Regionalisation of Rainfall Duration in Victoria for Design Flood Estimation Using Monte Carlo Simulation

K Haddad[1] **and A Rahman**[2]
[1]**Snowy Mountain Engineering Corporation, Australia,** [2]**School of Engineering and Industrial Design, University of Western Sydney, Australia, Email:** Khaled.Haddad@smec.com.au

Keywords: *Design Flood Estimation; Monte Carlo Simulation; Regionalisation; Rainfall Runoff Modelling; Ungaugued Catchments; Rainfall duration*

EXTENDED ABSTRACT

Design flood estimation is often required in hydrologic practice such as design of hydraulic structures and flood plain management. In flood estimation, use of rainfall runoff models is frequently adopted which convert selected rainfall events into the corresponding streamflow events. In Australia, runoff routing model is frequently adopted as the preferred method of rainfall runoff modeling. The national guideline for design flood estimation known as Australian Rainfall and Runoff (ARR) recommends Design Event Approach for adoption with the runoff routing models. The Design Event Approach considers the probabilistic nature of rainfall intensity but largely ignores the probabilistic behavior of other input variables in the rainfall runoff modeling such as rainfall temporal pattern and initial loss.

In recent years, there have been significant researches in Australia on the development and application of the Joint Probability Approach/ Monte Carlo Simulation technique to design flood estimation. The superiority of such an approach is based on the fact that this accounts explicitly for the probabilistic nature of the major input variables in the rainfall runoff modeling. The application of this approach so far has been limited to gauged catchments with reasonably long rainfall and streamflow records. However, in practical situations, many catchments are ungauged where there is no or little data available to identify the probability distributions of various input variables. To apply the Joint Probability Approach to ungauged catchments, it is necessary to regionalise the distributions of the input variables.

This paper presents the regionalisation of the distribution of rainfall duration in Victoria, which is provisionally divided into four zones, roughly cutting the state into quadrants along the Great Dividing Range and north from Melbourne The study uses pluviograph data from 91 stations across the region. The selected stations have an average 30 years of continuous pluviograph data.

The paper adopts three goodness-of-fit tests, Chi-squared test, Kolmogorov-Smirnov test and Anderson-Darling test and two candidate distributions are considered: Exponential and Gamma. Both at-site and regional analyses are undertaken. For regional analysis, the rainfall duration data across the region are pooled to estimate the parameters of the selected distribution.

For the at-site analysis, the Chi-squared test rejects the highest number of stations (87%). Based on the Kolmogorov-Smirnov and Anderson-Darling tests, 63% and 68% of the stations satisfy the at-site Exponential and Gamma distributions, respectively, considering Victoria as a single region. For Zone 1 (south-eastern Victoria), only about 40% of the stations satisfy either Exponential or Gamma distribution. For Zones 2 and 3 (north of Great Dividing Range), over 80% of the stations satisfy either Exponential or Gamma distribution. For Zone 4 (south-western Victoria), about 60% of the stations satisfy either Exponential or Gamma distribution.

Considering all the three tests, about 40% stations, satisfy a regional Exponential or Gamma distribution. The two-parameter Gamma distribution does not provide better fit than the one-parameter Exponential distribution. Zone 3 provides best result in that about 50% of the stations satisfy a regional Exponential or Gamma distribution.

Given that more than 50% of the selected stations do not satisfy either regional Exponential or Gamma distributions, other distributions should be examined to identify a more acceptable distribution to regionalize the distributions of storm durations in Victoria. It is also important to examine the effects of various regional distributions (e.g. Exponential and Gamma) on derived flood frequency curves to select an acceptable distribution so far the practical application of the Monte Carlo simulation technique for design flood estimation is concerned.

1. INTRODUCTION

Design flood estimation is often required in hydrologic practice such as design of hydraulic structures and flood plain management. In flood estimation, use of rainfall runoff models is frequently adopted which convert selected rainfall events into the corresponding streamflow events. In Australia, runoff routing model is frequently adopted as the preferred method of rainfall runoff modeling. The national guideline for design flood estimation known as Australian Rainfall and Runoff (ARR) recommends Design Event Approach for adoption with the runoff routing models (I. E. Aust., 1997).

The Design Event Approach considers the probabilistic nature of rainfall intensity but largely ignores the probabilistic behavior of other input variables in the rainfall runoff modeling such as rainfall temporal pattern and initial loss. In recent years, there have been notable researches in Australia on the application of a more holistic approach to design flood estimation such as Joint Probability Approach/ Monte Carlo Simulation (e.g. Rahman et al., 1998; Hoang et al., 1999; Rahman et al., 2001, 2002a, b, c, d; Weinmann et al., 2002; Kuczera et al., 2003; Nathan et al., 2003; Nathan and Weinmann, 2004; Rahman and Carroll, 2004).

The application of the Joint Probability Approach/ Monte Carlo Simulation technique such as by Rahman et al. (2002a) for ungauged catchments would require regionalisation of the parameters of the distributions of the key input variables such as rainfall duration, intensity, temporal pattern and losses. This paper examines the regionalisation of rainfall duration for the state of Victoria for application with the Monte Carlo Simulation technique for design flood estimation.

2. STUDY AREA AND DATA

The state of Victoria is provisionally divided into four hydrometeorological zones, roughly cutting the state into quadrants along the Great Dividing Range and north from Melbourne, as shown in Figure 1. This division is made similar to Rahman et al., 2001 to examine regional differences in rainfall duration characteristics in Victoria, if any. A total of 91 pluviograph stations were selected (27, 22, 16 and 26 stations) from Zones 1, 2, 3 and 4, respectively. These stations have an average of 30 years of continuous pluviograph data.

3. METHOD

Previous study on a smaller number of pluviograph stations (Rahman et al., 2002a) indicated that probability distribution of ‘storm-core’ (defined later in the section) durations in Victoria can be approximated by an Exponential Distribution. In this study, it was found that for many of the selected pluviograph stations, the mean and standard deviation values of ‘storm-core’ durations were quite different, suggesting a distribution other than Exponential. Thus, two candidate distributions are considered in this study: one-parameter Exponential distribution and two-parameter Gamma distribution. To test the statistical hypothesis that storm-core duration data in a particular pluviograph station follow either Exponential or Gamma distribution, three tests were applied: Chi-squared (C-S) test, Kolmogorov-Smirnov (K-S) Test and Anderson-Darling (A-D) test, at 5% level of significance.

The Chi-squared test is based on the Chi-squared statistic, which is related to the weighted sum of squared differences between the observed and theoretical frequencies. The test statistic is given by (Kottegoda and Rosso, 1997):

$$X^2 = \sum_{i=1}^{l} \frac{(O_i - E_i)^2}{E_i} \tag{1}$$

Where O_i is observed frequency and E_i is expected frequency for class i from a total of l classes. A large value of X^2 indicates a poor fit. The sampling distribution of X^2 tends, as sample size n approaches infinity, to a χ_v^2 distribution, where $v = l\text{-}1\text{-}k$ represents the degrees of freedom and k is the number of parameters estimated from the same data used for the test.

The Kolmogorov–Smirnov test is based on the maximum difference (D_{max}) between the observed cumulative distribution function $F_n(x)$ and expected cumulative distribution function $F_o(x)$. This test can be applied in two ways. The first method uses absolute D_{max} value which can be worked out by the following equation (Kottegoda and Rosso, 1997):

$$D_{\max} = \left[F_n(x) - F_o(x)\right] \tag{2}$$

Hence the maximum difference between the two sets of data, the critical D_{max} value is compared to the rejection region. In this study, since most stations has over 35 events *(n)*, following equation is used to compute the rejection region.

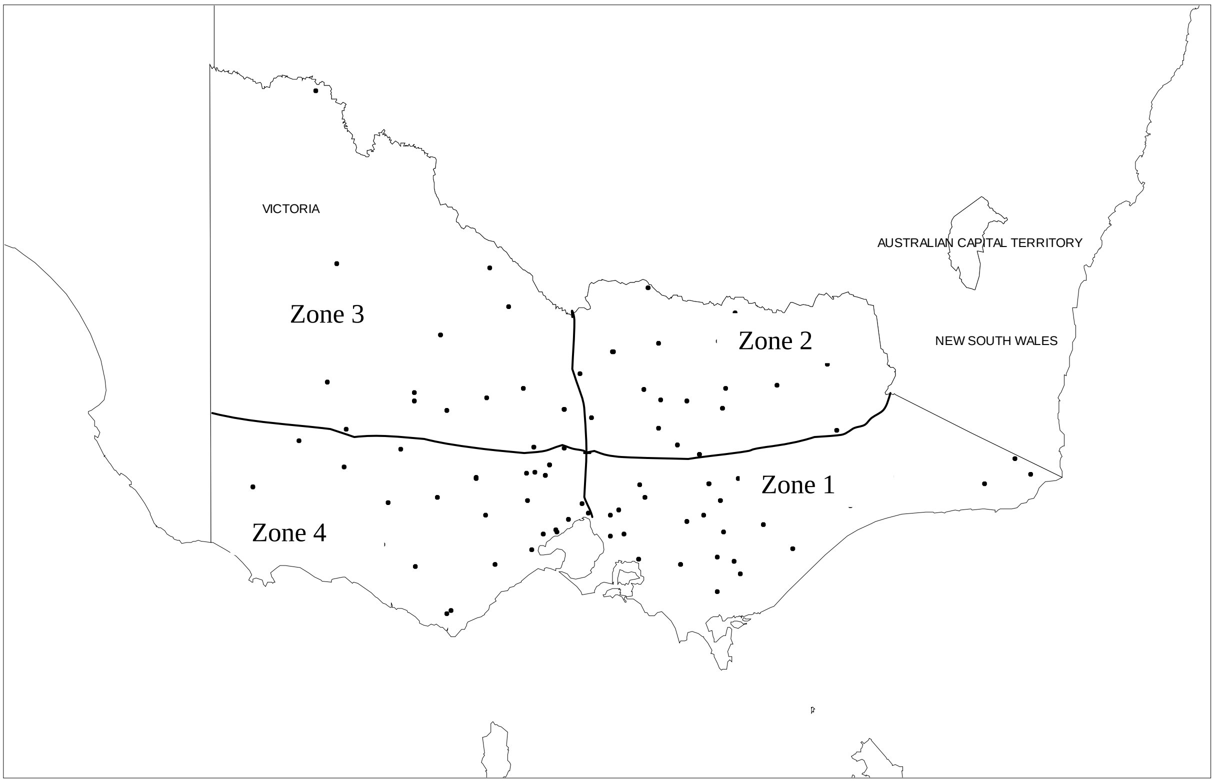

Figure 1. Selected pluviograph stations from Victoria

$$D_{n,\ 0.05} = 1.36/\sqrt{n} \qquad (3)$$

In the second approach, the D_{max} value is incorporated into the equation that uses the number of storm events *n.* This equation is shown below (Stephens, 1974):

$$D = D_{\max}\left(\sqrt{n} + 0.12 + 0.11/\sqrt{n}\right) \qquad (4)$$

To accept the null hypothesis, the value of *D* should be less than the rejection region, which is 1.358 for 5% significance level (Stephens, 1974). In this study, both the approaches are adopted i.e. if a station satisfies both the criteria, it passes the test. Here, the parameters of the distributions are obtained from the same sample that is used for the test.

The Anderson–Darling test is devised to give heavier weightings to the tails of a distribution where unexpectedly high or low values, called outliers are located (Kottegoda and Rosso, 1997). The rejection region for this test, with sample size greater than 5, at a 5 percent significance level is 2.492 (Stephens 1974).

These three tests were applied to the individual site and also to Zones 1, 2, 3 and 4 to assess the applicability of regional Exponential or Gamma distribution.

The probability density function of one-parameter Exponential distribution can be given by the following equation:

$$p(d_c) = \frac{1}{\sigma} e^{-d_c/\sigma} \qquad (5)$$

Where d_c is storm–core duration and σ is the mean value of d_c.

In the form of a cumulative density function this can be given by the following equation (Kottegoda and Rosso, 1997):

$$F(d_c) = 1 - e^{-d_c/\sigma} \qquad (6)$$

The probability density function of the two-parameter Gamma distribution can be given by the following equation (Kottegoda and Rosso, 1997):

$$p(d_c;\alpha;\beta) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} d_c^{\ \alpha-1} e^{-d_c/\beta} \quad d_c > 0 \qquad (7)$$

Where α and β are parameters of the Gamma distribution, and can be obtained from:

$$\beta = \frac{SD^2}{\bar{d}_c} \quad (8)$$

$$\alpha = \frac{\bar{d}_c^{\,2}}{SD^2} \quad (9)$$

Where $\bar{d}_c$ is the mean value of storm-core duration at a station or in a region and SD^2 is the variance of d_c values at a station or in a region.

In the Monte Carlo Simulation Technique, to provide the basis for a rigorous assessment of flood probabilities, a new storm event definition is required that produces rainfall events of random durations. Two different storm event definitions can be used: a 'complete storm' and a 'storm-core' within each complete storm (the most intense part of the storm) (Rahman et al., 2002a). A complete storm is defined as a period of significant rain preceded and followed by an arbitrarily defined period of dry hours (6 hours used here). These complete storm events are partial duration series events and are selected based on a threshold value of rainfall intensity in such way that on average, 4 to 7 top rainfall events are selected per year from each pluviograph station.

The corresponding storm-core is selected as the period within a complete storm that has the highest rainfall intensity ratio compared to the 2-year average recurrence interval (ARI) design rainfall. The selected storm-core events are then analysed to identify probability distributions of rainfall duration, intensity and temporal pattern. Following this approach, storm-core events were selected from each of the selected pluviograph stations.

4. RESULTS

The mean values of storm-core duration ($\bar{d}_c$) and its standard deviations for the 91 stations are given in Table 1.

The histograms of storm-core duration (d_c) values for each of the 91 stations were plotted at both 5 and 10 hours class intervals (sample shown in Figure 2). Plots of cumulative frequency distributions for the observed and fitted Exponential and Gamma distributions were prepared for visual assessment of the goodness-of-fit of a distribution. From the visual inspection, the fitting was rated on a criterion of 'poor', 'medium' and 'good'. Summarized results can be seen in Table 2, which shows that the Gamma distribution apparently fits the at–site and regional observed frequency of d_c data better than the Exponential distribution for all the four zones. Overall, the Gamma distribution provides a 'good fit' visually for 65% of the stations in Victoria as compared to 46% of the stations for the Exponential distribution.

Table 3 shows that for the at-site analysis, the Chi-squared test rejects the highest number of stations (87%). Based on the Kolmogorov-Smirnov and Anderson-Darling tests, 63% and 68% of the stations satisfy the at-site Exponential and Gamma distributions, respectively, across the four zones. For Zone 1 (south-eastern Victoria), only about 40% of the stations satisfy either Exponential or Gamma distribution. For Zones 2 and 3 (north of Great Dividing Range), over 80% of the stations satisfy either Exponential or Gamma distribution. For Zone 4 (south-western Victoria), about 60% of the stations satisfy either Exponential or Gamma distribution. These results show that there is a remarkable difference in the distributions of storm-core durations across the four zones of Victoria.

For regional analysis, the mean and standard deviation of the combined at-site d_c data for all the stations within the region were considered. The regional average mean d_c values are 14.1h, 13.3h, 11h and 12.9h respectively for Zones 1, 2, 3 and 4, respectively. The regional average standard deviation values of the d_c data are 17.2h, 14.5h, 11.2h and 16.6h respectively for Zones 1, 2, 3 and 4, respectively.

The regional average mean d_c value for a zone was used to fit the regional Exponential distribution and hypothesis testing was conducted against the at–site d_c data to assess the viability of a regional Exponential distribution. Similarly, the regional Gamma distribution was fitted using the regional average mean d_c value and regional average standard deviation value of the d_c data for a zone. The results of the hypothesis tests for regional distributions are summarized in Table 4.

Considering all the three tests, about 40% of the stations, satisfy a regional Exponential or Gamma distribution across the four zones. The two-parameter Gamma distribution does not provide remarkably better fit than the one-parameter Exponential distribution. Zone 3 provides the best result in that about 50% of the stations satisfy a regional Exponential or Gamma distribution based on the three tests. However, considering the Kolmogorov-Smirnov and Anderson-Darling test, about 70% of the stations in Zone 3 satisfy either a regional Exponential or Gamma distribution.

Table 1. Selected pluviograph stations and observed mean and standard deviation values of d_c values

Zone 1			Zone 2			Zone 3			Zone 4		
Station ID	$\bar{d}_c$ (h)	SD of d_c (h)	Station ID	$\bar{d}_c$ (h)	SD of d_c (h)	Station ID	$\bar{d}_c$ (h)	SD of d_c (h)	Station ID	$\bar{d}_c$ (h)	SD of d_c (h)
83033	18.5	19.4	80109	8.6	9.4	76031	9.8	8.5	79052	10.8	12.6
84005	13.6	16.4	81013	11.8	11.3	77087	8.7	8.2	86038	10.1	10.2
84015	10.6	15.4	81049	9.8	9.2	79046	11.0	12.1	86071	9.4	10.8
84078	11.8	13.4	81114	11.5	11.8	79079	10.7	10.8	87017	14.9	16.7
84112	23.3	29.1	81115	11.9	12.3	79082	9.6	8.4	87029	13.4	14.5
84122	17.8	20.1	82011	14.2	13.1	79086	10.9	10.2	87031	9.2	9.1
84123	10.4	10.6	82016	11.8	10.9	80006	8.6	5.6	87033	9.2	9.2
84125	16.7	20.6	82039	10.5	10.5	80102	8.9	9.6	87036	14.8	17.0
85000	11.0	14.9	84042	15.1	15.9	80110	9.5	8.6	87075	12.7	14.6
85026	14.3	15.3	82076	18.8	17.1	81003	11.9	11.8	87097	10.9	12.7
85034	11.7	14.0	82107	13.7	14.8	81026	11.0	10.2	87104	10.3	9.7
85072	12.9	13.2	82121	11.5	10.6	81038	12.1	12.1	87105	9.3	8.5
85103	20.3	21.2	83017	16.6	20.1	87036	14.6	17.0	87133	9.6	11.2
85106	19.9	20.2	83025	13.1	14.2	87153	8.1	9.4	89002	11.1	14.9
85170	12.3	14.0	83031	16.5	17.8	88029	11.5	10.2	89016	9.8	10.4
85176	25.3	25.2	83033	18.5	19.4	88037	11.7	12.5	89019	9.6	12.0
85236	10.7	13.3	83067	14.1	14.4				89025	14.5	15.8
85237	20.7	20.1	83074	20.9	20.9				89082	11.3	13.2
85240	12.8	15.4	88023	9.7	12.1				89085	10.2	10.5
85256	15.6	19.2	88029	10.9	9.4				89094	12.9	14.8
86074	10.9	13.5	88049	9.7	8.4				90058	12.9	15.7
86085	9.4	11.2	88153	9.1	10.1				90083	28.7	29.5
86142	11.6	14.4							90087	22.3	31.8
86219	18.5	22.6							90135	12.4	16.4
86224	7.6	9.5							90153	15.4	15.6
86234	10.2	12.6							90166	14.3	16.6
86314	13.3	16.4									

Table 2. Summary of Goodness–of–fit results, visual assessment (% of stations having a particular fit)

Distribution	Visual Assessment	Zone 1	Zone 2	Zone 3	Zone 4	**Average**
Exponential distribution	Good Fit	15%	64%	69%	35%	**46%**
	Medium Fit	44%	27%	13%	50%	**33%**
	Poor Fit	41%	9%	18%	15%	**21%**
Gamma Distribution	Good Fit	41%	82%	69%	69%	**65%**
	Medium Fit	44%	18%	25%	27%	**28%**
	Poor Fit	15%	0	6%	4%	**6%**

Table 3. Summary of hypothesis test results for at–site distributions (stations passed the test)

Distribution	Zone 1 (27 stations)	Zone 2 (22 stations)	Zone 3 (16 stations)	Zone 4 (26 stations)	**Average**
Exponential:					
C- S test	4 (15%)	2 (9%)	0 (0%)	6 (23%)	
K- S test	7 (26%)	16 (73%)	14 (88%)	16 (62%)	
A – D test	12 (44%)	17 (77%)	12 (75%)	16 (62%)	
Average	28%	53%	54%	49%	**46%**
Gamma:					
C- S test	6 (22%)	1 (5%)	0 (0%)	6 (23%)	
K- S test	11 (41%)	17 (77%)	13 (81%)	12 (46%)	
A – D test	15 (56%)	19 (86%)	15 (94%)	17 (65%)	
Average	40%	56%	58%	45%	**50%**

Table 4. Summary of hypothesis test results for regional distributions (stations passed the test)

Distribution	Zone 1 (27 stations)	Zone 2 (22 stations)	Zone 3 (16 stations)	Zone 4 (26 stations)	**Average**
Exponential:					
C- S test	14 (52%)	6 (27%)	3 (19%)	15 (58%)	
K- S test	5 (19%)	11 (50%)	12 (75%)	7 (27%)	
A – D test	5 (19%)	11 (50%)	10 (63%)	6 (23%)	
Average	30%	42%	52%	36%	**40%**
Gamma:					
C- S test	12 (44%)	4 (18%)	2 (13%)	6 (23%)	
K- S test	6 (22%)	8 (36%)	11 (69%)	2 (8%)	
A – D test	16 (59%)	13 (59%)	11 (69%)	19 (73%)	
Average	42%	38%	50%	35%	**41%**

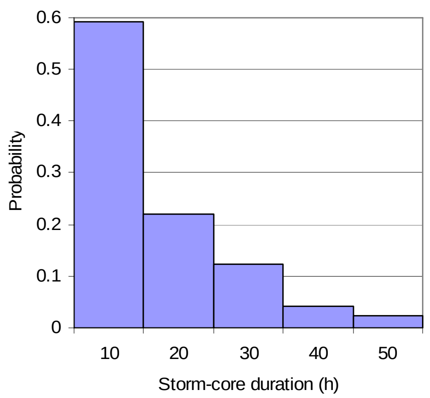

Figure 2. Histogram of storm-core duration for Station 82121

5. CONCLUSIONS

This paper compares one-parameter Exponential and two-parameter Gamma distributions for describing distribution of storm-core duration data in the state of Victoria Australia. The following conclusions can be drawn from this study:

- The two-parameter Gamma distribution does not provide better fit to the storm-core duration data in Victoria than the one-parameter Exponential distribution.

- For the at-site analysis, the Chi-squared test rejects the highest number of stations (87%). Based on the Kolmogorov-Smirnov test and Anderson-Darling A-D tests, 63% and 68% of the stations satisfy the at-site Exponential and

Gamma distributions, respectively, across the four zones. For Zone 1 (south-eastern Victoria), only about 40% of the stations satisfy either Exponential or Gamma distribution. For Zones 2 and 3 (north of Great Dividing Range), over 80% of the stations satisfy either Exponential or Gamma distribution. For Zone 4 (south-western Victoria), about 60% of the stations satisfy either Exponential or Gamma distribution.

- Considering all the three tests, on average, about 40% stations satisfy a regional Exponential or Gamma distribution.

- Given that 60% of the selected stations do not satisfy either regional Exponential or Gamma distribution, other distributions should be examined to identify a more acceptable distribution. It would be also important to examine the effects of various regional distributions (e.g. Exponential and Gamma) on derived flood frequency curves to select an acceptable distribution so far the practical application of the Monte Carlo simulation technique for design flood estimation is concerned.

6. ACKNOWLEDGMENTS

The authors would like to acknowledge the Australian Bureau of Meteorology for providing the data for the project, Mr. Erwin Weinmann of Monash University for providing constructive suggestions on the work and two anonymous reviewers for providing constructive suggestions on the draft paper.

7. REFERENCES

Hoang, T.M.T, Rahman, A., Weinmann, P.E., Laurenson, EM and Nathan, RJ 1999, Joint probability descriptions of design rainfalls, In *Proc. Intl. Hydro. and Water Resour Symp*, I.E. Aust., Brisbane, Australia, 6-8 July.

Institution of Engineers Australia (I. E. Aust.) 1997, Australian Rainfall and Runoff, Vol. 1.

Kottegoda, N.T. and Rosso, R. 1997, *Statistics, Probability, and Reliability for Civil and Environmental Engineers*, McGraw-Hill.

Kuczera, G., Lambert, M., Heneker, T., Jennings, S, Frost, A and Coombes, P 2003, Joint Probability and Design Storms at the Crossroads. Keynote paper. *28th Intl. Hydrology and Water Resources Symposium*, 11 – 13 Nov, Wollongong, Australia.

Nathan, R.J. and Weinmann, P.E. 2004, An improved framework for the characterisation of extreme floods and for the assessment of dam safety. In *Proc. British Hydrological Society Intl. Confer.*, Imperial College London, 1, 186-193.

Nathan, R.J., Weinmann, P.E. and Hill, P. 2003, Use of Monte Carlo Simulation to estimate the expected probability of large to extreme floods. In *Proc. 28th Intl. Hydrology and Water Resources Symp.*, 10-14 Nov, Wollongong, 1, 105-112.

Rahman, A. and Carroll, D. 2004, Appropriate spatial variability of flood producing variables in the joint probability approach to design flood estimation. In Proc. *British Hydrological Society Intl. Conference,* London, 1, 335-340.

Rahman, A. and Weinmann, P.E. 2002, Flood estimation in Northern Australian catchments using Monte Carlo Simulation technique. Proc. *27th National Hydrology and Water Resources Symp.*,I. E. Aust., 20-23 May, Melbourne.

Rahman, A., Hoang, T.M.T., Weinmann, P.E., Laurenson, E.M. 1998, *Joint Probability Approaches to Design Flood Estimation: A Review.* Research Report 98/8, CRC for Catchment Hydrology, Australia. 70 pp.

Rahman, A., Weinmann, P.E., Hoang, T.M.T. Laurenson, E.M., Nathan, R.J. 2001, *Monte Carlo Simulation of Flood Frequency Curves from Rainfall.* Technical Report 01/4. CRC for Catchment Hydrology, Australia. 63 pp.

Rahman, A., Weinmann, P.E., Hoang, T.M.T., Laurenson, E.M. 2002a, Monte Carlo Simulation of flood frequency curves from rainfall. *Journal of Hydrology*, 256 (3-4), 196-210.

Rahman, A., Carroll, D. and Weinmann, P.E. 2002b. Integration of Monte Carlo Simulation technique with URBS model for design flood estimation. *Proc. 27th National Hydrology and Water Resources Symp.*, I. E. Aust., 20-23 May, Melbourne. CD-ROM Publication. 7 pp.

Rahman, A., Weinmann, P.E. and Mein, R.G. 2002c, The use of probability-distributed initial losses in design flood estimation. *Australian Journal of Water Resources.* 6(1), 17-30.

Stephens, M.A. (1974). EDF statistics for goodness-of-fit and some comparisons. Journal of the American Statistical Association. 69, 377, 730-737.

Weinmann, P.E., Rahman, A., Hoang, T.M.T., Laurenson, E.M. and Nathan, R.J. 2002, Monte Carlo simulation of flood frequency curves from rainfall – the way ahead, *Australian Journal of Water Resources*, *6*(1), 71-80