Transactions on Ecology and the Environment vol 8, © 1994 WIT Press, www.witpress.com, ISSN 1743-3541

Ajuste II: hydrological frequency analysis software PART 1: theoretical aspects

L. Perreault, B. Bobée, H. Perron
INRS-Eau, University of Québec, Institut national de la recherche scientifique, NSERC – Hydro-Québec, Chair on statistical hydrology, 2800, rue Einstein, Sainte-Foy (Québec), Canada G1X 4N8

Abstract

Design of hydraulic structures, optimal operation of reservoirs, and flood forecasting involve estimation of flood flows X_T, with given prespecified return period. We present here some theoretical aspects concerning the fitting of statistical distributions to annual maximal flows, which is a privileged tool for determining X_T and its confidence intervals. We also present the Halphen distributions of type A, B and B^{-1} and a synthesis of the most recent theoretical developments about them.

Introduction

Engineering activities such as the design of hydraulic structures, the management of water resources systems, and the prevention of flood damage, all require an estimation of flood characteristics. This estimation must be as precise as possible because over-estimation can lead to a large increase in costs, while an under-estimation can lead to high risk of flood damage and to the loss of human life.

Hydrological frequency analysis deals with the estimation of flood characteristics such as the probability of occurrence of a given flood, or of the flood corresponding to a given return period.

To assist engineers in performing hydrological frequency analysis, we have developed the software ***Ajuste II***. This publication presents theoretical aspects of that software, while companion paper presents its software aspects and an application.

Definition of risk and return period

The main objective of hydrological frequency analysis is to establish the relation between extreme hydrological events (floods, droughts, etc.) and their exceedance (or non-exceedance) probability. An **extreme hydrological event** is a situation that can lead to a risk. In hydrology, such an event is related to a **return period *T***. The extreme hydrological event and the interpretation of the return period depend on the type of risk and the random variable considered .

For example, to determine the size of a hydraulic structure, it is important to evaluate the risk of failure (overflow). Let us suppose that the maximum annual flood corresponds to a random variable X having a probability density function (p.d.f.) denoted $f(x;\underline{\theta})$, where $\underline{\theta} = (\theta_1, \theta_2, ..., \theta_p)$ is a set of parameters. The cumulative distribution function (c.d.f) is given by :

$$F(x;\underline{\theta}) = \mathrm{Prob}\{X \leq x\} = \int_{-\infty}^{x} f(t;\underline{\theta})dt \qquad (1.)$$

and hydrological risk defined as the probability for a maximum annual flood X to be superior to some critical value x_c:

$$\mathrm{p} = \mathrm{Prob}\{X > x_c\} = 1 - F(x_c;\underline{\theta}) \qquad (2.)$$

Let us consider the following two complementary and mutually exclusive events:

$$A = \{X > x_c\} \text{ and } \overline{A} = \{X \leq x_c\} \qquad (3.)$$

where A represents a failure event (an overflow for example), and $\overline{A}$ a non-failure event. These two events describe a Bernoulli trial with :

$$\mathrm{Prob}\{A\} = p \text{ and } \mathrm{Prob}\{\overline{A}\} = 1 - p \qquad (4.)$$

If Z is the time (in years) between two failure events $A = \{X > x_c\}$, we can show that Z is a random variable distributed as a geometric distribution with p.d.f. $\mathrm{Prob}(Z = z) = (1-p)^{z-1}p$, (z = 1, 2, 3, ...). We can deduce that:

$$E[Z] = \frac{1}{p} = T, \qquad (5.)$$

where $E[Z] = T$ is the mean time between two events $A = \{X > x_c\}$. T is called the **return period** for the event $A = \{X > x_c\}$. If we define the critical value in terms of the return period T ($x_T = x_c$), we obtain from equations (2) and (5) :

$$p = \mathrm{Prob}\{X > x_T\} = 1 - F(x_T;\underline{\theta}) = \frac{1}{T} \qquad (6.)$$

Transactions on Ecology and the Environment vol 8, © 1994 WIT Press, www.witpress.com, ISSN 1743-3541

The value x_T is called the **quantile of return period *T*** and is a function of the parameters $\underline{\theta}$, the cumulative distribution function F and T given by :

$$x_T = F^{-1}\left(1-\frac{1}{T};\underline{\theta}\right) \tag{7.}$$

where F^{-1} is the inverse of the cumulative distribution function F.

In practice, x_T is estimated by fitting a statistical distribution to a sample of maximum annual floods constructed from historical records of daily flows. To apply this methodology, the sample has to meet some statistical hypotheses. In particular, the observations must be independant and identically distributed (iid):

a) independence means no autocorrelation and can be verified by Wald-Wolfowitz test [11];

b) identically distributed means that the observations come from the same statistical population $F(x;\underline{\theta})$, which implies :
   - Stationarity (Kendall test [8]);
   - Homogeneity (Mann and Whitney test [9]); and
   - Absence of outliers (discordancy test adapted to each distribution, for example the Grubbs and Beck test [6] for the normal distribution).

When the observations are independent and identically distributed, it is possible to obtain an estimate $\tilde{x}_T$ of x_T (given by eqn 7.) using the equation :

$$\tilde{x}_T = F^{-1}\left(1-\frac{1}{T};\tilde{\underline{\theta}}\right) \tag{8.}$$

Hence, it suffices to estimate the distribution parameters (cf. section on parameter estimation) by $\tilde{\underline{\theta}}$ and then use equation (8.) in order to obtain the estimation $\tilde{x}_T$ of the unknown value x_T.

Adaptation to droughts

Low flows (droughts) may affect hydropower production which may be insufficient to satisfy the demand. To manage adequately the hydropower production, it is necessary to understand this type of event. We must evaluate the non-exceedance probability of the critical value y_d.

In that case, hydrological risk is the probability for minimum flow Y to be inferior or equal to that critical value y_d, which is the probability that the event $B=\{Y \le y_d\}$ occurs. If we assume that the minimum annual flow Y is also distributed according to the distribution F, and that prob$\{B\}$ = 1 - π, it is possible to define (in an entirely analogous manner than for maximum annual flood) the hydrological risk and the annual minimum flow of return period T :

$$\text{Prob}\{Y \le y_d\} = F(y_d;\underline{\theta}) = 1 - \pi \qquad (9.)$$

If we consider the critical value in terms of the return period T ($y_d = y_T$), we obtain :

$$1 - \pi = \text{Prob}\{Y \le y_T\} = F(y_T;\underline{\theta}) = \frac{1}{T} \qquad (10.)$$

and

$$\tilde{y}_T = F^{-1}\left(\frac{1}{T};\underline{\tilde{\theta}}\right) \qquad (11.)$$

Table 1 summarizes the principal output from frequency analysis and illustrates the correspondence between flood and drought events.

Table 1. Output from frequency analysis.

Extreme event	Random variable	Failure	Risk	Flow of return period T
Flood	X	$A = (X \ge x_c)$	$p = \frac{1}{T}$	$\tilde{x}_T = F^{-1}\left(1 - \frac{1}{T};\underline{\tilde{\theta}}\right)$
Drought	Y	$B = (Y \le y_d)$	$1\text{-}\pi = \frac{1}{T}$	$\tilde{y}_T = F^{-1}\left(\frac{1}{T};\underline{\tilde{\theta}}\right)$

Parameter estimation

Usually, it is difficult, if not impossible, to identify exactly the form of the cumulative distribution function F of the data. The theoretical moments and the quantiles for a hydrological random variable are unknown. However, estimation of the parameters of a given distribution F can be obtained from a series of observations ($x_1, x_2, \ldots x_n$) of the random variable X. It is then possible to estimate the unknown quantiles x_T.

Many parameter estimation methods exist for most common distributions used in hydrology. In particular, the following can be considered :

- the maximum likelihood method;
- the methods of moments :
    - Direct method of moments (for example for log-Pearson type III, Bobée [2]) ;
    - Sundry averages method (Bobée [3]), which use moments of order; -1, 0 and 1;

Transactions on Ecology and the Environment vol 8, © 1994 WIT Press, www.witpress.com, ISSN 1743-3541

- Indirect method of moments on the Log of the observations (for example for log-Pearson type III, WRC [13]).

Many studies have been conducted to compare the estimation methods for a given distribution. We note in particular, the study related to the gamma family (Gamma, Log Gamma, Pearson, Log Pearson, Generalized Gamma) by Messaoudi [10]. Based on that study, we have chosen the estimation methods in function of their predictive ability (Cunnane [5]) :

Table 2 shows the estimation methods we have implemented for each distribution included in ***Ajuste II***..

Table 2. List of distributions and estimation methods in *Ajuste II*

Distributions	Estimation methods
Gamma	Maximum Likelihood, Method of moments
Generalized Gamma	Maximum Likelihood, Method of moments
Pearson Type 3	Maximum Likelihood, Method of moments
Log-Pearson Type 3	Direct method of moments, Sundry averages method, Method of moments on the Log of the observations
Inverted Gamma	Maximum likelihood
Generalised extreme value	Maximum Likelihood, Method of moments, Probability weighted moments
Gumbel	Maximum Likelihood, Method of moments
Weibull	Maximum Likelihood, Method of moments
Halphen (type A, B and B^{-1})	Maximum Likelihood
Normal	Maximum Likelihood
Log-normal (2 parameters)	Maximum Likelihood
Log-normal (3 parameters)	Maximum Likelihood, Method of moments
Exponential (2 parameters)	Maximum Likelihood

Halphen distributions

One of the most interesting characteristics of ***Ajuste II*** is the possibility to fit the Halphen distributions [7] These distributions, denoted type A, B and B^{-1}, have been presented by Bobée et al. [4] and are summarized here.

The p.d.f. for the type A is given by :

$$h_A(x;\alpha,m,\upsilon)=\frac{1}{2m^{\upsilon}K_{\upsilon}(2\alpha)}x^{\upsilon-1}\exp\left[-\alpha\left(\frac{x}{m}+\frac{m}{x}\right)\right],\qquad x>0 \qquad (12.)$$

where υ and $\alpha > 0$ are shape parameters, and $m > 0$ is a scale parameter. The constant $K_\upsilon(2\alpha)$ is a modified Bessel function of the second kind (Watson, 1966).

For the type B, the p.d.f. is given by :

$$h_B(x;\alpha,m,\upsilon) = \frac{2}{m^{2\upsilon} ef_\upsilon(\alpha)} x^{2\upsilon-1} \exp\left[-\left(\frac{x}{m}\right)^2 + \left(\frac{\alpha x}{m}\right)\right], \qquad x > 0 \qquad (13.)$$

where the constant $ef_\upsilon(\alpha)$ is the exponential factorial function defined by Halphen as :

$$ef_\upsilon(\alpha) = 2\int_0^{+\infty} \exp\left[-x^2 + \alpha x\right] \cdot x^{2\upsilon-1} dx \qquad (14.)$$

and for the type B^{-1}, the p.d.f. is given by :

$$h_{B^{-1}}(x;\alpha,m,\upsilon) = \frac{2m^{2\upsilon}}{ef_\upsilon(\alpha)} x^{-2\upsilon-1} \exp\left[-\left(\frac{m}{x}\right)^2 + \frac{\alpha m}{x}\right], \qquad x > 0 \qquad (15.)$$

One of the statistical characteristics of those distributions is that they possess joint sufficient and complete statistics, which lead to the existence of unbiased estimators with minimum variance for the parameters. Given that property, it can be shown (Bickel et Doksum [1]) that optimal estimators exist and are functions of the maximum likelihood estimators. The Halphen distributions are the only three parameter distribution used in hydrology that have that statistical characteristic.

For the type A, maximum likelihood estimators are the unique solution of the system :

$$\begin{cases} m\dfrac{K_{\upsilon+1}(2\alpha)}{K_\upsilon(2\alpha)} = A \\ \dfrac{1}{m}\left(\dfrac{K_{\upsilon-1}(2\alpha)}{K_\upsilon(2\alpha)}\right) = \dfrac{1}{H} \\ \ln m + \left(\dfrac{1}{K_\upsilon(2\alpha)}\right)\left(\dfrac{\partial K_\upsilon(2\alpha)}{\partial \upsilon}\right) = \ln G \end{cases} \qquad (16.)$$

where A is the arithmetic mean, H the harmonic mean, and G the geometric mean.

The systems of ML-equations associated to the Halphen distributions are non-linear and must be solved using iterative procedures. They involve the Bessel

Transactions on Ecology and the Environment vol 8, © 1994 WIT Press, www.witpress.com, ISSN 1743-3541

and exponential factorial functions and are difficult to manipulate. Numerical problems pertaining to the estimation of the parameters α and υ have led us to consider a two step estimation. This approach consists first to obtain some estimates $\tilde{\alpha}$ and $\tilde{m}$ of the parameters α and m, using a fixed parameter υ. Then, we maximize the partial maximum likelihood function $L(\upsilon; \tilde{\alpha}, \tilde{m})$ to determine an optimal value for $\tilde{\upsilon}$ and obtain the corresponding optimal values of $\tilde{\alpha}$ and $\tilde{m}$.

Finaly, since m is a scale parameter, it is possible to show that if $X > 0$ is Halphen of type A, B or B^{-1} distributed, then the random variable $Y=kX$ (where $k<0$) is of the same distribution. That is, if $X \sim h_i(x;\alpha,m,\upsilon)$ then $Y = kX \sim h_i(x;\alpha,km,\upsilon)$.

It is also possible to show that there exists a general relation between the distributions of type B and B^{-1} such that if $X \sim h_{B-1}(x;\alpha,m,\upsilon)$ then $Y = 1/X \sim h_B(x;\alpha,1/m,\upsilon)$

Conclusions

Hydrological frequency analysis is an important tool for engineers to design hydraulic structures, to manage water resources systems, and to prevent flood damage. This analysis must be as precise as possible to avoid large increase in costs and loss of human life.

In this paper, we have presented some theoretical aspects of hydrological frequency analysis, in particular how to establish the relation between extreme hydrological events (floods, droughts, etc.) and their exceedance (or non-exceedance) probability. We have shown how to estimate the critical value x_T associated with a return period T.

Also, we have presented some theoretical aspect of the Halphen distributions. Those distributions are complex since they involve the Bessel and exponential factorial functions. The systems of ML-equations are non-linear and must be solved using Newton-Raphson method. Furthermore, there is no explicit function for the estimation $\tilde{x}_T$ of x_T. Therefore, numerical approximations have to be used.

To help engineers perform hydrological frequency analysis, we have developed the software ***Ajuste II***. In the second part of this paper ("***Ajuste II*** : Hydrological frequency analysis software, Part 2 : Software aspects and applications") we present some features of that software, how we have solved some numerical problems encountered during the development, and an application.

Transactions on Ecology and the Environment vol 8, © 1994 WIT Press, www.witpress.com, ISSN 1743-3541

References

1. Bickel, P.J. and Doksum K.A. (1977). Mathematical Statistics : Basic Ideas and Selected Topics. Holden-Day.
2. Bobée, B. (1975). The log-Pearson type 3 distribution and its application in hydrology. Wat. Res. Res.11(5): 68-689
3. Bobée, B. (1988). The generalized method of moments applied to the log-Pearson 3 distribution. Journal of Hydraulic Engineering, ASCE, **114**(8): 899-909
4. Bobée, B., Rasmussen, P.F., Perreault, L., and Ashkar, F. (1993). Risk analysis of hydrologic data : Review and new developments. To appear in the proceedings of the Nato Advanced Study Institute on "Engineering Risk and Reliability in a Changing Physical Environment", Deauville, France, May 23 - June 5, 1993.
5. Cunnane, C. (1987). Review of statistical methods for flood frequency estimation. In V.P. Singh (ed.), Hydrologic Frequency Modeling, D. Reidel, Dordrecht, pp. 49-95.
6. Grubbs, F. and Beck G. (1972). Extension of sample sizes and percentage points for significance tests of outlying observations. Technometrics, 14(4): 847-854.
7. Halphen, E. (1941). Sur un nouveau type de courbe de fréquence. Comptes rendus de l'Académie des Sciences, tome 213:634-635. (Published under the name of D. Dugué due to the war)
8. Kendall, M.G. (1975). Rank Correlation Methods. Charles Griffin, London.
9. Mann, H.B. and Whitney, D.R. (1947). On the test of Whether one of two random variables is stochastically larger than the other. Ann. Math. Statist., 18: 50-60.
10. Messaoudi, H. (1994). Comparaison des méthodes d'estimation des distributions de la famille Gamma. PhD thesis at INRS-Eau (in preparation).
11. Wald , A. and Wolfowitz, J. (1943). An exact test for randomness in the non-parametric case based on serial correlation. Ann. Math. Statist., 14: 378-388.
12. Watson, G.N., (1966). A Treatise on the Theory of Bessel Functions. Cambridge University Press.
13. WRC (1967). Guidelines for Determining Flood Flow Frequency. US Water Resources Council Hydrology Committee, Washington.