

# Regional flood modelling in arid and semi-arid regions in Australia

M. Zaman <sup>a</sup>, A. Rahman <sup>a</sup> and K. Haddad <sup>a</sup>

<sup>a</sup> *School of Engineering, University of Western Sydney, Australia*  
Email: [A.Rahman@uws.edu.au](mailto:A.Rahman@uws.edu.au)

**Abstract:** Regional flood frequency analysis (RFFA) is often used in hydrological studies to estimate design floods for ungauged catchments. The RFFA techniques in arid and semi-arid regions are more difficult to develop, largely because of paucity of the recorded streamflow data, extreme temporal and spatial variability of floods and long periods in the streams without any significant flow. The current national design guideline, Australian Rainfall and Runoff (ARR), contains little information and guidance on RFFA in arid and semi-arid regions of Australia. This study develops a RFFA method based on the best available data set for the arid and semi-arid regions of Australia. Here, an index type RFFA procedure is adopted. Dimensionless regional flood frequency curves are developed in the form of regional growth factors.

The study uses streamflow and catchment characteristics data from 45 catchments in the arid and semi-arid regions of New South Wales (NSW), Victoria (VIC), Queensland (QLD), South Australia (SA) and Northern Territory (NT). The mean annual rainfall values for these selected catchments are smaller than 450 mm. The partial duration series flood data are used in this study to estimate the flood quantiles for average recurrence intervals (ARI) of 2 to 100 years using the generalised Pareto distribution fitted with the method of L-moments. The areas and streamflow record lengths of the selected catchments range from 3.8 to 5,975 km<sup>2</sup> and 10 to 46 years, respectively.

The index-flood, which is taken as the mean annual flood in this study, is related to catchment area and design rainfall intensity by ordinary least squares regression to provide a means of estimating the mean annual flood for an ungauged catchment. This estimated mean annual flood and the regional growth factors permit flood quantile estimation for different ARIs in the study regions.

Separate sets of prediction equations for mean annual flood are developed for VIC & NSW (combined), QLD, SA and NT. For the testing of the developed regional flood estimation methods, a one-at-a-time validation procedure is adopted. A number of evaluation statistics have been employed to assess the performances of the developed RFFA models. It has been found that the developed prediction equations provide relatively accurate design flood quantile estimates in the study regions with median relative error in the range of 20% to 110%. The steepness of the developed growth curves reflects the variability of the flood discharge with ARIs and it has been found that SA has the steepest growth curve and QLD the least one. This implies that SA design floods in arid and semi-arid regions show greater variability across different ARIs as compared to those of other Australian states. The developed RFFA methods in the arid and semi-arid regions are based on relatively shorter streamflow records and a smaller number of stations as compared to the coastal regions of Australia and hence have relatively lower degree of accuracy.

**Keywords:** *Arid and semi-arid regions, regional flood frequency analysis, ungauged catchments, Australian Rainfall and Runoff, index flood*

## 1. INTRODUCTION

Understanding of the probabilistic behaviour of extreme flows is important in the design of hydraulic structures such as weirs, barrage, dams and flood control facilities. Underestimation and overestimation of design flood discharge lead to increased flood risk and project costs, respectively, both of which are undesirable. It often happens that at the sites of interest the streamflow record length is much shorter than the return period of interest, and even worse when the site of interest is ungauged (i.e. no recorded streamflow data is available). Regional Flood Frequency Analysis (RFFA) can be used to obtain more reliable flood quantile estimates for ungauged catchments, which attempts to transfer flood characteristics information from gauged to ungauged catchments. One of the most commonly adopted RFFA methods is the Index Flood Method (IFM), which was initially developed by the United States Geological Survey (Dalrymple, 1960) and is based on the assumption that flood flows in a hydrologically similar region, when standardised by an appropriate index flood, are identically distributed. An alternative method to the IFM is the Quantile Regression Technique (QRT) (Benson, 1962; Thomas & Benson, 1970). In the QRT, flood quantiles are estimated from recorded streamflow data, which are then regressed against catchment characteristics that govern the flood generation process and which are relatively easy to obtain. To develop RFFA using the IFM and QRT, one needs long periods of recorded streamflow data for a large number of stations in a given region.

RFFA in arid and semi-arid regions is a difficult task due to information scarcity, extreme streamflow variability and long periods without flow and flood dominated regimes. There have been only a few RFFA studies in the arid regions worldwide. For example, Farquharson *et al.* (1992) developed a RFFA method based on the IFM considering data from the arid and semi-arid regions in a number of countries in the world. Ben-Zvi and Ben-Zvi (1971); Ben-Zvi and Cohen (1975) and Ben-Zvi (1982, 1996) developed event-based regional flood models in the arid regions of Israel applying the Poisson distribution for modelling the frequency of events, and the log-normal distribution for the magnitudes. Since individual convective storms dominate the flood regime in arid regions, many studies (e.g. Hjalmarsen and Thomas, 1992; Meirovich *et al.*, 1998) focused on relatively small catchments. Hjalmarsen and Thomas (1992) assumed that the annual maximum floods at different stations in the regions are independent from one another and they are proportional to the square root of the catchment area. An improved RFFA method was introduced by Meirovich *et al.* (1998), which included envelope curves for maximum observed discharges and for the volumes of events along with a new methodology of threshold selection for partial duration series flood data.

Australian Rainfall and Runoff (ARR) – the national guide of flood estimation provides little guidance on design flood estimation in arid and semi-arid regions of Australia (I. E. Aust., 1987). ARR87 developed RFFA methods for the arid and semi-arid regions in Australia based on a very limited data. Since then there has not been any update of the RFFA methods in the arid and semi-arid regions of Australia similar to other parts of Australia. This paper focuses on the development of a new RFFA method for arid and semi-arid regions in Australia using the currently available flood data in this region.

## 2. METHODOLOGY

An index flood method (IFM) is used in this study. The mean annual flood (MAF) is taken as the index-flood. In this method the Average Recurrence Interval (ARI) of  $T$ -years flood quantile ( $Q_T^i$ ) at any site  $i$  can be estimated as:

$$Q_T^i = \bar{Q}_i \times Z_T \quad (1)$$

Where,  $\bar{Q}_i$  = the MAF based on the partial duration series (PDS) at site  $i$ , and  $Z_T$  = the regional growth factor.

In this study, a prediction equation is developed for  $\bar{Q}_i$  as a function of catchment characteristics ( $X$ ). A regional growth curve is developed for a region. Here partial duration series (PDS) flood data is used. In the application, partial series based  $Q_T^i$  estimates may need to be converted to annual maximum flood series estimates using the Langbein transformation. The MAF is estimated from equation 2 where an Ordinary Least Squares (OLS) fitting procedure is adopted to estimate the regression coefficients:

$$\log_{10}(\bar{Q}) = b_0 + b_1 \log_{10}(X_1) + b_2 \log_{10}(X_2) + \dots \quad (2)$$

The flood series at each gauging station were reduced to dimensionless form by dividing each by the MAF before fitting the Generalised Pareto (GP) distribution. In the PDS model all events above a certain threshold level are considered. It is generally assumed that the occurrence of peaks can be described by a Poisson process with an annual periodic intensity, implying that the annual number of peaks is distributed according to the Poisson distribution. The Poisson parameter  $\lambda$  equals the expected number of threshold exceedances per year and is estimated as:

$$\lambda = N/t \quad (3)$$

where  $N$  is the number of observed exceedances in the period of  $t$  years. The exceedance magnitudes are assumed to be independent and identically distributed following the GP distribution. The GP distribution with the scale parameter  $\alpha$  and the shape parameter  $\kappa \neq 0$  has the cumulative distribution function (CDF):

$$F(q) = 1 - \left(1 - \kappa \frac{q - q_0}{\alpha}\right)^{1/\kappa} \quad (4)$$

The threshold level  $q_0$  is determined a priori, and a new variable,  $D = Q - q_0$ , is defined for estimation of the two remaining parameters  $\alpha$  and  $\kappa$ . The mean and the variance of  $D$  are as follows:

$$\mu = E\{D\} = \frac{\alpha}{1 + \kappa} \quad (5)$$

$$\sigma^2 = \text{var}\{D\} = \frac{\alpha^2}{(1 + \kappa)^2 (1 + 2\kappa)} \quad (6)$$

The  $T$ -year event in a PDS context is usually defined as the  $(1 - 1/\lambda T)$  quantile (e.g., Rosbjerg, 1985). Alternatively, the  $T$ -year event can be defined as the  $(1 - 1/T)$  quantile in the annual maximum distribution that corresponds to the parent PDS model. In this case one obtains (Madsen et al., 1997):

$$Q_T = q_0 + \frac{\alpha}{\kappa} [1 - \exp(-\kappa [\ln(\lambda) + y_T])] \quad (7)$$

where  $y_T = -\ln[-\ln(1 - 1/T)]$  is the Gumbel reduced variate.

In a regional model, the parameter estimation method that yields the least-biased parameter estimators is usually preferable, and in this respect the L-moments method is applied. The L-moments estimates, of the PDS/GP parameters are:

$$\hat{\alpha} = \hat{\lambda}_1 \left( \frac{1}{\hat{t}_2} - 1 \right) \quad (8)$$

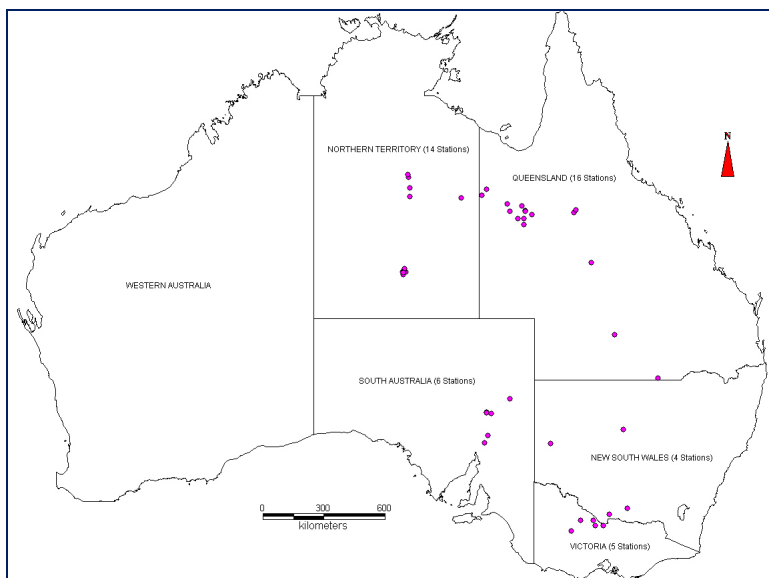
$$\hat{\kappa} = \frac{1}{\hat{t}_2} - 2 \quad (9)$$

where  $\hat{\lambda}_1$  is an estimate of the first L-moments (equal to the sample mean), and  $\hat{t}_2$  is an estimate of the L-coefficient of variation ( $L-C_v$ ); both are determined on the basis of the sample of exceedances  $d_i = q_i - q_0$ , where  $i = 1, 2, \dots, N$ .

### 3. STUDY AREA AND DATA

A total of 45 catchments were selected from arid and semi-arid regions of Australia, with locations shown in Figure 1. Based on the criteria by UNSECO (1999), 90% of the selected catchments are classified as semi-arid ( $0.2 < P/PET < 0.5$ ) and the remaining 10% as arid ( $0.03 < P/PET < 0.2$ ). Here,  $P$  is average annual precipitation and  $PET$  is potential evapotranspiration. The selected catchments have an average annual rainfall in the range of 209 mm to 454 mm. The catchment areas range from 3.8 to 5,975 km<sup>2</sup> (mean: 1,152 km<sup>2</sup> and median: 360 km<sup>2</sup>). The streamflow record lengths range from 10 to 46 years (mean: 25 years and

median: 22 years); these are the partial series maximum flood data, with one event per year on average being selected.



**Figure 1.** Locations of the selected catchments from the arid and semi-arid regions of Australia

#### 4. RESULTS

##### 4.1. Selection of regions

The formation of sub-regions in the arid and semi-arid region appears to be a difficult task, as there are only 45 catchments available from a vast region of interior Australia. To assess the hydrological similarity of various candidate regions, the heterogeneity test by Hosking and Wallis (1993) was applied. The result of the test reveals that none of the proposed regions are ‘acceptably homogeneous’. Since the data set is limited, formation of simple geographic regions appears to be a reasonable option. As NSW and VIC have only 4 and 5 catchments respectively, they are combined to achieve a sample size of 9. The selected catchments of these two states fall in the same drainage division (Drainage Division IV - Murray-Darling Division), and hence their combination appears to be reasonable. The states of QLD, NT and SA were treated as separate regions.

##### 4.2. Derivation of growth curves

The flood quantiles were estimated for  $T = 2, 5, 10, 20, 50$  and  $100$  years at each station by fitting a Generalised Pareto (GP) distribution using the method of L- moments (Madsen *et al.*, 1997). The  $Q_T/\bar{Q}$  values were estimated at individual stations; the weighted average of these values (weighting was done based on record length at individual sites) over all the sites in a region defines the growth curve, as shown in Figure 2. A smooth curve was drawn to represent the average growth curve for each of the four regions; the corresponding equations (Table 1) show a  $R^2$  value in the range of 0.95 to 0.99 indicating a very good fit. The growth factors for the selected ARIs, estimated from the fitted equations, are provided in Table 1. From Figure 2, it can be found that SA has the steepest growth curve followed by the NT. The QLD has the flattest growth curve. In the larger ARI range, the growth curve for SA is much higher than the other regions. The growth curves derived here generally agree with the world growth curves (except for SA at higher ARIs) derived by Farquharson *et al.* (1992) as shown in Table 1.

##### 4.3. Prediction equations for mean annual flood

The statistical package SPSS was used to develop prediction equations for the mean annual flood using an Ordinary Least Squares regression, where many different combinations of variables were examined to come up with the best possible model i.e. the one which had the highest coefficient of determination ( $R^2$ ) and the smallest number of predictor variables. The developed prediction equations are shown in Table 2.

The developed prediction equations contain two predictor variables, catchment area (*area*) and 12 hours duration and 2 years ARI rainfall intensity ( ${}^2I_{12}$ ), which are relatively easy to obtain. The values of the Durbin-Watson statistic ranges from 1.84 to 2.34, which is not far away from 2 (a value of 2 indicates no serial correlation between the predictor variables). The  $R^2$  value for the developed prediction equation for QLD is 0.88, which represents quite a good fit. However, the  $R^2$  values for the other three regions are moderate (in the range of 0.29 to 0.49). The quantile-quantile plots (QQ-plot) of the standardised residuals indicate that the residuals are near-normally distributed and there is no outlier data point. The QQ-plot of the standardised residuals for 20 years ARI for QLD is shown in the Figure 3. The plots of standardised residuals vs. predicted flood quantiles show that the residuals are within  $\pm 2.0 \times$  standard deviation, which indicate the absence of any true outlier data point.

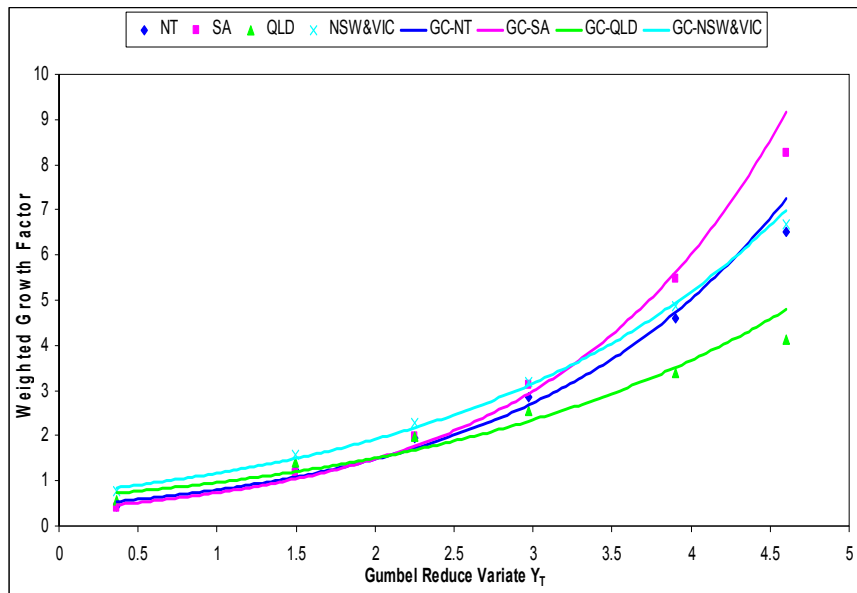


Figure 2. Growth curves (GC) for four selected arid and semi-arid regions in Australia

Table 1. Growth factors for arid and semi-arid regions in Australia ( $Y_T$  = Gumbel reduced variate)

Region		NSW-VIC	QLD	SA	NT	World data (Farquharson et al., 1992)
Equation		$GF = 0.71e^{0.5Y_T}$ $R^2 = 0.99$	$GF = 0.62e^{0.45Y_T}$ $R^2 = 0.95$	$GF = 0.37e^{0.7Y_T}$ $R^2 = 0.99$	$GF = 0.43e^{0.61Y_T}$ $R^2 = 0.99$	
$T$ (years)	$Y_T$					
2	0.37	0.85	0.73	0.48	0.54	0.6
5	1.50	1.50	1.22	1.06	1.07	1.5
10	2.25	2.19	1.71	1.79	1.70	2.1
20	2.97	3.13	2.36	2.96	2.63	3.0
50	3.90	5.00	3.59	5.68	4.65	4.2
100	4.60	7.08	4.91	9.26	7.11	5.9

Table 2. Summary of the prediction equations for the mean annual flood for arid and semi-arid regions

Region	Prediction equation ( $\bar{Q}$ = mean annual flood of partial duration series data in m <sup>3</sup> /s, <i>area</i> = catchment area in km <sup>2</sup> and <sup>2</sup> <i>I</i> <sub>12</sub> = design rainfall intensity of 2 years ARI and 12 hours duration in mm/h, obtained at catchment outlet)	R <sup>2</sup>	Durbin-Watson statistic
NSW-VIC (9 stations)	$\log_{10}(\bar{Q}) = -2.43 + 0.42\log_{10}(\text{area}) + 5.03\log_{10}({}^2I_{12})$	0.35	2.06
QLD (16 stations)	$\log_{10}(\bar{Q}) = -2.71 + 0.76\log_{10}(\text{area}) + 4.01\log_{10}({}^2I_{12})$	0.88	2.34
SA (6 stations)	$\log_{10}(\bar{Q}) = -0.19 + 0.85\log_{10}(\text{area})$	0.29	0.93
NT (14 stations)	$\log_{10}(\bar{Q}) = -0.50 + 0.60\log_{10}(\text{area}) + 1.65\log_{10}({}^2I_{12})$	0.45	1.84

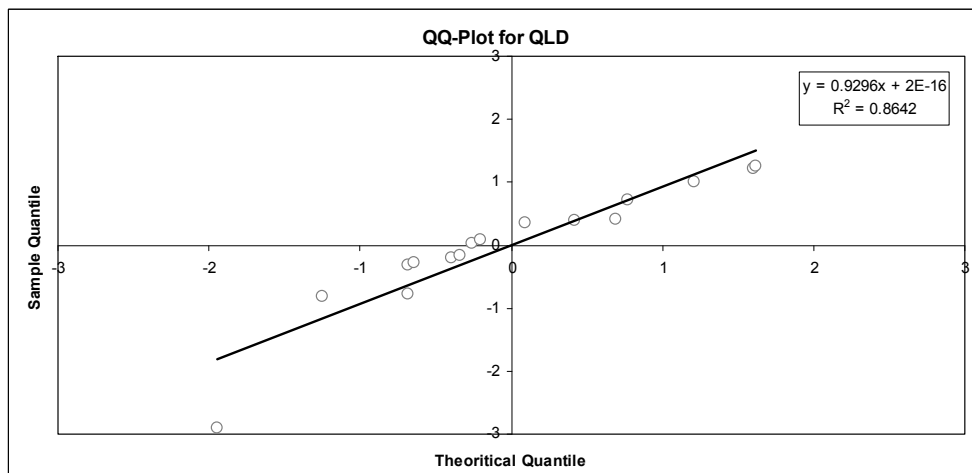


Figure 3. QQ-plot of the standardised residuals for  $Q_{20}$  (QLD)

#### 4.4. Validation and testing

For the testing of the developed regional flood estimation methods, a one-at-a-time (OAT) validation procedure is adopted for generating the mean annual flood (MAF) for each of the study catchments. Using these MAF and growth factors in Table 2, the predicted flood quantiles ( $Q_{pred}$ ) for each of the study catchments are calculated for different ARIs. The  $Q_{pred}$  from OAT are compared with the at-site flood frequency estimates ( $Q_{obs}$ ). The relative error values range from 52% to 110% for NSW-VIC, 30% to 46% for QLD, 20% to 70% for SA and 55% to 94% for the NT, which seem to be quite reasonable given the relatively high uncertainties in arid and semi-arid hydrology.

### 5. CONCLUSIONS

Catchments in the arid and semi-arid regions of Australia have a distinctly different flood hydrology from catchments in more humid/coastal regions, and thus warrant separate treatment for regional flood estimation. The limited availability of streamflow data at both temporal and spatial scales in the arid and semi-arid regions of Australia makes it difficult to develop statistically meaningful Regional Flood Frequency Analysis (RFFA) methods. Using the limited data, a simple RFFA method is developed here for the arid and semi-arid regions of Australia, which generally has a lower degree of accuracy than the humid/coastal regions of

Australia. The RFFA methods presented here for the arid and semi-arid regions should be used with caution; local hydrological knowledge must be exercised to interpret the results. It would be necessary to compare the results from these methods with other methods, such as rainfall runoff models and any locally available methods. The RFFA methods presented here will require further development and testing before they can be recommended for inclusion in the revised Australian Rainfall and Runoff.

## ACKNOWLEDGMENTS

The authors would like to acknowledge the Office of Water (New South Wales), Department of Sustainability and Environment (Victoria), Department of Environment and Resource Management (Queensland), Department of Water, Land and Biodiversity Conservation (South Australia) and NRETAS (the NT) for providing the streamflow data; the Bureau of Meteorology for providing climatic data CDs; Associate Professor James Ball, Mr. Mark Babister, Dr. William Weeks, Dr David Kemp and Mr Lakshman Rajaratnam for their suggestions and input to the project. This study is a part of ARR Revisions (Project 5), aimed at the development of more accurate regional flood estimation methods for Australia, which is being undertaken in association with Engineers Australia's National Committee on Water Engineering. Thanks to Engineers Australia and the Federal Department of Climate Change for providing financial support for the project.

## REFERENCES

- Benson, M.A. (1962). Evolution of methods for evaluating the occurrence of floods, *U.S. Geological Survey Water Supply Paper*, 1580-A, 30.
- Ben-Zvi, M. and Ben-Zvi, A. (1971). Probability distribution of the flow events in the Negev, *Journal of Hydrology*, 14, 348-353.
- Ben-Zvi, A. and Cohen, O. (1975). Frequency and magnitude of flows in the Negev, *Catena*, 2, 193-199.
- Ben-Zvi, A. (1982). Flow events in the Negev - a regional quantitative model, *Water International*, 7, 127-133.
- Ben-Zvi, A. (1996). Quantitative prediction of runoff events in the Negev desert. In: Issar, A.S., Resnick, S.D. (Eds.), *Runoff, Infiltration and Subsurface Flow of Water in Arid and Semi-Arid Regions*. Kluwer, Dordrecht, Netherlands, 121-130.
- Dalrymple, T. (1960). Flood frequency methods. *Water Supply Paper 1543-A, US Geological Survey*, 11-51.
- Farquharson, F.A.K., Meigh, J.R. and Sutcliffe, J.V. (1992). Regional flood frequency analysis in arid and semi-arid areas, *Journal of Hydrology*, 138, 487-501.
- Hjalmarson, H.W. and Thomas, B.E. (1992). New look at regional flood frequency relations for arid lands, *Journal of Hydraulic Engineering*, ASCE, 118, 868-886.
- Hosking, J.R.M and Wallis, J.R. (1993). Some Statistics Useful in Regional Frequency Analysis, *Water Resources Research*, 29(2), 271-281.
- Institute of Engineers, Australia (I.E. Aust.) (1987). Australian Rainfall and Runoff: A guide to flood estimation, Pilgrim, HD (ed.), *The Institute of Engineers Australia*, Canberra.
- Madsen, H., Pearson, C.P. and Rosbjerg, D. (1997). Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events. 2. Regional modelling, *Water Resources Research*, 33, 759-769.
- Meirovich, L., Ben-Zvi, A. and Shentsis, I. (1998). Frequency and magnitude of runoff events in the arid Negev of Israel, *Journal of Hydrology*, 207, 204-219.
- Rosbjerg, D. (1985). Estimation in partial duration series with independent and dependent peak values, *Journal of Hydrology*, 76, 183-195.
- Thomas, D.M. and Benson, M.A. (1970). Generalization of streamflow characteristics from drainage basin characteristics. *US Geological Survey Water Supply Paper*, 1975.
- UNESCO, (1999). Flush floods in arid and semi-arid zones, *Technical Documents in Hydrology*, UNESCO, Paris, 23.