# Database Underpinning the RFFE Model 2015 in the New Australian Rainfall and Runoff

Ataur Rahman
School of Computing, Engineering and Mathematics, Western Sydney University, Penrith (Kingswood) Campus, NSW 2751, Australia
E-mail: a.rahman@westernsydney.edu.au

Khaled Haddad
School of Computing, Engineering and Mathematics, Western Sydney University, Penrith (Kingswood) Campus, NSW 2751, Australia
E-mail: k.haddad@westernsydney.edu.au

Ayesha S. Rahman
School of Computing, Engineering and Mathematics, Western Sydney University, Penrith (Kingswood) Campus, NSW 2751, Australia
E-mail: ayesha.rahman@westernsydney.edu.au

Md. M. Haque
Institute for Infrastructure Engineering, Western Sydney University, Penrith (Kingswood) Campus, NSW 2751, Australia
E-mail: m.haque@westernsydney.edu.au

Wilfredo L. Caballero
School of Computing, Engineering and Mathematics, Western Sydney University, Penrith (Kingswood) Campus, NSW 2751, Australia
Email: w.caballero@westernsydney.edu.au

*Abstract*

*The new regional flood frequency estimation (RFFE) model for Australia, known as 'RFFE Model 2015' is based on an extensive database of gauged catchments. This paper presents important characteristics of the database so that the users of the RFFE Model are aware of the strengths and limitations of the database and how these could impact the outcomes of the RFFE Model. The RFFE Model data set consists of 798 gauged catchments from the humid coastal areas and 55 catchments from the arid/semi-arid areas. Humid coastal areas include most of the coastal belts of Australia where there are a good number of recorded streamflow stations. The arid/semi-arid areas include Pilbara and interior Australia. An upper limit of catchment size of 1,000 km2 is adopted; however, in Tasmania and in the Northern Territory, a few larger catchments are included as the numbers of catchments with areas less than 1,000 km2 are too small. The cut-off streamflow record length is selected as 19 years to maximise the number of eligible stations on the consideration that a higher cut-off would reduce that number. The annual maximum flood data sets for the selected catchments are collated based on a stringent data screening procedure e.g. gaps in the annual maximum flood series are in-filled as far as could be justified, outliers are detected using the multiple Grubbs-Beck test and error associated with rating curve extrapolation is investigated. The at-site flood frequency analyses are conducted using the FLIKE software. The potentially influential low flows are identified using multiple Grubbs-Beck test and are censored in the flood frequency analysis. A Bayesian parameter estimation procedure with LP3 distribution is used to estimate flood quantiles for each of the gauged site for annual exceedance probabilities of 50% to 1%. A total of five catchment predictor variables are included in the RFFE Model. The database prepared for ARR Project 5 is a valuable resource in Australian hydrology, which is expected to be progressively updated.*

## 1. INTRODUCTION

Design flood estimates are widely used in hydrologic design of bridge, culverts and levees and for planning and management of many other water resources, environmental and ecological management tasks. When recorded streamflow data of adequate length is available, at-site flood frequency analysis (FFA) is preferred to estimate design floods at the site of interest. For ungauged or poorly gauged catchments (e.g. streamflow record length is smaller than 10 years), flood quantiles are estimated using a regional flood frequency estimation (RFFE) method, which attempts to transfer information from a group of 'homogeneous' gauged catchments to the ungauged catchment of interest. A RFFE technique is expected to be simple, requiring readily obtainable input data to obtain design flood estimates relatively quickly. A RFFE model is a data driven approach and hence the quantity and quality of data is the core element of any RFFE model.

Two types of data are needed to develop a RFFE technique. The flood observations at gauged sites constitute the foundation of a RFFE model, and the quality and representativeness of the flood data determine to a large degree the accuracy and reliability of regional flood estimates. The challenge in collating a database for RFFE lies in maximising the amount of useful flood information, while practically minimising the random error component (or 'noise') that may be present in some flood data e.g. due to missing data, rating curve extrapolation and presence of outliers. In addition to gauged flood data, climatic and catchment characteristics data are needed to develop RFFE techniques, which serves as a 'bridge' between the gauged and ungauged sets of catchments.

In Australian Rainfall and Runoff (ARR) 1987, RFFE methods were based on streamflow data available till 1986 (I. E. Aust., 1987) and a good number of catchments had as little as ten years of streamflow data. Since 1987, there has been an additional streamflow data of over 25 years at many gauged locations in Australia. Haddad et al (2010) presented some of the relevant issues on streamflow data preparation in the development of a new RFFE model for Australia. To develop a national database, 'ARR Project 5 Regional flood methods' has been implemented in three stages (Stages 1, 2 and 3). The Stages 1 and 2 as reported in Rahman et al (2009 and 2012) covered the streamflow data length till 2005. In Stage 3, the streamflow data till 2011/2012 has been prepared at majority of the selected stations (Rahman et al., 2015a). Also, in the preparation of streamflow data and conducting at-site flood frequency analysis, a number of recent developments have been adopted in Stage 3. Based on these data, a regional flood estimation model, known as RFFE Model 2015 has been developed.

In this paper, we present an overview of the ARR Project 5 final data preparation so that the users of ARR RFFE Model 2015 have a better understanding of the strengths and limitations of the primary database on which the RFFE Model 2015 is founded. The final database used in developing the RFFE Model 2015 is archived with Engineers Australia (ARR Revision Team).

## 2. BACKGROUND OF THE RFFE TECHNIQUE IN THE 4TH EDITION OF ARR

The research on the development the RFFE technique started in 2006 by a team of Australian hydrologists and has been continued till 2015. This has produced a significant body of knowledge on RFFE in the form of technical reports (Rahman et al, 2009, 2012, 2015a) and refereed articles (e.g. Haddad et al, 2010; Rahman et al, 2011; Haddad et al, 2011; Haddad and Rahman, 2012; Haddad and Rahman, 2011a, 2011b, Haddad et al, 2012; Palmen and Weeks, 2011; Zaman et al, 2012; Haddad et al, 2015; Micevski et al, 2015; Rahman et al, 2015b; Ishak et al., 2011, 2013; Ishak and Rahman, 2015).

## 3. DATABASE USED IN DEVELOPING THE RFFE TECHNIQUE

In the new RFFE technique, Australia has been divided into humid coastal and arid/semi-arid areas. Humid coastal areas are situated in the high rainfall areas (with mean annual rainfall exceeding 500 mm) and include most of the coastal belts of Australia (Figure 1). The arid/semi-arid areas have mean annual rainfall values smaller than 500 mm.

An upper limit of catchment size of 1,000 km2 was adopted as a general guide to select eligible catchments. However, in Tasmania and the NT, a few larger catchments were included as the total numbers of catchments with areas less than 1,000 km2 were too small. The cut-off record length was selected as 19 years to maximise number of eligible stations on the consideration that a higher cut-off would reduce that number. The selected streams are essentially unregulated since major regulation affects the rainfall-runoff relationship significantly (e.g. storage effects). Streams with minor regulation, such as small farm dams and diversion weirs, are not excluded because this type of regulation is unlikely to have a significant effect on large annual floods.

The streamflow data for the initially selected potential catchments were further examined, as detailed in Haddad et al (2010), Ishak et al (2013) and Rahman et al (2015a): gaps in the annual maximum flood series were filled as far as could be justified, outliers were detected using the multiple Grubbs-Beck (MGB) test (Lamontagne et al, 2013), error associated with rating curve extrapolation was investigated and presence of trend was checked. From an initial number of about 1,200 catchments, a total of 798 catchments were finally adopted from all over Australia (excluding catchments in the arid regions, where some of the above criteria were relaxed). The record lengths of the annual maximum flood series of these 798 stations range from 19 to 102 years (median: 37 years). The catchment areas of the selected 798 catchments range from 0.5 km2 to 4,325 km2 (median: 178 km2). The geographical distribution of the selected 798 catchments is shown in Figure 1 and Table 1 presents a summary of the selected catchments.

In ARR1987, only a few catchments were used from arid/semi-arid areas to develop RFFE methods. Since the publication of ARR1987, there has been little improvement in terms of streamflow data availability in most of the arid/semi-arid areas of Australia. In the preparation of the regional flood estimation database, only a handful of catchments from the arid/semi-arid areas satisfy the selection criteria. To increase the number of stations from the arid/semi-arid areas to develop a ‘reasonably meaningful’ RFFE method, the selection criteria are relaxed i.e. the threshold streamflow record length is reduced to 10 years and the limit of catchment size is increased from 1,000 km2 to 6,000 km2. These criteria result in the selection of 55 catchments from the arid/semi-arid areas of Australia (Figure 1 and Table 1). The selected catchments have average annual rainfall smaller than 500 mm. The catchment areas range from 0.1 to 5,975 km2 (median: 259 km2) and streamflow record lengths range from 10 to 46 years (median: 27 years).

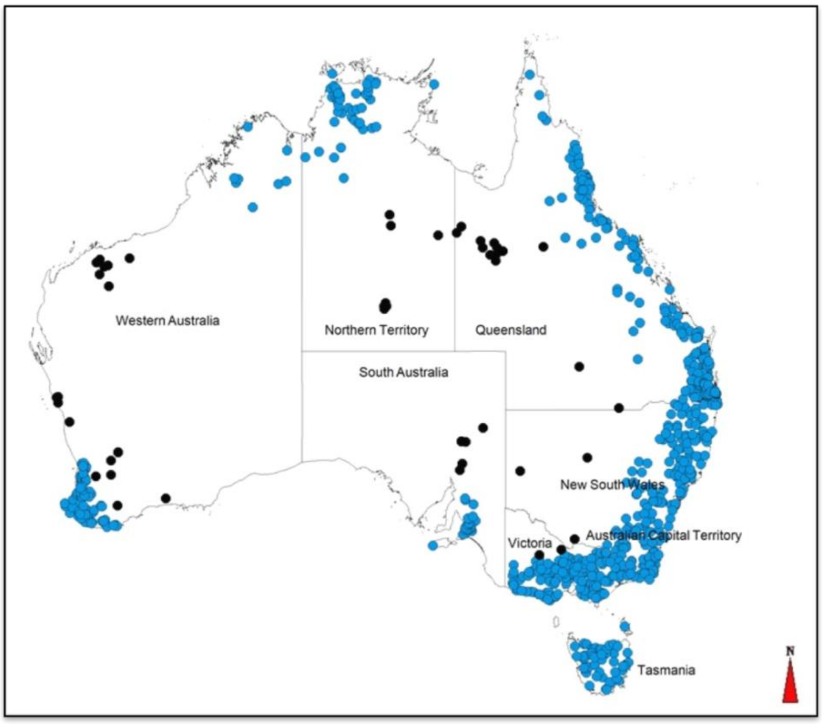


**Figure 1 Location of selected 853 catchments: humid coastal (blue circles) and arid/semi-arid areas (black)**

**Table 1. Summary of the selected 853 catchments**

| State | No. of stations | Streamflow record length (years) (range and median) | Catchment size (km$^2$) (range and median) |
|---|---|---|---|
| New South Wales & Australian Capital Territory | 176 | 20 – 82 (34) | 1 – 1036 (204) |
| Victoria | 186 | 20 – 60 (38) | 3 – 997 (209) |
| South Australia | 28 | 20 – 63 (37) | 0.6 – 708 (62.6) |
| Tasmania | 51 | 19 – 74 (28) | 1.3 – 1900 (158.1) |
| Queensland | 196 | 20 – 102 (42) | 7 - 963 (227) |
| Western Australia | 111 | 20 – 60 (30) | 0.5 – 1049.8 (49.2) |
| Northern Territory | 50 | 19 – 57 (42) | 1.4 - 4325 (352) |
| **Sub Total** | **798** | 19 – 102 (37) | 0.5 – 4325 (178.5) |
| Arid/semi-arid areas | 55 | 10 – 46 (27) | 0.1 - 5975 (259) |
| **TOTAL** | **853** | 10 – 102 (36) | 0.1 – 5975 (181) |

The at-site flood frequency analyses (for both the humid coastal and arid/semi-arid areas were conducted using the FLIKE software (Kuczera, 1999). The potentially influential low flows (PILFs) were identified using multiple Grubbs-Beck (MGB) test (Lamontagne et al, 2013) and were censored in the flood frequency analysis. A Bayesian parameter estimation procedure with LP3 distribution was used to estimate flood quantiles for at the gauged site for AEPs of 50%, 20%, 10%, 5%, 2% and 1%.

Over ten predictor variables were selected initially; however, it was found that the accuracy of a RFFE technique was not increased with the number of adopted predictor variables. A total of five predictor variables were finally adopted in the RFFE technique, which are (i) catchment area in km$^2$ (area); (ii) design rainfall intensity at catchment centroid (in mm/h) for 6-hour duration and AEP of 50% (I6,50); (iii) design rainfall intensity at catchment centroid (in mm/h) for 6-hour duration and AEP of 2% (I6,2); (iv) ratio of design rainfall intensities of AEPs of 2% and 50% for duration of 6-hour (I6,2/I6,50); and (v) shape factor, which is defined as the shortest distance between catchment outlet and centroid divided by the square root of catchment area. Design rainfall values were extracted using the 2013 intensity-frequency-duration (IFD) data from Australian Bureau of Meteorology website.

## 4. IMPORTANT FEATURES OF THE RFFE DATABASE

The Potentially Influential Low Flows (PILFs) are identified using the MGB test. In case of NSW, out of 188 sites, 66 do not require any censoring and for 20 catchments, 40% to 50% AM data points need censoring. Although, in few cases the results might be though to be ‘unusual’ e.g. about 50% of the annual maximum flood (AMF) data points need to be censored, in our investigation, it is found to be consistent with the judgement of experienced hydrologists who often adopt an interactive censoring. Interestingly, the flood quantile results based on MGB censoring agree very well with the GEV-L moments method, which provide an added assurance for the applicability of the MGB test.

We use FLIKE (Kuczera, 1999) to assess the impact of rating curve error on flood quantile estimates. We adopt two values of coefficient of variation (CV) (0%, which implies ‘no rating curve error case’; and CV = 20%). It is found that a higher CV value increases the band of the confidence limits of flood quantiles, in particular at smaller AEPs. The expected quantiles show notable differences between CV = 0% and CV = 20%. In most cases, the expected quantile estimates increase as CV increases. Moreover, the difference in quantile estimates between CV = 0% and CV = 20% increase with a decrease in AEP. In the development of RFFE Model 2015, it is decided to take flood quantile estimates with CV = 0%, since it is felt that more research needs to be undertaken to understand the implication of rating curve error on flood quantile estimates.

The distribution of standard deviation (SD) of loge(Q) values (where Q is AMF data) for humid coastal areas is shown in Figure 2. About 54% of the stations in the humid coastal areas have SD values in the range of 0.50 to 1.00 m3/s, 22% stations in the range of 0.014 to 0.49 m3/s, 21% stations in the range of 1.00 to 1.49 m3/s, and only 5% stations in the range of 1.50 to 2.201 m3/s. The SD values do not show any relationship with catchment size (Figure 3). The distribution of skew of loge(Q) values in the humid coastal areas is shown in Figure 4. About 82% of the stations have skew in the range of -

0.50 to 0.50. About 25% of the stations have skew in the range of -0.10 to 0.10. The skew values do not show any relationship with catchment size (Figure 5). It is found that most of the Australian stations do not show zero skew in the logarithm domain.

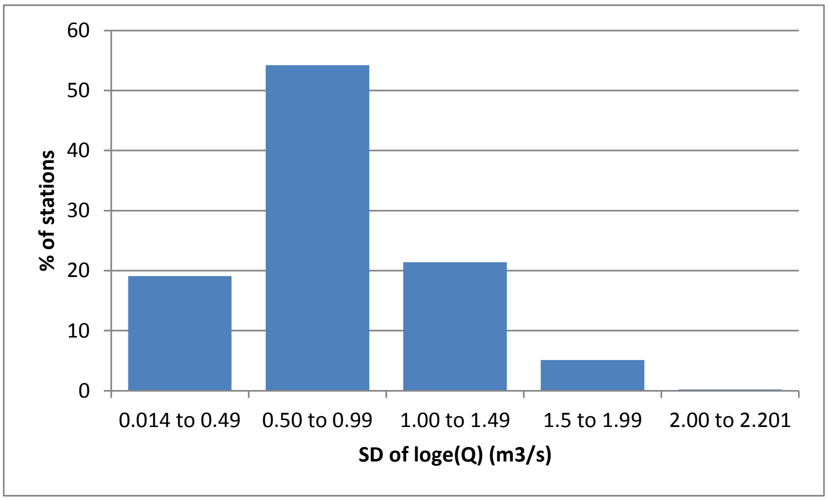


**Figure 2 Distributions of SD of loge(Q) in humid coastal areas (798 stations)**

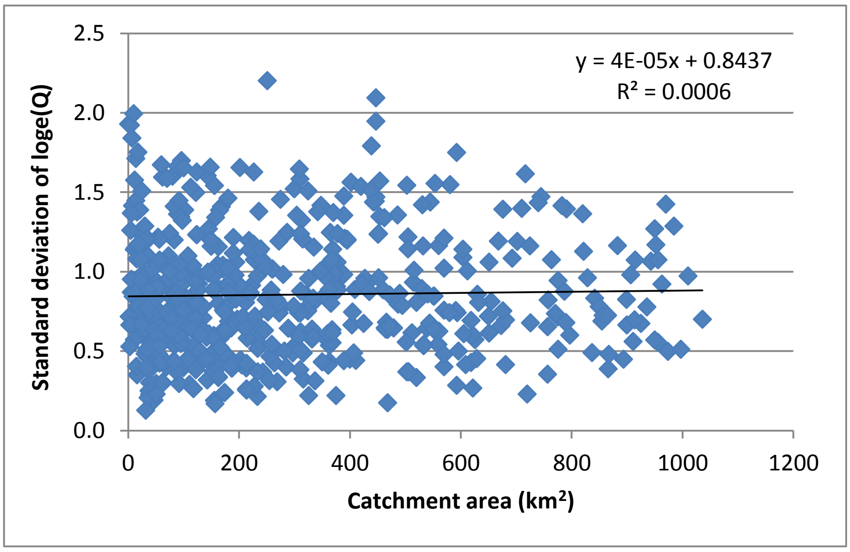


**Figure 3 Catchment area vs. standard deviation of loge(Q) (Q = AMF series) (east coast)**

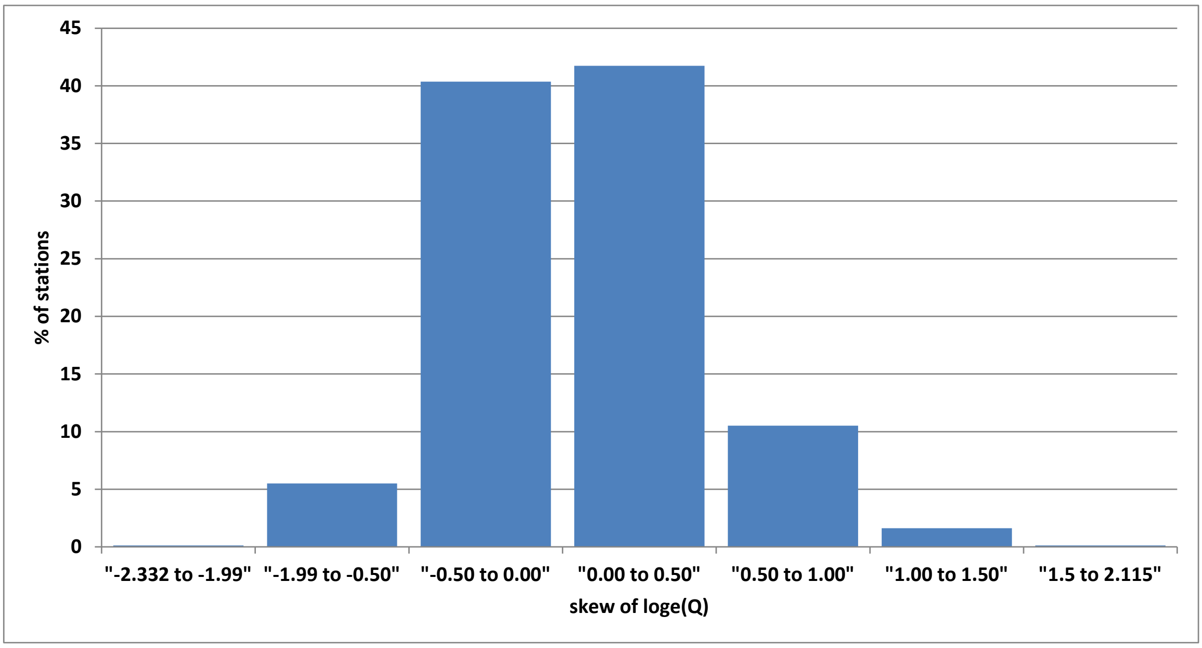


**Figure 4 Distributions of skew of loge(Q) (humid coastal areas: 798 stations)**

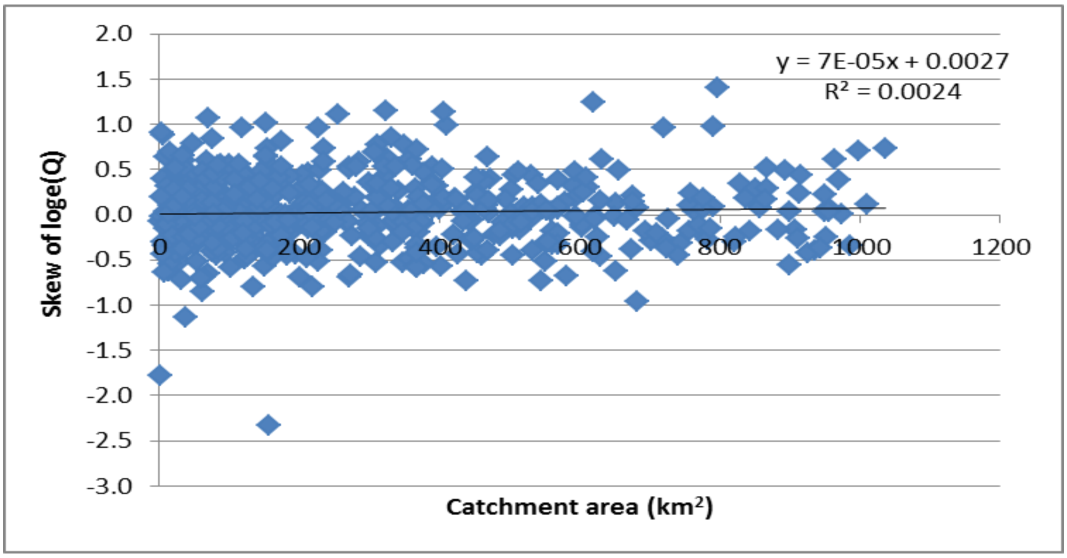


**Figure 5 Catchment area vs. skew of loge(Q) (Q = AMF flood series) (Region 1)**

Flood quantiles are found to have a strong correlation with catchment size (Figure 6 for east coast). The average exponent for catchment area of the log-log relationship between flood quantile and catchment area is found to be 0.75, the coefficient of determination ($R^2$) values of the regression equations range from 0.35 to 0.89 (average = 0.70). The smallest $R^2$ value is found for east coast region (0.35 to 0.38) and the highest $R^2$ for south west Western Australia (0.78 to 0.85).

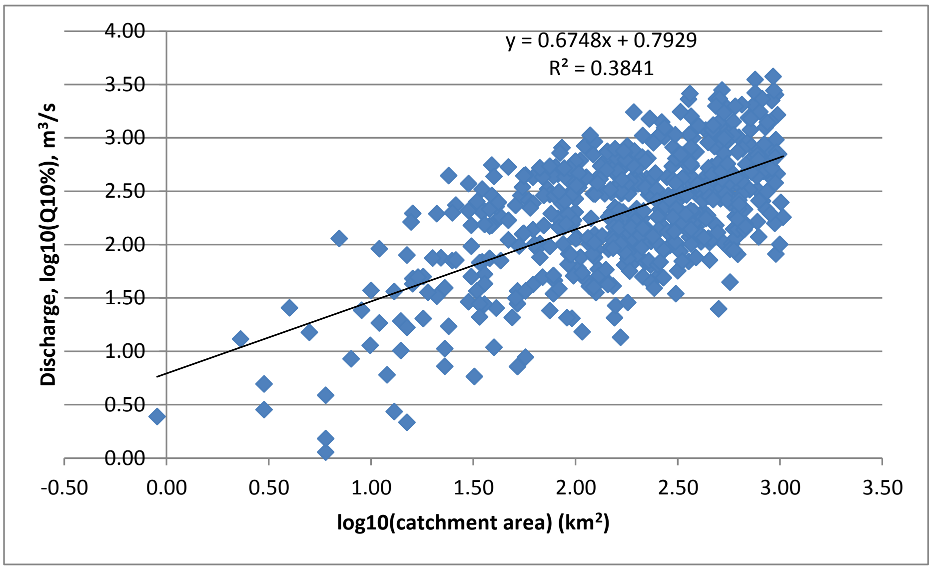


**Figure 6 Flood discharge (Q10% AEP) vs. catchment area (east coast)**

## 5. CONCLUSION

A database consisting of 853 gauged catchments has been prepared to develop and test the ARR RFFE Model 2015 for Australia. A total of 798 catchments are selected from the humid coastal areas and 55 catchments from the arid/semi-arid areas. It has been found that standard deviation and skew of the natural logarithms of annual maximum flood series do not show any significant link with catchment area. Flood quantiles are found to have a strong correlation with catchment size with coefficient of determination in the range of 0.35 to 0.89 with an average of 0.70. A total of five catchment characteristics are selected and data abstracted. The final database used in developing the ARR RFFE Model 2015 is archived with Engineers Australia (ARR Revision Team). This database is a valuable resource in Australian hydrology, which can be progressively updated as new records become available.

## 6. ACKNOWLEDGEMENTS

ARR Revision Project 5 was made possible by funding from the Federal Government through the Department of Climate Change and Energy Efficiency and Geoscience Australia. Project 5 reports and the associated publications including this paper are the result of a significant amount of in kind hours provided by Engineers Australia Members. The authors would also like to acknowledge various agencies in Australia for supplying data: Australian Bureau of Meteorology, DSE (Victoria), Thiess Services (Victoria), DTM (Qld), DERM (Qld), ENTURA, DPIPWE (TAS), DWLBC (SA), 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 (WA) and WMA Water (NSW).

## 7. REFERENCES

Haddad, K., Rahman, A., Weinmann, P.E., Kuczera, G., 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., Rahman, A. (2011a). Regional flood estimation in New South Wales Australia using Generalised Least Squares Quantile Regression, Journal of Hydrologic Engineering, ASCE, 16, 11, 920-925.

Haddad, K., Rahman, A. (2011b). 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.

Haddad, K., Rahman, A., Kuczera, G. (2011). Comparison of ordinary and generalised least squares regression models in regional flood frequency analysis: A case study for New South Wales, Australian Journal of Water Resources, 15, 2, 59-70.

Haddad, K., 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.

Haddad, K., Rahman, A., Stedinger, J.R. (2012). Regional Flood Frequency Analysis using Bayesian Generalized Least Squares: A Comparison between Quantile and Parameter Regression Techniques, Hydrological Processes, 26, 1008-1021.

Haddad, K., Rahman, A., Ling, F. (2015). Regional flood frequency analysis method for Tasmania, Australia: A case study on the comparison of fixed region and region-of-influence approaches, Hydrological Sciences Journal, DOI:10.1080/02626667.2014.950583.

Institution of Engineers Australia (I. E. Aust.) (1987). Australian Rainfall and Runoff: A Guide to Flood Estimation, Editor: Pilgrim, D.H., Engineers Australia, Canberra.

Ishak, E., Haddad, K., Zaman, Rahman (2011). Scaling property of regional floods in New South Wales Australia, Natural Hazards, 58: 1155-1167.

Ishak, E., Rahman, A., Westra, S., Sharma, A., Kuczera, G. (2013). Evaluating the Non-stationarity of Australian Annual Maximum Floods. Journal of Hydrology, 494, 134-145.

Ishak, E., Rahman, A. (2015). Detection of changes in flood data in Victoria, Australia over 1975-2011, Hydrology Research, doi:10.2166/nh.2014.064.

Kuczera, G. (1999). Comprehensive at-site flood frequency analysis using Monte Carlo Bayesian Inference, Water Resources Research, 35(5), 1551-1557.

Lamontagne, J. R., Stedinger, J. R., Cohn, T. A. and Barth, N.A. (2013*). Robust National Flood Frequency Guidelines: What is an Outlier?* World Environmental and Water Resources Congress 2013, 2013, pp. 2454-2466.

Micevski, T., Hackelbusch, A., Haddad, K., Kuczera, G., Rahman, A. (2015). Regionalisation of the parameters of the log-Pearson 3 distribution: a case study for New South Wales, Australia, Hydrological Processes, 29, 2, 250-260.

Palmen, L.B. and Weeks, W.D. (2011). Regional flood frequency for Queensland using the quantile regression technique, Australian Journal of Water Resources, 15, 1, 47-57.

Rahman, A., Haddad, K., Kuczera, G., Weinmann, E. (2009). *Australian Rainfall and Runoff Revision Project 5: Regional Flood Methods*. Stage 1 report, Engineers Australia.

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, 127-137.

Rahman, A., Haddad, K., Zaman, M., Ishak, E., Kuczera, G., Weinmann, P.E. (2012). Australian Rainfall and Runoff Revision Projects, Project 5 Regional flood methods, Stage 2 Report.

Rahman, A., Haddad, K., Rahman, A.S., Haque, M.M. (2015a). Australian Rainfall and Runoff Project 5: Regional flood methods: Database used to develop ARR REEF Technique 2015, Technical Report, No. P5/S3/026, Engineers Australia, Water Engineering, 68pp.

Rahman, A., Haddad, K., Haque, M., Kuczera, G., Weinmann, P.E. (2015b). Australian Rainfall and Runoff Project 5: Regional flood methods: Stage 3 Report, Technical Report, No. P5/S3/025, Engineers Australia, Water Engineering, 134pp.

Zaman, M., Rahman, A., Haddad, K. (2012). Regional flood frequency analysis in arid regions: A case study for Australia, Journal of Hydrology, 475, 74-83.