Regional Flood Estimation for NSW: Comparison of Quantile Regression and Parameter Regression Techniques

F. Kader[a], A. Derbas[b], K. Haddad[b] and <u>A. Rahman[b]</u>

[a]*Private Consultant*
[b] *School of Computing Engineering and Mathematics and Institute for Infrastructure Engineering, Western Sydney University, Australia*
Email: faruk.kader718@gmail.com

Abstract: Australian Rainfall Runoff (ARR) (1987) recommended a number of regional flood frequency estimation (RFFE) techniques to estimate design floods in ungauged catchments in Australia. These include Probabilistic Rational Method (PRM) for eastern New South Wales (NSW) and Victoria, and Index Flood Method (IFM) for western NSW. The PRM method uses a probabilistic runoff co-efficient map based on linear geographical interpolation, while the criterion of regional homogeneity required by the IFM method is rarely satisfied for Australian regional flood data. Recent studies on regression-based RFFE techniques have demonstrated that these can provide quite accurate flood quantile estimates in Australia using only a few predictor variables. Two regression-based methods, Quantile Regression Technique (QRT) and Parameter Regression Technique (PRT) have been tested recently in Australia. The QRT develops regression equation for individual quantile; however, PRT develops regression equations of the parameters of a probability distribution. For example, in the case log-Pearson Type 3 (LP3) distribution, the PRT develops regression equations for mean, standard deviation and skew of the logarithms of annual maximum flood series data.

This study presents a comparison between the PRT and QRT methods for NSW State in Australia using data from 96 small to medium sized gauged catchments. Out of 96 catchments, 12 are selected randomly as test catchments to make an independent validation of the PRT and QRT methods. The results show that both the PRT and QRT perform very similarly, with relative error values in the range of 36% to 45%. Further studies using data from different geographical regions of Australia need to be conducted before recommending an appropriate RFFE method for application in Australia.

Keywords: *Ungauged catchments, quantile regression technique, parameter regression technique, annual maximum flood series, LP3 distribution*

1. INTRODUCTION

Regional flood frequency estimation (RFFE) techniques are widely used for estimation of design floods in ungauged catchments. Australian Rainfall Runoff (ARR) (1987) recommended a number of RFFE techniques for design flood estimation in ungauged catchments, e.g. Probabilistic Rational Method (PRM) for Victoria and eastern NSW and Index Flood Method (IFM) for Australian Capital Territory (ACT) and Western NSW. The rational method (Mulvany, 1851) is one of the most widely used techniques in flood estimation; however, it has some limitations (e.g. Pegram, 2002 and Rahman et al., 2011a). The PRM, recommended by ARR 1987 (Pilgrim, 1982), uses a map of the runoff coefficient (C_{10}), based on a simple linear geographical interpolation method that ignores discontinuities at catchment boundaries and the existence of watercourses at many locations. On the other hand, the criterion of regional homogeneity assumed by the IFM method is rarely satisfied in the case of Australian regional flood data (Bates et al., 1998) and hence the IFM has limited applicability in Australia.

Since ARR1987, a significant number of RFFE studies have been undertaken in Australia (e.g. Weeks, 1991, Bates et al., 1998; Rahman et al., 1999, Rahman, 2005; Palmen and Weeks, 2011; Haddad and Rahman, 2012a; Haddad et al., 2010; Haddad et al., 2012a, Haddad et al., 2012b; Haddad et al., 2015; Rahman et al., 2015a, Rahman et al., 2015b). In 2006, the National Committee on Water Engineering initiated upgrade of the RFFA techniques in ARR. As a part of the ARR upgrade, two regression-based methods, Quantile Regression Technique (QRT) (Stedinger and Tasker, 1985 and Griffis and Stedinger, 2007) and Parameter Regression Technique (PRT) have been tested (Rahman et al., 2015b). The QRT develops regression equation for individual quantile; however, PRT develops regression equations of the parameters of a given probability distribution. One of the limitations of the QRT is that the flood quantiles given by the QRT may not increase smoothly with decreasing annual exceedance probability (AEP). The PRT provides a number of significant advantages over QRT including the capability of estimation of flood quantiles for any AEP in the range of interest and quantile estimates vary smoothly with AEPs for PRT. In this regard, QRT and PRT have been compared by number of recent studies (e.g. Haddad et al., 2012a; Haddad et al., 2012b; Haddad and Rahman, 2012; Micevski et al., 2015).

This study applies a PRT method, which involves regionalization of the first three moments/parameters of the LP3 distribution. The reason for considering LP3 distribution is that it has been widely used in Australia and USA for flood frequency analysis. For example, Vogel et al. (1993) and Rahman et al. (2013) found that the LP3 is one of the best-fit distributions in Australia in modeling annual maximum floods. The objective of this study is to compare the QRT and PRT methods for New South Wales (NSW) State in Australia using an independent validation method.

2. DATA USED IN THE STUDY

This study considers 96 small to medium sized gauged catchments in New South Wales (NSW) (Figure 1). The catchments have a minimum 25 years of streamflow records and areas range between 8 and 1000 km^2. For developing the QRT and PRT, we adopt the following seven catchments characteristics. The summary statistics of the catchment characteristics are provided in Table 1.

- Design rainfall intensities of various durations and AEPs, e.g. 12-hour duration and 1 in 2 AEP (I_{12_2}, mm/h);
- Mean annual rainfall (*rain*, mm);
- Mean annual areal evapotranspiration (*evap*, mm);
- Catchment area (*area*, km^2);
- Slope of central 75% of mainstream *S1085* (*slope*, m/km);
- Stream density (*sden*, km/km^2); and
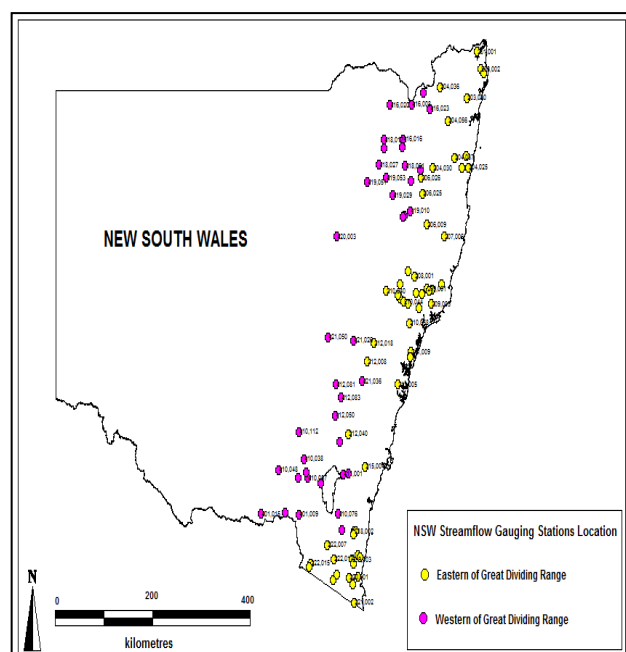- Fraction of basin covered by medium to dense forest (*forest*)

Figure 1. Study area and selected catchments.

From the selected 96 catchments, 12 catchments are selected randomly as test catchments, thus leaving 84 catchments for model development. This procedure is repeated 8 times to generate 8 different sets of model catchments (Model Set 1, Model Set 2, …) and 8 different sets of test catchments (Test Set 1, Test Set 2, …).

Table 1. Summary statistics of the catchment characteristics for the selected 96 gauged catchments.

Catchment Characteristics	Description	Mean	Std. Deviation
rain	Mean annual rainfall (mm)	983.0	323.0
evap	Mean annual area evapotranspiration (mm)	1222.04	134.0
area	Catchment area (km^2)	372.0	284.0
slope	Main stream slope S1085 (m/km)	11.88	9.40
sden	Stream density (km/km^2)	2.66	1.09
forest	Dense forest	0.49	0.3
I_{1_2}	1-hour duration, 1 in 2 AEP (mm/hr)	28.63	7.49
I_{12_2}	12-hour duration, 1 in 2 AEP (mm/hr)	6.06	2.05
I_{12_50}	12-hour duration, 1 in 50 AEP (mm/hr)	12.37	4.99

3. METHOD

Quantile Regression Technique (QRT) has the following general model form (Thomas and Benson, 1970):

$$Q_T = aB^b C^c D^d \dots$$
(1)

where *B, C, D,* … are catchment characteristic variables (predictors) and Q_T is the flood magnitude with AEP of 1 in *T*, and *a, b, c, d, ...* are regression coefficients.

QRT can yield design flood estimates that do not vary smoothly with AEP; however, adjustment can be made so that flood estimates increase smoothly with decreasing AEP; however, in this study, smoothing of the flood quantiles given by the QRT is not implemented.

Multiple linear regression technique in the SPSS is used to develop 8 sets of prediction equations for the 8 sets of model catchments for each of the seven AEPs (1 in 2, 1 in 5, 1 in 10, 1 in 20, 1 in 50, 1 in 100 and 1 in 200). The dependent variables are the corresponding seven flood quantiles, denoted by Q_2, Q_5, Q_{10}, Q_{20}, Q_{50}, Q_{100} and $Q_{200,}$ respectively, and the independent variables are the seven catchment characteristics (Table 1).

The LP3 distribution has the following model form:

$$\ln Q_T = M + K_T S \tag{2}$$

where Q_T = flood discharge having an AEP of 1 in *T* (design flood or flood quantile);

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 in *T*, which is a function of the AEP and skewness (*SK*) of the natural logarithms of the annual maximum flood series.

To develop the PRT, the moments of the LP3 distribution (*M, S* and *SK*) are regressed against catchment characteristics. The three regression equations in the PRT have the following model forms:

$$M = aB^b C^c D^d \dots \tag{3}$$

$$S = aB^b C^c D^d \dots \tag{4}$$

$$K_T = aB^b C^c D^d \dots \tag{5}$$

where *B, C, D, ...* are catchment characteristics variables (predictors) and *a, b, c, d, ...* are regression coefficients.

The multiple linear regression technique in the SPSS software is used to develop the prediction equations for both the PRT and QRT where an ordinary least squares method is used with 'backward' variable selection method.

4. RESULTS

The prediction equations developed by the QRT technique for Model Set 1 are provided in Table 2. The prediction equations are used to estimate flood quantiles for the test catchments and compared with the corresponding at-site flood quantiles, estimated by the LP3 distribution (Kuczera, 1999). The results show that the prediction equations for smaller AEPs (1 in 20 to 1 in 200) require a greater number of predictor variables than those of higher AEPs. The R^2 values of the developed prediction equations range from 0.618 to 0.748, while SEE values range from 0.2 to 0.3, which indicates a good linear fit for the developed prediction equations. Generally, R^2 values reduce as AEP decreases. Median Relative Error (MRE) between the estimated and observed flood quantiles for the 12 test catchments are found to vary between 37% and 45%. The comparison of the estimated and observed flood quantiles for AEP of 1 in 2 for the Test Set 1 is provided in Figure 2.

Table 2. Typical prediction equations for QRT (Model Set 1).

| ARI	Prediction equations
2	$\log(Q_2) = -3.702 + 0.683 \log(area) + 1.335 \log(I_{\_12\_50}) + 0.827 \log(rain)$ $R^2 = 0.727$ $SEE = 0.246$
5	$\log(Q_5) = -1.095 + 0.661 \log(area) + 1.663 \log(I_{\_12\_50})$ $R^2 = 0.742$ $SEE = 0.23$
10	$\log(Q_{10}) = 3.043 + 0.69 \log(area) + 1.809 \log(I_{\_12\_50}) - 1.411 \log(evap) + 0.338 \log(sden) -0.13\log(forest)$ $R^2 = 0.748$ $SEE = 0.3$
20	$\log(Q_{20}) = 4.278 + 0.685 \log(area) + 1.807 \log(I_{\_12\_50}) - 1.765 \log(evap) + 0.338 \log(sden)-0.13\log(forest)$ $R^2 = 0.748$ $SEE = 0.242$
50	$\log(Q_{50}) = 7.792 + 0.619 \log(area) + 2.725 \log(I_{\_12\_2}) - 0.911 \log(rain) - 1.968 \log(evap)+0.360\log(sden)-0.159\log(forest)$ $R^2 = 0.654$ $SEE = 0.278$
100	$\log(Q_{100}) = 9.031 + 0.6 \log(area) + 2.8 \log(I_{\_12\_2}) - 1.048 \log(rain) -2.218 \log(evap)+0.396 \log(sden)-0.18 \log(forest)$ $R^2 = 0.654$ $SEE = 0.278$
200	$\log(Q_{200}) = 10.189 + 0.58 \log(area) + 2.882 \log(I_{\_12\_2}) - 1.161 \log(rain) -2.467\log(evap)+0.432 \log(sden)-0.202 \log(forest)$ $R^2 = 0.618$ $SEE = 0.297$
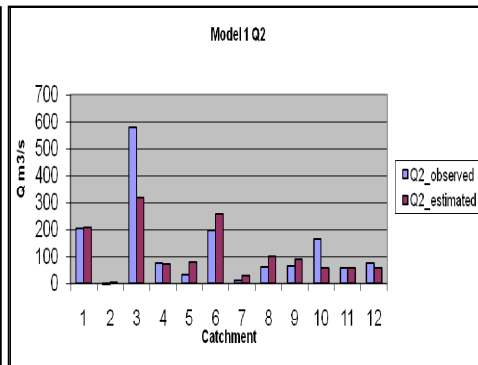
Figure 2. Comparison of flood quantiles by QRT and observed flood quantiles (AEP = 1 in 2) (Test Set 1).

Eight sets of regression equations for the parameters of the LP3 distribution (M, S and SK) for AEPs of 1 in 2 to 1 in 200 are developed (PRT). The regression equations for *M* are characterised by high R^2 values (0.70 to 0.76), followed by S (0.37 to 0.45) and SK (0.14 to 0.21). This indicates that prediction equations for S and SK are not strong, which is likely to introduce error in quantile estimation by the PRT. The prediction equations for 1 in 2 AEP for Model Set 1 are shown in Table 3, and the corresponding model coefficients are summarized in Table 4. Comparison of results of estimated flood quantiles by the PRT and the observed flood quantiles (by the LP3 distribution) for 1 in 2 and 1 in 50 AEPs for Model Set 1 are provided in Figures 3 and 4, which show a general tendency of overestimation of flood quantiles by the PRT, due to the uncertainty in the prediction equations for S and SK.

Table 3. PRT regression equations and goodness-of-fit statistics for Model Set 1.

Method	Prediction Equations
PRT	$M = -3.746 + 0.63 \log(area) + 1.245 \log(I_{\_12\_50}) +0.9 \log(rain)$ $R^2 = 0.72$ $SEE = 0.23$ $DWS = 1.91$ $S = 3.201 + 0.497 \log(I_{\_12\_50}) -1.044\log(rain) -0.075 \log(S1085)$ $R^2 = 0.39$ $SEE = 0.11$ $DWS = 2.25$ $SK = 3.064 -0.312 \log(area) -3.091\log(I_{\_1\_2}) +2.074 \log(I_{\_12\_2}) -0.207 \log(forest)$ $R^2 = 0.28$ $SEE = 0.44$ $DWS = 2.14$

Table 4. Regression coefficients for PRT and associated model statistics for Model Set 1.

| Model		Unstandardised Coefficients	Standardised Coefficients	Significance	R^2	*SEE*	DWS
Dependent Variables	Independent Variables						
M	(Constant)	-3.746		0.82			
	log (area)	0.63	0.68	0.05			
	log ($I_{\_12\_50}$)	1.245	0.43	0.23			
	log (rain)	0.9	0.25	0.31	0.72	0.23	1.91
S	(Constant)	3.201					
	log ($I_{\_12\_50}$)	0.497	0.526	0.3			
	log (rain)	-1.044	-0.876	0.07			
	log (sl-085)	-0.075	-0.168	0.27	0.39	0.11	2.25
SK	(Constant)	3.064					
	log (area)	-0.312	-0.31	0.32			
	log ($I_{\_1\_2}$)	-3.091	-0.63	1.4			
	log ($I_{\_12\_2}$)	2.074	0.56	0.1			
	log (*forest*)	-0.207	-0.21	1.54	0.28	0.44	2.14
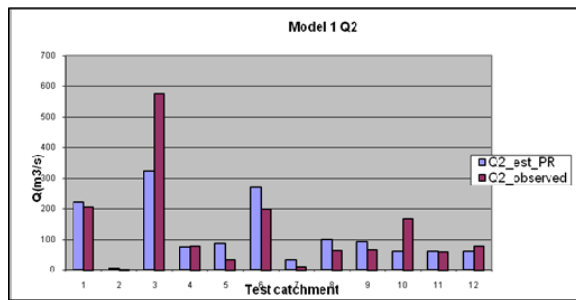
Figure 3. Comparison results for flood quantiles for 1 in 2 AEP by PRT.
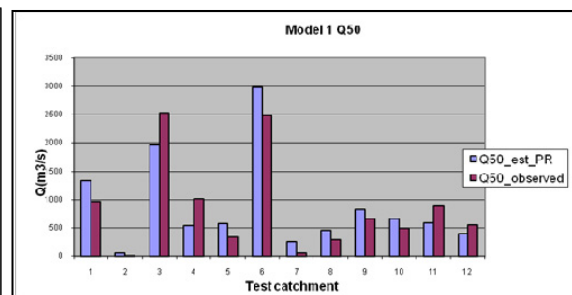

Figure 4. Comparison results for flood quantiles for 1 in 50 AEP.

The comparison of observed and predicted flood quantiles by the QRT and PRT techniques for 1 in 2 and 1 in 50 AEPs are presented in Figures 5 and 6, respectively. It shows that both QRT and PRT techniques in general overestimates the flood quantiles. This is not unexpected given that both techniques use the same catchment characteristics as predictor variables. In Table 5, the median relative error values of the flood quantiles for the QRT and PRT are presented considering all the 8 sets of test catchments. It shows that both the QRT and PRT have similar median relative error values across different AEPs.
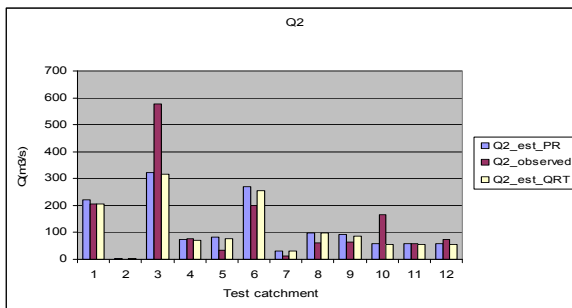

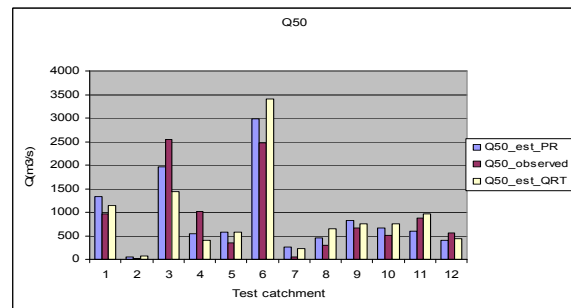**Figure 5.** Comparison of estimated (by QRT/PRT) and observed flood quantiles (AEP = 1 in 2).

Figure 6. Comparison of estimated (by QRT/PRT) and observed flood quantiles (AEP = 1 in 50).

Table 5. Comparison of median relative error values for QRT and PRT.

AEP (1 in *T*)	Median relative error (%)	
	PRT	QRT
2	39	37
5	40	38
10	37	39
20	36	37
50	42	41
100	42	44
200	45	45

5.   CONCLUSIONS

Since the publication of ARR1987, a significant number of studies on regional flood frequency estimation (RFFE) have been conducted in Australia. This study compares two commonly used RFFE methods, Quantile Regression Technique (QRT) and Parameter Regression Technique (PRT), using data from 96 gauged catchments in New South Wales. In the PRT, the first three moments of the LP3 distribution are regionalized. Based on a split-sample validation technique, it has been found that the QRT and PRT perform very similarly, with median relative error values in the range of 36% to 45% for flood quantiles in the range of 1 in 2 to 1 in 200 annual exceedance probabilities (AEPs). A more rigorous error analysis based on Monte Carlo cross validation or leave-one-out validation would provide better error estimation for the QRT and PRT. This study shows that both PRT and QRT techniques perform quite similarly in New South Wales. Further studies using data from different geographical regions of Australia need to be conducted before recommending an appropriate RFFE method for application in Australia.

REFERENCES

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.

Griffis, V.W. and Stedinger, J.R. (2007). The use of GLS regression in regional hydrologic analyses. Journal of Hydrology, 204, 82-95.

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., Rahman, A., and 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. and Rahman, A. (2012a). 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., and Stedinger, J.R. (2012b). 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 (1987) Australian rainfall and runoff: A guide to flood estimation, estimation of design peak discharges (ed D H Pilgrim); Vol. 1, I. E. Aust., Canberra.

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

Micevski, T., Hackelbusch, A., Haddad, K., Kuczera, G., and 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.

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.

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.

Pegram, G. (2002). Rainfall, rational formula and regional maximum flood – some scaling links. 27th National Hydrology and Water Resources Symp, 20-23 May, Melbourne.

Pilgrim, D.H. (1982). Assessment of derived rural and urban runoff coefficients. Civil Engineering Transactions, I. E. Aust., CE24, 235-241.

Rahman, A. (2005). A Quantile Regression Technique to estimate design floods for ungauged catchments in South-east Australia. Australian Journal of Water Resources, 9(1), 81-89.

Rahman, A., Bates, B.C., Mein, R.G. and Weinmann, P.E. (1999). Regional flood frequency analysis for ungauged basins in south-eastern Australia. Australian Journal of Water Resources, 3(2), 199-207.

Rahman, A., Haddad, K., Zaman, M., Kuczera, G. and Weinmann, P.E. (2011a). 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. (2011b). Regional flood methods for Australia: data preparation and exploratory analysis. Australian Rainfall and Runoff Revision Projects, Project 5 Regional Flood Methods, Stage 1 Report, Technical Report, No. P5/S1/003, Engineers Australia, Water Engineering, 181pp.

Rahman, A., Haddad, K., Zaman, M., Ishak, E., Kuczera, G., and Weinmann, P.E. (2012). Project 5 Regional flood methods, Stage 2 Report, Technical Report, No. P5/S2/015, Engineers Australia, Water Engineering, 319pp.

Rahman, S.A., Rahman, A., Zaman, M., Haddad, K., Ashan, A., and Imteaz, M. A. (2013). A study on selection of probability distributions for at-site flood frequency analysis in Australia. Natural Hazards, 69, 1803-1813.

Rahman, A., Haddad, K., Rahman, A.S., and 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., and 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.

Stedinger, J.R. and Tasker G.D. (1985). Regional hydrologic analysis, 1.Ordinary, weighted, and generalized least squares compared. Water Resources Research, 2209, 1421-1432.

Thomas, D.M., and Benson, M.A. (1970). Generalization of streamflow characteristics from drainage basin characteristics. *U.S. Geol. Water Supply Pap*. No.1975.

Vogel, R.M., McMahon, T.A., and Chiew, F.H.S. (1993). Floodflow frequency model selection in Australia. Journal of Hydrology, 146, 421-449.

Weeks, W.D. (1991). Design Floods for Small Rural Catchments in Queensland." Civil Engineering Transactions, I. E. Aust., CE33, 4, 249-260.