9th International Conference on Hydroinformatics
HIC 2010, Tianjin, CHINA

THE USE OF HISTORICAL FLOOD INFORMATION IN THE IOWA RIVER BASIN TO IMPROVE RISK ASSESSMENT

MEI XUEFEI

PhD candidate, Faculty of Civil Engineering and Geosciences, Delft University of technology; P.O.Box5048, 2600GA Delft, The Netherlands

P.H.A.J.M. VAN GELDER

Associate Professor, Faculty of Civil Engineering and Geosciences, Delft University of technology; P.O.Box5048, 2600GA Delft, The Netherlands

J.K.VRIJLING

Professor, Faculty of Civil Engineering and Geosciences, Delft University of technology; P.O.Box5048, 2600GA Delft, The Netherlands

Abstract: Iowa River, a tributary of the Mississippi River, is one of the main sources of flood in the Iowa State. The timescale of gauging record without regulation at Iowa City is 56 years. The 1918 flood is the largest flood event in the gauged record. Flood frequency analysis, using gauged records only with Maximum likelihood estimation (MLE) and L-moment estimation methods (LME), place the 1918 event at a return period of around 100 years. However, a review of historical (pre-gauged) floods gives a different perspective. This paper demonstrates the value of augmenting a gauged river flow record with historical data for flood frequency analysis. The L-moment estimation method, considering historical floods is applied to the parameters of Generalized Extreme Value (GEV) distribution and Generalized Pareto (GPD) distribution. The result shows that incorporation of historical data into flood frequency analysis increases the reliability of risk assessment and ultimately provides a better basis for planning decisions. Meanwhile, comparison between Maximum likelihood estimation method and L-moment estimation method is also carried out. The results suggest these two methods are comparable to the time series in Iowa City.

Key words: Flood risk; frequency analysis; historical flood data; L-moments; GEV; GPD

1 Introduction

Flood frequency analysis is important for designing flood related structures. The designer generally needs to determine the design flood of a given return period T. Traditional methods for determining design flood have been generally based on data from the systematic record alone. However, the gauged record of many sites has typically short duration while sites with long records are usually affected by human activity and natural elements. The small sample may not be representative of the population of floods and flood frequency analysis may yield unreliable estimates of the extreme flood.

In order to extend discharge records in time and improve the reliability of flood estimates, historical information is suggested to incorporate with the gauged flood record. During the last two decades, the value of using nonsystematic data in flood frequency analysis has received considerable attention. Recently, the method of L-moments estimation cooperate with historical information was developed by Chen,et al [1] , as an improved alternative to L-moment. This method based on GEV and GPD are adopted and compared in this paper to analyze extreme flood risk.

The paper is organized as follows. The methods in this paper will be applied to a case study in Iowa State, the USA, as described in Sec.2. In Sec.3, the data analysis will be presented. Data fitting and comparison metrics are followed in Sec.4 and 5. Sec.6 will list related result. Conclusions and recommendations will be given at last section as an ending.

2 Brief description of the watershed

The geographic focus of this study is the Iowa River (shown in Fig.1), a tributary of the Mississippi River with about 480 km length, which is one of major rivers in Iowa State. It proceeds roughly in a southeast direction and becomes impounded by the Coralville Dam in the Coralville Reservoir. The river runs generally south and passes through Iowa City, about 105 km from its mouth. South of Iowa City, the Iowa River is joined by its longest tributary, the Cedar River.

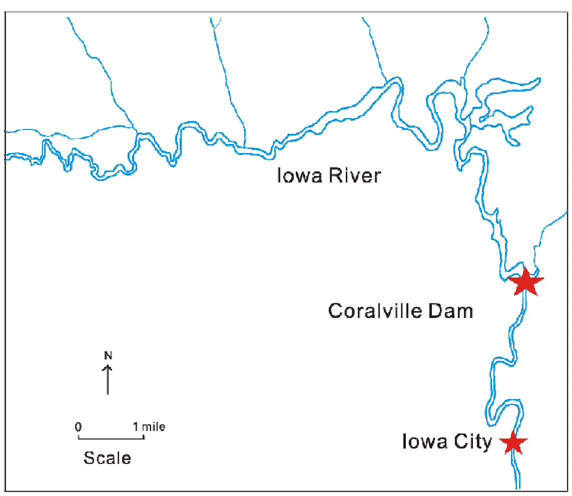

Fig1. Map of Iowa River

The data on the peak flows with historical information of Iowa River is analyzed in this paper to assess the affections of historical flood on flood frequency. Fig.2 shows the relatively long time series of the measured annual peak discharge, which has a significantly change since 1959 for the construction of Coralville Dam. The data are taken from the U.S.Geological Survey (USGS) Web sites.

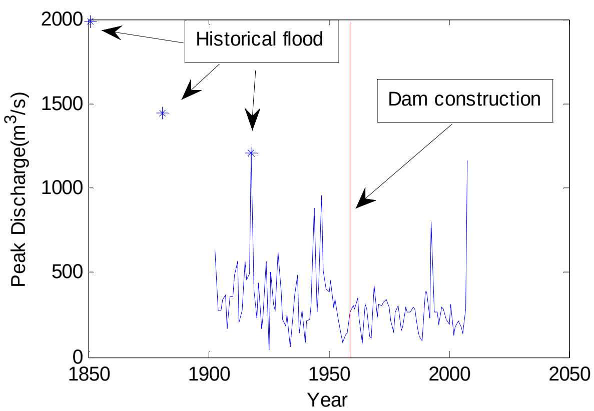

Fig2. Annual peak discharge of Iowa River at Iowa City

3 Data analysis

The behavior and course of a river may change considerably in long time period due to artificial or natural causes. When undertaking historical reviews, research must be undertaken to ensure that all observations from hydrologic events are come from the same population to be able to accurately perform frequency analysis. This study emphasizes test of randomness and homogeneities (both gradual trends and abrupt changes).

3.1 Test of Randomness

The runs test [2] is adopted in this paper to ask whether a data series comes from a random series. The null hypothesis that random process is rejected if the calculated u value is greater than the selected critical value obtained from the standard normal distribution table at significance level of 0.05. The result is shown in table1.

3.2 Test of homogeneities

3.2.1 Trend test

It is clearly from Fig.2 that the Iowa River was affected by the dam construction since 1959. As a result, our analysis time series has to reduce to 1903-1958 to do further test.

The Mann-Kendall trend test [3, 4] is applied to evaluate the absence of linear trend. The null hypothesis that there is no trend is accepted at significance level of 0.05 if the calculated z value is less than 1.96. In other words, the discharge series is suggested to be no trend. See the result in Table2.

3.2.2 Change test

Rapid shift may manifest when the time series transit from one state to another due to artificial or natural causes. For decision making, the Standard Normal Homogeneity Test (SNHT) is performed on annual series in this study to detect whether any changes have occurred. SNHT uses a statistic T_0 to compare the mean of the first a years of the record with that of the last $n-a$ years. T_0 will be small if the null hypothesis H_0 is true, whereas large value of T_0 make rejection of H_0 more probable. A possible shift is located at the year A when T_0 reaches a maximum at the year a=A [5]. The result is list in Table2.

4 Data fitting

4.1 Function distribution

A number of probability distribution functions have been utilized for flood frequency analysis. For instance, the log Pearson type 3 is a commonly used distribution in the United States [6], the general extreme value (GEV) distribution is recommended in the United Kingdom [7], and the Pearson type III distribution with a graphical curve fitting method is suggested by the Ministry of Water Resources, China [8]. In this study the there-parameter GEV distributions and Generalized Pareto (GPD) distribution are assumed as the desirable distribution of annual maximum floods.

4.1.1 General extreme value distribution

GEV [9] is widely used for modeling extremes of natural phenomena, which combines into a single form of the three possible types of limiting distribution for extreme values, as derived by Fisher and Tippett[10]. The distribution function is

$$
\begin{aligned}
F(x) &= \exp[-\{1-k(x-\xi)/\alpha\}^{1/k}] \quad k \neq 0 \\
&= \exp[-\exp\{-(x-\xi)/\alpha\}] \quad k = 0
\end{aligned}
\tag{1}
$$

With x bounded by $\xi + \alpha / k$ from above if $k > 0$ and from below if $k < 0$. Here ξ and α are location and scale parameters, respectively, and the shape parameter k determines which extreme value distribution is represented.

4.1.2 Generalized Pareto distribution

GPD is another approach widely used in the field of hydrology for extreme value analysis, which models the peaks of a time series exceeding a threshold. GPD studies in particular an extension of the classical extreme value analysis, instead of just annual maxima, several of the largest order statistics exceeding a sufficiently high threshold in the collected data [11]. The distribution function of GPD is given as:

$$F(x) = 1 - [1 - k(\frac{x-\beta}{\alpha})]^{1/k} \quad k \neq 0$$
$$= 1 - \exp[-(\frac{x-\beta}{\alpha})] \quad k = 0 \tag{2}$$

Where β, α and k denote the threshold, scale and shape parameters, respectively.

4.2 Parameter estimation for GEV and GPD

4.2.1 L-moment parameter estimate

L-moments are defined as linear combination of probability weighted moments (PWM) by Hosking [12]. The L-moment estimators have some desirable properties for parameter estimation. They are fast and straightforward to compute and always yield feasible values for the estimated parameters. The biases of the estimators are small. Method of L-moment estimation (LME) can often be used when the maximum likelihood estimation (MLE) are unavailable, difficult to compute, or have undesirable properties.

For an ordered sample $X_{1n} \leq X_{2n} \leq \ldots \leq X_{nn}$ unbiased estimators of the parameters for GEV are:

$$\hat{k} = 7.8590c + 2.9554c^2 \tag{3}$$

$$c = \frac{2b_1 - b_0}{3b_2 - b_0} - \frac{\log 2}{\log 3} \tag{4}$$

$$\alpha = \lambda_2 k / [(1 - 2^{-k})\Gamma(1+k)] \tag{5}$$

$$u = \lambda_1 - \alpha[(1 - \Gamma(1+k)] / k \tag{6}$$

Unbiased estimators of the parameters for GPD are:

$$a = (b_0 - \beta)(1 + \frac{b_0}{2b_1 - b_0} - \frac{\beta}{2b_1 - b_0} - 2) \tag{7}$$

$$\hat{k} = \frac{b_0}{2b_1 - b_0} - \frac{\beta}{2b_1 - b_0} - 2 \tag{8}$$

In this paper, double standard deviation is select as the threshold.

Unbiased sample estimators of the PWM are defined as:

$$b_r = \frac{1}{n} \sum_{j=1}^{n} \frac{(j-1)(j-2)\cdots(j-r)}{(n-1)(n-2)\cdots(n-r)} x_{j:n}$$

$$\lambda_1 = b_0 \tag{9}$$

$$\lambda_2 = 2b_1 - b_0 \tag{10}$$

4.2.2 L-moment parameter estimation with historical value (LMH)

Let X represent a series with historical flood in which the maximum return period is N; the length of recorded series(systematic) is n, the number of historical floods is a, the number of historical floods in the recorded series is l, $\{x_m, m = 1, 2, \cdots, (n + a - l)\}$ is the series in ascending order (from lowest to highest). The formulae of the probability weighted moments are as follow [1]:

$$b_0 = \frac{1}{N}\left[\frac{N-a}{n_0-a}\sum_{m=1}^{n_0-a} x_m + \sum_{m=n_0-a+1}^{n_0} x_m\right] \tag{11}$$

$$b_1 = \frac{1}{N}\left[\frac{N-a}{n_0-a}\sum_{m=1}^{n_0-a}\frac{(m-1)}{(n_0-a-1)}\frac{(N-a-1)}{(N-1)}x_m + \sum_{m=n_0-a+1}^{n_0}\frac{(N-n_0+m-1)}{(N-1)}x_m\right] \tag{12}$$

$$b_2 = \frac{1}{N}\left[\begin{array}{l}\frac{N-a}{n_0-a}\sum_{m=1}^{n_0-a}\frac{(m-1)}{(n_0-a-1)}\frac{(m-2)}{(n_0-a-2)}\frac{(N-a-1)}{(N-1)}\frac{(N-a-2)}{(N-2)}x_m \\ +\sum_{m=n_0-a+1}^{n_0}\frac{(N-n_0+m-1)}{(N-1)}\frac{(N-n_0+m-2)}{(N-2)}x_m\end{array}\right] \tag{13}$$

5. Data comparison metrics

For comparison estimation methods with empirical data sets used in the study, related coefficient(r), mean absolute deviation (MAD) and mean square deviation (MSD) proposed by Jain and Sing [13] was taken into account. Comparisons were made between computed flood discharges $\hat{Q}_i(T)$ and data (assumed true) values $Q_i(T)$ for the observed floods.

6. Results

The test results related to randomness and homogeneity of the peak flows on Iowa City site are given in Table 1 and 2, which suggest that the peak discharges between 1903 and 1958 are random and homogeneous while significant decline and change are tested around 1959 for the construction of Coralville Dam.

Table 1 Results for randomness test (at significance level of 5%)

Time	n_1	n_2	n	r	u	u*
1903-1958	25	31	56	32	0.907	1.96
1959-2008	23	27	50	31	1.48	1.96
1903-2008	35	71	106	63	3.34	1.96

Table 2 Results for homogeneity test

Time	Trend test		Change test	
	t	T*(95%)	t	T*(90%)
1903-1958	-0.002	1.96	2.462	7.34
1959-2008	-0.695	1.96	0.031	7.25
1903-2008	-2.930	1.96	8.376	7.87

Performance of the MLE, LME and LMH on GEV and GPD tested by r, MAD and MSD are presented in Table 3. The results reply that the LME is comparable with MLE for the discharge samples while GPD performs better than GEV on the whole. Although the performance of LMH is not so satisfied, it is still acceptable for its uncontinuity.

Table 3 Fitting Results for different models

Distribution	r			MSD			MAD		
	MLE	LME	LMH	MLE	LME	LMH	MLE	LME	LMH
GEV	0.9981	0.9985	0.9983	1.36	5.44	45.70	3.92	3.92	12.30
GPD	0.9971	0.9969	0.9951	0.33	3.93	3.81	0.99	2.48	5.68

In order to compare the methods more differently, Table 4 gives the quantile estimates for the return periods equal to 100.

Table 4 Comparison of 100-year flood quantile estimates for different models (m^3/s)

Distribution	MLE	LME	LMH
GEV	1085	821	736
GPD	1172	912	836
AVERAGE	1129	867	786

In the case of the River Iowa, proposed design flood based on historical information is 786 m^3/s averagely, which is 21.24% lower than just based on the gauged record. As for the extreme flood (1204 m^3/s) in 1918, it is ranked 3 in place the original 1 when extend the flood time from 56 to 108 years. Conclusion may be obtained from table 4 that the return period of 1918 extreme flood should be longer than now supported by actual data.

7 Conclusions and recommendations

Historical records may provide a more realistic assessment of the risk of extreme flooding than short periods of gauged flow as they could provide additional insights into the occurrence of flooding. However, the application of historical records should be undertaken with care and in relation to catchment changes. Meanwhile the ability to incorporate these data into quantitative flood estimates needs further research.

REFERENCES

[1] Chen Yuanfang, Xu Shengbin, Sha Zhigui, Van Gelder Pieter, Gu Shenghua, Study on L-moment Estimations for Log-normal Distributions with Historical Flood Data, in: GIS&RS in hydrology, Water Resources and Environment, Volume1, Chen et al.(eds), Sun Yat-sen University Press, 2003, ISBN 7-306-02142-7/P24.

[2] Bradley, James (1968). Distribution-free statistical tests. Englewood Cliffs, NJ: Prentice-Hall. Chapter 12.

[3] Mann,H.B.(1945). Nonparametric test against trend. Econometrica 13, 245-259.

[4] Kendall, M.G.(1975).Rank correlation methods. 4th ed. Charles Griffin, London

[5] Alexandersson,H. and Moberg, A.,1997. Homogenization of Swedish temperature data, Part I: Homogeneity test for linear trends. Int. J.Climatol., 17(1),25-34.

[6] Water Resources Coucil. 1981. Guidelines for determing floods flow frequency. Bull. 17B, Hydrol. Comm., Washington, DC,USA.

[7] Natural Environment Research Council. 1975. Flood Studies Reports. Natural Environmental Research Council, London.

[8] Ministry of water resources. 1980. Standard methods for flood quantile estimation procedure. SDJ 22-79, Water Resources Publishing House, Beijing, China (In Chinese).

[9] Embrechts, P., Kluppelberg, K., Miklosh, T. (1997). Modelling extremal events. Springer: Berlin, Heidelberg.

[10] Fisher, R.A., Tippett, L.H.C. 1928, Limiting forms of the frequency distribution of the largest or smallest member of a sample. Proceedings of the Cambridge Philosophical Society, v. 24, 180-190.

[11] M.D. Pandey , P.H.A.J.M. Van Gelder , J.K. Vrijling. 2001. The estimation of extreme quantiles of wind velocity using L-moments in the peaks-over-threshold approach . Structural Safety 23 (2001) 179–192

[12] Hosking. 1990.L-moments: Analysis and Estimation of Distribution using Linear Combinations of Order Statistics", Journal of the Royal Statistical Society, Series B, 52, pp. 105-124.

[13] Jain, D., Singh, V.P., 1987. Comparison of some flood frequency distributions using empirical data. In: Singh, V.P., (Ed.), Hydrologic Frequency Modeling, Reidel, Dordrecht, pp.467–485