

Identification of Outliers in Flood Frequency Analysis: Comparison of Original and Multiple Grubbs-Beck Test

Ayesha S. Rahman, Khaled Haddad, Ataur Rahman

Abstract—At-site flood frequency analysis is used to estimate flood quantiles when at-site record length is reasonably long. In Australia, FLIKE software has been introduced for at-site flood frequency analysis. 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. The new version of FLIKE has been incorporated with the multiple Grubbs and Beck test which can identify multiple numbers of potentially influential low flows. This paper presents a case study considering six catchments in eastern Australia which compares two outlier identification tests (original Grubbs and Beck test and multiple Grubbs and Beck test) and two commonly applied probability distributions (Generalized Extreme Value (GEV) and Log Pearson type 3 (LP3)) using FLIKE software. It has been found that the multiple Grubbs and Beck test when used with LP3 distribution provides more accurate flood quantile estimates than when LP3 distribution is used with the original Grubbs and Beck test. Between these two methods, the differences in flood quantile estimates have been found to be up to 61% for the six study catchments. It has also been found that GEV distribution (with L moments) and LP3 distribution with the multiple Grubbs and Beck test provide quite similar results in most of the cases; however, a difference up to 38% has been noted for flood quantiles for annual exceedance probability (AEP) of 1 in 100 for one catchment. This finding needs to be confirmed with a greater number of stations across other Australian states.

Keywords—Floods, FLIKE, probability distributions, flood frequency, outlier.

I. INTRODUCTION

FLOOD is one of the worst natural disasters that cause millions of dollars of damage each year including loss of human lives. Flood frequency analysis is the most direct method of estimating design floods, which is needed to design bridges, culverts, flood levees and other water infrastructure and in various water resources management tasks such as flood plain management and flood insurance studies.

Griffis and Stedinger [1] found that estimates of magnitude and frequency of floods using streamflow-gaging stations with shorter records of annual peakflow data had higher standard errors or uncertainties when compared to estimates using stream gauges with longer annual peakflow records. Flood

estimation should get the maximum information from the available data, be robust with respect to the distribution model and potentially influential low flows (PILFs). However, streamflow record length at many sites is often insufficient and identification of potentially influential low flows (PILFs) causes a great difficulty in fitting a probability distribution to available flood data.

Comparing various probability distributions and parameter estimation procedure had been done in numerous occasions in the past; however, due to the limited length of observed flood data as compared to the return period of interest, flood frequency analysis is deemed to be a challenging task and often associated with controversies [2]. The selection of an 'appropriate' probability distribution and associated parameter estimation procedure is an important step in flood frequency analysis. Flood frequency analysis has been widely researched in the past (e.g. [3]-[16]). 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 [2]. In practical applications, empirical suitability plays a much larger role in distributional choice than a priori reasoning [17], [18].

One of the earliest flood frequency analysis studies for New South Wales (NSW) coastal streams was carried out by Conway [19]. Kopittke et al. [20] did another study for Queensland (QLD). These studies established that the log Pearson type 3 (LP3) distribution was the most suitable distribution for the study catchments. To compare various probability distributions using the data from 172 catchments in Australia, McMahon and Srikanthan [3] used the moment ratio diagrams. They also concluded that the LP3 was the most suitable distribution for Australia. Based on the findings of these studies, it was recommended in Australian Rainfall and Runoff [21] that flood frequency analysis in Australia [22] should follow the footsteps of the USA i.e. to use LP3 distribution [23].

Since the publication of ARR [21], [22], there have been a number of studies to compare various probability distributions [24], [25]. For example, Nathan and Weinmann [26] 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. [4] compared a number of distributions using data from 61 stations in Australia, using the L-moments ratio diagram, they found that the generalized Pareto distribution (GPA) was the best-fit distribution followed by the GEV, three-parameter lognormal (LN3), and LP3. Haddad and Rahman [27] compared a number of distributions and

A. S. Rahman is a PhD student in Civil Engineering, School of Computing, Engineering and Mathematics, University of Western Sydney, Australia (phone: 61-2-47360533; e-mail: ayesha.rahman@uws.edu.au).

K. Haddad is a Research Fellow in the School of Computing, Engineering and Mathematics, University of Western Sydney, Australia (phone: 61-2-47360920; e-mail: k.haddad@uws.edu.au).

A. Rahman is an Associate Professor in Water Engineering in the School of Computing, Engineering and Mathematics, University of Western Sydney, Australia (phone: 61-2-47360145; e-mail: a.rahman@uws.edu.au).

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 [28] found that the two parameter distributions are preferable to Tasmania, with the lognormal appearing to be the best-fit distribution for Tasmania. As it seems an analyst might choose a different frequency model and fitting procedure for each catchment, but this could lead to inconsistencies in flood estimates across regions and among governmental agencies. National consistency in flood frequency estimates is important because these estimates are used in the allocation of resources and the implementation of the National Flood Insurance Program [29], [30]. For this reason, a national methodology should exhibit the characteristic of robustness. Robustness in this context means that the analysis does not perform poorly when its assumptions are violated.

An important step in flood frequency analysis is the detection of the PILFs in the flood data [31]. PILFs are unusually small observations of flood data which depart significantly from the trend of the rest of the data. Identification and treatment of PILFs are important issues in flood frequency analysis, because such observations can have a large influence on the estimate of extreme flood quantiles. In arid regions, even when it rains, channel losses can result in annual flood peaks that are zero or nearly zero, so that a LP3 distribution cannot fit the entire flood record without censoring zero values. Furthermore, unusually small values can result in relatively poor estimates of the large flood quantiles. In frequency analyses, one often uses a probability plot to examine if the sample data is consistent with a fitted curve [32], [33], unfortunately such decisions can be relatively subjective. The Bulletin 17B explicitly note that not dealing with this issue of PILFs would “significantly affect the [computed] statistical parameters.” Both Barnett and Lewis [34] and Beckman and Cook [32] discussed the notion of using outlier tests to identify unusual (high or low) data points that otherwise might have undue influence in a flood frequency analysis. Using a good low-outlier identification procedure has the potential for making low-outlier identification less subjective, by providing “rejection criteria which enable significance to be assessed” [34].

A wide range of test procedures for identifying PILFs have been examined in the past (e.g. [35]-[37], [34]), including methods for dealing with the case of multiple PILFs considered here [38]-[49]. Thompson [35] provided an early criterion for the rejection of an outlier based on the ratio of the sample standard deviation and an observation's deviation from the sample mean. An alternative test was proposed by Dixon [50], [51], who for high outliers proposed the test statistics for second most extreme observation in either tail of the distribution. Barnett and Lewis [34] also noted similar criteria as Dixon [50], [51]. Grubbs [36] and Grubbs and Beck [37] proposed a one sided 10% significance level criteria to identify PILFs. Rosner [39], [40] developed a sequential two sided outlier test, based on a generalization of the Grubbs [36] which usually detected outliers either too small or large but

this procedure was less computationally intensive and easy to apply in practice.

Bulletin 17B [23] was the guideline for flood frequency analysis in the United States for more than 30 years. Recently, there has been an attempt to revise Bulletin 17B to include recent advances in statistical techniques and computational resources [52] similar to the current revision of Australian Rainfall and Runoff (ARR). In Bulletin 17C, a new low outlier identification procedure, the multiple Grubbs-Beck (MGB) test [53] has been included. The MGB test is based on significance levels computed using the new approximations developed by Cohn et al. [29]. The MGB test is a generalization of the old Bulletin 17B original Grubbs-Beck (GB) test [36], [37].

In this paper the original GB test and the MGB test are compared using data from six Australian catchments. The Bulletin 17B's original GB test was based only on the distribution of the single smallest observation in a sample. As a result, even though multiple PILFs in flood data may exist, the original GB test rarely identifies more than a single PILF. The MGB test employs the actual distribution of the k^{th} smallest observation in a sample of n independent normal variates based upon significance levels provided by Cohn et al. [29], and is thus suited to test for multiple PILFs.

Kuczera [25] 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. It has additional advantages including the ability to (i) incorporate prior or regional information; (ii) incorporate stage-discharge uncertainty; (iii) assess parameter uncertainty obtained from regional information; and (iv) allow for threshold values (censoring). Recently a new version of FLIKE has been released. The older version of FLIKE needed manual identification of PILFs using the original Grubbs and Beck test. The new version of FLIKE is incorporated with the MGB test which attempts to identify multiple PILFs in the annual maximum flood series data.

The objective of this paper is to compare the performances of LP3 and GEV distributions for six selected stations in eastern Australia and also to explore the effects of censoring PILFs using the original GB test and MGB test.

II. STUDY AREA

For this study six catchments from New South Wales (NSW), Queensland (QLD) and Victoria (VIC) in Australia are considered (as shown in Fig. 1). Catchment area ranges from 87 to 900 km² with a mean of 348 km² and median of 140.5 km². Record length ranges from 33 to 60 years with a mean of 48 years and median of 49 years. All of the stations have log-space skew values significantly different from zero. Missing data points in the annual maximum flood series were in filled where possible by two methods. Method one involved comparing the monthly instantaneous maximum data (IMD)

with monthly maximum mean daily data (MMD) at the same station. If a missing month of IMD flow corresponded to a month of very low MMD flow, then that was taken to show that the annual maximum did not occur during that missing month. Method 2 involved a simple linear regression of the annual MMD flow against the annual IMD series of the same station. It must be mentioned that the regression equations

developed were used for filling gaps in the IMD record, but not to extend the overall period of record.

Rating curve extrapolation errors were identified by using a rating ratio test and treated using the in-built procedure 'rating curve error' case in FLIKE [25]. Table I shows the catchment area and record length for each station.

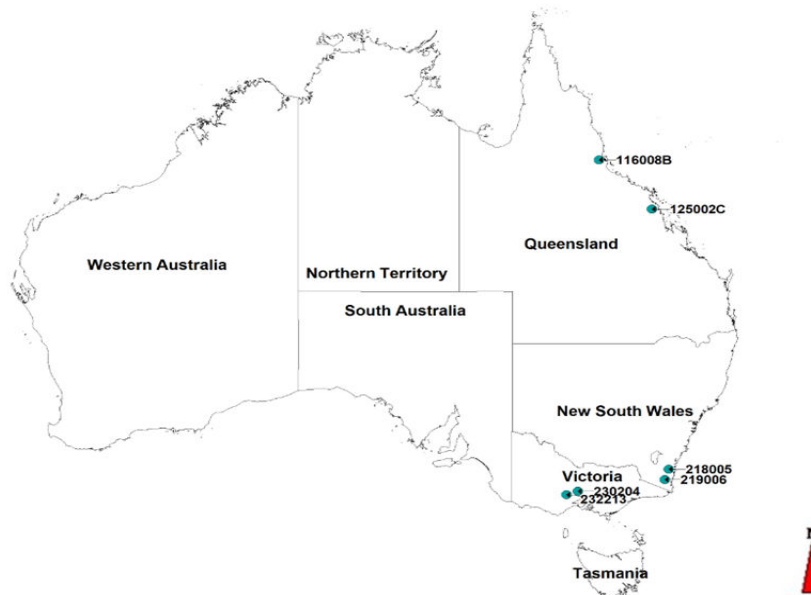


Fig. 1 Selected six catchments from eastern Australia

TABLE I
STATION DETAILS SHOWING STATION NAME, RIVER NAME, CATCHMENT AREA, RECORD LENGTH AND PERIOD OF RECORD

Station ID	Station name	River name	Catchment area (km ²)	Record length (years)	Period of record
218005	D/S Wadbilliga R Junct	Tuross	900	47	1965-2011
219006	Tantawangalo Mountain (Dam)	TantawangaloCk	87	60	1952-2011
116008B	Abergowrie	Gowrie Ck	124	58	1954-2011
125002C	Sarich's	Pioneer	740	51	1961-2011
230204	RiddellsCk	RiddellsCk	79	38	1974-2011
232213	U/S of Bungal Dam	LalLalCk	157	33	1977-2011

III. METHODOLOGY

The original GB test [36], [37] uses the at-site logarithms of the peak-flow data to calculate a one-sided, 10-percent significance-level critical value for a normally distributed sample. Although more than one recorded peak flow for a stream gage may be smaller than the Grubbs-Beck critical value, usually only one non-zero recorded peak flow is identified from the test as being a PILF. The original GB test which was recommended in Bulletin 17B [23], defines a low outlier (PILF) threshold as:

$$X_{crit} = \mu - k_n \sigma \quad (1)$$

where K_n is a one-sided, 10% significance-level critical value for an independent sample of n normal variates, and μ and σ denote the sample mean and standard deviation of the entire data set. Any observation less than X_{crit} is declared a 'low

outlier (PILF)' [23]. As per Bulletin 17B, PILFs are omitted from the sample and the frequency curve is adjusted, using a conditional probability adjustment [23]. K_n values are tabulated in section A4 of IACWD [23] based on Grubbs and Beck [37].

Stedinger et al. [33] provide an accurate approximation of K_n for $5 \leq n \leq 150$:

$$k_n \approx -0.9043 + 3.345\sqrt{\log_{10}(n)} - 0.4046 \log_{10}(n) \quad (2)$$

The original GB test only identifies one outlier/PILF from a particular data set, but there can be more numbers of PILFs available in the data. A method for statistically detecting multiple PILFs using a generalized Grubbs-Beck test has been developed [54]. The MBG test is also based on a one-sided, 10-percent significance-level critical value for a normally distributed sample, but the test is constructed so that groups of

ordered data are examined (for example, the eight smallest values) and excluded from the dataset when the critical value is calculated. If the critical value is greater than the eighth smallest value in the example, then all eight values are considered to be PILFs according to this new method.

Here, one considers whether $\{X_{[1:n]}, X_{[2:n]}, \dots, X_{[k:n]}\}$ are consistent with a normal distribution and the other observations in the sample by examining the statistic [29]:

$$\tilde{\omega}_{[k:n]} \equiv (k_{[k:n]} - \mu_k) / \sigma_k \quad (3)$$

where $X_{[k:n]}$ denotes the k^{th} smallest observation in the sample, and

$$0\mu_k \equiv \frac{1}{n-k} \sum_{j=k+1}^n X_{[j:n]} \quad (4)$$

$$\sigma_k \equiv \frac{1}{n-k-1} \sum_{j=k+1}^n (X_{[j:n]} - \mu_k)^2 \quad (5)$$

Here the partial mean μ_k and partial variance σ_k are computed using only the observations larger than $X_{[k:n]}$ to avoid swamping.

To implement the MGB test, recommended for Bulletin 17C, the following two steps are involved: (i) starting at the median and sweeping *outward* towards the smallest observation, each observation is tested with a MGB test significance level α_{out} . If the k^{th} smallest observation is identified as a low outlier, the outward sweep stops and all observations less than the k^{th} smallest (i.e. $j = 1, \dots, k$) are also identified as low outliers. (ii) An *inward* sweep always starts at the smallest observation and moves towards the median, with a significance level of α_{in} . If an observation $m \geq 1$ fails to be identified by the inward sweep, the inward sweep stops. The total number of low-outliers/PILFs identified by the MGB test is then the maximum of k and $m - 1$. The algorithm has two parameters that need to be specified [29]: (i) outward Sweep significance level for each comparison, α_{out} ; and (ii) inward Sweep significance level for each comparison, α_{in} .

Bulletin 17B used a 10% significance test with a single outlier threshold. The new outlier detection procedure uses two multiple threshold sweeps. Those thresholds are the Cohn et al. [29] $p(k;n)$ function which correctly describes if the k^{th} smallest observation in a normal sample of n variates is unusual. The first outward sweep seeks to determine if there is some break in the lower half of the data that would suggest the sample is best treated as if it had a number of low outliers. The second sweep using a less severe significance level, say $p(k;n) \leq 10\%$, mimics Bulletin 17B's willingness to identify one or more of the smallest observations as low outliers so that the frequency analysis is more robust.

A reasonable concern is that a flood record could contain more than one low outlier and the additional outliers can cause the original GB test statistic to fail to recognize the smallest observation as an outlier (by inflating the sample mean and variance). This effect is known as *masking* [38]. Inward sweep tests are particularly susceptible to masking [34], therefore an outward sweep is desirable to avoid the masking problem [47]. Rosner [40] uses a two-sided outward sweep. Spencer and

McCuen [55] recommended an outward sweep with their test for multiple outliers when fitting a LP3 distribution.

The generalized extreme value (GEV) distribution is a family of continuous probability distributions developed within extreme value theory to combine the Gumble, Fréchet and Weibull families also known as type I, II and III extreme value distributions. Here GEV distribution is used to compare the results with the LP3 distribution.

The LP3 distribution has been used for several decades to model annual maximum flood series. Estimation of the parameters of the distribution using a MOM estimator in log space was suggested by Beard [56]; this method was used presumably for computational ease. The only complication was the need for frequency factors to compute quantile estimates given the sample moments of the logs of the data. The needed frequency factors were tabulated in Benson [57] and in Bulletin 17B. Kirby [58] provided an excellent approximation. Now days they can be computed directly with built-in functions in many software packages, including Excel and MATLAB.

To fit the LP3 distribution it is required to calculate the mean, standard deviation, and skew coefficient of the logarithms of the annual maximum flood data. Estimate of the P percent annual exceedence probability (AEP) flood is computed by inserting the three statistics of the frequency distribution into the equation:

$$\log Q_p = \bar{X} + k_p S \quad (6)$$

where Q_p is the P-percent AEP flood or flood quantile; \bar{X} is the mean of the logarithms of the annual peak flows; K_p is a factor based on the skew coefficient and the given AEP and is obtained from [23]; and S is the standard deviation of the logarithms of the annual peak flows.

The mean, standard deviation and skew coefficient can be estimated from the available sample data (recorded annual-peak flows), but a skew coefficient calculated from small samples tends to be an unreliable estimator of the population skew coefficient. Accordingly, the guidelines in Bulletin 17B [23] indicate that the skew coefficient calculated from at-site sample data (station skew) needs to be weighted with a generalized, or regional, skew determined from an analysis of selected long-term stream gauges in the study region. The value of the skew coefficient used in equation 6 is the weighted skew that is based on station skew and regional skew. However, Australian Rainfall and Runoff 1987 did not adopt the weighted skew for application in Australia [22].

IV. RESULTS

For three stations (218005, 219006, 116008B), the original GB test did not find any PILF but the MGB test found 24, 27, 29 PILFs respectively and for the remaining three stations (125002C, 230204, 232213) the original GB test found only one PILF for each of them but the MGB test found 26, 17, 17 PILFs. These results show a remarkable difference between the results by the two methods.

TABLE II
NUMBER OF PILFs IDENTIFIED USING MGB TEST AND THE ORIGINAL GB TEST

Station ID	Number (%) of PILFs identified by MGB test	Number of PILFs identified by original GB test
218005	24 (51.06%)	None
219006	27 (45.76%)	None
116008B	29 (56.86%)	None
125002C	26 (49.01%)	1
230204	17 (44.74%)	1
232213	17 (51.52%)	1

Table III presents the log space skews of the original AM flood data set and after removing the PILFs using the original GB test and the MGB test. This table shows that application of MGB results in a greater reduction in log space skew than the original GB test. This is likely to affect the quantile estimation by LP3 distribution as skew plays an important role in the fitting of LP3 distribution.

Figs. 2 and 3 show how application of GB and MGB tests affects the fitting of a probability distribution to the AM flood series for Station 218004. The application of GB test did not identify any PILF for Station 218004; however, the

application of MGB test identifies 24 PILFs, the fitting of LP3 distribution is remarkably better in Fig. 3 (where MGB test is applied) than in Fig. 2 (where GB test is applied). In another example, Figs. 4-6 show the effects of PILFs on fitting a probability distribution to the AM flood series for Station 230204 where the application of GB test identified only one PILF; however, the MGB test identified 17 PILFs. Fig. 6 shows a better fit of the LP3 distribution to the AM flood data series (where MGB test is applied) than in Fig. 5 (where GB test is applied).

TABLE III
SKEW OF EACH STATION WITHOUT REMOVING ANY PILFs, PILFs REMOVED BY ORIGINAL GB TEST AND PILFs REMOVED BY MGB TEST RESPECTIVELY

Station ID	Skew (no PILFs removed)	Skew (PILFs removed by original GB test)	Skew (PILFs removed by MGB test)
218005	-0.375	-0.375	-0.553
219006	-0.514	-0.514	0.367
116008B	-0.310	-0.310	-0.079
125002C	-0.901	-0.757	0.095
230204	-0.671	-0.373	0.079
232213	-1.244	-1.197	-0.531

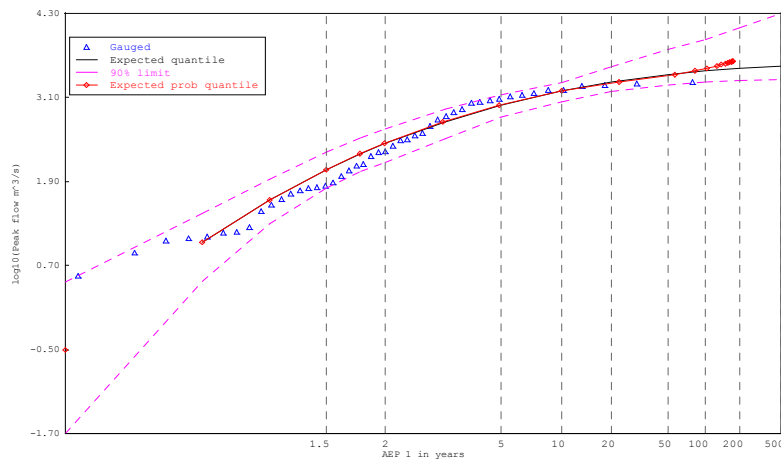


Fig. 2 Flood frequency curve for Station 218005 using LP3 distribution (there was no PILF as per original GB test)

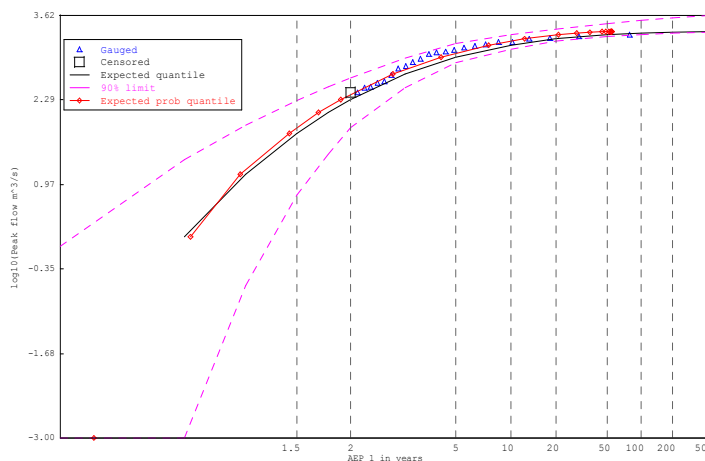


Fig. 3 Flood frequency curve for Station 218005 using LP3 distribution (24 PILFs censored as per MGB test)

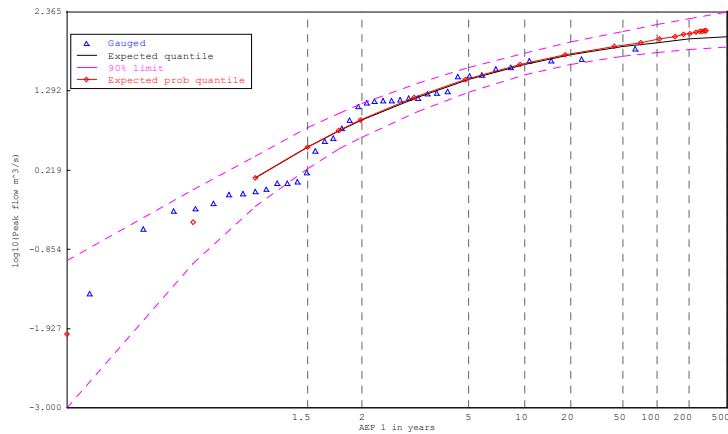


Fig. 4 Flood frequency curve for Station 230204 using LP3 distribution (no PILF censored)

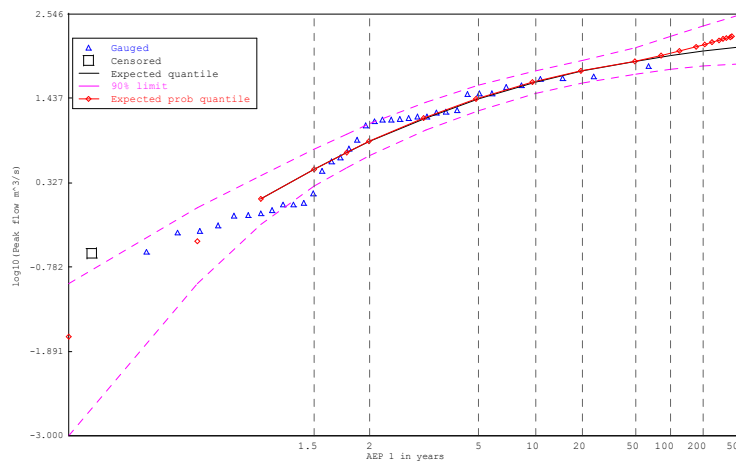


Fig. 5 Flood frequency curve for Station 230204 using LP3 distribution (one PILF censored as per original GB test)

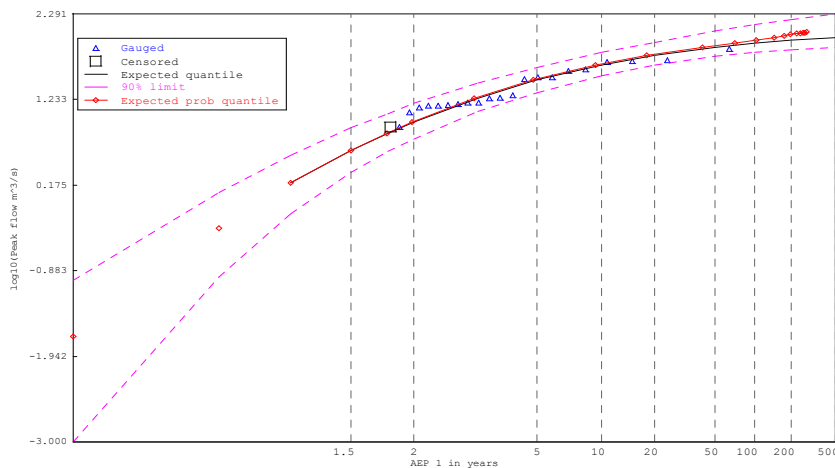


Fig. 6 Flood frequency curve for Station 230204 using LP3 distribution (17 PILFs censored as per MGB test)

Fig. 7 shows the fitting of the GEV distribution to Station230204 without removing any PILFs and Fig. 8 shows the fitting of LP3 distribution after removing 17 PILFs. From

these two plots it is evident that LP3 distribution (Fig. 8) shows a better fit to the AM flood series than the GEV distribution.

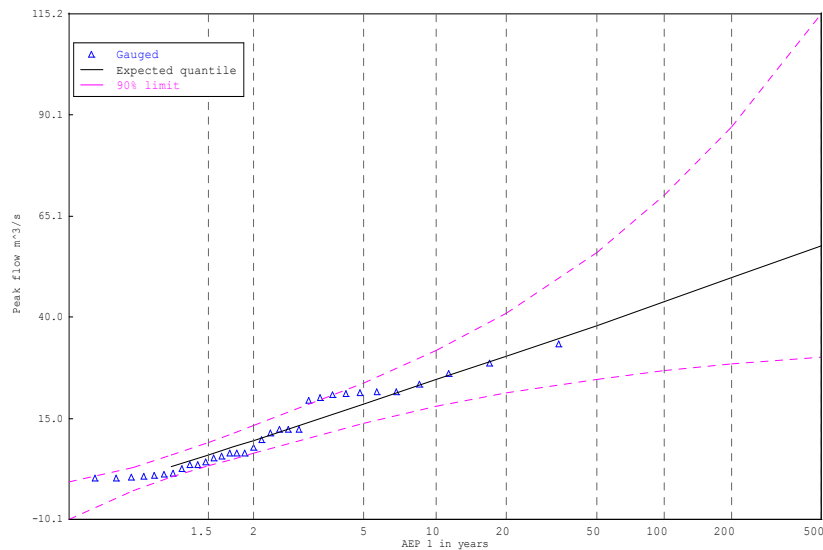


Fig. 7 Fitting of the GEV distribution to the AM flood data for Station 230204 (No PILF censored)

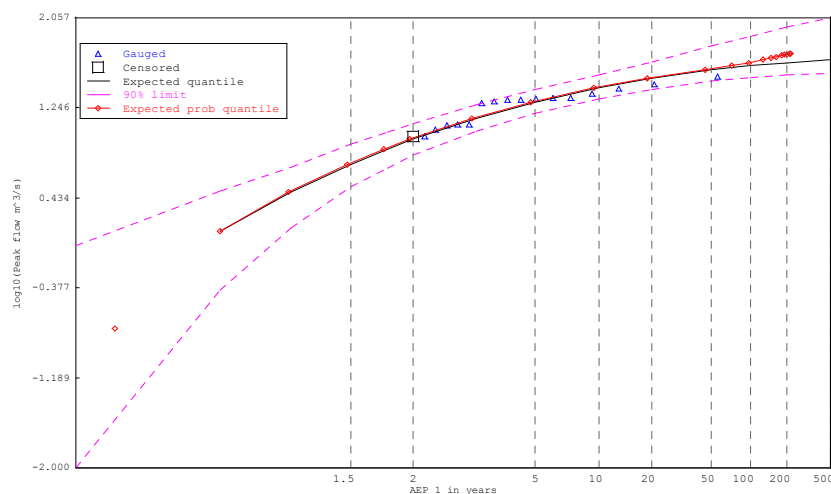


Fig. 8 Fitting of the LP3 distribution to the AM flood data for Station 230204 (17 PILFs censored by MGB test)

Table IV shows the flood quantiles using LP3 distribution where the PILFs are identified and censored by the original GB test and MGB test for AEPs of 1 in 10, 1 in 20, 1 in 50 and 1 in 100. It is found that there are notable differences between the two methods where flood quantiles show a variation in the range of -61% to 28%. Table V shows the variation between the flood quantiles estimated by two methods: LP3 with MGB test and GEV with L moments. It is found that for Station 218005 GEV distribution underestimates 1 in 10 AEP flood quantile by 6.68%, but for AEPs of 1 in 20, 1 in 50 and 1 in 100 GEV overestimates the flood quantiles by 4.4%, 24.1% and 38%, respectively. For other 5 stations the variations between the GEV and LP3 are mixed i.e. a combination of over- and under-estimation by 0.24% to 26.7%. These results highlight the expected differences in flood quantile estimates between the LP3 and GEV distributions in eastern Australia.

TABLE IV

ESTIMATED FLOOD QUANTILES AND PERCENTAGE DIFFERENCES BETWEEN THE TWO SETS OF QUANTILES: LP3 WITH MGB TEST AND LP3 WITH ORIGINAL GB TEST

Station ID	Estimated quantiles (m ³ /s) using LP3 distribution (PILFs removed by MGB test)				Estimated quantiles (m ³ /s) using LP3 distribution (PILFs removed by original GB test (% difference between LP3 with MGB test and LP3 with original GB test) AEPs (1 in Y)			
	10	20	50	100	10	20	50	100
218005	1422.74	1780.04	2063.82	2183.63	1072.7 (24.60%)	1704.52 (4.24%)	2690.8 (-30.38%)	3519.78 (-61.19%)
219006	184.37	279.38	427.86	555.27	133.51 (27.59%)	228.33 (18.27%)	402.83 (5.85%)	576.06 (-3.74%)
116008B	920.88	1136.21	1342.61	1451.81	886.91 (3.69%)	1141.43 (-0.46%)	1454.31 (-8.32%)	1670.81 (-15.08%)
125002C	3512.61	4198.86	4887.37	5276.62	2879.16 (18.03%)	3808.77 (9.29%)	4866.33 (0.43%)	5528.19 (-4.77%)
230204	46.02	60.56	76.41	85.86	34.92 (24.12%)	53.71 (11.31%)	81.32 (-6.43%)	103.21 (-20.21%)
232213	26.7	32.38	38.36	41.9	22.3 (16.48%)	28.38 (12.35%)	34.89 (9.05%)	38.75 (7.52%)

TABLE V

COMPARISON OF FLOOD QUANTILES BY LP3 WITH MGB TEST AND GEV WITH L MOMENTS

Station ID	Flood quantiles by GEV-L moments (m ³ /s)				Flood quantiles by LP3 with MGB test (m ³ /s) (% difference between LP3 with MGB test and GEV with L moments) AEPs (1 in Y)			
	10	20	50	100	10	20	50	100
218005	1333.6	1861.3	2719.2	3522.4	1422.74 (-6.68%)	1780.04(4.37%)	2063.82(24.10%)	2183.63(38.01%)
219006	151	220.6	344.1	469.8	184.37 (-22.10%)	279.38 (-26.65%)	427.86 (-24.34%)	555.27 (-18.19%)
116008B	819.1	1074.6	1457.8	1789.5	920.88 (-12.43%)	1136.21 (-5.73%)	1342.61(7.90%)	1451.81(18.87%)
125002C	3263.1	4188.7	5488.6	6544.1	3512.61 (-7.65%)	4198.86 (-0.24%)	4887.37(10.95%)	5276.62(19.37%)
230204	37.5	52.4	76.9	100	46.02 (-22.72%)	60.56 (-15.57%)	76.41(0.64%)	85.86(14.14%)
232213	24.5	30.3	38	43.9	26.7 (-8.98%)	32.38 (-6.86%)	38.36 (-0.95%)	41.9(4.56%)

V. CONCLUSION

This paper presents a case study using six catchments from eastern Australia which examines two outlier tests being the original Grubbs and Beck (GB) test and multiple Grubbs and Beck (MGB) test. Two different probability distributions i.e. the GEV and LP3 have been adopted in flood frequency analysis, which are incorporated into the new FLIKE software. For three stations, the original GB test did not detect any potentially influential low flows (PILFs); however, for these stations MGB test detected 46% to 57% of the annual maximum flood peaks as PILFs. For the remaining three stations, the original GB test identified one PILF from each station and the MGB test identified 45% to 51% as PILFs. Between these two methods, the differences in flood quantile estimates have been found to be up to 61% for the six study catchments. It has also been found that GEV distribution (with L moments) and LP3 distribution with the multiple Grubbs and Beck test provide quite similar results in most of the cases; however, a difference up to 38% has been noted for flood quantiles for annual exceedance probability (AEP) of 1 in 100 for one catchment. This finding needs to be confirmed with a greater number of stations across other Australian states.

ACKNOWLEDGMENT

Authors would like to acknowledge Australian Rainfall and Runoff revision Project 5 team for providing the data and FLIKE software for this study and Engineers Australia and Geosciences Australia for providing financial support for this project and Professor George Kuczera, Mr Erwin Weinmann, Assoc Professor James Ball, Mr Mark Babister and Dr William Weeks for their support and input to this study.

REFERENCES

- [1] V. W. Griffiths, and J. R. Stedinger, "The LP3 distribution and its application in flood frequency analysis, 2. Parameter estimation methods," *J. of Hydrol. Engineering*, vol. 12(5), pp. 492–500, 2007.
- [2] B. Bobée, G. Cavadas, F. Ashkar, J. Bernier, and P. Rasmussen, "Towards a systematic approach to comparing distributions used in flood frequency analysis," *J Hydrol*, vol. 142, pp. 121–136, 1993.
- [3] T. A. McMahon, and R. Srikanthan, "Log Pearson III distribution is it applicable to flood frequency analysis of Australian streams?," *J Hydrol*, vol. 52(139), pp. 147, 1981.
- [4] R. M. Vogel, T. A. McMahon, and F. H. S. Chiew, "Flood flow frequency model selection in Australia," *J Hydrol*, vol. 146(421), pp. 449, 1993.
- [5] B. C. Bates, A. Rahman, R. G. Mein, and P. E. Weinmann, "Climatic and physical factors that influence the homogeneity of regional floods in south-eastern Australia," *Water Resour Res*, vol. 34(12), pp. 3369–3381, 1998.
- [6] R. Merz, G. Blöschl, and G. Humer, "National flood discharge mapping in Austria," *Nat Hazards*, vol. 46, pp. 53–72, 2008.
- [7] A. Meshgi, and D. Khalili, "Comprehensive evaluation of regional flood frequency analysis by L-abdLH moments. 1. A re-visit to regional

- homogeneity,"*Stoch Environ Res Risk Assess*, vol. 23, pp. 119–135, 2009a.
- [8] A. Meshgi, and D. Khalili, "Comprehensive evaluation of regional flood frequency analysis by L-abdLHmoments. II. Development of LH-moments parameters for the generalized Pareto and generalized logistic distributions,"*Stoch Environ Res Risk Assess*, vol. 23, pp. 137–152, 2009b.
- [9] E. H. Ishak, A. Rahman, S. Westra, A. Sharma, and G. Kuczera, "Preliminary analysis of trends in Australian flood data," in *Proc. World Environmental and Water Resources Congress, American Society of Civil Engineers (ASCE)*, Providence, Rhode Island, USA, 2010, pp. 120–124.
- [10] E. H. Ishak, K. Haddad, M. Zaman, and A. Rahman, "Scaling property of regional floods in New South Wales Australia,"*Nat Hazards*, vol. 58, pp. 1155–1167, 2011.
- [11] K. Haddad, A. Rahman, and G. Kuczera, "Comparison of ordinary and generalised least squares regression models in regional flood frequency analysis: a case study for New South Wales,"*Aust J Water Resour*, vol. 15(2), pp. 59–70, 2011.
- [12] K. Haddad, A. Rahman, and J. R. Stedinger, "Regional Flood Frequency Analysis using Bayesian Generalized Least Squares: a Comparison between Quantile and Parameter Regression Techniques,"*Hydrol Process*, vol. 25(1), pp. 14, 2012.
- [13] K. Haddad, and A. Rahman, "Regional flood frequency analysis in eastern Australia: Bayesian GLS regression-based methods within fixed region and ROI framework—quantile regression versus parameter regression technique,"*J Hydrol*, vol. 430–431, pp. 142–161, 2012.
- [14] A. Rahman, K. Haddad, M. Zaman, G. Kuczera, P. E. Weinmann, and W. Weeks, "Regional flood estimation in Australia: an overview of the study for the upgrade of Australian Rainfall and Runoff," in *Proc. Hydrology and Water Resources Symposium, Engineers Australia*, Sydney, Australia, 2012, pp. 1441–1448.
- [15] K. Haddad, A. Rahman, M. Zaman, and S. Shrestha, "Applicability of Monte Carlo cross validation technique for model development and validation in hydrologic regression analysis using ordinary and generalised least squares regression,"*J Hydrol*, vol. 482, pp. 119–128, 2013.
- [16] A. S. Rahman, A. Rahman, M. Zaman, K. Haddad, A. Ashan, and M. A. Imteaz, "A Study on Selection of Probability Distributions for At-site Flood Frequency Analysis in Australia,"*Natural Hazards*, vol. 69, pp. 1803–1813, 2013.
- [17] C. Cunneane, "Factors affecting choice of distribution for flood series,"*Hydrol Sci J*, vol. 30, pp. 25–36, 1985.
- [18] C. Cunneane, "Statistical distributions for flood frequency analysis," in *Proc. Operational hydrological Report No. 5/33, World Meteorological Organization (WMO)*, Geneva, Switzerland, 1989.
- [19] K. M. Conway, "Flood frequency analysis of some NSW coastal rivers," *Thesis (M. Eng. Sc.)*, University of New South Wales, Australia, 1970.
- [20] R. A. Kopittke, B. J. Stewart, and K. S. Tickle, "Frequency analysis of flood data in Queensland," in *Proc. Hydrological Symposium, Institution of Engineers Australia, National Conference*, 1976, Publication No. 76/2, 20:24.
- [21] Institution of Engineers Australia (I.E. Aust.), *Australian Rainfall and Runoff: A Guide to Flood Estimation*, 2001.
- [22] Institution of Engineers Australia (I.E. Aust.), *Australian Rainfall and Runoff: A Guide to Flood Estimation*, 1987.
- [23] Interagency Advisory Committee on Water Data (IAWCD), *Guidelines for Determining Flood Flow Frequency: Bulletin 17-B. Hydrol. Subcomm.*, Washington, DC, 1982.
- [24] A. Rahman, P. E. Weinmann, and R. G. Mein, "At-site flood frequency analysis: LP3-product moment, GEV-L moment and GEV-LH moment procedures compared," in *Proc. Hydrology and Water Resource Symposium*, Brisbane, 1999, pp. 715–720.
- [25] G. Kuczera, "Comprehensive at-site flood frequency analysis using Monte Carlo Bayesian inference," *Water Resour Res*, vol. 35(5), pp. 1551–1557, 1999.
- [26] R. J. Nathan, and P. E. Weinmann, "Application of at-site and regional flood frequency analyses," in *Proc. International Hydrology Water Resources Symposium*, Perth, 1991, pp. 769–774.
- [27] K. Haddad, and A. Rahman, "Investigation on at-site flood frequency analysis in south-east Australia," *IEM Journal, The Journal of The Institution of Engineers*, Malaysia, vol. 69(3), pp. 59–64, 2008.
- [28] K. Haddad, A. Rahman, "Selection of the best fit flood frequency distribution and parameter estimation procedure: a case study for Tasmania in Australia,"*Stoch Environ Res Risk Assess*, vol. 25, pp. 415–428, 2010.
- [29] T. A. Cohn, J. F. England, C. E. Berenbrock, R. R. Mason, J. R. Stedinger, and J. R. Lamontagne, "A generalized Grubbs-Beck test statistic for detecting multiple potentially influential outliers in flood series,"*Water Resources Research*, vol. 49, pp. 5047–5058, 2013.
- [30] W. O. Thomas, "A uniform technique for flood frequency analysis,"*J. Water Resour. Plann. Manage*, vol. 111(3), pp. 321–337, 1985.
- [31] B. Saf, "Assessment of the effects of discordant sites on regional flood frequency analysis,"*J Hydrol*, vol. 380 (3–4), pp. 362–375, 2010.
- [32] R. J. Beckman, and R. D. Cook, "Outliers,"*Technometrics*, vol. 25(2), pp. 119–149, 1983.
- [33] J. R. Stedinger, R. M. Vogel, and E. Foufoula-Georgiou, *Frequency Analysis of Extreme Events*, chap. 18, pp. 99, McGraw Hill, New York, 1993.
- [34] V. Barnett, and T. Lewis, *Outliers in Statistical Data*, John Wiley, New York, 1994.
- [35] W. Thompson, "On a criterion for the rejection of observations and the distribution of the ratio of deviation to sample standard deviation,"*Ann. Math. Stat.*, vol. 6, pp. 214–219, 1935.
- [36] F. E. Grubbs, "Procedures for Detecting Outlying Observations in Samples,"*Technometrics*, vol. 11(1), pp. 1–21, 1969.
- [37] F. E. Grubbs, G. Beck, "Extension of sample sizes and percentage points for significance tests of outlying observations,"*Technometrics*, vol. 4(14), pp. 847–853, 1972.
- [38] G. Tietjen, and R. Moore, "Some Grubbs-type statistics for the detection of several outliers,"*Technometrics*, vol. 14(3), pp. 583–597, 1972.
- [39] B. Rosner, "On the detection of many outliers,"*Technometrics*, vol. 17(2), pp. 221–227, 1975.
- [40] B. Rosner, "Percentage points for a generalized ESD many outlier procedure,"*Technometrics*, vol. 25(2), pp. 165–172, 1983.
- [41] P. Prescott, "An approximate test for PILFs in linear models,"*Technometrics*, vol. 17(1), pp. 129–132, 1975.
- [42] P. Prescott, "Examination of the behaviour of tests for outliers when more than one outlier is present,"*Appl. Stat.*, vol. 27, pp. 10–25, 1978.
- [43] J. Gentleman, and M. Wilk, "Detecting PILFs. II. Supplementing the direct analysis of residuals,"*Biometrics*, vol. 31(2), pp. 387–410, 1975.
- [44] M. Marasinghe, "A multistage procedure for detecting several PILFs in linear regression,"*Technometrics*, vol. 27(4), pp. 395–399, 1985.
- [45] P. Rousseeuw, and B. V. Zomeren, "Unmasking multivariate PILFs and leverage points,"*J. Am. Stat. Assoc.*, vol. 85(411), pp. 633–639, 1990.
- [46] A. Hadi, and J. Simonoff, "Procedures for the identification of multiple outliers in linear models,"*J. Am. Stat. Assoc.*, vol. 88(424), pp. 1264–1272, 1993.
- [47] C. Spencer, and R. McCuen, "Detection of PILFs in Pearson type III data,"*J. Hydrol. Eng.*, vol. 1, pp. 2–10, 1996.
- [48] P. Rousseeuw, and A. Leroy, *Robust Regression and Outlier Detection*, John Wiley, Hoboken, N. J, 2003.
- [49] S. Verma, and A. Quiroz-Ruiz, "Critical values for six Dixon tests for outliers in normal samples up to sizes 100, and applications in science and engineering,"*Rev. Mex. Cienc. Geol.*, vol. 23(2), pp. 133–161, 2006.
- [50] W. J. Dixon, "Analysis of extreme values,"*Ann. Math. Stat.*, vol. 21(1), pp. 488–506, 1950.
- [51] W. J. Dixon, "Ratios involving extreme values,"*Ann. Math. Stat.*, vol. 22(1), pp. 68–78, 1951.
- [52] Interagency Advisory Committee on Water Data (IAWCD), *Robust National Flood Frequency Guidelines: What is an Outlier? Bulletin 17-C*, 2013.
- [53] J. R. Lamontagne, J. R. Stedinger, T. A. Cohn and N. A. Barth, "Robust national flood frequency guidelines: What is an outlier?" in *Proc. World Environmental and Water Resources Congress, ASCE*, 2013.
- [54] A. J. Gotvald, N. A. Barth, A. G. Veilleux, and C. Parrett C, "Methods for determining magnitude and frequency of floods in California, based on data through water year 2006," *U.S. Geological Survey*, Reston, Virginia, 2012.
- [55] R. H. McCuen, B. M. Ayyub, "Probability, statistics, and reliability for engineers and scientists," A Chapman & Hall book, USA, 1996.
- [56] L. R. Beard, "Statistical methods in hydrology, Civil works investigation project CW-151," *U.S. Army Corps of Engineers*, Sacramento, California, 1962.
- [57] M. A. Benson, "Uniform flood-frequency estimating methods for federal agencies,"*Water Resour. Res.*, vol. 4(5), pp. 891–908, 1968.
- [58] W. Kirby, "Computer-oriented Wilson-Hilferty transformation that preserves the first three moments and the lower bound of the Pearson type 3 distribution,"*Water Resour. Res.*, vol. 8(5), pp. 1251–1254, 1972.