models, respectively. What is clear is that the model errors are relatively high as all the EVRs are below 0.5. The Q_2 model has shown the highest heterogeneity in this case (i.e. EVR 0.39). The Q_{100} shows the highest EVR as expected.

Table 4.8.12 Pseudo ANOVA table for Q_2 model of Pilbara Region, WA (QRT)

Source	**Degrees of Freedom**	**Sum of Squares**
Model	*k=3*	$n(\sigma^2_{\delta 0} - \sigma^2_{\delta}) = 8.9$
Model error δ	*n-k-1=8*	$n(\sigma^2_{\delta}) = 2.4$
Sampling error η	*N = 12*	$tr[\Sigma(\hat{y})] = 0.9$
Total	*2n-1 = 23*	**Sum of the above = 12.2**
	EVR	**0.39**

Table 4.8.13 Pseudo ANOVA table for Q_{20} model of Pilbara Region, WA (QRT)

Source	**Degrees of Freedom**	**Sum of Squares**
Model	*k=3*	7.2
Model error δ	*n-k-1=8*	3.5
Sampling error η	*N = 12*	1.6
Total	*2n-1 = 23*	**12.3**
	EVR	**0.47**

Table 4.8.14 Pseudo ANOVA table for Q_{100} model of Pilbara Region, WA (QRT)

Source	Degrees of Freedom	Sum of Squares
Model	*k=3*	4.9
Model error δ	*n-k-1=8*	5.1
Sampling error η	*N = 12*	2.5
Total	*2n-1 = 23*	**12.5**
	EVR	**0.49**

Table 4.8.15 Pseudo ANOVA table for mean flood model of Pilbara Region, WA (PRT)

Source	Degrees of Freedom	Sum of Squares
Model	*k=3*	9.2
Model error δ	*n-k-1=8*	2.2
Sampling error η	*N = 12*	1
Total	*2n-1 = 23*	**12.4**
	EVR	**0.47**

Table 4.8.16 Pseudo ANOVA table for standard deviation model: Pilbara Region, WA (PRT)

Source	Degrees of Freedom	Sum of Squares
Model	*k=3*	0.4
Model error δ	*n-k-1=8*	1.6
Sampling error η	*N = 12*	1.1
Total	*2n-1 = 23*	**3.1**
	EVR	**0.68**

Table 4.8.17 Pseudo ANOVA table for skew model of Pilbara Region, WA (PRT)

Source	Degrees of Freedom	Sum of Squares
Model	*k=3*	0.06
Model error δ	*n-k-1=8*	0.96
Sampling error η	*N = 12*	4.1
Total	*2n-1 = 23*	**5.1**
	EVR	**4.2**

The underlying model assumptions are examined (i.e. the normality of residuals), using the plots of the standardised residuals vs. predicted values (see below). The predicted values were obtained from one-at-a-time cross validation. Figures 4.8.7 and 4.8.8 show the plots for Q_{20} and Q_{100}. It can be seen that most of the standardised residuals fall between the magnitudes of ± 2 for Q_{20} and Q_{100} respectively, hence the underlying model assumptions are satisfied for both QRT and PRT. Also no specific pattern can be identified, with the standardised values being almost equally distributed below and above zero. Similar results were obtained for the mean, standard deviation, skew and other flood quantile models.

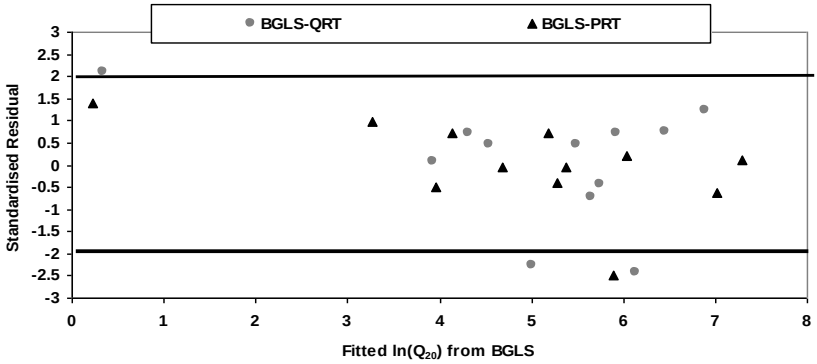


Figure 4.8.8 plots of standardised residuals vs. predicted values for ARI of 20 years (QRT and PRT, Pilbara Region, WA)

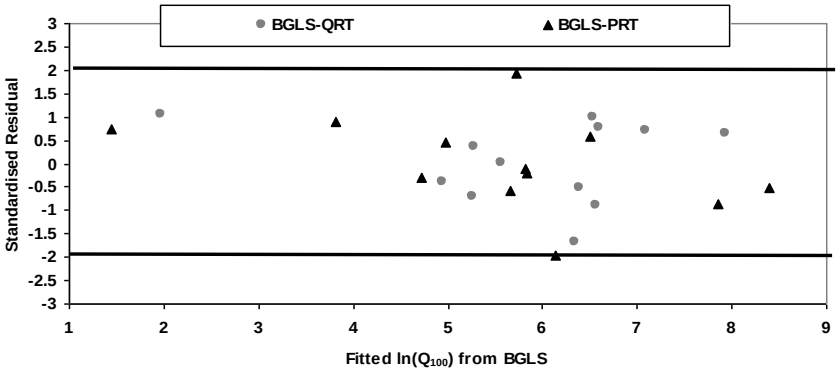


Figure 4.8.7 plots of standardised residuals vs. predicted values for ARI of 100 years (QRT and PRT, Pilbara Region, WA)

The summary of various regression diagnostics (the relevant equations are described in Section 4.2.2) is provided in Table 4.8.18. This shows that for the mean flood model the model error variance (MEV) and average standard error of prediction (SEP) are relatively higher than those of the standard deviation and skew models. For the standard deviation and

skew models, the SEP was 43% and 45% respectively again indicating that both the standard deviation and skew may be regionalised more accurately than the mean flood.

One can see from Table 4.8.18 that the SEP values for all the flood quanitle models ranged from 59% to 96%, which are quite large compared to the east coast of Australia. Again this indicates that the Pilbara region estimates are subject to greater uncertainty, due to the possible higher heterogeneity in the region. While the sampling error has had a slight impact on the analysis (see ANOVA results), the very large model errors have still dominated the regional analysis; this has been reflected in the very high AVPs and SEPs (see Table 4.8.18). The lowest SEP was found for the ARI of 2 years which subsequently has the highest R^2_{GLS} value (94%). Overall the R^2_{GLS} values obtained (80% to 94%) seem to be reasonable, however it must be kept in mind that the analysis has been undertaken on only 12 stations, thus little confidence can be placed on the above statistics (Table 4.8.18).

Table 4.8.18 Regression diagnostics for fixed region analysis: Pilbara Region, WA (12 stations)

Model	Fixed region			
	MEV	AVP	SEP (%)	R^2_{GLS} (%)
Mean	0.18	0.25	54	95
Stdev	0.13	0.17	43	31
Skew	0.080	0.18	45	57
Q_2	0.22	0.30	59	94
Q_5	0.25	0.34	64	92
Q_{10}	0.27	0.38	68	91
Q_{20}	0.32	0.45	75	88
Q_{50}	0.40	0.56	87	84
Q_{100}	0.47	0.66	96	80

Table 4.8.19 presents the root mean square error (RMSE) (Equation 4.2.16) and relative error (RE) (Equation 4.2.15) values for the PRT and QRT models. In terms of RMSE, PRT over all the ARIs provides relatively smaller values as compared to QRT. The best result was found for the 2-year ARI for QRT and PRT. In terms of RE, PRT gives 4 to 15% smaller values than the QRT. Overall it can be seen that PRT performs slightly better over all the ARIs.

Table 4.8.20 shows the results of counting the Q_{pred}/Q_{obs} ratios for the QRT and PRT. The use of this ratio has been discussed in Section 4.2.4. The desirable cases on average for the PRT and QRT are the same i.e. 86%. The PRT shows slightly more gross underestimation on average than the QRT (i.e. 6% and 3%, PRT and QRT respectively). The QRT shows more cases of overestimation on average as compared to PRT (11% and 7% for QRT and PRT respectively); however, these differences are relatively modest.

Table 4.8.19 Evaluation statistics (RMSE and RE) from one-at-a-time cross validation for Pilbara Region, WA

Model	RMSE (%)		RE (%)	
	PRT Fixed region	QRT Fixed region	PRT Fixed region	QRT Fixed region
Q_2	4.7	5.5	35	39
Q_5	5.6	5.7	25	40
Q_{10}	6.4	7.3	21	33
Q_{20}	6.6	7.4	24	34
Q_{50}	6.3	6.5	31	39
Q_{100}	5.9	7.1	36	41

Table 4.8.20 Summary of counts based on Q_{pred}/Q_{obs} ratio values for QRT and PRT for Pilbara Region, WA (fixed region). “U” = gross underestimation, “D” = desirable range and “O” = gross overestimation

Model	Count (QRT)			Percent (QRT)			Count (PRT)			Percent (PRT)		
	U	D	O	U	D	O	U	D	O	U	D	O
Q_2	0	10	2	0	83	17	0	12	0	0	100	0
Q_5	1	10	1	8	83	8	0	11	1	0	92	8
Q_{10}	1	10	1	8	83	8	1	10	1	8	83	8
Q_{20}	0	10	2	0	83	17	1	10	1	8	83	8
Q_{50}	0	11	1	0	92	8	1	10	1	8	83	8
Q_{100}	0	11	1	0	92	8	1	9	2	8	75	8
Sum / average	**2**	**62**	**8**	**3**	**86**	**11**	**4**	**62**	**5**	**6**	**86**	**7**

4.8.3 South-west region of WA (Drainage Division VI)

A total of 120 catchments were used from the south-west region of WA; these stations fall in Drainage Division VI (details can be found in the Appendix). This region has the best quality streamflow data, as well as a higher density of stations than the two other WA regions presented before. The annual maximum flood series record lengths of these 120 stations range from 20 to 56 years (mean 31 years, median 29 years and standard deviation 8 years). The catchment areas of these 120 stations range from 0.2 to 983 km^2 (mean 156 km^2, median 48 km^2 and standard deviation of 235 km^2).

In the fixed region approach, all the 120 catchments were considered to have formed one region, however, one catchment was left out for cross-validation and the procedure was repeated 120 times to implement one-at-a-time cross validation. In the region-of-influence (ROI) approach, an optimum region was formed for each of the 120 catchments by drawing an appropriate number of neighbouring stations based on the minimum model error variance.

Table 4.8.21 shows different combinations of predictor variables for the Q_{10} QRT model and the models for the first three moments of the LP3 distribution. Figure 4.8.8 and 4.8.9 show example plots of the statistics used in selecting the best set of predictor variables for the Q_{10} and skew models. All the model error variances for the different combinations are quite high, indicating the high heterogeneity of this region. For the Q_{10} model, combinations 20 and 18 have the lowest model error variances. However combinations 20 and 18 contained 4 to 5 predictor variables which in practice would not be practical. What is clear is all the combinations had very similar R^2_{GLS} values. Combination 10 with 2 predictor variables (area and $I_{tc,ARI}$) showed a smaller standard error of the model error variance as compared to the other combinations, indicating better accuracy with the model error variance estimate. The regression coefficients for the combination 10 variables were 4 and 5 times the posterior standard deviation away from zero, which was the best among all the combinations. The AVPO, AVPN, AIC and BIC values were not that different between combinations 20, 18 and 10, hence combination 10 was finally selected as the best set of predictor variables for the Q_{10} mode, which includes area and design rainfall intensity $I_{tc,10.}$

For the skew model, combination 16 showed the lowest model error variance and the highest R^2_{GLS} (see Figure 4.8.9) as well as the lowest AIC and BIC. Combination 1 without any explanatory variables showed higher AVPO and AVPN as compared to combination 16.

A similar procedure was adopted in selecting the best set of predictor values for other models with the QRT and PRT. The sets of predictor variables selected as above were used in the one-at-a-time cross validation (with fixed regions) and the region-of-influence (ROI) approach.

The significance of the estimated regression coefficient values shown in Equations 4.8.19 to 4.8.27 was evaluated using the Bayesian plausibility value (BPV). The BPVs for the regression coefficients associated with variables area and design rainfall intensity $I_{tc,ARI}$ for the QRT over all the ARIs ranged from 0% to 8%. The BPVs for the skew model were 2% and 1% for rain and evaporation, respectively, indicating that these are reasonably good predictors for skew in this application. The BPVs for the mean flood model were 0% and 1% for both the predictor variables respectively (area and $^2I_{12}$). For the standard deviation model, the BPVs for the predictor variables area and $^2I_{12}$ were greater than 10%, indicating that these variables are not good predictors of the standard deviation of annual maximum floods in this application.

Regression equations developed for the QRT and PRT for the fixed region are given by Equations 4.8.19 to 4.8.27.

Table 4.8.21 Different combinations of predictor variables considered for the QRT models and the parameters of the LP3 distribution of the South-west Region of WA (QRT and PRT fixed region)

Combination	Combinations for mean, standard deviation & skew models	Combinations for flood quantile model
1	Const	Const
2	Const, area	Const, area
3	Const, area, $^{2}I_{1}$	Const, area, $^{2}I_{1}$
4	Const, area, $^{50}I_{1}$	Const, area, $^{2}I_{12}$
5	Const, area, $^{50}I_{12}$	Const, area, $^{50}I_{1}$
6	Const, area, $^{2}I_{12}$	Const, area, $^{50}I_{12}$
7	Const, area, rain	Const, area, rain
8	Const, area, forest	Const, area, for
9	Const, area, evap	Const, area, forest, evap
10	Const, area, sden	Const, area, $I_{tc,ARI}$
11	Const, area, S1085	Const, area, evap
12	Const, rain, S1085	Const, area, sden
13	Const, forest, rain	Const, area, S1085
14	Const, sden, forest	Const, sden, rain
15	Const, evap	Const, for, rain
16	Const, rain, evap	Const, area, $^{50}I_{12}$, rain
17	-	Const, area, $^{50}I_{12}$, S1085
18	-	Const, area, $^{50}I_{12}$, rain, evap
19	-	Const, area, $^{50}I_{12}$, $I_{tc,ARI}$, evap
20	-	Const, area, $^{50}I_{12}$, $I_{tc,ARI}$, rain, evap
21	-	Const, area, $^{50}I_{12}$, $I_{tc,ARI}$, S1085
22	-	Const, area, $^{50}I_{12}$, $I_{tc,ARI}$, sden
23	-	Const, area, $I_{tc,ARI}$, evap
24	-	Const, area, $I_{tc,ARI}$, rain
25	-	Const, area, $^{2}I_{1}$, $I_{tc,ARI}$

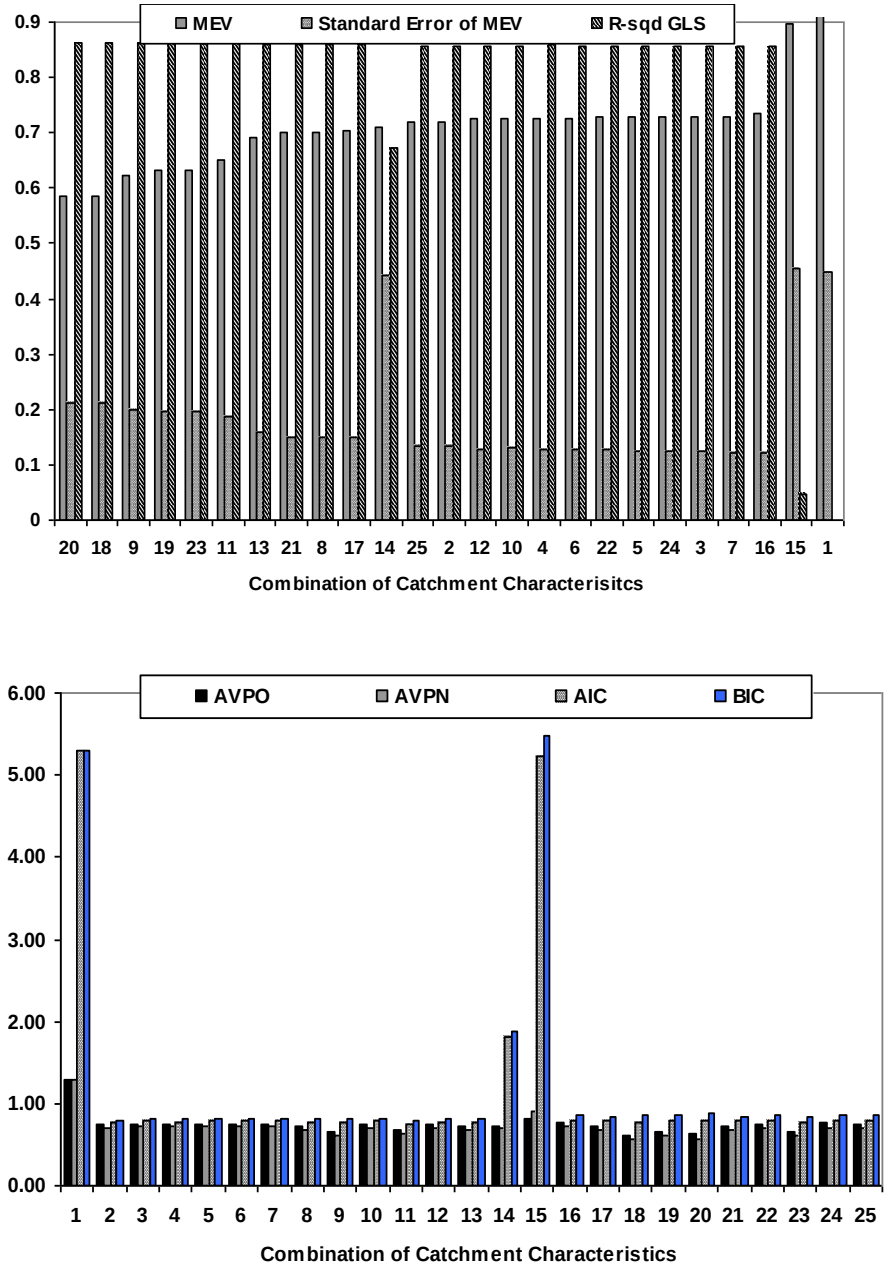


Figure 4.8.8 Selection of predictor variables for the BGLS regression model for Q_{10} of the South-west Region of WA (QRT, fixed region), MEV = model error variance, AVPO = average variance of prediction (old), AVPN = average variance of prediction (new) AIC = Akaike information criterion, BIC = Bayesian information criterion

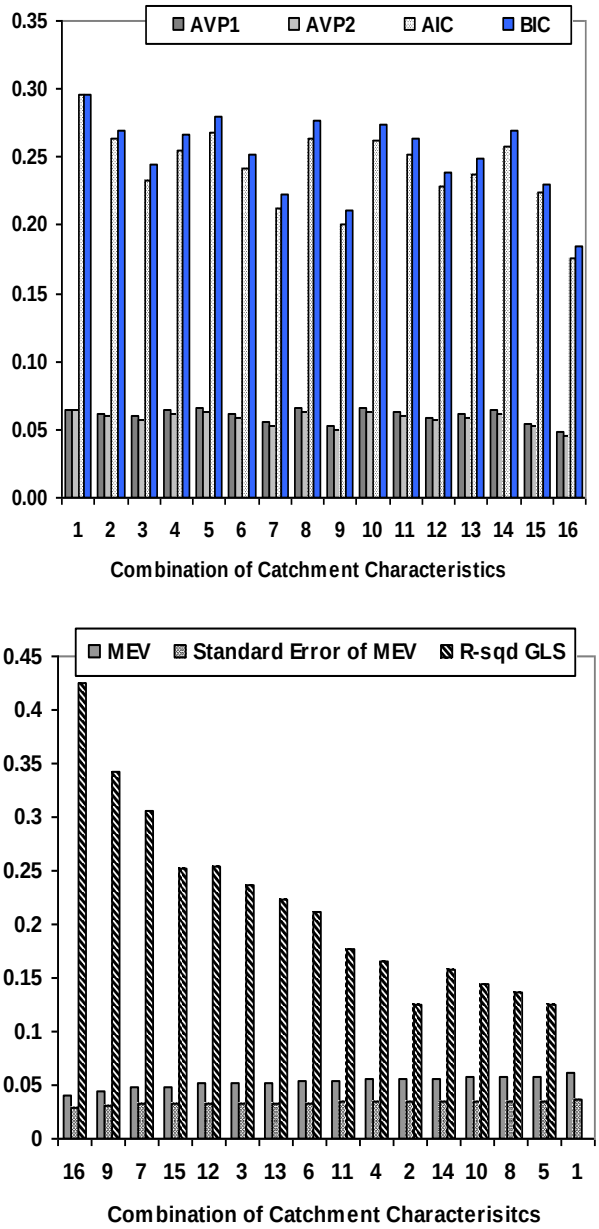


Figure 4.8.9 Selection of predictor variables for the BGLS regression model for skew

Regression equations for the QRT method:

$$\ln(Q_2) = 0.33 + 1.01(\boldsymbol{z}area) + 0.90(\boldsymbol{z}I_{tc,2}) \quad (4.8.19)$$

$$\ln(Q_5) = 1.07 + 1.02(\boldsymbol{z}area) + 0.79(\boldsymbol{z}I_{tc,5}) \quad (4.8.20)$$

$$\ln(Q_{10}) = 1.42 + 0.86(\boldsymbol{z}area) + 0.23(\boldsymbol{z}I_{tc,10}) \quad (4.8.21)$$

$$\ln(Q_{20}) = 1.83 + 0.54(\boldsymbol{z}area) - 0.96(\boldsymbol{z}I_{tc,20}) \quad (4.8.22)$$

$\ln(Q_{50}) = 2.23 + 0.48(\boldsymbol{z}area) - 1.28(\boldsymbol{z}I_{tc,50})$ (4.8.23)

$\ln(Q_{100}) = 2.47 + 0.48(\boldsymbol{z}area) - 1.35(\boldsymbol{z}I_{tc,100})$ (4.8.24)

Regression equations for the PRT method:

$M = 0.32 + 0.82(\boldsymbol{z}area) + 1.19(\boldsymbol{z}I_{12,2})$ (4.8.25)

$stdev = 0.80 - 0.01(\boldsymbol{z}area) - 1.17(\boldsymbol{z}I_{12,2})$ (4.8.26)

$skew = -0.08 - 0.37(\boldsymbol{z}rain) + 1.70(\boldsymbol{z}evap)$ (4.8.27)

where z () is explained in Equation 4.3.10.

It is reassuring to observe that the regression coefficients in the QRT set of equations vary in a regular fashion with increasing ARI.

The Pseudo Analysis of Variance (ANOVA) tables for the Q_{20} and Q_{100} models and the parameters of the LP3 distribution are presented in Tables 4.8.22 – 4.8.26 for the fixed regions and ROI. A Pseudo ANOVA presented here is an extension of the ANOVA in the OLS regression which does not recognize and correct for the expected sampling variance (Reis et al., 2005).

For the LP3 parameters, the EVR (i.e. the sampling error) increases with the order of the moments. The ROI shows a reduced model error variance (i.e. a reduced heterogeneity) as compared to the fixed regions, if only just slightly. The model error has clearly dominated the regional analysis for the mean flood and the standard deviation models for both the fixed regions and ROI. This is more pronounced for the mean flood. The ROI shows a slightly higher EVR than the fixed regions, e.g. for the mean flood model the EVR is 0.04 for the ROI and 0.03 for the fixed region (Table 4.8.24). For the standard deviation model the increase in EVR is more pronounced for ROI (0.36) compared to 0.25 for the fixed region (Table 4.8.25). These results indicate that the mean flood has the greater level of heterogeneity associated with it compared to the standard deviation and skew.

The EVR values for the skew model are 7 and 7.2 for the fixed regions and ROI respectively (Tables 4.8.26), which are much higher than the recommended limit of 0.20. This clearly indicates that the GLS regression is the preferred modeling choice for the skew model. An OLS model would have clearly given misleading results, as the sampling error has clearly dominated the regional analysis. As far as the ROI is concerned, there is little change in the EVR compared to the fixed region, as the skew model tends to include more stations in the regional analysis because of the low model error variance and higher sampling error.

Pseudo ANOVA tables were also prepared for the flood quantile models. Tables 4.8.22 and 4.8.23 show the results for the Q_{20} and Q_{100} models, respectively. Here the ROI shows a slight improvement in EVR over the fixed region.

Table 4.8.22 Pseudo ANOVA table for Q_{20} model of the South-west Region of WA (QRT, fixed region and ROI)

Source	**Degrees of Freedom**		**Sum of Squares**		
	Fixed region	ROI	*Equations*	Fixed region	ROI
Model	*k=3*	*k=3*	$n(\sigma^2_{\delta 0} - \sigma^2_{\delta})$=	23	23
Model error δ	*n-k-1=116*	*n-k-1=61*	$n(\sigma^2_{\delta})$=	25	67
Sampling error η	*N = 120*	*N = 65*	$tr[\Sigma(\hat{y})]$ =	6.7	6
Total	*2n-1 = 239*	*2n-1 = 130*	**Sum of the above =**	**114**	**95**
			EVR	**0.08**	**0.10**

Table 4.8.23 Pseudo ANOVA table for Q_{100} model of the South-west Region of WA (QRT, fixed region and ROI)

Source	**Degrees of Freedom**		**Sum of Squares**		
	Fixed region	ROI		Fixed region	ROI
Model	*k=3*	*k=3*		14.6	14.6
Model error δ	*n-k-1=116*	*n-k-1=67*		97	72
Sampling error η	*N = 120*	*N = 71*		10.5	9.3
Total	*2n-1 = 239*	*2n-1 = 142*	**Sum of the above =**	**121**	**96**
			EVR	**0.11**	**0.13**

Table 4.8.24 Pseudo ANOVA table for mean flood model of the South-west Region of WA (PRT, fixed region and ROI)

Source	**Degrees of Freedom**		**Sum of Squares**		
	Fixed region	ROI		Fixed region	ROI
Model	*k=3*	*k=3*		43	78
Model error δ	*n-k-1=116*	*n-k-1=38*		105	71
Sampling error η	*N = 120*	*N = 42*		3.5	3
Total	*2n-1 = 239*	*2n-1 = 84*	**Sum of the above =**	**151**	**151**
			EVR	**0.03**	**0.04**

Table 4.8.25 Pseudo ANOVA table for the standard deviation model of the South-west Region of WA (PRT, fixed region and ROI)

Source	**Degrees of Freedom**		**Sum of Squares**		
	Fixed region	ROI		Fixed region	ROI
Model	*k=3*	*k=3*		0.13	2.9
Model error δ	*n-k-1=116*	*n-k-1=73*		8.0	5.3
Sampling error η	*N = 120*	*N = 77*		2	1.9
Total	*2n-1 = 239*	*2n-1 = 154*	**Sum of the above =**	**10**	**10**
			EVR	**0.25**	**0.36**

Table 4.8.26 Pseudo ANOVA table for the skew model of the South-west Region of WA (PRT, fixed region and ROI)

Source	**Degrees of Freedom**		**Sum of Squares**		
	Fixed region	ROI		Fixed region	ROI
Model	*k=3*	*k=3*		2.6	3.3
Model error δ	*n-k-1=116*	*n-k-1=93*		4.8	4.2
Sampling error η	*N = 120*	*N = 97*		34	30
Total	*2n-1 = 239*	*2n-1 = 184*	**Sum of the above =**	**41**	**38**
			EVR	**7**	**7.2**

To assess the underlying model assumptions (i.e. the normality of residuals), the plots of standardised residuals vs. predicted values were examined. The predicted values were obtained from one-at-a-time cross validation. Figures 4.8.10 and 4.8.11 show the plots for the Q_{20} flood models with the fixed region and ROI. If the underlying model assumption is satisfied to a large extent, the standardised residual values should not be of greater magnitude than ± 2; in practice, 95% of the standardised residuals should fall between ± 2. The results in Figures 4.8.10 and 4.8.11 reveal that the developed equations satisfy the normality of residual assumption quite satisfactorily. Also no specific pattern (heteroscedasicity) can be identified, with the standardised values being almost equally distributed below and above zero. Similar results were obtained for the mean flood, skew, standard deviation and other flood quantile models.

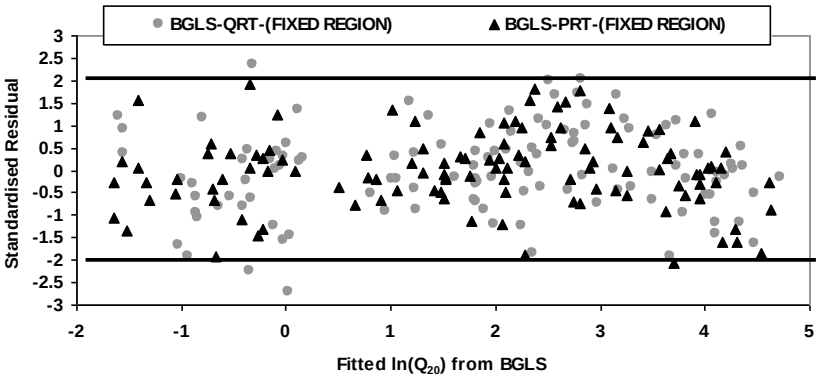


Figure 4.8.10 plots of standardised residuals vs. predicted values for ARI of 20 years (QRT and PRT, fixed region, WA– south-west region)

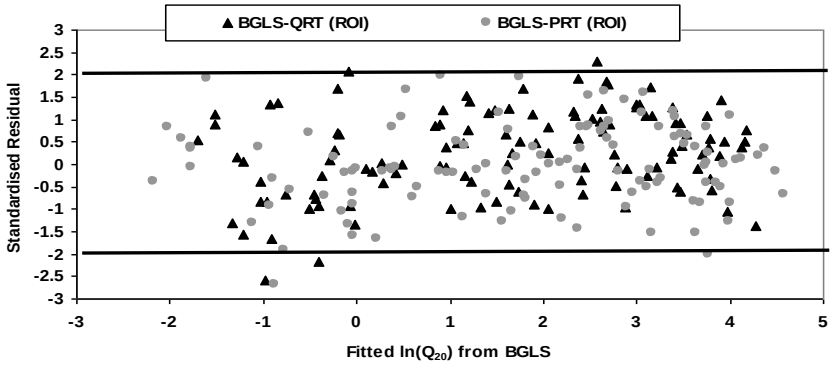


Figure 4.8.11 plots of standardised residuals vs. predicted values for ARI of 20 years (QRT and PRT, ROI, WA– south-west region)

The QQ-plots of the standardised residuals (Equation 4.2.13) vs. normal score (Equation 4.2.14) for the fixed region (based on one-at-a-time cross validation) and ROI were examined. Figures 4.8.12 to 4.8.14 present results for the Q_{20} and skew models, which show that all the points closely follow a straight line. This indicates that the assumption of normality and the homogeneity of variance of the standardised residuals have largely been satisfied. If the standardised residuals are indeed normally and independently distributed N(0,1) with mean 0 and variance 1, then the slope of the best fit line in the QQ-plot, which can be interpreted as the standard deviation of the normal score (Z score) of the quantile, should approach 1 and the intercept, which is the mean of the normal score of the quantile should approach 0 as the number of sites increases. Figures 4.8.12 to 4.8.14 indeed show that the fitted lines for the developed models pass through the origin (0, 0) and have a slope approximately equal to one. The ROI approach approximates the normality of the residuals

slightly better (i.e. a better match with the fitted line) than the fixed region approach. Similar results were also found for the mean, standard deviation and other flood quantile models.

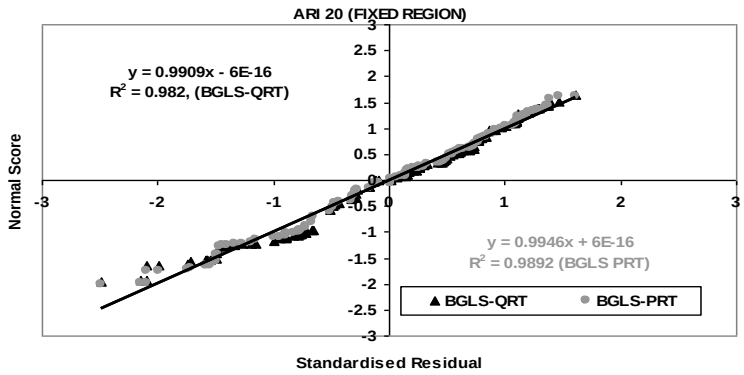


Figure 4.8.12 QQ plot of the standardised residuals vs. Z score for ARI of 20 years (QRT and PRT, fixed region, WA– south-west region)

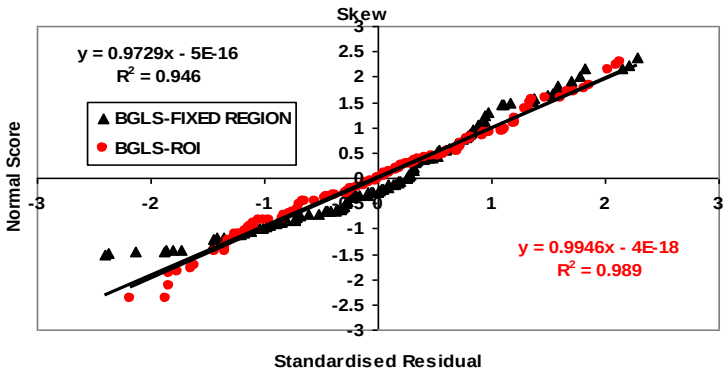


Figure 4.8.13 QQ plot of the standardised residuals vs. Z score for the skew model (PRT, fixed region, ROI, WA – south-west region)

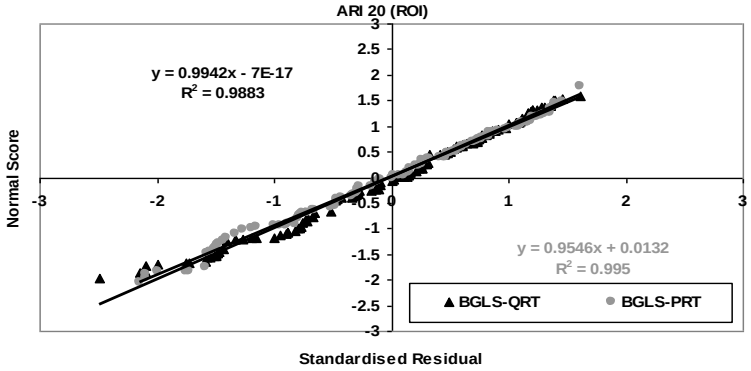


Figure 4.8.14 QQ plot of the standardised residuals vs. Z score for ARI of 20 years (QRT and PRT, ROI, WA– south-west region)

The summary of various regression diagnostics (the relevant equations are described in Section 4.2.2) is provided in Table 4.8.27. This shows that for the mean flood model the model error variance (MEV) and average standard error of prediction (SEP) are much higher than those of the standard deviation and skew models, as expected. For the mean flood model, the ROI shows smaller model error variance than the fixed region. Also, the R^2_{GLS} value for the mean flood model with the ROI is 2% higher than the fixed region. These indicate that the ROI should be preferred over the fixed region for developing the mean flood model. For the standard deviation model, ROI shows no difference in SEP, while a 2% increase in R^2_{GLS} is gained. The SEP and R^2_{GLS} values for the skew model are the same for the fixed region and ROI, as the number of sites in the skew model is nearly all the sites in the region (i.e. 97 from 120 sites). There was also no notable difference in the model error and sampling error variances.

Interestingly one can see from Table 4.8.27 that the SEP values for all the flood quanitle models are 16% to 24% smaller for the ROI cases than the fixed region, the lowest SEP% results being for the 10 and 20 years ARIs. The R^2_{GLS} values for ROI cases are much the same as the fixed region ones. These show that the ROI approach performs better than the fixed region approach in terms of reducing the SEP% in estimates (i.e. heterogeneity).

Table 4.8.27 Regression diagnostics for fixed region and ROI for WA– south-west region

Model	Fixed region				ROI			
	MEV	AVP	SEP (%)	R^2_{GLS} (%)	MEV	AVP	SEP (%)	R^2_{GLS} (%)
Mean	0.88	0.91	122	80	0.59	0.66	96	82
Stdev	0.050	0.052	23	65	0.044	0.056	23	67
Skew	0.041	0.048	22	42	0.035	0.045	22	42
Q_2	0.90	0.92	123	80	0.61	0.68	99	81
Q_5	0.80	0.82	113	82	0.56	0.61	92	83
Q_{10}	0.73	0.75	106	82	0.55	0.59	89	82
Q_{20}	0.71	0.74	105	82	0.55	0.59	89	82
Q_{50}	0.76	0.79	110	81	0.58	0.62	93	82
Q_{100}	0.81	0.84	115	80	0.60	0.65	96	81

Table 4.8.28 shows number of sites and associated model error variances for the ROI and fixed region models. This shows that the ROI mean flood model has fewer sites (42 out of 120 i.e. 35%) than the standard deviation and skew models. The ROI skew model has the highest number of sites, which includes nearly all the sites in WA. The model error variance for the fixed region mean flood model is 33% higher than the corresponding ROI model. The model error variances for all the ROI models are smaller than the fixed region models. This shows that the fixed region models experience a greater heterogeneity than the ROI. If the fixed regions are made too large, the model error will be inflated by heterogeneity unaccounted for by the catchment characteristics. Figure 4.8.15 shows the resulting spatial variation in the mean flood for the minimum model error variances from the ROI analysis. The significance of this finding is that if sub-regions do exist, they are most likely to be captured by the ROI. Interestingly enough, these spatial variations compare well to the regions formed for the south west WA region in ARR 1987. This may allow for more efficient design flood estimation based on local information surrounding the ungauged catchment in question.

Table 4.8.28 Model error variances associated with fixed region and ROI for WA – south-west region (n = number of sites of the parameters and flood quantiles

Parameter/ Quantiles	Mean	Stdev	Skew	Q_2	Q_5	Q_{10}	Q_{20}	Q_{50}	Q_{100}
ROI (n)	42	77	97	37	51	62	65	67	71
$\hat{\sigma}_{\delta}^{2}$	0.59	0.044	0.035	0.61	0.56	0.55	0.55	0.58	0.60
Fixed region (n)	120	120	120	120	120	120	120	120	120
$\hat{\sigma}_{\delta}^{2}$	0.88	0.050	0.041	0.90	0.80	0.73	0.71	0.76	0.81

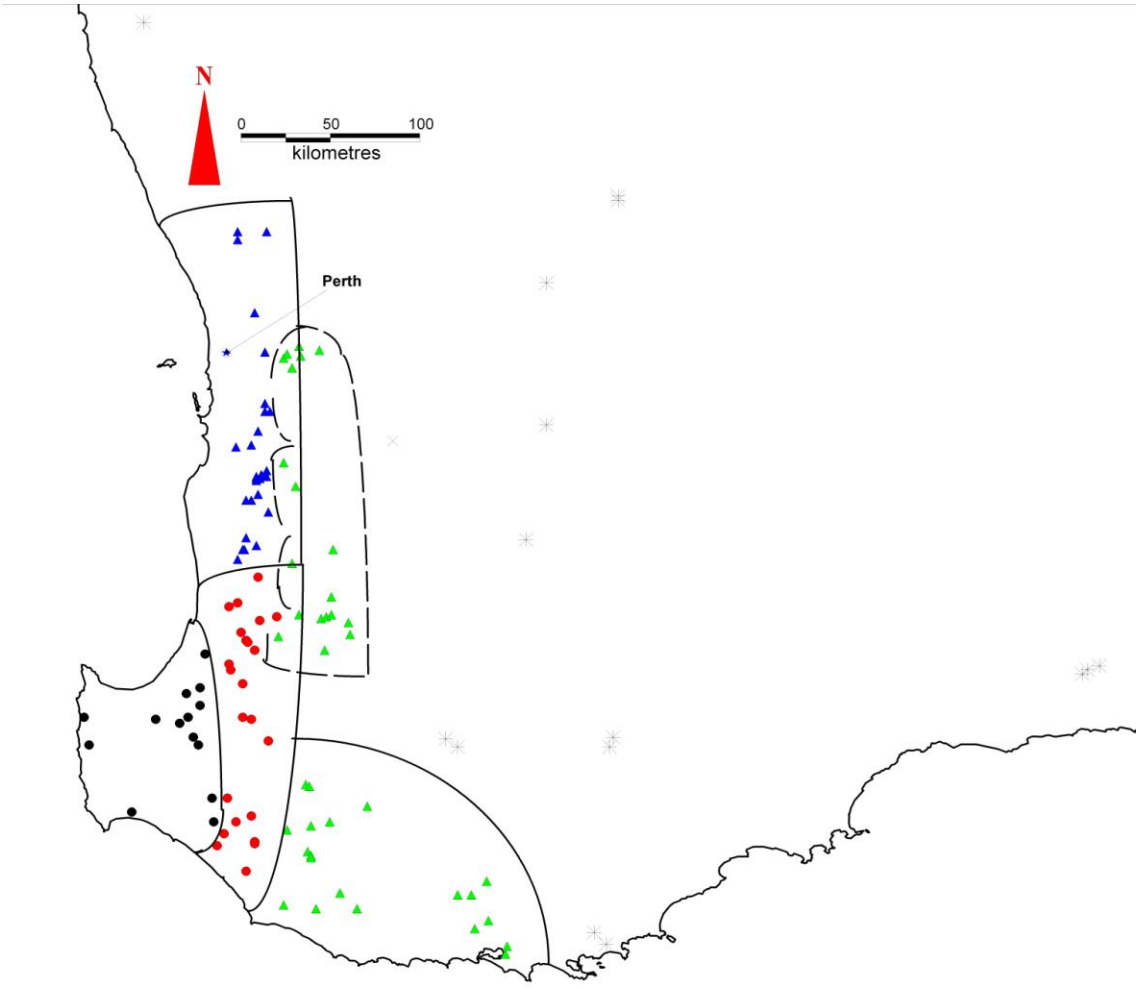


Figure 4.8.15 Binned minimum model error variance for the mean flood model

Table 4.8.29 presents the root mean square error (RMSE) (Equation 4.2.16) and relative error (RE) (Equation 4.2.15) values for the PRT and QRT models with both the fixed region and ROI. In terms of RMSE, ROI always gives smaller values than the fixed regions for all the ARIs. The QRT-ROI shows smaller RMSE values (2% to 9%) than the PRT-ROI for all the ARIs. In terms of RE, ROI gives 0 to 9% smaller values than the fixed regions. The QRT-ROI gives smaller values of RE (by 2% to 3%) for ARIs of 2 and 5 years than the PRT-ROI. For the 10-year ARI, both QRT-ROI and PRT-ROI perform very similarly. However, for ARIs of 20 to 100 years, the PRT-ROI gives smaller RE values (by 2% to 6%) than the QRT-ROI. These results show that there are only modest differences between the performances of the QRT and PRT.

Table 4.8.30 shows the results of counting the Q_{pred}/Q_{obs} ratios for the QRT and PRT ROI. The use of this ratio has been discussed in Section 4.2.4. There are slightly more desirable cases for the PRT on average (PRT 62% and QRT 61%). The QRT method shows gross underestimation with a greater number of cases, on average i.e. 24% for QRT as compared to 20% for PRT. There is also slightly more gross overestimation on average with the PRT (i.e. 19% for PRT and 15% for QRT).

Table 4.8.29 Evaluation statistics (RMSE and RE) from one-at-a-time cross validation for the South-west Region of WA

Model	RMSE (%)				RE (%)			
	PRT		QRT		PRT		QRT	
	Fixed region	ROI	Fixed region	ROI	Fixed region	ROI	Fixed region	ROI
Q_2	30	19	34	15	46	46	46	43
Q_5	28	17	28	8	51	44	43	41
Q_{10}	23	15	23	10	50	41	43	41
Q_{20}	22	15	23	12	52	43	52	49
Q_{50}	19	14	22	12	52	47	54	50
Q_{100}	16	13	19	10	54	51	60	53

Table 4.8.30 Summary of counts based on Q_{pred}/Q_{obs} ratio values for QRT and PRT for the South-west Region of WA (ROI). “U” = gross underestimation, “D” = desirable range and “O” = gross overestimation

Model	Count (QRT)			Percent (QRT)			Count (PRT)			Percent (PRT)		
	U	A	O	U	A	O	U	A	O	U	A	O
Q_2	25	65	30	21	54	25	19	72	29	16	60	24
Q_5	29	79	12	24	66	10	23	74	23	19	62	19
Q_{10}	33	71	16	28	59	13	26	74	20	22	62	17
Q_{20}	28	76	16	23	63	13	25	75	20	21	63	17
Q_{50}	29	75	16	24	63	13	25	75	20	21	63	17
Q_{100}	31	74	15	26	62	13	25	73	22	21	61	18
Sum / average	**175**	**440**	**105**	**24**	**61**	**15**	**143**	**443**	**134**	**20**	**62**	**19**

4.9 Results from South Australia

4.9.1 Preliminary investigation by SA team

The SA team developed preliminary prediction equations for SA using GLS regression. These equations contained two predictor variables, catchment area (AREA) and design rainfall intensity as shown in Table 4.9.1. The R^2 values of these equations range from 0.30 to 0.42, which appear to be quite low as compared to other Australian states.

Table 4.9.1 Preliminary regional prediction equations developed for SA by SA team (Hewa et al.) (Here *area* is catchment area in km^2 and *I* is design rainfall intensity in mm/h for various ARIs and durations)

Model	R^2	Adjusted R^2
$Q_{100} = 0.69(area) + 7.58\ (^{100}I_{24})$	0.42	0.36
$Q_{50} = 0.53(area) + 12.68\ (^{50}I_{48})$	0.42	0.35
$Q_{20} = 0.36(area) + 8.04\ (^{20}I_{24})$	0.43	0.37
$Q_{10} = 0.25(area) + 8.24\ (^{10}I_{24})$	0.40	0.32
$Q_{5} = 0.17(area) + 4.47\ (^{5}I_{12})$	0.38	0.31
$Q_{2} = 0.065(area) + 3.3\ (^{2}I_{12})$	0.30	0.20

4.9.2 Development of RFFA method for SA by UWS team

For the SA stations, an LP3-Bayesian parameter fitting procedure (FLIKE) was applied similar to other Australian states to estimate flood quantiles for ARIs of 2 to 100 years. Out of the 30 catchments, one catchment was located far north and was removed from this data set and was placed in the SA arid-region. These catchments are listed in Table A3 in the appendix. The locations of these catchments are shown in Figure 2.8. The annual maximum flood series record lengths of these 29 stations range from 17 to 67 years (mean 35 years, median 34 years and standard deviation 10 years). The catchment areas of these 29 stations range from 0.6 to 708 km^2 (mean 170 km^2, median 77 km^2 and standard deviation of 202 km^2).

The Bayesian GLS regression was adopted to develop the prediction equations based on the Parameter Regression Technique (PRT). A fixed region approach was used to form the region where all the 29 catchments were considered to have formed one region, however, one catchment was left out for cross-validation and the procedure was repeated 29 times to implement the one-at-a-time cross validation. No region-of-influence (ROI) approach was used for SA as the sample size was too small to apply the ROI approach.

Table 4.9.2 shows different possible combination of predictor variables for the first three moments of the LP3 distribution. Figures 4.9.1 and 4.9.2 show example plots of the statistics used in selecting the best set of predictor variables for the mean flood and the skew models. According to the model error variance, combinations 4 and 6 (with catchment area and design rainfall intensities) were the top 2 potential sets of predictor variables for the mean flood model. Combinations 4 and 6 showed no significant difference in $R^2_{GLS.}$ The AVPO, AVPN, AIC and BIC for combinations 4 and 6 were very comparable. The regression coefficients for both the intensities ($^{50}I_1$ and $^{2}I_{12}$) were approximately 4 times the posterior standard deviation away from zero. Finally combination 6 (area and $^{2}I_{12}$) was selected to be consistent with other Australian states.

For the skew model, combination 1 with no explanatory variables showed one of the lowest model error variances (0.037) and the lowest AVPO and AVPN. (see Figure 4.9.2). The next best combination was 15 with a slightly smaller model error variance; however the R^2_{GLS} was very poor (< 14%). Therefore there was enough evidence to stay with combination 1 as the other models did not show major improvement in model error variance with the additional explanatory variables. In this case given the relatively small differences in the model error variance, it may be argued that a regional average skew is applicable for SA. A similar outcome was obtained for the standard deviation model.

The set of predictor variables selected as above were used in the one-at-a-time cross validation approach. The BPVs for the mean model were 0% for both the predictor variables area and $^{2}I_{12}$. Regression equations developed for the PRT for the fixed region are given by Equations 4.9.1 to 4.9.3.

Table 4.9.2 Different combinations of predictor variables considered for the parameters of the LP3 distribution (PRT fixed region SA)

Combination	Combinations for mean, standard deviation & skew models
1	Const
2	Const, area
3	Const, area, $^{2}I_{1}$
4	Const, area, $^{50}I_{1}$
5	Const, area, $^{50}I_{12}$
6	Const, area, $^{2}I_{12}$
7	Const, area, rain
8	Const, area, SL*
9	Const, area, evap
10	Const, area, sden
11	Const, area, S1085
12	Const, S1085, rain
13	Const, SL, rain
14	Const, sden, SL
15	Const, evap
16	Const, rain, evap

* SL = Stream length

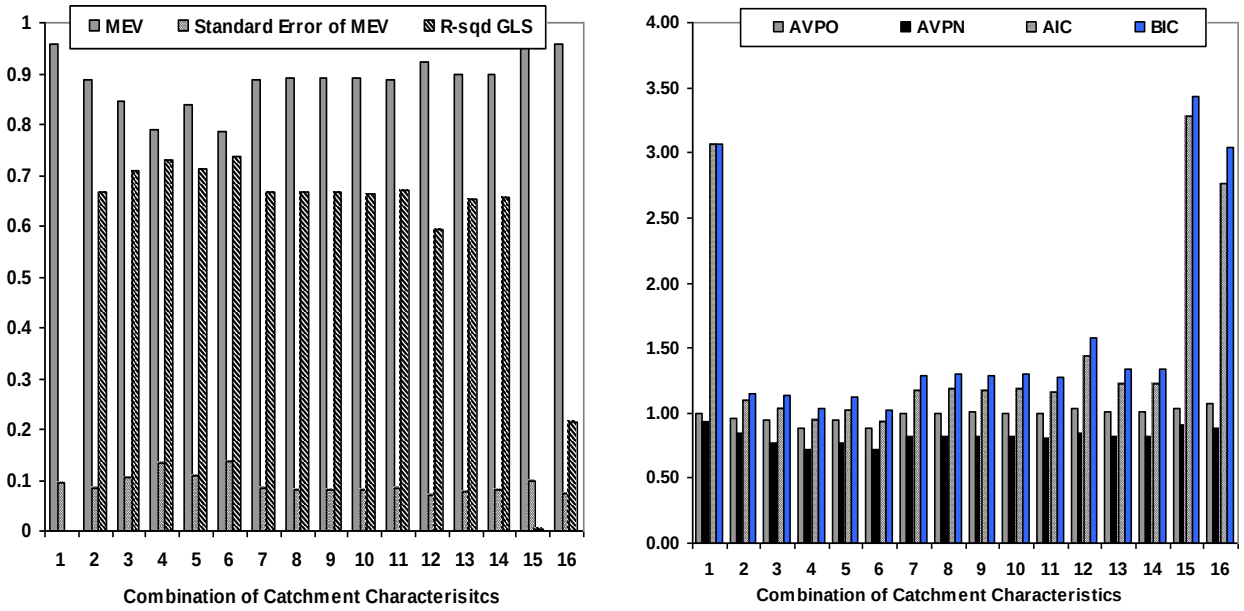


Figure 4.9.1 Selection of predictor variables for the BGLS regression model for the mean model (PRT, fixed region SA), MEV = model error variance, AVPO = average variance of prediction (old), AVPN = average variance of prediction (new) AIC = Akaike information criterion, BIC = Bayesian information criterion

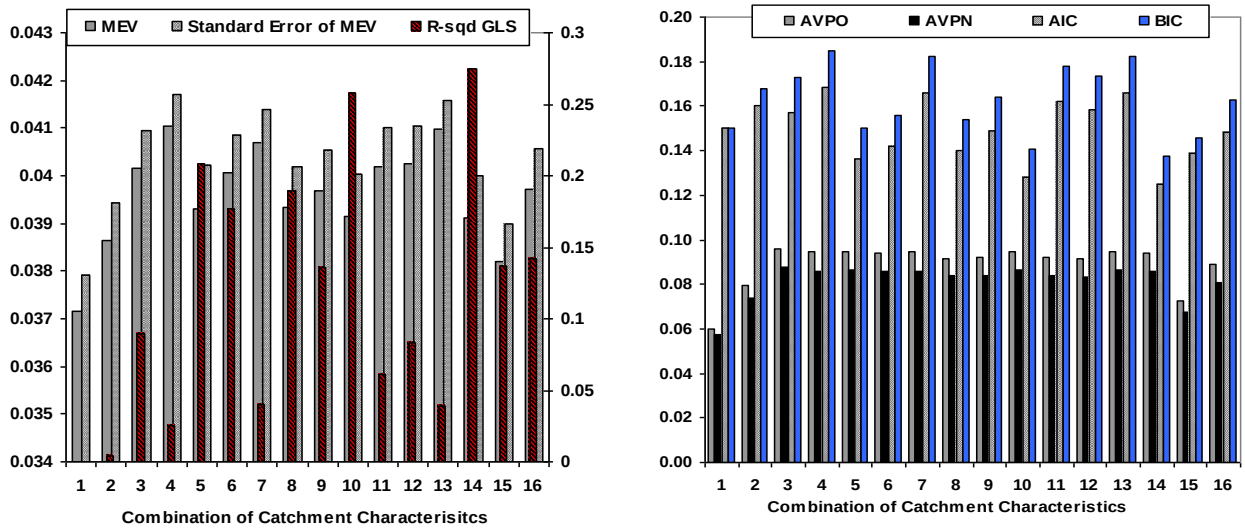


Figure 4.9.2 Selection of predictor variables for the BGLS regression model for skew (R-sqd GLS uses right hand axis)

The developed prediction equations are:

$$M = 1.40 + 1.15[\ln(area) - 4.14] + 5.75[\ln(I_{12,2}) - 1.26] \quad (4.9.1)$$

$$stdev = 1.36 \quad (4.9.2)$$

$$skew = -0.75 \quad (4.9.3)$$

The Pseudo Analysis of Variance (ANOVA) tables for the parameters of the LP3 distribution are presented in Tables 4.9.3 to 4.9.5. For the LP3 parameters, the sampling error increases as the order of moment increases i.e. the EVR (sampling error to model error ratio) increases with the order of moments. The model error dominates the regional analysis for the mean flood and the standard deviation models; this is more pronounced for the mean flood model (0.09 compared to 0.18). The EVR value for the skew model is 8.4 (Table 4.9.4) which is much higher than that of the mean flood and standard deviation models. This indicates that the skew model is dominated by sampling error and in this case the GLS regression modeling framework should be the preferred. These results indicate as found for all the Australian states that the mean flood shows greater levels of heterogeneity when compared to the standard deviation and skew models.

Table 4.9.3 Pseudo ANOVA table for the mean flood model (PRT, SA)

Source	Degrees of Freedom	Sum of Squares
Model	*k=3*	5.0
Model error δ	*n-k-1=25*	21.8
Sampling error η	*N = 29*	2.02
Total	*2n-1 =57*	**28.8**
	EVR	**0.09**

Table 4.9.4 Pseudo ANOVA table for the standard deviation model (PRT, SA)

Source	Degrees of Freedom	Sum of Squares
Model	*k=1*	0.0
Model error δ	*n-k-1=27*	13.1
Sampling error η	*N = 29*	2.32
Total	*2n-1 = 57*	**15.4**
	EVR	**0.18**

Table 4.9.5 Pseudo ANOVA table for the skew model (PRT, SA)

Source	Degrees of Freedom	Sum of Squares
Model	*k=1*	0.0
Model error δ	*n-k-1=27*	1.07
Sampling error η	*N = 29*	9.0
Total	*2n-1 = 57*	**10.1**
	EVR	**8.4**

The underlying model assumptions are examined (i.e. the normality of residuals) using the plots of the standardised residuals vs. predicted values. The predicted values were obtained from one-at-a-time cross validation. Figure 4.9.3 shows the plots for the Q_{20} and Q_{50} estimated by the PRT. It can be seen that most of the standardised residuals fall between the magnitudes of ± 2, hence the underlying model assumptions are satisfied satisfactorily. Also no specific pattern (heteroscedasicity) can be identified with the standardised values being almost equally distributed below and above zero. Similar results were obtained for the mean, standard deviation, skew and other flood quantiles estimated by the PRT.

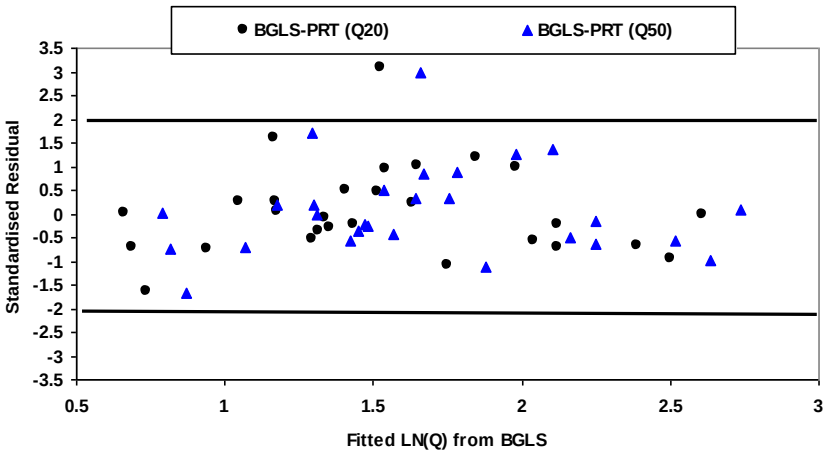


Figure 4.9.3 plots of standardised residuals vs. predicted values for ARI of 20 and 50 years (PRT, SA)

The QQ-plots of the standardised residuals (Equation 4.2.13) vs. normal score (Equation 4.2.14) for the one-at-a-time cross validation were examined. Figures 4.9.4 and 4.9.5 present the results for Q_{20}, Q_{50} and the mean flood models, which show that most of the points closely follow a straight line, while some points also fall away from the line. This indicates

that the assumption of normality and the homogeneity of variance of the standardised residuals have been satisfied reasonably well. The standardised residuals are indeed normally and independently distributed N(0,1) (with mean 0 and variance 1) as the slope of the best fit line in the QQ-plot, which can be interpreted as the standard deviation of the normal score (Z score) of the quantile, should approach 1 and the intercept, which is the mean of the normal score of the quantile should approach 0 as the number of sites increases. It can be observed from Figures 4.9.4 and 4.9.5 show that the fitted lines for the developed models pass through the origin (0, 0) and they have a slope approximately equal to one. Similar results were also found for the standard deviation and skew models and other flood quantiles estimated by the PRT.

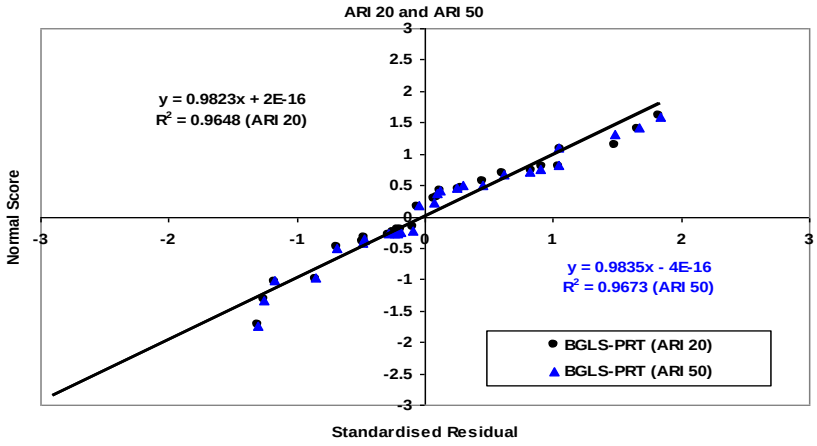


Figure 4.9.4 QQ plot of the standardised residuals vs. Z score for ARI of 20 and 50 years (PRT, SA)

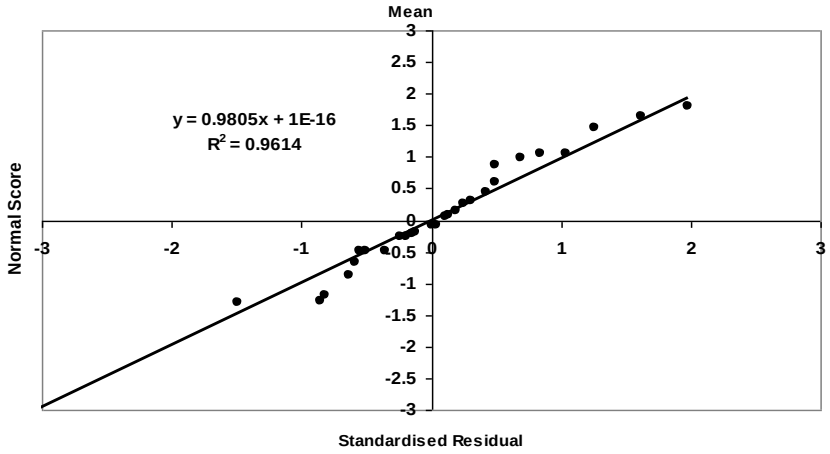


Figure 4.9.5 QQ plot of the standardised residuals vs. Z score for the mean flood model (PRT, SA)

The summary of various regression diagnostics (the relevant equations are described in Section 4.2.2) is provided in Table 4.9.6. This shows that for the mean flood model the model error variance (MEV) and average standard error of prediction (SEP) are much higher than those of the standard deviation and skew models. The large model error variance indicates that the mean flood model has the highest uncertainty associated with its estimation; this was found with all other Australian states as well. The result is inline with the ANOVA analysis. For the standard deviation and skew models, the SEP was 79% and 25% respectively again indicating that both the standard deviation and skew may be regionalised more accurately than the mean flood.

Table 4.9.6 Regression diagnostics for fixed region analysis for SA

Model	Fixed region			
	MEV	AVP	SEP (%)	R^2_{GLS} (%)
Mean	0.78	0.88	119	74
Stdev	0.47	0.49	79	0
Skew	0.038	0.062	25	0

Table 4.9.7 presents the root mean square error (RMSE) (Equation 4.2.16) and relative error (RE) (Equation 4.2.15) values for the quantiles estimated by the PRT. In terms of RMSE, the 100 year ARI shows the highest value (19%). The smallest values are found for the 5 and 10 years ARIs. South Australia showed some of the highest RMSE as compared to the other Australian states. The RE values ranged from 54% to 73%, the smallest value being for the 5 year ARI, while the highest being for the 100 year ARI. Again, the RE values are considered reasonable for the ARIs considered in this study. However they are a bit larger as compared to the RE values for the other Australian states.

Table 4.9.8 shows the results of counting the Q_{pred}/Q_{obs} ratios for the PRT method. The use of this ratio has been discussed in Section 4.2.4. There are on average 57% of cases that are in the desirable estimation range. The cases for overestimation and underestimation on average are 26% and 16% respectively.

Table 4.9.7 Evaluation statistics (RMSE and RE) from one-at-a-time cross validation for SA

Model	RMSE (%)	RE (%)
	PRT	PRT
Q_2	17	58
Q_5	15	54
Q_{10}	15	57
Q_{20}	16	57
Q_{50}	18	66
Q_{100}	19	73

Table 4.9.8 Summary of counts based on Q_{pred}/Q_{obs} ratio values for PRT for SA (fixed region). "U" = gross underestimation, "D" = desirable and "O" = gross overestimation

Model	Count (PRT)			Percent (PRT)		
	U	D	O	U	D	O
Q_2	6	18	5	21	62	17
Q_5	4	18	7	14	62	24
Q_{10}	5	17	7	17	59	24
Q_{20}	5	17	7	17	59	24
Q_{50}	4	15	10	14	52	34
Q_{100}	4	15	10	14	52	34
Sum / average	**28**	**100**	**46**	**16**	**57**	**26**

4.10 Summary

In this chapter, fixed region and the region-of-influence (ROI) approach has been compared for various Australian states (Victoria, NSW, Qld, WA (south west region)). It has been found that ROI outperforms the fixed region for all the states tested.

The QRT and PRT based on Bayesian GLS regression have been applied to all the Australian states to develop prediction equations for 2 to 100 years flood quantiles. It has been found that the QRT and PRT provide very similar results. Since the PRT offers additional advantages over the QRT, it should be preferable to the QRT.

In the PRT methods, prediction equations for mean flood have been developed for various states/regions, which require two predictor variables, catchment area (*area*) and design rainfall intensity except for the NT where design rainfall intensity is replaced by the mean annual rainfall (*rain*). Also, for the PRT method it has been shown that no separate prediction equation needs to be developed for estimating standard deviation and skew of ln(*Q*), instesd the regional average values weighted by error covariance matrix can be adopted.

For Victoria, NSW and Qld the state boundaries should be removed and the data of these states are to be combined to apply the Bayesian GLS PRT method; for this, the coefficients of the prediction equations should be derived at about 10km grid intervals and tabulated for industry use. A software can be developed to facilitate easy application of this technique, which will be discussed in more detail in Chapter 11.

For some regions, fixed region-PRT methods have been developed which include west Tasmania, east Tasmania, SA, NT, WA (Pilbara region) and WA (Kimberley region).

To apply the developed prediction equations developed here the predictor variables need to be standardised using the data shown in Table 4.10.1. A natural logarithm transformation has been used for developing the prediction equations and the predictor variables were centred around the mean as explained in Equation 4.3.10.

Table 4.10.1 RFFA methods for various states/regions and data needed for standardising the predictor variables

State	Region	Mean of ln(*area*)	Mean of $\ln(^{2}I_{12})$	Mean of ln(*rain*)	Mean of $\ln(^{12}I_{50})$
NSW & ACT	ROI	5.43	1.77		
Victoria	ROI	5.37	1.46		
SA	Fixed region	4.14	1.26		
Tasmania	east Tasmania Fixed region	4.74	1.40		
	west Tasmania Fixed region	5.00	1.55		
Qld	ROI	5.38			2.90
Western Australia	South west region (Drainage Division VI) (ROI)	3.39	1.41		
	Kimberley (Drainage Division VIII) Fixed region	6.46	1.87		
	Pilbara (Drainage Division VII) Fixed region	4.71	1.47		
NT	North-western part (Drainage Division VIII) Fixed region	5.41		7.17	

5. Applicability of regional flood prediction equations to small catchments

5.1 General

Most RFFA techniques are developed based on small to medium sized catchments where sufficiently long streamflow records are available; however, these techniques are often applied to very small catchments as well. The question then arises whether the regional prediction equations developed based on small to medium sized catchments (5 to 1000 km^2) are at all applicable to very small catchments say down to 0.1 km^2. This has been investigated in this chapter. Firstly, the applicability of the QRT and PRT methods in the range of smaller catchments within the available data set is investigated. Secondly, the possibility of extrapolating the developed RFFA methods to very small catchments beyond the available data set is explored.

5.2 Performances of the QRT and PRT for small catchments

The performances of the Bayesian-GLS-PRT method in predicting flood quantiles for smaller catchments (having good streamflow record lengths) are investigated in this section. These catchments form part of the database used to develop and test the RFFA method as presented in Chapter 4. The error statistics presented here are based on the one-at-a-time cross validation method discussed in Chapter 4.

The plot of absolute relative error vs. catchment size (Figure 5.2.1) did not show any evidence that the Bayesian-GLS-PRT method performs more poorly for smaller catchments in the database than for the medium and larger catchments – the RE does not show any noticeable increase for smaller catchments. The median relative error values (for Q_{20}) for different catchment sizes for different states are presented in Table 5.2.1, which show that except for SA, the relative error values for catchments smaller than 50 km^2 are not noticeably higher than for the medium and larger catchments in the database (in fact the medium sized catchments tend to have the highest median RE values).

The plots of predicted vs. observed flood quantiles for smaller catchments (examples shown in Figures 5.2.2 to 5.2.4) generally show quite satisfactory results, with the NSW catchments showing the best results.

The overall conclusion from these investigations is that there is no evidence from the catchments in the dataset to indicate inferior performance of the derived regional flood estimation models for the smaller catchments (2 to 50 km^2)

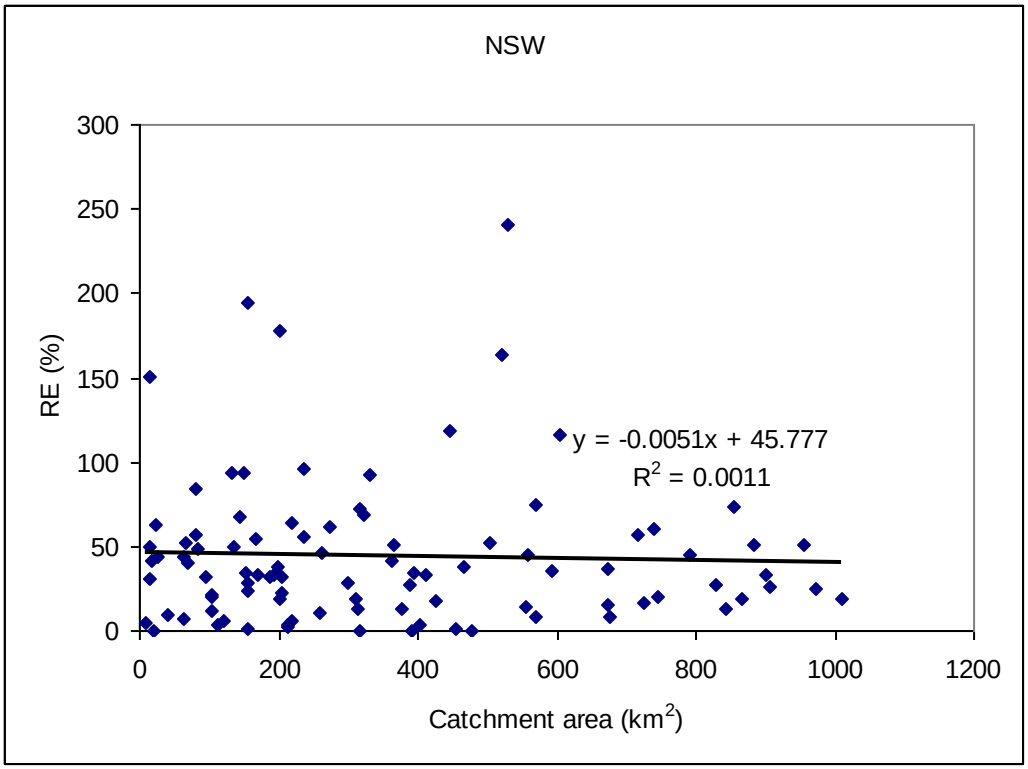


Figure 5.2.1 Plot of relative error (RE %) vs. catchment size (Q_{20}, NSW)

Table 5.2.1 Median relative error (RE) values for different catchment sizes in different states (Q_{20})

Catchment area (km^2)	Median RE (%)						
	NSW	VIC	Qld	TAS	NT	WA (south-west region)	SA
≤ 50 km^2	42	26	28	45	44	50	71
51 to 200 km^2	45	50	36	54	41	58	70
> 201 km^2	32	43	33	40	41	45	53
All data	33	43	35	44	41	52	57

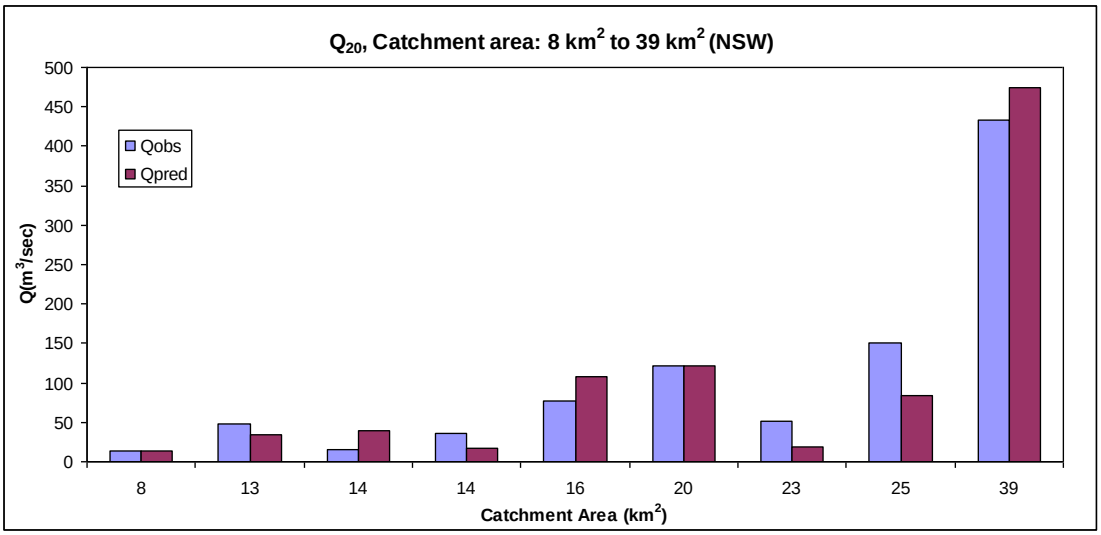


Figure 5.2.2 Comparison of Q_{20} flood quantiles for smaller catchments (NSW)

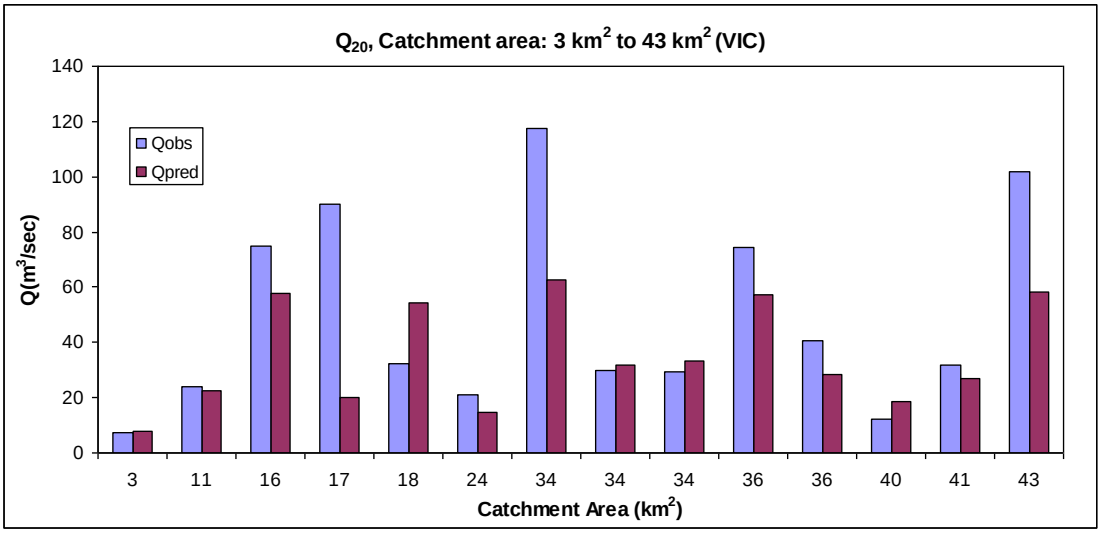


Figure 5.2.3 Comparison of Q_{20} flood quantiles for smaller catchments (Vic)

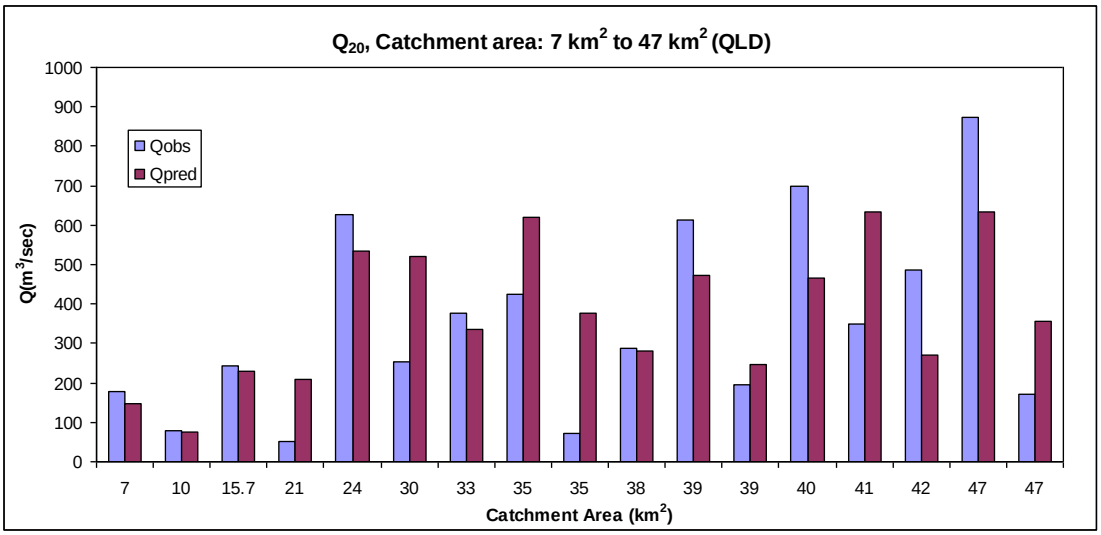


Figure 5.2.4 Comparison of Q_{20} flood quantiles for smaller catchments (Qld)

5.3 Extrapolation of the regional flood estimation methods to very small catchments

5.3.1 General

There is a notion that smaller catchments produce a higher unit area runoff than the larger catchments, which should be accounted for by the exponent of catchment area (area) in the regression equation (this exponent is generally smaller than 1 in the log domain regression model). This observation gets complicated when additional predictor variables are added to the prediction equation. The question though is whether the exponent with catchment area should remain constant for all the area ranges or should vary with catchment area, i.e. whether smaller catchment ranges should have a higher exponent value in the prediction equations. These issues are investigated in this section using data from eastern Australia.

5.3.2 Data used

A total of 429 catchments were selected for this investigation from the eastern part of Australia i.e. from states of New South Wales (NSW), Victoria and Queensland, as shown in Figure 5.3.1. These catchments are mainly unregulated with no major known land use changes over the period of streamflow records. The distribution of catchment areas shown in Figure 5.3.2 indicates that most of the selected catchments are smaller than 400 km^2. The range of catchment areas is 2.3 km^2 to 1010 km^2, with a mean and median value of 309 km^2 and 241 km^2, respectively. About 1.2% of catchments are smaller than 5 km^2, 3.1% are smaller than 10 km^2 and 7% are smaller than 20 km^2, which implies that the data set contains only a small proportion of very small catchments. The record length of the annual

maximum flood series ranges from 20 to 94 years with mean and median values of 36 and 33 years respectively. The distribution of streamflow record lengths of the selected 429 stations is shown in Figure 5.3.3.

The streamflow data of these stations were prepared for RFFA through a stringent procedure, e.g. gaps were filled, outliers were detected, rating curves error was detected and accounted for in at-site flood frequency analysis and trend analysis was conducted, as detailed in Haddad et al. (2010). Two predictor variables are used for this investigation (Table 5.3.1).

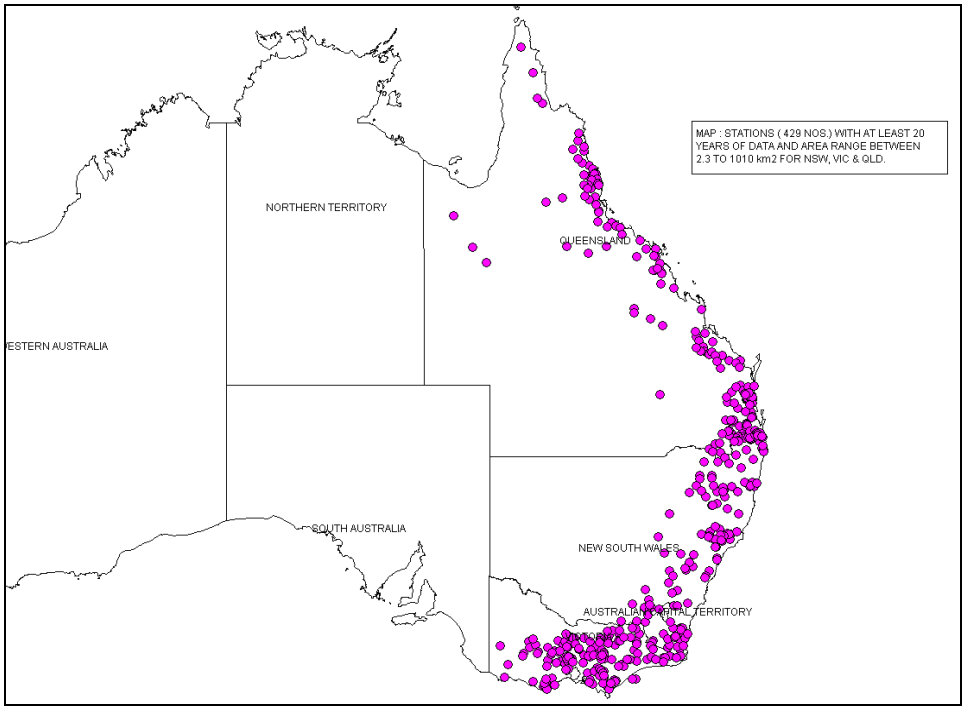


Figure 5.3.1 Locations of catchments used for scaling study

Table 5.3.1 Catchment characteristics variables used

Catchment Characteristics
1. *area*: Catchment area (km^2) 2. *I*: Design rainfall values in mm/h: I_{tc_Y} (where Y = 2, 5, 10, 20, 50 and 100 years and and t_c = time of concentration (hours), estimated from $t_c = 0.76(\text{area})^{0.38}$)

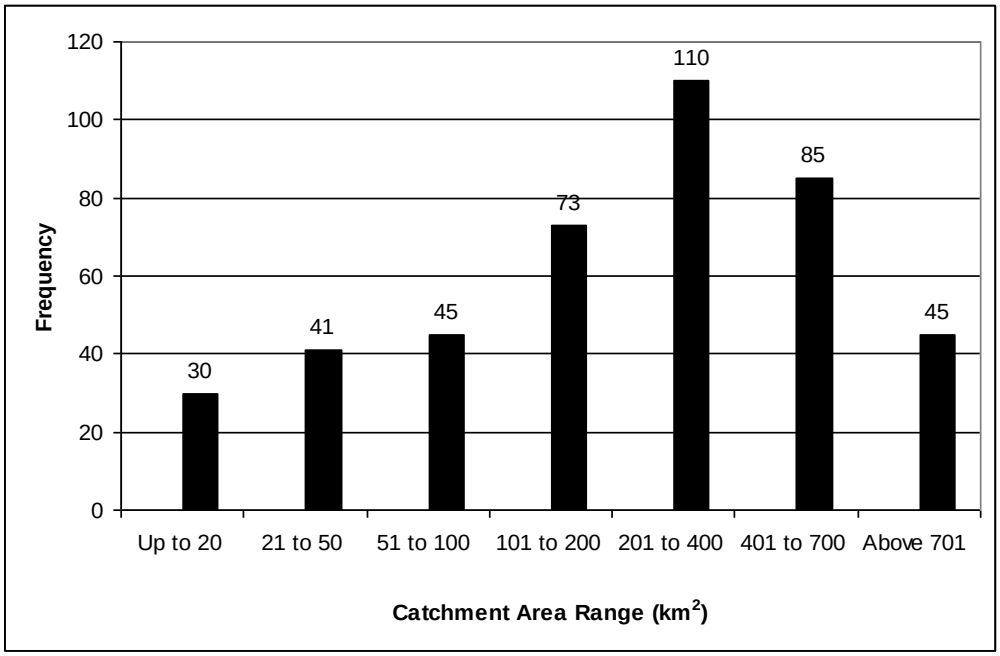


Figure 5.3.2 Distribution of catchment sizes in selected data set

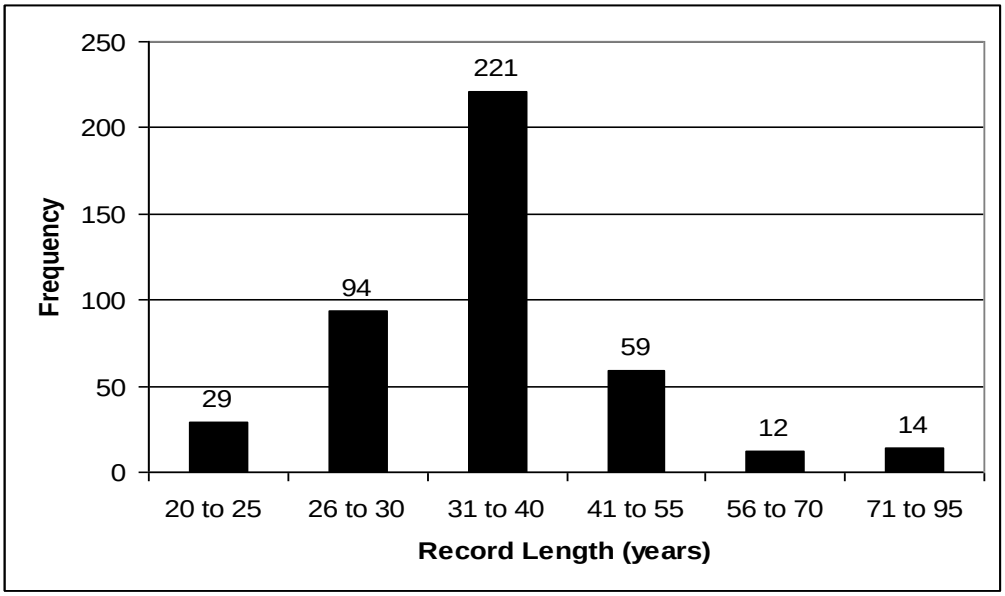


Figure 5.3.3 Distribution of streamflow record lengths of the selected stations

5.3.3 Method

To account for the effects of catchment size on flood quantile estimates, development of a simple scale correction factor (SCF) is adopted here. The method assumes that flood quantiles estimated from the QRT/PRT can be corrected by applying a simple SCF:

$$SCF \propto \frac{1}{area} \quad (5.1)$$

To apply Equation 5.1, regional flood prediction equations are developed using QRT. An empirical analysis is then undertaken to evaluate Equation 5.1, as discussed in Section 5.3.4.

5.3.4 Results

To visualize the scaling property between the flood quantiles and the catchment size, a preliminary data analysis was carried out, which involved grouping the selected catchments into three class intervals based on catchment size: up to 100 km^2, between 101 km^2 and 400 km^2, and above 400 km^2. The median standardised peak discharge per unit area corresponding to each class interval was calculated for various ARIs and plotted against catchment area, as shown in Figure 5.3.4. This figure shows a clear pattern, i.e. smaller catchments produce greater discharge per unit area, which is more prominent as the ARI increases. Figure 5.3.5 shows that smaller catchments generally produce larger unit discharges. Some of these smaller catchments showing higher unit discharge might be located in the wetter parts of the region but the general trend that the smaller the catchment the greater the unit discharge is clearly evident in Figures 5.3.4 and 5.3.5.

Each of the flood quantiles Q_2, Q_5, Q_{10}, Q_{20}, Q_{50} and Q_{100} was regressed (using ordinary least squares regression) against 2 predictor variables (A and I_{tc_Y}) using statistical package SPSS. A number of alternative models were developed for each of the quantiles and the ones showing the highest coefficient of determination (R^2) and satisfying the model assumptions quite closely were selected as the final models (Equations 5.2 to 5.7). The regression coefficients in the prediction equations were found to be significantly different from zero (at significance level of 0.05 or less). The values of R^2 are reasonably high (range: 0.68-0.80) for Australian conditions.

The selected prediction equations are given below:

$$\log_{10}(Q_2) = -3.055 + 1.186\log_{10}(A) + 2.103\log_{10}(I_{tc,2}) \quad (5.2)$$

R^2 = 0.78, adjusted R^2= 0.779

$$\log_{10}(Q_5) = -2.847 + 1.182\log_{10}(A) + 2.089\log_{10}(I_{tc,5}) \quad (5.3)$$

R^2 = 0.805, adjusted R^2= 0.804

$$\log_{10}(Q_{10}) = -2.476 + 1.13\log_{10}(A) + 1.932\log_{10}(I_{tc,10}) \quad (5.4)$$

R^2 = 0.764, adjusted R^2= 0.763

$$\log_{10}(Q_{20}) = -2.766 + 1.173\log_{10}(A) + 2.108\log_{10}(I_{tc,20}) \quad (5.5)$$

R^2 = 0.763, adjusted R^2= 0.762

$$\log_{10}(Q_{50}) = -2.793 + 1.169\log_{10}(A) + 2.132\log_{10}(I_{tc,50}) \quad (5.6)$$

R^2 = 0.722, adjusted R^2= 0.720

$$\log_{10}(Q_{100}) = -2.789 + 1.159\log_{10}(A) + 2.135\log_{10}(I_{tc,100}) \quad (5.7)$$

R^2 = 0.684, adjusted R^2= 0.682

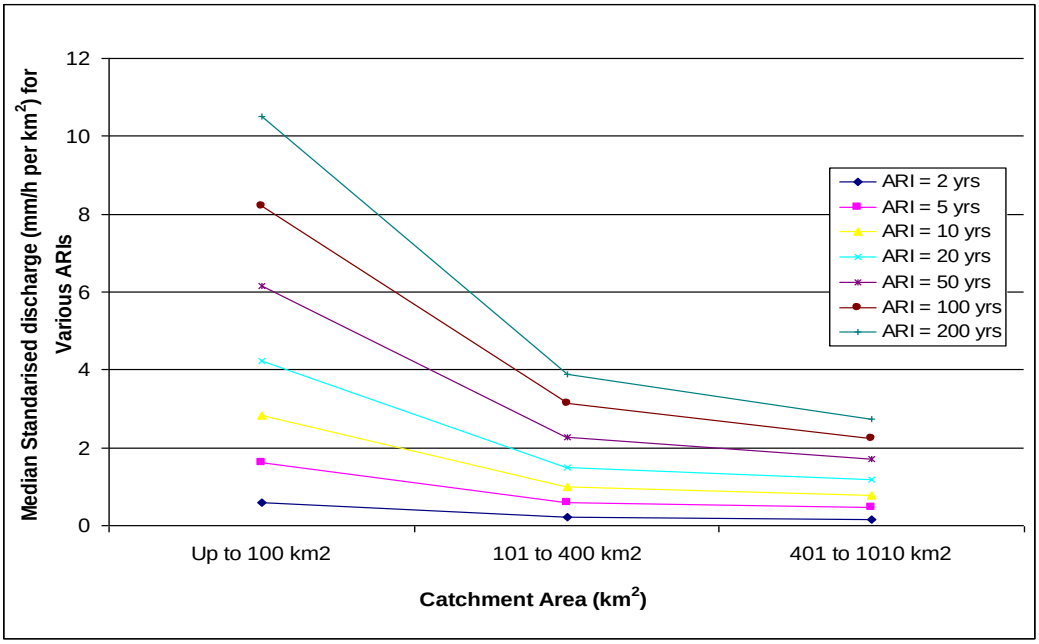


Figure 5.3.4 Relationship between median standardised discharge per unit area and catchment size

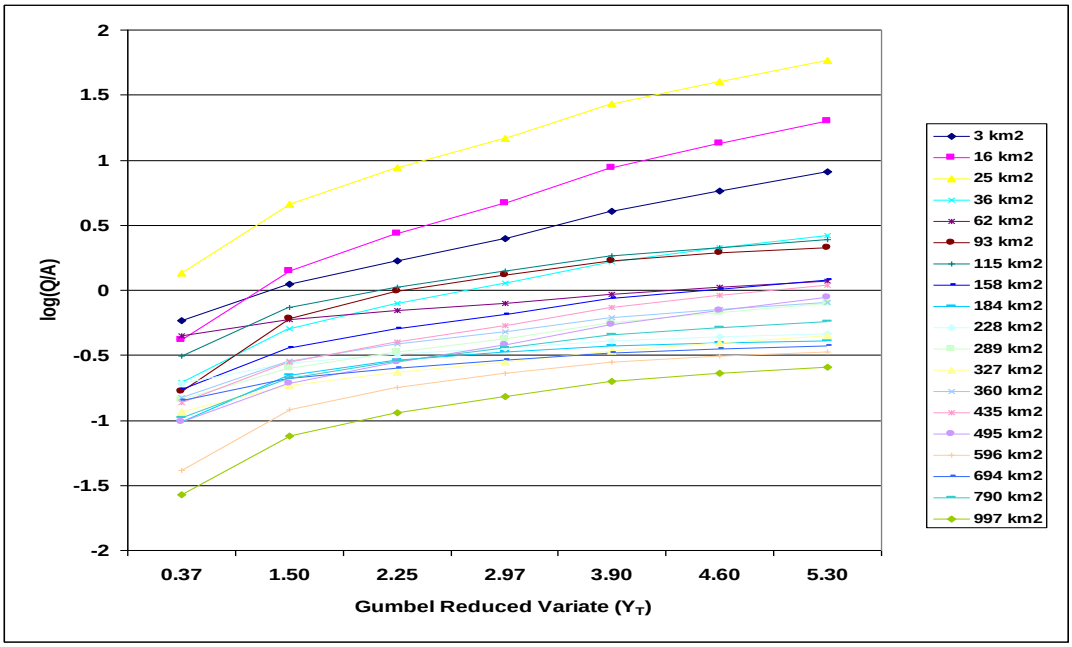


Figure 5.3.5 Peak discharge per unit area for catchments of various size

To account for the scaling factor, i.e. to adjust the flood quantile estimates obtained from Equations 5.3 to 5.8, the following equation is proposed:

$$Q_{T\ \text{scale corrected}} = Q_T \times SCF \qquad (5.8)$$

where Q_T is to be obtained from developed RFFA models like Equations 5.2 to 5.7 and the scale correction factor (SCF) is to be estimated from following equation:

$$SCF = a(b + area)^c \qquad (5.9)$$

where *a*, *b* and *c* are coefficients to be estimated from empirical analysis of the at-site flood quantiles and catchment size data; these coefficients may be used as estimators of the scale corrected values of Q_T. In this study, the values of *a*, *b* and *c* have been estimated using data from 429 catchments from Victoria, NSW and Qld and are presented in Table 5.3.2. Equation 5.9 is plotted in Figure 5.3.6 for various ARIs and catchment size. Figure 5.3.6 shows that the proposed SCF increases as the catchment area decreases and SCF = 1 for *area* =200 km^2. This implies that for catchments greater than 200 km^2, no scale correction is needed.

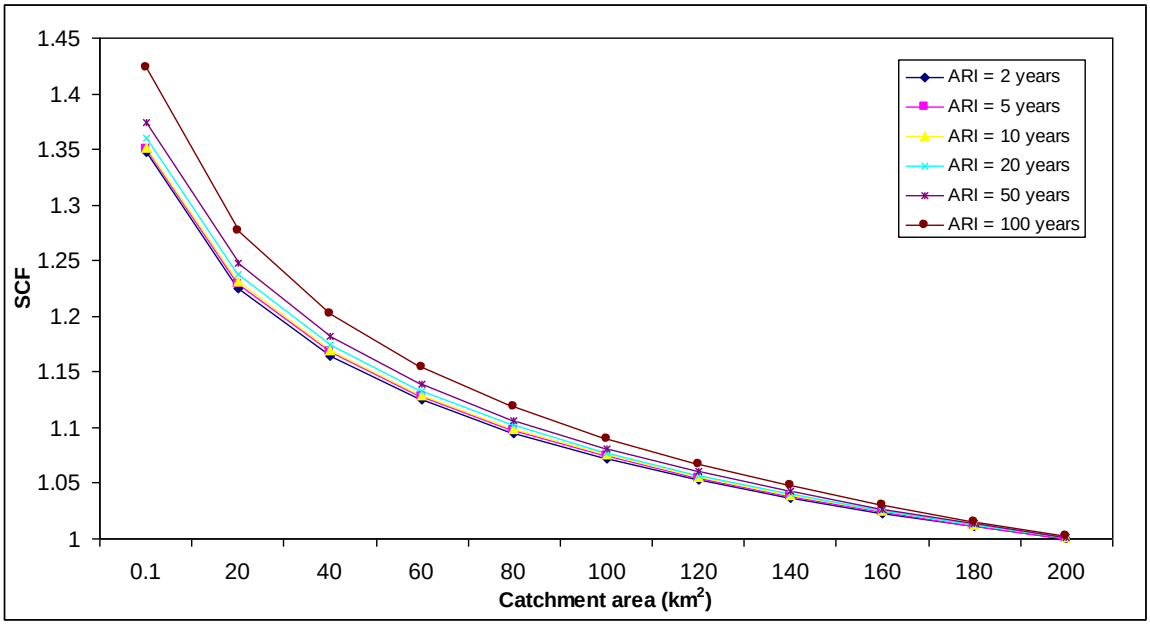


Figure 5.3.6 Relationship between scale correction factor (SCF) and catchment area

Table 5.3.2 Values of *a*, *b* and *c* (Equation 5.9) for different ARIs

ARI (years)	Estimated coefficients		
	a	*b*	*c*
2	1.813	14.43	-0.1108
5	1.842	15.29	-0.1136
10	1.859	15.78	-0.1152
20	1.89	16.02	-0.1182
50	1.952	16.82	-0.1241
100	2.07	16.00	-0.1349

A total of 33 catchments were selected randomly from the database of the 429 catchments and the scale correction factors (Equations 5.8 and 5.9) were applied and the median relative error values were estimated with respect to at-site flood frequency estimates, as shown in Table 5.3.3. It is seen that application of the scale correction factor has reduced the median relative error in the estimation. The plots of uncorrected and corrected flood quantile estimates in Figures 5.3.7 and 5.3.8 demonstrate the improvement in the corrected flood quantiles as compared to the observed flood quantiles.

Table 5.3.3 Median relative error (RE) for uncorrected and corrected flood quantiles for different area range (Q_{20})

Catchment area range (km^2)	RE (%) Uncorrected quantiles	RE (%) Scale corrected quantiles
2 - 5	21	19
6 - 10	49	34
11 - 50	31	27
51 - 200	37	34

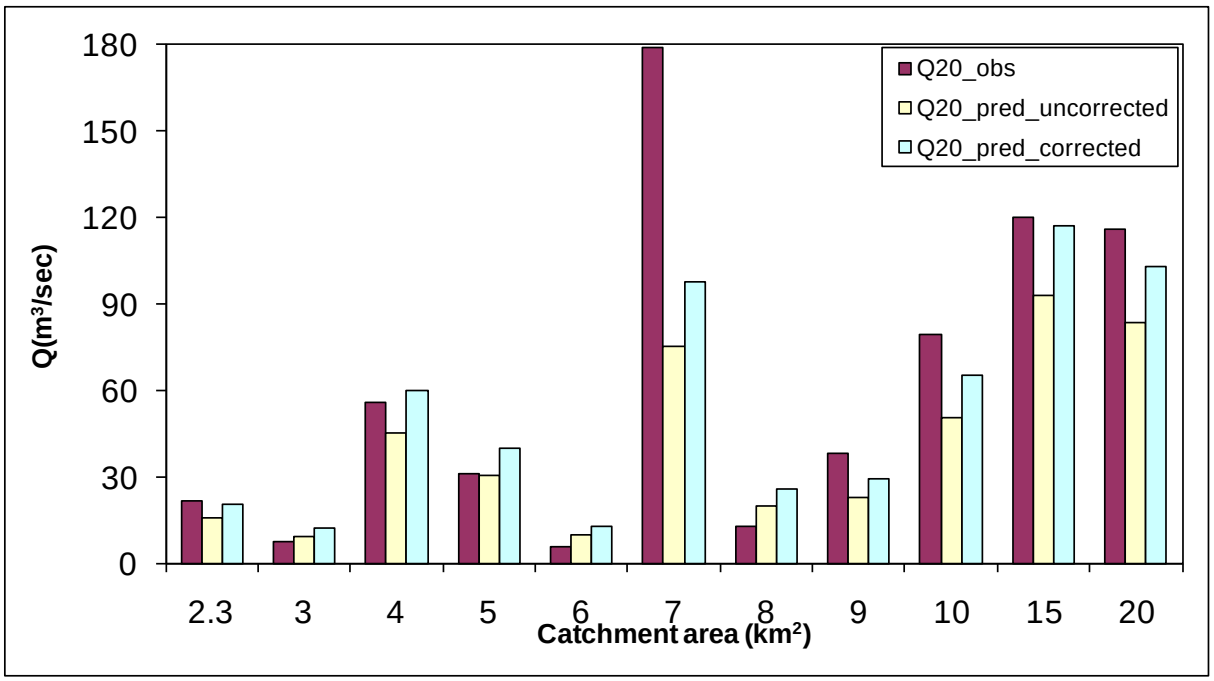


Figure 5.3.7 Comparison of scale corrected flood quantiles with observed flood quantiles (Q_{20}, catchment area range: 2.3 km^2 to 20 km^2)

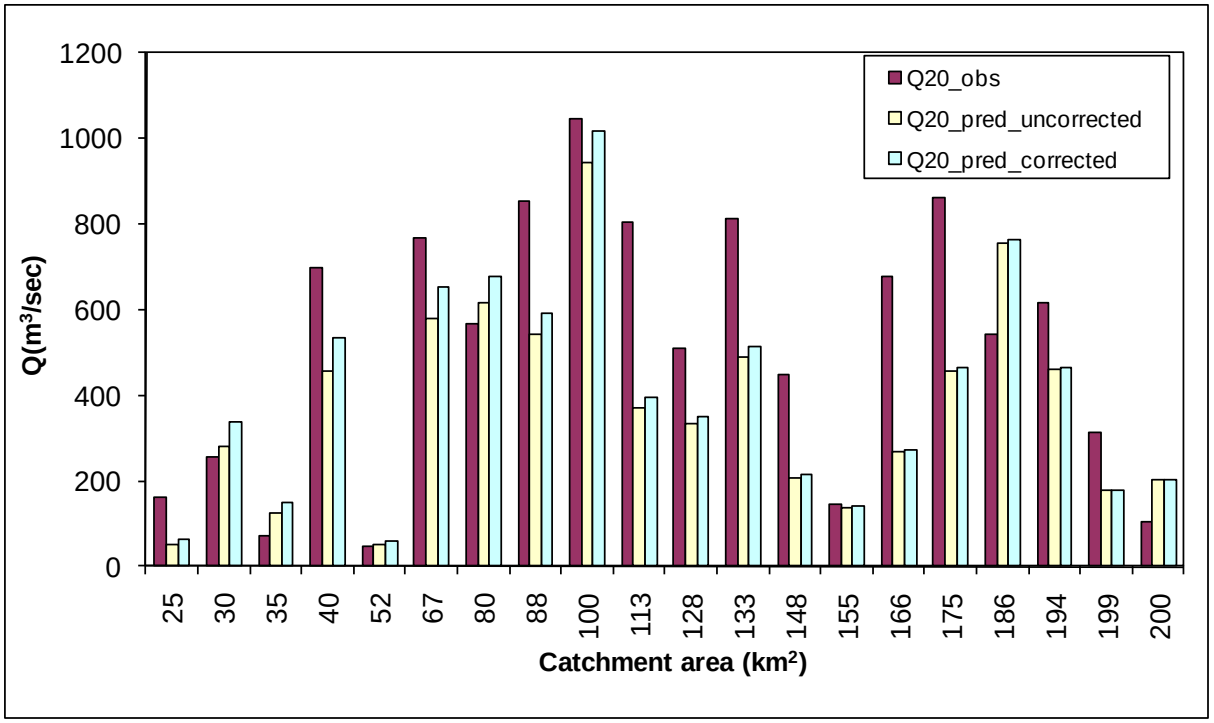


Figure 5.3.8 Comparison of scale corrected flood quantiles with observed flood quantiles (Q_{20}, catchment area range: 21 km^2 to 200 km^2)

5.4 Sensitivity analysis

To carry out a sensitivity analysis on how the estimated flood quantiles may vary with catchment size, two different regression models are considered: Model 1 contains the two predictor variables of catchment area (*area*) and rainfall intensity for 12 hours duration and 2 years ARI (I_{12_2}), while Model 2 contains catchment area (*area*) and design rainfall intensity I_{tc_ARI}. The objective of using two different types of prediction equations as mentioned above is to assess how predicted flood quantiles reduce with catchment size. For Model 1, the design rainfall intensity (I_{12_2}) does not change with catchment area, i.e. it has fixed duration of 12 hours; however, for Model 2, the design rainfall intensity I_{tc_ARI} changes with t_c, where t_c varies with catchment size according to $t_c = 0.76(area)^{0.38}$, i.e. as catchment size reduces, I_{tc_ARI} increases. The use of these two types of prediction equations allows examining the nature of variation in the predicted flood quantiles when catchment size is reduced progressively from a high value to a very small value.

To carry out the sensitivity how estimated flood quantiles may vary with catchment size, two different regression models are considered: Model 1 contains two predictor variables, catchment area (*area*) and rainfall intensity for 12 hours duration and 2 years ARI (I_{12_2}), and Model 2 contains catchment area (*area*) and design rainfall intensity I_{tc_ARI}. The objectives of

using two different types of prediction equations as mentioned above are to assess how predicted flood quantiles reduce with catchment size. For Model 1, the design rainfall intensity (I_{12_2}) does not change with catchment area i.e. it has fixed duration of 12 hours; however, for Model 2, the design rainfall intensity I_{tc_ARI} changes with t_c, where t_c varies with catchment size according to Equation 5.2 i.e. as catchment size reduces I_{tc_ARI} increases. The use of these two types of prediction equations allows examining the nature of variation in the predicted flood quantiles when catchment size is reduced progressively from a high value to a very small value.

To develop Model 1 and Model 2, we used data from the 96 NSW catchments (these have been used to develop RFFA models for NSW in Chapter 4). Here, we considered Q_{20} and prediction equations were developed using OLS regression procedure in the SPSS software. The developed prediction equations are shown in Table 5.4.1. The R^2 values of these equations are 0.70 and 0.71, which are comparable to the results from the Bayesian GLS regression in Chapter 4. The Durbin-Watson statistic values are 1.84 and 1.96 for the two predictor variables, which are close to 2; this indicates that the predictor variables are not highly correlated.

To examine the effects of reduced catchment size on predicted flood quantile, Models 1 and 2 were applied to two NSW catchments (412063 & 419054) having catchment areas of 570 km^2 and 391 km^2 respectively. The catchment area is reduced progressively from 570 km^2 (or 391 km^2) to 0.1 km^2; the corresponding $I_{tc,20}$ values are extracted for each of the reduced catchment sizes from ARR Volume 2 (using the BOM IFD calculator and AUS-IFD software). The predicted flood quantiles for 20 years ARI (Q_{pred}) based on Models 1 and 2 are then plotted against the catchment size in Figures 5.4.1 and 5.4.2. From these two figures it is clearly found that Model 1 exhibits a much smoother curve than Model 2 when catchment size is progressively reduced to 0.10 km^2. The discontinuities in flood estimates in Figures 5.4.1 and 5.4.2 are due to the fact that the IFD values have only been defined for a discrete set of durations. While they could be interpolated for any t_c value, the routine application by practitioners would result in the step function as shown in Figure 5.4.1 and 5.4.2. This in essence indicates that for developing scaling factors for smaller catchments, use of Model 1 (which includes area and I_{12_2} as predictor variables) would be preferable.

Table 5.4.1 Model 1 and Model 2 for Q_{20} based on 96 NSW catchments

Region	Prediction equation (Q_{20} in m^3/s, *area* = catchment area in km^2 and $I_{12,2}$ = design rainfall intensity for 2 years ARI and 12 hours duration in mm/h, $I_{tc,20}$ = design rainfall intensity in mm/h for 20 years ARI and duration of t_c hours)	R^2	Durbin-Watson statistic
NSW (96 stations)	Model 1 : $\ln(Q_{20}) = -1.054 + 0.692\ln(area) + 1.856\ln(I_{12\_2})$	0.70	1.84
	Model 2 : $\ln(Q_{20}) = -5.128 + 1.127\ln(area) + 1.828\ln(I_{tc,20})$	0.71	1.96

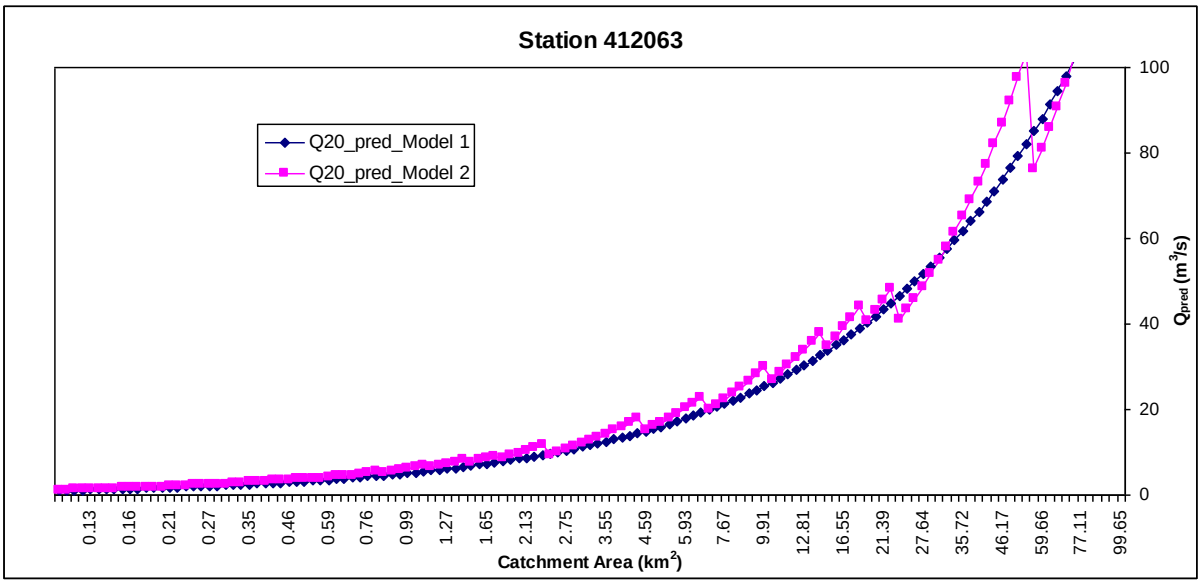


Figure 5.4.1 Plot of Q_{pred} vs. catchment area for two different types of model forms (Station 412063 NSW) (Q_{20})

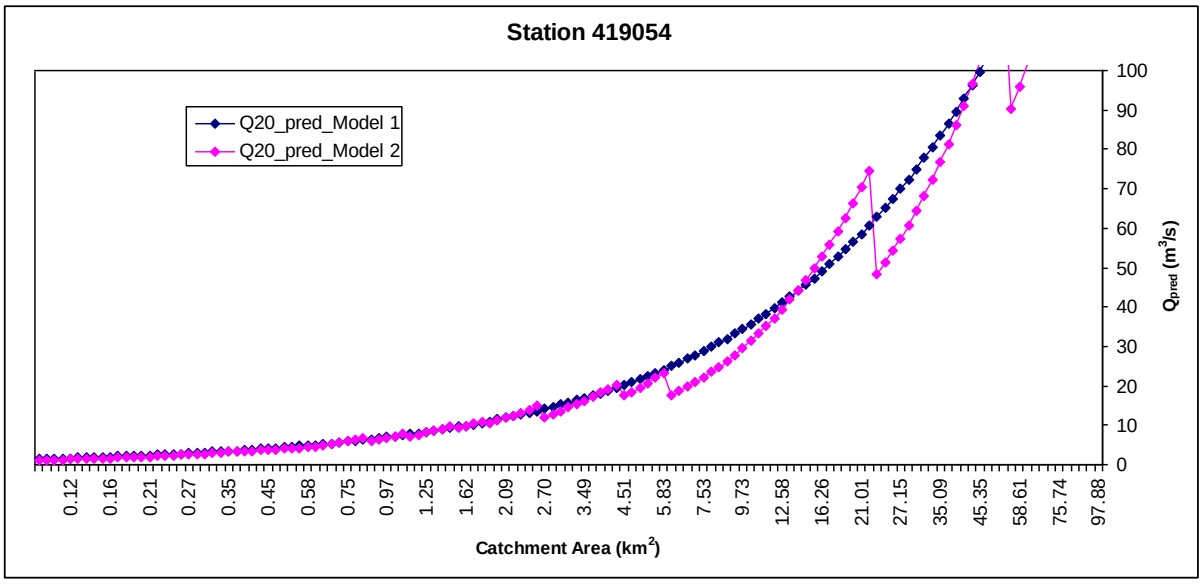


Figure 5.4.2 Plot of Q_{pred} vs. catchment area for two different types of model forms (Station 419054 NSW) (Q_{20})

5.5 Summary

From the investigations presented in this chapter, based on a data set that includes only a limited number of small catchments, it has been found that the Bayesian-GLS-PRT based predictions equations developed in Chapter 4 can provide design flood estimates for smaller catchments (in the range of 2 to 50 km^2) with the same level of accuracy as for the medium to larger size catchments.

However, it has further been found that smaller catchments in general produce a larger unit area flood discharge, and this increase is only partly covered by the exponent for area in the regression equations. A method has been explored here to demonstrate how this finding could possibly be considered in the prediction equations developed using a quantile regression technique.

A scale correction factor is proposed to account for the effects of scale on flood quantile estimates to be applied to smaller catchments. The application of this scale correction factor shows that this provides more accurate flood quantiles for smaller catchments.

Prediction equations with two different design rainfall variables have been applied to examine how smoothly the predicted flood quantiles vary with reduced catchment size. It has been found that the model with fixed design rainfall intensity (e.g. $I_{12,2}$) produces a smoother

variation of flood estimates with catchment area than the model with design rainfall intensity that varies with catchment size and ARI (e.g. $I_{tc,ARI}$).

Further investigation is needed to find whether the RFFA models developed in Chapter 4 can be applied to very small catchments (e.g. smaller than 3 km^2) or if the estimates obtained from these models need to be scaled up for smaller catchments using some empirical relationship, as proposed in this chapter. This has been left for future research, as a part of Stage III of Project 5.

Also it should be noted here that equations 5.2 to 5.7 have been developed using the current IFD data (ARR87). A check will need to be made if the revised IFD data (up-coming ARR 4^{th} edition) will change the nature and magnitude of the correction needed for small catchments.

6. Regional flood estimation technique for major floods: Applicability of a simple Large Flood Regionalisation Model (LFRM)

6.1 Introduction

Estimation of major floods is a necessity in the design of large water infrastructure such as large detention basins, urban trunk drainage, large bridges, dam spillways and other major hydraulic structures. This problem in the past has been addressed by many researchers (e.g. Pilgrim, 1986, Rowbottom et al., 1986; Pilgrim & Rowbottom, 1987; Stedinger et al., 1992; Nathan & Weinmann, 1995). Book VI of Australian Rainfall and Runoff (ARR) was upgraded in 1999 with guidance for estimation of large to probable maximum floods (PMF); in this context, the term 'large' floods refers to floods with 50 to 100 years average recurrence intervals (ARIs) (Nathan & Weinmann, 2001). Floods in the range from 100 years ARI to the 'credible limit of extrapolation' (ARI in the order of 2000 years) are referred to as 'rare' floods, while floods from the credible limit of extrapolation to the PMF are termed 'extreme' floods. The procedures outlined in ARR2001 include flood frequency analysis and various rainfall-based methods. For flood frequency estimates in the range of 'rare' floods, use of regional information plus paleohydrological information was suggested and for rainfall-based methods, an annual exceedance probability (AEP) neutral approach was recommended (Nathan & Weinmann, 2001).

In Project 5, Stage I report (Rahman et al., 2009) and in Haddad, Rahman and Weinmann (2010), a simple Probabilistic Model was presented that can exploit regional flood information over a large region when developing 'easy to apply' prediction equations to estimate major floods. The Probabilistic Model is referred in this report as Large Flood Regionalisation Model (LFRM). The proposed method was intended to offer an alternative flood estimation approach that can be applied for feasibility studies and the design of hydraulic structures in situations where a slightly larger degree of uncertainty in the flood estimates is acceptable. The regionalisation procedure adopted by the Probabilistic Model /LFRM (Majone & Tomirotti, 2004; Majone et al., 2007; Haddad, Rahman and Weinmann, 2010) is based on the assumption that the standardised maximum values of the annual maximum flood series from a large number of individual sites in a region can be pooled (after standardising to allow for the across-sites variations in the mean and coefficient of variation (CV) values of annual maximum floods). The concept is similar to the CRC-FORGE method (Nandakumar et al.,

1997) where design rainfall estimates are based on pooled standardised rainfall extremes from a large region that includes up to several hundred gauges. The particular advantage of the LFRM is that, in contrast to the 'index flood method' approach, it does not assume a constant CV across the sites. This feature, in particular, allows the LFRM to pool data more effectively over a very large region.

The main focus of the LFRM being investigated in ARR Project 5 is to couple this approach with Generalised Least Squares (GLS) regression so that the method can be applied to ungauged catchments. The advantages of GLS regression are that this approach accounts for the variation in flood record lengths across various sites in the region, inter-site correlations of the concurrent flood records and cross-correlated residuals. As a result, GLS estimators are more efficient than those of the ordinary least squares (OLS) and provide more realistic measures of estimation.

6.2 Independence of the data in the simple Large Flood Regionalisation Model

The Probabilistic Model/Large Flood Regionalisation Model (LFRM) (referred to LFRM here) presented by Majone et al. (2007) and applied by Haddad, Rahman and Weinmann (2010) ignores the cross-correlation of the pooled standardised data, where the highest data point from each station's annual maximum (AM) flood series (after standardisation) is combined with all the stations in the region to form a database referred to as 'LFRM data series'. It was assumed that the individual values in the LFRM data series are independent. This assumption may be valid if the data being pooled come from stations that are spread over a large region. However, examination shows (Figure 6.1) that values in the LFRM data series used in Project 5 tend to cluster in some years, with very few events in other years. This appears to violate the assumption of independent distribution of the events in time and indicates that some of the events occurring in the same year might have resulted from the same hydro-meteorological events. However, if the events are separated by at least a few months, they may be treated as being independent.

Significant correlation between events in the pooled series of maxima used for regional flood frequency analysis will result in the effective size of the sample being over-estimated, and the exceedance probabilities of given flood magnitudes being underestimated. The testing of the PM/ LFRM by Haddad, Rahman and Weinmann (2010) has demonstrated that if the LFRM data series is assumed to be independent, the LFRM underestimates the at-site flood frequency estimates, as shown in Table 6.1. Here, 17 out of the 18 test catchments show an

underestimation by 7% to 40%. This result indicates that the issue of cross-correlation needs to be addressed for successful application of the LFRM in Australia.

The cross-correlation among the concurrent annual maximum flood series of all the possible pairs of sites (irrespective of their ranks) (these data have been prepared as a part of ARR Project 5) was examined and it was found that the cross-correlation coefficients are quite high for the nearby pairs of sites. An example is shown in Figure 6.2 where two nearby Victorian stations (Stations 221201 and 221207) show a cross-correlation coefficient of 0.96). The correlations vs. distance between pairs of stations in south-east Australia is shown in Figure 6.3, which indicates that AM flood series have cross-correlation close to 1 for some nearby stations, but cross-correlation reduces with distance sharply. Also high correlation is a dominant issue only for a limited number of pairs of stations.

The cross-correlation between two stations based on all the concurrent AM flood data has little relevance to the LFRM model as this model uses only rank 1 data i.e. the highest flood value from the AM series of each station. A viable approach would be to use average cross-correlation considering all the concurrent AM data from all the possible pairs of stations in the database and develop a spatial dependence model similar to CRC FORGE method (Nandakumar et al., 1997). This model can then be used to account for the cross-correlation in the LFRM data series in flood quantile estimates using the Large Flood Regionalisation Model.

Another approach might be to examine the starts of the individual events which contain the annual maxima for all the sites plotted against the same year (e.g. as in Figure 6.1); if the starts of the events are a few months apart from each other they may be treated independent. If they have resulted from the events which have occurred in the same day or week, only one data point from these can be retained to establish an independent series. Here, if the stations are far away (e.g. one station from Victoria and another from Qld) they can be treated independent, although plotted against the same year, as they are most likely resulted from different hydro-meteorological events. This approach requires the examination of the distances between pairs of stations and the start and end of the individual events, which is time demanding.

Any significant degree of correlation between the events in a regional sample reduces the effective sample size drastically, so the most productive approach might be to establish essentially independent networks of stations (perhaps by using the concept of de-correlation distance as an indicator) and then only pool the maxima from such a network of stations.

Some form of constrained random sampling may need to be used to establish a number of alternative networks of independent stations.

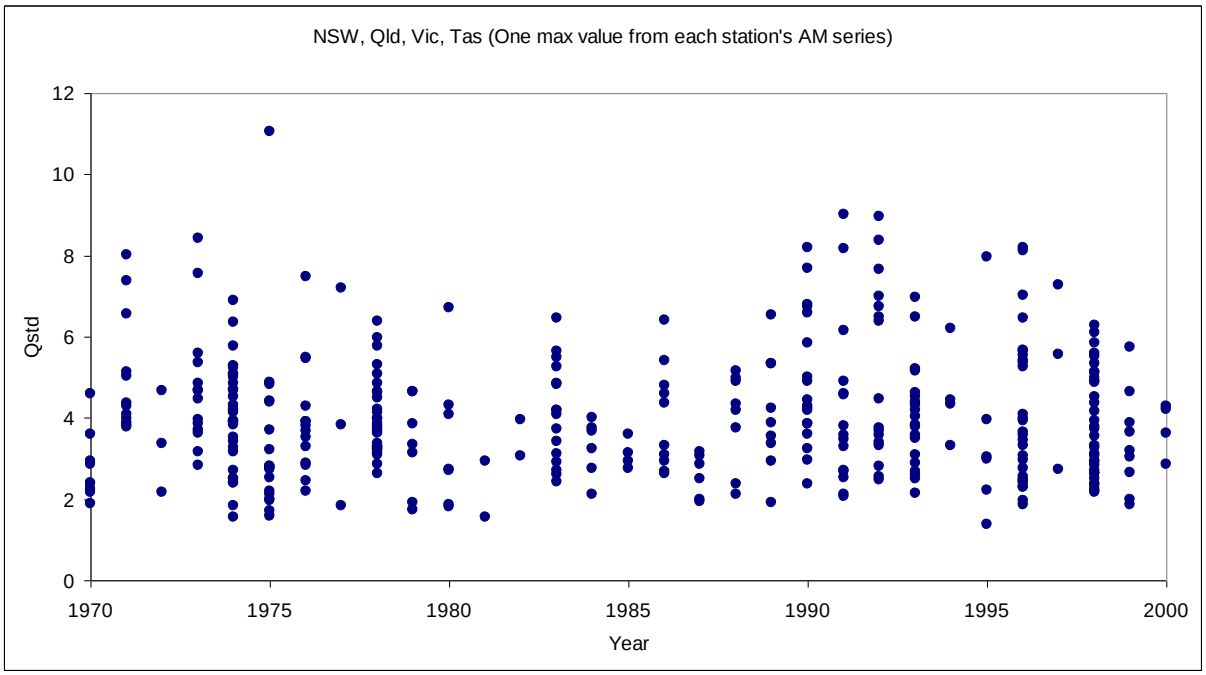


Figure 6.1 Occurrences of the highest floods – data from NSW, Qld, Vic and Tasmania are combined (only the highest value from each station's AM series is taken to form the LFRM data series)

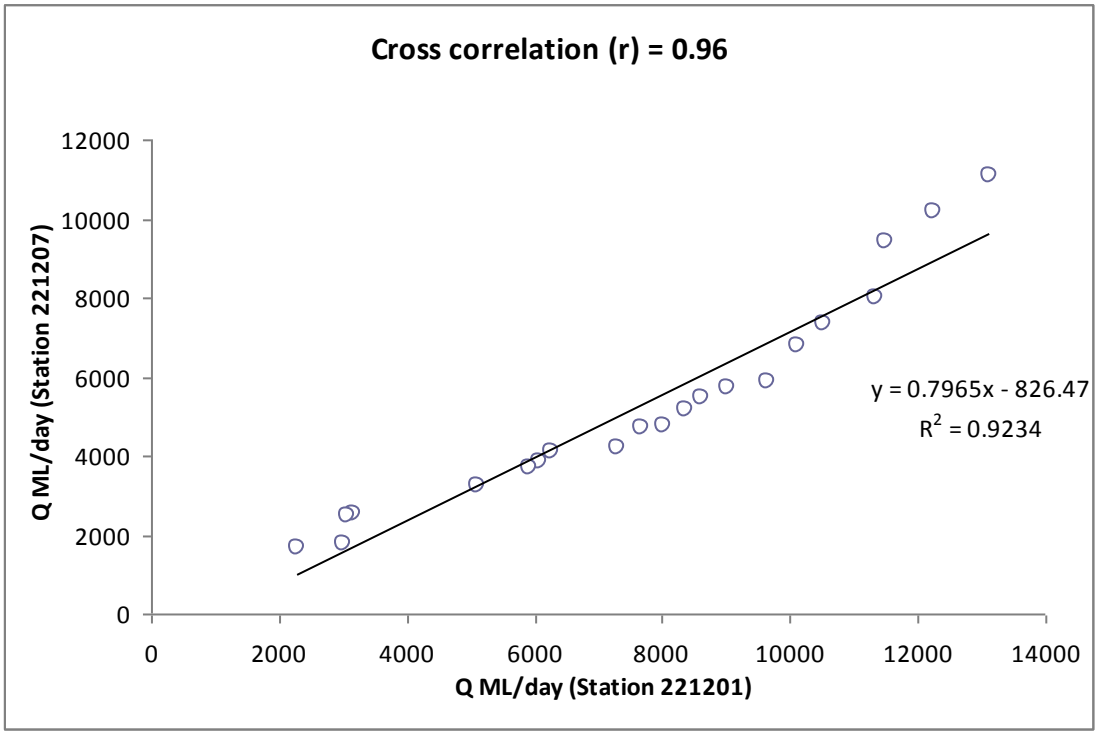


Figure 6.2 Cross-correlation between two nearby Victorian Stations 221201 and 221207 (considering all concurrent AM flood data over the period of records – only 21 data points are concurrent for the pair of stations)

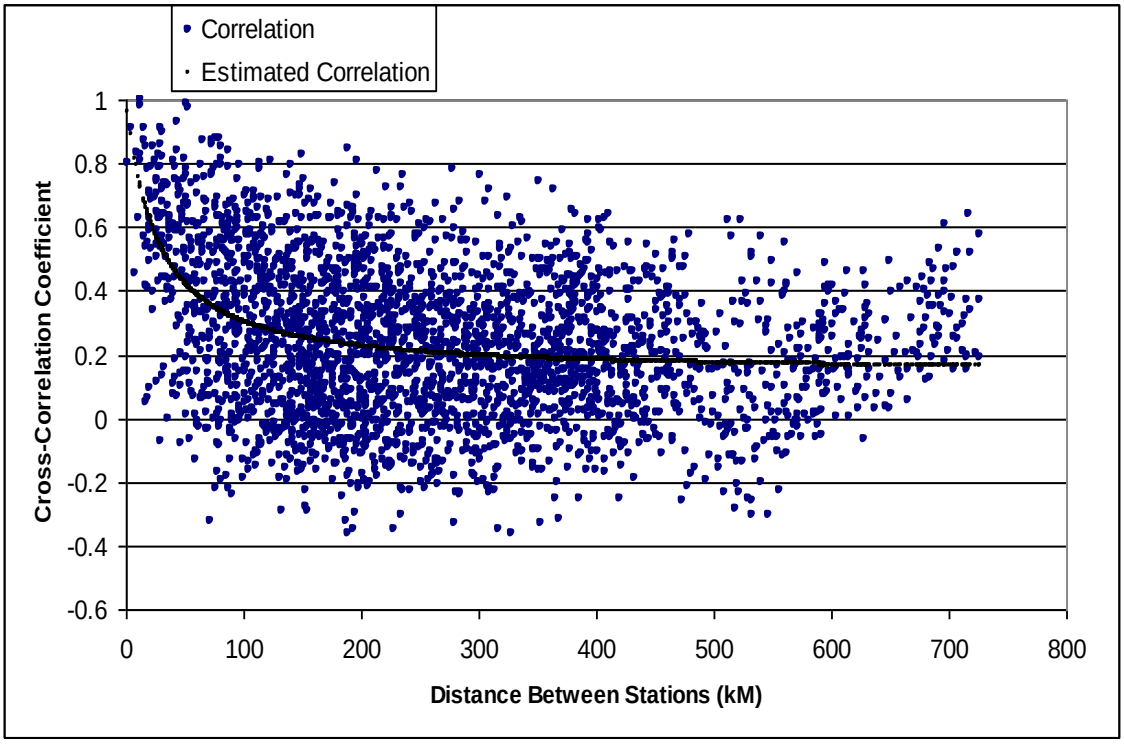


Figure 6.3 Relationship between the cross-correlations among AM data and distance between pairs of stations in south - east Australia

Table 6.1 Comparison of predicted flood quantiles by the AM/LFRM with at-site flood frequency analysis (FFA) estimates for ARI = 100 years (CLL and CLU refer respectively to lower and upper 90% confidence limits of at-site FFA, RE refers to relative error) (Haddad, Rahman and Weinmann, 2010)

Test Catchment / Area (km^2)	LFRM (m^3/s)	CLL (m^3/s)	FFA (m^3/s)	CLU (m^3/s)	RE %
TC4 / 18	50	33	67	217	-25.4
TC12 / 20	150	133	234	497	-35.9
TC16 / 23	68	65	92	233	-26.1
TC5 / 36	50	42	75	283	-33.3
TC6 / 95	200	133	333	981	-39.9
TC15 / 105	830	629	917	1004	-9.5
TC8 / 108	165	94	177	459	-6.8
TC10 / 141	94	71	82	118	14.6
TC11 / 200	118	106	165	400	-28.5
TC3 / 214	59	71	80	118	-26.3
TC13 / 395	990	683	1294	1388	-23.5
TC17 / 402	779	542	1129	1393	-31.0
TC7 / 407	157	189	251	313	-37.5
TC9 / 629	379	375	438	813	-13.5
TC18 / 829	688	500	1000	2875	-31.2
TC1 / 837	1500	1313	2063	3688	-27.3
TC14 / 900	2813	2000	3250	3788	-13.4
TC2 / 943	375	250	438	1063	-14.4

Table 6.2 Concurrent peaks in the top 50% events for Victorian Stations 221201 and 221207

Year	Station 221201, Q (ML/day)	Station 221207, Q (ML/day)	Rank
1998	32909	20118	1
1978	23901	12980	2
1992	8624	5464	11
1994	7682	4733	14

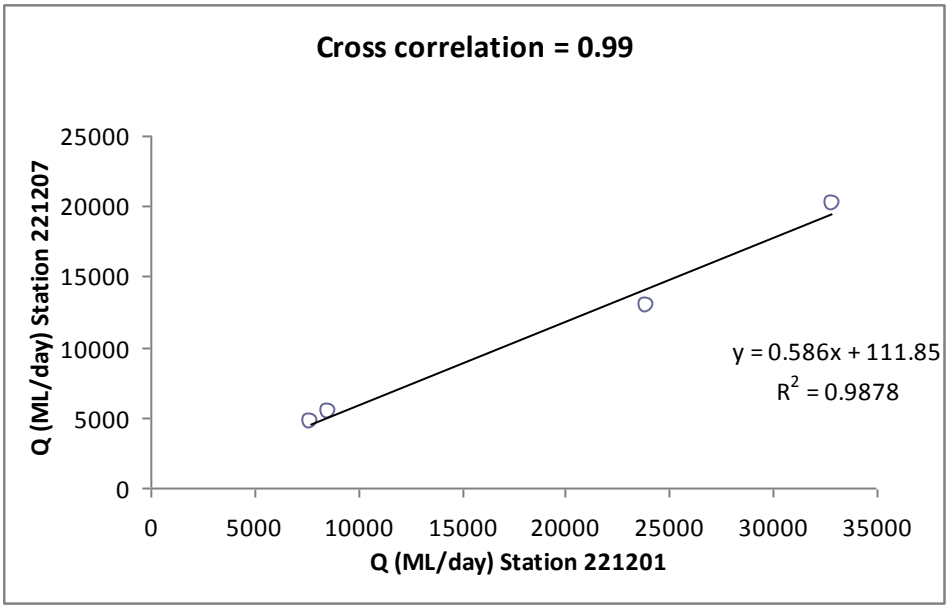


Figure 6.4 Cross-correlation of AM floods between two nearby Victorian Stations 221201 and 221207(considering the top 50% of the concurrent data points)

6.3 Summary

The development of a simple Large Flood Regionalisation Model (LFRM) for regional flood estimation in the major flood range for Australian catchments needs consideration of the cross-correlations among the highest data points from each station's AM series, which however appears to be a difficult task. A number of possible ways to deal with this problem have been discussed in this chapter but additional research and development work is required to identify the most effective approach for practical application to estimate large/rare floods in Australia.. This is left for future research as a part of Project 5 Stage III.

The possible future research tasks on the LFRM are outlined below:

- Form a LFRM data series using the data from all Australian states and territories.

- Assess the degree of dependence of the LFRM data series.
- Develop a method to account for any significant cross-correlation of the LFRM data series, either by an appropriate probability adjustment or by identifying essentially independent maximum events for inclusion in the LFRM data series.
- Develop and test the updated LFRM for flood quantile estimation in the range of 100 to 2000 years ARIs.
- Compare the updated LFRM with alternative methods.
  Recommend a preferred method for application in Australia and state its limits of application in terms of catchment size, location and ARI.

7. Development of regional flood estimation methods for arid and semi-arid regions in Australia

7.1 General

Arid and semi-arid regions are characterised by low mean annual rainfall in relation to mean annual potential evaporation (UNESCO, 1999). Rainfall events tend to be infrequent and their occurrence and severity are highly variable. Typically dry antecedent conditions may result in many rainfall events not producing any significant runoff. However, severe rainfall events can still result in significant flooding with serious consequences for a range of activities. Large transmission losses may also results in discharge reducing in a downstream direction, particularly in the lower river reaches of larger catchments in arid regions (Costello et al., 2003). The special flooding characteristics of catchments in arid and semi-arid regions make it desirable to treat them separately form catchments in more humid regions.

Design flood estimation in arid and semi-arid region is a difficult task due to the episodic nature of flood events and the limitations of recorded streamflow data of acceptable quality. In ARR87, only few catchments were used from arid and semi-arid regions to develop RFFA methods, which had a lower degree of accuracy and limited applicability.

Since the publication of ARR87 (I. E. Aust., 1987), there has been little improvement in terms of streamflow data availability in most of the arid and semi-arid regions of Australia. In the preparation of regional flood estimation database for Australia as a part of Project 5 'Regional flood methods in Australia', only a handful of catchments from the arid and semi-arid regions satisfy the selection criteria (see in Rahman et al., 2009; Haddad et al., 2010). To increase the number of stations from the arid and semi-arid regions to develop a 'reasonably meaningful' RFFA method, the selection criteria were relaxed i.e. the threshold streamflow record length was reduced from 25 years to 10 years and the limit of catchment size was increased from 1000 km^2 to 6000 km^2. These criteria resulted in the selection of 45 catchments from the arid and semi-arid regions of Australia. Based on this limited data set, approximate RFFA methods are developed for arid and semi-arid regions of Australia, as discussed in this chapter. This method will need further development and testing before this can be recommended for inclusion in the revised ARR.

7.2 Special issues for consideration in the arid and semi-arid regions

In the arid and semi-arid regions, most of the streams are ill-defined and non-perennial. The geomorphology of arid and semi-arid regions is notably different from coastal Australia e.g.:

- the river is shallow and lacks a defined channel;
- transmission loss is very high;
- a tendency of the river to shift its course remarkably during high flow;
- lack of vegetation in the catchment; and
- excessive erosion rate in the catchment and stream.

Another problem in arid regions is the difficulty in measuring the flood flow (Parks and Sutcliffe, 1987) because floods are in general flashy, and the problem of defining the peak level accurately by water level recorder or maximum level gauge is aggravated by siltation of the channel. The difficulty of establishing a reasonable rating, particularly at high flow levels, is made worse by the problems of access for gauging near the peak of a short flood (as arid and semi-arid regions are located far from major Australian cities), the long periods without flow, and the instability of the channel control and cross-section area due to the scouring effects of floods.

Floods in arid regions are generally caused by storms of high intensity and are often of relatively limited extent where rainfall rates exceed the infiltration capacity of at least part of the catchment. Thus, the variability in flood magnitude from year to year and from site to site in arid areas is much greater than non-arid areas. This implies that a longer record length would be needed in arid regions to estimate the flood frequency curve with reasonable accuracy; however, in the arid regions of Australia, streamflow record lengths are even shorter than those of coastal Australia.

Some catchments in the arid and semi-arid regions (e.g. the NT) may have a significant proportion of karst geology which can cause the disappearance and reappearance of the stream channel within a short distance, thus affecting the flood flow. It is often quite difficult to define the catchment boundary, and the contributing catchment area for a particular flood event may be very small compared to the whole of the catchment.

In arid regions, annual maximum flood series generally contain many zero values and hence it is more appropriate to use the partial duration series in flood frequency analysis, which has been adopted in this chapter.

There have been only few previous RFFA studies in arid and semi-arid regions. Farquharson et al. (1992) presented a study which uses 162 catchments from 12 countries of the world including 30 catchments from Queensland. Their catchments were located in the regions with an average annual rainfall of 600 mm or less, and with catchment areas in the range of 1 to 357,000 km^2. The use of very large catchments by Farquharson et al. (1992) limits the application of their method to small and medium sized catchments. It is expected that the growth curves in the arid region will show greater uncertainty and steeper slopes compared to those of the coastal regions.

7.3 Method

The application of QRT and PRT seem to be difficult in arid regions as these techniques require relatively longer periods of streamflow data. It appears that a simpler RFFA method will be more appropriate for the arid regions. Here, an index type approach as suggested by Farquharson et al. (1992) is adopted. In the index flood method, the mean annual flood ($\overline{Q}$) and a dimensionless growth factor is used to estimate Q_T:

$$Q_T = \overline{Q} \times X_T \qquad (7.1)$$

where Q_T is the flood quantile for an ARI of T years, $\overline{Q}$ = the mean annual flood based on the partial duration series, X_T = the regional growth factor. In this study, a prediction equation is developed for $\overline{Q}$ as a function of catchment characteristics and regional growth factors are developed based on the observed partial duration series data. In the application, partial series–based Q_T estimates may need to be converted to annual maximum flood series estimates using the Langbein transformation.

7.4 Data

The UNESCO (1999) has suggested the ratio of precipitation (*P*) to potential evapotranspiration (*ET*) as an aridity index:

P/PET < 0.03	hyper-arid zone
0.03 < *P/PET* < 0.2	arid zone
0.2 < *P/PET* < 0.5	semi-arid zone

A total of 45 catchments were selected from arid and semi-arid regions of Australia, with locations shown in Figure 7.1. Based on the above criteria by UNSECO (1999), 90% of the selected catchments are classified as semi-arid ($0.2 < P/PET < 0.5$) and the remaining 10% as arid ($0.03 < P/PET < 0.2$).

The selected catchments have average annual rainfall in the range of 209 mm to 454 mm (Table 7.1). The catchment areas range from 3.8 to 5,975 km^2 (mean: 1152 km^2 and median: 360 km^2). Streamflow record lengths range from 10 to 46 years (mean: 25 years and median: 22 years); these are the partial series maximum flood data, with one event per year on average being selected.

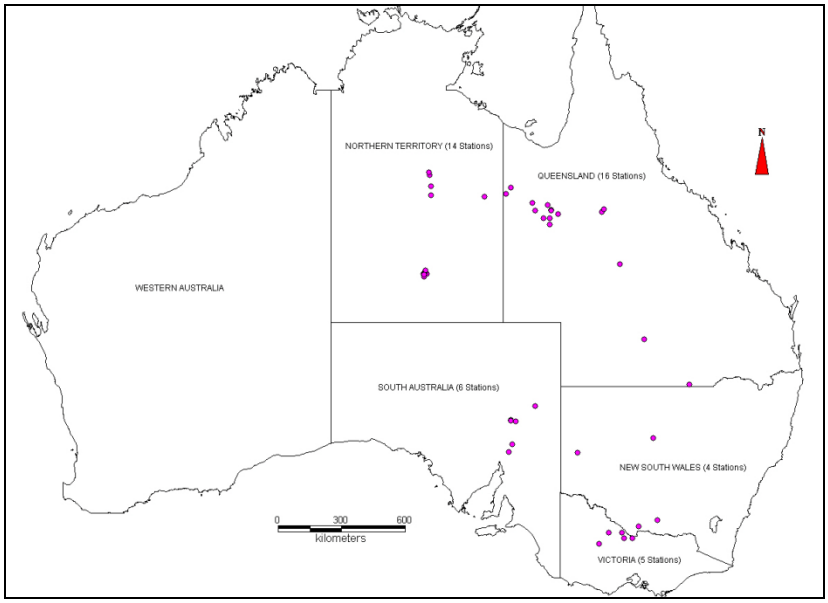


Figure 7.1 Locations of the selected catchments from the arid and semi-arid regions of Australia

Table 7.1 Selected stations for arid and semi-arid regions of Australia

State	Number of stations selected	Range of catchment areas (km^2)	Range of streamflow record length (partial series) (years)	Range of mean annual rainfall (mm)
NSW	4	15 to 4660 (Median: 175)	16 to 35	243 to 434
Vic	5	32 to 4740 (Median: 1629)	16 to 41	407 to 453
Qld	16	6 to 5975 (Median: 757)	17 to 38	393 to 454
SA	6	170 to 448 (Median: 275)	10 to 36	209 to 302
NT	14	3.8 to 4360 (Median: 141)	10 to 46	290 to 429

7.5 Results

7.5.1 Selection of regions

The formation of sub-regions in the arid and semi-arid region is a difficult task, as there are only 45 catchments available from a vast region of interior Australia. There are a few alternatives: (i) all the 45 stations to form one region, which however appears to be unreasonable given the areal extent of the region (over 5000 km × 5000 km), there are likely to be different hydro-geo-meteorological processes over this vast region that affect the flood generation process; (ii) formation of small sub-regions based on geographical proximity; however, too small a region makes the developed RFFA methods of little statistical significance.

To assess the hydrological similarity of various candidate regions, the test by Hosking and Wallis (1993) was applied. This uses the *H* statistic to test for the degree of heterogeneity in a proposed region, where $H < 1$ indicates an 'acceptably homogeneous region', $1 \leq H < 2$ indicates a 'possibly heteorogeneous' region and $H \geq 2$ a 'definitely heterogeneous' region. The results of this test are summarised in Table 7.2, which shows that none of the proposed regions are 'acceptably homogeneous'. It should be noted here that Australian catchments generally exhibit a high level of heterogeneity and application of the test by Hosking and Wallis (1993) did not generate acceptable homogeneous regions in Australia (e.g. Bates et al., 1993 and Rahman et al., 1999).

Since there is limited data in the arid and semi-arid regions in Australia, formation of simple geographic regions appears to be a reasonable option. As NSW and Vic have only 4 and 5 catchments they are combined together to achieve a sample size of 9. The selected catchments of these two states fall in the same drainage division (Drainage Division IV - Murray-Darling Division), and hence their combination is not unreasonable. The states of Qld, NT and SA were treated as separate regions.

Table 7.2 Heterogeneity statistics for candidate regions in the arid and semi-arid regions

State	Nos. of station	H_1	H_2	H_3
All the 45 catchments	45	24.98	12.45	4.88
NSW & VIC	9	18.68	9.24	5.21
Queensland	16	7.41	3.11	2.40
South Australia	6	3.51	2.58	1.12
Northern Territory	14	11.53	5.93	2.03

7.5.2 Derivation of growth curves

The flood quantiles were estimated for T = 2, 5, 10, 20, 50 and 100 years at each station by fitting a Genralised Pareto (GPA) distribution using L moments (Madsen et al., 1997). The $Q_T/\overline{Q}$ values were estimated at individual stations; the weighted average of these values (weighting was done based on record length at individual sites) over all the stations in a region defines the growth curve, as shown in Figure 7.2. A smooth curve was drawn to represent the average growth curve for each of the four regions; the corresponding equations (Table 7.3) show a R^2 value in the range of 0.95 to 0.99 indicating quite a good fit. The growth factors for the selected ARIs, estimated from the fitted equations, are provided in Table 7.3

The approximate 95% confidence interval of the derived growth curves for the four regions are shown in Figure 7.3. From Figures 7.2 and 7.3, it can be found that SA has the steepest growth curve followed by the NT. The Qld has the flattest growth curve. In the larger ARI range, the growth curve for SA is much higher than the other regions (e.g. for 100 years ARI, SA growth factor is approximately double to that of Qld). At higher ARIs, NT and NSW &Vic growth curves are very similar. The combination of growth curves for the four selected regions seems to be unviable, in particular, when the differences are so high between the SA and Qld growth curves. The growth curves derived here generally agree with the world

growth curves (except for SA at higher ARIs) derived by Farquharson et al. (1992) as shown in Table 7.3.

The growth curve for SA catchments found in this study is considerably steeper (particularly for higher ARIs) than the ones implied by Table 1.8 in ARR87 (page 13, Book IV) (but smaller than the ones recommended in ARR87 for the WA Wheatbelt region (page 18, Book IV)). Table 1.8 in ARR87 was based on annual maximum flood series data from three catchments from the Alice Spring area and a 1 year ARI.

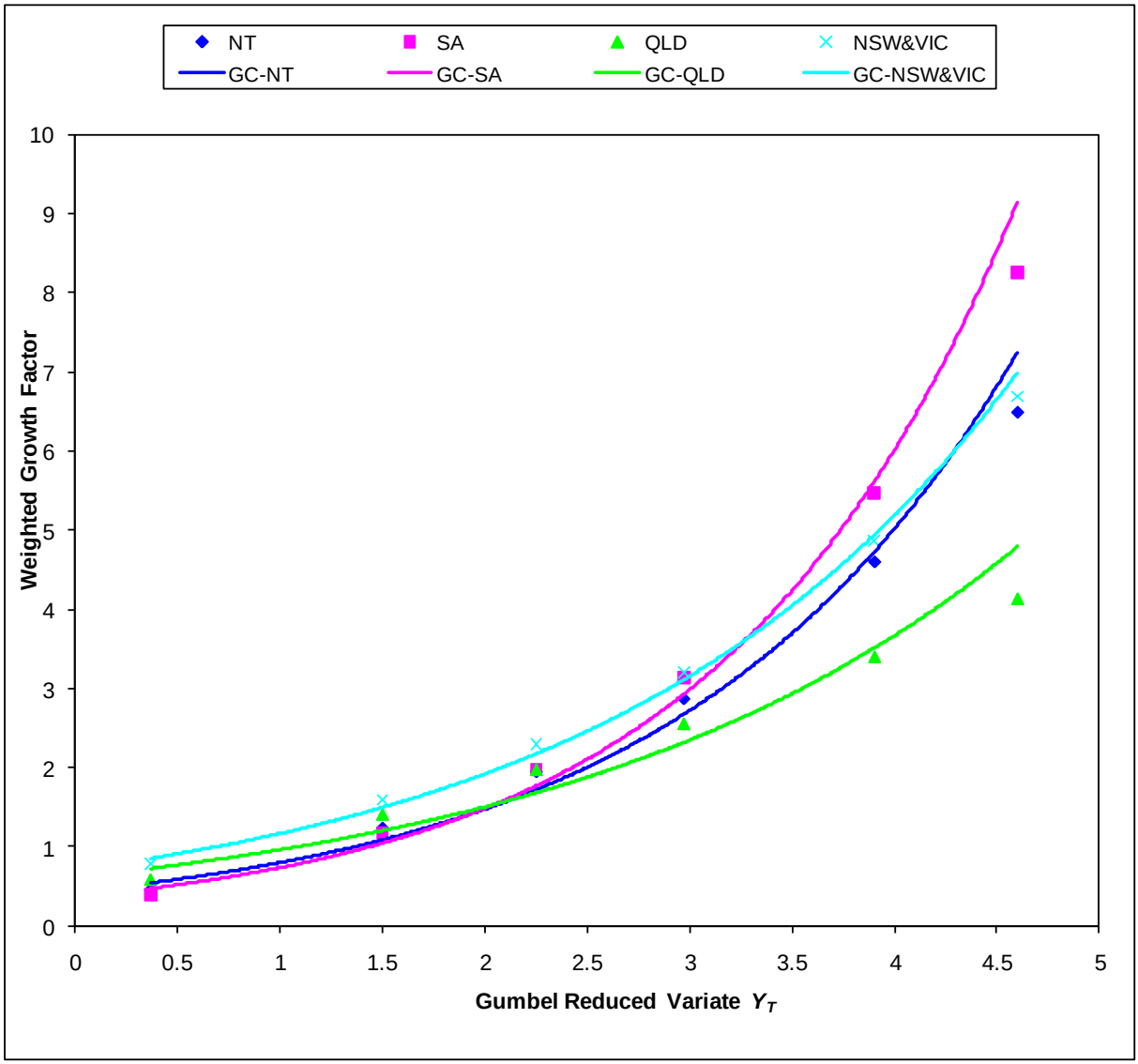


Figure 7.2 Growth curves (GC) for four selected arid and semi-arid regions in Australia

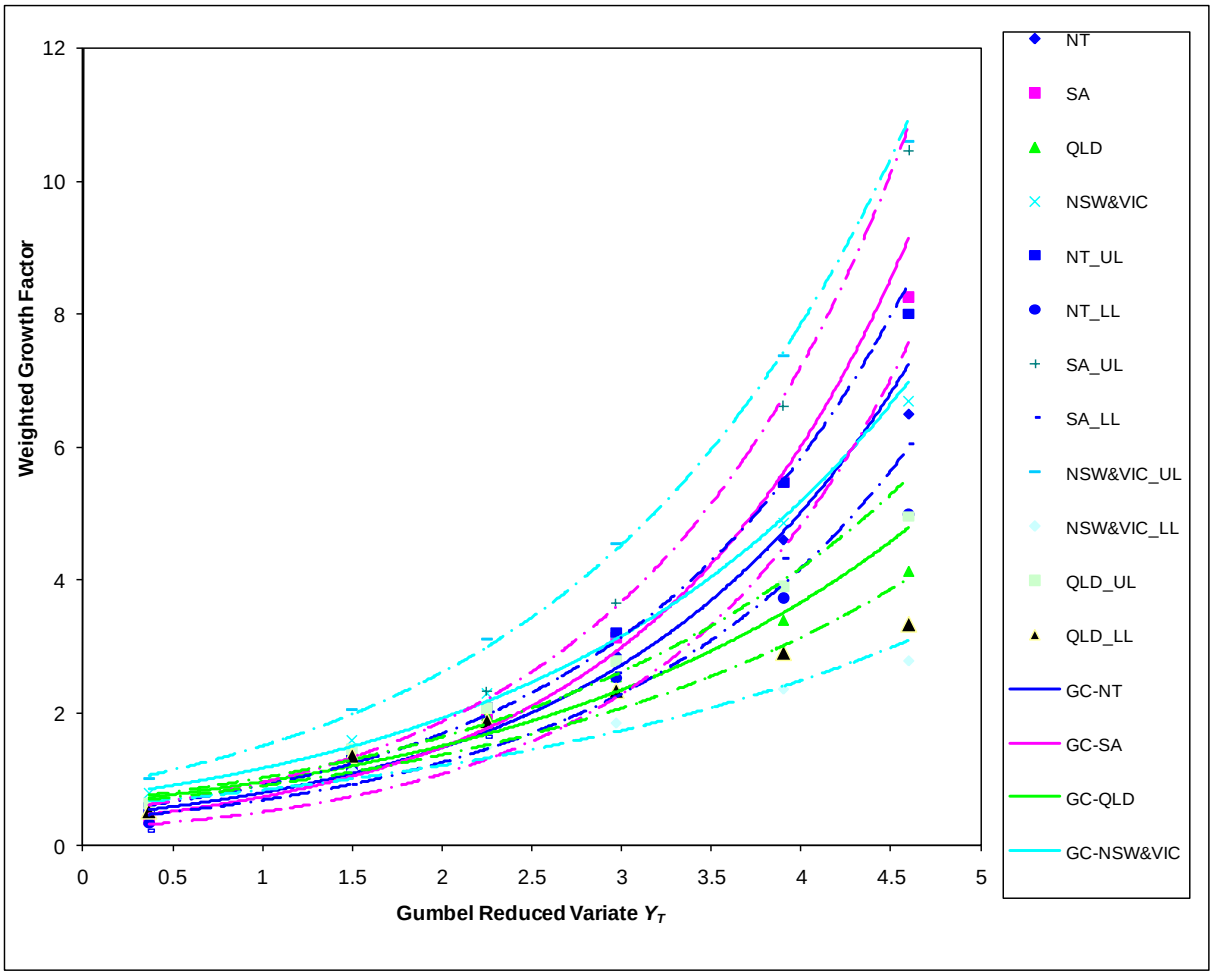


Figure 7.3 Growth curves with 95% confidence intervals (UL- upper limit, LL- lower limit) for four selected arid and semi-arid regions in Australia

Table 7.3 Growth factors for arid and semi-arid regions in Australia (Y_T = Gumbel reduced variate = - ln(- ln(1 - 1/T)) (The growth factors in the table are obtained from the prediction equations shown in the second row of the table)

Region		NSW-Vic	Qld	SA	NT	World data (Farquharson et al., 1992)
Equation		$GF = 0.71e^{0.5Y_T}$ $R^2 = 0.99$	$GF = 0.62e^{0.45Y_T}$ $R^2 = 0.95$	$GF = 0.37e^{0.7Y_T}$ $R^2 = 0.99$	$GF = 0.43e^{0.61Y_T}$ $R^2 = 0.99$	
T (years)	Y_T					
2	0.37	0.85	0.73	0.48	0.54	0.6
5	1.50	1.50	1.22	1.06	1.07	1.5
10	2.25	2.19	1.71	1.79	1.70	2.1
20	2.97	3.13	2.36	2.96	2.63	3.0
50	3.90	5.00	3.59	5.68	4.65	4.2
100	4.60	7.08	4.91	9.26	7.11	5.9

7.5.3 Development of prediction equations for mean annual flood

The statistical package SPSS was used to develop prediction equations for the mean annual flood using an ordinary least squares (OLS) regression, where many different combinations of variables were examined to come up with the best possible model i.e. the one which had the highest coefficient of determination (R^2) and the smallest number of predictor variables. The developed prediction equations are shown in Table 7.4.

The developed prediction equations contain two predictor variables, catchment area (*area*) and 12 hours 2 year rainfall intensity (I_{12_2}), which are relatively easy to obtain. The values of the Durbin-Watson statistic range from 1.84 to 2.34, which are not far away from 2 (a value of 2 indicates no correlation between predictor variables). For all the four regions except for SA, the regression coefficients associated with the predictor variables area and I_{12_2} are positive which indicate that mean flood increases with increasing area and rainfall intensity, which is as expected. However, for SA, the regression coefficient for I_{12_2} is negative, which appears to be counter-intutive, i.e. how can the mean flood decrease with increasing rainfall intensity. To investigate this further, plots of the mean flood vs. area and I_{12_2} are prepared (Figures 7.4 and 7.5), which show that the mean flood increases with area as expected but mean flood decreases with increasing I_{12_2}. This unexpected behaviour of I_{12_2} may be due to very small sample size and due to possible problems in the design rainfall data, which might have been derived in ARR87 using a very limited data set, thus having a high degree of uncertainty associated with these. Here, I_{12_2} was finally ignored for SA and only area was considered in the prediction equation, which showed a smaller R^2 value than the equation with area and I_{12_2}.

The R^2 value for Qld is 0.88, which represents quite a good fit. However, the R^2 values for the other three regions are moderate (in the range of 0.29 to 0.49). The R^2 values obtained here are comparable to the values obtained by Farquharson et al. (1992) (their average R^2 value was 0.57). The quantile-quantile plots (QQ-plot) of the standardised residuals indicate that the residuals are near-normally distributed and there is no outlier data point for the Qld (Figure 7.6) and the NT (Figure 7.8) data sets. The plots of standardised residuals vs. predicted flood quantiles (examples in Figures 7.7 and 7.9 for Qld and the NT, respectively) show that the residuals are within $\pm$ 2.0 $\times$ standard deviation, which indicate the absence of any true outlier data point. The plots for other states were not examined due to very small sample sizes.

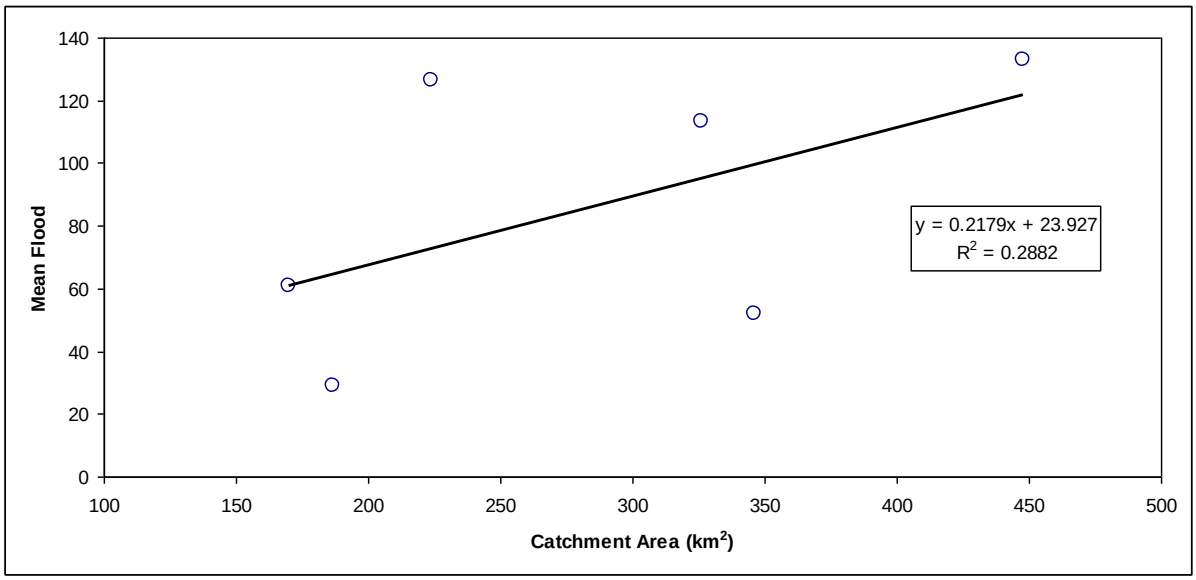


Figure 7.4 Plots of mean flood ($\overline{Q}$) vs. catchment area for South Australia

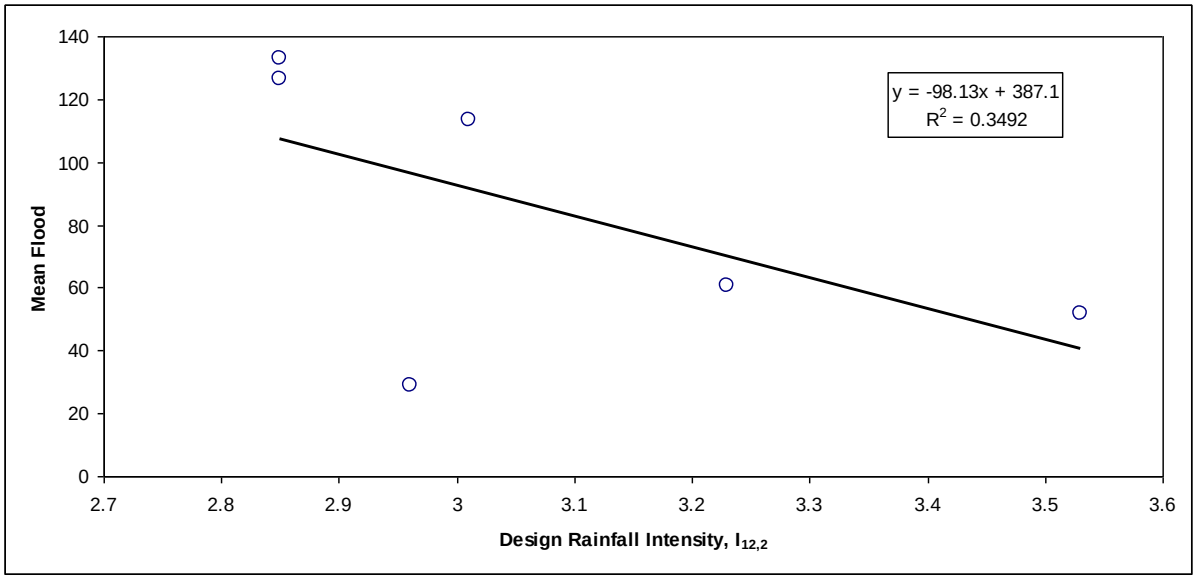


Figure 7.5 Plots of mean flood ($\overline{Q}$) vs. design rainfall intensity $I_{12,2}$ for South Australia

Table 7.4 Summary of prediction equations for mean annual flood for arid and semi-arid regions

Region	Prediction equation ($\overline{Q}$ = mean annual flood of partial duration series data in m^3/s, A = catchment area in km^2 and $I_{12,2}$ = design rainfall intensity of 2 years ARI and 12 hours duration in mm/h, obtained at catchment outlet)	R^2	Durbin-Watson statistic
NSW-Vic (9 stations)	$\log_{10}(\overline{Q}) = -2.43 + 0.42\log_{10}(A) + 5.03\log_{10}(I_{12\_2})$	0.35	2.06
Qld (16 stations)	$\log_{10}(\overline{Q}) = -2.71 + 0.76\log_{10}(A) + 4.01\log_{10}(I_{12\_2})$	0.88	2.34
SA (6 stations)	$\log_{10}(\overline{Q}) = 1.44 + 0.84\log_{10}(A) - 3.29\log_{10}(I_{12\_2})$	0.49	2.12
	$\log_{10}(\overline{Q}) = -0.19 + 0.85\log_{10}(A)$	0.29	-
NT (14 stations)	$\log_{10}(\overline{Q}) = -0.50 + 0.60\log_{10}(A) + 1.65\log_{10}(I_{12\_2})$	0.45	1.84

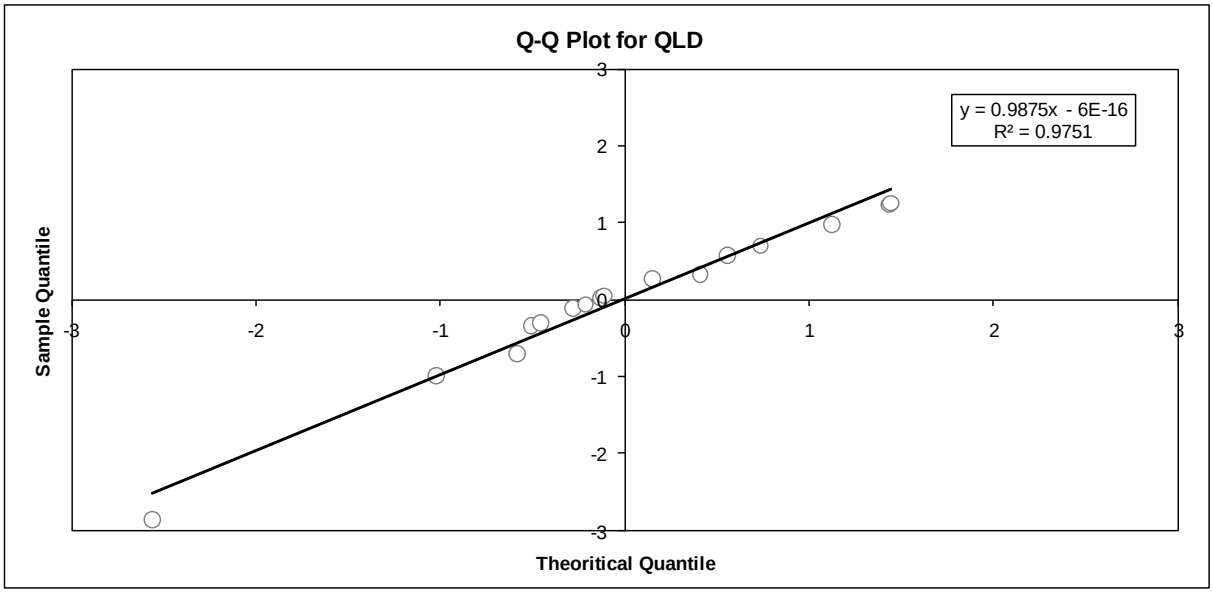


Figure 7.6 QQ-plot of the standardised residuals for Q_{20} (Qld)

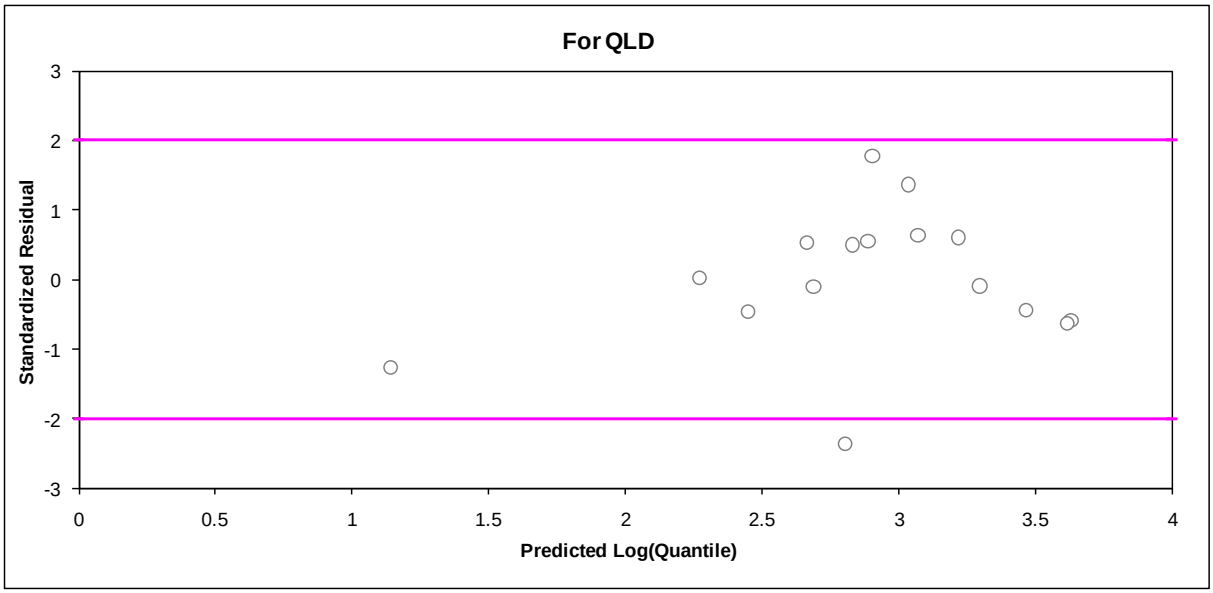


Figure 7.7 Standardised residuals vs. predicted quantiles for Q_{20} (the red marks show the bound of $\pm$ 2.0 $\times$ standard deviation) (Qld)

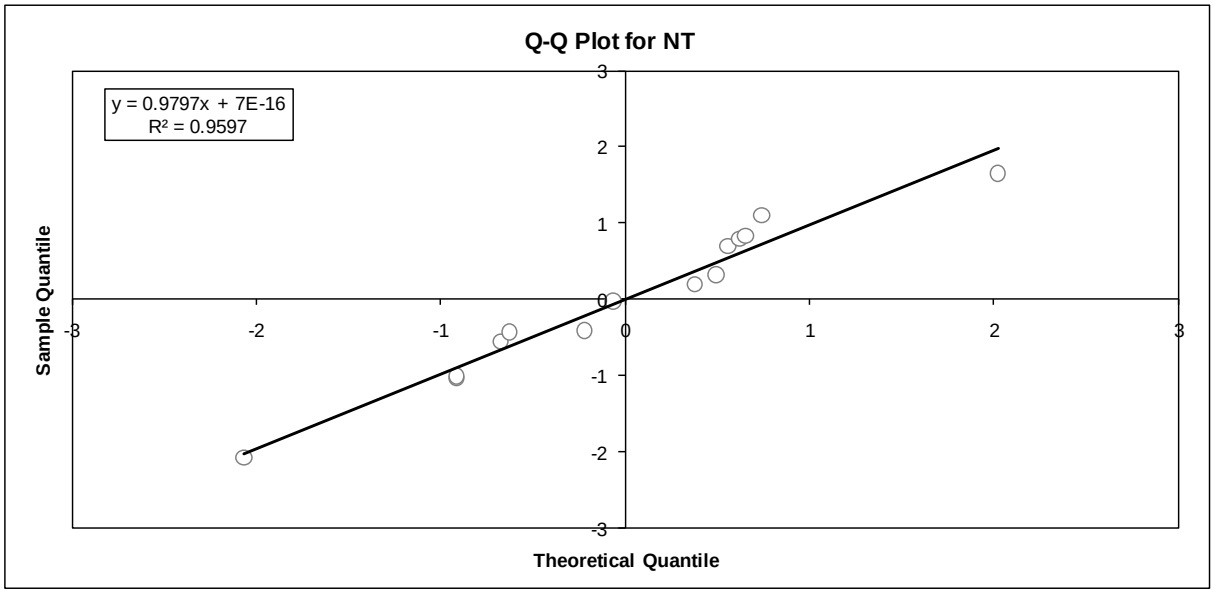


Figure 7.8 *QQ*-plot of the standardised residuals for Q_{20} (NT)

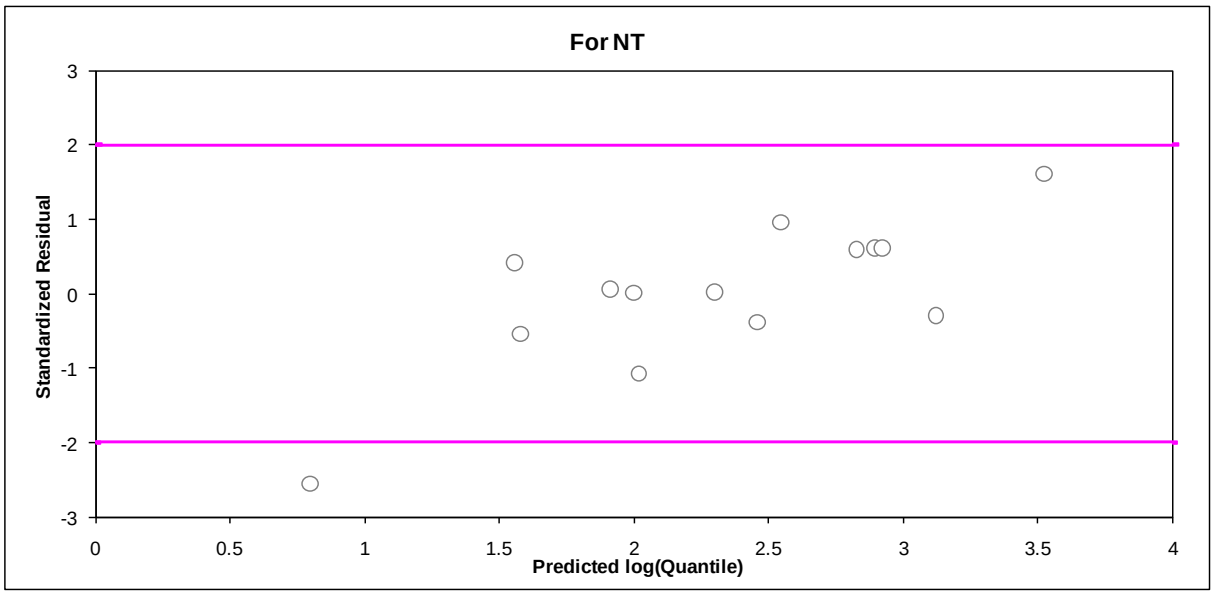


Figure 7.9 Standardised residuals vs. predicted quantiles for Q_{20} (the red marks show the bound of $\pm$ 2.0 $\times$ standard deviation) (NT)

7.5.4 Validation and testing

The predicted flood quantiles (using Equations in Table 7.4 and growth factors in Table 7.3) for each of the study catchments were compared with the at-site flood frequency estimates (Q_{obs}). The relative error values (Tables 7.5 to 7.8) range from 46% to 71% for NSW-Vic, 22% to 31% for Qld, 12% to 50% for SA and 49% to 59% for the NT, which seem to be quite reasonable. The count of the ratio Q_{pred}/Q_{obs} values are also presented in Tables 7.5 to 7.8. The ratio values in the 'desirable' range of 0.5 to 2 are 50%, 85%, 92% and 51% for NSW-Vic, Qld, SA & NT respectively; the ratio values for Qld and SA appear to be quite satisfactory. It should be noted here that due to small sample size, the error statistics discussed here may have little statistical validity. The observed and predicted flood quantiles for Q_{20} in Queensland match very well (Figures 7.10 and 7.11).

Table 7.5 Summary of model testing (NSW-Vic)

ARI (years)	Median RE (%)	Count Q_{pred}/Q_{obs} (ratio)		
		ratio < 0.5	0.5 ≤ ratio ≤ 2.0	Ratio > 2.0
2	47	1	6	2
5	46	1	5	3
10	55	2	4	3
20	71	3	3	3
50	66	2	4	3
100	67	2	4	3
Total count		10	27	17
% of count		19%	50%	31%

Table 7.6 Summary of model testing (Queensland)

ARI (years)	Median RE (%)	Count Q_{pred}/Q_{obs} (ratio)		
		ratio < 0.5	0.5 ≤ ratio ≤ 2.0	Ratio > 2.0
2	31	4	10	2
5	26	0	14	2
10	23	0	15	1
20	22	0	14	2
50	22	1	14	1
100	18	1	14	1
Total count		6	82	8
% of count		7%	85%	8%

Table 7.7 Summary of model testing (South Australia)

ARI (years)	Median RE (%)	Count Q_{pred}/Q_{obs} (ratio)		
		ratio < 0.5	0.5 ≤ ratio ≤ 2.0	Ratio > 2.0
2	50	1	3	2
5	19	0	6	0
10	14	0	6	0
20	21	0	6	0
50	12	0	6	0
100	12	0	6	0
Total count		1	33	2
% of count		3%	92%	5%

Table 7.8 Summary of model testing (Northern Territory)

ARI (years)	Median RE (%)	Count Q_{pred}/Q_{obs} (ratio)		
		ratio < 0.5	0.5 ≤ ratio ≤ 2.0	Ratio > 2.0
2	59	6	4	4
5	49	2	9	3
10	58	2	9	3
20	58	5	6	3
50	54	5	7	2
100	56	5	7	2
Total count		25	43	16
% of count		30%	51%	19%

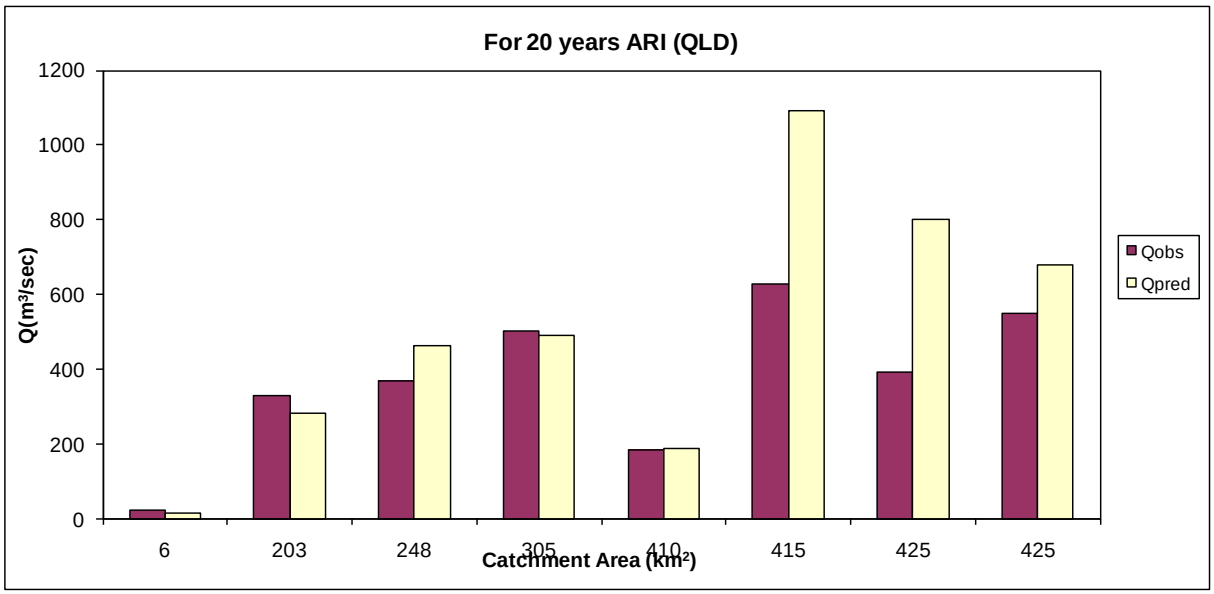


Figure 7.10 Predicted vs. observed floods for test catchments in the arid and semi-arid region in Qld for Q_{20} (catchment areas in the range of 6 km^2 to 425 km^2)

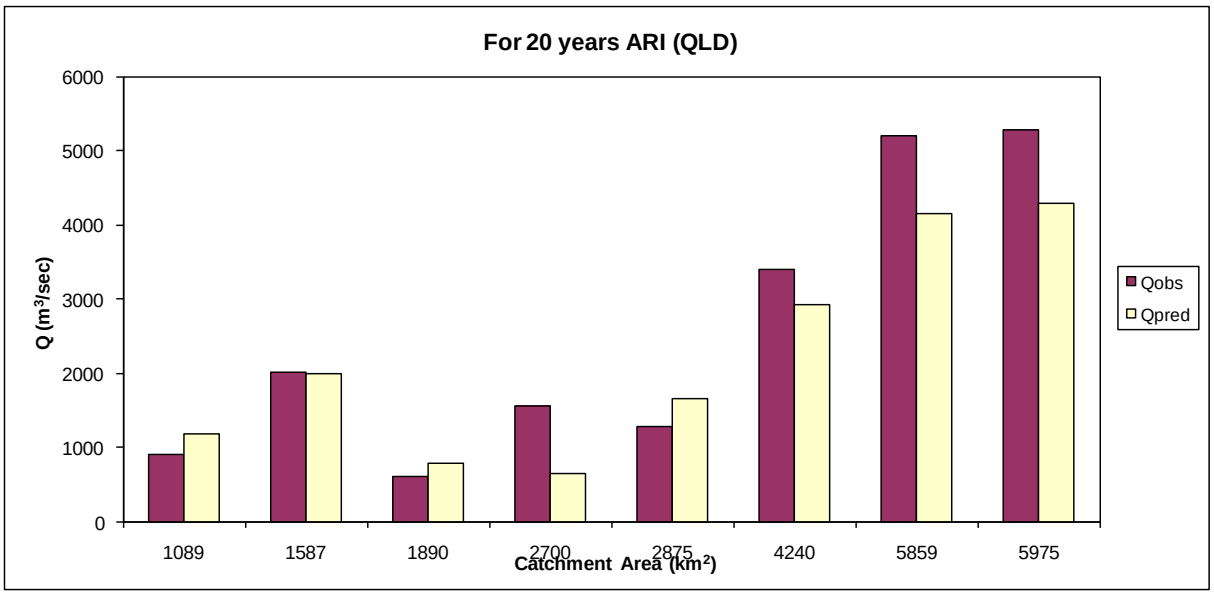


Figure 7.11 Predicted vs. observed floods for test catchments in the arid and semi-arid region in Qld for Q_{20} (catchment areas in the range of 1089 km^2 to 5975 km^2)

7.6 Summary

Catchments in the arid and semi-arid regions of Australia have a distinctly different flood hydrology from catchments in more humid/coastal regions, and they thus warrant separate treatment for regional flood estimation. The limited availability of streamflow data at both temporal and spatial scales in the arid and semi-arid regions of Australia makes it difficult to

develop statistically meaningful RFFA methods. Using the limited data, a simple RFFA method is developed here for the arid and semi-arid regions, which has generally a lower degree of accuracy than the RFFA methods developed in Chapter 4 for the coastal regions of Australia, which are based on a larger data set with good quality. The RFFA methods presented here for the arid and semi-arid regions should be used with caution; local hydrological knowledge must be exercised to interpret the results. It would be necessary to compare the results from these methods with other methods, such as rainfall runoff models and any locally available methods. The RFFA methods presented here will require further development and testing before they can be recommended for inclusion in the revised ARR.

It is recommended that some representative stream gauges be established in the arid and semi-arid regions for 'long term monitoring' to develop a comprehensive database which will assist in upgrading the RFFA methods presented here in the future. A more comprehensive flood data base is required to develop a better understanding of how the special climate, catchment and stream characteristics of the arid and semi-arid regions of Australia interact to produce distinctly different flood responses. It is to be noted here that a high degree of uncertainty associated with a RFFA method results in inaccurate design flood estimates, which increase the capital cost of the infrastructure in the case of over-design, or the average annual flood damage cost in the case of under-design. The cost of streamflow monitoring and data collection is expected to be far less than the cost associated with grossly inaccurate RFFA methods that are developed based on inadequate data.

8. Time trends in Australian flood data

8.1 General

The initial results of trend analyses in Australian annual maximum flood series data had shown a trend for over 30% of the stations investigated (Ishak *et al.,* 2010). Many of these stations had record lengths in the range of 30 to 97 years and were affected by the exceptionally dry weather regime since 1990s, and hence it was not possible to confirm whether the detected trends were due to climate change or due to climate variability. This initial study also disregarded the correlation structure within the annual maximum flood series data in the trend detection analysis. The objective of this section is to provide a brief review on trend analysis in rainfall and flood data, focusing on the most recent literatures and to investigate the impacts of serial and cross-correlation on trend analysis.

8.2 A review of trend analysis for hydrological data

The potential impacts of climate change and natural climate variability on the hydrologic regime have received great attention in contemporary hydrology and water resources management research. Climate change in the context of hydrology can be defined as any change in the hydrologic cycle which is attributable to human activities, most notably those associated with increasing greenhouse gas concentrations in the atmosphere and the corresponding increases in global mean temperature. The effects of anthropogenic emissions of aerosols also fall within this category, although their effects on climate are likely to be more regional and shorter-lived. Climate variability, on the other hand, is generally viewed as resulting from 'natural' sources, and may be due to internal dynamics of the climate system (e.g. ENSO/IPO) or external forcing (e.g. periodic fluctuations in solar radiation, and 'spikes' due to volcanic eruptions).

Assessment of hydrological records collected in different parts of the world has provided evidence of regime-like or quasi-periodic climate behaviour, and of systematic trends in key climate variables due to climate change and/or climate variability (Gallant *et al.,* 2007; Fu *et al.,* 2008; Ma *et al.,* 2008; Zhang and Lu, 2009; Chowdhury and Beecham, 2009; Villarini *et al.,* 2009). In Australia average surface temperature has increased over the past 98 years, where the last two decades have been particularly warm, with the warmest year on record occurring during 2005, as shown in Figure 8.1. Furthermore, the Intergovernmental Panel on

Climate Change (IPCC) fourth assessment report acknowledged that the global surface temperature is expected to continue to warm up over the 21st century, affecting all aspects of the hydrological cycle (IPCC, 2007). The implications for flood hydrology are expected to be significant, with projections of changing mean temperature and rainfall intensities leading to a change in the flood frequency regime.

It has been reported in the literature that the frequency and magnitude of extreme flood events are expected to rise in the near future due to climate change, even in cases where the long-term annual average rainfall is expected to decline (IPCC, 2007). Changing climate will have notable impacts on the rainfall runoff process, and thus the assumption of stationary hydrology (i.e. the idea that the past is the key to the future) will have to be revised if the stationarity assumption is not met. Otherwise, the effectiveness of the return period concept can be undermined, and can lead to underestimation/overestimation of the design flood (Khaliq *et al.,* 2006), which in turn will have important implications on the hydrologic design and operation processes. Recent research carried out in some regions of the world has questioned the validity of the traditional flood risk assumptions of stationarity and homogeneity (Power *et al.,* 1999; Douglas *et al.,* 2000; Strupczewski *et al.,* 2001a; Franks and Kuczera, 2002; Cunderlik and Burn, 2003; Prudhomme *et al.,* 2003; Micevski *et al.,* 2006; Leclerc and Ouarda, 2007; among many others), especially with the recognition that climate naturally varies at all scales. Accordingly, design flood estimation techniques are required to consider the changing flood regimes in the presence of trends in hydrological variables, e.g. by assuming time-varying parameters of the flood frequency distribution (e.g., Strupczewski *et al.,* 2001a, b).

In attempts to address the impacts of climate change on hydrological time series, numerous assessments have been undertaken worldwide to investigate if abnormalities in the form of trends exist in time series of hydrological variables. For instance, in North America, Olsen *et al. (*1999) have reported positive trends in flood risk over time for gauged sites within the Mississippi, Missouri, and Illinois River basins. Douglas *et al*. (2000) discovered no evidence of trends in flood flows but they did find evidence of upward trends in low flows at larger scale in the Midwest and at a smaller scale in Ohio, the north central and the upper Midwest regions. Negative trends in total streamflow were most common for the analysed Pennsylvanian streamflow time series from 1971 to 2001 due to climate variability (Zhu and Day, 2005). Novotny & Stefan (2007) investigated the streamflow records from 36 gauging stations in five major river basins of Minnesota, USA, for trend and correlations using the Mann-Kendall (MK) test and moving averages method. The authors found that trends differed significantly from one river basin to another, and became more prominent for shorter

time windows. Pasquini & Depetris (2007) presented an overview of discharge trends and flow dynamics of South American rivers draining the southern Atlantic seaboard. Juckem *et al. (*2008) found a decrease in annual flood peaks for stream gauging stations in the Driftless Area of Wisconsin.

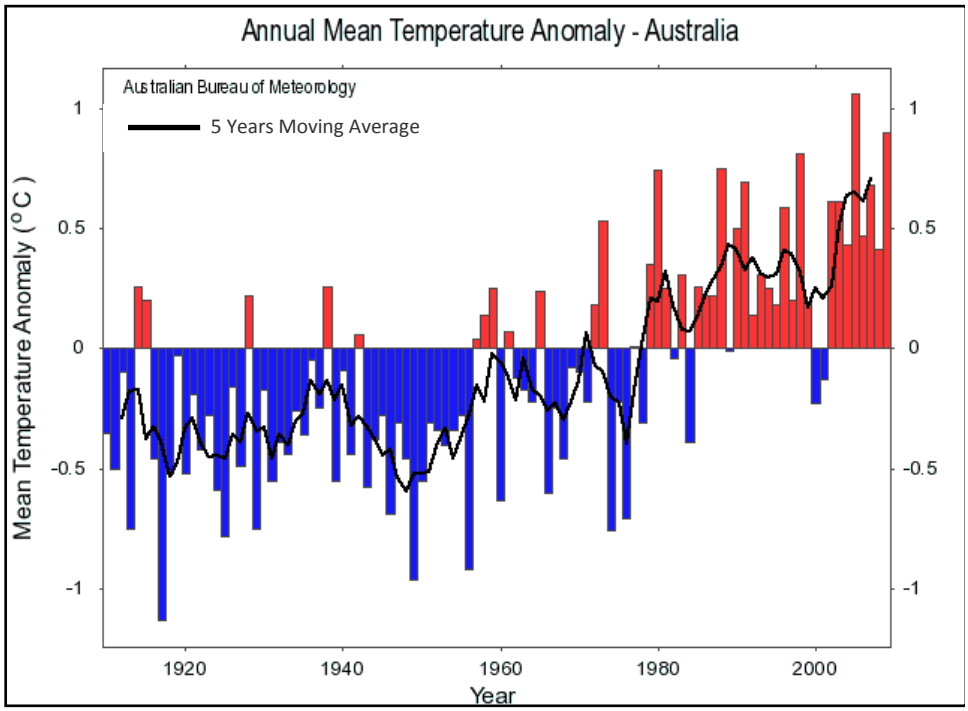


Figure 8.1 Annual mean temperature anomalies for Australia based on 1961-2009 (Source: Australian Bureau of Meteorology dated 10/03/2010)

Similarly, several studies have been undertaken in Canadian basins to assess climate change impacts on hydro-meteorological variables. Zhang *et al. (*2001) analysed hydrological data from a network of 54 hydrometric stations and meteorological data from a network of 10 stations using the Mann-Kendall test associated with the pre-whitening approach. They found that overall Canadian streamflows experienced negative trends for the past 30 to 50 years. The temperature, precipitation and streamflow data for sites in British Columbia and the Yukon were examined by Whitfield (2001). Burn & Hag Elnur (2002) analysed the trends and variability in the hydrological regime for the Mackenzie Basin in northern Canada. The authors identified similarities in trends and patterns in the hydrological and meteorological variables at chosen locations in Canada, implying a relationship between the two groups of variables. The trends in annual streamflow volume in northern British Columbia and the Yukon have been investigated by Fleming and Clarke (2003). Burn *et al.* (2004a, b) assessed the trends in streamflow data in the Liard and Athabasca River Basins in northern Canada. Abdul Aziz & Burn (2006) applied the non-parametric Mann-Kendall test with the

trend-free pre-whitening approach to identify trends in hydrological variables. Winter month flows exhibited strong increasing trends, and an earlier inception of the spring freshet was noted over the basin.

Swedish annual runoff volumes, annual and seasonal flood peak time series have been analysed by Lindstrom & Bergstrom (2004), where they found that the 1970s were very dry and, in a short perspective, both runoff volumes and flood magnitude increased substantially between 1970 and 2002. Birsan *et al. (*2005) have studied the trends in the streamflows in Switzerland using the Mann–Kendall nonparametric test in three study periods. The trends in water levels and streamflow in the Yangtze River basin in China have also been investigated by Zhang *et al.* (2006). The temporal trends of annual and seasonal precipitation and temperature in the Hanjiang basin in China have been analysed by Hua *et al.* (2007) using Mann-Kendall and linear regression techniques. Petrow and Merz (2009) analysed the trends in the flood time series in Germany using the MK test. The analysis detected significant upward trends in flood data for a considerable fraction of basins. Also they found that most changes were detected for sites in the west, south and centre of Germany. Petrow and Merz (2009) concluded that the missing relation between significant changes and basin area suggested that the observed changes in flood behaviour are climate-driven.

In Australia, Chiew & McMahon (1993) examined trends in annual streamflow of 30 unregulated Australian rivers to identify changes in streamflow in relation to the changes in climate. The authors did not find evidence of changes in streamflow resulting from climate change. They also indicated that the detection of statistically significant trends in streamflow is largely affected by inter-annual variability in streamflow and to a lesser degree the length of streamflow record. Hennessy *et al. (*1999), Plummer *et al. (*1999) and Collins *et al. (*2000) reported that Australia's continental average rainfall and temperature have an increasing trend since the beginning of the 20th century, while Smith (2004) and Alexander *et al. (*2007) reported some decreases in the rainfall in the southeast and along the east coast of the country after 1950. Similarly, Murphy and Timbal (2008) found that the South-eastern Australia region has been experiencing an annual rainfall downward trend at the rate of 20.6 mm per decade since 1950.

Taschetto and England (2008) investigated the post 1970 Australian rainfall trends, and they found an increasing trend to the west (except coastlines) and a decreasing trend on the northeast coast. This is consistent with the trend in annual total rainfall maps issued by the Australian Bureau of Meteorology, as shown in Figure 8.2. In general, the spatial pattern of the trends in annual precipitation can be separated into two main regions: to the west where the rain has been increasing, and the east where precipitation has been decreasing,

especially during the last 30 years. Chowdhury and Beecham (2009) investigated the monthly rainfall trends and their relation to the southern oscillation index (SOI) at ten rainfall stations across Australia covering all the state capital cities. The outcomes of their assessment revealed decreasing trends of rainfall depth at two stations (Perth airport and Sydney Observatory Hill); no significant trends were found in the Melbourne, Alice Springs and Townsville rainfall data, while the remaining five stations showed increasing trends of monthly rainfall depth. Furthermore, they found that SOI accounted for the increasing trends for the Adelaide and Cairns rainfall data and the decreasing trends for Sydney rainfall. On a short time scale, Haddad *et al. (*2008) reported a decreasing trend in Victorian observed annual maximum flood series data (for a quite large number of stations), particularly after the 1990s. Ishak *et al.* (2010), based on a preliminary study, found that 30% of the study stations exhibited local significant trend in annual maximum flood series data.

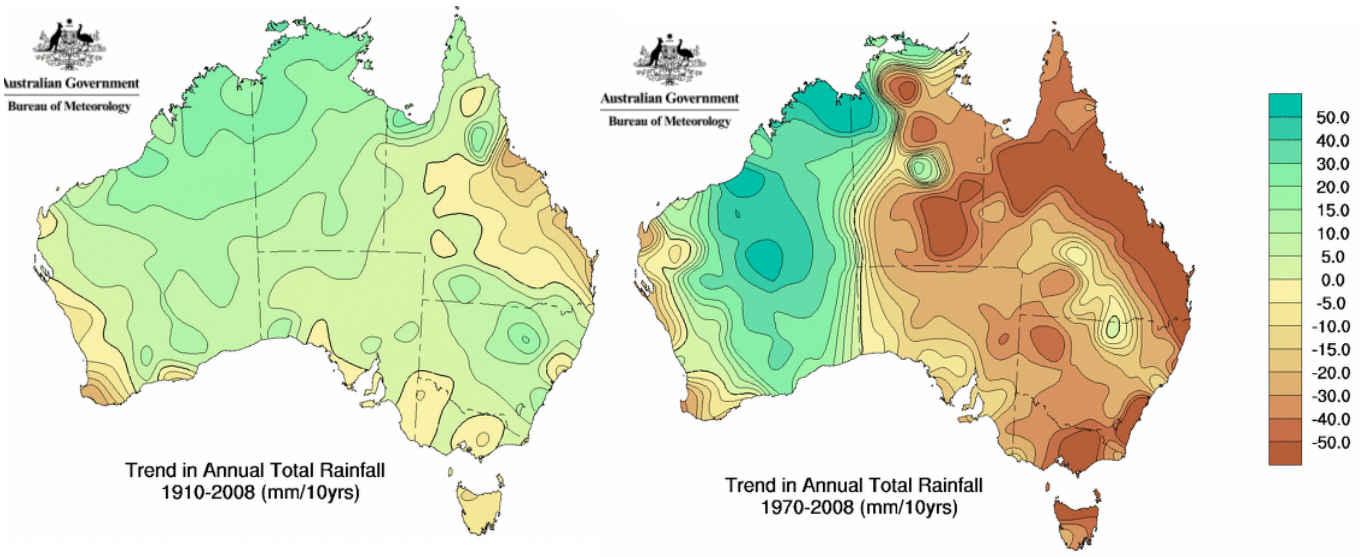


Figure 8.2 Rainfall trends in Australia for (a) 1910 to 2008 and (b) 1970-2008. Trends are shown in mm per decade. (Source: Australian Bureau of Meteorology dated 05/01/2009)

The review of literatures on the identification of trends in hydrological data reveals that the non-parametric Mann-Kendall (MK) and Spearman Rho (SR) associated with Sen's robust slope estimator tests have been favoured for identifying temporal changes in observational records. The non-parametric tests are more robust with respect to non-normality, nonlinearity, missing values, serial dependence, sensitivity to outliers (extremes), and seasonality (Yue *et al.,* 2002). Furthermore, the majority of trend investigations in Australia were concentrated on the evaluation of the trends in the Australian rainfall time series, with

limited investigation on the impact of climate changes on flood risk appraisal. Thus the emphasis in this chapter is on the assessment of trends in the annual maximum flood series for the Australian continent with and without the consideration of serial and cross-correlation. The spatial distribution of catchments exhibiting trends and not exhibiting trends is also investigated. The relationships between trends in annual maximum flood series and the catchments attributes are also investigated.

8.3 Adopted Methodology

The first step in the trend analysis is to investigate trends in the hydrological data at local scales. Thus, the methodology used for exploring the trends and variability started with the preliminary evaluation of trends in hydrological variables for individual stations, with the assumption of serial independence of recorded observations, using the Mann-Kendall and the Spearman Rho nonparametric trend tests without the consideration of the correlation structure. Both tests are rank-based methods, where the MK test assesses whether a random response monotonically increases or decreases, and the SR test examines whether the correlation between time steps and streamflow observations is significant. The results of the trend analysis can be used to establish whether the observed streamflow time series from the selected sites exhibit trends for a number of sites that is greater than the number that is expected to occur by chance. All the trend outcomes have been appraised using a local significance level of 10% to ensure an effective exploration of the trend characteristics in the study area.

A vital part of the trend identification procedure is to consider the correlation structure of the time series under assessment. The correlation structure consists of the serial correlation of the data series, and the cross-correlation between hydrological variables at different locations (Khaliq *et al.,* 2009). The existence of positive serial correlation within a time series increases the possibility of the null hypothesis of no trend being rejected while the null hypothesis is actually true (von Storch, 1995). Similarly, the occurrence of positive cross-correlation within a hydrological homogeneous region or within a stream gauging network will enhance the possibility of the null hypothesis of no field significance of identified trends being rejected (Douglas *et al.*, 2000). The field significance analysis of identified trends helps to establish whether the stations recognized with significant trends at local scales are real or just coincidental because of cross-correlation among the set of stations studied. Hence, failure to take this into consideration in the trend detection process could result in erroneous conclusions.

The impact of serial correlation on the trend detection analysis has been addressed by applying the trend-free-pre-whitening (TFPW) approach proposed by Yue *et al.* (2002) to the hydrological variables. The TFPW process involves an estimation of the slope of the trend by using Sen's robust slope estimator method, then the series is detrended by assuming a linear trend and the lag-1 serial correlation coefficient is evaluated from the detrended series. If the lag-1 serial correlation coefficient is non-significant at the 95% significance level, then the MK test is applied to the original time series, as it is considered to be serially independent; otherwise the trend identification test is applied to the detrended pre-whitened series recombined with the previously estimated slope of trend. Note that while the TFPW process requires fitting and removing a linear trend, the overall MK trend analysis does not make any assumptions about the nature of the trend in the data set.

In a similar manner, the presence of positive cross-correlation among a stream gauging network will inflate the rate of rejecting the null hypothesis of no field significance of trends while it is true (Douglas *et al.*, 2000). Therefore, in this study, the cross-correlation was incorporated by evaluating the field significance of the trend results using the group block bootstrap resampling approach from Yue *et al.* (2003) by preserving the cross-correlation within the stations network. The resampling approach determines the critical value for the percentage of stations exhibiting an upward or a downward trend separately by chance. Based on this critical value, it is possible to determine whether the observed number of upward or downward trends within the stream gauging network exceeds what is expected to occur by chance. For illustration, at the field significance level of 0.1, if the observed number of sites with significant upward trend is greater than or equal to the 90^{th} quantile value of the simulated distribution of the number of sites with significant upward trend developed by the bootstrap approach, then the observed number of sites with significant upward trends over the network is judged to be field-significant at 0.1 significance level. Similarly, the observed number of sites with significant downward trend was assessed. Results obtained from the trend test were analysed using a local significance level of 10% and a field significance level of 10%.

8.4 Study period and database

The preliminary trend analysis using MK and SP tests was performed on a study period corresponding to all the available records at each station to allow for an optimal spatial coverage. This initial assessment provides general guidance about the trend behaviour for individual stations, although the periods of records reflected at each station potentially fluctuate, making interpretation of the results more difficult. Further, the methodology described for the consideration of the correlation structure was carried out on three study periods starting in 1955, 1965, and 1975 and ending in 2004. The different fixed study periods selected correspond to a trade-off between the temporal and spatial coverage offered by the selected data set. The selection of a common period of record in this way facilitates investigation of variable climate conditions during the common prescribed period. It should be prominent that for a station to be included in any of the three given study periods, it should have a continuous record during the study period, whereas two years of no data were allowable during the preliminary analysis.

Australian annual maximum flood series (AMS) data collected from river monitoring stations throughout Australia were used for the trend identification analysis. The data base consists of 491 streamflow stations selected with a minimum record length of 30 years, where the average record length is 38 years and the longest one is 97 years, to ensure statistical validity of the trend results. Although there are a total of 491 stations with streamflow data, the selected study periods determined which of the stations were available for the investigation. On average, the numbers of stations available were 330 for the 30-year period, 77 for the 40-year period, and 21 for the 50-year period. The selected stations have catchments with only minor anthropogenic influence and high quality measurements. Catchment sizes range from 1.3 km^2 to 4,360 km^2, with a median value of about 280 km^2. About 3% of the catchments are greater than 1,000 km^2 in size; about 22% are less than 100 km^2. Most of the selected stations were spatially located near the coast line, as shown in Figure 8.3.

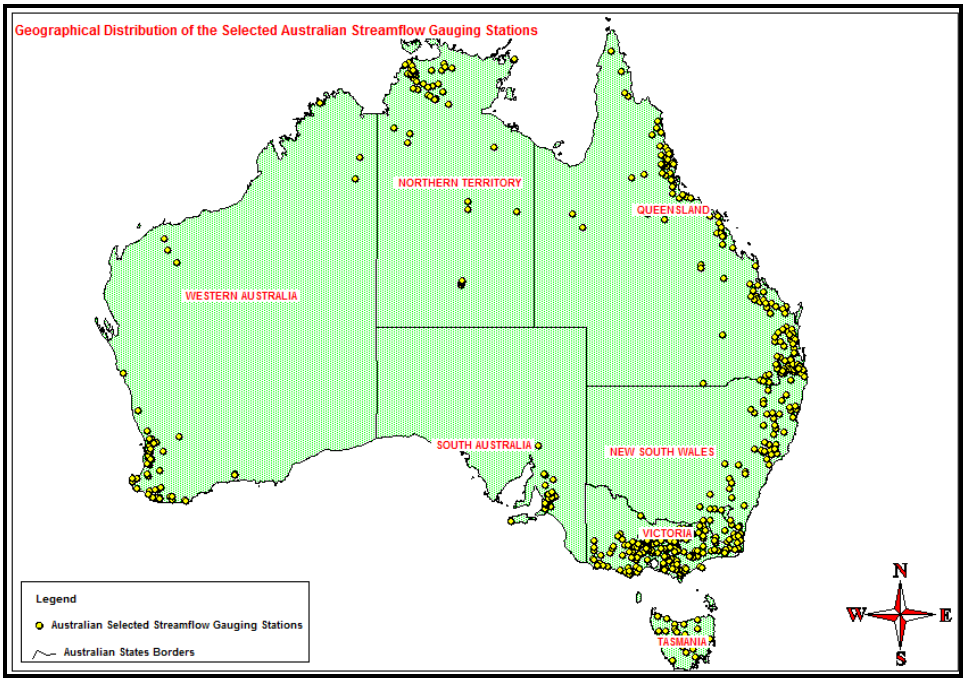


Figure 8.3 Geographical distributions of the selected catchments

8.5 Results of trend analysis

This section presents the preliminary results from the MK and SR tests for the study period corresponding to all the available records at each station, without consideration of the correlation structure. The outcomes of statistically significant trends observed at the 90% local confidence level are summarised in Table 8.1. Presented is the percentage of stations with significant trend for the annual maximum (AM) flood series; results are given separately for upward and downward trends. Among the 491 stations, the total numbers of stations exhibiting either a positive or negative trend by the MK and SR tests are 156 and 172 stations respectively. Prominently, the numbers of trends for the AM flood data far exceed the critical level for establishing on-site significance. As a consequence, the preliminary conclusion is that the Australian AM flood series are exhibiting substantially more trends (32% and 35% from MK and SR tests respectively) than would be expected to occur by chance (10%). Further, Table 8.1 displays that the direction of the trends is, in general, downward, as recognized by the two trend tests. However, the shortness in record length for the majority of the selected stations (average record length of 38 years) might have an impact on these results, particularly the dry period experienced in the last decade in the south-eastern and south-western parts of the continent.

Table 8.1 Trend analysis results and percentage of stations with a significant trend

Trend Tests	Number of Stations	Number of stations with Decreasing Trends	Number of stations with Increasing Trends	Percent Significant Trends (%)
Mann-Kendall	491	127	29	32
Spearman's Rho		140	32	35

Additionally, the spatial distribution of trends in the annual maximum flood series for the preliminary study period from MK and SR tests for the Australian region is visually displayed in Figures 8.4 and 8.5 respectively. On these maps, a yellow circle, a blue circle and a red circle represent a station with no significant trend, significant upward trend, and significant downward trend, respectively. Notable from these figures are the basins located in south-eastern Australia and in the south-west of Western Australia region that exhibit downward trends only, suggesting a decrease in the AM flood series with time within these regions. Contrary are the basins located in the north-western part of the continent, which display upward trends, suggesting an increase in the AM flood series (with time) for these basins, while combined decreasing and increasing trend patterns were detected in the north-eastern region mostly in Queensland. It is noteworthy that this preliminary trend analysis allows for an optimal spatial coverage, although the period of records reflected at each station fluctuates, making interpretation of the results quite difficult.

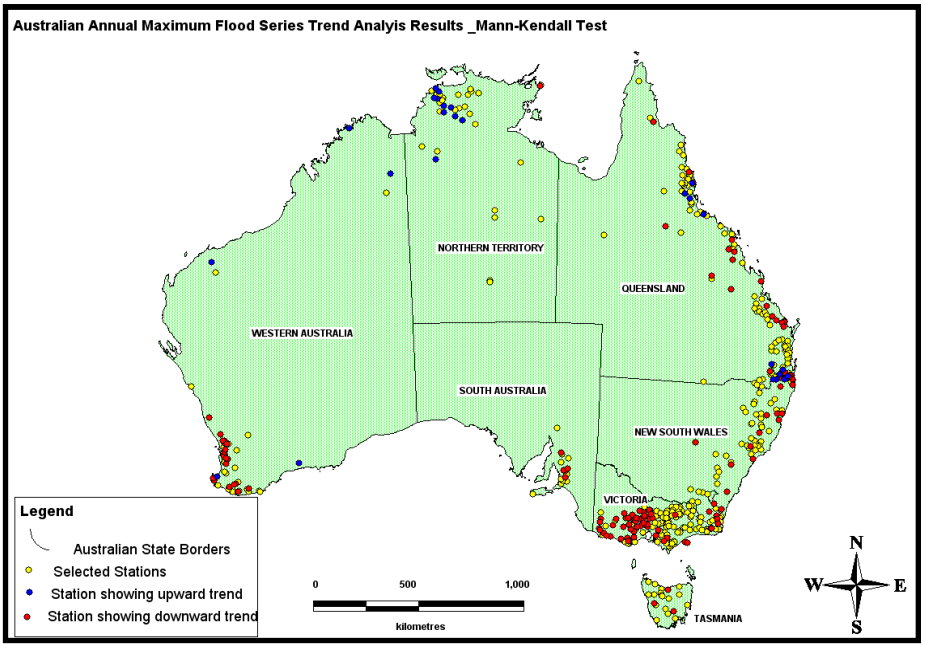


Figure 8.4 Results of trend analysis based on Mann-Kendall test. Red and blue circles represent downward and upward trends respectively

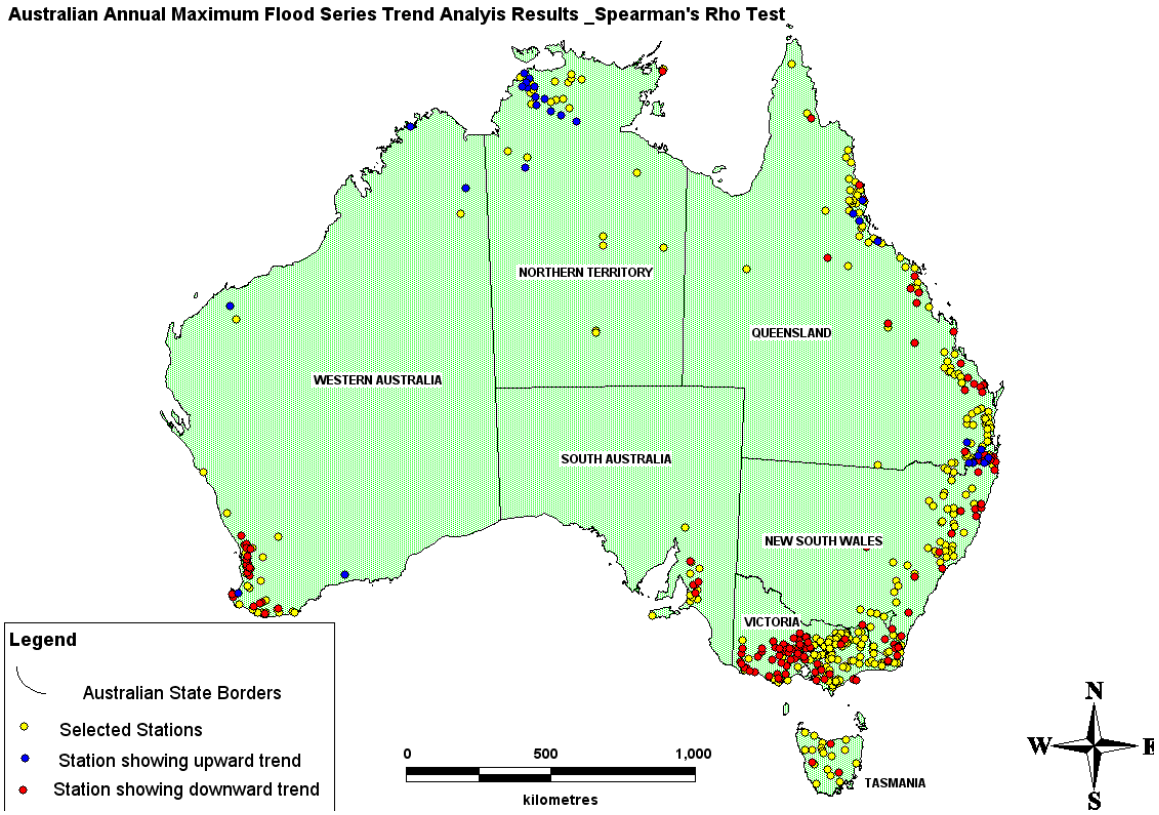


Figure 8.5 Results of trend analysis based on Spearman's Rho. Red and blue circles represent downward and upward trends, respectively

8.6 Impact of serial and spatial correlation on trend results

8.6.1 Site significance assessment

As a first step, the annual maximum (AM) flow series at all stations for the three selected study periods were subjected to serial correlation (or autocorrelation) analysis. It was found that the majority of the stations (with 40 and 50 years study periods) had no significant lag-1 serial correlation coefficient. However, for the 30 years study period, it was found that 23 out of 330 stations (7%) showed positive serial correlation, whereas 8 stations (2%) had negative serial correlation. The mapped results in Figure 8.6 indicated no regional pattern for positive serial correlation. The stations with a significant serial correlation are subjected to the TFPW processes before applying the MK test.

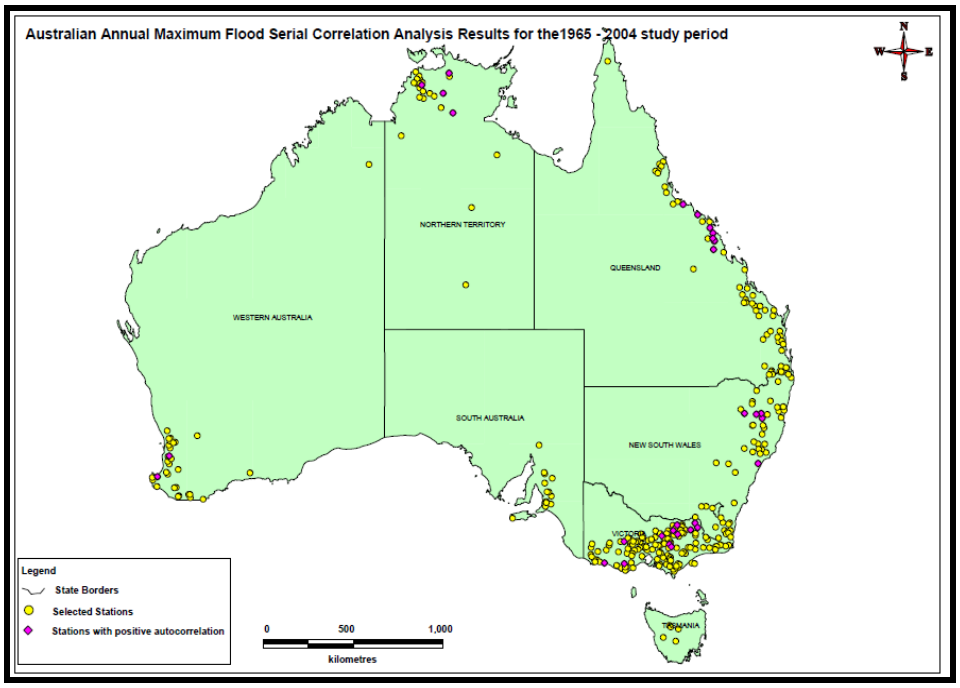


Figure 8.6 Serial correlation analysis. Purple lozenge represents the stations with positive serial correlation

The MK test with the trend free pre-whitening procedures was then applied to assess the significance of trend at the significance level of 0.10 in the annual maximum flow series for 30-, 40-, and 50-year periods. Table 8.2 presents the number of available stations and the percentages of the stations with significant upward and downward trends for the three study periods. Additionally, the last two columns in Table 2 present the total number and the percentage of the stations with significant trends. Apparent from Table 8.2 is the decrease in the density of the station network with the increasing length of the study period. The outcomes indicate that for all the three time frames considered here, the percentages of stations with significant downward trends are higher than the expected number at the significance level of 0.10. For example, consider the 1965-2004 period, here 16 stations out of 77 show downward trend, which is 21% of the total stations, and greater than 7.7 stations, expected to happen by chance at the 10% significance level. The percentages of stations with significant upward trends are not considered to be of statistically significant, as they are smaller than the expected number at the 10% significance level. However, if the stations showing significant trend (irrespective of increasing or decreasing trend) are combined, the percentages of total sites with significant trends are higher than the expected number at the 10% significance level.

Table 8.2 Percentage of Stations with significant upward and downward trends at the significance level of 0.10 (based on trend free pre-whitening procedure)

Study period	Total stations tested	Downward trend		Upward trend		Total stations showing trend	
		Number of stations	%	Number of stations	%	Number of stations	%
1955 - 2004	21	6	29	1	5	7	33
1965 - 2004	77	16	21	2	3	18	23
1975 - 2004	330	75	23	7	2	82	25

Furthermore, the spatial distributions of the stations showing significant trends in annual maximum flows based on TFPW_MK test are shown in Figure 8.7(a, b, c). On the maps, a yellow, blue and a red circle represent a station with no significant trend, significant upward trend and significant downward trend, respectively. The density of stations for the 1955-2004 study periods is quite low and results in a very uneven spatial distribution of gauging stations across the country, making the interpretation of the results quite difficult. Hence, the focus will be on the findings from the 1965-2004 and 1975-2004 study periods, since these periods provide a good spatial coverage and present a reasonably long record length. Figure 8.7(b) shows a decreasing trend for a good number of stations in the south-west of Western Australia, and in the south-east of New South Wales and south-east of Queensland regions. A significant upward trend is found for only a few stations in the north-east of Queensland and in the north of the Northern Territory regions.

The spatial patterns of the trends for the 30-year period are displayed in Figure 8.7(c), which shows an upward trend along the north-west regions, and a downward trend along the south-east and the eastern regions of the country. Interesting is the detection of both upward and downward trends in the south-west of Western Australia. Furthermore, Figures 8.4, 8.5 and Figure 8.7 substantiate that the identified trends in the annual maximum flood series are spatially consistent with the trends in regional mean annual rainfall and other average rainfall characteristics identified previously (e.g., Murphy and Timbal, 2008; Taschetto and England, 2008). These studies have identified an increasing trend of mean annual rainfall and other average rainfall characteristics in the western region (except near the coastline) and a decreasing trend along the eastern coastlines. This similarity in trends and patterns in the annual maximum flow and rainfall variables for the study area implies that the trends in streamflow might be related to the trends in rainfall. However, the flood behaviour is likely to be more related to extreme rainfalls rather than mean annual rainfall. Further study will be

conducted in Stage III of Project 5 to investigate the link between the trends of extreme rainfalls and flood data in Australia.

8.6.2 Field significance of trends

As mentioned earlier, the presence of cross-correlation in a station-network affects the ability of a test to assess the field significance of trends over a network. The cross-correlation coefficient between two sites can be computed after Salas *et al*. (1980). The histograms of the cross-correlation coefficients among the sites of the AM flood series for the three time frames (1955–2004, 1965–2004 & 1975–2004) were plotted in Figure 8.8. This shows that the number of pairs of sites with positive cross-correlation is much greater than those with negative cross-correlation, and that positive cross-correlation dominates the streamflow observation network. The bootstrap test with preserving the cross-correlation structure in the network and with the TFPW approach to remove serial correlation at sites was applied to assess the field significance of trends in the AM flood series over the whole country.

For the purpose of illustration, the bootstrap empirical cumulative distributions (BECDs) of the number of significant upward and downward trends for the AM flood series for the three study timeframes, with preserving the cross-correlation structure of the network, are displayed in Figure 8.9(a), (b), (c). Further, the 90^{th} quantile values of the simulated distribution of the number of sites with significant upward/downward trends developed by the bootstrap resampling approach for the three time periods are presented in Table 8.3. The field assessment results with the consideration of the influence of the cross-correlation among the sites, for the three timeframes, 1955-2004, 1965-2004 and 1975-2004, show that the downward trends in the real data network are statistically significant at the 10% significance level. To clarify, the 90^{th} quantile values of the empirical cumulative distribution for the number of sites with significant downward trend are found to be 3, 9, and 40 for the 50-, 40- and 30-year periods respectively, compared to 6, 16, and 75 sites with significant downward trends for the real data network. On the other hand, none of the three study periods demonstrates that the upward trend in the data network is statistically significant at the 10% level of significance with the consideration of the cross-correlation among the sites in the network.

Figure 8.7 Spatial illustration of significant trends for the annual maximum flows: (a) 1955 - 2004; (b) 1965 - 2004; and (c) 1975 – 2004

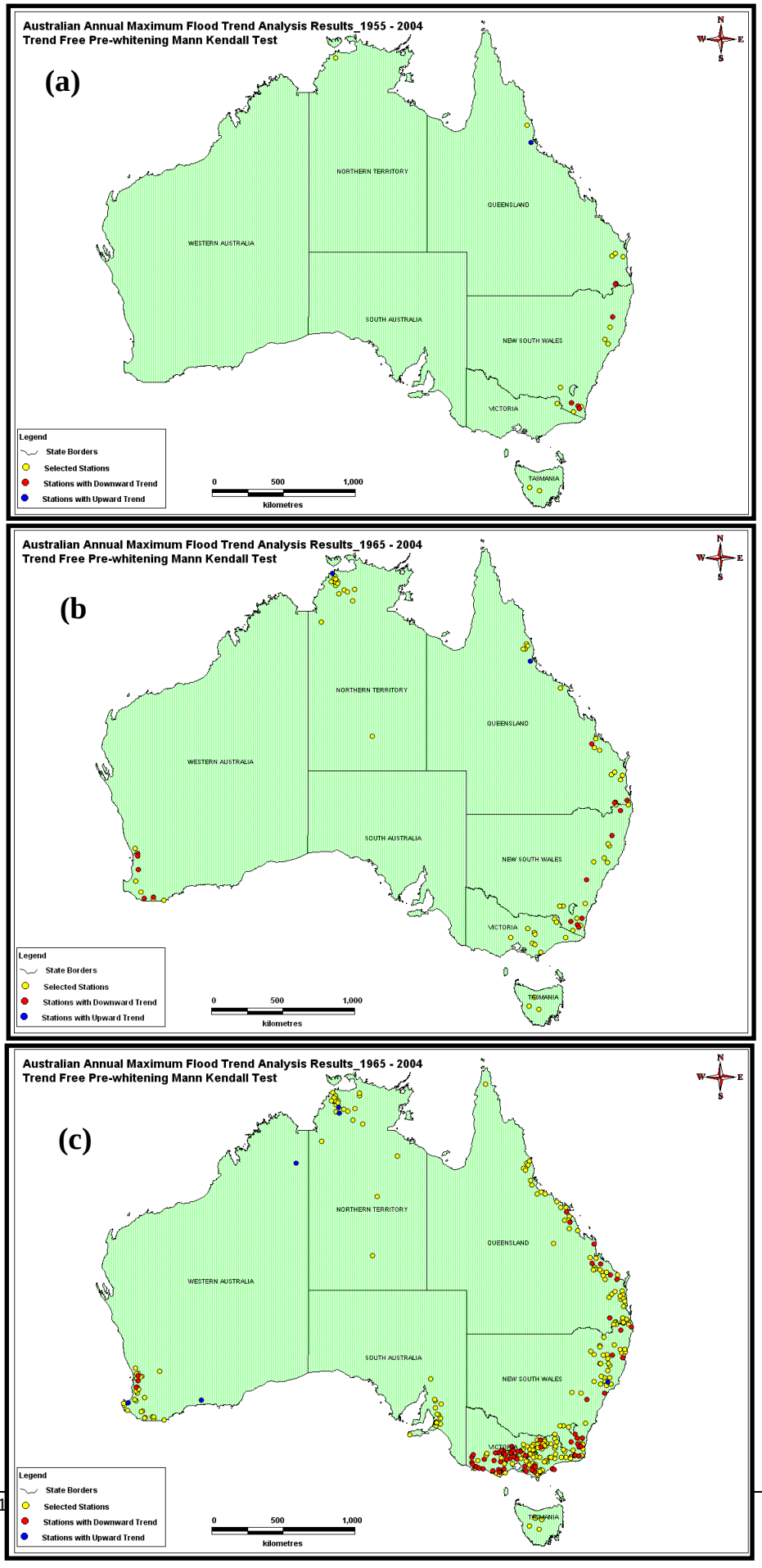
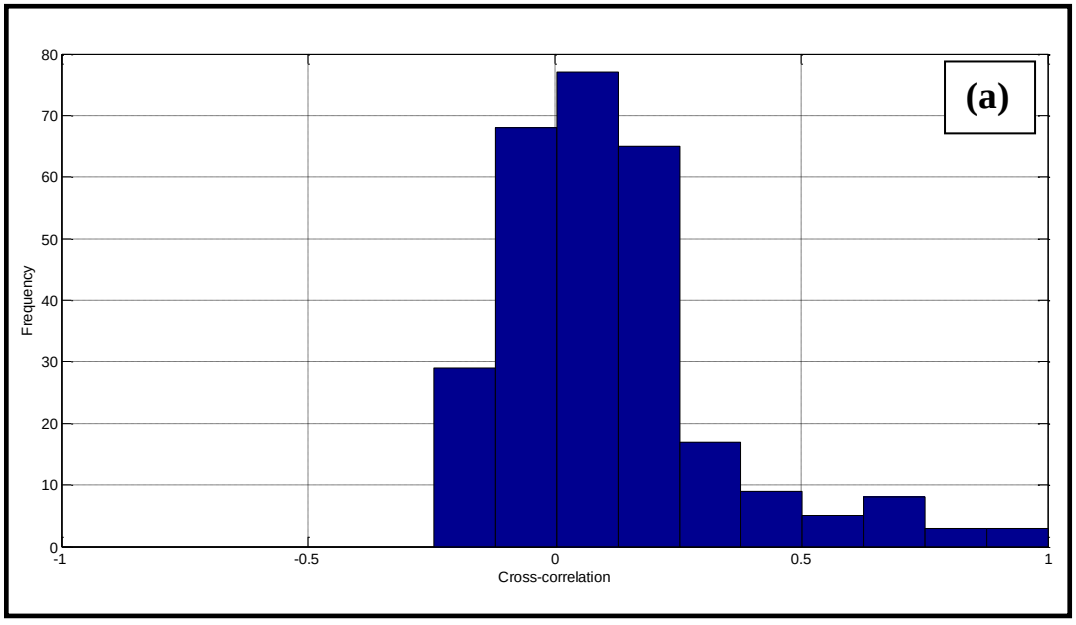
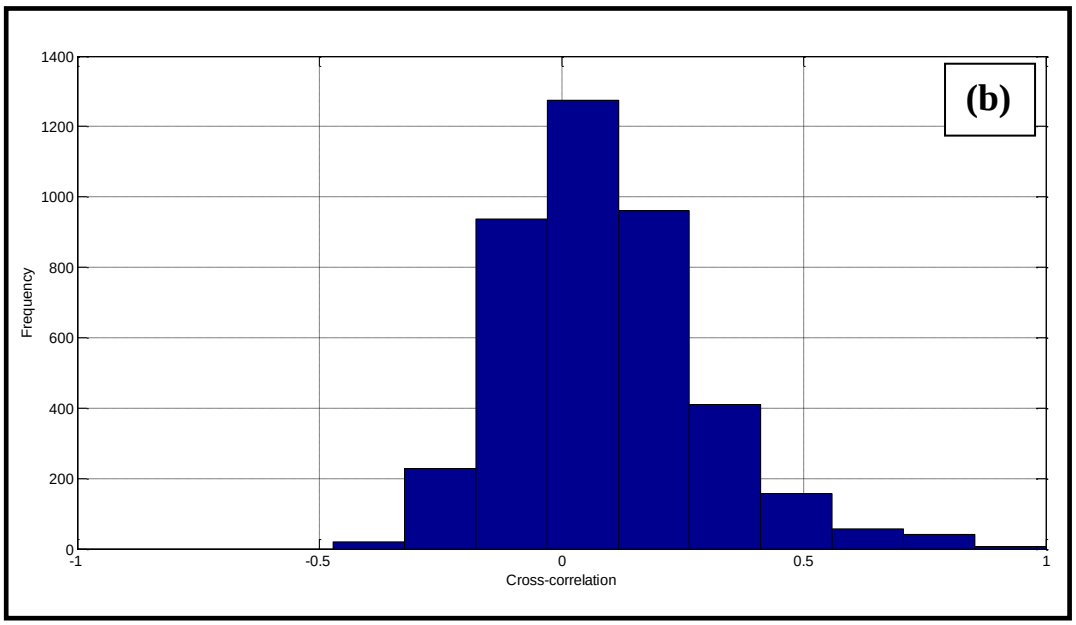


Figure 8.8 Histograms of cross-correlation coefficients of the network for different time frames: (a) 1955-2004; (b) 1965-2004; and (c) 1975-2004





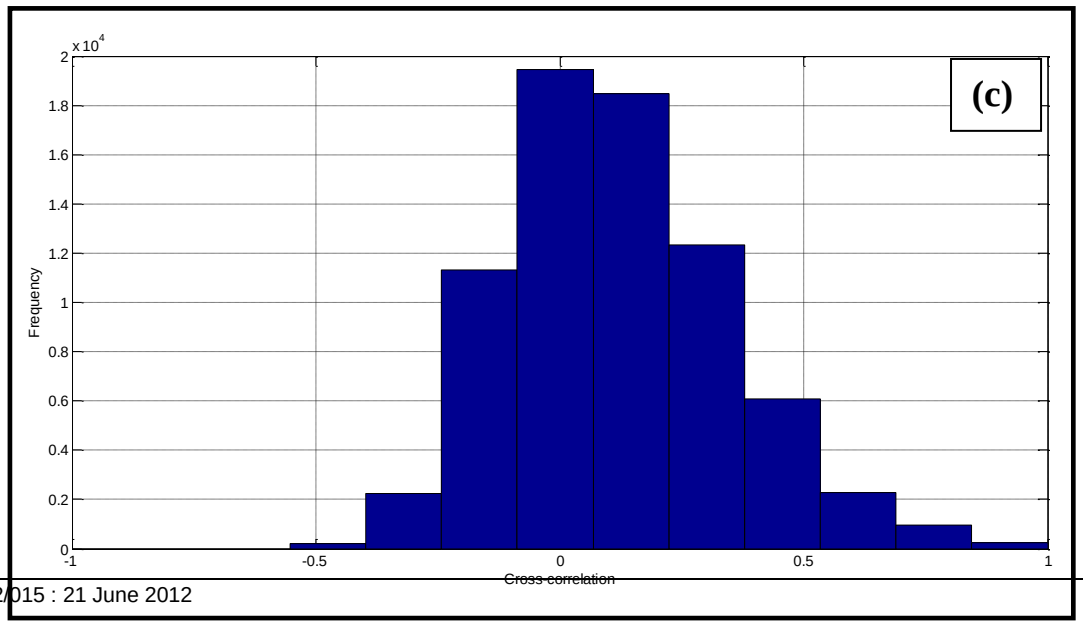


Figure 8.9: BECDs of the number of significant trends for AMFS with preserving the cross-correlation structure of the network: (a) 1955-2004; (b) 1965-2004; and (c) 1975-2004

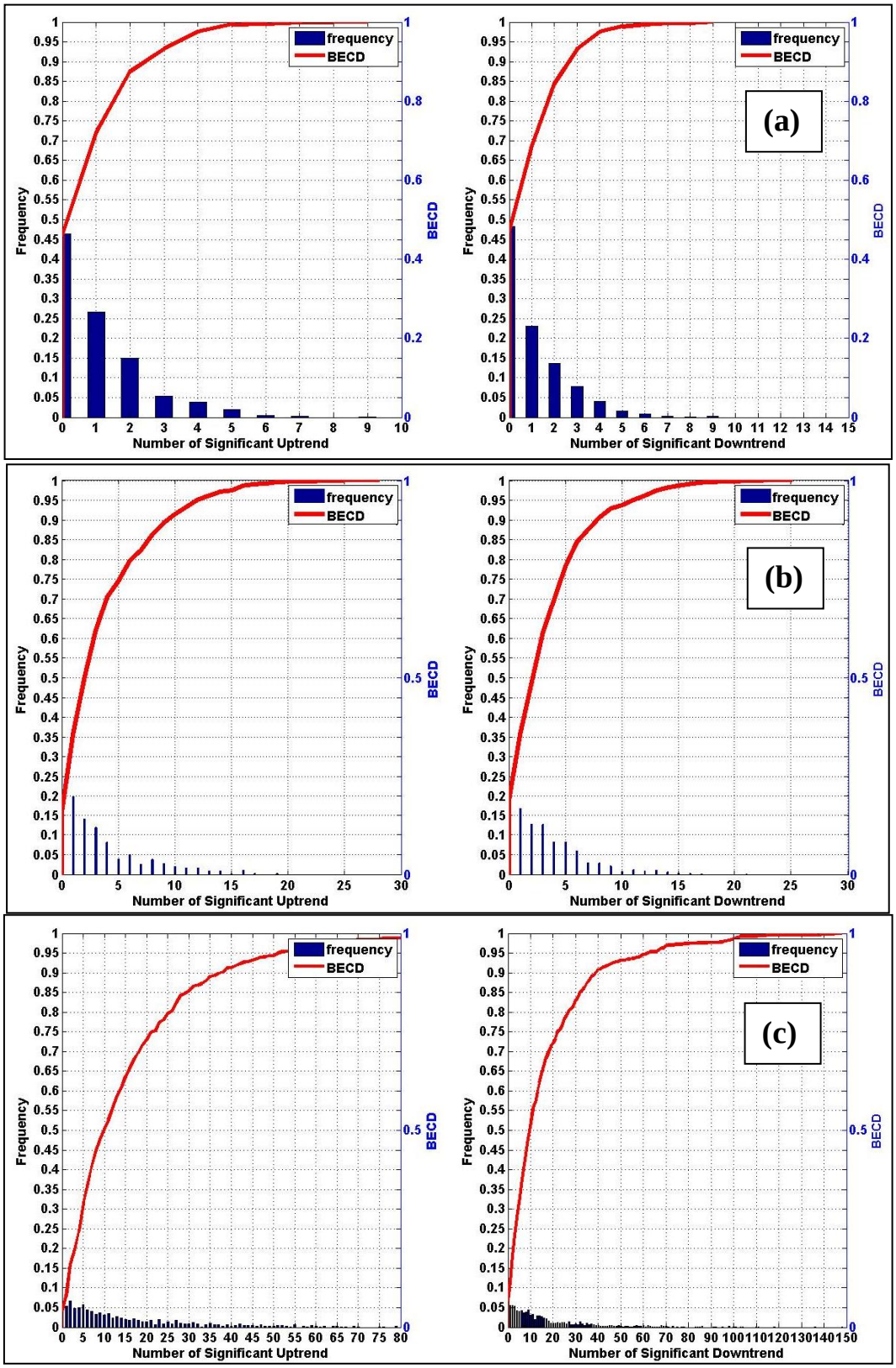


Table 8.3 Field significance assessment results by the bootstrap test.

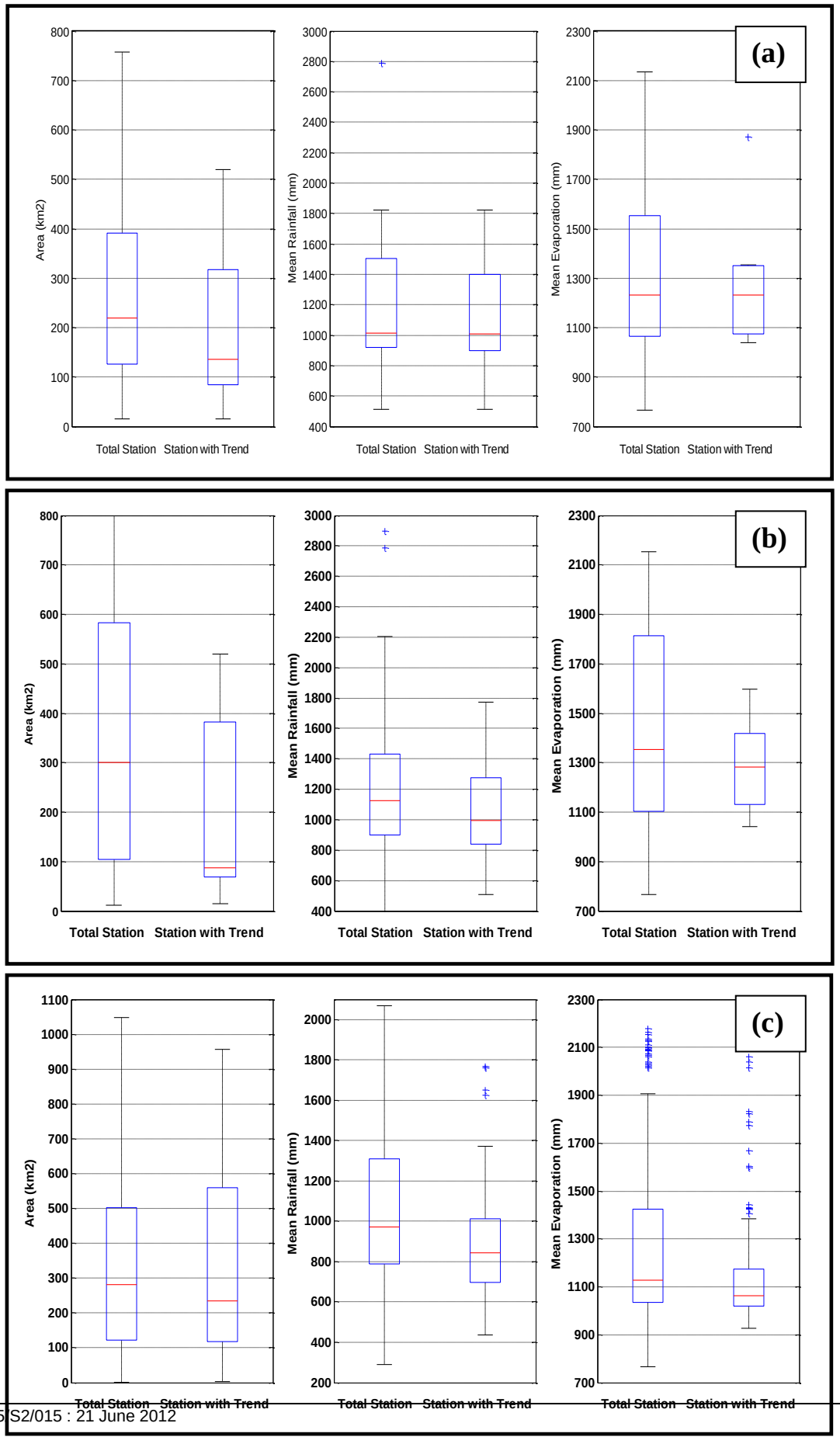
Period	Total Stations Tested	90th quantile value of the simulated distribution	
		Upward	Downward
1955 - 2004	21	3	3
1965 - 2004	77	8	9
1975 - 2004	330	40	40

8.7 Impacts of catchment attributes on trends

In this section the relationships between trends in the AM flow and the catchments attributes are investigated. The catchment attributes are summarised by the catchment areas, mean annual rainfall and mean annual evaporation. **Figure 8.10(a, b, c)** presents box plots for the stations showing significant trends (at the 10% significance level) compared to their parent data set for the three study periods. Apparent from these box plots is that the catchments with smaller size are generally exhibiting more trend than the bigger catchments.

An investigation was also made to see whether the stations showing trends have any unusual physical features e.g. storage, mining, land clearing; however, nothing remarkable was found.

Figure 8.10: Relationships between trends in the AM flow and the catchments attributes: (a) 1955-2004; (b) 1965-2004; and (c) 1975-2004



8.8 Impact of trend on regional flood estimates

To assess the impacts of time trends on regional flood estimates, it is proposed that a non-stationary generalised extreme value (GEV) distribution be fitted to selected stations' AM flood series in different parts of the country and compared with the quantile estimates obtained from the stationary GEV distribution. This will enable identification of the level of expected differences in flood quantile estimates over various ARIs and locations due to the presence of expected levels of upward or downward trends. This has been left for future study.

8.9 Summary

This chapter presents the results of trend analyses for the Australian annual maximum flood series. Firstly, a preliminary trend assessment using the rank-based non-parametric Mann-Kendal and Spearman Rho tests using all the available records for the selected stations was undertaken. However, to eliminate the effect of serial correlation on the Mann-Kendall test for assessing the site significance of a trend, the trend-free pre-whitening (TFPW) procedure was applied to the annual maximum flood series. Also, a bootstrap test with preserving the cross-correlation structure in a station network and with the TFPW to remove serial correlation at a site for assessing separately the field significance of upward and downward trends over the network was used.

There is a good agreement between the outcomes from the site significance assessment based on the TFPW procedures and the results from the preliminary trend analysis. In general, prominent from both the analyses is the geographical distribution of the stations with significant upward and downward trend at the significance level of 0.1; specifically, the findings of negative trend in the annual maximum flood series in the south-east and positive trend in the north-west of the continent. Furthermore, the additional bootstrap test was applied to assess the field significance of the upward and downward trends in the annual maximum flood series for the three timeframes over the country. At the significance level of 10% and with the consideration of the cross-correlation among the sites in the network, field significance of downward trends in the annual maximum flood series for the three time periods was detected over the whole country. Conversely, the field significance of upward trends for the three time periods was found to be statistically non-significant at the 10%

significance level. It has also been found that smaller catchments are affected by trends in greater proportion than the larger ones.

Based on the results of this investigation, it can be concluded that the annual maximum flood series in a large number of Australian stations are affected by a time trend at the 10% level of significance. However, given the length of records and the existence of exceptionally dry weather in the last decade or so, it is not possible to state whether this trend is due to climate change. A further investigation is needed to identify the causes of the identified trends and their possible links with climate indices and physical catchment attributes.

The impacts of the identified trends on regional flood quantile estimates for ARIs in the range of 2 to 100 years will be investigated in Stage III of the project. This is expected to produce climate change adjustment factors as a function of ARIs and locations across Australia.

9. Summary of Project 5 Stage II investigations

The summary of research investigations from Project 5 Stage II, as presented in this report, is provided below.

1) **Updated national database:** The initially prepared streamflow and catchment attributes database in Stage I of the project has been upgraded, as summarised below. The national database now contains 682 stations from coastal regions of Australia and 45 stations from arid/ semi-arid regions, giving a total of 727 stations.

   a) New database for WA has been prepared consisting of 146 stations:
      i) 120 stations from south-west region (Drainage Division VI)
      ii) 12 stations from Pilbara region (Drainage Division VII)
      iii) 14 stations from Kimberley region (Drainage Division VIII – WA part)

   b) Database for Tasmania has been updated which now contains 53 stations:
      i) 32 stations from western Tasmania and
      ii) 21 stations for east Tasmania.
   c) Database for the NT has been updated, now contains 55 stations.
   d) Database for Victoria, NSW, Qld and SA have been further tested and updated where needed. The number of stations for these states are:
      i) Victoria – 131 stations
      ii) NSW - 96 stations
      iii) Qld – 172 stations
      iv) SA - 29 stations
   e) Database for arid/semi-arid regions has been prepared, containing 45 stations from SA, Vic, NSW, Qld and NT.

   Some important details of the national database are provided in Table 9.1. The locations of the selected 727 stations are plotted in Figure 9.1.

Table 9.1 Summary of national database (Project 5 Regional flood methods in Australia)

State	No. of stations	Median annual maximum flood record length (years)	Median catchment size (km^2)
NSW & ACT	96	34	267
Victoria	131	33	289
South Australia	29	34	76.5
Tasmania	53	28	158
Queensland	172	36	254
Western Australia	146	30	60
Northern Territory	55	33	360
Sub Total	**682**	-	-
Arid semi-arid region	45	22	360
TOTAL	**727**	-	-

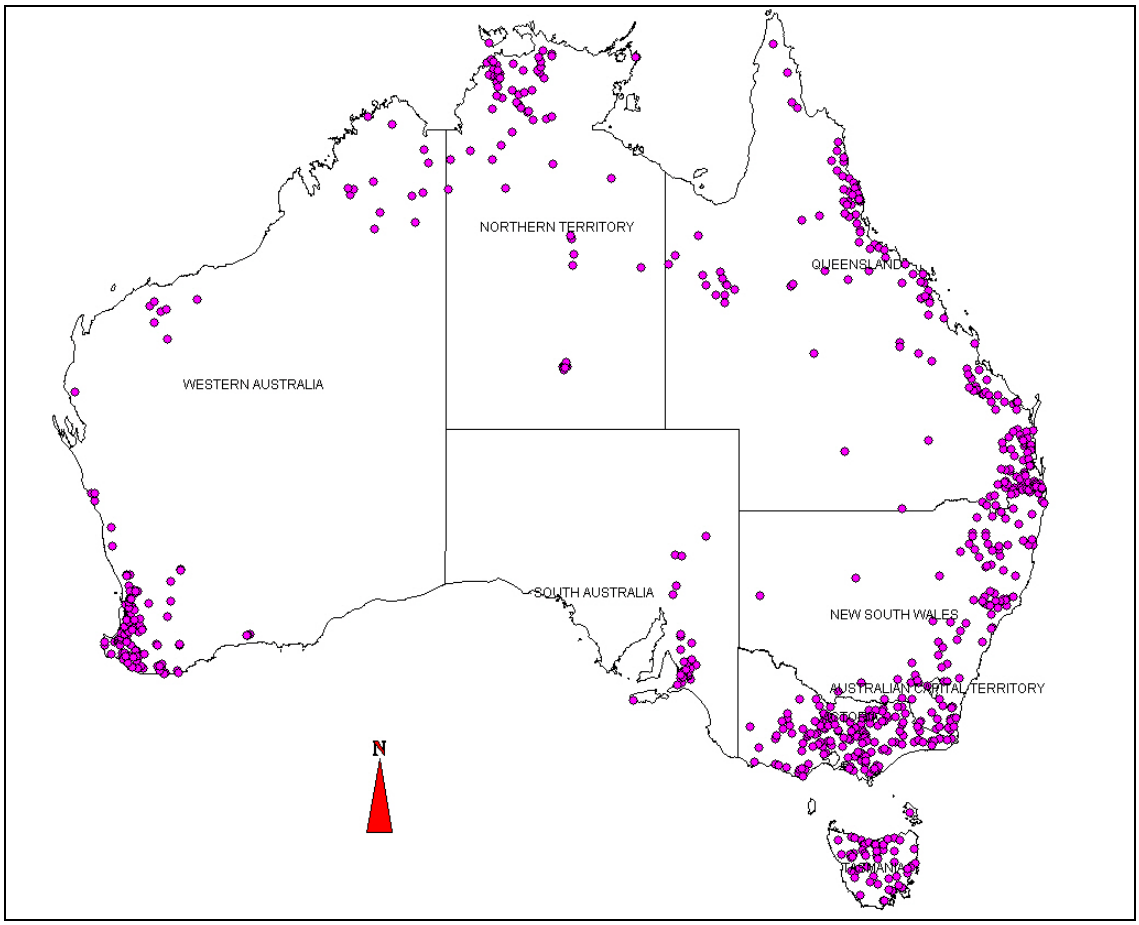


Figure 9.1 Geographical distributions of the selected 727 stations for Project 5

2) **GLS based Quantile Regression Technique (QRT) vs. Probabilistic Rational Method (PRM):** The two methods have been compared for the state of NSW. To make a valid comparison, the same predictor variables and data set have been used with both the methods. The comparison has examined the specific features of each method and assessed its performance using a one-at-a-time validation approach, where each of the study catchments is tested independently, as well as a split sample approach, leaving a randomly selected 20% of catchments for independent testing. Based on a range of evaluation statistics (such as root mean squared error, median relative error and ratio of predicted and observed flood quantiles), the QRT has been found to outperform the PRM. No evidence has been found that PRM and QRT perform more poorly for smaller catchments, as far as the range of smaller catchments used in this study is concerned. The applicability of these methods for catchments smaller than 8 km^2 could not be tested due to limitations of streamflow data for these catchments.

   The particular advantage of the QRT is that it does not require a contour map of the runoff coefficient as with the PRM. The GLS-based QRT also offers rigorous uncertainty analysis of the estimated flood quantiles by differentiating the sampling and model error. The QRT can also be integrated with the region-of-influence approach where a region can be formed around an ungauged catchment by selecting an 'appropriate number' of neighbouring gauged catchments based on the criterion of minimum model error variance. Hence, QRT offers much greater flexibility and potential in terms of error analysis and further development.

   Based on these findings, the PRM method was not considered for further testing. Subsequently, the QRT has been further developed and tested with the region-of-influence approach and compared with the Parameter Regression Technique (PRT). In the PRT, the first three moments of the LP-3 distribution have been regionalised.

3) **QRT vs. PRT and fixed regions vs. region-of-influence approach:** The GLS regression method developed in the Stage I project has been enhanced by adding Bayesian analysis, which is referred to as Bayesian GLS regression. The Bayesian Quantile Regression Technique (QRT) and Bayesian Parameter Regression Technique (PRT) have been compared for various Australian states (i.e. Victoria, NSW & ACT, Tasmania, Qld, SA, NT and WA).

It has been found that Bayesian QRT and Bayesian PRT methods perform very similarly for various Australian states. Since the PRT method offers several additional advantages over the QRT (namely, in PRT, flood quantiles increase smoothly with increasing ARIs and floods of any ARI in the range of 2 to 100 years can be estimated), this has been recommended for further testing for inclusion in the 4th edition of ARR. It has been found that for the PRT, only two predictor variables (catchment area and design rainfall intensity of 2 years ARI and 12 hours duration) are needed to estimate the mean annual flood and regional average values of the standard deviation (SD) and skew of ln(Q), which can then be used to fit the regional LP3 distribution. This will make the application of the PRT method relatively easy in practice.

From the comparison of fixed regions and region-of-influence (ROI) approaches, it has been found that the ROI approach outperforms the use of fixed regions; the ROI reduces the model error variance by reducing the size of the region i.e. this provides a region with a lower level of heterogeneity. It has been found that the mean flood model has the highest model error as compared to the SD and skew models. In the ROI approach, the mean flood, SD and skew models typically require about 40, 60 and 100 sites, respectively.

4) **Applicability of the RFFA method to small catchments:** It has been found that the recommended RFFA methods i.e. GLS-PRT-ROI and GLS-PRT-fixed-region perform quite well for the smaller catchments in the database, where there is no evidence that smaller catchments perform more poorly than for the medium and larger catchments. The possibility of extending the RFFA method to very small catchments beyond the limit of the current Project 5 database has been examined; however, further study is needed to develop an acceptable method.

5) **RFFA method for flood estimation in the large flood range:** It has been found in Stage 1 of Project 5 that the development of a simple Large Flood Regionalisation Model for regional flood estimation in the 'rare' flood range (ARIs 100 to 2000 years) needs consideration of the cross-correlations among the highest data points from each station's AM series. A number of possible ways to deal with this problem have been discussed however, this aspect needs further investigation.

6) **RFFA method for arid/semi-arid regions:** Catchments in the arid and semi-arid regions of Australia have a distinctly different flood hydrology from catchments in more humid/coastal regions, and they thus warrant separate treatment for regional flood

estimation. The limited availability of streamflow data at both temporal and spatial scales in the arid and semi-arid regions of Australia makes it difficult to develop statistically meaningful RFFA methods. Using the limited data, a simple RFFA method has been developed for the arid and semi-arid regions, which has generally a lower degree of accuracy than the RFFA methods developed in Chapter 4 for the coastal regions of Australia, which are based on a larger data set with good quality. The RFFA methods presented here for the arid and semi-arid regions should be used with caution; local hydrological knowledge must be exercised to interpret the results. It would be necessary to compare the results from these methods with other methods, such as rainfall runoff models and any locally available methods. The RFFA methods presented in this report for arid/semi-arid region will require further development and testing before they can be recommended for inclusion in the revised ARR.

It is recommended that some representative stream gauges be established in the arid and semi-arid regions for 'long term monitoring' to develop a more comprehensive database, which will assist in upgrading the RFFA methods presented here in the future. A more comprehensive flood data base is required to develop a better understanding of how the special climate, catchment and stream characteristics of the arid and semi-arid regions of Australia interact to produce distinctly different flood responses. It is to be noted here that a high degree of uncertainty associated with a RFFA method results in inaccurate design flood estimates, which increase the capital cost of the infrastructure in the case of over-design, or the average annual flood damage cost in the case of under-design. The cost of streamflow monitoring and data collection is expected to be far less than the cost associated with grossly inaccurate RFFA methods developed based on inadequate data.

Trends in the annual maximum flood data: The impacts of serial and cross-correlation on trend analysis in the annual maximum flood series have been investigated. At the significance level of 10% and with the consideration of the cross-correlation among the sites in the network, field significance of downward trends in the annual maximum flood series was detected over the whole country. Conversely, the field significance of upward trends was discovered to be not statistically significant at the 10% level. It has also been found that smaller catchments are affected by trends in greater proportion than the larger catchments. Based on the results of this investigation, it can be stated that the annual maximum flood series in a large number of Australian stations are affected by a time trend at 10% level of significance. However, given the length of records and the existence of exceptionally dry weather conditions in the last decade or so, it is not possible to state whether this trend is

due to climate change. A further investigation is needed to identify the causes of the identified trend and their possible links with climate indices and physical catchment attributes. The impacts of the identified trends on regional flood quantile estimates for ARIs in the range of 2 to 100 years will be investigated in Stage III of the project. This is expected to produce climate change adjustment factors as a function of ARIs and locations across Australia.

7)

10 Recommended RFFA methods for inclusion in the ARR and further testing and development

10.1 Recommended RFFA methods for ARR

1. Based on the findings of Stage I and Stage II of Project 5, it is recommended that the Bayesian-GLS-PRT method be applied as a *general RFFA method* for Australia. Also, the region-of-influence (ROI) approach should be used if there are enough stations in a region (say at least 50 stations). For the arid/semi-arid region, a simple index type RFFA method is recommended for use.

2. Given the availability and geographical contiguity of a sufficient number of stations, the Bayesian-GLS-PRT-ROI method is recommended for the states of Victoria, NSW/ACT, Qld and south-west WA. Here, the state boundaries between Victoria, NSW/ACT and Qld should be disregarded and the stations from these regions combined into a single database to apply the Bayesian-GLS-PRT method.

3. For Tasmania, two separate fixed regions are recommended: east Tasmania with 21 stations and western Tasmania with 32 stations.

4. For the NT, two separate fixed regions are recommended, one for north-west NT (Drainage Division VIII – NT part, containing 51 stations) and the other for Drainage Division IX (NT part). The arid/semi-arid part of the NT is to be treated as a separate region (possibly to be combined with data from other arid/semi-arid regions).

5. For SA, a fixed region is recommended for Drainage Division V containing 29 stations. The arid/semi-arid part of SA should be treated as a separate region (possibly to be combined with data from other arid/semi-arid regions).

6. For WA, the Kimberley region (containing 14 stations) should be treated as a separate fixed region and, similarly, the Pilbara region (containing 12 stations) should be treated as a separate fixed region.

7. A simplified index-type RFFA method is recommended for arid/semi-arid regions of Australia, where four separate regions are recommended at this stage (this needs further development and testing before inclusion in ARR):
   (1) Arid/semi-arid parts of NSW and Victoria – one region (containing 9 stations)
   (2) Arid/semi-arid part of SA – one region (containing 6 stations)
   (3) Arid/semi-arid part of Qld – one region (containing 16 stations)
   (4) Arid/semi-arid part of NT – one region (containing 14 stations)

Recommended RFFA methods for various Australian regions are summarised in Table 9.2.

Table 9.2 Recommended RFFA methods for inclusion in ARR (subject to further testing)

State	Region	Number of stations	Method of forming region	Estimation model
NSW, ACT, Vic, Qld		399	ROI	Bayesian GLS-PRT-ROI
Tasmania	west Tasmania	32	Fixed region	Bayesian GLS-PRT
	east Tasmania	21	Fixed region	Bayesian GLS-PRT
South Australia		29	Fixed region	Bayesian GLS-PRT
Northern Territory	North-west NT (Drainage Division VIII – NT part)	51	Fixed region	Bayesian GLS-PRT
	North east NT (Drainage Division IX – NT part)	4	Fixed region	Bayesian GLS-PRT
WA	Kimberley region	14	Fixed region	Bayesian GLS-PRT
	Pilbara region	12	Fixed region	Bayesian GLS-PRT
	South-west WA	120	ROI	Bayesian GLS-PRT-ROI
Arid/semi-arid region	Arid/ semi-arid parts of NSW and Vic	9	Fixed region	Simple index flood method

	Arid/ semi-arid part of SA	6		
	Arid/ semi-arid part of Qld	16		
	Arid/ semi-arid part of NT	14		
	Arid/semi-arid part of WA	-	To be identified	To be identified

10.2 Further development and testing of the RFFA methods to be included in ARR

The RFFA methods identified for inclusion in the ARR chapter (as discussed in Section 10.1) need further testing/development to come up with the final set of methods, design databases, user guidelines and application tools. This will form the scope of Stage III project, as summarised in Table 10.2.

Table 10.2 Further development and testing of the RFFA methods to be included in ARR (Stage III proposed scope)

Task	Description
1	Further testing and development of the recommended regional flood frequency analysis (RFFA) method, which is the parameter regression technique (PRT) using Bayesian GLS regression with region-of-influence (ROI) and fixed regions: 1.1 Testing for smaller catchments 1.2 Investigating the outlier catchments 1.3 Examining the effects of removing the state boundaries for the combined data set of Vic, NSW and Qld on the results of the PRT-GLS-ROI method 1.4 Testing the recommended RFFA method with randomly selected ungauged catchments in different states and compare the results with alternative methods 1.5 Linking the method with at-site flood data and historical information, if any 1.5 Documenting the results of testing and any modifications to the recommended methods.
2	Development of a database and user instructions for pilot application of the recommended methods and benchmark testing by various state agencies/stake holders 2.1 Deriving gridded coefficient values for PRT-GLS-ROI method, producing map in GIS format 2.2 Preparing a set of guidelines on application of method (with a simple to use

	spreadsheet) 2.3 Analysing results of benchmark testing and adjusting method/guidelines for application.
3	Development of a functional specification for the Windows-based GIS application tool for routine application of the recommended Bayesian-GLS-PRT-ROI method, including: * clear delineation of the limits of application in terms of the range of predictor variables used in the regressions and any other constraints * procedures for flood estimation in boundary areas between regions * calculation and presentation of uncertainties in flood estimates * generation of warning messages if uncertainties exceed nominal limits.
4	Development and testing of Windows-based GIS application tool for routine application of the recommended RFFA methods.
5	Recalibrating the adopted RFFA methods with new design rainfall data.
6	Examination of trends in annual maximum flood data and identification of the impacts of climate variability and change on regional flood estimates.
7	Development of a simple regional flood estimation techniques for large to extreme floods (Application of Large Flood Regionalisation Model) in the range of 100 to 1000 years.
8	Preparation of technical reports and refereed papers.
9	ARR chapter drafting and finalisation.

11. Development of application tools

The RFFA models based on fixed regions can be applied without the need of any software/application tool. However, the application of the Bayesian GLS-PRT-ROI method [for NSW, ACT, Vic, Qld and WA (south-west part)] would preferably require user-friendly software. The **specifications of the Windows-based GIS application tool/software** for routine application of the recommended Bayesian GLS-PRT-ROI method are provided below and illustrated in the flow chart in Figure 11.1.

1. **Locating the ungauged catchment of interest:** The software will ask the user to enter the name, latitude and longitude of the ungauged catchment. The software will then plot the location of the ungauged catchment on a map of Australia. The software will check whether the ungauged catchment falls within the limits of application defined by the design data set that were used to develop the RFFA method which underpins the software.

2. **Obtaining regression coefficients and values of regional average standard deviation and skew of ln(*Q*):** Regression coefficient values for estimation of the mean flood and regional average values of standard deviation and skew at each of the pre-determined grid points will be provided to the software developer in the form of a database table. Based on the location of the ungauged catchment, the software will select the regression coefficients from the grid point which is closest to the location of the ungauged catchment in question. It will then calculate the mean annual flood and display the mean annual flood and regional average values of SD and skew of ln(*Q*) in a window.

3. **Abstraction of predictor variables at the ungauged site of interest:** Catchment area will be estimated by the user outside the software or it can be obtained from a linked GIS-based catchment map. The software may be linked with the BOM design rainfall calculator so that the user can obtain the required design rainfall intensity value at the ungauged site of interest directly. The software will then check whether the estimated area and rainfall intensity values are within the range of design data used to derive the gridded coefficient values e.g. whether 3 km^2 <= area <= 1000 km^2 and 0.80×Imin <= I <= 1.20×Imax (where Imin and Imax are minimum and maximum

design rainfall intensity values respectively from the model database) (these limits will be confirmed after further testing).

4. **Fitting the regional LP3 distribution/PRT:** Once the mean flood, SD and skew are calculated for the ungauged catchment, the frequency factors K_T will be estimated and flood quantiles Q_T will be obtained for the specified range of T values and reported in a table. A flood frequency curve will be plotted by the software.

5. **Poorly gauged catchments with some data availability:** If at-site flood data or historical flood information are available, the regional estimate obtained in step 4 will be updated.

6. **Confidence limits:** The confidence limits of the estimated Q_T values will be derived and plotted by the software.

7. **Upgrade of gridded coefficients:** Once a sufficient amount of new flood data are available, the gridded coefficients should be upgraded and fed into the software. This may need to be done every 5 years or so.

8. **Hosting of the software:** The software should be web-based and hosted by an organisation like EA or BOM.

9. **Continuing feedback during initial period of application:** The Project 5 team will be available for providing feedback during the initial application period of the software by the industry.

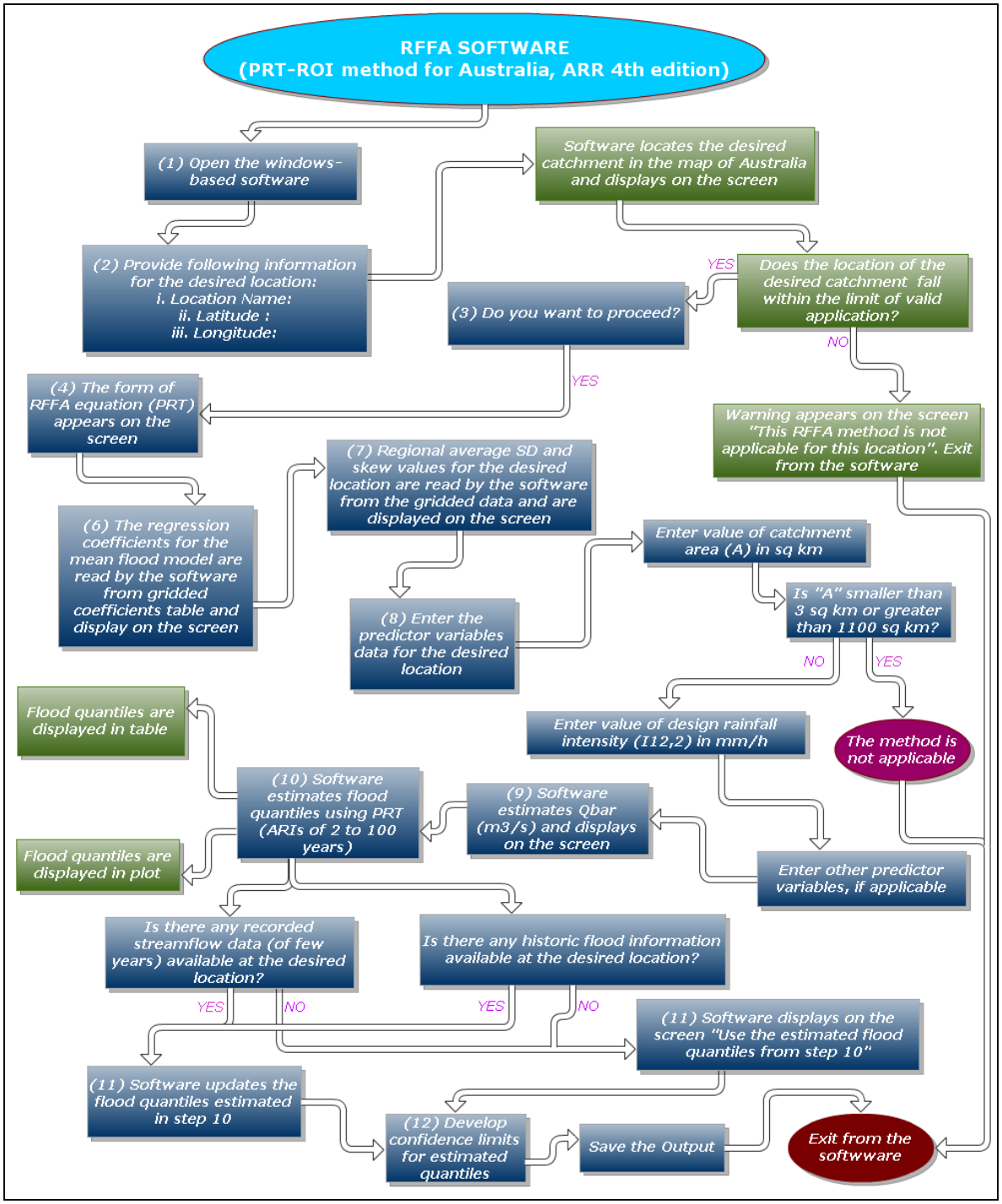


Figure 11.1 Flow chart showing the desirable features of the application tools/software for implementing the Bayesian GLS-PRT-ROI regional flood frequency analysis method

12. Conclusions

This report presents research and investigations carried out for ARR Project 5 Regional flood methods in Australia (Stage II). The following conclusions can be drawn from this report:

1. A quality controlled national database consisting of 727 stations has been prepared for development and testing of regional flood frequency analysis (RFFA) methods to be included in the upcoming 4th edition of Australian Rainfall and Runoff (ARR). This database covers all of mainland Australia and Tasmania.

2. It has been found that regression-based RFFA methods (such as the Quantile Regression Technique (QRT) or the Parameter Regression Technique (PRT)) are preferable to the Probabilistic Rational Method. The particular advantage of the QRT and PRT is that they do not require a contour map of the runoff coefficient as with the PRM. The GLS-based QRT/PRT methods also offer rigorous uncertainty analysis of the estimated flood quantiles by differentiating between the sampling and model errors. The QRT/PRT can also be integrated with the region-of-influence approach, where a region can be formed around an ungauged catchment by selecting an 'appropriate number' of neighbouring gauged catchments, based on the criterion of minimum model error variance. Hence, QRT/PRT offer much greater flexibility and potential in terms of error analysis and further development.

3. It has been found that Bayesian QRT and Bayesian PRT methods perform very similarly for various Australian states. Since the PRT method offers several additional advantages over the QRT (namely, in the PRT flood quantiles increase smoothly with increasing ARIs and flood quantiles of any ARI (in the range of 2 to 100 years) can be estimated from the regional LP3 distribution), this has been recommended for general application in Australia.

4. From the comparison of the fixed regions and region-of-influence (ROI) approaches, it has been found that the ROI approach outperforms the fixed regions approach; the ROI reduces the model error, i.e. regional heterogeneity. It has also been found that the mean annual flood model has the highest model error as compared to the SD and skew models. In the ROI approach, the mean flood, SD and skew models typically require about 40, 60 and 100 sites, respectively.

5. The developed RFFA methods require data of two or three climatic and physical catchment characteristics (i.e. catchment area and design rainfall intensity or mean annual rainfall), which are easy to obtain. This would make the application of the recommended RFFA methods easy and simple.

6. It has been found that the recommended RFFA methods, i.e. GLS-PRT-ROI and GLS-PRT-fixed-region, perform quite well for the smaller catchments in the database where there is no evidence that smaller catchments perform more poorly than the medium and larger catchments. The possibility of extending the RFFA method to very small catchments beyond the limit of the current Project 5 database has been examined; however, further study is needed to develop an acceptable method.

7. The development of a simple Large Flood Regionalisation Model for regional flood estimation in the major flood range needs consideration of the cross-correlations among the highest data points from each station's AM series. A number of possible ways to deal with this problem have been discussed; however this aspect needs further investigation.

8. There is insufficient streamflow data available at both temporal and spatial scales in the arid and semi-arid regions of Australia that can be used to develop statistically meaningful RFFA methods. A simplified index type RFFA method is recommended for arid/semi-arid regions of Australia where four separate regions have been identified at this stage (this needs further development and testing before inclusion in ARR).

9. The impacts of serial and cross-correlation on trend analysis in the annual maximum flood series have been investigated. At the significance level of 10% and with the consideration of the cross-correlation among the sites in the network, field significance of downward trends in the annual maximum flood series was detected over the whole country. Conversely, the field significance of upward trends was discovered to be not statistically significant at the 10% level. It has also been found that smaller catchments are affected by trends in greater proportion than the larger catchments. Based on the results of this investigation, it can be concluded that the annual maximum flood series at a large number of Australian stations are affected by a time trend at the 10% level of significance. However, given the length of records and the existence of exceptionally dry weather conditions in the last decade or so, it is not possible to state whether this trend is due to climate change. A further

investigation is needed to identify the causes of the identified trends and their possible links with climate indices and physical catchment attributes. The impacts of the identified trends on regional flood quantile estimates for ARIs in the range of 2 to 100 years will be investigated in Stage III of the project. This is expected to produce climate change adjustment factors as a function of ARIs and locations across Australia.

10. The testing of the recommended RFFA methods by various states/stakeholders in cooperation with the Project 5 team has been recommended. A set of future tasks has been identified. Also, the scope of developing an application tool/software has been indicated.

Based on the research work related to Project 5, a total of 22 peer-reviewed technical papers and research reports have been published/ prepared as of Dec 2010 (Listed in Appendix B). This has provided the opportunity of getting important comments and feedbacks from experts in the field of regional flood estimation and of subsequent improvement of the methods reported here.

Project 5 Stage I built a national database, established a network of relevant researchers and professionals and generated preliminary results for recommending potential RFFA methods for detailed investigation. Subsequently, Stage II has developed a firm basis for recommendations on the RFFA methods to be included in the revised ARR Chapter (4^{th} edition). It has also identified future research and development work in Stage III of the Project, required to develop the Stage II findings into a final set of methods, design databases, user guidelines and application tools.

The results presented in this report are applicable to the rural catchments in the vicinity of the catchments selected in this study; this should not be applied to urban catchments.

References

Abdul Aziz, O.I. and Burn, D.H. (2006). Trends and variability in the hydrological regime of the Mackenzie River Basin. Journal of Hydrology, 319, 282-294.

Acreman, M. C. and Wiltshire, S. E. (1987). Identification of regions for regional flood frequency analysis. Abstract, EOS, 68, 44, 1262.

Alexander, L.V., Hope, P., Collins, D., Trewin, B., Lynch, A. and Nicholls, N. (2007). Trends in Australia's climate means and extremes: a global context. Australian Meteorological Magazine, 56, 1-18.

Australian Bureau of Statistics (ABS) (2009). Consumer Price Index. Accessed from http://www.abs.gov.au/.

Aziz, K., Rahman, A., Fang, G., Haddad, K. and Shrestha, S. (2010). Design flood estimation for ungauged catchments: Application of artificial neural networks for eastern Australia. World Environmental and Water Resources Congress 2010, American Society of Civil Engineers (ASCE), 16-20 May 2010, Providence, Rhode Island, USA, 2841-2850.

Bates, B.C., Rahman, A., Mein, R. G. and Weinmann, P. E. (1998). Climatic and physical factors that influence the homogeneity of regional floods in south-eastern Australia. Water Resources Research, 34, 12, 3369-3381.

Benson, M.A. (1962). Evolution of methods for evaluating the occurrence of floods. U.S. Geological Survey, Water Supply Paper, 1580-A, 30.

Birsan, M.V., Molna´r, P. Burlando, P. and Pfaundler, M. (2005). Streamflow trends in Switzerland. Journal of Hydrology, 314, 312–329.

Burn, D.H. Cunderlik, J. M. and Pietroniro, A. (2004a). Hydrological trends and variability in the Liard River Basin. Hydrological Sciences Journal, 49, 1, 53-67.

Burn, D.H., Abdul Aziz, O.I. and Pietroniro, A. (2004b). A comparison of trends in hydrometeorological variables for two watersheds in the Mackenzie River Basin. Canadian Water Resources Journal, 29, 4, 283-298.

Burn, D.H. and Hag Elnur, M.A. (2002). Detection of hydrologic trends and variability. Journal of Hydrology, 255, 107–122.

Burn, D.H. (1990a). An appraisal of the region of influence approach to flood frequency analysis, Hydrological Sciences Journal, 35, 2, 149-165.

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

Chiew, F.H. and McMahon, T.A. (1993). Detection of trend or change in annual flow of Australian rivers. International Journal of Climatology, 13, 643–653.

Chowdhury, R.K. and Beecham, S. (2009). Australian rainfall trends and their relation to the southern oscillation index. Hydrological Processes, (doi: 10.1002/hyp.7504).

Collins, D.A., Della-Marta, P.M., Plummer, N. and Trewin, B.C. (2000). Trends in annual frequencies of extreme temperatures events in Australia. Australian Meteorological Magazine, 49, 277-292.

Costelloe, J.F., Grayson, R.B., Argent, R.M. and McMahon, T.A. (2003). Modelling the flow regime of an arid zone floodplain river, Diamantina River, Australia. Environmental Modelling & Software, 18, 693–703.

Cunderlik, J.M. and Burn, D.H. (2003) Non-stationary pooled flood frequency analysis. Journal of Hydrology, 276, 210-223.

Douglas, E.M., Vogel, R.M. and Kroll, C.N. (2000). Trends in floods and low flows in the United States: impact of spatial correlation. Journal of Hydrology, 240, 1–2, 90–105.

Eng, K., Tasker, G.D. and Milly, P.C.D. (2005). An Analysis of Region-of-Influence Methods for Flood Regionalization in the Gulf-Atlantic Rolling Plains. Journal of the American Water Resources Association, 41, 1, 135-143.

Farquharson, F.A.K., Meigh, J.R. and Sutcliffe, J.V. (1992). Regional flood frequency analysis in arid and semi-arid areas. Journal of Hydrology, 138, 487-501.

Fleming, S.W. and Clarke, G.K.C. (2003). Glacial control of water resources and related environmental responses to climatic warming: empirical analysis using historical streamflow data from Northwestern Canada. Canadian Water Resources Journal, 28, 1, 69-86.

Franks, S.W. and Kuczera, G. (2002). Flood frequency analysis: evidence and implications of secular climate variability, New South Wales. Water Resources Research, 38, 5, 1062, (doi:10.1029/2001WR000232).

French, R. (2002). Flaws in the rational method. 27th National Hydrology and Water Resources Symp, Melbourene, 20-23 May.

Fu, G., Viney, N.R. and Charles, S.P. (2008). Temporal variation of extreme precipitation events in Australia; 1910-2006. Hydrology and Water Symposium, Adelaide.

Gallant, A.J.E., Hennessy, K.J. and Risbey, J. (2007). Trends in rainfall indices for six Australian regions: 1910-2005. Australian Meteorological Magazine, 56, 223-239.

Gamble, S.K., Turner, K. and Smythe, C. (1998). Application of the Focussed Rainfall Growth Estimation Technique in Tasmania, Hydro Electric Corporation, Tasmania, Internal Report.

Gelman, A., Carlin, J.B., Stern H.S., Rubin, D.B. (2004). Bayesian Data Analysis, Chapman and Hall/CRC, Boca Raton, FL.

Griffis, V.W. and Stedinger, J.R. (2007). The use of GLS regression in regional hydrologic analyses. Journal of Hydrology, 344, 82-95.

Gruber, A.M., Stedinger, J.R. (2008). Models of LP3 regional skew, data selection and Bayesian GLS regression. International World Environmental and Water Resources Congress, Ahupua'a.

Gruber, A.M., Reis, D.S. and Stedinger, J.R. (2007). Models of regional skew based on Bayesian GLS Regression. International World Environmental & Water Resources Conference, Tampa, Florida, May 15-18, 2007.

Gupta, V.K., Mesa, O.J. and Dawdy, D.R. (1994). Multiscaling theory of flood peaks: regional quantile analysis. Water Resources Research, 30, 12, 3405-3421.

Gupta, V.K. and Waymire, E.C. (1989). Statistical Self-Similarity in River Networks Parameterized by Elevation. Water Resources Research, 25, 3, 463-476.

Hackelbusch, A., Micevski, T., Kuczera, G., Rahman, A. and Haddad, K. (2009), Regional flood frequency analysis for eastern New South Wales: A region of influence approach using generalised least squares log-Pearson 3 parameter regression. 32nd Hydrology and Water Resources Symp., Newcastle, 30 Nov to 3 Dec, 603-615.

Haddad, K., Rahman, A., Ling, F. and Weinmann, P.E. (2011). Towards a new regional flood frequency analysis method for Tasmania, 34th IAHR World Congress, 26 June – 1 July 2011, Brisbane (Accepted).

Haddad, K., Rahman, A., Kuczera, G. and Micevski, T. (2011). Regional Flood Frequency Analysis in New South Wales Using Bayesian GLS Regression: Comparison of Fixed Region and Region-of-influence Approaches, 34th IAHR World Congress, 26 June–1 July 2011, Brisbane(Accepted).

Haddad, K. and Rahman, A. (2011). A Regional Flood Estimation Method Based on Generalised Least Squares Based Quantile Regression Technique for New South Wales Australia, Journal of Hydrologic Engineering, ASCE (Accepted).

Haddad, K. and Rahman, A. (2011). Regional Flood Frequency Analysis using Bayesian Generalized Least Squares: A Comparison between Quantile and Parameter Regression Techniques, Hydrological Processes (Under review).

Haddad, K. and Rahman, A. (2011). Application of GLS regression for regional flood methods in NSW, Australian Journal of Water Resources (Under review).

Haddad, K., Rahman, A. and Weinmann, P.E. (2011). Estimation of major floods: applicability of a simple probabilistic model, Australian Journal of Water Resources, 14, 2, 117-126.

Haddad, K., Rahman, A., Weinmann, P. E., Kuczera, G. and Ball, J. E. (2010) Streamflow data preparation for regional flood frequency analysis: Lessons from south-east Australia. Australian Journal of Water Resources, 14, 1, 17-32.

Haddad, K., Zaman, M. and Rahman, A. (2010). Regionalisation of skew for flood frequency analysis: a case study for eastern NSW. Australian Journal of Water Resources, 14, 1, 33-41.

Haddad, K. and Rahman, A. (2010). Selection of the best fit flood frequency distribution and parameter estimation procedure – A case study for Tasmania in Australia, Stochastic Environmental Research & Risk Assessment, (doi: 10.1007/s00477-010-0412-1).

Haddad, K., Pirozzi, J., McPherson, G. Rahman, A. and Kuczera, G. (2009). Regional Flood Estimation Technique for NSW: Application of Generalised Least Squares Quantile Regression Technique. In Proc. 32nd Hydrology and Water Resources Symposium, 30 Nov to 3 Dec, Newcastle, Australia, 829-840.

Haddad, K., Aziz, K.. Rahman, A. and Ishak, E.H. and Weinmann, P.E. (2009). A Probabilistic Model for Estimation of Large Floods in Ungauged Catchments: Application to South-east Australia. In Proc. 32nd Hydrology and Water Resources Symposium, 30 Nov to 3 Dec, Newcastle, Australia, 817-828.

Haddad, K., Pirozzi, J., McPherson, G., Rahman, A. and Kuczera, G. (2009). Regional flood estimation technique for NSW: Application of generalised least squares quantile regression technique. 32nd Hydrology and Water Resources Symposium, Newcastle, 30 Nov to 3 Dec, 829-840.

Haddad, K. and Rahman, A. (2008). Investigation on at-site flood frequency analysis in south-east Australia. IEM Journal, The Journal of The Institution of Engineers, Malaysia, 69(3), 59-64.

Haddad, K. Rahman, A. and Weinmann, P. E. (2008). Development of a Generalised Least Squares Based Quantile Regression Technique for design flood estimation in Victoria, 31st Hydrology and Water Resources Symposium, Adelaide, 15-17 April 2008, 2546-2557.

Haddad, K., Rahman, A. and Weinmann, P. E. (2008). Development of a generalised least squares based quantile regression technique for design flood estimation in Victoria. 31st Hydrology and Water Resources Symposium, Adelaide, 5-17 April, 2546-2557.

Haddad, K., Rahman, A. and Weinmann, E. (2006). Design flood estimation in ungauged catchments by quantile regression technique: ordinary least squares and generalised least squares compared. In Proc. 30th Hydrology and Water Resources Symposium, The Institution of Engineers Australia, 4-7 Dec 2006, Launceston, 6, ISBN 0-8582579-0-4.

Hamed, K.H. and Rao, A. R. (1998). A modified Mann-Kendall trend test for autocorrelated data. Journal of Hydrology, 204, 182–196.

Hennessy, K.J. Suppiah, R. and Page, C.M. (1999). Australian rainfall changes, 1910-1995. Australian Meteorological Magagazine, 48, 1-13.

Hua, C., Guo, S. Xu, Chong-yu. and Singh, V.P. (2007). Historical temporal trends of hydroclimatic variables and runoff response to climatic variability and their relevance in water resource management in the Hanjiang basin. Journal of Hydrology, 344, 171–184.

Institution of Engineers Australia (IEAust) (1987, 2001) Australian Rainfall and Runoff: A Guide to Flood Estimation. Pilgrim, D. H. (editor), Vol. 1, IEAust, Canberra.

IPCC (2007). The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC).

Ishak, I., Rahman, A. Westra, S. Sharma, A. and Kuczera, G. (2011). Trends in Peak Streamflow Data in Australia: Impacts of Serial and Cross-correlation, 34th IAHR World Congress, 26 June – 1 July 2011, Brisbane (Accepted).

Ishak, E. Haddad, K. Zaman, M. and Rahman, A. (2011). Scaling property of regional floods in New South Wales Australia, Natural Hazards (Accepted).

Ishak, E., Rahman, A., Wetra, S., Sharma, A. and Kuczera, G. (2010). Preliminary Analysis of Trends in Australian Flood Data World Environmental and Water Resources Congress, American Society of Civil Engineers (ASCE), 16-20 May 2010, Providence, Rhode Island, USA, (doi: 10.1061/41114(371)14).

Ishak, E.H., Aziz, K., Rahman, A. and Haddad, K. (2009). Scaling Behaviour of Regional Floods in New South Wales Australia. In Proc. 32nd Hydrology and Water Resources Symposium, 30 Nov to 3 Dec, Newcastle, Australia, 400-408.

Juckem, P.F., Hunt, R.J., Anderson, M. P. and Robertson, D. M. (2008). Effects of climate and land management change on streamflow in the driftless area of Wisconsin, Journal of Hydrology, 355, 123–130.

Khaliq, M.N., Ouarda, T.B.M.J., Gachon, P., Sushama, L. and St-Hilaire, A. (2009). Identification of hydrological trends in the presence of serial and cross-correlations: A review of selected methods and their application to annual flow regimes of Canadian rivers. Jounal of Hydrology, 368, 117-130.

Khaliq, M.N., Ouarda, T.B.M.J. Ondo, J.-C. Gachon, P. and Bobée, B. (2006). Frequency analysis of a sequence of dependent and/or non-stationary hydro- meteorological observations: A review. Journal of Hydrology, 329, 3-4, 534-552.

Kjeldsen, T.R. and Jones, D. (2009). An exploratory analysis of error components in hydrological regression modelling. Water Resources Research, 45, W02407, (doi:10.1029/2007WR006283).

Kjeldsen, T. R. and Jones, D. A. (2006). Prediction uncertainty in a median based index flood method using L-moments, Water Resources Research, 42, W07414, (doi:10.1029/2005WR004069).

Kottegoda, N.T. & Rosso, R. (1997). Statistics, Probability, and Reliability for Civil and Environmental Engineers, McGraw Hill, 735.

Kroll, C.N. and Stedinger, J.R. (1998). Regional hydrologic analysis: Ordinary and generalized least squares revisited, Water Resources Research, 34, 1, 121-128.

Kuczera, G. (1999). Comprehensive at-site flood frequency analysis using Monte Carlo Bayesian inference. Water Resources Research, 35, 5, 1551-1557.

Leclerc, M. and Ouarda T.B. (2007). Non-stationary regional flood frequency analysis at ungauged sites. Journal of Hydrology, 343, 254–265.

Lindstrom, G. and Bergstrom, S. (2004). Runoff trends in Sweden 1807-2002. Hydrological Sciences, 49, 1, 69-83.

Ma, Z., Kang, S., Zhang, L., Tong, T. and Su, X. (2008). Analysis of impacts of climate variability and human activity on streamflow for a river basin in arid region of northwest China. Journal of Hydrology, 352, 239-249.

Madsen, H., Mikkelsen, P.S., Rosbjerg, D. and Harremoes, P. (2002). Regional estimation of rainfall intensity duration curves using generalized least squares regression of partial duration series statistics. Water Resources Research, 38, 11, 1-11.

Madsen, H., Pearson, C.P. and Rosbjerg, D. (1997). Comparison of annual maximum series and partial duration series for modelling extreme hydrologic events, 2. Regional modelling. Water Resources Research, 33, 4, 759-769.

Madsen, H., Rosbjerg, D. and Harremoes, P. (1995). Application of the Bayesian approach in regional analysis of extreme rainfalls. Stochastic Hydrology and Hydraulics, 9, 77-88.

Majone, U., Tomirotti, M. and Galimberti, G. (2007). A probabilistic model for the estimation of peak flood flows. Special Session 10, 32nd Congress of IAHR, Venice, Italy, July 1-6.

Majone, U. and Tomirotti, M. (2004). A trans-national regional frequency analysis of peak flood flows. L'Aqua, 2, 9-17.

McConachy, F.L.N., Xuereb, K., Smythe, C.J. and Gamble, S.K. (2003). Homogeneity of Rare to Extreme Rainfalls over Tasmania. 28th International Hydrology and Water Resources Symposium, Wollongong, The Institution of Engineers, Australia.

Merz, R. and Blöschl, G. (2005). Flood frequency regionalisation–spatial proximity vs. Catchment attributes. Journal of Hydrology 302, 283–306.

Micevski, T. and Kuczera, G. (2009). Combining site and regional flood information using a Bayesian Monte Carlo approach. Water Resources Research, 45, W04405, doi: 10.1029/2008WR007173.

Micevski, T. Franks, S.W. and Kuczera, G. (2006). Multidecadal variability in coastal eastern Australian flood data. Journal of Hydrology, 327, 219-225.

Murphy, B.F. and Timbal, B. (2008). A review of recent climate variability and climate change in south-eastern Australia. International Journal of Climatology 28: 859–879.

Mulvany, T.J. (1851). On the use of self registering rain and flood gauges in making observations of the relation of rainfall and of flood discharge in a given catchment. Trans ICE Ire, 4, 18-31.

Nandakumar, N. Weinmann, P.E. Mein, R.G. and Nathan, R.J. (1997). Estimation of Extreme Rainfalls for Victoria Using the CRC-FORGE Method (For Rainfall Durations 24 to 72 Hours), Report 97, 4.

Nathan, R.J. and Weinmann, P.E. (2004). An improved framework for the characterisation of extreme floods and for the assessment of dam safety. In Proc. British Hydrological Society Intl. Confer., Imperial College London, 1, 186-193.

Nathan, R.J. and Weinmann, P.E. (1992). Practical Aspects of At-Site and Regional Flood Frequency Analyses. Civil Engineering Transactions, CE34, 3, 255-262.

Nathan, R.J. and Weinmann, P.E. (1991). Application of at-site and regional flood frequency analysis, Proceedings of the International Hydrology and Water Resources Symposium, 769–774.

Novotny, E.V. and Stefan, H.G. (2007). Stream flow in Minnesota: Indicator of climate change. Journal of Hydrology, 334, 319– 333.

Olsen, J.R., Lambert, J.H. and Haimes, Y.Y. (1998). Risk of extreme event under nonstationary conditions. Risk Analysis 18, 4, 497–510.

Pandey, G.R. and Nguyen, V.T.V. (1999). A comparative study of regression based methods in regional flood frequency analysis. Journal of Hydrology, 225, 92-101.

Parks, Y.P. and Sutcliffe, J.V. (1987). The development of hydrological yield assessment in NE Botswana, Brit. Hydrol. Soc. Symposium, 14-16 September, Hull, Brit. Hydrol. Soc. I.C.E.,London, 12.1-12.11.

Pasquini, A.I. and Depetris, P.J. (2007). Discharge trends and flow dynamics of South American rivers draining the southern Atlantic seaboard: An overview. Journal of Hydrology, 333, 385– 399.

Pegram, G. (2002). Rainfall, rational formula and regional maximum flood – some scaling links. 27th National Hydrology and Water Resources Symposium, Melbourne, 20-23 May.

Petrow, T. and Merz, B. (2009). Trends in flood magnitude, frequency and seasonality in Germany in the period 1951-2002. Journal of Hydrology, 371, 129–141.

Pilgrim, D.H. (1987). Estimation of peak flows for small to medium sized rural catchments. Book 4, Section 1, Australian Rainfall and Runoff, The Institution of Engineers Australia.

Pilgrim, D.H., Kennedy, M.R. Rowbottom, I.A. Cordery, I. Canterford, R.P. and Turner, L.H. (1987) Temporal Patterns of Rainfall Bursts, Australian Rainfall and Runoff, Engineers Australia.

Pilgrim, D. H. (1986) Bridging the gap between flood research and design practice. Water Resources Research, Vol. 22, Supplement, No. 9, 165S-176S.

Pirozzi, J. and Rahman, A. (2010). Design Streamflow Estimation for Ungauged Catchments in Eastern NSW: Identification of Important Predictor Variables, Australian Water Association National Conference, Ozwater 2010, 8-10 March, Brisbane.

Pirozzi, J., Ashraf, M., Rahman, A. and Haddad, K. (2009). Design Flood Estimation for Ungauged Catchments in Eastern NSW: Evaluation of the Probabilistic Rational Method. In Proc. 32nd Hydrology and Water Resources symposium, 30 Nov to 3 Dec, Newcastle, Australia, 805-816.

Plummer, N., Salinger, M.J., Nicholls, N., Suppiah, R., Hennessy, K.J., Leighton, R.M., Trewin, B., Page, C.M. and Lough, J.M. (1999). Changes in climate extremes over the Australian region and New Zealand during the twentieth century. Climatic Change, 42, 183-202.

Power, S. Casey, T. Folland, C. Colman, A. and Mehta, V. (1999). Inter-decadal modulation of the impact of ENSO on Australia. Clim. Dynamics, 15, 319–324.

Prudhomme, C. Jakob, D. and Svensson, C. (2003). Uncertainty and climate change impact on the flood regime of small UK catchments. Journal of Hydrology, 277, 1-23.

Rahman, A., Haddad, K., Zaman, M., Kuczera, G. and Weinmann, P.E. (2011). Design flood estimation in ungauged catchments: A comparison between the Probabilistic Rational Method and Quantile Regression Technique for NSW. Australian Journal of Water Resources, 14, 2, 127-140.

Rahman, A., Zaman, M., Fotos, M., Haddad, K., Rajaratnam, L. and Weeks, B. (2011). Towards a New Regional Flood Estimation Method for the Northern Territory, 34th IAHR World Congress, 26 June–1 July 2011, Brisbane (Accepted).

Rahman, A., Haddad, K., Ishak, E., Weinmann, P.E. and Kuczera, G. (2010). Regional Flood Estimation in Australia: An Overview of the Study in Relation to the Upgrade of Australian

Rainfall and Runoff. 50th Annual Floodplain Management Authorities Conference Gosford 2010 FMA, 23-29 Feb, Gosford, NSW, 2010.

Rahman, A., Haddad, K., Kuczera, G. and Weinmann, P. E. (2009). Regional flood methods for Australia: data preparation and exploratory analysis, ARR Revision Project 5 Stage I Report, Engineers Australia, Report No. P5/S1/003, 179.

Rahman, A., Rima, K. and Weeks, W. (2008). Development of regional flood estimation methods using quantile regression technique: A case study for North-eastern part of Queensland. 31st Hydrology and Water Resources Symposium, Adelaide, 15-17 April, 329-340.

Reis, Jr.D.S., Stedinger, J.R. and Martins, E.S. (2005). Bayesian GLS regression with application to LP3 regional skew estimation. Water Resources Research, 41, W10419, doi:10.1029/2004WR00344.

Reis, D.S., Stedinger, J.R. and Martins E.S. (2003). Bayesian GLS regression with application to LP3 regional skew estimation, In: Bizier, P., DeBarry P. (Eds.), Proceedings of the World Water and Environmental Resources Congress, Philadelphia, PA, June 23-26, 2003. American Society of Civil Engineers, Reston, VA.

Rowbottom, I.A., Pilgrim, D.H. and Wright, G. (1986). Estimation of rare floods (between the Probable Maximim Floof and the I-in0100 flood). Trans. Instn Engrs, Australia. CE28, 92-108.

Salas, J.D. Delleur, J.W. Yevjevich, V. and Lane, W. L. (1980) Applied Modelling of Hydrologic Time Series. Water Resources Publications, Littleton, Colorado, USA.

Smith, I. (2004). An Assessment of Recent Trends in Australian Rainfall. Australian Meteorological Magazine, 53, 163-173.

Stedinger, J.R. and Lu., L.H. (1995). Appraisal of regional and index flood quantiles estimators. Stochastic Hydrology and Hydraulics, 9, 1, 49-75.

Stedinger, J.R., Vogel, R.M. and Foufoula-Georgiou, E. (1992). Frequency analysis of extreme events. In: Handbook of Hydrology (ed. by D. R. Maidment), Chapter 18. McGraw-Hill, New York, USA.

Stedinger, J.R. and Tasker, G.D. (1986a). Correction to Regional hydrologic analysis, 1.Ordinary, weighted, and generalised least squares compared. Water Resources Research, 22, 5, 844.

Stedinger J.R. and Tasker, G.D. (1986b). Regional hydrologic analysis, 2. Model error estimators, estimation of sigma and log – Pearson type 3 distributions. Water Resources Research, 22, 10, 1487-1499.

Stedinger, J.R. and Tasker G.D. (1985). Regional hydrologic analysis, 1.Ordinary, weighted, and generalised least squares compared, Water Resources Research, 21, 9, 1421-1432.

Strupczewski, W.G., Singh, V.P. and W. Feluch (2001a), Non-stationarity approach to at-site flood frequency modeling I. Maximum likelihood estimation, Journal of Hydrology, 248, 123–142.

Strupczewski, W.G., Singh, V.P. and Mitosek, H.T. (2001b). Non-stationary approach to at-site flood frequency modelling. III. Flood analysis of Polish rivers. Journal of Hydrology, 248, 123–142.

Taschetto, A.S. and England, M.H. (2008). An analysis of late twentieth century trends in Australian rainfall. International Journal of Climatology 29, 791–807.

Tasker, G.D., Hodge, S.A. and Barks, C.S. (1996). Region of influence regression for estimating the 50-year flood at ungauged sites. Water Resources Bulletin, American Water Resources Association, 32, 1, 63-170.

Tasker, G.D. and Stedinger, J.R. (1989). An operational GLS model for hydrologic regression. Journal of Hydrology, 111, 361-375.

Tasker, G.D. and Stedinger, J.R. (1986). Estimating generalized skew with weighted least squares regression. Journal of Water Resources Planning and Management, 112, 2, 225–237.

Tasker, G.D., Eychaner, J.H. and Stedinger, J.R. (1986). Application of generalised least squares in regional hydrologic regression analysis. US Geological Survey Water Supply Paper, 2310, 107-115.

Thomas, D.M., and Benson, M.A. (1970). Generalization of streamflow characteristics from drainage basin characteristics. US Geological Survey Water Supply Paper, No.1975.

UNESCO (1990). Flash floods in the arid and semi-arid zones. International Hydrology Program, Technical Documents in Hydrology, No. 23, 60.

Villarini, G. Serinaldi, F. Smith, J.A. and Krajewski, W.F. (2009). On the stationarity of annual flood peaks in the continental United States during the 20th century. Water Resource Research, 45, W08417, doi: 10.1029/2008WR007645.

Von Storch, H. (1995). Misuses of statistical analysis in climate research. In: Analysis of Climate Variability: Applications of Statistical Techniques (ed. by H. von Storch & A. Navarra), 11–26. Springer-Verlag, Berlin, Germany.

Weeks, W.D. (1991). Design floods for small rural catchments in Queensland. Civil Engineering Transactions, IEAust, Vol. CE33, 4, 249-260.

Whitfield, P.H. (2001). Linked hydrologic and climatic variations in British Columbia and Yukon. Environmental Monitoring and Assessment, 67, 1-2, 217-238.

Xuereb, K.C. Moore, G.J. and Taylor, B.F. (2001). Development of the Method of Storm Transposition and Maximisation for the West Coast of Tasmania. HRS Report No. 7.

Yue, S.P., Pilon, P.J. and Phinney, B. (2003). Canadian streamflow trend detection: Impacts of serial and cross-correlation. Hydrological Sciences Journal, 48, 1, 51–63.

Yue, S., Pilon, P.J., Phinney, B. and Cavadias, G. (2002). The influence of autocorrelation on the ability to detect trend in hydrological series. Hydrological Processes, 16, 1807–1829.

Zaman, M., Rahman, I., Haddad, K. and Rahman, A. (2010). Scaling issues in design flood estimation for ungauged catchments: A case study for eastern Australia. World Environmental and Water Resources Congress 2010, American Society of Civil Engineers (ASCE), 16-20 May 2010, Providence, Rhode Island, USA, 2860-2869.

Zhang, S. and Lu, X.X. (2009). Hydrological responses to precipitation variation and diverse human activities in amountainous tributary of the lower Xijiang, China. Catena, 77, 130-142.

Zrinji, Z. And Burn, D.H. (1994). Flood frequency analysis for unguaged sites using a region of influence approach. Journal of Hydrology, 153, 1–21.

Zhang, Q., Liu, C., Xu, C.Y., Xu, Y.P. and Jiang, T. (2006). Observed trends of water level and streamflow during past 100 years in the Yangtze River basin, China. Journal of Hydrology, 324, 1-4, 255-265

Zhang, X., Harvey, K.D., Hogg, W.D. and Yuzyk, T.R. (2001). Trends in Canadian streamflow. Water Resources Research, 37, 4, 987–998.

Zhu, Y. and Day, R.L. (2005). Analysis of Streamflow Trends and the Effects of Climate in Pennsylvania, 1971 to 2001. Journal of the American Water Resources Association, 41, 6, 1393-1405.

Appendices

Table A2 Selected catchments from New South Wales

Station ID	Station Name	River Name	Lat (°S)	Long (°E)	Area (km^2)	Record Length (years)	Period of Record
201001	Eungella	Oxley	-28.36	153.29	213	49	1958 - 2006
203002	Repentance	Coopers Ck	-28.64	153.41	62	30	1977 - 2006
203012	Binna Burra	Byron Ck	-28.71	153.50	39	29	1978 - 2006
203030	Rappville	Myrtle Ck	-29.11	153.00	332	27	1980 - 2006
204025	Karangi	Orara	-30.26	153.03	135	37	1970 - 2006
204026	Bobo Nursery	Bobo	-30.25	152.85	80	29	1956 - 1984
204030	Aberfoyle	Aberfoyle	-30.26	152.01	200	29	1978 - 2006
204036	Sandy Hill(below Snake Cre	Cataract Ck	-28.93	152.22	236	54	1953 - 2006
204037	Clouds Ck	Clouds Ck	-30.09	152.63	62	35	1972 - 2006
204056	Gibraltar Range	Dandahra Ck	-29.49	152.45	104	31	1976 - 2006
204906	Glenreagh	Orara	-30.07	152.99	446	34	1973 - 2006
206009	Tia	Tia	-31.19	151.83	261	53	1955 - 2007
206025	near Dangar Falls	Salisbury Waters	-30.68	151.71	594	34	1973 - 2006
206026	Newholme	Sandy Ck	-30.42	151.66	8	33	1975 - 2007
207006	Birdwood(Filly Flat)	Forbes	-31.39	152.33	363	32	1976 - 2007
208001	Bobs Crossing	Barrington	-32.03	151.47	20	52	1955 - 2006
209001	Monkerai	Karuah	-32.24	151.82	203	34	1946 - 1979
209002	Crossing	Mammy Johnsons	-32.25	151.98	156	31	1976 - 2006
209003	Booral	Karuah	-32.48	151.95	974	38	1969 - 2006
209006	Willina	Wang Wauk	-32.16	152.26	150	36	1970 - 2005
209018	Dam Site	Karuah	-32.28	151.90	300	27	1980 - 2006
210011	Tillegra	Williams	-32.32	151.69	194	75	1932 - 2006
210014	Rouchel Brook (The Vale)	Rouchel Brook	-32.15	151.05	395	42	1960 - 2001
210017	Moonan Brook	Moonan Brook	-31.94	151.28	103	67	1941 - 2007

Station ID	Station Name	River Name	Lat (°S)	Long (°E)	Area (km^2)	Record Length (years)	Period of Record
210022	Halton	Allyn	-32.31	151.51	205	65	1941 - 2005
210040	Wybong	Wybong Ck	-32.27	150.64	676	50	1956 - 2005
210042	Ravensworth	Foy Brook	-32.40	151.05	170	30	1967 - 1996
210044	Middle Falbrook(Fal Dam Si	Glennies Ck	-32.45	151.15	466	51	1957 - 2007
210068	Pokolbin Site 3	Pokolbin Ck	-32.80	151.33	25	41	1965 - 2005
210076	Liddell	Antiene Ck	-32.34	150.98	13	37	1969 - 2005
210079	Gostwyck	Paterson	-32.55	151.59	956	33	1975 - 2007
210080	U/S Glendon Brook	West Brook	-32.47	151.28	80	31	1977 - 2007
211009	Gracemere	Wyong	-33.27	151.36	236	35	1973 - 2007
211013	U/S Weir	Ourimbah Ck	-33.35	151.34	83	30	1977 - 2006
212008	Bathurst Rd	Coxs	-33.43	150.08	199	55	1952 - 2006
212018	Glen Davis	Capertee	-33.12	150.28	1010	35	1972 - 2006
212040	Pomeroy	Kialla Ck	-34.61	149.54	96	27	1980 - 2004
213005	Briens Rd	Toongabbie Ck	-33.80	150.98	70	27	1980 - 2006
215004	Hockeys	Corang	-35.15	150.03	166	75	1930 - 2004
218002	Belowra	Tuross	-36.20	149.71	556	29	1955 - 1983
218005	D/S Wadbilliga R Junct	Tuross	-36.20	149.76	900	42	1965 - 2006
218007	Wadbilliga	Wadbilliga	-36.26	149.69	122	33	1975 - 2005
219003	Morans Crossing	Bemboka	-36.67	149.65	316	64	1944 - 2007
219017	near Brogo	Double Ck	-36.60	149.81	152	41	1967 - 2007
219022	Candelo Dam Site	Tantawangalo Ck	-36.73	149.68	202	36	1972 - 2007
219025	Angledale	Brogo	-36.62	149.88	717	30	1977 - 2006
220001	New Buildings Br	Towamba	-36.96	149.56	272	26	1955 - 1980
220003	Lochiel	Pambula	-36.94	149.82	105	41	1967 - 2005

Station ID	Station Name	River Name	Lat (°S)	Long (°E)	Area (km^2)	Record Length (years)	Period of Record
220004	Towamba	Towamba	-37.07	149.66	745	37	1971 - 2007
221002	Princes HWY	Wallagaraugh	-37.37	149.71	479	36	1972 - 2007
222004	Wellesley (Rowes)	Little Plains	-37.00	149.09	604	65	1942 - 2006
222007	Woolway	Wullwye Ck	-36.42	148.91	520	57	1950 - 2006
222009	The Falls	Bombala	-36.92	149.21	559	43	1952 - 1994
222015	Jacobs Ladder	Jacobs	-36.73	148.43	187	27	1976 - 2002
222016	The Barry Way	Pinch	-36.79	148.40	155	31	1976 - 2006
222017	The Hut	Maclaughlin	-36.66	149.11	313	28	1979 - 2006
401009	Maragle	Maragle Ck	-35.93	148.10	220	56	1950 - 2005
401013	Jingellic	Jingellic Ck	-35.90	147.69	378	33	1973 - 2005
401015	Yambla	Bowna Ck	-35.92	146.98	316	31	1975 - 2005
410038	Darbalara	Adjungbilly Ck	-35.02	148.25	411	37	1969 - 2005
410048	Ladysmith	Kyeamba Ck	-35.20	147.51	530	48	1939 - 1986
410057	Lacmalac	Goobarragandra	-35.33	148.35	673	49	1958 - 2006
410061	Batlow Rd	Adelong Ck	-35.33	148.07	155	60	1948 - 2007
410062	Numeralla School	Numeralla	-36.18	149.35	673	43	1965 - 2007
410076	Jerangle Rd	Strike-A-Light C	-35.92	149.24	212	31	1975 - 2005
410088	Brindabella (No.2&No.3-Cab	Goodradigbee	-35.42	148.73	427	38	1968 - 2005
410112	Jindalee	Jindalee Ck	-34.58	148.09	14	30	1976 - 2005
410114	Wyangle	Killimcat Ck	-35.24	148.31	23	30	1977 - 2006
411001	Bungendore	Mill Post Ck	-35.28	149.39	16	25	1960 - 1984
411003	Butmaroo	Butmaroo Ck	-35.26	149.54	65	28	1979 - 2006
412050	Narrawa North	Crookwell	-34.31	149.17	740	34	1970 - 2003
412063	Gunning	Lachlan	-34.74	149.29	570	39	1961 - 1999

Station ID	Station Name	River Name	Lat (°S)	Long (°E)	Area (km^2)	Record Length (years)	Period of Record
412081	near Neville	Rocky Br Ck	-33.80	149.19	145	33	1969 - 2001
412083	Tuena	Tuena Ck	-34.02	149.33	321	33	1969 - 2001
416003	Clifton	Tenterfield Ck	-29.03	151.72	570	28	1979 - 2006
416008	Haystack	Beardy	-29.22	151.38	866	35	1972 - 2006
416016	Inverell (Middle Ck)	Macintyre	-29.79	151.13	726	35	1972 - 2006
416020	Coolatai	Ottleys Ck	-29.23	150.76	402	28	1979 - 2006
416023	Bolivia	Deepwater	-29.29	151.92	505	28	1979 - 2006
418005	Kimberley	Copes Ck	-29.92	151.11	259	35	1972 - 2006
418014	Yarrowyck	Gwydir	-30.47	151.36	855	37	1971 - 2007
418017	Molroy	Myall Ck	-29.80	150.58	842	29	1979 - 2007
418021	Laura	Laura Ck	-30.23	151.19	311	29	1978 - 2006
418025	Bingara	Halls Ck	-29.94	150.57	156	28	1980 - 2007
418027	Horton Dam Site	Horton	-30.21	150.43	220	36	1972 - 2007
418034	Black Mountain	Boorolong Ck	-30.30	151.64	14	29	1976 - 2004
419010	Woolbrook	Macdonald	-30.97	151.35	829	28	1980 - 2007
419016	Mulla Crossing	Cockburn	-31.06	151.13	907	33	1974 - 2006
419029	Ukolan	Halls Ck	-30.71	150.83	389	27	1979 - 2005
419051	Avoca East	Maules Ck	-30.50	150.08	454	31	1977 - 2007
419053	Black Springs	Manilla	-30.42	150.65	791	33	1975 - 2007
419054	Limbri	Swamp Oak Ck	-31.04	151.17	391	33	1975 - 2007
420003	Warkton (Blackburns)	Belar Ck	-31.39	149.20	133	30	1976 - 2005
421026	Sofala	Turon	-33.08	149.69	883	34	1974 - 2007
421036	below Dam Site	Duckmaloi	-33.75	149.94	112	25	1956 - 1980
421050	Molong	Bell	-33.03	148.95	365	33	1975 - 2007

Table A2 Selected catchments from Victoria

Station ID	Station Name	River Name	Lat (°S)	Long (°E)	Area (km^2)	Record Length (years)	Period of Record
221207	Errinundra	Errinundra	-37.45	148.91	158	35	1971 - 2005
221209	Weeragua	Cann(East Branch	-37.37	149.20	154	33	1973 - 2005
221210	The Gorge	Genoa	-37.43	149.53	837	34	1972 - 2005
221211	Combienbar	Combienbar	-37.44	148.98	179	32	1974 - 2005
221212	Princes HWY	Bemm	-37.61	148.90	725	31	1975 - 2005
222202	Sardine Ck	Brodribb	-37.51	148.55	650	41	1965 - 2005
222206	Buchan	Buchan	-37.50	148.18	822	32	1974 - 2005
222210	Deddick (Caseys)	Deddick	-37.09	148.43	857	35	1970 - 2005
222213	Suggan Buggan	Suggan Buggan	-36.95	148.33	357	35	1971 - 2005
222217	Jacksons Crossing	Rodger	-37.41	148.36	447	30	1976 - 2005
223202	Swifts Ck	Tambo	-37.26	147.72	943	32	1974 - 2005
223204	Deptford	Nicholson	-37.60	147.70	287	32	1974 - 2005
224213	Lower Dargo Rd	Dargo	-37.50	147.27	676	33	1973 - 2005
224214	Tabberabbera	Wentworth	-37.50	147.39	443	32	1974 - 2005
225213	Beardmore	Aberfeldy	-37.85	146.43	311	33	1973 - 2005
225218	Briagalong	Freestone Ck	-37.81	147.09	309	35	1971 - 2005
225219	Glencairn	Macalister	-37.52	146.57	570	39	1967 - 2005
225223	Gillio Rd	Valencia Ck	-37.73	146.98	195	35	1971 - 2005
225224	The Channel	Avon	-37.80	146.88	554	34	1972 - 2005
226204	Willow Grove	Latrobe	-38.09	146.16	580	35	1971 - 2005
226205	Noojee	Latrobe	-37.91	146.02	290	46	1960 - 2005
226209	Darnum	Moe	-38.21	146.00	214	34	1972 - 2005
226217	Hawthorn Br	Latrobe	-37.98	146.08	440	34	1955 - 1988
226218	Thorpdale	Narracan Ck	-38.27	146.19	66	35	1971 - 2005

Station ID	Station Name	River Name	Lat (°S)	Long (°E)	Area (km^2)	Record Length (years)	Period of Record
226222	Near Noojee (U/S Ada R Jun	Latrobe	-37.88	145.89	62	35	1971 - 2005
226226	Tanjil Junction	Tanjil	-38.01	146.20	289	46	1960 - 2005
226402	Trafalgar East	Moe Drain	-38.18	146.21	622	31	1975 - 2005
227200	Yarram	Tarra	-38.46	146.69	25	41	1965 - 2005
227205	Calignee South	Merriman Ck	-38.36	146.65	36	31	1975 - 2005
227210	Carrajung Lower	Bruthen Ck	-38.40	146.74	18	33	1973 - 2005
227211	Toora	Agnes	-38.64	146.37	67	32	1974 - 2005
227213	Jack	Jack	-38.53	146.53	34	36	1970 - 2005
227219	Loch	Bass	-38.38	145.56	52	32	1973 - 2004
227225	Fischers	Tarra	-38.47	146.56	16	33	1973 - 2005
227226	Dumbalk North	Tarwineast Branc	-38.50	146.16	127	36	1970 - 2005
227231	Glen Forbes South	Bass	-38.47	145.51	233	32	1974 - 2005
227236	D/S Foster Ck Jun	Powlett	-38.56	145.71	228	27	1979 - 2005
228212	Tonimbuk	Bunyip	-38.03	145.76	174	30	1975 - 2004
228217	Pakenham	Toomuc Ck	-38.07	145.46	41	29	1974 - 2002
229218	Watsons Ck	Watsons Ck	-37.67	145.26	36	26	1974 - 1999
230202	Sunbury	Jackson Ck	-37.58	144.74	337	31	1975 - 2005
230204	Riddells Ck	Riddells Ck	-37.47	144.67	79	32	1974 - 2005
230205	Bulla (D/S of Emu Ck Jun)	Deep Ck	-37.63	144.80	865	32	1974 - 2005
230211	Clarkefield	Emu Ck	-37.47	144.75	93	31	1975 - 2005
231200	Bacchus Marsh	Werribee Ck	-37.68	144.43	363	28	1978 - 2005
231213	Sardine Ck- O'Brien Cro	Lerderderg Ck	-37.50	144.36	153	47	1959 - 2005
231225	Ballan (U/S Old Western H)	Werribee Ck	-37.60	144.25	71	33	1973 - 2005
231231	Melton South	Toolern Ck	-37.91	144.58	95	27	1979 - 2005

Station ID	Station Name	River Name	Lat (°S)	Long (°E)	Area (km^2)	Record Length (years)	Period of Record
232200	Little	Little Ck	-37.96	144.48	417	32	1974 - 2005
232210	Lal Lal	Mooraboolwest Br	-37.65	144.04	83	33	1973 - 2005
232213	U/S of Bungal Dam	Lal Lal Ck	-37.66	144.03	157	29	1977 - 2005
233211	Ricketts Marsh	Birregurra Ck	-38.30	143.84	245	31	1975 - 2005
233214	Forrest (above Tunnel)	Barwoneast Branc	-38.53	143.73	17	28	1978 - 2005
234200	Pitfield	Woady Yaloak	-37.81	143.59	324	34	1972 - 2005
235202	Upper Gellibrand	Gellibrand	-37.56	143.64	53	31	1975 - 2005
235203	Curdie	Curdies	-38.45	142.96	790	31	1975 - 2005
235204	Beech Forest	Little Aire Ck	-38.66	143.53	11	30	1976 - 2005
235205	Wyelangta	Arkins Ck West B	-38.65	143.44	3	28	1978 - 2005
235227	Bunkers Hill	Gellibrand	-38.53	143.48	311	32	1974 - 2005
235233	Apollo Bay- Paradise	Barhameast Branc	-38.76	143.62	43	29	1977 - 2005
235234	Gellibrand	Love Ck	-38.49	143.57	75	27	1979 - 2005
236205	Woodford	Merri	-38.32	142.48	899	32	1974 - 2005
236212	Cudgee	Brucknell Ck	-38.35	142.65	570	31	1975 - 2005
237207	Heathmere	Surry	-38.25	141.66	310	31	1975 - 2005
238207	Jimmy Ck	Wannon	-37.37	142.50	40	32	1974 - 2005
238219	Morgiana	Grange Burn	-37.71	141.83	997	33	1973 - 2005
401208	Berringama	Cudgewa Ck	-36.21	147.68	350	41	1965 - 2005
401209	Omeo	Livingstone Ck	-37.11	147.57	243	27	1968 - 1994
401210	below Granite Flat	Snowy Ck	-36.57	147.41	407	38	1968 - 2005
401212	Upper Nariel	Nariel Ck	-36.45	147.83	252	52	1954 - 2005
401215	Uplands	Morass Ck	-36.87	147.70	471	35	1971 - 2005
401216	Jokers Ck	Big	-36.95	141.47	356	52	1952 - 2005

Station ID	Station Name	River Name	Lat (°S)	Long (°E)	Area (km^2)	Record Length (years)	Period of Record
401217	Gibbo Park	Gibbo	-36.75	147.71	389	35	1971 - 2005
401220	McCallums	Tallangatta Ck	-36.21	147.50	464	30	1976 - 2005
402203	Mongans Br	Kiewa	-36.60	147.10	552	36	1970 - 2005
402204	Osbornes Flat	Yackandandah Ck	-36.31	146.90	255	39	1967 - 2005
402206	Running Ck	Running Ck	-36.54	147.05	126	31	1975 - 2005
402217	Myrtleford Rd Br	Flaggy Ck	-36.39	146.88	24	36	1970 - 2005
403205	Bright	Ovens Rivers	-36.73	146.95	495	35	1971 - 2005
403209	Wangaratta North	Reedy Ck	-36.33	146.34	368	33	1973 - 2005
403213	Greta South	Fifteen Mile Ck	-36.62	146.24	229	33	1973 - 2005
403221	Woolshed	Reedy Ck	-36.31	146.60	214	30	1975 - 2004
403222	Abbeyard	Buffalo	-36.91	146.70	425	33	1973 - 2005
403224	Bobinawarrah	Hurdle Ck	-36.52	146.45	158	31	1975 - 2005
403226	Angleside	Boggy Ck	-36.61	146.36	108	32	1974 - 2005
403227	Cheshunt	King	-36.83	146.40	453	33	1973 - 2005
403233	Harris Lane	Buckland	-36.72	146.88	435	34	1972 - 2005
404206	Moorngag	Broken	-36.80	146.02	497	33	1973 - 2005
404207	Kelfeera	Holland Ck	-36.61	146.06	451	31	1975 - 2005
405205	Murrindindi above Colwells	Murrindindi	-37.41	145.56	108	31	1975 - 2005
405209	Taggerty	Acheron	-37.32	145.71	619	33	1973 - 2005
405212	Tallarook	Sunday Ck	-37.10	145.05	337	31	1975 - 2005
405214	Tonga Br	Delatite	-37.15	146.13	368	49	1957 - 2005
405215	Glen Esk	Howqua	-37.23	146.21	368	32	1974 - 2005
405217	Devlins Br	Yea	-37.38	145.48	360	31	1975 - 2005
405218	Gerrang Br	Jamieson	-37.29	146.19	368	47	1959 - 2005

Station ID	Station Name	River Name	Lat (°S)	Long (°E)	Area (km^2)	Record Length (years)	Period of Record
405219	Dohertys	Goulburn	-37.33	146.13	694	39	1967 - 2005
405226	Moorilim	Pranjip Ck	-36.62	145.31	787	32	1974 - 2005
405227	Jamieson	Big Ck	-37.37	146.06	619	36	1970 - 2005
405229	Wanalta	Wanalta Ck	-36.64	144.87	108	36	1969 - 2005
405230	Colbinabbin	Cornella Ck	-36.61	144.80	259	33	1973 - 2005
405231	Flowerdale	King Parrot Ck	-37.35	145.29	181	32	1974 - 2005
405237	Euroa Township	Seven Creeks	-36.76	145.58	332	33	1973 - 2005
405240	Ash Br	Sugarloaf Ck	-37.06	145.05	609	33	1973 - 2005
405241	Rubicon	Rubicon	-37.29	145.83	129	33	1973 - 2005
405245	Mansfield	Ford Ck	-37.04	146.05	115	36	1970 - 2005
405248	Graytown	Major Ck	-36.86	144.91	282	35	1971 - 2005
405251	Ancona	Brankeet Ck	-36.97	145.78	121	33	1973 - 2005
405263	U/S of Snake Ck Jun	Goulburn	-37.46	146.25	327	31	1975 - 2005
405264	D/S of Frenchman Ck Jun	Big	-37.52	146.08	333	31	1975 - 2005
405274	Yarck	Home Ck	-37.11	145.60	187	29	1977 - 2005
406213	Redesdale	Campaspe	-37.02	144.54	629	30	1975 - 2004
406214	Longlea	Axe Ck	-36.78	144.43	234	34	1972 - 2005
406215	Lyal	Coliban	-36.96	144.49	717	32	1974 - 2005
406216	Sedgewick	Axe Ck	-36.90	144.36	34	26	1975 - 2005
406224	Runnymede	Mount Pleasant C	-36.55	144.64	248	30	1975 - 2004
406226	Derrinal	Mount Ida Ck	-36.88	144.65	174	28	1978 - 2005
407214	Clunes	Creswick Ck	-37.30	143.79	308	31	1975 - 2005
407217	Vaughan atD/S Fryers Ck	Loddon	-37.16	144.21	299	38	1968 - 2005
407220	Norwood	Bet Bet Ck	-37.00	143.64	347	33	1973 - 2005

Station ID	Station Name	River Name	Lat (°S)	Long (°E)	Area (km^2)	Record Length (years)	Period of Record
407221	Yandoit	Jim Crow Ck	-37.21	144.10	166	33	1973 - 2005
407222	Clunes	Tullaroop Ck	-37.23	143.83	632	33	1973 - 2005
407230	Strathlea	Joyces Ck	-37.17	143.96	153	33	1973 - 2005
407246	Marong	Bullock Ck	-36.73	144.13	184	33	1973 - 2005
407253	Minto	Piccaninny Ck	-36.45	144.47	668	33	1973 - 2005
415207	Eversley	Wimmera	-37.19	143.19	304	31	1975 - 2005
415217	Grampians Rd Br	Fyans Ck	-37.26	142.53	34	33	1973 - 2005
415220	Wimmera HWY	Avon	-36.64	142.98	596	32	1974 - 2005
415226	Carrs Plains	Richardson	-36.75	142.79	130	31	1971 - 2001
415237	Stawell	Concongella Ck	-37.02	142.82	239	29	1977 - 2005
415238	Navarre	Wattle Ck	-36.90	143.10	141	30	1976 - 2005

Table A3 Selected catchments from South Australia

Station ID	Station Name	River Name	Lat (°S)	Long (°E)	Area (km^2)	Record Length (years)	Period of Record
A4260504	4km East of Yundi	Finniss River	-35.32	138.67	191	38	1971-2008
A4260529	Cambrai	Marne River upstream	-34.68	139.23	239	33	1974-2006
A4260533	Hartley	Bremer River	-35.21	139.01	473	34	1975-2008
A4260536	Worlds End	Burra Creek	-33.84	139.09	704	31	1975-2005
A4260557	Mount Barker	Mount Barker Creek dowstream	-35.09	138.92	88	28	1981-2008
A4260558	Dawesley	Dawesley Creek	-35.04	138.95	43	29	1980-2008
A5020502	Dam And Road Bridge	Myponga River upstream	-35.38	138.48	76.5	29	1980-2008
A5030502	Scott Bottom	Scott Creek	-35.1	138.68	26.8	38	1971-2008
A5030503	4.5km WNW Kangarilla	Baker Gully	-35.14	138.61	48.7	38	1971-2008
A5030504	Houlgrave	Onkaparinga River	-35.08	138.73	321	34	1975-2008
A5030506	Mount Bold Reservoir	Echunga Creek upstream	-35.13	138.73	34.2	34	1975-2008
A5030507	Lenswood	Lenswood Creek	-34.94	138.82	16.5	35	1974-2008
A5030508	Craigbank	Inverbrackie Creek	-34.95	138.93	8.4	35	1974-2008
A5030509	Aldgate Railway Station	Aldgate Ck	-35.02	138.73	7.8	35	1974-2008
A5030526	Uraidla	Cox Creek	-34.97	138.74	4.3	31	1978-2008
A5030529	Mount Bold Reservoir	Burnt Out Creek upstream	-35.13	138.71	0.6	29	1980-2008
A5040500	Gumeracha Weir	River Torrens	-34.82	138.85	194	67	1942-2008
A5040512	Mount Pleasant	Torrens River	-34.79	139.03	26	34	1975-2008
A5040517	Waterfall Gully	First Creek	-34.97	138.68	5	28	1978-2005
A5040518	Minno Creek Junction	Sturt River upstream M	-35.04	138.63	19	30	1979-2008
A5040523	Castambul	Sixth Creek	-34.87	138.76	44	30	1979-2008
A5040525	Millbrook Reservoir	Kersbrook Ck upstream	-34.81	138.84	23	18	1991-2008
A5050502	Yaldara	North Para River	-34.57	138.88	384	64	1945-2008

Station ID	Station Name	River Name	Lat (°S)	Long (°E)	Area (km^2)	Record Length (years)	Period of Record
A5050504	Turretfield	North Para River	-34.56	138.77	708	35	1974-2008
A5050517	Penrice	North Para River	-34.46	139.06	118	30	1979-2008
A5060500	Rhynie	Wakefield River	-34.1	138.63	417	67	1943-2009
A5070500	Andrews	Hill River	-33.61	138.63	235	38	1971-2008
A5070501	Spalding	Hutt River	-33.54	138.6	280	37	1971-2007
A5130501	Gorge Falls	Rocky River upstream	-35.96	136.7	190	34	1975-2008

Table A4 Selected catchments from Tasmania

Station ID	Station Name	River Name	Lat (°S)	Long (°E)	Area (km^2)	Record Length (years)	Period of Record
76	at Ballroom Offtake	North Esk	-41.50	147.39	335.0	74	1923 - 1996
159	D/S Rapid	Arthur	-41.12	145.08	1600.0	42	1955 - 1996
473	D/S Crossing Rv	Davey	-43.14	145.95	680.0	34	1964 - 1997
499	at Newbury	Tyenna	-42.71	146.71	198.0	33	1965 - 1997
852	at Strathbridge	Meander	-41.49	146.91	1025.0	24	1985 - 2008
1012	3.5 Km U/S Esperance	Peak Rivulet	-43.32	146.90	35.0	23	1975 - 1997
1200	at Whitemark Water Supply	South Pats	-40.09	148.02	21.0	22	1969 - 1990
2200	at The Grange	Swan	-42.05	148.07	440.0	33	1964 - 1996
2204	U/S Coles Bay Rd Bdg	Apsley	-41.94	148.24	157.0	24	1969 - 1992
2206	U/S Scamander Water Supply	Scamander	-41.45	148.18	265.0	28	1969 - 1996
2207	3 Km U/S Tasman Hwy	Little Swanport	-42.34	147.90	600.0	19	1971 - 1989
2208	at Swansea	Meredith	-42.12	148.04	88.0	27	1970 - 1996
2209	Tidal Limit	Carlton	-42.87	147.70	136.0	28	1969 - 1996
2211	U/S Brinktop Rd	Orielton Rivulet	-42.76	147.54	46.0	24	1973 - 1996
2213	D/S McNeils Rd	Goatrock Ck	-42.14	147.92	1.3	22	1975 - 1996
3203	at Baden	Coal	-42.43	147.45	55.0	26	1971 - 1996
4201	at Mauriceton	Jordan	-42.53	147.12	730.0	36	1966 - 2001
5200	at Summerleas Rd Br	Browns	-42.96	147.27	15.0	30	1963 - 1992
6200	D/S Grundys Ck	Mountain	-42.94	147.13	42.0	29	1968 - 1996
7200	Dover Ws Intake	Esperance	-43.34	146.96	174.0	29	1965 - 1993
14206	1.5 Km U/S of Mouth	Sulphur Ck	-41.11	146.03	23.0	29	1964 - 1992
14207	at Bannons Br	Leven	-41.25	146.09	495.0	35	1963 - 1997
14210	U/S Flowerdale R Juncti	Inglis	-41.00	145.63	170.0	21	1968 - 1988
14215	at Moorleah	Flowerdale	-40.97	145.61	150.0	31	1966 - 1996

Station ID	Station Name	River Name	Lat (°S)	Long (°E)	Area (km^2)	Record Length (years)	Period of Record
14217	at Sprent	Claytons Rivulet	-41.26	146.17	13.5	26	1970 - 1995
14220	U/S Bass HWY	Seabrook Ck	-41.01	145.77	40.0	20	1977 - 1996
16200	U/S Old Bass Hwy	Don	-41.19	146.31	130.0	24	1967 - 1990
17200	at Tidal Limit	Rubicon	-41.26	146.57	255.0	31	1967 - 1997
17201	1.5KM U/S Tidal Limit	Franklin Rivulet	-41.26	146.61	131.0	20	1975 - 1994
18201	0.5 Km U/S Tamar	Supply	-41.26	146.94	135.0	19	1965 - 1983
18221	D/S Jackeys Marsh	Jackeys Ck	-41.68	146.66	29.0	27	1982 - 2008
18312	D/S Elizabeth R Junctio	Macquarie	-41.91	147.39	1900.0	19	1989 - 2007
19200	2.6KM U/S Tidal Limit	Brid	-41.02	147.37	134.0	32	1965 - 1996
19201	2KM U/S Forester Rd Bdg	Great Forester	-41.11	147.61	195.0	27	1970 - 1996
19204	D/S Yarrow Ck	Pipers	-41.07	147.11	292.0	25	1972 - 1996
304040	U/S Derwent Junction	Florentine River	-42.44	146.52	435.8	58	1951 - 2008
304125	Below Lagoon	Travellers Rest River	-42.07	146.25	43.6	25	1949 - 1973
304597	At Lake Highway	Pine Tree Rivulet Ck	-41.80	146.68	19.4	40	1969 - 2008
308145	At Mount Ficham Track	Franklin River	-42.24	145.77	757.0	56	1953 - 2008
308183	Below Jane River	Franklin River	-42.47	145.76	1590.3	22	1957 - 1978
308225	Below Darwin Dam	Andrew River	-42.22	145.62	5.3	21	1988 - 2008
308446	Below Huntley	Gordon River	-42.66	146.37	458.0	27	1953 - 1979
308799	B/L Alma	Collingwood Ck	-42.16	145.93	292.5	28	1981 - 2008
308819	Above Kelly Basin Rd	Andrew River	-42.22	145.62	4.6	26	1983 - 2008
310061	At Murchison Highway	Que River	-41.58	145.68	18.4	22	1987 - 2008
310148	Above Sterling	Murchison River	-41.76	145.62	756.3	28	1955 - 1982
310149	Below Sophia River	Mackintosh River	-41.72	145.63	523.2	27	1954 - 1980
310472	Below Bulgobac Creek	Que River	-41.62	145.58	119.1	32	1964 - 1995

Station ID	Station Name	River Name	Lat (°S)	Long (°E)	Area (km^2)	Record Length (years)	Period of Record
315074	At Moina	Wilmot River	-41.47	146.07	158.1	46	1923 - 1968
315450	U/S Lemonthyme	Forth River	-41.61	146.13	311.0	46	1963 - 2008
316624	Above Mersey	Arm River	-41.69	146.21	86.0	37	1972 - 2008
318065	Below Deloraine	Meander River	-41.53	146.66	474.0	28	1969 - 1996
318350	Above Rocky Creek	Whyte River	-41.63	145.19	310.8	33	1960 - 1992

Table A5 Selected catchments from Queensland

Station ID	Station Name	River Name	Lat (°S)	Long (°E)	Area (km^2)	Record Length (years)	Period of Record
102101A	Fall Ck	Pascoe	-12.88	142.98	651	33	1968 - 2005
104001A	Telegraph Rd	Stewart	-14.17	143.39	470	32	1970 - 2005
105105A	Developmental Rd	East Normanby	-15.77	145.01	297	34	1970 - 2005
107001B	Flaggy	Endeavour	-15.42	145.07	337	43	1959 - 2004
108002A	Bairds	Daintree	-16.18	145.28	911	29	1969 - 2000
108003A	China Camp	Bloomfield	-15.99	145.29	264	32	1971 - 2004
110003A	Picnic Crossing	Barron	-17.26	145.54	228	80	1926 - 2005
110011B	Recorder	Flaggy Ck	-16.78	145.53	150	44	1956 - 2003
110101B	Freshwater	Freshwater Ck	-16.94	145.70	70	37	1922 - 1958
111001A	Gordonvale	Mulgrave	-17.10	145.79	552	43	1917 - 1972
111003C	Aloomba	Behana Ck	-17.13	145.84	86	28	1943 - 1970
111005A	The Fisheries	Mulgrave	-17.19	145.72	357	34	1967 - 2004
111007A	Peets Br	Mulgrave	-17.14	145.76	520	31	1973 - 2004
111105A	The Boulders	Babinda Ck	-17.35	145.87	39	29	1967 - 2003
112001A	Goondi	North Johnstone	-17.53	145.97	936	39	1929 - 1967
112002A	Nerada	Fisher Ck	-17.57	145.91	15.7	75	1929 - 2004
112003A	Glen Allyn	North Johnstone	-17.38	145.65	165	46	1959 - 2004
112004A	Tung Oil	North Johnstone	-17.55	145.93	925	31	1967 - 2004
112101B	U/S Central Mill	South Johnstone	-17.61	145.98	400	81	1917 - 2003
113004A	Powerline	Cochable Ck	-17.75	145.63	95	32	1967 - 2001
114001A	Upper Murray	Murray	-18.11	145.80	156	31	1971 - 2003
116005B	Peacocks Siding	Stone	-18.69	145.98	368	36	1936 - 1971
116008B	Abergowrie	Gowrie Ck	-18.45	145.85	124	51	1954 - 2004
116010A	Blencoe Falls	Blencoe Ck	-18.20	145.54	226	40	1961 - 2000

Station ID	Station Name	River Name	Lat (°S)	Long (°E)	Area (km^2)	Record Length (years)	Period of Record
116011A	Ravenshoe	Millstream	-17.60	145.48	89	42	1963 - 2004
116012A	8.7KM	Cameron Ck	-18.07	145.34	360	41	1962 - 2002
116013A	Archer Ck	Millstream	-17.65	145.34	308	42	1962 - 2003
116014A	Silver Valley	Wild	-17.63	145.30	591	44	1962 - 2005
116015A	Wooroora	Blunder Ck	-17.74	145.44	127	38	1967 - 2004
116017A	Running Ck	Stone	-18.77	145.95	157	33	1971 - 2004
117002A	Bruce HWY	Black	-19.24	146.63	256	31	1974 - 2004
117003A	Bluewater	Bluewater Ck	-19.18	146.55	86	30	1974 - 2003
118101A	Gleesons Weir	Ross	-19.32	146.74	797	44	1916 - 1959
118106A	Allendale	Alligator Ck	-19.39	146.96	69	30	1975 - 2004
119006A	Damsite	Major Ck	-19.67	147.02	468	25	1979 - 2003
120014A	Oak Meadows	Broughton	-20.18	146.32	182	28	1971 - 1998
120102A	Keelbottom	Keelbottom Ck	-19.37	146.36	193	38	1968 - 2005
120120A	Mt. Bradley	Running	-19.13	145.91	490	30	1976 - 2005
120204B	Crediton Recorder	Broken	-21.17	148.51	41	31	1957 - 1987
120206A	Mt Jimmy	Pelican Ck	-20.60	147.69	545	27	1961 - 1987
120216A	Old Racecourse	Broken	-21.19	148.45	100	36	1970 - 2005
120307A	Pentland	Cape	-20.48	145.47	775	34	1970 - 2003
121001A	Ida Ck	Don	-20.29	148.12	604	48	1958 - 2005
121002A	Guthalungra	Elliot	-19.94	147.84	273	32	1974 - 2005
122004A	Lower Gregory	Gregory	-20.30	148.55	47	33	1973 - 2005
124001A	Caping Siding	O'Connell	-20.63	148.57	363	35	1970 - 2004
124002A	Calen	StHelens Ck	-20.91	148.76	118	32	1974 - 2005
124003A	Jochheims	Andromache	-20.58	148.47	230	29	1977 - 2005

Station ID	Station Name	River Name	Lat (°S)	Long (°E)	Area (km^2)	Record Length (years)	Period of Record
125002C	Sarich's	Pioneer	-21.27	148.82	757	43	1961 - 2005
125004B	Gargett	Cattle Ck	-21.18	148.74	326	38	1968 - 2005
125005A	Whitefords	Blacks Ck	-21.33	148.83	506	32	1974 - 2005
125006A	Dam Site	Finch Hatton Ck	-21.11	148.63	35	29	1977 - 2005
126003A	Carmila	Carmila Ck	-21.92	149.40	84	31	1974 - 2004
129001A	Byfield	Waterpark Ck	-22.84	150.67	212	48	1953 - 2005
130004A	Old Stn	Raglan Ck	-23.82	150.82	389	41	1964 - 2004
130108B	Curragh	Blackwater Ck	-23.50	148.88	776	31	1973 - 2005
130207A	Clermont	Sandy Ck	-22.80	147.58	409	40	1966 - 2005
130208A	Ellendale	Theresa Ck	-22.98	147.58	758	37	1965 - 2001
130215A	Lilyvale Lagoon	Crinum Ck	-23.21	148.34	252	29	1977 - 2005
130319A	Craiglands	Bell Ck	-24.15	150.52	300	44	1961 - 2004
130321A	Mt. Kroombit	Kroombit Ck	-24.41	150.72	373	41	1964 - 2004
130334A	Pump Stn	South Kariboe Ck	-24.56	150.75	284	33	1973 - 2005
130335A	Wura	Dee	-23.77	150.36	472	34	1972 - 2005
130336A	Folding Hills	Grevillea Ck	-24.58	150.62	233	33	1973 - 2005
130348A	Red Hill	Prospect Ck	-24.45	150.42	369	30	1976 - 2005
130349A	Kingsborough	Don	-23.97	150.39	593	28	1977 - 2005
130413A	Braeside	Denison Ck	-21.77	148.79	757	34	1972 - 2005
133003A	Marlua	Diglum Ck	-24.19	151.16	203	36	1969 - 2004
135002A	Springfield	Kolan	-24.75	151.59	551	40	1966 - 2005
135004A	Dam Site	Gin Gin Ck	-24.97	151.89	531	40	1966 - 2005
136006A	Dam Site	Reid Ck	-25.27	151.52	219	40	1966 - 2005
136102A	Meldale	Three Moon Ck	-24.69	150.96	310	32	1949 - 1980

Station ID	Station Name	River Name	Lat (°S)	Long (°E)	Area (km^2)	Record Length (years)	Period of Record
136107A	Cania Gorge	Three Moon Ck	-24.73	151.01	370	26	1963 - 1988
136108A	Upper Monal	Monal Ck	-24.61	151.11	92	43	1963 - 2005
136111A	Dakiel	Splinter Ck	-24.75	151.26	139	41	1965 - 2005
136112A	Yarrol	Burnett	-24.99	151.35	370	40	1966 - 2005
136202D	Litzows	Barambah Ck	-26.30	152.04	681	85	1921 - 2005
136203A	Brooklands	Barker Ck	-26.74	151.82	249	64	1941 - 2005
136301B	Weens Br	Stuart	-26.50	151.77	512	66	1936 - 2005
137001B	Elliott	Elliott	-24.99	152.37	220	52	1949 - 2004
137003A	Dr Mays Crossing	Elliott	-24.97	152.42	251	30	1975 - 2004
137101A	Burrum HWY	Gregory	-25.09	152.24	454	36	1967 - 2004
137201A	Bruce HWY	Isis	-25.27	152.37	446	38	1967 - 2004
138002C	Brooyar	Wide Bay Ck	-26.01	152.41	655	94	1910 - 2005
138003D	Glastonbury	Glastonbury Ck	-26.22	152.52	113	81	1921 - 2006
138009A	Tagigan Rd	Tinana Ck	-26.08	152.78	100	31	1975 - 2005
138010A	Kilkivan	Wide Bay Ck	-26.08	152.22	322	97	1910 - 2006
138101B	Kenilworth	Mary	-26.60	152.73	720	52	1921 - 1972
138102C	Zachariah	Amamoor Ck	-26.37	152.62	133	83	1921 - 2005
138103A	Knockdomny	Kandanga Ck	-26.40	152.64	142	34	1921 - 1954
138104A	Kidaman	Obi Obi Ck	-26.63	152.77	174	42	1921 - 1963
138106A	Baroon Pocket	Obi Obi Ck	-26.71	152.86	67	39	1941 - 1986
138107B	Cooran	Six Mile Ck	-26.33	152.81	186	58	1948 - 2005
138110A	Bellbird Ck	Mary	-26.63	152.70	486	45	1960 - 2004
138111A	Moy Pocket	Mary	-26.53	152.74	820	39	1964 - 2004
138113A	Hygait	Kandanga Ck	-26.39	152.64	143	34	1972 - 2005

Station ID	Station Name	River Name	Lat (°S)	Long (°E)	Area (km^2)	Record Length (years)	Period of Record
140002A	Coops Corner	Teewah Ck	-26.06	153.04	53	27	1975 - 2005
141001B	Kiamba	South Maroochy	-26.59	152.90	33	65	1938 - 2004
141003C	Warana Br	Petrie Ck	-26.62	152.96	38	41	1959 - 2004
141004B	Yandina	South Maroochy	-26.56	152.94	75	27	1959 - 2004
141006A	Mooloolah	Mooloolah	-26.76	152.98	39	33	1972 - 2004
142001A	Upper Caboolture	Caboolture	-27.10	152.89	94	40	1966 - 2005
142201D	Cashs Crossing	South Pine	-27.34	152.96	178	46	1918 - 1963
142202A	Drapers Crossing	South Pine	-27.35	152.92	156	39	1966 - 2005
143010B	Boat Mountain	Emu Ck	-26.98	152.29	915	31	1967 - 2005
143015B	Damsite	Cooyar Ck	-26.74	152.14	963	35	1969 - 2005
143101A	Mutdapily	Warrill Ck	-27.75	152.69	771	39	1915 - 1953
143102B	Kalbar No.2	Warrill Ck	-27.92	152.60	468	55	1913 - 1970
143103A	Moogerah	Reynolds Ck	-28.04	152.55	190	36	1918 - 1953
143107A	Walloon	Bremer	-27.60	152.69	622	36	1962 - 1999
143108A	Amberley	Warrill Ck	-27.67	152.70	914	36	1962 - 2004
143110A	Adams Br	Bremer	-27.83	152.51	125	29	1972 - 2004
143113A	Loamside	Purga Ck	-27.68	152.73	215	28	1974 - 2004
143203C	Helidon Number 3	Lockyer Ck	-27.54	152.11	357	74	1927 - 2004
143208A	Dam Site	Fifteen Mile Ck	-27.46	152.10	87	26	1957 - 1985
143209B	Mulgowie2	Laidley Ck	-27.73	152.36	167	31	1958 - 2004
143303A	Peachester	Stanley	-26.84	152.84	104	77	1928 - 2005
143307A	Causeway	Byron Ck	-27.13	152.65	79	26	1976 - 2005
145002A	Lamington No.1	Christmas Ck	-28.24	152.99	95	43	1910 - 1953
145003B	Forest Home	Logan	-28.20	152.77	175	83	1918 - 2005

Station ID	Station Name	River Name	Lat (°S)	Long (°E)	Area (km^2)	Record Length (years)	Period of Record
145005A	Avonmore	Running Ck	-28.30	152.91	89	30	1923 - 1952
145010A	5.8KM Deickmans Br	Running Ckreek	-28.25	152.89	128	40	1966 - 2005
145011A	Croftby	Teviot Brook	-28.15	152.57	83	38	1967 - 2005
145012A	The Overflow	Teviot Brook	-27.93	152.86	503	39	1967 - 2005
145018A	Up Stream Maroon Dam	Burnett Ck	-28.22	152.61	82	32	1971 - 2005
145020A	Rathdowney	Logan	-28.22	152.87	533	32	1974 - 2005
145101D	Lumeah Number 2	Albert	-28.06	153.04	169	43	1911 - 1953
145102B	Bromfleet	Albert	-27.91	153.11	544	85	1919 - 2005
145103A	Good Dam Site	Cainbable Ck	-28.09	153.08	42	32	1963 - 2004
145107A	Main Rd Br	Canungra Ck	-28.00	153.16	101	32	1974 - 2005
146002B	Glenhurst	Nerang	-28.00	153.31	241	85	1920 - 2005
146003B	Camberra Number 2	Currumbin Ck	-28.20	153.41	24	55	1928 - 1982
146004A	Neranwood	Little Nerang Ck	-28.13	153.29	40	35	1927 - 1961
146005A	Chippendale	Tallebudgera Ck	-28.16	153.40	55	27	1927 -1953
146010A	Army Camp	Coomera	-28.03	153.19	88	43	1963 - 2005
146012A	Nicolls Br	Currumbin Ck	-28.18	153.42	30	31	1971 - 2005
146014A	Beechmont	Back Ck	-28.12	153.19	7	31	1972 - 2004
146095A	Tallebudgera Ck Rd	Tallebudgera Ck	-28.15	153.40	56	29	1971 - 2004
416303C	Clearview	Pike Ck	-28.81	151.52	950	48	1935 - 1987
416305B	Beebo	Brush Ck	-28.69	150.98	335	36	1969 - 2005
416312A	Texas	Oaky Ck	-28.81	151.15	422	35	1970 - 2004
416404C	Terraine	Bracker Ck	-28.49	151.28	685	45	1953 - 2001
416410A	Barongarook	Macintyre Brook	-28.44	151.46	465	32	1968 - 2001
422210A	Tabers	Bungil Ck	-26.41	148.78	710	32	1967 - 2004

Station ID	Station Name	River Name	Lat (°S)	Long (°E)	Area (km^2)	Record Length (years)	Period of Record
422301A	Long Crossing	Condamine	-28.32	152.34	85	66	1912 - 1977
422302A	Killarney	Spring Ck	-28.35	152.34	21	45	1910 - 1954
422303A	Killarney	Spring Ck South	-28.36	152.34	10	45	1910 - 1954
422304A	Elbow Valley	Condamine	-28.37	152.16	275	56	1916 - 1971
422306A	Swanfels	Swan Ck	-28.16	152.28	83	85	1920 - 2004
422307A	Kings Ck	Kings Ck	-27.90	151.91	334	42	1921 - 1966
422313B	Emu Vale	Emu Ck	-28.23	152.23	148	58	1948 - 2005
422317B	Rocky Pond	Glengallan Ck	-28.13	151.92	520	38	1954 - 1991
422319B	Allora	Dalrymple Ck	-28.04	152.01	246	36	1969 - 2005
422321B	Killarney	Spring Ck	-28.35	152.33	35	45	1960 - 2004
422326A	Cranley	Gowrie Ck	-27.52	151.94	47	34	1970 - 2004
422332B	Oakey	Gowrie Ck	-27.47	151.74	142	25	1969 - 2006
422334A	Aides Br	Kings Ck	-27.93	151.86	516	35	1970 - 2004
422338A	Leyburn	Canal Ck	-28.03	151.59	395	27	1975 - 2004
422341A	Brosnans Barn	Condamine	-28.33	152.31	92	29	1977 - 2005
422394A	Elbow Valley	Condamine	-28.37	152.14	325	32	1973 - 2004
913010A	16 Mile Waterhole	Fiery Ck	-18.88	139.36	722	29	1973 - 2004
915011A	Mt Emu Plains	Porcupine Ck	-20.18	144.52	540	31	1972 - 2004
915206A	Railway Crossing	Dugald	-20.20	140.22	660	31	1970 - 2004
915211A	Landsborough HWY	Williams	-20.87	140.83	415	31	1971 - 2003
917104A	Roseglen	Etheridge	-18.31	143.58	867	32	1967 - 2005
917107A	Mount Surprise	Elizabeth Ck	-18.13	144.31	651	32	1969 - 2002
919005A	Fonthill	Rifle Ck	-16.68	145.23	366	32	1969 - 2004
919013A	Mulligan HWY	McLeod	-16.50	145.00	532	25	1973 - 2005

Station ID	Station Name	River Name	Lat (°S)	Long (°E)	Area (km^2)	Record Length (years)	Period of Record
919201A	Goldfields	Palmer	-16.11	144.78	533	30	1968 - 2004
919305B	Nullinga	Walsh	-17.18	145.30	326	35	1957 - 1991
922101B	Racecourse	Coen	-13.96	143.17	172	32	1968 - 2004
926002A	Dougs Pad	Dulhunty	-11.83	142.42	332	30	1971 - 2004

Table A6 Selected catchments from Western Australia

Station ID	Station Name	River Name	Lat (°S)	Long (°E)	Area (km^2)	Record Length (years)	Period of Record
601005	Cascades	Young	-33.54	120.97	88.9	25	1974 - 1998
601006	Munglinup	Young	-33.56	120.9	11.5	33	1974 - 2006
601600	Melaleuka	Young	-33.58	120.87	3.5	34	1975 - 2008
602003	Wellards	Jackitup Ck	-33.95	118.1	88	28	1980 - 2007
602005	Anderson Farm	Chelgiup Ck	-34.89	118.01	48	32	1977 - 2008
602199	Black Cat	Goodga	-34.95	118.08	49.2	45	1964 - 2008
602600	Hinkleys Farm	Jackitup Ck	-33.9	118.12	0.5	27	1972 - 1998
603003	Kompup	Denmark	-34.7	117.21	241.9	35	1974 - 2008
603005	Beigpiegup	Mitchell	-34.83	117.39	51.4	23	1986 - 2008
603007	Sleeman Rd Bridge	Sleeman	-34.96	117.5	75.7	24	1985 - 2008
603008	Pardelup Prison Farm	Upper Hay Trib	-34.63	117.38	1.3	20	1989 - 2008
603013	Eden Rd	Cuppup	-35	117.49	61.1	20	1989 - 2008
603136	Mt Lindesay	Denmark	-34.87	117.31	502.4	49	1960 - 2008
603190	Woonanup	Yate Flat Ck	-34.7	117.29	56.3	46	1963 - 2008
606001	Teds Pool	Deep	-34.77	116.62	467.8	33	1975 - 2007
606002	Wattle Block	Weld	-34.69	116.52	24.2	27	1982 - 2008
606185	Dog Pool	Shannon	-34.77	116.38	407.6	35	1964 - 1998
606218	Baldania Ck Conflu	Gardner	-34.75	116.19	392.4	33	1966 - 1998
607004	Quabicup Hill	Perup	-34.33	116.46	666.7	35	1974 - 2008
607005	North Catch. B	Yerraminnup Ck	-34.14	116.32	2.4	23	1975 - 1997
607006	South Catch.B	Yerraminnup Ck	-34.15	116.34	2	23	1975 - 1997
607007	Bullilup	Tone	-34.25	116.68	983.1	31	1978 - 2008
607009	Pemberton Weir	Lefroy Brook	-34.44	116.02	253.6	30	1952 - 1981
607010	March Rd Catch.E	Six Mile Brook Trib	-34.48	116.33	2.9	24	1976 - 1999

Station ID	Station Name	River Name	Lat (°S)	Long (°E)	Area (km^2)	Record Length (years)	Period of Record
607011	April Rd North Catch.F	Quininup Brook Trib	-34.5	116.35	2.5	23	1976 - 1998
607012	April Rd South Catch.G	Quininup Brook Trib	-34.51	116.35	1.6	24	1976 - 1999
607013	Rainbow Trail	Lefroy Brook	-34.43	116.02	249.4	30	1979 - 2008
607014	Netic Rd	Four Mile Brook	-34.3	116	13.1	20	1979 - 1998
607144	Quintarrup	Wilgarup	-34.35	116.35	460.5	48	1961 - 2008
607155	Malimup Track	Dombakup Brook	-34.58	115.97	118.5	39	1961 - 1999
607600	Manjimup Research Stn	Smith Brook Trib	-34.37	116.21	0.5	31	1970 - 2007
608001	Upper Iffley	Barlee Brook	-34.21	115.77	159.1	28	1972 - 1999
608002	Staircase Rd	Carey Brook	-34.39	115.84	30.3	34	1975 - 2008
608004	Lewin North Catch C	Easter Brook Trib	-34.21	115.86	1.2	22	1976 - 1997
608006	Lease Rd	Carey Brook	-34.33	115.91	2.4	24	1976 - 1999
608151	Strickland	Donnelly	-34.33	115.78	782.1	57	1952 - 2008
608171	Boat Landing Rd	Fly Brook	-34.45	115.8	62.9	39	1962 - 2008
609002	Brennans Ford	Scott	-34.28	115.3	627.7	40	1969 - 2008
609003	Cambray	St Paul Brook	-33.9	115.66	161.6	26	1974 - 1999
609004	Dido Rd	St Paul Brook	-33.83	115.58	26	26	1974 - 1999
609005	Mandelup Pool	Balgarup	-33.91	117.14	82.4	34	1975 - 2008
609006	Balgarup	Weenup Ck	-33.95	117.21	13.3	25	1975 - 1999
609008	Millbrook	Apostle Brook	-33.8	115.63	27.6	24	1976 - 1999
609010	Lake Toolibin Inflow	Northern Arthur	-32.9	117.61	438.5	31	1978 - 2008
609011	Padbury Rd	Balingup Brook Trib	-33.81	116	1.7	21	1978 - 1998
609016	Hester Hill	Hester Brook	-33.92	116.1	176.6	21	1983 - 2005
609017	Brooklands	Balingup Brook	-33.8	115.95	548.9	26	1983 - 2008
609018	Barrabup Pool	St John Brook	-33.94	115.69	552.3	26	1983 - 2008

Station ID	Station Name	River Name	Lat (°S)	Long (°E)	Area (km^2)	Record Length (years)	Period of Record
610001	Willmots Farm	Margaret	-33.94	115.05	443	39	1970 - 2008
610005	Happy Valley	Ludlow	-33.68	115.62	109.2	26	1973 - 1998
610006	Woodlands	Wilyabrup Brook	-33.8	115.02	82.3	36	1973 - 2008
610007	Claymore	Ludlow	-33.74	115.7	9.5	22	1977 - 1998
610008	Whicher Range	Margaret R North	-33.81	115.44	15.5	23	1977 - 1999
610219	Yates Bridge	Capel	-33.65	115.7	315.1	23	1966 - 2008
611004	Boyanup Bridge	Preston	-33.48	115.73	808.4	29	1980 - 2008
611049	Beelerup	Preston	-33.56	115.88	597.5	21	1955 - 1975
611111	Woodperry Homestead	Thomson Brook	-33.63	115.95	102.1	51	1958 - 2008
611221	Pesconeris Farm	Coolingutup Brook	-33.53	115.87	3.9	43	1966 - 2008
612004	Worsley	Hamilton	-33.31	116.05	32.3	37	1972 - 2008
612005	Mast View	Stones Brook	-33.37	115.94	12.9	27	1972 - 1998
612007	Dons Catchment	Bingham R Trib	-33.28	116.47	3.5	35	1974 - 2008
612008	Ernies Catchment	Bingham R Trib	-33.29	116.44	2.7	35	1974 - 2008
612009	Lemon Catchment	Pollard Brook Trib	-33.3	116.41	3.5	29	1974 - 2005
612010	Wights Catchment	Salmon Brook Trib	-33.42	115.98	0.9	34	1974 - 2007
612011	Salmon Catchment	Salmon Brook	-33.42	115.98	0.8	25	1974 - 1998
612012	Falcon Rd	Falcon Brook	-33.41	115.97	5.4	23	1974 - 1996
612014	Palmer	Bingham	-33.28	116.28	366.1	34	1975 - 2008
612016	Maxon Farm	Batalling Ck	-33.32	116.57	16.8	33	1976 - 2008
612019	Duces Farm	Bussell Brook	-33.46	116.02	37.5	22	1977 - 1998
612021	Stenwood	Bingham	-33.19	116.47	48.4	21	1978 - 1998
612022	Sandalwood	Brunswick	-33.22	115.92	116.2	29	1980 - 2008
612025	James Well	Camballan Ck	-33.46	116.43	170	27	1982 - 2008

Station ID	Station Name	River Name	Lat (°S)	Long (°E)	Area (km^2)	Record Length (years)	Period of Record
612034	South Branch	Collie	-33.39	116.16	661.6	53	1952 - 2008
612036	Stubbs Farm	Harris	-33.29	116.15	382.2	25	1952 - 1976
612152	Olive Hill	Brunswick	-33.24	115.87	225.4	21	1962 - 1982
612230	James Crossing	Collie R East Trib	-33.38	116.58	170.6	42	1967 - 2008
613002	Dingo Rd	Harvey	-33.09	116.04	147.2	39	1970 - 2008
613007	Waterous	Bancell Brook	-32.95	115.95	13.6	34	1975 - 2008
613013	Wagerup	Bancell Brook	-32.95	115.96	12.5	25	1954 - 1978
613018	Urquharts	McKnoes Brook	-32.89	115.97	24.4	22	1980 - 2001
613020	Mt William	Samson Brook	-32.93	116.03	4	21	1981 - 2001
613146	Hillview Farm	Clarke Brook	-33	115.92	17.1	39	1962 - 2000
614003	Brookdale Siding	Marrinup Brook	-32.7	115.97	45.6	36	1972 - 2007
614005	Kentish Farm	Dirk Brook	-32.42	116	35.1	30	1971 - 2000
614007	Del Park	South Dandalup Trib	-32.67	116.04	1.3	34	1975 - 2008
614011	Tunnel Rd	Mooradung Bk Trib	-32.95	116.48	2.1	22	1976 - 1997
614017	Warren Catchment	Little Dandalup Trib	-32.59	116.03	0.9	32	1977 - 2008
614018	Bennetts Catchment	Little Dandalup Trib	-32.6	116.03	0.9	32	1977 - 2008
614019	Hansens Catchment	Little Dandalup Trib	-32.59	116.05	0.7	22	1977 - 1998
614020	Higgens Catchment	Little Dandalup Trib	-32.58	116.09	0.6	21	1978 - 1998
614021	Lewis Catchment	North Dandalup Trib	-32.57	116.06	2	32	1977 - 2008
614024	Jones Catchment	North Dandalup Trib	-32.55	116.09	0.7	21	1978 - 1998
614025	Umbucks Catchment	Marrinup Brook Trib	-32.7	116	3.3	20	1979 - 1998
614028	Hopelands Rd	Dirk Brook	-32.43	115.91	63.8	22	1979 - 2000
614037	O'Neil Rd	Big Brook	-32.51	116.19	149.4	26	1983 - 2008
614047	Murray Valley Plntn	Davis Brook	-32.76	116.1	65.7	46	1956 - 2001

Station ID	Station Name	River Name	Lat (°S)	Long (°E)	Area (km^2)	Record Length (years)	Period of Record
614060	Gordon Catchment	South Dandalup R Trib	-32.63	116.26	2.1	21	1988 - 2008
614062	Bates Catchment	Little Dandalup Trib	-32.58	116.03	2.2	20	1989 - 2008
614073	Mundlimup	Gooralong Brook	-32.35	116.04	51.5	47	1952 - 1998
614123	Quindanning Rd	Chalk Brook	-33.02	116.24	57.1	26	1960 - 1985
615011	Mooranoppin Rock	Mooranoppin Ck	-31.6	117.73	83.1	34	1975 - 2008
615222	Brookton Highway	Dale R South	-32.4	116.83	286	32	1967 - 1998
615600	North	Kunjin	-32.32	117.73	0.2	30	1969 - 1998
615604	Homestead	North Nungarin	-31.16	118.15	0.2	26	1972 - 1997
615605	Jollys Farm	South Nungarin	-31.18	118.15	0.2	27	1972 - 1998
616006	Tanamerah	Brockman	-31.34	116.09	961.2	28	1981 - 2008
616007	Byfield Rd	Rushy Ck (Manns Gully)	-31.96	116.21	39.2	30	1969 - 1998
616009	Slavery Lane	Pickering Brook	-31.98	116.19	29.4	27	1972 - 1998
616010	Hairpin Bend Rd	Little Darkin	-32.03	116.24	37.8	27	1972 - 1998
616012	Trewd Rd	Helena Brook	-31.92	116.28	26.7	27	1972 - 1998
616013	Ngangaguringuring	Helena	-31.94	116.4	327	36	1973 - 2008
616014	Furfaros Orchard	Piesse Brook	-31.95	116.08	55.2	24	1975- 1998
616022	Ceriani Farm	More Seldom Seen Ck	-32.25	116.08	3.4	39	1970 - 2008
616023	Mount Curtis	Waterfall Gully	-32.21	116.08	8.6	43	1966 - 2008
616041	Vardi Rd	Wungong Brook	-32.25	116.11	80.8	27	1982 - 2008
616189	Railway Parade	Ellen Brook	-31.75	116.02	581.4	44	1965 - 2008
616216	Poison Lease	Helena	-31.97	116.29	590.9	42	1966 - 2007
617002	Hill R Springs	Hill	-30.28	115.37	925.9	37	1971 - 2007
617058	Gingin	Gingin Brook	-31.34	115.92	105.8	51	1958 - 2008
617165	Molecap Hill	Lennard Brook	-31.38	115.92	59	40	1962 - 2001

Station ID	Station Name	River Name	Lat (°S)	Long (°E)	Area (km^2)	Record Length (years)	Period of Record
701003	Wootachooka	Nokanena Brook	-28.37	114.52	235.2	30	1972 - 2001
701005	Robb Crossing	Arrowsmith	-29.62	115.29	809.8	29	1972 - 2000
701006	Buller	Buller	-28.64	114.62	33.9	26	1975 - 2000
701601	Wearbe	Nokanena Brook Catch	-28.33	114.62	0.1	28	1971 - 1998
704001	Boolathana	Yandoo Ck	-24.63	113.82	1000	20	1983 - 2002
706207	Mt Samson	Hardey	-22.67	117.61	250.3	34	1967 - 2000
707001	Palra Springs	Robe	-22.06	117.06	174.3	31	1969 - 1999
708009	Fish Pool	Kanjenjie Ck Trib.	-21.66	117.33	41.1	28	1975 - 2002
708227	Recorder Pool	Portland	-21.45	116.88	553.4	34	1967 - 2000
709006	Blue Dog Pool	Tanberry Ck	-21.59	117.55	128.1	22	1975 - 1996
709007	Marmurrina Pool U-South	Harding	-21.3	117.07	49.4	24	1975 - 1998
709010	Pincunah	Turner	-21.23	118.83	885	24	1985 - 2008
802002	Mt Pierre Gorge	Mount Pierre Ck	-18.62	126.09	318.4	28	1971 - 1998
802202	Mt Winifred	Leopold	-18.02	126.31	5115.4	40	1966 - 2008
802213	Phillips Range	Hann	-16.87	126.05	5069.9	42	1967 - 2008
803001	Mt Joseph	Lennard	-17.37	125.11	1049.8	42	1967 - 2008
803002	Mt Herbert	Lennard	-17.17	125.23	441.4	31	1968 - 1998
803003	Dromedary	Fletcher	-17.12	124.99	67	31	1968 - 1998
806003	Crystal Head	Crystal Ck	-14.49	125.8	68.2	30	1969 - 1998
806004	Old Theda	Carson	-14.79	126.79	1288.6	30	1971 - 2000
809310	Bedford Downs	Ord	-17.43	127.6	552.2	29	1970 - 1998
809312	Frog Hollow	Fletcher Ck Trib	-17.28	128.06	30.6	41	1968 - 2008
809314	Cockburn North	King R	-15.7	128.12	850.3	23	1986 - 2008
809315	Mistake Ck Homestead	Negri	-17.18	129.09	7405.7	38	1970 - 2008

Station ID	Station Name	River Name	Lat (°S)	Long (°E)	Area (km^2)	Record Length (years)	Period of Record
809317	Koongie Park	Black Elvire R Trib	-18.39	127.77	456.8	30	1971 - 2000
809321	Dunham Gorge	Dunham	-16.19	128.3	1631.5	34	1975 - 2008

Table A7 Selected catchments from Northern Territory

Station ID	Station Name	River Name	Lat (°S)	Long (°E)	Area (km^2)	Record Length (years)	Period of Record
G8100189	Victoria HWY	Moriarty Ck	-16.07	129.19	88	19	1967 - 1985
G8110004	Victoria HWY	East Baines	-15.77	130.00	2342	46	1963 - 2008
G8110014	U/S Fig Tree Yard	Sullivan's Ck	-15.57	131.29	143	22	1970 - 1992
G8110110	V.R.D. Rd Crossing	Surprise Ck	-16.08	130.90	361	31	1960 - 1990
G8110263	1.5 Miles D/S Bore	Bullock Ck	-17.13	131.45	474	22	1971 - 1992
G8140008	Old Railway Br	Fergusson	-14.07	131.98	1490	51	1957 - 2008
G8140061	Blue Hole	Copperfield Ck	-13.99	131.90	306	20	1958 - 1977
G8140063	D/S Old Douglas H/S	Douglas	-13.80	131.34	842	51	1957 - 2007
G8140086	D/S Stuart HWY	King	-14.63	132.59	484	23	1964 - 1986
G8140151	Victoria HWY	Mathieson Ck	-15.07	131.74	725	22	1964 - 1986
G8140152	Dam Site	Edith	-14.17	132.08	590	41	1962 - 2008
G8140158	Dam Site	McAdden Ck	-14.35	132.34	133	41	1963 - 2006
G8140159	Waterfall View	Seventeen Mile C	-14.28	132.40	619	45	1963 - 2007
G8140161	Tipperary	Green Ant Ck	-13.74	131.10	435	42	1966 - 2007
G8140166	Gorge	Fish	-14.24	130.90	992	23	1963 - 1985
G8150010	Batchelor Damsite	Finniss	-13.03	130.95	360	32	1975 - 2006
G8150018	Stuart HWY	Elizabeth	-12.61	131.07	101	54	1954 - 2007
G8150096	Cox Peninsula	Carawarra Ck	-12.53	130.67	38.5	42	1965 - 2007
G8150097	Rum Jungle +Ansto Eb4	East Finniss	-12.97	130.97	71	43	1965 - 2007
G8150098	Tumbling Waters	Blackmore	-12.77	130.95	174	48	1960 - 2007
G8150127	D/S McMillans Rd	Rapid Ck	-12.39	130.87	18.3	44	1964 - 2007
G8150151	U/S Darwin R Dam	Celia Ck	-12.91	131.05	52	19	1989 - 2007
G8150180	Gitchams	Finniss	-12.97	130.76	1048	47	1961 - 2007
G8150200	Rum Jungle Rd Crossing	East Finniss	-12.99	131.00	52	23	1982 - 2007

Station ID	Station Name	River Name	Lat (°S)	Long (°E)	Area (km^2)	Record Length (years)	Period of Record
G8150233	McArthur Park	Palmerston Catc.	-12.49	130.98	1.4	20	1984 - 2003
G8160235	Damsite	Takamprimili	-11.78	130.78	166	20	1967 - 1986
G8170002	Railway Br	Adelaide	-13.24	131.11	632	45	1963 - 2007
G8170020	Dirty Lagoon	Adelaide	-12.91	131.24	4325	38	1963 - 2007
G8170062	Eighty-Seven Mile Jump Up	Burrell Ck	-13.42	131.15	36.8	28	1958 - 1985
G8170066	Stuart HWY	Coomalie Ck	-13.01	131.12	82	50	1958 - 2007
G8170075	U/S Manton Dam	Manton	-12.88	131.13	28	37	1965 - 2007
G8170084	Tortilla Flats	Adelaide	-13.09	131.24	1246	49	1959 - 2007
G8170085	Stuart HWY	Acacia Ck	-12.78	131.12	11	45	1963 - 2007
G8180026	El Sherana Rd Crossing	Mary	-13.60	132.22	466	47	1961 - 2008
G8180065	Old Point Stuart Rd Crossing	Opium Ck	-12.55	131.79	15.5	22	1964 - 1985
G8180069	near Burrundie	McKinlay	-13.53	131.72	352	51	1958 - 2008
G8180252	D/S El Sherana Rd	Harriet Ck	-13.68	131.99	122	44	1965 - 2008
G8190001	U/S Arnhem HWY	West Alligator	-12.79	132.18	316	33	1976 - 2008
G8200045	El Sherana (C)	South Alligator	-13.52	132.52	1300	51	1958 - 2008
G8200046	Coljon (C Part)	Deaf Adder Ck	-13.10	133.02	513	20	1972 - 1991
G8200049	near Nourlangie Rock	Koongarra Ck	-12.88	132.83	15.4	28	1978 - 2005
G8200112	Kakadu HWY	Nourlangie Ck	-12.82	132.74	2220	43	1961 - 2005
G8210001	Nimbuwah (C)	Cooper Ck	-12.19	133.35	645	22	1971 - 1992
G8210009	D/S Jabiru	Magela Ck	-12.64	132.90	605	37	1971 - 2007
G8210012	George Town Crossing	Gulungul Ck (Bog	-12.69	132.89	47	21	1972 - 1992
G8210016	Mt. Borradaile	Cooper Ck	-12.08	132.97	1650	26	1980 - 2005
G8210017	Jabiluka Billabong	Magela Ck Plains	-12.46	132.88	1134	33	1973 - 2005
G8210019	Outflow Main Channel	Magela Plains	-12.30	132.82	1435	29	1975 - 2003

Station ID	Station Name	River Name	Lat (°S)	Long (°E)	Area (km^2)	Record Length (years)	Period of Record
G8210024	D/S Nabarlek	Cooper Ck	-12.29	133.34	225	28	1978 - 2005
G8260052	U/S Eldo Rd Crossing	Upper Latram	-12.32	136.82	31	32	1971 - 2004
G8260053	above Tidal Reach	Lower Latram	-12.31	136.78	85	21	1964 - 1984
G9030089	Rd Br	Waterhouse	-14.56	133.11	3110	36	1973 - 2008
G9030090	Wattle Hill	Chambers Ck	-14.50	133.36	89	19	1974 - 1992
G9030124	Daly Waters	Daly Waters Ck	-16.26	133.38	777	29	1977 - 2007
G9070142	Bailey's Grave	McArthur	-16.78	135.76	3885	43	1965 - 2008

Table A8 Selected catchments for arid/semi arid region from all over Australia

Station ID	Station Name	River Name	Lat (°S)	Long (°E)	Area (km^2)	Record Length (years)	Period of Record	Average Annual Rainfall (mm)
410012	Cocketgedong	Billabong Ck	-35.31	146.04	4660	33	1973 - 2005	434.35
409056	Aratula Rd	Tuppal Ck	-35.63	145.06	300	18	1986 - 2000	412.38
425028	Quondong	Wireyards Ck	-32.13	141.85	50	16	1983 - 1999	243.02
425016	Cobar	Box Ck	-31.46	145.81	15	35	1974 - 2008	407.93
408203B	Quambatook	Avoca	-35.91	143.51	4740	29	1973 - 1996	407.9
407287B	U/S Box Ck	Bullock Ck	-35.92	144.18	1231	16	1990 - 2001	412.75
407236B	Mitiamo	Mount Hope Ck	-36.17	144.29	1629	41	1968 - 1996	425.6
415257A	Donald	Richardson	-36.43	142.98	1831	40	1989 - 1999	433.74
406264A	Northern HWY- Echuca	Millewa Ck	-36.19	144.73	32	17	1992 - 2005	452.85
G0060008	South Rd Crossing	Roe Ck	-23.82	133.84	560	41	1967 - 2008	290.56
G0060003	Soil Erosion Project	Gillen Ck	-23.70	133.82	3.8	27	1967 - 1993	295
G0060047	Big Dipper	Charles	-23.65	133.86	52	14	1973 - 1986	304.96
G0060012	Bond Springs (CSIRO Site 6	Stn Ck	-23.53	133.92	34	10	1974 - 1982	306.49
G0060017	U/S	Emily Ck	-23.69	133.98	60	28	1981 - 2008	318.05
G0060046	Wigley Gorge	Todd	-23.64	133.88	360	46	1963 - 2001	318.6
G0060009	Anzac Oval	Todd	-23.70	133.89	443	35	1973 - 2007	320.58
G0060015	Bond Springs	Stn Ck	-23.53	133.92	34	18	1979 - 1995	326.33
G0060126	Heavitree Gap	Todd	-23.73	133.87	502	37	1973 - 2007	329.88
G0010005	Soudan Homestead	Ranken	-20.05	137.02	4360	45	1965 - 2009	381.07
G0290240	Old Telegraph Stn	Tennant Ck	-19.56	134.23	72.3	37	1973 - 2007	391.42
G0290012	Kelly Well. Stuart HWY	Kelly Ck	-19.97	134.21	62	34	1975 - 2008	401.36
G0290242	Stuart HWY	Attack Ck	-19.01	134.15	259	22	1967 - 1986	414.48
G0290228	D/S	Morphett Ck	-18.88	134.09	211	29	1981 - 2007	428.93
001204A	Camooweal	Georgina	-19.93	138.11	2875	19	1971 - 1988	393.47

Station ID	Station Name	River Name	Lat (°S)	Long (°E)	Area (km^2)	Record Length (years)	Period of Record	Average Annual Rainfall (mm)
424202A	Yarronvale	Paroo	-26.79	145.34	1890	20	1968 - 1987	397.53
915204A	Damsite	Cloncurry	-21.08	140.42	4240	33	1969 - 1994	398.48
915210A	Agate Downs	Cloncurry	-21.36	140.41	1089	17	1971 - 1987	411.71
003205A	Darr	Darr	-23.22	144.08	2700	38	1971 - 2006	415.69
915211A	Landsborough HWY	Williams	-20.87	140.83	415	36	1971 - 2006	417.56
915205A	Black Gorge	Malbon	-21.06	140.08	425	17	1971 - 1987	423.63
912115A	Morestone	O Shannassy	-19.60	138.38	425	18	1971 - 1988	431.19
422211A	Woolerbilla-Hebel Rd	Briarie Ck	-28.91	147.68	410	32	1968 - 2004	436.01
915203A	Cloncurry	Cloncurry	-20.67	140.49	5975	33	1969 - 1997	439.12
915203B	Cloncurry	Cloncurry	-20.70	140.50	5859	37	1969 - 2006	440.8
915209A	Main Rd	Corella	-20.45	140.32	1587	17	1972 - 1987	442.72
915001A	Richmond	Mitchell Grass C	-20.76	143.14	6	22	1969 - 1990	443.63
913009A	Flinders HWY	Gorge Ck	-20.69	139.65	248	17	1971 - 1987	444.2
913005A	Damsite	Paroo Ck	-20.34	139.52	305	19	1969 - 1987	450.59
915006A	Revenue Downs	Mountain Ck	-20.64	143.22	203	17	1972 - 1988	454.65
A0040502	Terrapinna Springs	Hamilton Ck	-29.92	139.67	326	10	1984 - 1990	209.43
A5100510	Leigh Creek	Windy Ck	-30.61	138.39	448	18	1986 - 2006	226.58
A5100511	Leigh Creek	Emu Ck	-30.62	138.39	224	18	1986 - 2006	226.58
A5100507	Maynards Well	Windy Ck	-30.64	138.65	170	15	1974 - 1988	288.09
A5090503	Old Kanyaka Ruins	Kanyaka Creek	-32.09	138.29	186.7	36	1977 - 2008	289.54
A5100502	Sugarloaf Hill	Mernmerna Creek	-31.75	138.45	346	18	1973 - 1989	302.34

Appendix B: Papers and technical reports published from research related to Project 5 Regional Flood Methods for Australia

1. Aziz, K., Rahman, A., Fang, G. and Shrestha, S. (2011). Application of Artificial Neural Networks in Regional Flood Estimation in Australia: Formation of Regions Based on Catchment Attributes, The Thirteenth International Conference on Civil, Structural and Environmental Engineering Computing and CSC2011: The Second International Conference on Soft Computing Technology in Civil, Structural and Environmental Engineering, Chania, Crete, Greece, 6-9 September, 2011, 13 pp.
2. Aziz, K., Rahman, A., Fang, G., Haddad, K. and Shrestha, S. (2010). Design flood estimation for ungauged catchments: Application of artificial neural networks for eastern Australia, World Environment and Water Resources Congress, American Society of Civil Engineers (ASCE), 16-20 May 2010, Providence, Rhode Island, USA, pp. 2841-2850.
3. Hackelbusch, A., Micevski, T. Kuczera, G., Rahman, A. and Haddad, K. (2009). Regional Flood Frequency Analysis for Eastern New South Wales: A Region of Influence Approach using Generalised Least Squares Based Parameter Regression. In Proc. 31st Hydrology and Water Resources Symp., Newcastle.
4. Haddad, K. and Rahman, A. (2012). Regional flood frequency analysis in eastern Australia: Bayesian GLS regression-based methods within fixed region and ROI framework – Quantile Regression vs. Parameter Regression Technique, Journal of Hydrology, 430-431 (2012), 142-161.
5. Haddad, K., Rahman, A. And Stedinger, J.R. (2011a). Regional Flood Frequency Analysis using Bayesian Generalised Least Squares: A Comparison between Quantile and Parameter Regression Techniques, Hydrological Processes, 25, 1-14.
6. Haddad, K., Rahman, A. And Kuczera, G. (2011b). Comparison of Ordinary and Generalised Least Squares Regression in Regional Flood Estimation: A Case Study for New South Wales. Australian Journal of Water Resources, 15, 2, 59-70.
7. Haddad, K., and Rahman, A. (2011c). Regional flood estimation in New South Wales Australia using Generalised Least Squares Quantile Regression. Journal of Hydrologic Engineering, ASCE, 16, 11, 920-925.
8. Haddad, K., Rahman, A. and Weinmann, P.E. (2011d). Estimation of major floods: applicability of a simple probabilistic model, Australian Journal of Water Resources, 14, 2, 117-126.

9. Haddad, K., Rahman, A., Ling, F. and Weinmann, P.E. (2011e). Towards a new regional flood frequency analysis method for Tasmania, 34th IAHR World Congress, 26 June – 1 July 2011, Brisbane.
10. Haddad, K., Rahman, A., Kuczera, G. and Micevski, T. (2011f). Regional Flood Frequency Analysis in New South Wales Using Bayesian GLS Regression: Comparison of Fixed Region and Region-of-influence Approaches, 34th IAHR World Congress, 26 June – 1 July 2011, Brisbane.
11. Haddad, K. and Rahman, A. (2011g). Selection of the best fit flood frequency distribution and parameter estimation procedure – A case study for Tasmania in Australia, Stochastic Environmental Research & Risk Assessment, 25, 415-428.
12. Haddad, K., Rahman, A., Weeks, W., Kuczera, G. and Weinmann, P.E. (2011h).Towards a new regional flood frequency analysis method for Western Australia, The 19th International Congress on Modelling and Simulation, 12-16 Dec 2011, Perth, Australia, 3788-3795.
13. Haddad, K., Uddin, J., Rahman, A., Kuczera, G., and Weinmann, P.E. (2011i). A new flood regionalisation model for large flood estimation in Australia, 11th International Multidisciplinary Scientific Geo-Conference and Expo SGEM 2011, Bulgaria, 19-25 June, 2, 761-768.
14. Haddad, K., Rahman, A., Weinmann, P.E., Kuczera, G. and Ball, J.E. (2010a). Streamflow data preparation for regional flood frequency analysis: Lessons from south-east Australia. Australian Journal of Water Resources, 14, 1, 17-32.
15. Haddad, K., Zaman, M. and Rahman, A. (2010b). Regionalisation of skew for flood frequency analysis: a case study for eastern NSW. Australian Journal of Water Resources, 14, 1, 33-41.
16. Haddad, K., Rahman, A. and Weinmann, P.E. (2010c). Estimation of major floods: applicability of a simple probabilistic model, Australian Journal of Water Resources, 14, 2, 117-126.
17. Haddad, K., Zaman, M. Rahman, A. and Shrestha, S. (2010d). Regional Flood Modelling: Use of Monte Carlo cross-validation for the Best Model Selection. World Environment and Water Resources Congress, American Society of Civil Engineers (ASCE), 16-20 May 2010, Providence, Rhode Island, USA, pp. 2831-2840.
18. Haddad, K., Aziz, K., Rahman, A., and Ishak, E.H. and Weinmann, P.E. (2009a). A Probabilistic Model for Estimation of Large Floods in Ungauged Catchments: Application to South-east Australia. In Proc. 32nd Hydrology and Water Resources Symp., 30 Nov to 3 Dec, Newcastle, Australia, pp. 817-828.
19. Haddad, K., Pirozzi, J., McPherson, G., Rahman, A. and Kuczera, G. (2009b). Regional Flood Estimation Technique for NSW: Application of Generalised Least

Squares Quantile Regression Technique. In Proc. 31st Hydrology and Water Resources Symp., Newcastle.

20. Haddad, K., Rahman, A. and Weinmann, P. E. (2008a). Development of a Generalised Least Squares Based Quantile Regression Technique for design flood estimation in Victoria, 31st Hydrology and Water Resources Symp., Adelaide, 15-17 April 2008, 2546-2557.
21. Haddad, K. and Rahman, A. (2008b). Investigation on at-site flood frequency analysis in south-east Australia, IEM Journal, The Journal of The Institution of Engineers, Malaysia, 69(3), 59-64.
22. Haddad, K., Rahman, A. and Weinmann, E. (2006). Design flood estimation in ungauged catchments by quantile regression technique: ordinary least squares and generalised least squares compared. In Proc. 30th Hydrology and Water Resources Symp., The Institution of Engineers Australia, 4-7 Dec 2006, Launceston, 6pp. ISBN 0-8582579-0-4.
23. Ishak, E., Haddad, K., Zaman and Rahman (2011a). Scaling property of regional floods in New South Wales Australia, Natural Hazards, 58: 1155-1167.
24. Ishak, I., Rahman, A., Westra, S., Sharma, A. and Kuczera, G. (2011b). Trends in Peak Streamflow Data in Australia: Impacts of Serial and Cross-correlation, 34th IAHR World Congress, 26 June – 1 July 2011, Brisbane.
25. Ishak, E.H., Rahman, A., Westra, S., Sharma, A. and Kuczera, G. (2010). Preliminary analysis of trends in flood data in Australian continent. World Environment and Water Resources Congress, American Society of Civil Engineers (ASCE), 16-20 May 2010, Providence, Rhode Island, USA, pp. 120-124.
26. Ishak, E.H., Aziz, K., Rahman, A. and Haddad, K. (2009). Scaling Behaviour of Regional Floods in New South Wales Australia. In Proc. 32nd Hydrology and Water Resources Symp., 30 Nov to 3 Dec, Newcastle, Australia, 400-408.
27. Palmen, L.B. and Weeks, W.D. (2009). Regional flood frequency for Queensland using the Quantile Regression Technique. 32nd Hydrology and Water resources Symp, Newcastle, 30th Nov to 3rd Dec 2009
28. Pirozzi, J. and Rahman, A. (2010). Design Streamflow Estimation for Ungauged Catchments in Eastern NSW: Identification of Important Predictor Variables, Australian Water Association National Conference, Ozwater 2010, 8-10 March, Brisbane.
29. Pirozzi, J., Ashraf, M., Rahman, A., and Haddad, K. (2009). Design Flood Estimation for Ungauged Catchments in Eastern NSW: Evaluation of the Probabilistic Rational Method. In Proc. 32nd Hydrology and Water Resources Symp., 30 Nov to 3 Dec, Newcastle, Australia, pp. 805-816.

30. Pirozzi, J., Ashraf, M., Rahman, A., and Haddad, K. (2009a). Design Flood Estimation for Ungauged Catchments in Eastern NSW: Evaluation of the Probabilistic Rational Method. In Proc. 31st Hydrology and Water Resources Symp., Newcastle.
31. Rahman, A., Zaman, M., Fotos, M., Haddad, K. Rajaratnam, L., Weeks, B. (2011a). Towards a New Regional Flood Estimation Method for the Northern Territory, 34th IAHR World Congress, 26 June – 1 July 2011, Brisbane.
32. Rahman, A., Haddad, K., Zaman, M., Kuczera, G. and Weinmann, P.E. (2011b). Design flood estimation in ungauged catchments: A comparison between the Probabilistic Rational Method and Quantile Regression Technique for NSW, Australian Journal of Water Resources, 14, 2, 127-139.
33. Rahman, A., Haddad, K., Ishak, E., Weinmann, P.E., Kuczera, G. (2010). Regional Flood Estimation in Australia: An Overview of the Study in Relation to the Upgrade of Australian Rainfall and Runoff. 50th Annual Floodplain Management Authorities Conference Gosford 2010 FMA, 23-29 Feb, Gosford, NSW, 2010.
34. Rahman, A., Haddad, K., Kuczera, G. and Weinmann, P.E. (2009). Regional flood methods for Australia: data preparation and exploratory analysis. School of Engineering, UWS. Research Report prepared for Engineers Australia, Report No. P5/S1/003, 179 pp.
35. Rahman, A., Haddad, K., Pirozzi, J. and McPherson G. (2009a). Upgradation of regional design flood estimation method for New South Wales. Research Report. Prepared for Engineers Australia, School of Engineering, UWS, 64 pp.
36. Rahman, A., Rima, K. and Weeks, W. (2008). Development of Regional Flood Estimation Methods Using Quantile Regression Technique: A Case Study for North-eastern Part of Queensland, 31st Hydrology and Water Resources Symp., Adelaide, 15-17 April 2008, 329-340
37. Rahman, A., Haddad, K. and Weinmann, P.E. (2008a). Regional Flood Frequency Analysis for Victorian Catchments: Sensitivity to Data Availability and Catchment Selection. Research Report Prepared for Department of Sustainability and Environment Victoria, School of Engineering, UWS, 34pp.
38. Taylor, M., Haddad, K., Zaman, M. and Rahman, A. (2011). Regional flood modelling in Western Australia: Application of regression based methods using ordinary least squares, The 19th International Congress on Modelling and Simulation, 12-16 Dec 2011, Perth, Australia, 3803-3810.
39. Zaman, M., Rahman, A. and Haddad, K. (2011). Regional flood modelling in arid and semi-arid regions in Australia, The 19th International Congress on Modelling and Simulation, 12-16 Dec 2011, Perth, Australia, 3811-3817.

40. Zaman, M., Rahman, I., Haddad, K., and Rahman, A. (2010). Scaling issues in design flood estimation for ungauged catchments: A case study for eastern Australia. World Environment and Water Resources Congress, American Society of Civil Engineers (ASCE), 16-20 May 2010, Providence, Rhode Island, USA, pp. 2860-2869.