Proc. 3rd International Conference on Water and Environmental Engineering (iCWEE2022)
27-30 Nov 2022, Sydney, Australia
ISBN: 978-0-6456692-0-6

Low Flows in Annual Maximum Flood Data: Does It Matter in Flood Frequency Analysis?

Laura Rima[1], Ataur Rahman[2], Khaled Haddad[3], Mohammed A Alim[4]

[1]PhD Student, School of Engineering, Design and Built Environment, Western Sydney University, Sydney, Australia
[2]Professor, School of Engineering, Design and Built Environment, Western Sydney University, Sydney, Australia
[3] Drainage, Marine & Floodplain Engineer, City of Canada Bay, Sydney, Australia
[4]First Year Course Advisor, School of Engineering, Design and Built Environment, Western Sydney University, Sydney, Australia

Corresponding author's E-mail: 19274818@student.westernsydney.edu.au

Abstract

Flood frequency analysis (FFA) is a widely used statistical technique for estimating design floods, which is required for the management of water projects and the design of hydraulic structures. Sometimes, the annual maximum flood (AMF) data series contains low flow observations, and the presence of these small observations can cause major problems in fitting an appropriate probability distribution and selecting a statistical model in AMF data series modelling. This paper examines the effects of low flows on flood frequency distributions and how censoring these values can affect the flood quantiles estimates. A total of 183 catchments from Victoria in Eastern Australia are selected and Generalised Extreme Value (GEV) and Log Pearson type III (LP3) distributions are used. Furthermore, the multiple Grubbs and Beck (MGB) test is used to identify low flows in the selected AMF data using FLIKE software. It was found that 72% of the catchments required censoring of low flows using LP3 distribution. Also, censoring low flows with MGB test provided a more accurate fitting of the LP3 distribution to the AMF data. Thus, it is strongly recommended that low flows be censored in FFA using LP3 distribution since they can affect the accuracy of the quantile estimates.

Keywords: Flood, Flood frequency analysis, Low flows, Multiple Grubbs test, Censoring.

1. INTRODUCTION

Flooding is a natural hazard that happens frequently and has a serious impact on both human lives and infrastructure (Haddad and Rahman, 2019). Hydrologists face great challenges due to the increase in flooding which affects the entire society (Stojkovic et al., 2017). Flood frequency analysis (FFA) is the preferred method to design flood estimates which are needed for the design of hydraulic structures to prevent flooding and save lives. FFA requires long periods of streamflow data to produce accurate quantile estimates (Rahman et al., 2018). However, many catchments in Australia have poor or no recorded streamflow data available. In these areas, regional flood frequency analysis (RFFA) is adopted where information on flood characteristics is transferred from gauged catchments to ungauged catchments based on regional homogeneity (Cunnane, 1989).

Streamflow data preparation and selection of best-fit distribution are important steps in FFA as good quality data and appropriate selection of distribution help to generate accurate outcomes (Rahman et al., 2015). Numerous RFFA techniques have been developed throughout the years, each with its own assumptions, data requirements, and restrictions. There is currently no unique RFFA method that has been used globally, and most of them have a significant error margin (Haddad and Rahman, 2012). In order to construct more flood-safe infrastructures that limit flood damage, it is desirable to develop new and more accurate RFFA techniques. In Australia, many studies have identified the Log Pearson III (LP3) and the Generalised Extreme Value (GEV) distribution as the best fit distributions in FFA (Haddad et al., 2009; Rahman et al., 2013 and Ahn and Palmer, 2016).

Low flow values are the lowest flow events in the annual maximum flood (AMF) data series that can occur during a given period of record and present significant deviation from the trend of the other values. This can happen due to data collection errors and methods adopted or just because of natural factors (Stojkovic et al., 2017). One of the major challenges in data preparation is identifying these low-flow values. Several test procedures have been applied for identifying low flows in AMF data (such as Thompson, 1935; Grubbs and Beck, 1972, and Barnett and Lewis, 1994). Among these tests, the Grubbs-Beck (GB) test, which was introduced by Grubbs (1969) and recommended by the federal guidelines in the United States Bulletin 17B (IACWD, 1982), has been widely adopted as a censored-data statistical technique for the detecting of low flows in FFA. Cohn et al. (2013) presented a generalised GB (multiple GB tests (MGB)) that can detect multiple low flows in the AMF data series, and it has been included in Bulletin 17C (England et al., 2008). Moreover, Rahman et al. (2014) compared flood quantiles derived from GEV and LP3 distributions using two outlier detection tests (GB and MGB) and concluded that MGB provides better results with LP3.

The objective of this paper is to investigate the presence of low flows in the new AMF data series and their impacts on FFA.

2. STUDY AREA AND DATA

This study focuses on 183 catchments from Victoria (VIC) state of Australia. These catchments were used in the Australian Rainfall and Runoff (ARR) Project 5 (Rahman et al., 2015). Table 1 summarises the geographical information of the selected catchments. The catchment areas of the selected catchments range from 3 km^2 to 997 km^2 with a mean of 272 km^2 where 50% of the catchments have an area of more than 50 km^2. The AMF record lengths of these 183 catchments range from 20 to 67 years with a mean of 43 years where 88% of the catchments have a record length of more than 30 years.

Table 1. Summary of catchment areas and AMF record lengths for 183 catchments in Victoria.

Catchment area (km^2)			**Record length (years)**	
Minimum	3		**Minimum**	20
Maximum	997		**Maximum**	67
Average	272		**Average**	43
	Number of stations			**Number of stations**
Less than 200 km^2	91		Less than 30 years	22
Between 200 and 400	48		Between 30 and 40	18
Between 400 and 600	22		Between 40 and 50	126
Between 600 and 800	14		Between 50 and 60	14
More than 800 km^2	8		More than 60 years	3

3. METHODOLOGY

At-site FFA for each of the 183 catchments in Victoria was conducted using the FLIKE software (Kuczera and Franks, 2016). Flood quantiles were estimated using GEV and LP3 distributions for annual exceedance probabilities (AEPs) of 1%, 2%, 5% and 10%. Low flows were censored using MGB test with the LP3 distribution.

3.1. Generalised Extreme Value (GEV) distribution

GEV distribution is a three parameters distribution: location μ, scale α and shape κ. The cumulative density function (CDF) is defined in Chowdhury et al. (1991) as:

$$\boldsymbol{F(x) = \exp\{-\left[1 - \kappa\left(\frac{x}{\propto} - \mu\right)\right]^{1/\kappa}\}} \qquad \kappa \neq 0 \qquad (1)$$

$$\boldsymbol{F(x) = \exp\{-\exp\left[-\left(\frac{x}{\propto} - \mu\right)\right]\}} \qquad \kappa = 0 \qquad (2)$$

3.2. Log Person type III (LP3) distribution

LP3 distributions uses three parameters: the location μ, the scale α and the shape κ which are based respectively on the mean, variance and skewness of the data (Millington et al., 2011).

The flood quantile (Q_T) for an AEP of 1 in T years is defined from Chow (1951) as:

$$lnQ_T = M + K_T S \qquad (3)$$

where M is the mean of the natural logarithms of AMF data series; S is the standard deviation of the natural logarithms of the AMF data series; and K_T is the frequency factor for the LP3 distribution for AEP of T%, which is a function of the AEP and the skewness of the natural logarithms of the AMF data series.

3.3. Multiple Grubbs-Beck test

Grubbs (1969) and Grubbs and Beck (1972) define a low outlier threshold in GB test as:

$$X_{crit} = \boldsymbol{\hat{\mu} - K_n \hat{\sigma}} \qquad (4)$$

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. Cohn et al. (2013) and Stedinger (1993) provide details for the GB and MGB tests.

4. RESULTS AND DISCUSSION

4.1. Identification of low flows in AMF data

Low flows are censored using MGB test with the LP3 distribution via FLIKE software. Figure 1 illustrates the censoring output of low flows in AMF data in Victoria. The catchments that did not require censoring are shown in yellow shapes, while the remaining shapes show where the other catchments are distributed based on the percentage of low flows that required censoring. 51 out of 183 catchments did not require any censoring and 40 out of 183 catchments required censoring for more than 40% of the AMF data points.

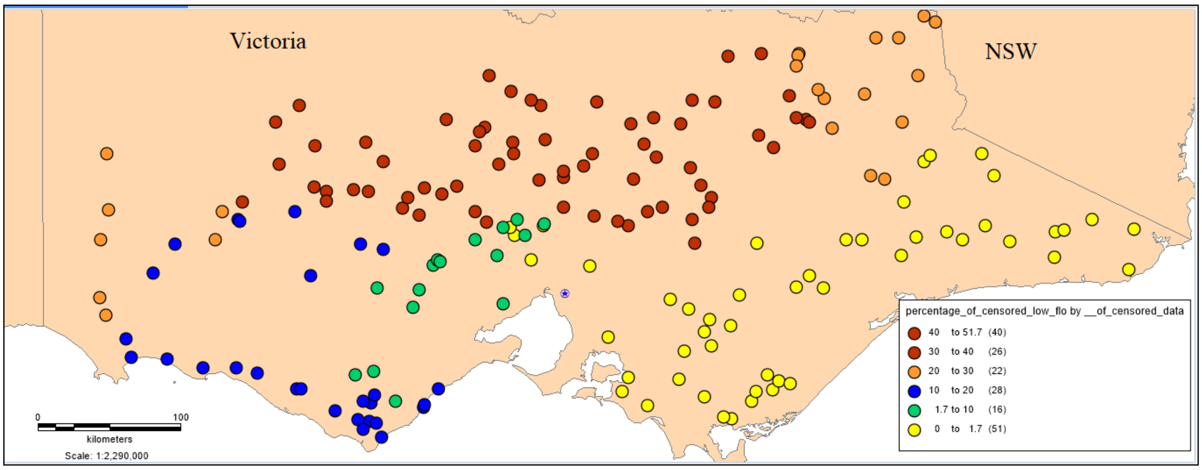

Figure 1. Location of the catchments that required censoring of low flows in AMF data.

Figure 2 illustrates the percentage of catchments that required low-flow censoring. It was found that for 132 stations, MGB test detected 1.2% to 51.7% of the AMF data series as low flows. It is shown that out of 183 catchments in Victoria, 28% of the catchments required no censoring, 11% of the catchments required censoring of low flows between 1% and 10% of the AMF data points, 14% of the catchments required 11% to 20%, 12% of the catchments required 21% to 30%, 15% required 31% to 40%, and 20% required 41% to 51%.

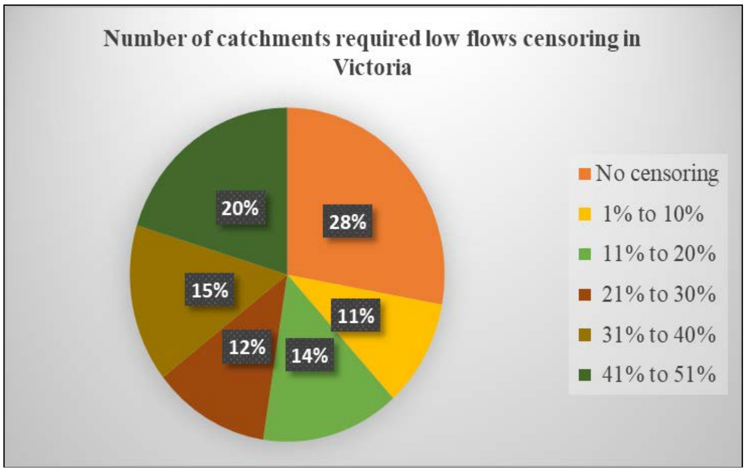

Figure 2. Percentage of catchments required low flow censoring in Victoria.

Figure 3 represents four flood frequency curves for station 226222 in Victoria. The first graph shows the fitting of the GEV distribution using L-moments and without censoring any low flows. The second, third and fourth graphs show the fitting of the LP3 distribution after censoring 0, 6 and 13 low flows respectively. Among the last three, it is obviously clear that fitting of GEV distribution and LP3 distribution after censoring 13 low flows present the best frequency curves. It also shows how censoring low flows affect the fitting of LP3 distribution for Station 22622. The application of MGB test identifies in total 13 low flows, the fitting of LP3 distribution is the best in the last graph (after censoring 13 low flows) than in the second graph (with 0 censored points).

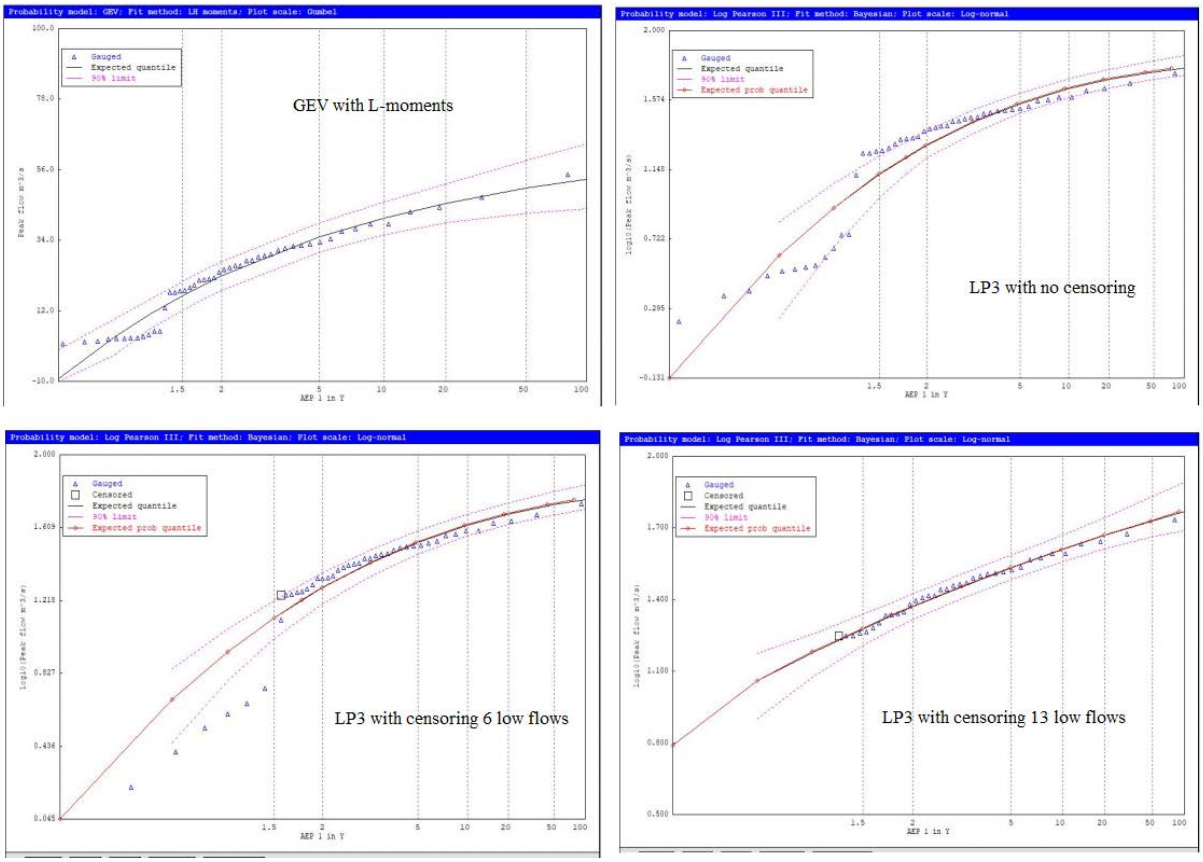

Figure 3. Flood frequency curves for station 226222 using GEV and LP3 distributions

4.2. Comparison of flood quantiles estimates between GEV and LP3

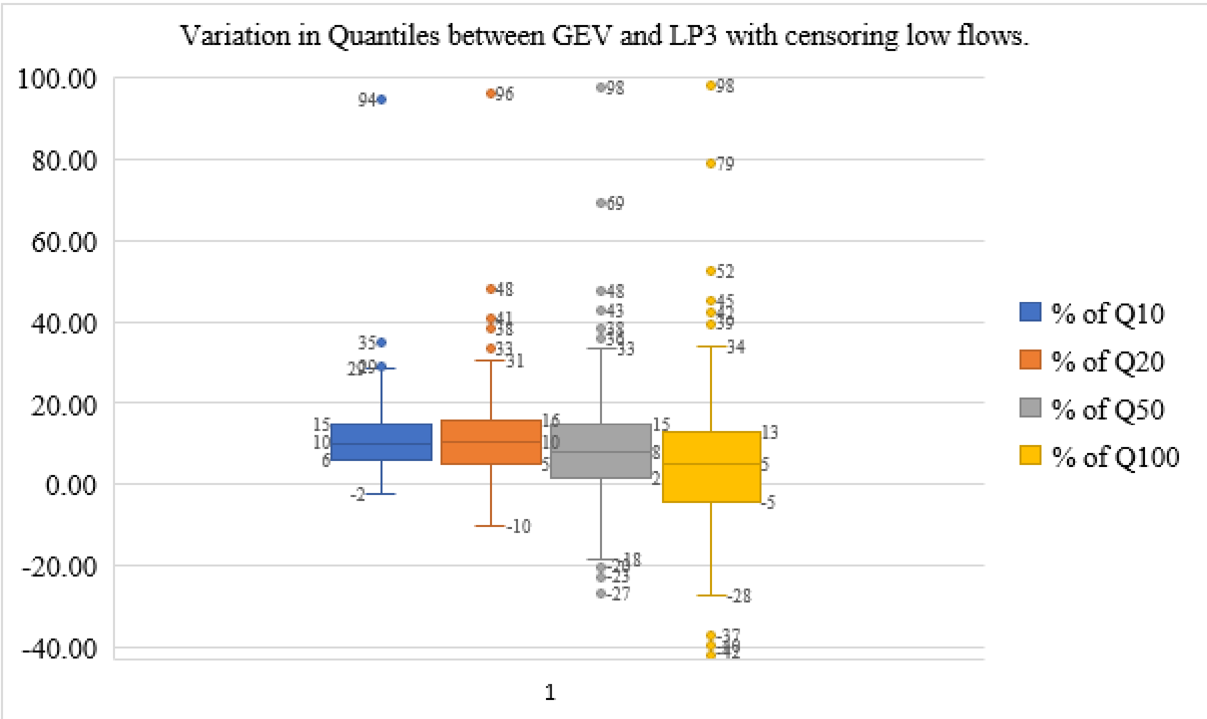

Figure 4. Comparison of flood quantiles by LP3 with censoring and GEV with L-moments.

Figure 4 presents boxplot illustrating the variations of flood quantiles using GEV distribution with L-moments and LP3 distribution with MGB for 1%, 2%, 5% and 10% AEPs. It was found that variation of flood quantiles range between -2% and 94% for Q_{10}, between -10% and 96% for Q_{20}, between -27% and 98% for Q_{50} and between -42% and 98% for Q_{100}. In addition, the plot shows that the majority of the variations falls within the range of 6% to 15% for Q_{10}, 5% to 16% for Q_{20}, 2% to 15% for Q_{50}, and -5% to 13% for Q_{100}. The variation between GEV and LP3 increases with the return period.

4.3. Regression analysis between censoring points and catchments characteristics

A multiple linear regression analysis was conducted to examine if the catchment characteristics, mean annual rainfall (MAR) and shape factor (SF), significantly affect the percentage of censored low flows. Table 2 illustrates the regression outputs between the percentage of censored low flows and the characteristics of the catchment MAR and SF. It is shown that the MAR predictor variable is statistically significant because its p-values is less than the usual significance level of 0.1 while the shape factor is not statistically significant. On the other hand, the value 0.28 of multiple R means that the linear relationship is weak. Consequently, the percentage of the number of censored low flows is found to have a weak correlation with MAR.

Table 2. Linear regression outputs between censored low flows and selected catchments characteristics

Regression Statistics				**Coefficients**	**P-value**
Multiple R	0.2844		Intercept	48.2931	6.68E-10
R Square	0.0809		MAR	-0.0172	0.001131
Adjusted R Square	0.0666		SF	-6.5209	0.311994
Standard Error	14.8444				
Observations	132				

5. CONCLUSION

This paper examined the effects of low flows on FFA using MBG test for 183 catchments in Victoria. In FFA, FLIKE software was used with two common distributions GEV and LP3. It was found that around 28 percent of the 183 catchments did not require any censoring of low flows. The application of MGB has improved the fitting of LP3 distribution where all censored points are applied. In addition, the variation in flood quantiles between GEV (with L-moments) and LP3 (with MGB) increases with the return period. As a result, it is strongly recommended to censor low flows in FFA since these values can affect the quantile estimates. Moreover, a weak relationship was found between the percentage of censored low flows and the catchment characteristics (MAR and SF).

ACKNOWLEDGMENTS

The authors would like to acknowledge ARR Project 5 team for providing the primary AMF data and Victorian Government for providing the updated streamflow data used in this study.

REFERENCES

Ahn, K. H. and Palmer, R. (2016). Regional flood frequency analysis using spatial proximity and basin characteristics: Quantile regression vs. parameter regression technique. Journal of Hydrology, 540, 515-526.

Barnett, V., & Lewis, T. (1994). Outliers in Statistical Data, John Wiley. New York.

Chow, V. T. (1951). A general formula for hydrologic frequency analysis. *Eos, Transactions American Geophysical Union*, *32*(2), 231-237.

Chowdhury, J. U., Stedinger, J. R., & Lu, L. H. (1991). Goodness-of-fit tests for regional generalized extreme value flood distributions. *Water Resources Research*, *27*(7), 1765-1776.

Cunnane, C. (1989). Statistical Distributions for Flood Frequency Analysis. World Meteorological Organisation, *Operational Hydrology Report. No. 33.*

Cohn, T. A., England, J. F., Berenbrock, C. E., Mason, R. R., Stedinger, J. R., & Lamontagne, J. R. (2013). A generalized Grubbs-Beck test statistic for detecting multiple potentially influential low outliers in flood series. *Water Resources Research*, *49*(8), 5047-5058.

England, Jr, J. F., & Cohn, T. A. (2008). Bulletin 17B flood frequency revisions: Practical software and test comparison results. In *World Environmental and Water Resources Congress 2008: Ahupua'A* (pp. 1-11).

Grubbs, F. E. (1969). Procedures for detecting outlying observations in samples. *Technometrics*, *11*(1), 1-21.

Grubbs, F. E., & Beck, G. (1972). Extension of sample sizes and percentage points for significance tests of outlying observations. *Technometrics, 14*(4), 847-854.

Haddad, K., Pirozzi, J., McPherson, G., Zaman, M., Rahman, A. and Kuczera, G. (2009). Regional flood estimation technique for NSW: application of generalised least squares quantile regression technique. In H2009: 32nd Hydrology and Water Resources Symposium, Newcastle: Adapting to Change (p. 829). Engineers Australia.

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, 142-161.

Haddad, K., & Rahman, A. (2019). Development of a Large Flood Regionalisation Model Considering Spatial Dependence—Application to Ungauged Catchments in Australia. *Water*, *11*(4), 677.

Interagency Advisory Committee on Water Data (IAWCD) (1982), Guidelines for Determining Flood Flow Frequency: Bulletin 17-B. technical report, 28 pp., Hydrol. Subcomm, Washington, DC.

Kuczera, G., & Franks, S. (2016). At-site flood frequency analysis. *Australian rainfall and runoff: A guide to flood estimation*, 5-99.

Millington, N., Das, S., & Simonovic, S. P. (2011). The comparison of GEV, log-Pearson type 3 and Gumbel distributions in the Upper Thames River watershed under global climate models.

Rahman, A., Charron, C., Ouarda, T. B., & Chebana, F. (2018). Development of regional flood frequency analysis techniques using generalized additive models for Australia. *Stochastic environmental research and risk assessment*, *32*(1), 123-139.

Rahman, A. S., Rahman, A., Zaman, M. A., Haddad, K., Ahsan, A., & Imteaz, M. (2013). A study on selection of probability distributions for at-site flood frequency analysis in Australia. *Natural hazards*, *69*(3), 1803-1813.

Rahman, A. S., Haddad, K., & Rahman, A. (2014). Impacts of outliers in flood frequency analysis: A case study for Eastern Australia. *J. Hydrol. Environ. Res*, *2*(1), 17-13.

Rahman, A., Haddad, K., & Rahman, A. S. (2015). Australian rainfall and runoff project 5: regional flood methods: database used to develop ARR RFFE technique 2015. *Commonwealth of Australia (Geoscience Australia).*

Stojkovic, M., Prohaska, S., & Zlatanovic, N. (2017). Estimation of flood frequencies from data sets with outliers using mixed distribution functions. *Journal of Applied Statistics*, *44*(11), 2017-2035.

Stedinger, J. R. (1993). Frequency analysis of extreme events. *in Handbook of Hydrology*.

Thompson, W. R. (1935). On a criterion for the rejection of observations and the distribution of the ratio of deviation to sample standard deviation. The Annals of Mathematical Statistics, 6(4), 214-219.