

Regional Flood Modelling in the New Australian Rainfall and Runoff

A.S. Rahman^a, K. Haddad^a, A. Rahman^a

^a *School of Computing, Engineering and Mathematics, University of Western Sydney, Australia*
E-mail: a.rahman@uws.edu.au

Abstract: This paper presents the regional flood estimation model being developed for the 4th edition of Australian Rainfall and Runoff (ARR). This covers the basic steps in the adopted modelling framework i.e. data collation, formation of regions, model form, and estimation of the parameters of the model. The adopted modeling framework revolves around the principles of reduction of uncertainty and the best possible use of the regional data by the reduction of the regional heterogeneity to enhance the accuracy of at-site design flood estimation. The ARR regional flood estimation model uses a Bayesian generalised least squares regression-based approach in the data-rich regions of Australia that consists of 619 stations. In the formation of the regions, it divides Australia into a number of regions. A region-of-influence approach is adopted to form regions where there are over 50 stations in the region. It also adopts a parameter regression technique where three parameters of the Log Pearson Type 3 distribution are regionalized. A Monte Carlo simulation technique is adopted to estimate the model uncertainty. The RFFA model is being re-calibrated with the new IFD data and the most up-to-date streamflow records and is expected to be released by Dec 2013.

Keywords: *Regional floods, ungauged catchments, Australian Rainfall and Runoff, GLS regression*

1. INTRODUCTION

Flood is one of the worst natural disasters that causes remarkable damage and inconveniences to both urban and rural communities. To plan for new developments and to safeguard existing ones, flood estimation is needed. This is also needed for many other purposes such as flood insurance studies, emergency evacuation and ecological studies. Flood estimation is made by hydrologists based on the best available data and information at hand. Most important of these data is the recorded flood data at the point of interest; however, that may not be available at many locations where flood estimation is needed. Moreover, the length and quality of data may not be adequate to make reliable estimation. Additional challenges are imposed by land use changes over the period of historical data availability and the possible non-stationarity caused by climate change. The time, expertise and tools available for flood estimation can also make a difference.

The most direct method of flood estimation is at-site flood frequency analysis, which requires a sufficiently long period of recorded flood data at the site of interest. At many locations of interest, recorded streamflow data are limited or completely absent (ungauged situation). Regional flood frequency analysis (RFFA) methods are used to estimate design floods in such ungauged catchments. Design flood estimates at ungauged locations has many direct applications e.g. design of culverts, small to medium sized bridges, causeways, farm dams, soil conservation works and for many other water resources management tasks. This can also serve as a 'benchmark' for checking the consistency of rainfall-based design flood estimates. Also, estimates from ungauged locations may serve as prior information in Bayesian type at-site flood frequency analysis. A RFFA is essentially a data-based empirical procedure which attempts to compensate for the lack of temporal data at a given location by spatial data i.e. the flood data collected at other locations within a 'homogeneous region'.

Australian Rainfall and Runoff (ARR) 1987 recommended various regional flood estimation techniques for small to medium sized ungauged catchments for different regions of Australia (I. E. Aust., 1987). Since 1987, the regional flood estimation methods in ARR have not been upgraded, although there have been an additional 25 years of streamflow data available and notable developments in both at-site and regional flood frequency analyses techniques. As a part of the current revision of the ARR (4th Edition), Project 5 Regional Flood Methods for Australia focuses on the development, testing and recommendation of new regional flood estimation methods for Australia by incorporating the latest data and techniques. This paper presents an overview of the RFFA model that is being investigated as a part of ARR Project 5.

2. DATABASE USED IN DEVELOPING THE NEW RFFA MODEL IN AUSTRALIA

Initially, the streamflow database of each individual state was consulted. Data for each station was checked and prepared using an appropriate gap-filling method, the outliers were identified, the rating curve error was assessed, and time trend in the annual maximum flood series data was examined. A detailed description of the data preparation procedure can be found in ARR Project 5 Stage 1 and Stage 2 reports (Rahman *et al.*, 2009 and 2012a) and in Haddad *et al.* (2010). The annual maximum flood series data may be affected by multi-decadal climate variability and climate change, which are not easy to deal with. The effects of multi-decadal climate variability can be accounted for by increasing the cut-off record length at individual stations; however, the impacts of climate change present a serious problem in terms of the applicability of the past data in predicting future flood risk. In this study, the assumption that the annual maximum flood series data at an individual candidate station satisfies the assumption of stationarity was checked by applying trend tests (Ishak *et al.*, 2010; 2013). About 15% stations exhibited significant time trend, which were left for further analysis and were not used to develop the prediction equations presented in this paper. A total of 619 catchments were finally adopted from the data rich regions of Australia (shown in Figure 1).

For the data poor semi-arid and arid region, some of the above criteria were relaxed, as discussed in Rahman *et al.* (2012a and 2012b). The record lengths of the annual maximum flood series of the selected stations from NSW, ACT, VIC and QLD states range from 25 to 97 years and the catchment areas range from 3 to 1,010 km². For the remaining states, the threshold record length was reduced to 18 years and maximum catchment size was increased to 7,405 km² to increase the size of the database, as summarised in Table 1. The geographical distribution of the selected catchments is shown in Figure 1.

The at-site flood frequency analyses were conducted using the FLIKE software (Kuczera, 1999). A Bayesian parameter estimation procedure with the Log-Pearson Type 3 (LP3) distribution was used to estimate flood quantiles for average recurrence intervals (ARIs) of 2 to 100 years. The following seven predictor variables were selected: (i) catchment area expressed in km² (area); (ii) design rainfall intensities (mm/h) for fixed durations of 1 and 12 hours and ARIs of 2 and 50 years ($I_{1,2}$), ($I_{12,2}$), ($I_{1,50}$) and ($I_{12,50}$), or design rainfall

intensity values $I_{t_c,ARI}$ for durations t_c (hours) matched to the catchment size (estimated from $t_c = 0.76(area)^{0.38}$) and ARI corresponding to the flood estimate (ARI = 2, 5, 10, 20, 50 and 100 years); (iii) mean annual rainfall expressed in mm/y (rain); (iv) mean annual areal potential evapo-transpiration expressed in mm/y (evap); (v) stream density expressed in km/km² (sden); (vi) main stream slope expressed in m/km (S1085); and (vii) forest cover expressed as a fraction of catchment area (forest). The forested area in a catchment was obtained from 1:100,000 topographic maps where forest area includes tropical rainforest, large gatherings of trees, high-density scattered shrubs, and medium-density scattered shrubs. On a topographic map, these were located by looking for the green areas indicating forests. The slope S1085 excludes the extremes of slope found at either end of the mainstream. It is the ratio of the difference in elevation of the stream bed at 10% and 85% of its length from the catchment outlet, and 75% of the main stream length. This slope was determined from 1:100,000 topographic maps, using an opisometer to measure the stream length. The mean annual rainfall and mean annual areal potential evapo-transpiration values were obtained from a CD-ROM provided by the Australian Bureau of Meteorology. The design rainfall intensity values were obtained at the catchment outlet location from ARR Volume II. The mean annual rainfall, mean annual areal potential evapo-transpiration and design rainfall intensity values were obtained at the catchment outlet location. For catchments in QLD, NT and WA some of the predictor variables data such as stream density and forest were not obtained.

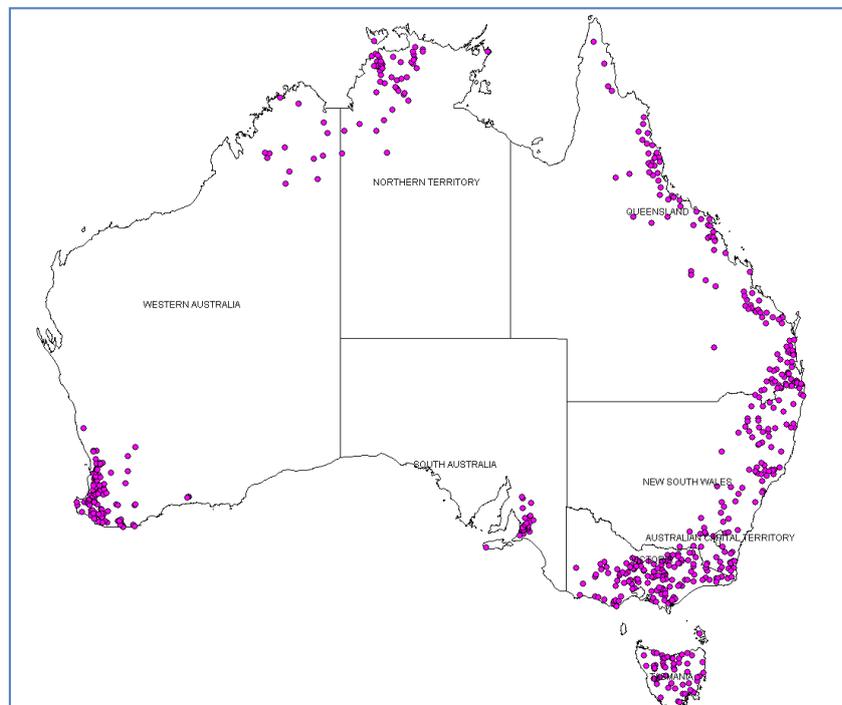


Figure 1. Streamflow data used in ARR RFFA from data-rich region of Australia.

3. ADOPTED RFFA MODEL

A number of RFFA models were developed and tested using the national database of 619 stations. These include the Probabilistic Rational Method (PRM) (IE Aust., 1987) and various regression based techniques: Quantile Regression Technique (QRT) based on Ordinary Least Squares (QRT-OLS) and Generalised Least Squares (QRT-GLS) (Tasker and Stedinger, 1989; Reis *et al.*, 2005), and Parameter Regression Technique (PRT) based on GLS regression (PRT-GLS). In the PRT, prediction equations were developed for the three parameters (i.e. mean, standard deviation and skewness) of the log-transformed annual maximum flood series to estimate flood quantiles from a regional LP3 distribution. Rahman *et al.* (2011) and Palmen and Weeks (2011) compared the QRT against the PRM and found that QRT outperformed the PRM. Haddad *et al.* (2011a, b, c), Haddad and Rahman (2011, 2012) and Hackelbusch *et al.* (2009) found that the QRT-GLS method outperformed the QRT-OLS method while a region-of-influence (ROI) approach (Burn, 1990)

outperformed the fixed region approach. The selected ROI contained the nearest N stations in geographical space, where N was selected so as to minimize the predictive error which accounts for both model and regression coefficients uncertainty. This strategy seeks to minimise the heterogeneity unaccounted for by the regression predictors. It should be noted here that the GLS regression offers a powerful statistical technique to estimate the flood quantiles which accounts for the inter-station correlation of annual maximum flood series and across-site variation in flood series record lengths. It also differentiates between sampling error and model error and thus provides a more realistic framework for error analysis. It was found that the QRT and PRT methods performed very similarly for various Australian states (Haddad et al, 2012). However, the PRT method offers several practical advantages over the QRT: (i) PRT flood quantiles increase smoothly with decreasing annual exceedance probability (AEP); (ii) flood quantiles of any ARI (in the range of AEPs of 50% to 1%) can be estimated from the regional LP3 distribution; and (iii) 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. For these reasons, the PRT coupled with Bayesian GLS regression was finally adopted for general application to Australia in the data-rich region.

Table 1. Summary of adopted stations from data rich regions of Australia (the figures in parentheses indicate the median values)

State	No. of stations	Streamflow record length (years)	Catchment size (km ²)
NSW & ACT	96	25 - 75 (34)	8 - 1,010 (267)
Victoria	131	26 - 52 (33)	3 - 997 (289)
Queensland	126	25 - 97 (36)	7 - 963 (254)
Tasmania	52	19 - 74 (28)	1.3 - 1,900 (158)
South Australia	29	18 - 67 (34)	0.6 - 708 (76)
Northern Territory	55	19 - 54 (33)	1.4 - 4,325 (360)
Western Australia	134	20 - 57 (30)	0.1 - 7,405 (60)
TOTAL	619		

In the ROI approach, for each of the gauged sites in the region, the regional prediction equation has a different set of model coefficients; hence a single regional prediction equation cannot be pre-specified. To overcome this problem, the parameters of the regional prediction model for all the gauged catchment locations in a ROI region have been pre-estimated and integrated with an application tool. In the application tool, for a given location of the ungauged catchment, the parameter set of the nearest ROI sub-region is automatically selected for flood quantile estimation. For the region of South Australia there are fewer than 50 gauged stations, and also these stations are located far away from the stations of the adjacent states. In this case, the application of ROI is not feasible, and hence a fixed region approach has been adopted where all the available stations within the state are included in a single fixed region.

In the adopted PRT, the first three moments of the log-transformed annual maximum flood series (i.e. the mean, standard deviation and skewness) are regionalised and used to estimate flood quantiles from the Log Pearson Type 3 (LP3) distribution, which is described by the following equation:

$$\ln Q_T = M + K_T S \quad (1)$$

where Q_T = the discharge having an AEP of $1/T$ (where T is return period); M = mean of the natural logarithms of the annual maximum flood series; S = standard deviation of the natural logarithms of the annual maximum flood series; and K_T = frequency factor for the LP3 distribution for AEP of $1/T$, which is a function of AEP and skewness). The prediction equations for the mean, standard deviation and skewness are developed for the fixed region and at all the gauged catchment locations in the ROI regions using Bayesian GLS regression. 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 the LP3 distribution and to estimate the flood quantiles in the AEPs of 50% to 1%. The adopted method is referred to as the 'ARR RFFE 2012' method.

4. RESULTS

In the ARR RFFE 2012 method, four ROI regions and two fixed regions are adopted (as shown in Table 2): (i) VIC + ACT + NSW + QLD: 353 sites (Region 1); (ii) TAS: 52 sites (Region 2); (iii) NT + Kimberley region of WA (Region 4): 65 sites; and (iv) South-West WA (Region 5): 120 sites. For SA, a fixed region is adopted containing 29 stations (Region 3). For the data poor semi-arid and arid region, a fixed region is adopted containing 57 stations. The boundaries between the semi-arid and arid (data-poor) and coastal (data-rich) regions have been drawn approximately based on the 500 mm mean annual rainfall contour (Rahman *et al.*, 2013).

Table 2. ARR RFFE 2012 method: regions in Australia

Region Code	Description	Method to form region	Number of gauged stations	Estimation model
Region 1	VIC + ACT + NSW + QLD	ROI	353	Bayesian GLS-PRT
Region 2	TAS	ROI	52	
Region 3	SA	Fixed region	29	
Region 4	NT + Kimberley of WA	ROI	65	
Region 5	South-west WA	ROI	120	
Region 6	Data poor semi-arid and arid region	Fixed region	57	Index type RFFE method
Fringe A	Between Region 1 and Region 6	ROI	-	Inverse distance weighted average
Fringe B	Between Region 3 and Region 6	ROI	-	
Fringe C	Between Region 4 and Region 6	ROI	-	
Fringe D	Between Region 5 and Region 6	ROI	-	
Fringe E	Between Region 1 and Region 4	ROI	-	

To reduce the effects of sharp variation in quantile estimates for the ungauged catchments located close to these regional boundaries, five fringe zones are delineated as shown in Table 2. For these fringe zones, flood quantiles at an ungauged catchment location are taken as the ‘inverse distance weighted average value’ of the two nearby regional estimates.

For all the data-rich regions (shown in Table 2), the adopted prediction model form is given below:

$$M = b_0 + b_1(\text{area}) + b_2(I_{12,2}) \quad (2)$$

$$S = c_0 \quad (3)$$

$$\text{skewness} = d_0 \quad (4)$$

Here, M is mean of the natural logarithms of the annual maximum flood series; S = standard deviation of the natural logarithms of the annual maximum flood series; b_0 , b_1 and b_2 are regression coefficients, estimated using the method and database noted in Table 2; area represents catchment area in km^2 , $I_{12,2}$ represents design rainfall intensity for 12 hours duration and 2 years ARI. The c_0 and d_0 are the regional weighted average values of standard deviation and skewness of the natural logarithms of the annual maximum flood series; for a fixed region, all the gauged catchments within the region are used to obtain the regional average values; and for the ROI, an appropriate number of sites (based on ROI analysis) are used to obtain the regional average values at a gauged catchment location. The regional average values were used, as no significant predictor variables were found to model the standard deviation and skewness. The values of b_0 , b_1 , b_2 , c_0 , and

d_0 for the adopted region (in the case of fixed regions) or at all the individual gauged catchment locations (in the case of ROI) will be embedded in the application tool. The application tool selects the appropriate set of values for b_0 , b_1 , b_2 , c_0 , and d_0 for the ungauged catchment of interest given its region and location (in terms of latitude and longitude). For the ROI regions, the values for b_0 , b_1 , b_2 , c_0 , and d_0 for the ROI sub-region located closest to the ungauged catchment of interest are selected in application tool (Rahman *et al.*, 2013).

In developing the confidence limits for the estimated flood quantiles, a Monte Carlo simulation approach is adopted by assuming that the first three moments of the log-transformed annual maximum flood series can be specified by a multivariate normal distribution. Here the correlations of the three moments are estimated from the model data set. The mean and standard error values of log-transformed annual maximum flood series are estimated from the Bayesian GLS regression. Based on 10,000 simulated values of the moments, 10,000 Q_T values are estimated, which are then used to develop the 90% confidence intervals. An application tool has been developed called ARR RFEM 2012, which facilitates the application of the adopted RFFE model. The application is designed to be simple but flexible to use.

There are number of updates to be implemented by 2013 to finalise the ARR RFFA model. (i) At present, the regional flood estimation model with the ARR RFFE 2012 method has been based on annual maximum flow data up to 2005 and calibrated with the design rainfall ($I_{12,2}$) data from ARR1987 (I. E. Aust., 1987). An extended annual maximum flow data base is currently being prepared with data to the year 2011. With the new IFD data (as part of ARR Project 1) and the updated annual maximum flow data, the regional flood estimation model is being re-calibrated and model coefficients for various ROI and fixed regions are being updated in the application tool.

5. CONCLUSIONS

This paper presents the RFFA model being developed as a part of ARR revision Project 5 Regional Flood Methods. In the adopted ARR RFFA model, Australia has been divided into six regions and five fringe zones. A total of 619 catchments have been used to develop prediction equations for the data rich regions. A region-of-influence approach has been adopted to form regions where sufficient data was available and a Bayesian GLS regression technique has been adopted to develop prediction equations. A regionalised LP3 distribution is recommended to derive design flood estimates for ungauged catchments in the range of AEPs of 50% to 1%. For all the regions, the derived prediction equations use only two predictor variables, catchment area and a representative design rainfall intensity. An application tool has been developed which automates the application of the new RFFA model. The user is required to input latitude, longitude and catchment area to obtain design flood quantiles and associated uncertainty estimates. The RFFA model is being calibrated with the new IFD data and the most up-to-date streamflow records and is expected to be finalised by Dec 2013.

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, in particular, would like to acknowledge Professor George Kuczera, Mr Erwin Weiamann, Associate Professor James Ball, Dr William Weeks, Professor Ashish Sharma, Mr Peter Stensmyr, Mr Elias Ishak, Dr Tom Micevski, Mr Andre Hackelbusch, Ms Monique Retallick, Mr Md Jalal Uddin, Mr James Pirozzi, Mr Gavin McPherson, Mr Chris Randall, Mr Wilfredo Caballero, Mr Khaled Rima, Mr Tarik Ahmed, Mr Kashif Aziz, Ms Melanie Taylor, Mr Ahmed Derbas, Mr Fotos Melaisis, Mr Luke Palmen, Dr Seth Westra, Mr Syed Quddusi, Mr Robert French, Mohammed Abedin, Ms Tanmila Islam, Dr Guna Hewa, Ms Sithara Gamage, Ms Subhashini Hewage, Mr Trevor Daniell, Dr David Kemp, Dr Fiona Ling, Mr Crispin Smythe, Mr Chris MacGeorge, Mr Bryce Graham, Mr Lakshman Rajaratnam, Mr Mark Pearcey, Mr Jerome Goh, Dr Neil Coles, Ms Leanne Pearce, Mr John Ruprecht, Dr Robin Connolly, Mr Patrick Thompson. 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).

REFERENCES

- Burn, D.H. (1990). Evaluation of Regional Flood Frequency Analysis with a Region of Influence Approach. *Water Resources Research*, 26(10), 2257-2265.
- 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. *Hydrology and Water Resources Symp.*, Newcastle, 603-615.
- 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. and Rahman, A. (2011). 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., and Kuczera, G. (2011a). 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., and Stedinger, J.R. (2011b). Regional Flood Frequency Analysis using Bayesian Generalized Least Squares: A Comparison between Quantile and Parameter Regression Techniques, *Hydrological Processes*, 25, 1-14. 26, 1008-1021.
- Haddad, K., Rahman, A., Kuczera, G. and Micevski, T. (2011c). 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, 162-169.
- 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, 142-161.
- 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.H., Rahman, A., Westra, S., Sharma, A. and Kuczera, G. (2010). Preliminary analysis of trends in Australian flood data. *World Environmental and Water Resources Congress 2010*, American Society of Civil Engineers (ASCE), 16-20 May 2010, Providence, Rhode Island, USA, pp. 120-124.
- Ishak, E., Rahman, A., Westra, S., Sharma, A. and Kuczera, G. (2013). Evaluating the Non-stationarity of Australian Annual Maximum Floods. *Journal of Hydrology*, 494, 134-145.
- Kuczera, G. (1999). Comprehensive at-site flood frequency analysis using Monte Carlo Bayesian Inference. *Water Resources Research*, 35(5), 1551-1557.
- Micevski, T and Kuczera, G. (2009). Combining site and regional flood information using a Bayesian Monte Carlo approach. *Water Resources Research*, W04405, doi:10.1029/2008WR007173.
- 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., Zaman, M., Kuczera, G., Weinmann, P. E., Stensmyr, P., Babister, M. (2013). New regional flood frequency estimation method for the whole of Australia: Overview of progress, *Floodplain Management Association National Conference*, 28-31 May 2013, Tweed Heads, NSW.
- Rahman, A., Haddad, K., Zaman, M., Ishak, E., Kuczera, G. And Weinmann, P.E. (2012a). *Australian Rainfall and Runoff Revision Projects, Project 5 Regional flood methods, Stage 2 Report No. P5/S2/015*, Engineers Australia, 319pp.
- Rahman, A., Zaman, M., Haddad, K., Kuczera, G, Weinmann, P. E., Weeks, W., Rajaratnam, L. and Kemp, D. (2012b). Development of a New Regional Flood Frequency Analysis Method for Semi-arid and Arid Regions of Australia, *Hydrology and Water Resources Symposium*, Engineers Australia, 19-22 Nov 2012, Sydney, 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., Kuczera, G. and Weinmann, P.E. (2009). Regional flood methods for Australia: data preparation and exploratory analysis. *Australian Rainfall and Runoff Revision Projects, Project 5 Regional Flood Methods, Stage 1 Report No. P5/S1/003*, Engineers Australia, Water Engineering, 181pp.
- 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, (1) 1029.
- Tasker, G.D. and Stedinger, J.R. (1989). An operational GLS model for hydrologic regression. *Journal of Hydrology*, 111, 361-375.