Features of Regional Flood Frequency Estimation (RFFE) Model in Australian Rainfall and Runoff

A. Rahman[a, b], K. Haddad[b] and G. Kuczera[c]

[a] *Institute for Infrastructure Engineering, Western Sydney University*
[b] *School of Computing, Engineering and Mathematics, Western Sydney University*
[c] *School of Engineering, The University of Newcastle*
Email: a.rahman@westernsydney.edu.au

Abstract: Flood estimates at ungauged catchments are generally associated with a high degree of uncertainty. For the upcoming Australian Rainfall and Runoff (ARR) 2015 (4th edition), a new regional frequency estimation (RFFE) model known as ARR RFFE Model 2015 has been developed. This RFFE Model 2015 is based on the concept of regionalization, which is a data driven approach where data from gauged catchments are utilized to make flood quantile estimates at ungauged locations. This paper presents the modelling approach that underpins the RFFE Model 2015.

In the RFFE Model 2015, Australia is divided into humid coastal areas and arid/semi-arid areas. In the humid coastal areas, a region of influence approach is adopted to form sub-region by drawing a number of nearby gauged catchments for a location of interest. This in essence attempts to reduce the degree of heterogeneity in forming local region at the location of interest. For estimating flood quantiles, a regional log Pearson Type 3 (LP3) distribution is adopted where the location, scale and shape parameters are estimated based on prediction equations. A Bayesian Generalized Least Squares (GLS) regression approach is adopted to develop prediction equations in the humid coastal areas. The main advantages of GLS regression is that this accounts for inter-station correlation and variation in streamflow data lengths across different sites in a region. Furthermore, GLS regression differentiates between the sampling and model errors and hence provides a more rigorous approach of dealing with uncertainty in regional flood modelling compared with the ordinary least squares regression.

For the arid/semi-arid areas, an index flood method is adopted where 10% annual exceedance probability (AEP) flood is taken as the index variable. The prediction equation for the index variable is developed using an ordinary least squares regression approach in the arid/semi-arid areas. The regional growth factors are estimated from the at-site flood frequency analysis where LP3 distribution is fitted to the annual maximum flood series data. The Multiple Grubb-Beck test is used to censor the zero and low annual maximum flood series data points, which is typical in the arid regions. A total of 798 gauged catchments are used from the humid coastal areas and 55 catchments from the arid/semi-arid areas to develop and test the ARR RFFE model. The data from these catchments are prepared adopting a stringent quality control procedure that involved infilling the gaps in the data, checking for outliers, trends and rating curve extrapolation error.

In developing the confidence limits for the estimated flood quantiles, a Monte Carlo simulation approach is adopted by assuming that the uncertainty in the first three parameters of the LP3 distribution (i.e. the mean, standard deviation and skewness of the logarithms of the annual maximum flood series) can be specified by a multivariate normal distribution. In the ARR RFFE model, the model coefficients have been embedded in an application software (known as RFFE Model 2015), which enables the user obtaining design flood estimates relatively easily using simple input data such as latitude, longitude and catchment area of the ungauged catchment of interest.

The RFFE Model 2015 is applicable to any catchment that has similar attributes and flood producing characteristics as the catchments used in the derivation of the flood estimation equations embedded in the RFFE model. Catchments which do not satisfy this requirement can be divided into three groups: (i) catchments which have been substantially modified from their natural characteristics and for which the RFFE model is not applicable and should thus not be used (ii) catchments for which flood estimates must be expected to have lower accuracy such as arid region catchments; and (iii) 'atypical catchments' where additional catchment attributes need to be considered and adjusted for such as catchments with large natural flood plain area.

Keywords: *RFFE, GLS regression, ARR, ungauged catchments, index flood method*

1.   INTRODUCTION

Regional flood frequency estimation (RFFE) is a data driven approach, which attempts to transfer flood characteristics information from a group of gauged catchments to the catchment location of interest. A RFFE technique is expected to be simple so that design flood estimates can be obtained from readily available input data. In RFFE, the region is considered to be homogeneous i.e. the gauged catchments are hydrologically similar.

A RFFE method essentially consists of two principal steps: (i) formation of regions based on the available streamflow gauging stations; and (ii) development of regional prediction equations to be used for derivation for flood quantile estimation. Formation of regions can be based on proximity in geographic or catchment attributes space; however, regions in geographic space are more common in RFFE. A region can be fixed, having a definite boundary (such as based on state boundary) or it can be formed in geographic or catchment characteristics data space.

In RFFE, the regions are based on 'homogeneity' assumption, which has different meaning to different RFFE approaches. For example, in the index flood method, homogeneity means that all the stations in a region have the same standardized frequency curves within the margin of 'acceptable' sampling variability (Hosking and Wallis, 1993). The regression based approach may not need to satisfy the criteria of homogeneity; however, a heterogeneous region may fail to generate powerful predictions equations for the region. Burn (1990) proposed the region-of-influence (ROI) approach where a site of interest (i.e. catchment where flood quantiles are to be estimated) can form its own region by selecting a group of 'nearby catchments' in the geographical or catchment characteristics space. In developing the regional prediction equations, index flood method, quantile regression technique (QRT) and artificial intelligence based methods can be adopted. In the regression approach, ordinary least squares (OLS) and generalized least squares regression approaches can be adopted (e.g. Stedinger and Tasker, 1985; Rahman, 2005; Griffis and Stedinger, 2007 Haddad and Rahman, 2012; Haddad et al., 2012).

Australian Rainfall and Runoff (ARR) 1987 recommended a number of different RFFE methods for Australia. For example, the Probabilistic Rational Method (PRM) was recommended for Victoria and eastern NSW (I. E. Aust., 1987). As a part of the recent upgrade of ARR, a number of RFFE methods have been examined such as QRT based on OLS and GLS regression techniques, PRM and Parameter Regression Technique (PRT) (e.g. Stedinger and Tasker, 1985; Griffis and Stedinger, 2007; Rahman et al., 2009, 2012, 2015a; Rahman et al., 2011; Haddad et al., 2012; Haddad et al., 2012; Palmen and Weeks, 2011; Zaman et al., 2012; Micevski et al., 2015).

The objective of this paper is to present the modelling framework adopted in developing the ARR RFFE Model 2015 so that the users have an understanding of the strengths and limitations of the RFFE model when they apply this model in practice.

2.   FORMATION OF REGIONS IN ARR RFFE MODEL 2015

In ARR RFFE Model 2015, Australia has been divided into humid coastal and arid/semi-arid areas (Figure 1). A total of 798 catchments are adopted from the humid coastal areas and 55 catchments are selected from the arid/semi-arid areas. Further information on the preparation of these data can be found in Haddad et al. (2010). A total of five humid coastal regions, two arid/semi-arid regions and seven fringe zones are formed (Table 2, Figure 2 and Table 3). The boundary of the fringe zone located near humid coastal region was approximately defined by the 500 mm mean annual rainfall isohyet, while the other side was defined by 400 mm mean annual rainfall isohyet to establish a fringe zone. In drawing the boundary, some minor adjustment was made to make the boundary as smooth as possible.

For the humid coastal regions, an ROI approach is adopted to form sub-regions at each of the 798 gauged locations. The arid/semi-arid areas have very few gauging stations (only 55) and hence in these areas, application of ROI is deemed infeasible and hence a fixed region approach is adopted.

In forming the ROI sub-regions, in the first iteration, a ROI sub-region consisting of the ten nearest stations to the site of interest is formed, the regional prediction equation is then developed and its prediction error variance is noted. At each of the subsequent iterations, the radius of the ROI sub-region is increased by 10 km and new stations are added to the previously selected stations. This procedure ends when all the eligible stations are included in the ROI sub-region. The final ROI sub-region for the site of interest is then selected as the one that gives the lowest prediction error variance.

There are seven fringe zones (Figure 2) which is intended to provide a smooth variation in regional flood estimates across regional boundaries. When an ungauged catchment is located in a fringe zone, the flood estimate is made considering the inverse distance weighted average value of the two flood estimates based on the two nearest regions. This interpolation method is embedded into the ARR RFFE Model 2015.
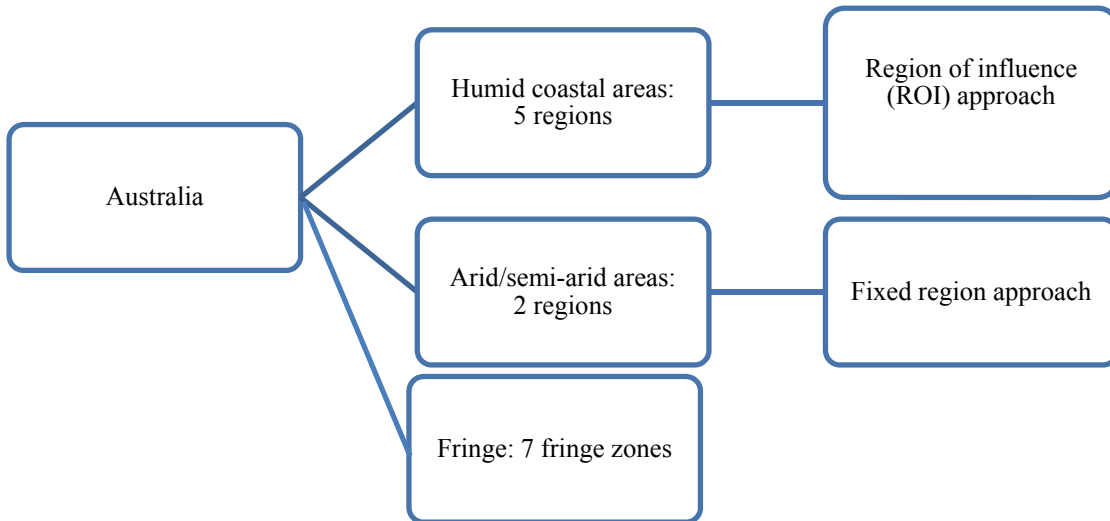

Figure 1. Formations of regions in ARR RFFE Model 2015.

Table 2. Regions in humid coastal areas of Australia.

Region	Method to form region	Number of stations
Region 1: East Coast	ROI (formed based on geographical proximity)	558
Region 2: Tasmania		51
Region 3: Humid SA		28
Region 4: Top End NT and Kimberly		58
Region 5: SW WA		103

3.   ESTIMATION EQUATIONS

In the humid coastal regions, a parameter regression technique (PRT) is adopted where the first three moments of the Log Pearson Type 3 (LP3) distribution (i.e. the mean, standard deviation and skewness of the natural logarithms of the annual maximum floods) are regionalized. The PRT though not as common as the QRT, PRT offers three significant advantages over the QRT: (i) it ensures that flood quantiles increase smoothly with decreasing AEP, an outcome that may not always be achieved with the QRT; (ii) it is straightforward to combine any at-site flood information with regional estimates to produce more accurate quantile estimates (Kuczera, 1999); and (iii) it permits quantiles to be estimated for any AEP in the range of interest i.e. it offers an easy method of interpolation between AEPs.

The flood quantile estimates from the LP3 distribution are described by the following equation:

$$\ln Q_x = M + K_x S \tag{1}$$

where Q_x = the discharge having an AEP of $x\%$ (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_x = frequency factor for the LP3 distribution for AEP of x%, which is a function of the AEP and the skewness (*SK*) of the natural logarithms of the annual maximum flood series
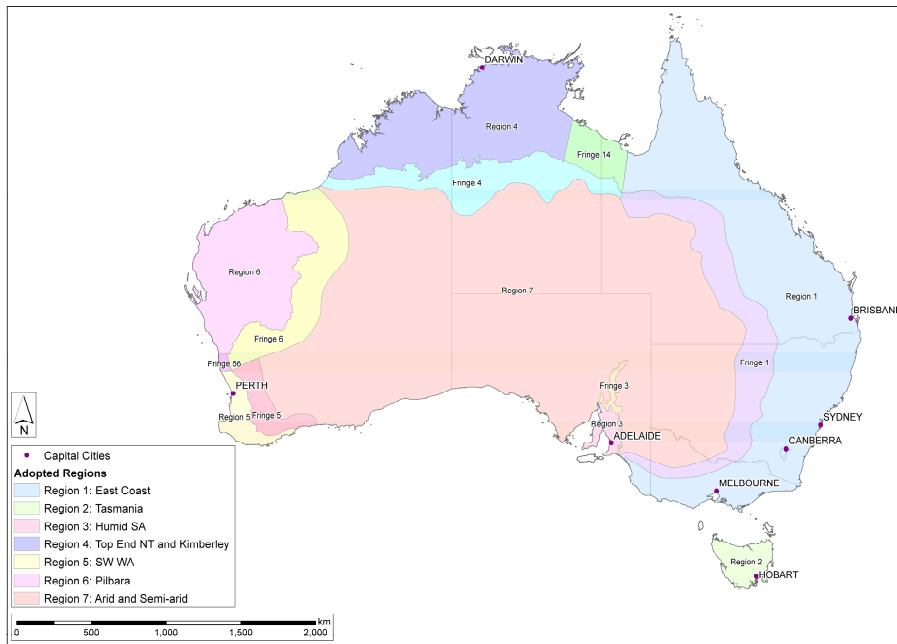

Figure 2. Adopted regions in the ARR RFFE Model 2015.

Table 3. Regions in arid/semi-arid areas of Australia.

Location	No. of stations	Streamflow record length (years) (range and median)	Catchment size (km^2) (range and median)
Pilbara	11	22 – 34 (28)	0.2 – 5975 (303)
Arid and Semi-arid	44	10 – 46 (27)	3 – 997 (209)
TOTAL	55	10 – 46 (27)	0.1 – 5975 (259)

A Bayesian Generalized Least Squares (GLS) regression is adopted in the humid coastal regions to develop prediction equations for M, S and SK. Further information on this Bayesian GLS regression can be found in Haddad et al. (2012) and Haddad and Rahman (2012).

For the five humid coastal regions provided in Table 2, the adopted estimation equations for *M*, *S* and *SK* for the regional LP3 model (equation 1) have the following general form:

$$M = b_0 + b_1(\ln(area)) + b_2(\ln(I_{6,50})) + b_3 (\ln(shape\ factor)) \tag{2}$$

$$S = c_0 + c_1(\ln(I_{6,2}/ I_{6,50})) \tag{3}$$

$$SK = d_0 + d_1(\ln(area)) + d_2(\ln(I_{6,2}/ I_{6,50})) + d_3(\ln(I_{6,2})) \tag{4}$$

where, *area* = catchment area (km^2);

$I_{6,50}$ = design rainfall intensity (mm/h) at catchment centroid for 6-hour duration and AEP of 50%;

shape factor = shortest distance between catchment outlet and centroid/*area*$^{0.5}$; and

$I_{6,2}$ = design rainfall intensity (mm/h) at catchment centroid for 6-hour duration and AEP of 2%.

For Region 1 and Region 2, equations 9 and 10 have only the model intercepts, i.e. for these two regions, the weighted average values of *S* and *SK* are adopted, which are determined on the basis of record lengths at the stations within the ROI sub-region. The values of regression coefficients b_0, b_1, b_2, b_3, c_0, c_1, d_0, d_1, d_2 and d_3 at all the 798 individual gauged catchment locations (in the humid coastal regions) are estimated and embedded in the REEF Model 2015 (Figure 4). To derive flood quantile estimate at an ungauged catchment of interest, RFFE Model 2015 uses a natural neighbor interpolation approach (using regional flood quantiles) based on up to 15 nearest gauged catchment locations within 300 km radius from the catchment of interest. This ensures a smooth variation of flood quantile estimates over the space.
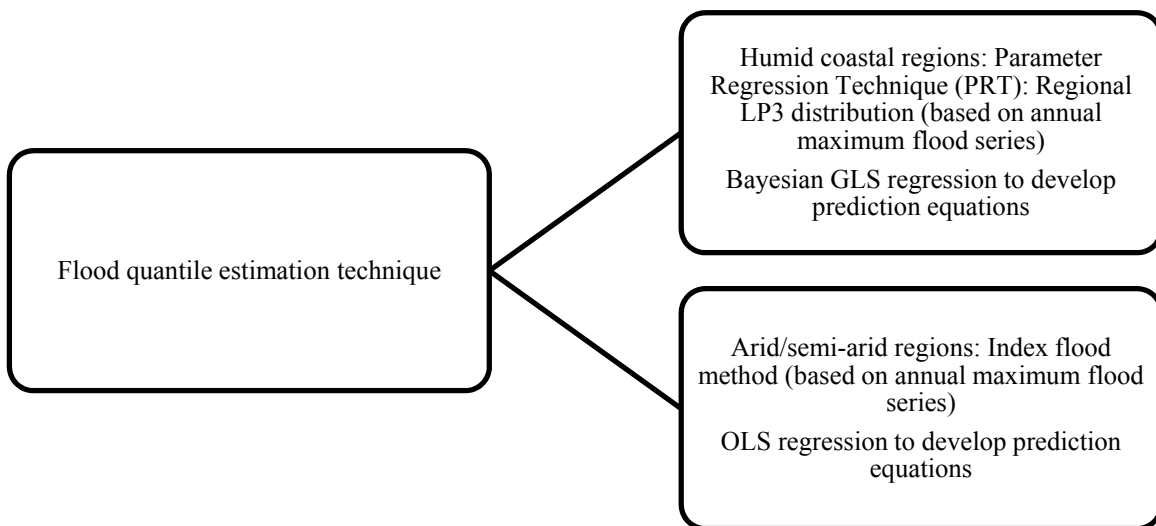

Figure 3. Flood quantile estimation methods in RFFE Model 2015.

For arid/semi-arid regions, an index flood approach is adopted where 10% AEP flood quantile (Q_{10}) is used as the index variable and a dimensionless growth factor for AEP of *x*% (GF_x) is used to estimate Q_x:

$$Q_x = Q_{10} \times GF_x \tag{5}$$

A prediction equation is developed for Q_{10} as a function of catchment characteristics, and regional growth factors are developed based on the observed annual maximum flood series data. It should be noted that in arid regions both the partial duration series and annual maximum flood series are compared and it is found that annual maximum flood series (with Multiple Grubb-Beck test proposed by Lamontagne et al. (2013) to censor zero and low flows) with the LP3 distribution provides more accurate flood quantile prediction.

To derive the design growth curves, the Q_x/Q_{10} values are first estimated at individual stations from the fitted LP3 distribution; the weighted average of these values (weighting is done based on record length at individual sites) over all the stations in a region finally gives the growth factors (GF_x) for the region.

The adopted prediction equation for the index variable Q_{10} has the following form:

$$\log_{10}(Q_{10}) = b_0 + b_1(\log_{10}(area)) + b_2(\log_{10}(I_{6,50})) \tag{12}$$

where b_0, b_1 and b_2 are regression coefficients, estimated using OLS regression; *area* represents catchment area in km^2, and $I_{6,50}$ is the design rainfall intensity (mm/h) at catchment centroid for 6-hour duration and 50% AEP. The values of b_0, b_1 and b_2 and the regional growth factors (GF_x) are embedded into the REEF Model 2015.

A leave-one-out validation method is adopted to assess the relative accuracy of the RFFE model (Haddad et al., 2013). In this validation, each of the gauged catchments is removed from the data set in iteration *i* and the developed RFFE Model is then applied on this catchment and relative error is noted. The procedure is then repeated *n* times to estimate relative error and other error statistics (where *n* = number of gauged catchments in the modelling data set). RFFE Model 2015 testing results can be seen in Rahman et al. (2015b).

4.   BUILDING CONFIDENCE LIMITS ON THE ESTIMATED FLOOD QUANTILES

To build the confidence limits on the flood quantiles, a Monte Carlo simulation approach is adopted by assuming that the uncertainty in the first three parameters of the LP3 distribution (i.e. the mean, standard deviation and skewness of the logarithms of the annual maximum flood series) can be specified by a multivariate normal distribution (MND).

To apply the MND, the correlations of the mean, standard deviation and skewness are needed, which are estimated from the residuals of the GLS regression models of the LP3 parameters. The mean of the LP3 parameter is specified by its regional predicted value and the standard deviation of the LP3 parameter is taken as the square root of the average variance of prediction of the parameter at the nearest gauged site. Based on 10,000 simulated values of the LP3 parameters from the MND as defined above, 10,000 Q_x values are estimated, which are then used to develop the 90% confidence intervals of the estimated flood quantiles.
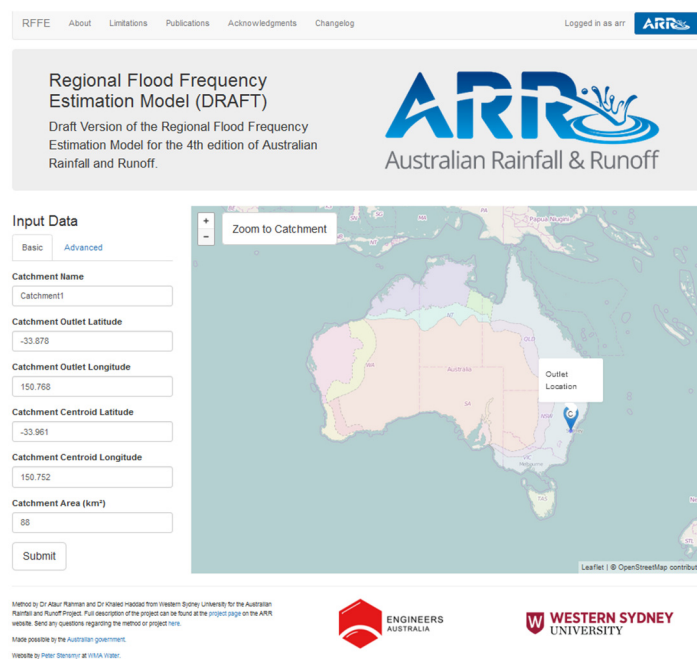

Figure 4. RFFE Model 2015 (input data window).

5.   CONCLUSION

A new RFFE Model is developed for Australian Rainfall and Runoff (ARR) 2015 (4th edition) for application in Australia. In the RFFE Model, Australia has been divided into seven regions, five are humid coastal regions and two are arid/semi-arid regions. There are seven fringe zones which are adopted to avoid a sharp variation in flood quantile estimates for an ungauged catchment located near regional boundary. For the humid coastal regions, a region of influence approach is adopted to form a sub-region around the location of interest. For flood quantile estimation, a regional Log Pearson Type 3 (LP3) distribution is developed which enables estimation of design floods at ungauged catchments. In the arid regions, an index flood method is applied. To develop the confidence interval, a Monte Carlo simulation technique is adopted. A tool called ARR RFFE Model 2015 is developed integrating the RFFE modelling techniques, which can be applied in practice to obtain design flood estimates at any location in Australia using simple data input like catchment area and latitudes and longitudes of the ungauged catchment location.

(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). The authors would also like to acknowledge Md Mahmudul Haque, Mohammad Zaman, Erwin Weinmann, James Ball, Mark Babister and William Weeks and many other people (can be seen in Rahman et al., 2015b) who have contributed to ARR Project 5.

REFERENCES

Burn, D.H. (1990). Evaluation of regional flood frequency analysis with a region of influence approach, *Water Resources Research*, 26, 10, 2257-2265.

Griffis, V.W. and Stedinger, J.R. (2007). The use of GLS regression in regional hydrologic analyses, *Journal of Hydrology*, 344, 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. (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., and 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., Zaman, M., and Shrestha, S. (2013). Applicability of Monte Carlo cross validation technique for model development and validation using generalised least squares regression, *Journal of Hydrology*, 482, 119-128.

Hosking, J.R. and Wallis, J.R. (1993). Some statistics useful in regional frequency analysis, *Water Resources Research*, 29(2), 271-281.

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

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, 2454-2466.

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.

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., and 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., and Weinmann, P.E. (2012). Australian Rainfall and Runoff Revision Projects, Project 5 Regional flood methods, Stage 2 Report.

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

Rahman, A., Haddad, K., Kuczera, G., Weinmann, P.E. (2015b). Regional flood methods, in Australian Rainfall & Runoff, Chapter 3, Book 3, edited by J. E. Ball, Engineers Australia, 78-114 (In press).

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.

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.

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