Maximum entropy method for flood frequency analysis: A case study of the Grand River in Ontario, Canada

J Deng¹,²
¹Department of Civil Engineering, Lakehead University, Thunder Bay, ON, P7B 5E1, Canada
E-mail: jian.deng@lakeheadu.ca

Abstract. Flood frequency analysis provides the fundamental information for the design of hydraulic and civil infrastructure facilities as well as water resources management. The available procedures to determine flood distributions and their parameters from historical data suffer from restriction on the family of presumed standard classical distributions, such as gamma distribution and extreme value distributions. This paper presents a distribution-free approach for flood frequency analysis by using the combination of maximum entropy method and Akaike’s information criterion. The determination procedure is given in a flowchart. It is distribution-free because no classical distributions were presumed in advance and the inference result gives a universal form of probability curves. Comparative studies from historical flood data show that the proposed method can, accurately and reasonably, characterize the probabilistic information of observed or transformed flood data.

1. Introduction
The primary objective of flood frequency analysis is to determine flood quantiles corresponding to probability of non-exceedance (or return periods) of specific relevance by using the theory of probability and statistics. The estimation of flood recurrence can be used in designing various civil structures such as urban drainage, culverts, road embankments, bridges, spillways, and dams [1]. In order to prevent under-designing or over-designing, it is imperative to perform accurate flood frequency analysis statistically due to ubiquitous randomness in hydrology.

In practice, the true probability distribution of the flood at a region or a site was usually unknown. The type of probability distribution was selected from existing classical distributions such as lognormal, normal, Weibull, Gumbel, and Pearson. The distribution parameters were estimated from observed sample data. The fitted distribution was then used to calculate the magnitudes of flood frequencies. It has been criticized, however, that the selection of the initial probability distribution type in the traditional methods is always arbitrary, and there is rarely any verification or justification [2]. The Bayesian methods also involve the assumption of the prior parametric distribution and the selection of the type of probability distribution as well.

An alternative technique is based on the maximum entropy method, which gives a universal form of probability distributions. The maximum entropy is based on Shannon’s information entropy, which has been adopted in several disciplines of engineering for estimating distribution functions. The entropy is defined as an uncertainty measure of information associated with a random variable. The methodology is to determine a probability density function that maximizes Shannon’s entropy subjected to constraints in terms of sample moments [3].
Entropy theory has been successfully applied in probabilistic modelling of environmental and water engineering [4]. A cross entropy method was used efficiently for flood frequency analysis [2]. The maximum entropy method was employed to estimate the probabilistic distribution parameters [5]. More recently, the method of Halphen distribution, coupled with the principle of maximum entropy (ME), was presented for flood frequency analysis [6]. However, all these studies have not solved the problem of optimal order of probabilistic models, i.e., how many orders of mathematical moments one should use. It was Baker [7] who presented a procedure for estimation of probability density function based on information concepts that researchers joined together Jaynes’ maximum entropy method with Akaike’s information criterion for the selection of the optimal member of a group of model order. However, Baker did not use statistical testing to establish and justify the underlying distribution. Later, Deng and Pandey [8] applied the maximum entropy method to determine the quantile function based on fractional probability weighted moments.

In the present paper, Jaynes’ maximum entropy method and Akaike’s information criterion are combined to determine how many and what statistical moments from a sample have to be used to determine the probability density function of floods. A case study of flood frequency analysis on the Grand River in Ontario, Canada, was performed as an illustration.

2. Methodology
First, a probabilistic function is set up to quantify the random property of the flood data based on the maximum entropy method. The result is a family of functions with respect to the order of moments being used. The probabilistic function is distribution-free because it is based on sample moments only, and no classic distributions are presumed a priori. The Akaike’s information criterion is then applied to determine the optimal order of the probabilistic function. The whole procedure is to generate an optimal probabilistic distribution function in accordance with the quantity and nature of the available sample data.

2.1. Maximum entropy method
In information theory, entropy is defined as a measure of the uncertainty involved in a random variable. Mathematically, entropy is given as

$$H = - \int_R f(x) \ln [f(x)] dx,$$

(1)

where $X$ is a continuous random variable in the domain $R$ and $f(x)$ is the probability density function. To obtain a probability distribution, it is necessary to maximizes the entropy constrained by moments of a random variable. The formulation of the maximum entropy method is presented as follows

$$\text{maximize } H = - \int_R f(x) \ln [f(x)] dx,$$

subject to $\int_R x^k f(x) dx = \mu_k, \ k = 0, 1, \cdots, K,$

(2)

where $K$ is the highest order of moments, $\mu_k$ is the $k$th moment. $\mu_0 = 1$ represents the normalization condition. To solve the maximization problem in equation (2), an entropy function is formulated as

$$\bar{H} = -H + (\lambda_0 - 1) \int_R f(x) dx + \sum_{k=1}^{K} \lambda_k \left[ \int_R x^k f(x) dx - \mu_k \right],$$

(3)

where $\lambda_k$ is the Lagrangian multiplier. $(\lambda_0 - 1)$ is used as the coefficient instead of $\lambda_0$ for ease of calculation. The Lagrangian maximization demands

$$\frac{\partial \bar{H}}{\partial f(x)} = 0,$$

which results in

$$f_K(x) = \exp \left[ - \sum_{k=0}^{K} \lambda_k x^k \right].$$

(4)

As $f_K(x)$ is a function of $\lambda_k (k = 0, 1, \cdots, K)$, it can also be written as $f_K(x, \lambda)$. If the moments are
estimated from a sample data,
\[
\mu_k = \frac{1}{N} \sum_{i=1}^{N} x_i^k ,
\]

then the \( f_K(x) \) is an approximate of the real probability density function \( f(x) \) in the random space, where \( x_i (i = 1, 2, \cdots, N) \) signifies \( N \) measured sample values, and \( N \) is the sample size. Substituting equation (4) into the term \( \ln \left[ f(x) \right] \) of equation (1) and considering equation (2) yield an approximate of \( H \)
\[
\hat{H}[f(x)] = \sum_{k=0}^{K} \lambda_k \mu_k .
\]

The \( K + 1 \) Lagrange parameters, \( \lambda_k \), can be obtained by solving a set of \( K + 1 \) nonlinear equations
\[
\int x^j \exp \left[ -\sum_{k=0}^{K} \lambda_k x^k \right] dx = \mu_j , \quad j = 0, 1, 2, \cdots, K .
\]

The maximum entropy method has been considered as a rational principle to select a consistent probability distribution that contains minimum spurious information [3,5]. It is seen that the above procedure to determine \( f(x) \) is completely based upon the available sample information and no any presumptions were made about the potential probabilistic distributions, thus, the inference shall result in a universal form of probability models. In the next section, Akaike’s information criterion is utilized to decide how many order of sample moments one should use in the above procedure.

2.2. Akaike’s information criterion

A measure of distance between \( f_K(x) \) and \( f(x) \) is represented in the Kullback-Leibler (KL) entropy
\[
KL[f(x), f_K(x, \lambda)] = \int_R f(x) \ln \frac{f(x)}{f_K(x, \lambda)} dx = C - L(\lambda, K) ,
\]

\[
C = \int_R f(x) \ln f(x) dx , \quad L(\lambda, K) = \int_R f(x) \ln f_K(x, \lambda) dx .
\]

Here \( C \) is irrelevant of the model function \( f_K(x, \lambda) \) and therefore acts as a constant while minimizing the KL entropy with respect to \( \lambda \). The term \( L(\lambda, K) \) is the expected value of \( \ln f_K(x, \lambda) \), therefore from \( N \) measurements of a sample, one can obtain a natural estimate \( \hat{L}(\lambda, K) \) of \( L(\lambda, K) \)
\[
\hat{L}(\lambda, K) = \frac{1}{N} \sum_{i=1}^{N} \ln f_K(x_i, \lambda) , \quad \hat{KL}[f(x), f_K(x, \lambda)] = C - \hat{L}(\lambda, K) ,
\]

where \( x_i (i = 1, 2, \cdots, N) \) signifies \( N \) measured sample values, \( \hat{KL}[\lambda, K] \) is a sample estimate of the KL entropy \( KL[f(x), f_K(x, \lambda)] \). The best choice of \( \lambda \) is determined by minimizing \( \hat{KL} \) with respect to the vector of unknown parameters \( \lambda \),
\[
\min_{\lambda} \{ \hat{KL}[\lambda, K] \} = C + \min_{\lambda} \left\{ -\hat{L}(\lambda, K) = C + \min_{\lambda} \left\{ -\frac{1}{N} \sum_{i=1}^{N} \ln f_K(x_i, \lambda) \right\} \right\} .
\]

The term \( \sum_{i=1}^{N} \ln f_K(x_i, \lambda) \) is a log likelihood function. Baker [1990] introduced Akaike’s information criterion and suggested that one of the best estimate of \( \lambda \) can be obtained if we minimize an unbiased estimate of \( -\hat{L}(\lambda, K) \), instead of the natural estimate \( -\hat{L}(\lambda, K) \) of biased likelihood function. The unbiased estimate is given as,
\[
\hat{\Gamma}(\lambda, K) = -\hat{L}(\lambda, K) + \frac{K}{N} .
\]

When applying to the family of maximum entropy functions, one can substitute equation (4) into (10) and (12) to obtain the unbiased estimate of \( \hat{\Gamma}(\lambda, K) \)
\[
\hat{\Gamma}(\lambda, K) = \sum_{k=0}^{K} \lambda_k \left[ \frac{1}{N} \sum_{i=1}^{N} (x_i)^k \right] + \frac{K}{N} \sum_{k=0}^{K} \lambda_k \mu_k + \frac{K}{N} = \hat{H}[f(x)] + \frac{K}{N} .
\]

Given a sample and a specific order of \( K \), one can determine the unknown parameters \( \lambda_k \) of the
entropy distribution function by using the algorithm in Section 2.1. For a series of moment order numbers starting from 0, there does exist a number $K$ to minimize the term $\hat{f}(\lambda, K)$ in equation (12), which in turn minimizes the term $\tilde{KL}[f(x), f_K(x, \lambda)]$ in equation (11). Consequently, this $K$ is the optimal order of maximum entropy distribution. The distribution obtained by maximum entropy is said to be the most unbiased, as it is derived from a systematic minimization by using Akaike’s Information Criterion. The flowchart for maximum entropy method together with Akaike’s information criterion is given in figure 1, in which the conventional method is also provided for comparison.

![Flowchart for maximum entropy method together with Akaike’s information criterion.](image)

**Figure 1.** Flowchart for maximum entropy method together with Akaike’s information criterion.

### 3. Application to flood frequency analysis

In flood frequency analysis, three different data sets are usually considered: (i) the partial duration (PD) series or peaks over a threshold (POT) model, (ii) the annual maximum (AM) series model, and (iii) the time series model. The time series flood model is described best by a stochastic process in continuous time. The PD (or POT) model is criticized as dependent observations were involved. The AM model is statistically more efficient than the PD model when $m$ is small ($m < 1.65$) where $m$ is the mean number of peaks per year included in the PD series [1]. The following discussion is based on the historical annual maximum daily discharge data of the Grand River in Ontario, Canada [2], which is listed in table 1. The Grand river watershed, spanning a length of 290km, covers an area of 6965 km$^2$ and is the largest of the watersheds in Southwestern Ontario that drain into Lake Erie (figure 2 [9]).

| No. | Data | No. | Data | No. | Data | No. | Data | No. | Data |
|-----|------|-----|------|-----|------|-----|------|-----|------|
| 1   | 331  | 16  | 841  | 31  | 379  | 46  | 490  | 61  | 855  |
| 2   | 253  | 17  | 323  | 32  | 476  | 47  | 680  | 62  | 561.5|
| 3   | 558  | 18  | 351  | 33  | 530  | 48  | 173  | 63  | 501  |
| 4   | 719  | 19  | 653.5| 34  | 1040 | 49  | 311  | 64  | 445  |
| 5   | 493  | 20  | 439  | 35  | 1070 | 50  | 351.5| 65  | 428  |
| 6   | 685  | 21  | 430  | 36  | 657  | 51  | 127  | 66  | 697  |
| 7   | 459  | 22  | 527  | 37  | 759.5| 52  | 470  | 67  | 425  |
| 8   | 402  | 23  | 374  | 38  | 524  | 53  | 179  | 68  | 367  |
| 9   | 538  | 24  | 343  | 39  | 382  | 54  | 657.5| 69  | 478  |
| 10  | 498  | 25  | 561  | 40  | 357  | 55  | 264  | 70  | 248  |
| 11  | 456  | 26  | 654  | 41  | 1140 | 56  | 487  | 71  | 244  |
| 12  | 473  | 27  | 600  | 42  | 292  | 57  | 228  | 72  | 636  |
| 13  | 411  | 28  | 388  | 43  | 654.5| 58  | 345  | 73  | 564  |

**Table 1.** Annual maximum daily discharge of the Grand River, Ontario, Canada (m$^3$/s).
Table 2. Kullback-Leibler (KL) entropy of entropy distributions.

| Sample moment order | KL entropy, $\hat{f}(\lambda, K)$ |
|---------------------|----------------------------------|
| $K=1$               | -0.076                           |
| $K=2$               | -0.211                           |
| $K=3$               | -0.234                           |
| $K=4$ (optimum)     | -0.247                           |
| $K=5$               | -0.234                           |
| $K=6$               | -0.232                           |

Application of the maximum entropy method together with Akaike’s information criterion in Section 2 renders the Kullback-Leibler (KL) entropy for various orders of entropy distributions in Table 2. It is found that the smallest $KL$ entropy corresponds to the maximum entropy distribution of order $K=4$, which indicates that the following distribution is the most unbiased and optimal distribution as follows

$$f_K(x) = \exp\left[-6.87403633 - 11.92280057\times10^{-3}x + 72.84334585\times10^{-6}x^2 - 120.32851466\times10^{-9}x^3 + 57.29891827\times10^{-12}x^4\right].$$

Table 3. The parameters of Gumbel, lognormal and gamma distributions.

| Distribution | parameters | Maximum likelihood estimate |
|--------------|------------|-----------------------------|
| Gumbel       | $\mu$      | -405.410729                 |
|              | $\sigma$   | 172.927108                  |
| lognormal    | $\mu$      | 6.126279                    |
|              | $\sigma$   | 0.447489                    |
| gamma        | $a$        | 5.735789                    |
|              | $b$        | 87.291384                   |

For the purpose of comparison, three classical probabilistic distributions commonly used in flood
frequency analysis are fitted by the same flood data. They are Gumbel, lognormal and gamma distributions [2,5]. These distribution function parameters are estimated by the method of maximum likelihood, which are listed in table 3. The probability density function for Gumbel distribution or the extreme value type I with location parameter $\mu$ and scale parameter $\sigma$ is

$$f(x, \mu, \sigma) = \frac{1}{\sigma} \exp\left(\frac{x-\mu}{\sigma}\right) \exp\left(-\exp\left(\frac{x-\mu}{\sigma}\right)\right), -\infty < x < +\infty. \quad (15)$$

The gamma probability density function is

$$f(x, a, b) = \frac{1}{b^a \Gamma(a)} x^{a-1} \exp\left(-\frac{x}{b}\right), 0 \leq x < +\infty, \quad (16)$$

where $\Gamma(a)$ is the Gamma function, $a$ is a shape parameter, $b$ is a scale parameter.

The lognormal probability density function is

$$f(x, \mu, \sigma) = \frac{1}{\sigma x \sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right), x > 0. \quad (17)$$

Comparison of entropy density distribution with Gumbel, gamma and lognormal distributions are illustrated in figure 3. The flood frequency curves are depicted in figure 4, which can relate flood discharge values to probability of non-exceedance and then to provide an estimate of the intensity of a flood event. The observed frequency curve was obtained by using the Gringorten plotting position formula, because the Gringorten formula is known to be nearly unbiased in discharge for a variety of flood-like distributions [10]. It is noticed that all distributions fit the observed flood data good. However, for evaluating the predictive ability of the distributions, relative root mean square error (RRMSE) and relative absolute error (RAE) were introduced as follows:

$$\text{RRMSE} = \frac{1}{N-1} \sum_{k=1}^{N} \left(\frac{Q_{oi} - Q_{ci}}{Q_{oi}}\right)^2 \right)^{1/2}; \quad \text{RAE} = \frac{1}{N} \sum_{k=1}^{N} \left|\frac{Q_{oi} - Q_{ci}}{Q_{oi}}\right|, \quad (18)$$

in which $N$ is the sample size; $Q_{oi}$ is the $i$th observed annual peak discharge of the $i$th given probability; and $Q_{ci}$ is the $i$th computed annual peak discharge of the same probability. The smaller RRMSE and RAE values represent a better performance of the distribution [5]. Table 4 shows that the entropy distribution with order $k=4$ boasts the smallest with both RRMSE and RAE, and thus is the best distribution among all the distributions being considered.
Figure 3. Comparison of entropy distribution with Gumbel, lognormal and gamma distributions.

![Figure 3](image-url)

Figure 4. Flood frequency curves for the Grand river.

Table 4. The RRMSE and RAE of the fitted distributions.

| Distribution  | RRMSE | RAE   |
|---------------|-------|-------|
| Gumbel        | 3.091 | 2.512 |
| lognormal     | 0.0978| 0.0634|
| gamma         | 0.0659| 0.0406|
| Entropy (K=4) | 0.0372(smallest) | 0.0268(smallest) |

One advantage of entropy distribution is that one can adjust the distribution’s sophistication by adopting different order of sample moments. Only the entropy distribution that minimizes the unbiased KL entropy in equation (13) is able to adequately and optimally characterize the sample information. The term \( K/N \) is able to avoid adopting too elaborate models which cannot be validated by the sample data, as it is directly proportional to the model number \( K \) but is inversely proportional to the sample length \( N \). The modern world is immersed in information explosion. Therefore, the Akaike’s information criterion can parsimoniously retain only useful information and abandon the irrelevance. Since the entropy distribution is an exponential function shown in equation (4), the present method seems difficult to model a distribution with a sharp spike shape at the top point.

4. Conclusions
This paper presents a distribution-free approach for flood frequency analysis using the maximum entropy method together with Akaike’s information criterion. It is distribution-free because no classical distributions were presumed in advance, the derivation was based on sample moments only, and the inference result provides a universal form of probability curves. The most unbiased entropy distribution is the one having the smallest Kullback-Leibler entropy. The determination procedure is given in a flowchart. As an application example, the historical annual maximum daily discharge data of the Grand river in Ontario, Canada was considered for the flood frequency curves. Three classical probabilistic distributions commonly used in flood frequency analysis were fitted by the same flood data for comparison, i.e., Gumbel, lognormal and gamma distributions. The entropy distribution
boasted the lowest the relative root mean square error (RRMSE) and relative absolute error (RAE), which suggests the distribution-free method can be robustly used in flood frequency analysis.

References
[1] Rao A R and Hamed K H 2000 *Flood Frequency Analysis* (New York: CRC Press)
[2] Lind N C, Hong H P and Solana V 1989 A Cross entropy method for flood frequency analysis *Stoch. Hydrol. Hydraulic.* 3 191-202
[3] Shore J E and Johnson R W 1980 Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy *IEEE T. Inform. Theory* 26 26-37
[4] Singh V P 2013 *Entropy Theory and Its Applications in Environmental and Water Engineering* (New York: John Wiley)
[5] Singh V P 1998 *Entropy-Based Parameter Estimation in Hydrology* (Netherlands: Springer)
[6] Xiong F, Guo S, Chen L, Yin J and Liu P 2018 Flood frequency analysis using Halphen distribution and maximum entropy *J. Hydrol. Eng.* 23 04018012
[7] Baker D R 1990 Probability estimation and information principles *Struct. Saf.* 9 97-116
[8] Deng J and Pandey M D 2008 Estimation of the maximum entropy quantile function using fractional probability weighted moments *Struct. Saf.* 30 307-19
[9] Hanief A and Laursen A E 2019 Modeling the natural drainage network of the grand river in southern *Ontario. Geosci.* 9 46
[10] Hirsch R M 1987 Probability plotting position formulas for flood records with historical information *J. Hydrol.* 96 18-199