Sampling Variability in Flood Frequency Analysis: How Important Is It?

Ayesha S Rahman [a,b], **Fazlul Karim** [b] and **Ataur Rahman** [a,c]

[a]*School of Computing, Engineering and Mathematics, Western Sydney University, Australia*
[b]*CSIRO Land and Water Flagship, Canberra, ACT 2601, Australia*
[c]*Institute for Infrastructure Engineering, Western Sydney University, Australia*
Email: ayesha.rahman@westernsydney.edu.au

Abstract: Flood quantile estimation using available streamflow records, known as at-site flood frequency analysis (FFA), are widely used in hydrology. The estimated flood quantiles by at-site FFA are used in the planning and design of many water resources management tasks. However, FFA estimates often suffer from high sampling variability, in particular when length of streamflow record is relatively short. This aspect of FFA has not been fully examined for Australian catchments. As the hydrology in Australia suffers from a very high degree of variability, it is likely that the sampling variability in FFA is also very high.

This paper presents results from a case study based on three different gauged stations located in New South Wales, Queensland and Victoria using the FLIKE (an extreme value analysis package) software. These stations represent different hydrological regimes (e.g. Victoria is dominated by winter rainfall and Queensland is by summer rainfall). Two widely used probability distribution functions, Generalized Extreme Value (GEV) and Log Pearson type 3 (LP3) distributions are adopted in this case study. We have used updated flood data which have been prepared for Australian Rainfall and Runoff Project 5. The selected streamflow data length ranges from 58 to 102 years. The annual maximum flood data at each of these stations have been sub-divided into three sub-sets: full data set, 50% split and 25% split, which enables to carry out this test with sample sizes in the range of 14 years to 51 years.

The study shows that for all the three stations, at-site flood quantile estimates are more affected by the sampling variability in the case of the LP3 (Bayesian) distribution than the GEV (L moments) distribution. Based on the results of this empirical study, it has been found that for 50, 40, 30, 20 and 15 years of annual maximum flood data lengths, the sampling variability estimates are in between -41% to 326% (for LP3 distribution) and -42% to 39% (for GEV distribution) relative to the full data length.

The findings of this study have crucial implications in the field of FFA as at-site FFA estimates are generally taken 'accurate' in decision making. Furthermore, in assessing the performances of the regional flood frequency estimation models and calibration of runoff routing model, at-site flood frequency analysis estimates based on about 25 years of data are considered 'robust' and 'accurate', which seems to be not the case. This exercise is being conducted to a greater number of stations by applying boot-strap and Monte Carlo simulation techniques, which will enable to generalize the findings of this study.

Keywords: *Flood frequency analysis, LP3 distribution, GEV distribution, sampling variability, FLIKE*

1. INTRODUCTION

Estimation of design flood is often needed in hydrologic design. Flood frequency analysis is the most commonly adopted method to obtain design floods; however, it is filled with controversies (Bobée et al. 1993). Numerous studies have been carried out in the past to compare various probability distributions for at-site flood frequency analysis. However, due to the limited length of observed flood data as compared to the return period of interest, flood frequency analysis is subject to high degree of uncertainty. The selection of an 'appropriate' probability distribution and associated parameter estimation procedure is an important step in flood frequency analysis and has been widely researched (Srikanthan and McMahon, 1981; Vogel et al., 1993; Onoz and Bayazit, 1995; Bates et al., 1998; Merz et al., 2008; Meshgi and Khalili, 2009a, b; Merz and Thieken, 2009; Ishak et al., 2010; Haddad et al., 2011, Haddad et al., 2012; Haddad and Rahman, 2012; Zaman et al., 2012; Haddad et al., 2013; Rahman et al., 2013). In flood frequency analysis, a probability distribution is often selected on the basis of statistical tests or by graphical methods, and convenience plays an important role in this choice (Bobee et al., 1993). In practical applications, empirical suitability plays a much larger role in distribution choice than a priori reasoning (Cunnane, 1985, 1989).

Based on the findings from the studies by Conway (1970), Kopittke et al. (1976) and McMahon and Srikanthan (1981), Australian Rainfall and Runoff (ARR) 1987 recommended the LP3 distribution with the product moment method for parameter estimation for at-site flood frequency analysis in Australia (I.E. Aust. 1987) following the footsteps of the USA (IAWCD 1982).

Since the publication of ARR 1987 (I. E. Aust., 1987), there have been a number of studies to compare various probability distributions (Rahman et al., 1999; Kuczera, 1999). For example, Nathan and Weinmann (1991) examined 53 catchments from Central Victoria (VIC), with L-moments-based goodness-of-fit test, and found that the generalized extreme value (GEV) distribution was the best-fit distribution. Vogel et al. (1993) compared a number of distributions using data from 61 stations in Australia. By using the L-moments ratio diagram, they concluded that the generalized Pareto distribution (GPA) was the best-fit distribution followed by the GEV, three-parameter lognormal, and LP3. Kuczera (1999) presented a comprehensive study on flood frequency analysis using Bayesian method and incorporated a number of probability distributions in his FLIKE software. The advantage of FLIKE is that, for a given application, the user can compare a number of most commonly adopted probability distributions and parameter estimation methods relatively quickly using a windows interface. Haddad and Rahman (2008) compared a number of distributions and parameter estimation procedures for 18 catchments in southeast Australia and found that the GEV distribution was the best-fit distribution for the selected catchments. In another study, Haddad and Rahman (2010) found that the two parameter distributions are preferable to Tasmania, with the lognormal appearing to be the best-fit distribution for Tasmania.

Since the inception of the above-mentioned studies, there has been a significant increase in recorded streamflow data at many stations in Australia. In particular, ongoing ARR revision Project 5 Regional Flood Methods has compiled a national database of over 600 Australian stations, which represents the most comprehensive annual maximum flood database in Australia to date (Rahman et al., 2014). This database can be used to compare various probability distributions for different regions of Australia. The results from such a comparison could provide useful guidance on the suitability of candidate probability distributions for a given region/state in Australia and in other similar regions around the world.

Now the questions are: (i) how accurate the estimates obtained by flood frequency analysis technique are and (ii) how sampling variability affect flood frequency analysis estimates. This study investigates the effects of sampling variability on estimated flood quantiles using the most commonly adopted flood frequency analysis methods in Australia.

2. DATA SELECTION

For this study, data from stream gauging stations, 215004, 138002 and 401216 in NSW, QLD and VIC respectively have been selected (Table 1). These stations have reasonably longer annual maximum flood (AMF) record lengths (58 to 102 years). The AMF data series used in this study have been prepared as part of the on-going revision of the regional flood estimation methods in ARR (Project 5 Regional flood methods). These stations have not been affected by major land use change or any major regulation.

Table 1. Stations used in this study.

Station ID	Station name	State	Catchment area (km²)	AMF data Period and data length (years)	Mean annual rainfall (mm)
138002	Brooyar	QLD	655	1910-2011 (102)	1756
215004	Hockeys	NSW	166	1930-2011 (82)	1216
401216	Jokers CK	VIC	356	1952-2011 (60)	584

3.   METHODOLOGY

The AMF data series for each of the stations is divided into three sub-sets: full data set, 50% split and 25% split. In case of full data set, the maximum period of record is 102 years for Station 138002 and when the full data set is divided into 4 parts (i.e. 25% split), the smallest period of record becomes 14 years for Station 401216. Two commonly adopted probability distributions, Log Pearson Type 3 (LP3) and Generalized Extreme Value (GEV) distribution are used to estimate flood quantiles from the AMF data. The reasons for selecting LP3 and GEV distributions are that these are the most commonly used distributions in at-site flood frequency analysis in Australia (I. E. Aust., 1987; Haddad and Rahman, 2011).

The Multiple Grubbs-Beck (MGB) test is adopted in this study to identify outliers. The MGB test is a statistical method designed to detect multiple low outliers in a data series and has been recommended in Bulletin 17C for use in the USA (Cohn et al., 2013; Lamontagne et al., 2013). The MGB test uses an approximation of the probability of the kth smallest sample in a normal sample sized n, is smaller than the value observed. The MGB test has been incorporated into the FLIKE software for application in Australia (Kuczera, 1999) and this has been adopted in ARR Project 5 to detect low outliers in the AM flood series (Rahman et al., 2014).

The GEV distribution is a family of continuous probability distributions developed within extreme value theory to combine the Gumble, Frèhet and Weibull distributions. This distribution has been adopted in this study. The method of L-moments has been adopted to estimate its parameters.

4.   RESULTS

Table 2 reports the differences between the mean, standard deviation (SD) and skew values of the Q and X ($X = \log_e(Q)$) AMF series based on the full data set ($n = 102$ years) and each of the sub-periods ($n = 51, 51, 26, 25, 26$ and 25 years). Table 2 presents interesting results where mean of the AMF series for Station 138002 (for 50% split cases) exhibits a difference of 9.4% to 25.7% with respect to the full data set. For the 25% split cases, the differences in the mean range from 2.7% to 44.6% with respect to the full data set. For Station 215004, the differences in the mean range from 2.3% to 61.9% for 50% split case, and 0.2% to 48.7% for 25% split cases. For Station 401216, the differences in the mean are found to be much smaller, i.e. 1.25% to 3.06% for the 50% split cases, and 0.78% to 6.53% for the 25% split cases. Since flood quantile estimates are highly influenced by the estimated sample mean, it is highly likely that these differences will influence the flood quantile estimates greatly for Stations 138002 and 215004.

Table 3 presents the percentage differences in flood quantile estimates by the LP3 Bayesian distribution for Station 138002 when the AMF data is divided into sub periods. For the full data set, out of the 102 data points, 29 are identified as potentially influential low flows (PILF) by the MGB test. For the 50% split, out of the 51 data points, 8 are identified as PILFs in the second half of the 25% split. In the case of 25% split, 1 and 5 PILFs are identified in the 2nd and 4th splits. These PILFs are censored in flood frequency analysis by the LP3 distribution using the inbuilt MGB test facility in the FLIKE.

Table 3 presents the percentage differences in flood quantile estimates by the LP3 distribution for Station 138002 when the AMF data is divided into sub periods. These differences are between the flood quantiles based on the full data set ($n = 102$ years) and each of the sub-periods ($n = 51, 51, 26, 25, 26$ and 25 years). It can be seen that for the 50% split cases, the differences in flood quantiles are in the range of -23.1% to 25.1%. For the 25% of split cases, the differences are in the range of -41% to 326%. Figure 1 exhibits the differences in flood quantile estimates for Station 138002, which shows that in one case the expected flood quantiles fall out of the 90% confidence interval (CI). These are unusually high differences and highlight the severity of

sampling variability in flood quantile estimates in Australian catchments e.g. even with 25 years of AMF series data, the sample estimates of quantiles can suffer by a sampling error up to about 300%. In the case of GEV L moments method, the differences in quantile estimates range from -9.2% to 12.3% for 50% split cases, and -42.4% to 39.4% for 25% split cases. These results show that for Station 138002, the LP3 Bayesian method has been more affected by the sampling variability compared with the GEV L-moments method.

Table 2. Percentage difference in mean, SD and skew values in Station 138002 (full data set with $n = 102$ years vs. split data sets ($n = 51, 51, 26, 25, 26$ and 25 years).

AEP (50% - 1% AEPs)	50% split ($n = 51$ years)	50% split ($n = 51$ years)	25% split ($n = 26$ years)	25% split ($n = 25$ years)	25% split ($n = 26$ years)	25% split ($n = 25$ years)
Mean(Q)	-25.7	-15.1	-41.4	-7.6	-44.6	10.7
Mean (X)	-10.8	-8.6	-16.4	-2.7	-16.2	-8.3
SD (Q)	-16.9	9.4	-33.2	-2.2	-43.7	46.8
SD (X)	52.2	41.0	61.2	18.8	68.5	66.3
Skew (Q)	-13.1	17.9	-23.6	-20.6	-12.1	-14.3
Skew (X)	-275.1	-99.9	-244.9	-143.0	-406.9	-27.2

Table 3. Percentage differences in flood quantile estimates in the case of LP3 distribution: Station 138002 (full data set with $n = 102$ years vs. split data sets ($n = 51, 51, 26, 25, 26$ and 25 years) (PILFs are disregarded in these computation).

AEP	50% split (1st half) ($n = 51$ years)	50% split (2nd half) ($n = 51$ years)	25% split ($n = 26$ years)	25% split ($n = 25$ years)	25% split ($n = 25$ years)	25% split ($n = 25$ years)
50%	16.9	-23.1	-14.9	52.0	-6.2	-41.0
20%	7.6	-1.3	-12.0	32.8	-15.5	11.7
10%	1.1	8.5	-12.3	26.4	-22.2	61.8
5%	-4.9	15.5	-13.4	22.6	-28.2	123.5
2%	-11.9	21.9	-15.4	19.7	-35.2	227.3
1%	-16.6	25.1	-17.2	18.5	-39.8	326.3

Table 4. Percentage differences in flood quantile estimates in the case of GEV distribution: Station 138002 (full data set with $n = 102$ years vs. split data sets ($n = 51, 51, 26, 25, 26$ and 25 years).

AEP	50% split (1st half) ($n = 51$ years)	50% split (2nd half) ($n = 51$ years)	25% split ($n = 26$ years)	25% split ($n = 25$ years)	25% split ($n = 25$ years)	25% split ($n = 25$ years)
50%	12.3	-15.2	-15.0	39.4	-6.2	L- moments should
20%	4.5	-7.2	-17.1	24.9	-18.0	be positive
10%	0.8	-2.8	-18.8	19.7	-24.5	
5%	-2.4	1.4	-20.4	15.7	-30.3	
2%	-6.3	7.3	-22.7	11.1	-37.4	
1%	-9.2	12.0	-24.4	7.9	-42.4	

For Station 215004, differences in flood quantile estimates by the LP3 distribution based on the full data set ($n = 81$ years) and each of the sub-periods ($n = 41, 40, 21, 20, 20$ and 20 years) are found to be -23.8% to 48.1% for the 50% split cases, and -38.1% to 152.3% for the 25% split cases. For Station 401216, differences in flood quantile estimates by the LP3 distribution based on the full data set ($n = 58$ years) and each of the sub-periods ($n = 29, 29, 15, 14, 15$ and 14 years) are found to be -5.9% to 13.6% for the 50% split cases, and -18.2% to 84.1% for the 25% split cases.

For Station 215004, differences in flood quantile estimates by the GEV distribution based on the full data set (*n* = 81 years) and each of the sub-periods (*n* = 41, 40, 21, 20, 20 and 20 years) are found to be -23.3% to 22.3% for the 50% split cases, and -34.8% to 61.3% for the 25% split cases. For Station 401216, differences in flood quantile estimates by the GEV distribution based on the full data set (*n* = 58 years) and each of the sub-periods (*n* = 29, 29, 15, 14, 15 and 14 years) are found to be -5.3% to 5.1% for the 50% split cases, and -28.2% to 40.3% for the 25% split cases. Hence it can be seen that for all the three stations, sampling variability in flood quantile estimates are much higher for the LP3 Bayesian distribution than the GEV L- moments distribution. However, this analysis needs to be repeated for a good number of stations (e.g. 50 stations) from different parts of Australia to make any firm recommendation on the desirable minimum record length in at-site flood frequency analysis.
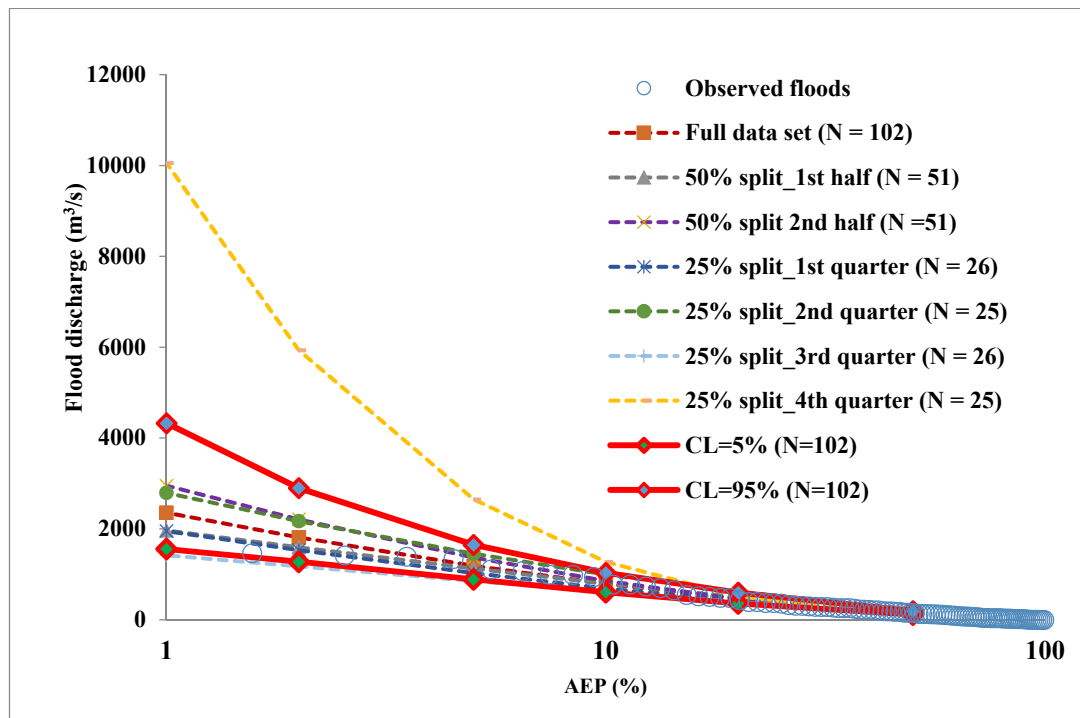
Figure 1. Differences in flood quantile estimates for Station 138002 using LP3 Bayesian distribution (full data, 50% split and 25% split cases).

5. CONCLUSIONS

This study examines how sampling variability is likely to affect at-site flood quantile estimates in Australian catchments. For this purpose, three different gauging stations with at least 60 years of data have been selected from New South Wales, Queensland and Victoria states of Australia. Two widely used probability distributions have been utilised from ARR FLIKE software (Kuczera, 1999) i.e. Log Pearson type 3 (LP3) distribution with Bayesian procedure and Generalized Extreme Value (GEV) distribution with L-moments. The selected streamflow data length ranges from 58 to 102 years. The annual maximum flood data series at each of these stations have been sub-divided into three sub-sets: full data set, 50% split and 25% split, which enables to carry out this test with sample sizes in the range of 14 years to 51 years. The study shows that for all the three stations, at-site flood quantile estimates are more affected by the sampling variability in the case of the LP3 (Bayesian) distribution than the GEV (L moments) distribution. Based on the results of this empirical study, it has been found that the sampling variability could be -41% to 326% for LP3 distribution and -42% to 39% for GEV distribution relative to the full data length. The findings of this study have crucial implications as at-site flood frequency analysis estimates are generally taken 'accurate' in decision making. Furthermore, in assessing the performances of the regional flood frequency estimation models and calibration of runoff routing model, at-site flood frequency analysis estimates based on about 25 years of data are considered 'robust' and 'accurate', which seems to be not true. This exercise is being conducted to a greater number of stations by applying boot-strap and Monte Carlo simulation techniques, which will enable to generalise the findings of this study.

ACKNOWLEDGMENTS

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.

Bobée, B., Cavidas, G., Ashkar, F., Bernier, J. and Rasmussen, P. (1993). Towards a systematic approach to comparing distributions used in flood frequency analysis, *Journal of Hydrology*, 142, 121–136.

Conway, K.M. (1970). Flood frequency analysis of some NSW coastal rivers. *Thesis (M. Eng. Sc.)*, University of New South Wales, Australia.

Cunnane, C. (1985). Factors affecting choice of distribution for flood series, *Journal of Hydrology,* 30, 25–36.

Cunnane, C. (1989). Statistical distributions for flood frequency analysis, Operational hydrological Report No. 5/33, *World Meteorological Organization (WMO)*, Geneva, Switzerland.

Haddad, K. and Rahman, A. (2008). Investigation on at-site flood frequency analysis in south-east Australia, *IEM Journal of Institution of Engineers, Malaysia,* 69(3), 59–64.

Haddad, K. and Rahman, A. (2011). 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., 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 versus parameter regression technique, *Journal of Hydrology,* 430–431, 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 in hydrologic regression analysis using ordinary and generalized least squares regression, *Journal of Hydrology*, 482, 119–128.

Institution of Engineers Australia (I.E. Aust.) (1987). *Australian rainfall and runoff: a guide to flood estimation*, in: Pilgrim HD (ed), The Institute of Engineers Australia, Canberra.

Interagency Advisory Committee on Water Data (IAWCD) (1982). *Guidelines for determining flood flow frequency: Bulletin 17-B*. Hydrology Sub-committee, Washington, DC, March, 1982.

Ishak, E.H., Rahman, A., Westra, S., Sharma, A. and Kuczera, G. (2010) Preliminary analysis of trends in Australian flood data. In: World Environmental and Water Resources Congress 2010, American Society of Civil Engineers (ASCE), 16–20 May 2010, Providence, Rhode Island, USA, 120–124.

Kopittke, R.A., Stewart, B.J. and Tickle, K.S. (1976). Frequency analysis of flood data in Queensland. In: Hydrological Symposium, Institution of Engineers Australia, National Conference, 76(2), 20-24.

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.

McMahon, T.A. and Srikanthan, R. (1981). Log Pearson III distribution is it applicable to flood frequency analysis of Australian streams? *Journal of Hydrology,* 52, 1-2, 139-147.

Merz, B. and Thieken, A.H. (2009). Flood risk curves and uncertainty bounds, *Natural Hazards*, 51, 437–458.

Merz, R., Bloschil, G. and Humer, G. (2008). National flood discharge mapping in Austria, *Natural Hazards,* 46, 53-72.

Meshgi, A. and Khalili, D. (2009a). Comprehensive evaluation of regional flood frequency analysis by L-ab LHmoments. 1. A re-visit to regional homogeneity, *Stochastic Environmental Resources and Risk Assessment*, 23, 119–135.

Meshgi, A. and Khalili, D. (2009b). Comprehensive evaluation of regional flood frequency analysis by L-abd LHmoments. II. Devewlopment of LH-moments parameters for the generalized Pareto and generalized logistic distributions, *Stochastic Environmental Research and Risk Assessment,* 23, 137–152.

Nathan, R.J. and Weinmann, P.E. (1991). Application of at-site and regional flood frequency analyses. In: Proceedings International Hydrology Water Resources Symposium, Perth, 2–4 October, 769-774.

Onoz, B. and Bayazit, M. (1995). Best-fit distribution of largest available flood samples, *Journal of Hydrology,* 167, 1-4, 195-208.

Rahman, A., Weinmann, P.E. and Mein, R.G. (1999). At-site flood frequency analysis: LP3-product moment, GEV-L moment and GEV-LH moment procedures compared. In: Proceeding Hydrology and Water Resource Symposium, Brisbane, 6–8 July, 2, 715–720.

Rahman, A., Haddad, K., Rahman, A.S., Haque, M.M. Kuczera, G. and Weinmann, P.E. (2014). An Overview of Preparation of Streamflow Database for ARR Project 5 Regional Flood Method, 35th Hydrology and Water Resources Symposium, Perth, Engineers Australia, 24-27 Feb, 2014, 678-685.

Rahman, S.A., Rahman, A., Zaman, M., Haddad, K., Ahsan, 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.

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

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.