National Estimates of Rare Design Rainfall

Janice Green
IFD Revision Project Manager, Environment and Research Division, Bureau of Meteorology, Sydney, Australia

Catherine Beesley
Hydrologist, Environment and Research Division, Bureau of Meteorology, Sydney, Australia

Andrew Frost
Hydrologist, Environment and Research Division, Bureau of Meteorology, Sydney, Australia

Scott Podger
Hydrologist, Environment and Research Division, Bureau of Meteorology, Sydney, Australia

Cynthia The
Hydrologist, Environment and Research Division, Bureau of Meteorology, Sydney, Australia

Abstract

In 2013 the Bureau of Meteorology released new Intensity-Frequency-Duration (IFD) design rainfalls as part of the broader revision of Australian Rainfall and Runoff (ARR). In addition to point IFDs for probabilities from 1 Exceedance per Year (EY) to 1% Annual Exceedance Probability (AEP), a number of enhancements have been made to the new IFDs including the incorporation of lower probability design rainfall estimates. The Bureau's new IFD website provides practitioners with a centralised location for easy access to improved, consistent point design rainfall estimates for probabilities from 12EY to 0.05% AEP across a range of durations.

Design rainfall depths less frequent than 1% AEP are typically required for the design of bridges, small dams and other critical infrastructure. With the exception of most of the Northern Territory, gridded 2% to 0.05% AEP design rainfalls have previously been generated across Australia using a method developed by the Cooperative Research Centre for Catchment Hydrology (CRCCH) in the 1990's. The Cooperative Research Centre – Focussed Rainfall Growth Estimation (CRC-FORGE) method was applied on a state-by-state basis and as such there are variations in the approach used between regions. Practitioners have previously had to approach individual state agencies for access to the data. Given the differences between the state methods and the availability of an extensive, quality-controlled rainfall database established during the IFD revision project, there was an opportunity to review methods and achieve national consistency. Consideration was given to three potential approaches based on regional frequency analysis which use pooling of standardised annual maxima series based on spatial or climatically similar regions to reduce at site uncertainty. While the CRC-FORGE method has been demonstrated to provide reasonable rare design rainfall estimates, this paper recommends the adoption of a regional LH-moment approach as being more appropriate and consistent with the approach adopted for the new IFDs.

1. INTRODUCTION

Rainfall estimates in the large to rare range are defined in Book VI of Australian Rainfall and Runoff (Nathan and Weinmann, 1999) as 2% annual exceedance probability (AEP) to the credible extrapolation limit (around 0.05% AEP). Design rainfall depths less frequent than 1% AEP are typically required for the design of bridges, small dams and other critical infrastructure. There can be significant costs associated with building, upgrading or with the failure of such infrastructure and therefore reliable design estimates are required for risk and economic analyses.

As part of Phase 2 of the IFD revision project, nationally consistent rare design rainfalls have been derived and will be made freely accessible to users via the Bureau of Meteorology website (http://www.bom.gov.au/water/designRainfalls/ifd/) for durations of 1 to 7 days.

Prior to this project, Irish (1994) was commissioned to provide recommendations for a method to improve the extrapolation of rainfall frequency curves for estimation of rare design rainfalls. Under that review, the author suggested that the approach of Reed and Stewart (1989) or Schaefer (1990) may be feasible options for regional frequency analyses of rainfall in Australia.

Following those studies, rainfall estimates in the large to rare range were generated for each state, excluding much of the Northern Territory, using a method developed by the Cooperative Research Centre for Catchment Hydrology (CRCCH) further discussed below. As the method was applied on a state-by-state basis, there are some variations in the approach adopted for each region. Differences in the cut-off period for data due to the staggered timing of the different studies, the base used for the 2% AEP estimates (that is, whether anchored to the 1987 IFD estimates using other procedures or not), gridding settings and smoothing processes created inconsistencies which are particularly apparent in overlapping state border areas. Also, currently there is no central point of dissemination and it is necessary to contact the individual or organisation responsible for preparing each of the state CRC-FORGE estimates for access.

The development of nationally consistent rare design rainfalls provided the opportunity to reassess the CRC-FORGE method and to review work undertaken by Wallis et al (2007) on a modification of the original approach developed Schaefer (1990) (hereafter referred to as the modified Schaefer method) as well as the possible adaptation of the method used to derive the new IFDs (Green et al., 2012) using regional LH-moments. Therefore, the three candidate methods for deriving the revised rare design rainfalls were:

- CRC-FORGE
- Modified Schaefer
- Regional LH-moments adaptation of the new IFDs

This paper describes these methods and discusses the selection and application of the regional LH-moments approach to derive the rare design rainfall estimates.

2. REGIONAL FREQUENCY ANALYSES

Regional frequency analyses are commonly used to provide design rainfall estimates for particular probabilities as the use of information from multiple sites improves accuracy and consistency, provided inter-site correlation is taken into account (Nandakumar et al, 1997). Regional approaches can either involve averaging site values within homogeneous regions (Schaefer, 1990) or use of a station-year method which effectively pools data from several sites to create a longer record (Reed and Stewart, 1989; Nandakumar et al, 1997). The majority of methods require identification of homogenous regions (by cluster analysis or less formal approaches) for site grouping (Kjeldsen & Jones 2009), while other analyses use a Region of Influence approach (ROI; Burn 1990), such that for each station an individual homogenous region is defined. Each ROI will contain a potentially unique set of sites, with each site possibly contributing to multiple ROIs.

For each homogeneous region it is assumed that those sites have the same frequency distribution or that any differences are a result of sampling issues. For the CRC-FORGE method, these areas were mostly defined by state boundaries. The Schaefer approach obviates the need for defining spatially contiguous homogeneous regions by finding physical site characteristics such as mean annual rainfall that result in sites with similar characteristics having similar frequency distributions. The new IFD method uses a ROI approach with weighting of at-site estimates according to record length (Johnson et al, 2012).

2.1. CRC-FORGE Method

In the mid 1990’s, the Cooperative Research Centre for Catchment Hydrology developed the CRC-FORGE method based on the FORGE concept devised by Reed and Stewart (1989). The approach is a station-year method in that site data surrounding a focus station within a homogeneous region can be progressively pooled into one record and fitted with a non-parametric frequency curve. The method assumes that the extreme rainfall distribution at each site in a homogeneous region is the same once the data has been standardised by an index variable which is the at-site annual maximum rainfall mean in the case of CRC-FORGE. To address the additional assumption that each site is independent, a spatial dependence model is used to derive the effective number of independent stations (Ne) which is then used to derive the plotting position of the pooled rainfall data. The GEV distribution was used with L-moments for parameter estimation although the growth curves were fitted empirically. In order to maintain consistency with the ARR87 IFDs, the growth curves were anchored to the 1% AEP ARR87 IFD estimates.

The spatial dependence model is derived using synthetically generated data and is an inherent source of uncertainty which consequently increases the uncertainty in the growth curve since it is used to determine how the regional rainfall data is plotted. Nandakumar et al (1997) investigated the use of a constant Ne model and an Ne model that varied with AEP and suggested the latter performed better. The overall method is described in detail in Nandakumar et al (1997). The method was consecutively applied on a state by state basis across the country with each state report reference provided in Table 1.

Table 1. State CRC-FORGE estimate references

State	Reference
New South Wales / ACT	Nandakumar et al, 2012
Queensland	Hargraves, 2004
South Australia	Hill et al, 2000
Tasmania	Gamble et al, 1998
Victoria	Nandakumar et al, 1997
Western Australia	Durrant et al, 2004

The key assumptions of the CRC-FORGE method are that:

- The region is homogeneous: Pooling data collectively over large regions as is done here, when site data is not homogeneous, would have the effect of biasing the growth curve at different locations.
- The method for estimating spatial dependence is correct: as this determines the plotting position (and AEP) of points on the growth. Any error in the empirical method (based on synthetic studies) used to derive this value could cause large biases.

2.2. Modified Schaefer method

An alternative method of calculating rare design rainfalls is to use an approach similar to that applied in Schaefer (1990), and more recently within Wallis et al (2007). The method assumes a certain probability distribution describes the data and a statistical method can be used to estimate the distribution parameters and is analogous to regional L-moments (Hosking and Wallis, 1990).

While undertaking a regional rainfall frequency analysis for the Washington area, Schaefer found difficulty forming homogenous regions due to significantly varying topography and climate across the study area. Schaefer discovered that by treating the study area as a heterogeneous ‘superregion’, he could group smaller numbers of sites together based on site characteristics rather than proximity to form homogeneous subregions which avoids the problematic requirement of spatially contiguous regions. He found that the coefficient of variation and skewness had a relatively constant relationship with mean annual rainfall (MAR). By plotting mean MAR against regional weighted average moments for each of the sub-regions and determining a relationship using a line of best fit, GEV distribution parameters could be derived for any location in the study area based on MAR. Quantile estimates can then be calculated through scaling by the index rainfall. Schaefer determined the AEP to which the curves could reasonably be extrapolated by using an equivalent independent record length (EIRL) method which takes into consideration the reduced total record in the dataset due to spatial dependence. This method is described in more detail in Schaefer (1990).

Following the recommendations from Irish (1994), McConachy (1996) investigated the method and its application to Victorian 24 hour rainfall data. Rather than probability weighted moments, McConachy used L-moments which are linear combinations of ranked observations developed by Hosking and Wallis (1990), commonly used since the 1990s for estimating distribution parameters due to low bias and sampling variance. This approach was also adopted later by Wallis et al (2007) in an extension of Schaefer's previous work. While having the advantage of being robust to outliers, this can also be an issue for very rare rainfall events as they have less influence on the tail of the distribution (Nandakumar et al, 1997). The method applied assumes a GEV distribution describes the data, and L-moments (Hosking and Wallis, 1990) are used to estimate the distribution parameters at each site. Following Schaefer (1990), these L-moment parameters are then averaged over a sub-region, and related to physical/climate characteristics (e.g. mean annual rainfall). This has the benefit that the requirement for homogeneity across the whole region is not required.

The key assumptions of the Schaefer method are:

- Smoothing: That the averaging of distributional parameters across a subregion does not smooth the at-site results too much, but rather adds information on extremes by space-for-time substitution. If intersite correlation is not taken into consideration, underestimation of uncertainty will result.
- Homogeneous subregions: That the sub-region distributions are homogeneous
- Regression relationships accurate: Relationships exist between hydro-climatic data distributional parameters that can be used to predict the L-moment parameters spatially.

2.3. Regional LH-moments

In 2013 the Bureau of Meteorology released new IFD design rainfalls for probabilities of 1 EY to 1% AEP (Green et al, 2012). Annual maxima series (AMS) from all sites within the Bureau of Meteorology's Australian Data Archive for Meteorology (ADAM: containing approximately 20 000 daily read rainfall stations starting in 1800) along with other sites available through the terms of the Water Regulations 2008, were used in deriving the new IFD estimates. Daily read stations with greater than 30 AMS years were used. Linear combinations of the order statistics (L-moments: Hosking & Wallis 2005) were used to characterise the distributional properties based on at-site data. This approach has been shown to have low bias and be robust to outliers. The ROI regionalisation approach (Burn 1990; Johnson et al., 2012a) has been used, which pools data from surrounding sites until 500 station years of record are used, to reduce uncertainty in the estimated L-moments. This approach, and example of an index flood procedure, assumes that all sites within the region of influence are homogeneous such that all sites in the region have the same probability distribution, other than a scaling factor. Estimates were derived for ungauged locations by spatial interpolation using ANUSPLIN (Hutchinson, 2007) with three independent variables: latitude, longitude and elevation. (The et al., 2012; 2014).

In order to extrapolate this method reliably beyond 1% AEP, some modifications to the method are needed. The proposed adaptations are:

- Restricting the analysis to sites with a long AMS: site distributions are considered more reliable as they are based on a significant amount of data and a longer sample of the climate record.
- Using LH-moments: LH-moments allow the distribution to increasingly focus on the tail of the distribution which is considered appropriate for rare design rainfall estimation.
- Increasing the ROI: the index rainfall method relies on a sufficient pooling of records for extrapolation to lower probabilities.
- Using regression to estimate at ungauged locations: a regression based approach using various site descriptor variables to improve accuracy over spline interpolation.

The key assumptions of the new IFD method are:

- Homogeneous regions: the selected of sites within the region of influence are homogeneous.
- The method modifications are sufficient for extrapolating out to 0.05% AEP.

2.4. Selected method

It is difficult to assess the veracity of any method adopted for extending design rainfalls beyond the 1% AEP as there is insufficient length of observed rainfall data to validate the results. Therefore, the selection of the most appropriate method needs to be based on a qualitative assessment that the assumptions made in the method are reasonable and that the adopted approach is consistent with methods used to derive more frequent design rainfalls where the results can be validated.

The regional LH-moments adaptation of the new IFDs was selected for deriving the new design rainfalls for the following reasons:

- The assumptions made in the approach are reasonable
- It is consistency with not only the data but also the method adopted for the new IFDs.
- It has the ability to leverage off work undertaken for the new IFDs
- It incorporates of new statistical and gridding techniques that have been developed in recent years.
- It builds on work undertaken recently overseas into the estimation of rare design rainfalls.

3. APPLICATION OF THE REGIONAL LH-MOMENTS METHOD

3.1. Data

As this analysis builds on the work undertaken by the Bureau of Meteorology to derive the new IFD design rainfalls, a quality-controlled database of annual maximum series (AMS) was available for all daily rainfall gauges with more than 30 years of record up to December 2012 (Green et al., 2012). The restricted 9am to 9am totals recorded at these rainfall gauges had previously been converted to unrestricted totals using the factors 1.15, 1.11, 1.07, 1.05, 1.04, 1.03, 1.02 for 1, 2, 3, 4, 5, 6 and 7 day durations respectively.

As the data are used to estimate rare events, long station records provide a larger sample and, therefore, a more reliable regional frequency analysis. However, the spatial coverage of daily gauges with long records varies considerably across Australia. Figure 1 shows the coverage of sites with a) 100 years and b) 60 years of record.

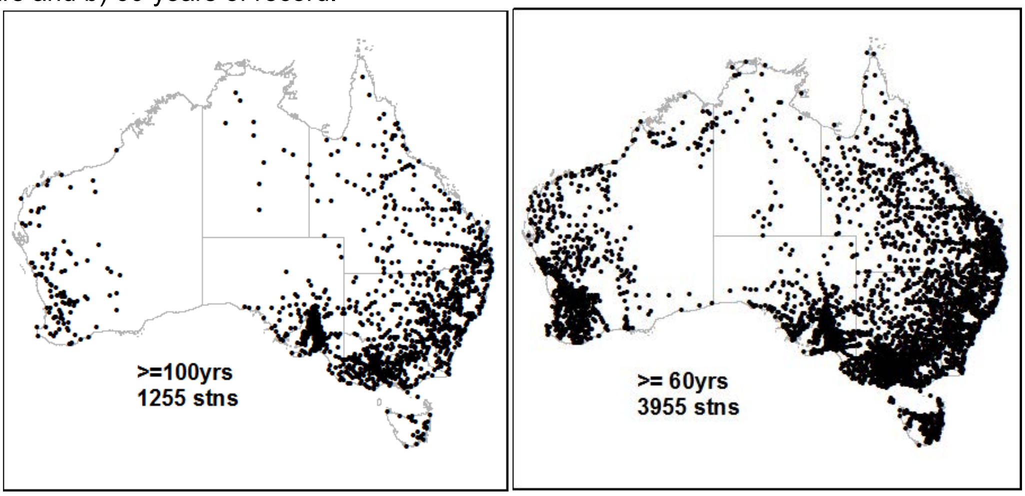

Figure 1 Coverage of daily rainfall gauges for more than a) 100 years and b) 60 years of record.

As a compromise between more accurate estimation by using longer records and having adequate spatial coverage, a minimum AMS of 60 years has been adopted with station weighting ensuring that longer records are incorporated preferentially where available.

3.2. Stationarity

A requirement of a regional frequency analysis is that the data must be homogeneous with respect to space and time. Extensive stationarity testing was carried out on the AMS as part of the IFD revision project (Green and Johnson, 2011). This involved assessing the AMS and fitted distributions for individual sites as well as regional averages for rainfall trends over time for 6 minute to 72 hour duration events. This testing found that while a number of sites failed one or more of the test, there was no clear spatial pattern associated with these sites. The recommendation from the results of this study was to use the entire length of record for generating the new IFD estimates. Since the data set used for the rare design rainfall analysis is a subset of the data assessed by Green and Johnson (2011), the assumption of stationarity based on that work was adopted.

3.3. LH-moments

In order to more accurately fit the lower frequency, upper tail of the distribution, it was decided that LH-moments introduced by Wang (1997) be used. LH-moments are a generalisation of L-moments and allow the distribution to be increasingly focused on the larger data values depending on the value of η, where $\eta=0$ is equivalent to L-moments. The equations for deriving LH-moments and the associated GEV parameters are given in Wang (1998). LH-moments with a $\eta=2$ were selected as a compromise between providing a better fit to the tail of the site distributions without giving too much influence to the high outliers. LH-moments ($\eta=2$) were derived for all stations with greater than 60 years AMS.

3.4. Distribution

In recent years, the 3 parameter Generalised Extreme Value (GEV) distribution has been found in many studies to most adequately describe the AMS data and is often adopted for regional frequency analyses (Green et al, 2012; Irish, 1994; McConarchy, 1996; Nandakumar et al, 1997; Schaefer, 1990). Based on the findings from the comparison of distributions undertaken for the new IFDs and given that the rare design rainfall analysis is based on a subset of those data, a GEV distribution has been shown to be appropriate.

3.5. Regionalisation

For the rare design rainfalls, the ROI approach adopted for the IFDs was used to reduce the uncertainty in the estimated LH-moments by regionalising the station point estimates. However, where for the IFDs 500 station years was found to be an optimum pool size, as the rare design rainfalls are provided for probabilities out to 1 in 2000 AEP, the ROI needed to be increased. The tradeoff between gaining improved accuracy from a larger pool of data was that the assumption of homogeneity may not be satisfied. Based on testing for the pool size that reduced uncertainty without introducing significant homogeneity, a minimum of 2000 station years was adopted. However, where necessary, the number of pooled station years was increased above this number to maximize the available record used, while ensuring homogeneity.

The average LH-CV for each region was calculated using a weighted average of the LH-CV at all stations in the region, with the weights proportional to the station lengths. This was repeated for the LH-Skewness.

3.6. Regression and quantile estimation

In light of the rarer probabilities being estimated; the smaller number of rainfalls stations being used; and the resultant decrease in station density, a regression based approached was adopted for deriving GEV parameters at ungauged locations, rather than the gridding undertaken for the IFDs.

The regression was based on the BGLSR approach used with the frequent and infrequent design rainfalls to derive sub-daily rainfall statistics from daily rainfall statistics. However, for the rare design rainfalls the BGLSR was used to define the relationship between regionalised LH-CV and LH-skew and hydro-climatic predictors. To provide spatial continuity, regression relationships were defined using a ROI approach such that each station had its own ROI. Regression coefficients were interpolated for predictands at ungauged sites to provide values for each 2.5km grid point.

Quantiles were estimated for each grid using the GEV parameters. These were then scaled to ensure consistency between the infrequent and the rare design rainfalls. The depths were smoothed across durations and AEPs to ensure consistency.

4. DISSEMINATION

The rare design rainfall depths are incorporated into the Bureau of Meteorology's new IFD website (http://www.bom.gov.au/water/designRainfalls/ifd/). By entering the co-ordinate of a location of interest, users can obtain design rainfall estimates from 12EY to 0.05% AEP for durations from 24 to 168 hours for any point in Australia in a single query.

5. ACKNOWLEDGEMENTS

The authors acknowledge the technical contributions of Rory Nathan, Erwin Weinmann and Nanda Nandakumar and George Kuczera.

6. REFERENCES

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

Durrant, J.M. and Bowman, S. (2004), Estimation of Rare Design Rainfalls for Western Australia: Application of the CRC-FORGE Method, Department of Environment, Government of Western Australia, Surface Water Hydrology Report Series Report No. HY17.

Gamble, S.K., Turner, K.J., Smythe, C.J. (1998), Application of the Focussed Rainfall Growth Estimation Technique in Tasmania, Hydro Tasmania Internal Report Number EN-0004-010213-CR-01

Green, J. and Johnson, F.J. (2011), *Stationarity Assessment of Australian Rainfall*, Bureau of Meteorology Report.

Green, J., Xuereb, K., Johnson, F., Moore, G. and The, C. (2012), *The Revised Intensity-Frequency-Duration (IFD) Design Rainfall Estimates for Australia – An Overview*, 34th Hydrology & Water Resources Symposium. 19-22 November 2012, Engineers Australia, Sydney, Australia, pp 808-815.

Hargraves, G., (2004), CRC FORGE and (CRC) ARF techniques development and application to Queensland and Border Locations. Extreme Rainfall Estimation Project, Water Assessment and Planning Resource Sciences Centre, Queensland.

Hill, P., Nathan, R., Rahman, A., Lee, B., Crowe, P., Weinmann, E. (2000), Estimation of Extreme Design Rainfalls for South Australia using the CRC-FORGE Method. Proc. Hydro 2000 3rd International Hydrology and Water Resources Symposium, Perth.

Hosking, J.R.M. and Wallis, J.R. (1990), *Regional frequency analysis using L-moments*. (RC 15658). IBM Research Division, Yorktown Heights, NY.

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

Hutchinson, M.F. (2007). *ANUSPLIN version 4.37 User Guide*, The Australian National University,

Centre for Resources and Environmental Studies, Canberra.

Irish, J. (1994). *Review of methodology for estimation of flood frequency curves for extreme events*, Report prepared for the Cooperative Research Centre for Catchment Hydrology, Water Resources Consulting Service, 35pp.

Johnson, F., K. Xuereb, E. Jeremiah and J. Green (2012a), *Regionalisation of rainfall statistics for the IFD Revision Project,* 34th Hydrology & Water Resources Symposium. Sydney, Australia, Engineers Australia.

Johnson, F., Haddad, K., Rahman, A., and Green, J. (2012b), *Application of Bayesian GLSR to Estimate Sub Daily Rainfall Parameters for the IFD Revision Project.* Hydrology and Water Resources Symposium, Sydney, NSW, November 2012.

Kjeldsen, T. R. and D. A. Jones (2009). *A formal statistical model for pooled analysis of extreme floods*, Hydrology Research 40(5): 465-480.

Lu, L.H. and Stedinger, J.R. (1992), *Sampling variance of normalized GEV/PWM quantile estimators and a regional homogeneity test*, J. Hydrol., 138, pp 223-245.

McConarchy, F.L.N (1996), *Estimation of the rainfall frequency curve in the extreme range using regional methods*, PhD Thesis, Dept. of Civil Eng., Monash University, Australia.

Nandakumar, N., Weinmann, P.E., Mein, R.G. and R.J. Nathan (1997), *Estimation of Extreme Rainfalls for Victoria Using the CRC-FORGE Method (For Rainfall Durations 24 to 72 Hours)*, Report 97/4, CRC for Catchment Hydrology, Monash University, Australia.

Nandakumar, N., Jordan, P., Weinmann, E., Sih, K., Hill, P and Nathan, R. (2012), *Estimation of Rare Design Rainfalls for New South Wales and the Australian Capital Territory: An Application of the CRC-FORGE Method.* Proceedings of 34th Hydrology and Water Resource Symposium, Engineers Australia.

Nathan, R.J. and P.E. Weinmann (1999), *Book VI Estimation of Large to Extreme Floods*, Australian Rainfall and Runoff, Institution of Engineers, Australia.

Reed, D.W., and Stewart, E.J. (1989), *Focus on rainfall growth estimation*, Proc. 2nd National Hydrology Symposium, British Hydrological Society.

Schaefer, M.G. (1990), *Regional analysis of precipitation annual maxima in Washington State*, Water Resour. Res., 26(1), pp119-131.

The, C., Johnson, F., Hutchinson, M., and Green, J. (2012), *Gridding of Design Rainfall Parameters for the IFD Revision Project for Australia*, 34th Hydrology & Water Resources Symposium. 19-22 November 2012, Engineers Australia, Sydney, Australia.

The, C., Hutchinson, M., Johnson, F., Beesley, C. and Green, J. (2014). *Application of ANUSPLIN to produce new Intensity-Frequency-Duration (IFD) design rainfalls across Australia*. Hydrology and Water Resources Symposium, Perth, WA, February 2014.

Wallis, J.R., Schaefer, M.G., Barker, B.L. and Taylor, G.H. Taylor (2007), *Regional precipitation-frequency analysis and spatial mapping for 24-hour and 2-hour durations for Washington State*, Hydrol. Earth Syst. Sci., 11(1), 415-442, 2007.

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

Wang, QJ. (1998) *Approximate goodness-of-fit tests of fitted generalised extreme value distributions using LH moments.* Water Resources Research, 34(12) 3497-3502.