JFE JOURNAL OF FLOOD ENGINEERING
1(1) January–June 2010; pp. 67–76

REGIONAL FLOOD FREQUENCY ANALYSIS USING L-MOMENTS FOR NORTH KAROON BASIN, IRAN

Sattar Chavoshi Borujeni[1], Wan Nor Azmin Sulaiman[1] & Saeid Eslamian[2]
[1]*Dept. of Environmental Science, University Putra, Malaysia*
[2]*Dept. of Water Engineering, Isfahan University of Technology, Iran*

ABSTRACT: L-moment theory is increasingly being used in hydrology of floods for regional frequency analysis. The L-moments have been used for parameter estimation, homogeneity testing and selection of the regional distribution. Records of peak floods, observed in North-Karoon basin, are analyzed using five distributions: generalized logistic, generalized extreme value, lognormal, Pearson type III and generalized Pareto. Each of these three-parameters distributions are estimated by the L-moment method. The discordancy index and homogeneity testing show that 5 out of the 7 study sites belong to a homogenous region. Based on the L-moment ratios diagram and the goodness-of-fit measure, the three-parameter lognormal distribution is identified as the most appropriate distribution in the homogeneous study region. The regional peak flood estimates for each return period are obtained based on this distribution.

Keywords: L-moments, Regional Flood Frequency, Probability Distribution, Statistical Hydrology.

1. INTRODUCTION

Planning, designing and operation of any water resources projects such as dams, spillways, road and railway bridges, culverts, urban drainage systems, flood plain zoning and economic evaluation of flood protection projects are based on estimation of design flood. The reasonable estimation of flood has remained as one of the main problematic issues where hydrological data and information are limited. This is typical issue in many basins of Iran, where most rivers are ungauged. This problem is intensified due to shortness of records, incomplete records, and inaccuracy of flow rating curves. In such cases, regionalization can be very helpful in pooling flood data such that design flood estimations can be made at ungauged basins. In this paper the application of L-moment theory in regional flood frequency analysis of North-Karoon catchment is presented. The main objective of this research is providing flood quantiles with different return periods in the study area, that faces with mentioned problems (Chavoshi, 1998).

2. REGIONAL FLOOD FREQUENCY ANALYSIS

One of the most important areas of flood hydrology is flood frequency analysis. The main objective of flood frequency analysis is to estimate flood quantiles at a specific site. There are several methods for flood frequency analysis, which can be categorized into two main classes of at-site, and regional flood frequency analysis. Recent research on flood suggest regional analysis to make flood frequency estimates at sites that have no flow records or to improve estimates at sites with relatively short records. The statistical estimation of design floods by regionalization was first presented by Dalrymple (1960) then developed by others. It involves two major steps of classifying sites into similar hydrological regions, and regional estimation of flood quantiles at the site of interest. A review of different methods of index flood was conducted by Bocchiola *et al.* (2003).

2.1. L-Moments Approach

The approach based on the theory of L-moments was firstly proposed by Wallis (1989) then developed by Hosking and Wallis (1997). L-moments are summary statistics for probability distributions and data samples. They are analogous to ordinary moments, because of providing measures of location, dispersion, skewness, kurtosis, and other aspects of the shape of probability distributions or data samples, but are computed from linear combinations of the ordered data values. L-moments may be applied in four steps for the regional frequency analysis including screening of the data, identification of homogeneous regions, choice of a frequency distribution and estimation of the frequency distribution (Hosking and Wallis, 1997). The main advantages of L-moments over conventional (product) moments are that they are able to characterize a wider range of distributions, and (when estimated from a sample) are less subject to bias in estimation and more robust to the presence of outliers in the data. The latter is because ordinary moments (unlike L-moments) require involution of the data which causes disproportionate weight to be given to the outlying values. The identification of a distribution from which the sample was drawn is more easily achieved (particularly for skewed distributions) using L-moments than conventional moments. Basically, L-moments are linear functions of probability weighted moments (PWMs).

$$Br = E\{X[F(x)]^r\} \tag{1}$$

where $F(x)$ is the cumulative distribution function of x. The first four L-moments expressed as linear combinations of PWM are:

$$\lambda_1 = \beta_0 \tag{2}$$

$$\lambda_2 = 2\beta_1 - \beta_0 \tag{3}$$

$$\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0 \tag{4}$$

$$\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \tag{5}$$

The L-mean, λ_1, is a measure of central tendency and the L-standard deviation, λ_2, is a measure of dispersion. Their ratio, λ_2 / λ_1, is termed the L-coefficient of variation, τ_2. The ratio λ_3 / λ_2 is referred to as τ_3 or L-skewness, while the ratio λ_4 / λ_2 is referred to as τ_4 or L-kurtosis.

2.2. A Homogeneity Test

In an attempt to regionalized frequency analysis, many studies have been carried out to identify homogeneous regions. There are three principal methods commonly used to delineate a study region into homogeneous regions. The first is based on geographical or administratively defined regions such as provincial, national, rivers, valley and latitude/longitude boundaries. The second is based on the similarity of basin and climate characteristics such as geology, land use, drainage characteristics and rainfall/runoff similarity. The third is based on flood characteristics such as the homogeneity test of Dalrymple (1960), similarity of the coefficient of variation (CV) test within the region (Cunnane, 1987) or a heterogeneity measure based on L-moments (Hosking and Wallis, 1993).

2.3. Discordance Test

A test of discordance is an L-moments-based test to determine weather a particular site, which does not appear to belong to the group of (T_3, T_4) points on the L-moment diagram, should be removed from the study area. In other words, a point which is far from the center of the diagram can be discarded.

Hosking and Wallis (1993) defined the discordance measure statistic, D-statistic, as:

$$\bar{U} = \frac{1}{N}\sum_{i=1}^{N} U_i \tag{6}$$

$$S = \frac{1}{N}\sum_{i=1}^{N} (U_i - \bar{U})(U_i - \bar{U})^T \tag{7}$$

$$D_i = \frac{1}{3}(U_i - \bar{U})^T S^{-1}(U_i - \bar{U}) \tag{8}$$

Where N is the total number of sites and U_i is defined as a vector containing the L-moment ratios for site i. $\bar{U}$ and S are the group averages and sample covariance matrix of u, respectively. Generally, sites with D-statistic greater than 3 are considered to be discordant from the rest of the region (Wallis, 1989).

2.4. Heterogeneity Test

For testing the regional homogeneity, a test statistic H, termed as heterogeneity measure was proposed by Hosking and Wallis (1993). It compares the inter-site variations in sample L-moments for the group of sites with what would be expected of a homogeneous region. The inter-site variation of L-moment ratio is measured as the standard deviation of the at-site L-CV's weighted proportionally to the record length at each site. To establish what would be expected of a homogeneous region, simulations are used. A number of 500 data regions are generated based on the regional weighted average statistics using a four-parameter distribution i.e. Kappa distribution. The inter-site variation of each generated region is computed and the mean (μ_v) and standard deviation (σ_v) of the computed inter-site variation is obtained. Then, heterogeneity measure H is computed as below (Kumar *et al.*, 2003).

$$H = V - \frac{\mu_v}{\delta_v} \tag{9}$$

and

$$V_1 = \sum_{i=1}^{N} \left\{ n_i \left[(Lcv_i - \bar{L}cv)^2 \right]^{1/2} \right\} / \sum_{i=1}^{N} n_i \tag{10}$$

$$V_2 = \sum_{i=1}^{N} \left\{ n_i \left[(Lcv_i - \bar{L}cv)^2 + (\tau_{3i} - \bar{\tau}_3\}^2 \right]^{1/2} \right\} / \sum_{i=1}^{N} n_i \tag{11}$$

$$V_3 = \sum_{i=1}^{N} \left\{ n_i \left[(\tau_{3i} - \bar{\tau}_3\}^2 + (\tau_{4i} - \bar{\tau}_4)^2 \right]^{1/2} \right\} / \sum_{i=1}^{N} n_i \tag{12}$$

The criteria established through Monte Carlo experiments by Hosking and Wallis (1993) for assessing heterogeneity of a region is as follows:

If $H < 1$, Region is acceptably homogeneous

If $1 \leq H < 2$, Region is possibly heterogeneous

If $H \geq 2$, Region is definitely heterogeneous

2.4. Goodness-of-Fit Test for Selecting the Best Distribution

This test is employed to select one of various unimodal distributions and to estimate its parameters. The goodness-of-fit criterion, Z-statistic, is defined as:

$$Z = (T_4 - \bar{T}_4 + \beta_4) / \delta_4 \tag{13}$$

$$\beta_4 = \frac{1}{N_{sim}} \sum_{m=1}^{N_{sim}} (T_{4m} - T_4) \tag{14}$$

$$\delta_4 = \sqrt{\frac{1}{N_{sim^{-1}}} \sum_{m=1}^{N_{sim}} (\bar{T}_{4m} - \bar{T}_4)^2 - N_{sim}\, \beta_4^2} \tag{15}$$

where β_4 and δ_4 are the bias and standard deviation of T_4, respectively. N_{sim} is the number of simulated regional data sets generated using a kappa distribution (Adamowski, 2000). The goodness-of-fit measure, Z, judges how well the simulated L-skewness and L-kurtosis of a fitted distribution matches the regional average L-skewness and L-kurtosis values obtained from the observed data (Isameldin *et al.*, 2006). A probability distribution with the smallest value of $|Z|$ is considered the best choice among possible distributions. At the significance level of $\alpha = 0.10$, the critical value of Z is 1.64, i.e., if a probability distribution whose $|Z| \leq 1.64$, then it is assessed to be an acceptable distribution for representing sample data at $\alpha = 0.10$.

2.5. L-Moments Ratio Diagram

An L-moment ratio diagram of L-Kurtosis versus L-skewness compares sample estimates of the dimensionless ratios T_3 and T_4 with their population counterparts for a range of statistical distributions. L-moment diagrams are useful for discribing grouping of sites with their similar flood frequency behavior and identifying the statistical distribution likely to adequately describe this behavior (Pearson, 1991).

3. APPLICATION

3.1. Data Set and Some Basic Characteristics of Study Area

North–Karoon basin, located between 49-34 to 51-47 longitudes and 31-18 to 32-40 latitude, with 14476 square kilometers area is the main sub-basin of the Great Karoon basin. It has a humid to semi-humid climate, high annual precipitation and associated discharge. There are several main rivers in the region such as Behesht-Abad, Koohrang, Sabz-Kooh, Vanak, Bazoft and Lordegan (Figure 1). Most part of the region is occupied with high-elevated mountains.

Figure 1: Location of North Karoon Catchment in Iran and Sites in the Study Area

The minimum, maximum and average elevation of the study area is 780, 4221 and 2279 meters, respectively. The longest river in the catchment with 137 kilometers length has divided it into two western and eastern parts. Main annual precipitation is due to the humid currents of Mediterranean, which affects the region for 8 months. The long-term maximum and minimum annual rate of precipitation is about 1600 and 300 millimeters, respectively. The average annual precipitation of the region is 707 millimeters while about 75 percent of the study area receives more than 500 millimeters annual rainfall and only 25 per cent of the catchment has more than 800 millimeters annual rainfall. The mean annual volume of precipitation is 10.2 billion cubic meters, which varies between 6.8 to 12.4. Mean annual evaporation is 1620 millimeters while in some years it gets to 2200 millimeters in south of the region. There are a number of 30 hydrometric sites in the study area, each of them with more than 5 years annual peak flood data. Among them, sites which are independent i.e. not subjected to upper catchments or any practices were selected for this study (Table 1).

Table 1
List of Studied Sites in the Region

Site	*River*	*Area (km^2)*	*Elevation (m)*
Zarinderakht	Khan Mirza	397	1770
Koohesookhteh	Kiar	2909	1980
Dezak	Biregan	630	2160
Tangedarkesh	Jooneghan	899	2000
Godarkabk	Agh Bolagh	716	2150
Marghak	Bazoft	34221	980
Lordegan	Lordegan	374	1650

3.2. L-Moment Software

In order to apply the L-moment method for analysis of data in the study area, the Fortran Package programmed by Hosking (1996) was used. This package has the ability to be applied for different stages of regional flood frequency analysis including screening of data, delineation of homogeneous region, tests of regional homogeneity, selection and estimation of regional frequency distribution and estimation of flow magnitude.

3.3. Checking on Homogeneity

Several methods have been proposed to identify homogeneous basins in the study area. Most of them use catchment's attributes to delineate pooled groups. In order to delineate a pooling group, all sites that have a high similarity with the site of interest are grouped together. The Euclidean distance in a multidimensional attribute space is widely used in similarity measures. The test of homogeneity of the study area has been studied using several methods such as Cluster Analysis and Andrew's Curve (Chavoshi, 1998).

3.4. Discordancy Test

The discordancy measure is considered as a means of screening analysis, and the aim is to identify those sites that are grossly discordant with the group as a whole (Isameldin *et al.*, 2006).

Discordancy measure for each site of the basin has been calculated and found between 0.30 to 2.08. The data range shows no discordancy for the study sites. However, the values of different heterogeneity measures, *H*1, *H*2 and *H*3 are found 5.93, 2.64 and 0.48, respectively. Since the first two heterogeneity measures, *H*1 and *H*2, are more than 2, dealing with heterogeneity of the studied sites, two suspected sites were removed and the process repeated. The results for the five remaining sites confirm that the rest of the sites may be considered as homogeneous region (Table 2).

Table 2
The L-moment's Properties and Discordance Measure

Site	*N*	*Name*	*L-CV*	*L-SKEW*	*L-KURT*	*D (I)*
1	15	Zarinderakht	0.4671	0.2864	0.1404	0.81
2	14	Koohesookhteh	0.3117	0.2744	0.1091	1.32
3	17	Dezak	0.4617	0.3993	0.2227	0.52
4	15	Tangedarkesh	0.2318	-0.0026	0.0927	1.28
5	21	Godarkabk	0.3226	0.3564	0.2551	1.06
Weighted means			0.3594	0.2728	0.1728	-
Parameters of regional Kappa distribution			0. 5539	0. 5447	-.0659	0. .3719

3.5. Homogeneity Testing Using the Heterogeneity Measure

The heterogeneity measure compares the between-site variations in sample L-moments for the group of sites with what would be expected for a homogeneous region (Isameldin *et al.*, 2006). Discordancy measure for each site of the basin has been calculated. The values of different heterogeneity measures H_1, H_2, H_3 are found 1.94, 0.84 and 0.28, respectively. Therefore, the study region demonstrates acceptable homogeneity.

3.6. Distribution Selection Using the Goodness-of-Fit Measure

After confirming the homogeneity of the study region, an appropriate distribution needs to be selected for the regional frequency analysis. In other words, in a homogeneous region all sites should have the same population L-moments. The selection was carried out by comparing the moments of the candidate distributions to the average moments statistics derived from the regional data. The best fit to the observed data indicates the most appropriate distribution (Yongqin et al., 2006). A number of five three-parameter distributions, i.e. Generalized Logistic (GL), Generalized Extreme Value (GEV), Generalized Pareto (GP), General Normal (LNIII)

Table 3
***Z*–Statistic for Various Distributions**

Distribution	*L-KURTOSIS*	*Z–statistic*
GEN. LOGISTIC	L-KURTOSIS = 0.229	Z VALUE = 0.88 *
GEN. EXTREME VALUE	L-KURTOSIS = 0.199	Z VALUE = 0.35 *
Log NORMAL III	L-KURTOSIS = 0.181	Z VALUE = 0.02 *
PEARSON TYPE III	L-KURTOSIS = 0.150	Z VALUE = -0.54 *
GEN. PARETO	L-KURTOSIS = 0.123	Z VALUE = -1.03 *

and Pearson Type III (PE3) were fitted to the region. The value of ZDIST statistic for the study area for each three-parameter distribution is presented in the Table 3. It can be seen that all of the candidates are acceptable; however LNIII is the most appropriate one.

3.7. L-moment Ratio Diagram

In an attempt to assign the sites to regions i.e. a set of sites whose flood frequency distributions are approximately the same, the diagrams of L-moment ratio were constructed (Figure 2). An advantage of L-moment diagrams is that one can compare the fit of several distributions using a single graphical instrument. Figure 2, compares the theoretical relationship between L-kurtosis and L-skewness for the Exponential, Uniform, Normal, Gumbel, Generalized Logistic (GL), Generalized Extreme Value (GEV), Log-normal 3 (LN3), Pearson Type 3 (P3), Generalized Pareto (GP) and lower bound of Wakeby. Figure 2, is a plot of L-moments of streamflow in real space; hence, it is not possible to represent the Log-Pearson Type 3 distribution. In this diagram, sites are illustrated as the circle points. The L-moment ratios diagram is also a very effective tool for distribution selection. It shows that the point for regional average L-moments $\tau_3 = 0.2728$ and $\tau_4 = 0.1728$, lies closest to the LN3, which supports our selection.

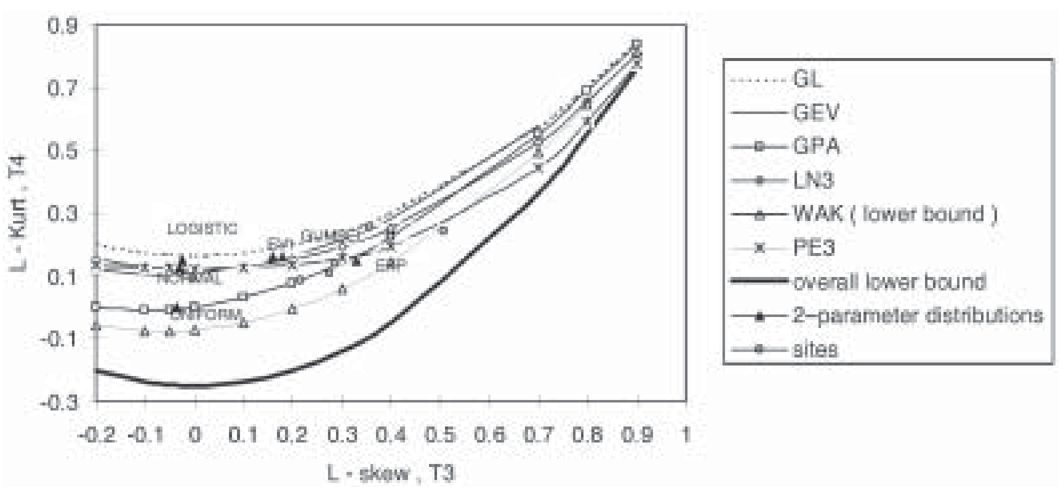

Figure 2: L-moment Ratios of Some Common Statistical Distributions in the Study Area

3.8. Regional Flood Quantile Estimation

The next step in regional flood frequency is to estimate flood quantiles in the region. In this paper, flood quantiles for each distribution are presented at the 90 percent level (Table 4).

Moreover estimated parameters of each distribution are presented (Table 5). The regional parameters of the Wakeby distribution have been included in the table because it has five parameters which is more than most of the common distributions so it can attain a wider range of distributional shapes than the common distributions. This makes the Wakeby distribution particularly useful for simulating artificial data for use in studying the robustness, under changes in distributional form of methods of data analysis (Kumar *et al.*, 2003).

Table 4
Quantile Estimates with Different Probabilities for Distributions Accepted at the 90% Significance Level

Probability	*0.010*	*0.020*	*0.050*	*0.100*	*0.200*	*0.500*	*0.900*	*0.950*	*0.990*	*0.999*
GL	0.014	0.084	0.203	0.321	0.479	0.845	1.799	2.277	3.753	7.330
GEV	0.069	0.126	0.223	0.323	0.466	0.834	1.851	2.325	3.614	6.090
* LN3	0.110	0.154	0.234	0.322	0.456	0.828	1.878	2.343	3.523	5.521
PE3	0.188	0.207	0.254	0.319	0.437	0.820	1.918	2.360	3.360	4.754
GEP	0.239	0.248	0.275	0.322	0.423	0.811	1.957	2.375	3.200	4.096
WAKEBY	0.162	0.180	0.230	0.308	0.448	0.829	1.894	2.362	3.480	5.164

* –The most appropriate distribution in the study area.

Table 5
Parameters of the Distribution

Distribution	*Parameters*
GEN. LOGISTIC	$\xi = 0.845$ $\alpha = 0.317$ $k = -0.273$
GEN. EXTREME VALUE	$\xi = 0.668$ $\alpha = 0.440$ $k = -0.154$
LOG NORMAL III	$\xi = 0.828$ $\alpha = 0.557$ $k = -0.568$
PEARSON TYPE III	$\mu = 1.000$ $\sigma = 0.692$ $\gamma = 1.640$
GEN. PARETO	$\mu = 0.230$ $\sigma = 0.880$ $\gamma = 0.143$
WAKEBY	$\xi = 0.145$ $\alpha = 1.161$ $\beta = 4.570$ $\gamma = 0.630$ $\delta = 0.026$

4. SUMMARY AND CONCLUSION

In Iran, a number of studies has been carried out for estimation of design floods for various purposes by different organizations. The insufficient number of sites and its uneven spatial distribution, short record of flood, outliers and the years of no-flow cause these studies problematic. Furthermore, available data may not represent to the basin flood due to the changes in watershed characteristics, such as changing vegetation, and land use and urbanization. Hence, regionalization methods are applied to make estimates of flood statistics at ungauged catchments using regional characteristics. In cases where no point data are available for flood estimation, data may be used from gauged neighboring catchments or data from catchments with similar hydrologic condition. Previous study in North-Karoon (Chavoshi, 1998) addresses these problems and recommends new methodology to overcome them. The method of L-moments was employed to identify the possible regional probability distributions of peak discharge in North-Karoon basin.

Discordance measure for study sites indicates no discordancy; however, the values of different heterogeneity measures show that the region is heterogeneous. So, two suspected sites were removed and the process repeated. The results for the remained sites confirm that the rest of the sites are homogeneous. In order to select the parent distribution for the study area, L-moment ratio diagram was applied. The *Z*-statistic was also applied to select a possible distribution type of peak discharge in the region. The possible distribution types identified by the *Z*-statistic are Generalized Logistic, Generalized Extreme Value, Generalized Pareto, General Normal (LNIII) and Pearson Type III. Both goodness-of-fit analysis and L-moment ratio diagram analysis indicated that the three-parameter Log Normal distribution is suitable for flood frequency analysis in the study area. Therefore, the flood quantiles for different return periods were obtained which are especially useful for water resources planning, design, and management in ungauged watersheds. The methodology used in this study can be adopted for other regions provided that sufficient flood records are available.

5. ACKNOWLEDGEMENTS

Most of the flow data used in this study was supplied by Iranian Water Resources Research Company (TAMAB), Iran. The useful comments of two anonymous reviewers are gratefully acknowledged.

References

[1] Adamowski, K., Regional Analysis of Annual Maximum and Partial Duration Flood Data by Nonparametric and L-moment Methods, *J. Hydrology,* **229** (2000) 219-231.

[2] Bocchiola, D., De Michele C. and R. Rosso, Review of Recent Advances in Index Flood Estimation. *Hydrology and Earth System Sciences,* **7**(3) (2003) 283-296.

[3] Cunnane, C., Review of Statistical Methods for Flood Frequency Estimation, in *Hydrologic Frequency Modeling,* edited by V. P. Singh, (1987) 49-95.

[4] Chavoshi, S., *Regionalization of Flood Estimation in Arid Lands using Hybrid Method,* Master Thesis, Dept. of Natural Resources, Isfahan University of Technology, (1998) 117.

[5] Dalrymple, T., Flood Frequency Analysis, US Geological Survey Water Supply Paper, (1960) 543A.

[6] Hosking, J. R. M., and J. R. Wallis, Some Statistics useful in Regional Frequency Analysis, *J. Water Resource Resources,* **20** (1993) 271-281.

[7] Hosking, J. K. M., Fortran Routines for use with the Method of L-moments,Version 3, Research Report RC205225, IBM Research Division, Yorktown Heights, New York. (1996).

[8] Hosking, J. K. M., and J. R. Wallis, *Regional Frequency Analysis, an Approach Based on L-moments,* Cambridge University Press. (1997).

[9] Isameldin, A. Atiem and Nilgun B. Harmancioglu, Assessment of Regional Floods using L-Moments Approach: *The Case of the River Nile, Water Resources Management*, **20** (2006) 723–747.

[10] Kumar, R., Chatterjee, C., Sanjay Kumar, A., Lohani, K., and R. D. Singh, Development of Regional Flood Frequency Relationships using L-moments for Middle Ganga Plains Subzone 1(f) of India. *Water Resources Management,* **17** (2003) 243–257.

[11] Pearson, C. P., New Zealand Regional Flood Frequency Analysis using L-moments. *J. Hydrol,* **30**(2) (1991) 53-63.

[12] Wallis, J. R., Regional Frequency Studies using L-moments. IBM Research Report RC14597 (1989) 17.

[13] Yongqin, D. C., Guoru, H., Quanxi S., and Y. X. Chong, Regional Analysis of Low Flow using L-moments for Dongjiang Basin, South China. *Hydrological Sciences Journal–des Sciences Hydrologiques,* **51**(6) (2006) 1051-1064.