How Significant is the BIAS in Low Flow Quantiles Estimated by L- and LH-Moments?

Hewa, G. A. [1]**, Wang, Q. J.** [2]**, Peel, M. C.** [3]**, McMahon, T. A.** [3] **and Nathan, R. J.** [4]

[1] University of South Australia, Mawson Lakes, SA 5095, Australia
[2] CSIRO (Land and Water), Tatura, Victoria, Australia
[3] The University of Melbourne, Parkville, Victoria, 3010, Australia
[4] Sinclair Knight Merz, Melbourne, Australia
e-mail: guna.hewa@unisa.edu.au

Keywords: *Low flow, frequency analyses, GEV, LH-moment*

EXTENDED ABSTRACT

The relationship between magnitude and the annual recurrence interval of a low flow variable can be expressed using a Low Flow Frequency Curve (LFFC). The goal of frequency analysis is to estimate accurately the quantiles of the distribution of a random variable. As many researchers; for instance Velz and Gannon (1953), Institute of Hydrology (1980), McMahon and Diaz (1982), Tollow (1987), Nathan and McMahon (1990) and Durrans (1996) have noticed, there can be situations where high low flows which occur frequently do not always follow the same trend as the remainder of the observed series. In such situations, a commonly used single distribution cannot provide a good fit over the entire range of the sample. In low flow studies the less frequent flows are of primary interest. Hewa et al (2007) demonstrated how Generalised Extreme Value (GEV) distribution fitted by the LH-moments can be used to avoid the undue influence that more frequent observations can cause on less frequent observations of a minima series. LH-moments (Wang, 1997) are a generalization of L-moments (Hosking 1990). The first three orders of LH-moments at h= 0, 1 and 2 are named as L-moments, L1-moments and L2-moments respectively (Wang, 1997).

The aim of this study was to investigate how significant is the *BIAS* (Equation 1) of low flow quantiles estimated by the GEV/LH-moment at different orders of LH-moments. Two orders of LH-moments; L-moments and L2-moments were selected for the comparison. In this study, annual minimum 7–day flow volume (named as 7-day minima series) series of 84 selected catchments from Victoria were used for the analysis. Two at-site estimates of the 7-day 10-yr low flow quantile were made by fitting the GEV distribution using L-moments and L2-moments respectively. Monte Carlo simulation was used to estimate the average *BIAS* of the quantile estimates.

As can be observed in Figure 1, the 7-day minima series of station 222202 follows a single trend whereas station 228206 exhibits two different trends, one for the high flow portion and another for the low flow portion.

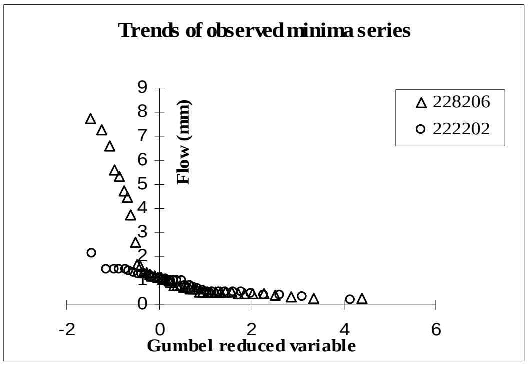

Figure 1. Trends of observed minima series for two sites in Victoria, Australia.

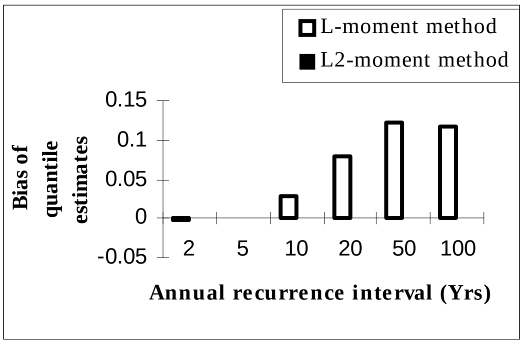

Figure 2. Dimensionless *BIAS* of low flow quantile estimates made by using L-moments and L2-moments for the Station 228206.

It is clear from the results shown in Figure 2 that when low flows and frequent high flows are not derived from a single distribution function, the effect of *BIAS* on L-moment quantile estimates was significantly greater compared to the *BIAS* of L2-moment quantile estimates.

1. INTRODUCTION

A low flow frequency estimate indicates the annual probability of non-exceedance of a low flow event of specified magnitude. The relationship between magnitude and the annual recurrence interval of a low flow variable can be expressed using a Low Flow Frequency Curve (LFFC). The goal of frequency analysis is to estimate accurately the quantiles of the distribution of a random variable.

Fitting a probability distribution function to observed data provides a compact and smoothed representation of the frequency distribution revealed by the available data, and leads to a systematic procedure for extrapolation to frequencies beyond the range of the data set (Stedinger et al., 1992). The uncertainties in the extrapolation are significant and depend on how well the data constrain the possible formulation to the adopted distribution function. Hence, parameters for the selected distribution function need to be estimated using an appropriate method so that extrapolation of required quantiles and expectations are reliable.

The usual practice of estimating parameters is to equate estimates of the first two or three sample moments (central tendency, spread and symmetry) to the theoretical moments of the distribution function. The most efficient estimate of a parameter is defined as the estimate which has the minimum variance. There are a number of techniques that can be used for estimating the parameters of a distribution function, including the Method of Moment (MOM), the method of Maximum Likelihood (ML) and the method of Probability Weighted Moments (PWM). These methods have been widely used and their performance compared (Fischer, 1929; Matalas, 1963; Wallis et al., 1974; Vogel and Fennessey, 1993; Nathan and McMahon, 1990 and Wang, 1996)

A more recent and superior method, based on a linear combination of order statistics, known as L-moments (Hosking, 1990) is now available. The L-moments method is nearly unbiased relative to the other estimation methods. Wang (1997) introduced LH-moments, which is a generalization of L-moments for estimating extreme floods. LH-moments was proved to be superior to other estimation methods in avoiding undue influence that more frequent observations can cause on less frequent observations. Recently, Hewa et al. (2007) demonstrated how the LH-moments can be used in low flow frequency analyses.

The aim of this study is to investigate how significant is the BIAS of low flow quantiles estimated by L-moments and L2-moments. In this study, the GEV distribution is used to estimate low flow quantiles for 84 selected catchments from Victoria and the BIAS of the estimated quantiles by the two methods (L-moment and L2-moments) is quantified through Monte Carlo simulation.

1.1. Fitting a distribution function to a minima series

There are a number of probability distributions that have been suggested as being suitable for modelling low flows. The most commonly used distribution functions in low flow frequency analysis are the Normal, Log-Normal, Gamma, Log-Pearson Type III, Extreme Value Type I (Gumbel) and Extreme Value Type III (Weibull). Matalas (1963) and Eratakulan (1970) found the Weibull and Pearson Type III alternatives perform well over a range of skewness values, although O'Conner (1964) and Vogel and Kroll (1989) indicated preferences for the Log-Normal choice for their study area. Nathan and McMahon (1990) and Durrans (1996) used Weibull models for censored low flow data. In fitting a distribution to the annual minima series it is assumed that the sample considered represents the low flow extremes of observed streamflows, and thus the usual practice is to fit a suitable probability distribution function to all the observed minima of the required duration.

Nevertheless as can be seen in Figure 1, high low flows which occur frequently are not always derived from the same probability distribution as the remaining observations. This has been observed by many researchers; for instance Velz and Gannon (1953), Institute of Hydrology (1980), McMahon and Diaz (1982), Tollow (1987), Nathan and McMahon (1990) and Durrans (1996). According to McMahon and Diaz (1982), the break in the low flow frequency curve shows the point where the higher frequency flows are no longer drought flows but rather are flows approaching normal conditions.

In low flow studies the less frequent flows are of primary interest. When model parameter estimation is performed on the basis of the entire data sample, it can lead to a deterioration of the statistical properties (Durrans, 1996). When the largest low flow appears to deviate appreciably from the second largest low flow, it results in a large skewness, which in turn, can lead to a theoretical distribution underestimating the severity of the extreme droughts (Matalas, 1963). Nathan and McMahon (1990) indicated that certain low flow quantile estimates were usually affected by up to 22% when large data values were retained in the analysis. Similar observations have been made in the context of flood frequency analysis, where the presence of low flow data values can exhibit adverse effects (Klemes, 1986). Hence, estimation should be concentrated on the less frequent observations of the sample. The most common practice in dealing with this is to censor the data series at the break point and fit a distribution function to the remainder.

Many researchers, for instance Gilliom and Helsel (1986), Helsel and Gilliom (1986) and Kroll and Stedinger (1996) investigated various estimation methods such as Log Probability Plot Regression (LPPR), Log-Normal Maximum Likelihood (LN-ML) and Log-Normal Partial Probability Weighted Moments (LN-PPWM) for censored data samples. Recently, Hewa et al. (2007) demonstrated the capability of the Generalised Extreme Value distribution, fitted by LH-moments, for giving more emphasis to the less frequent observation, without practically censoring the data sample.

1.2. GEV distribution and LH-moments

According to Hosking (1990) and Wang (1997), L-moments and LH-moments are linear functions of PWMs and LH-moments are equivalent to higher order PWMs. Detailed discussion of the GEV distribution function, its parameters as well as estimation using LH-moments of a minima series through PWMs are presented in Hewa et al. (2007).

2. STUDY AREA AND DATA

The catchments selected for this study are from Victoria, Australia. They are within two drainage divisions: South East Coast and Murray-Darling. A total of 84 catchments from 30 drainage basins in these two divisions are used. The geographical distribution of the selected catchments and the main criteria used to select those catchments from the drainage divisions are given in Hewa et al. (2007).

3. METHODOLOGY

The *BIAS* of a low flow quantile estimate ($\hat{\theta}$) in Equation 1 is a measure of the average error in the estimate from the population value and is used to compare the reliability of quantile estimates made by the two methods.

$$BIAS\left(\hat{\theta}\right) = E\left[\hat{\theta} - \theta\right] \qquad (1)$$

where θ is the population value of $\hat{\theta}$

In this study, the GEV is assumed to be the parent distribution. For each of the 84 selected catchments, the average *BIAS* of a range of quantiles is estimated from 10,000 simulated samples by using L-moments and L2-moments. The steps in the Monte Carlo Simulation are as follows:

1. Fit the GEV to the observed minima series using L2-moments and estimate the low flow quantiles for the Average Recurrence Intervals of interest (ARIs) of interest.
2. Assume that the estimated GEV parameters are the population values.
3. Using a random number generator, generate a GEV distributed sample of length n, where n is the length of the observed minima series in step 1, using the GEV population parameters of step 2.
4. Fit the GEV distribution to the generated sample using L2-moments.
5. Estimate low flow quantiles for the same ARI's of step 1.
6. Repeat steps 3 to 5 for N times, where N is number of simulations (in this paper N =10,000).
7. Estimate the *BIAS* of the low flow quantiles, by taking the quantile estimates in step1 as the population values.
8. Repeat the procedure from step 1 to step 7 for L-moments

4. RESULTS AND DISCUSSION

Absolute *BIAS* is expressed as Ml/km^2 (= mm). Figure 3 presents the number of catchments at which the *BIAS* of a quantile estimate for one method is smaller than that of the other estimation method.

It can be observed in Figure 3 that the larger the annual recurrence interval the greater the number of catchments for which absolute *BIAS* of L2-moments estimates is smaller than that of the L-moments estimates. This suggests that, on average, L2-moments estimates of low flow quantiles (7-day series) at large annual recurrence intervals are more reliable than that of the L-moments estimates.

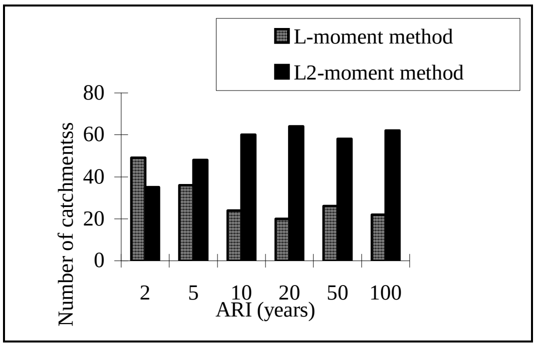

Figure 3. Comparing the number of catchments with smaller absolute *BIAS* of quantiles estimated via L-moments method and L2-moments method

Having noticed that the majority of the L2-moments estimates are with less *BIAS* than L-moments estimates, it was decided to investigate how the *BIAS* varies at varying ARIs

Figures 2 and 4 compare the dimensionless *BIAS* (*BIAS* divided by the mean daily flow of the station) estimates of the low flow quantiles made using the L-moment and L2-moment methods for two sample stations 228206 and 222202 presented in Figure 1. The difference is greater for Station 228206 in which more frequent and less frequent observations had two different trends. These observations suggests that, L2-moments is capable of giving more emphasis towards the less frequent observations when extreme low flows are derived from a different distribution to that of the more frequent observations.

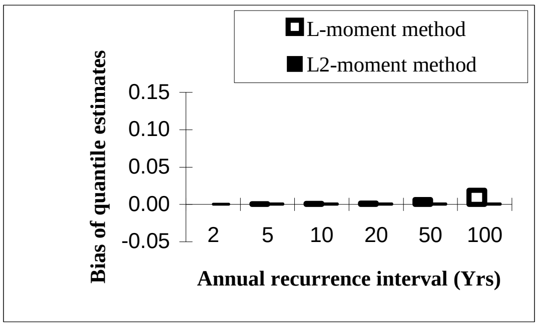

Figure 4. Dimensionless BIAS of low flow quantile estimates made by using L-moments and L2-moments for the Station 222202

Though, it was evident that L2-moment estimates in general have smaller *BIAS* when compared to L-moment estimates, it was interesting to see how significant the effect of this *BIAS* on mean quantile estimates. Therefore, how significant is the effect of absolute *BIAS* on the low flow quantile is further investigated by using the Station 228206 in whish observed minima series had more than one trend. The results are presented in Figure 5 and 6.

QARI is the true low flow quantile at the average recurrence interval (ARI) of interest while Qm is the mean quantile estimate at the same ARI.

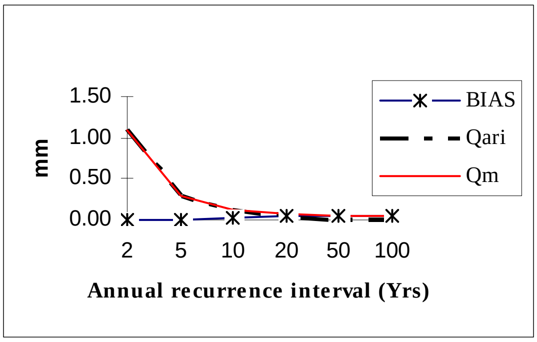

Figure 5. Effect of *BIAS* on mean quantile estimates made by the L-moments method for the Station 228206

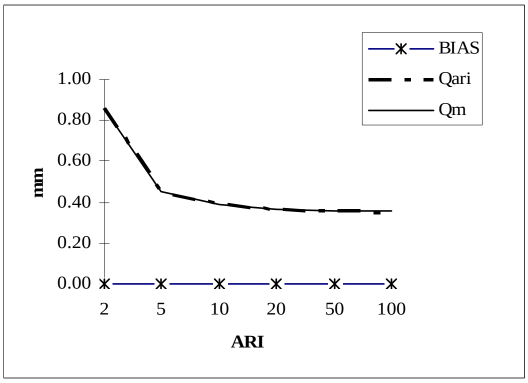

Figure 6. Effect of *BIAS* on mean quantile estimates made by the L2-moments method for the Station 228206

It can be observed from Figures 5 and 6 that the effect of *BIAS* on Q_m of the L2-moment method is negligible, while that of the L-moment method makes a difference at large annual recurrence intervals. Hence, when more frequent observations are not derived from the same distribution as the less frequent observations, the L2-moment method is capable of reducing undue influence of more frequent events in estimating low flow quantiles at large annual recurrence intervals. Consequently, L2-moment estimates of low flow quantiles at large annual recurrence intervals are more reliable than the L-moments estimates.

5. CONCLUSIONS

BIAS of L-moment estimates in general is greater than that of L2-moment estimates. When the more frequent and less frequent observations follow two different trends, the *BIAS* of L-moment estimates is significantly greater at high ARIs as opposed to that of L2-moment estimates. The ability of LH-moments to give more emphasis to less frequent observations in minima series helped to make more reliable low flow quantiles at higher average recurrence intervals. The benefit of LH-moments is greater when the less frequent and high frequent low flows follow different trends (eg. Station 228206 in Figure 1).

6. REFERENCES

Durrans, S. R. (1996), Low-Flow analysis with a conditional Weibull tail model, *Water Resources Research*, 32(6): 1749-1760.

Eratakulan, S. J. (1970), Probability distribution of annual droughts, *Journal of Irrigation and Drainage Engineering*, ASCE (96): 461-474.

Fischer, R. A. (1929), Moments and product moments of sampling distributions, Royal Society of London, Series A, Containing papers of a mathematical and physical character, 2(30): 199.

Gilliom, R. J. and Helsel, D. R. (1986), Estimation of distributional parameters for censored trace level water quality data, 1, Estimation techniques, *Water Resources Research* 22(2), 135-146.

Helsel, D. R. and Gilliom, R. J (1986), Estimation of distributional parameters for censored trace level, Water quality data, 2, Verification and applications, *Water Resources Research*, 22(2) 146-155.

Hewa, G. A., Wang, Q. J., McMahon, T. A. Nathan, R. J. and Peel, M. C. (2007), Suitability of the Generalized Extreme Value (GEV) distribution fitted by LH-moments for low flow frequency analysis, *Water Resources Research*, 43 (6), W06301, doi: 10.1029.

Hosking, J. R. M. (1990), L-moments: Analysis and estimation of distributions using linear combinations of order statistics, *Journal of the Royal Statistical Society*, Series B 52: 105-124.

Institute of Hydrology (1980), Low Flow studies: (1) Research report, (2) Flow duration curve estimation; (3) Flow frequency curve estimation; (4) Catchment Characteristics Estimation Manual, Volumes 1, 2, 3 and 4. Institution of Hydrology, Wallingford.

Klemes, V. (1986), Dilettantism in hydrology: Transition or Destiny?. *Water Resources Research* 22(1): 177-188.

Kroll, C. N. and Stedinger, J. R. (1998), Regional hydrologic analysis: Ordinary and Generalized Least Squares revisited. *Water Resources Research* 34(1): 121-128.

Kroll, C. N. and Stedinger, J. R. (1999), Development of regional regression relationships with censored data. *Water Resources Research* 35(3): 775-784.

Matalas, N. C. (1963), Probability distributions of low flows. USGS Professional Papers 434-A, 27pp.

McMahon, T. A. and Diaz, A. A. (1982), Method of computation of low streamflow, No.36. UNESCO, Paris, 95pp.

Nathan, R. J. and McMahon, T. A. (1990), Practical aspects of low flow hydrology, *Water Resources Research*, 26(9): 2135-2141.

O'Conner, D. J. (1964), Comparison of probability distributions in the analysis of drought flows, Water Sewage Works (111): 180pp.

Stedinger, J. R., Vogel, R. M. and Foufoula-Georgiou, E. (1992), Frequency analysis of extreme events. Handbook of Hydrology, D. R. Maidment. McGraw Hill, New York: 18.1-18.58.

Tollow, A. J. (1987), Methods used in assessing low flow probabilities and their application. National Hydrology symposium, British Hydrological society: 15.1-15.5.

Velz, C. J. and Gannon, J. J., Eds. (1953), Low Flow Characteristics of Streams. Ohio State University Engineering Series. Ohio State University, 138-154.

Vogel, R. M. and Fennessey, N. M. (1993), L-Moments Diagram Should Replace Product Moment Diagrams. *Water Resources Research*, 29(6): 1745-1752.

Vogel, R. M. and Kroll, C. N. (1989), Low Flow Frequency analysis using probability plot correlation coefficients. *Water Resources Planning and Management*, (115): 338-357.

Wang, Q. J. (1996), Using partial probability weighted moments to fit the extreme value distributions to censored samples, *Water Resources Research*, 32(6): 1767-1771.

Wang, Q. J. (1997), LH-moments for statistical analysis of extreme events, *Water Resources Research*, 33(12): 2841-2848.