Chapter 10

An Update on Consolidated Frequency Analysis

Kumaraswamy Ponnambalam and Ravi M. Phatarfod

Frequency analysis is necessary to estimate the exceedance probability of flood flows or low flows. The return periods are calculated from these exceedance probabilities. In this chapter, a state-of-the-art computer program available for the Windows environment is described. The program can extract 7 or 30-day average low flows or r -maximum floods in a year, r ranging from 1 to 12, from Canadian Daily Streamflow Records. The frequency analysis can be carried out for various commonly used probability distribution functions. A brief description of an alternative flow analysis based on deriving a common form for all distributions is also described in this chapter. The result of this alternative analysis provides, in the case of flood flows for example, the expected value of the maximum flood for given lengths of time.

10.1 Introduction

The magnitude and frequency of occurrence of extreme events (floods, droughts or low flows) affect a hydrologic system adversely. Analysis of such events is necessary for both the design and the operation of hydrologic systems. Frequency analysis (FA) can be used to estimate the exceedance of certain flows, that is $P(X \geq x_T)$ where x_T is the flow value, and T is the return period. The return period T is defined as the average recurrence interval between events, where the

Ponnambalam, K. and R.M. Phatarfod. 1996. "An Update on Consolidated Frequency Analysis."

Journal of Water Management Modeling R191-10. doi: 10.14796/JWMM.R191-10.

© CHI 1996 www.chijournal.org ISSN: 2292-6062 (Formerly in *Advances in Modeling the Management of Stormwater Impacts*. ISBN: 0-9697422-5-8)

event may be defined as the flow equaling or exceeding a given magnitude x_T (Chow et al., 1988). The above analysis is suitable for flood flows. Low flow is characterized by an index such as 7-day 20 year low flow (7Q20) which is the discharge having a 20-year return period derived from a frequency analysis of the lowest average flow for seven consecutive days in a year (Logan, 1980).

We have developed a MS-Windows based computer software program (Figure 10.1) that performs the following:

1. analysis of daily flow record to determine the sequence of either the required maximum flows or low flows, and
2. frequency analysis using a selected probability distribution function. The results are tabulated and graphed in a single screen.

Because the program is truly Windows-based, the results can be easily transferred to other Windows programs such as a word processor, spreadsheet, among others.

Although frequency analysis provides necessary probabilities and return periods, the exact application of the results to hydrologic system design and operation is still debatable for many reasons. One debate concerns the appropriateness of the selected probability distribution function; another is the meaning of return periods. However, practitioners rarely have to worry about these questions because of existing standards in each country, that stipulate the use of certain distributions and return periods depending upon the type of hydrologic systems involved. We briefly discuss an alternative analysis of floods (similar ideas can be applied for low flows) that provides an estimate of the expected maximum flood in a given length of time [Phatarfod, 1995].

10.2 State of the Art

Methodology

Current state-of-the-art FA requires:

1. that random flows be collected from an existing record,
2. adoption of a probability distribution function from any one of the most likely distributions, for example, Gumbel, Lognormal, Log Pearson Type III,
3. selection of a parameter estimation method, most likely from the method of moments, maximum-likelihood method, probability weighted moments method, L-moment method, etc.,
4. a fit to the collected data in (1.) with the combination of steps (2.) and (3.), and
5. use of extrapolation or interpolation to compute flows x_T of required return period. The reader may refer to any one of the hydrology text books (for example, Chow et al., 1988) for the description of the

methodology, or Kite (1977) for a description with computer programs.

Computer Program

Many computer programs are available that implement the above FA although the authors have not found one that implements both the preprocessing necessary for collecting the random flows from daily flow records available from the HYDAT database of Environment Canada and the required FA. Moreover, most existing programs do not take advantage of the Windows environment. Therefore, we decided to implement these two features in a single program, and have developed it in the Windows environment. Due to space limitations we present just a single screen (Figure 10.1) that shows the standard Windows-type menu with both numerical results and the graph shown in a single screen. The program is under continuing development to add additional features.

Figure 10.1 Screen shot of final results.

10.3 Critique of State-of-the-Art Methodologies

The FA methodology described above has been in practice for many decades despite the short-comings that are described below. The main reason for using FA, in spite of the well-known short-comings, is the lack of alternative

techniques. In the next section an alternative technique to FA is proposed. In due time, the alternative technique will also be available in a single computer program with FA as described above.

The four main short-comings of FA are as follows.

1. The connection between the x_T and the design horizon is not direct. That is, the value of x_T is such that, in any year, the probability of flow exceeding x_T is $1/T$. If N is the design horizon, then the probability of flow exceeding x_T at least once during the design horizon is $1 - (1 - 1/T)^N$. This value, even for large T (rare events), and reasonable N is very large. Moreover, this value is rarely used in combination with design costs unlike the probability of exceedance.
2. The expected probability of exceedance is usually greater than $1/T$ (Chow et al., 1988).
3. Arbitrary choice in the probability distribution function and a somewhat arbitrary choice in the choice of method of fitting the distribution (including in the plotting positions when graphing the fitted data with computed distribution).
4. The confidence intervals for estimate of flows of a given return period depends heavily on the size of the data. These short-comings are well known and hence are not discussed in detail.

10.4 An Alternative to Frequency Analysis

Much research time in the past has been spent in overcoming the last two short-comings of FA with no definitive conclusions. We attempt to solve the problem of design and extreme events by asking an alternative question that will be useful for design. The alternative question is (from here on we use maximum flows as an example) *why not determine the expected maximum value in n samples of independent flows?* That is, if $x_{(n)}$ = the maximum flow in n samples, then we want to determine $\mu_n = E[x_{(n)}]$. Note that the n here simply means independent samples and therefore there can be r samples per year. Secondly, using the equation for μ_n given below, we can extrapolate for large values of n . The major advantage is that, for all practical purposes, for most commonly used distributions, there is a single equation for μ_n , which is, $\mu_n = a + b [\log n]^c$ where a , b , and c are parameters that will be determined from a regression analysis on the data (Phatarfod, 1995). Secondly, the parameters are found using a common regression analysis, a single and robust method, and hence confidence intervals are easily derived using a standard analysis.

Table 10.1 presents the actual form of the equation for commonly used distributions. The equations for μ_n in Table 10.1 were theoretically derived and were statistically verified as discussed in the next section and in Phatarfod [1995].

The theory can prove the validity of the equations for $n = 50$ to 5000. But statistical verification has shown good estimation even for higher values of n .

Table 10.1 Commonly used distributions and form of μ_n

Distribution Name	Distribution	Form of μ_n
Exponential	$\exp(-x)$	$\mu_n = \gamma + \log n$
EV1 (Gumbel)	$\exp(\exp(-x))$	$\mu_n = \gamma + \log n$
Logistic	$\exp(-x)/(1+\exp(-x))^2$	$\mu_n = \gamma + \log n$
Lognormal	$\exp(-.5(\log x)^2)/(x \sqrt{\pi})$	$\mu_n = a + b[\log n]^c$ $c=2$ (high skew) and 1 (low skew)
Gamma and Pearson Type 3	$x^{\beta-1}\exp(-x)/\Gamma(\beta)$	$\mu_n = a + b[\log n]$

Note that for exponential distribution Gumbel (1958) proved $\mu_n = \gamma + \log n$ where γ is Euler's constant. An example derivation for EV1 (Gumbel) is presented below.

The density function for EV1 is

$$f(x) = \exp(-x - e^{-x})$$

and the CDF is

$$F(x) = \exp(-e^{-x})$$

Therefore, the CDF of $x_{(n)}$ is given by

$$\begin{aligned} F_n(x) &= \exp(-ne^{-x}) \\ &= \exp(-e^{-(x-b_n)}) \end{aligned}$$

where $b_n = \log n$. That is, b_n displaces the distribution to the right. However, $E(x) = \gamma$ and hence $E[x_{(n)}] = \gamma + \log n$.

10.5 Results

We conducted the following experiment to test the validity of the equations for μ_n . For space limitations we present results corresponding to a gamma distribution with shape factor (β)=3 only. We generated short ($n=500$ or $n=1000$) samples and using a regression analysis we estimated the parameters a and b (Table 10.2). Note that with a single sample, permutation was used to shuffle the samples to get various estimates of a and b . This is justified because the individual values in the sample are independent. The mean values of the parameters were used in extrapolating to $n=5000$. That is, using $n=500$ and the equation for μ_n

extrapolation is used to estimate μ_n for $n=5000$ and we compared this result with an actual value generated from $n=5000$. Similar results were obtained for $n=10000$. The above experiments were conducted 20 times to get $\min(\mu_n)$, $\max(\mu_n)$ and $E(\mu_n)$. Note that we show here the minimum and maximum values of μ_n to demonstrate the limits of the method but 95% confidence limits were closer and are usually sufficient for all practical purposes. Table 10.2 presents the results.

Table 10.2 Actual and predicted values of $\text{Min}(\mu_n)$, $E(\mu_n)$ and $\text{Max}(\mu_n)$ from 20 experiments.

n	$\text{Min}(\mu_n)$	$E(\mu_n)$	$\text{Max}(\mu_n)$	Remarks
5000	11.30	13.75	17.16	True
1000	10.32	14.08	19.92	Predicted for $n=5000$
500	10.49	14.63	20.63	Predicted for $n=5000$
10000	12.68	15.09	20.41	True
1000	10.08	14.37	19.52	Predicted for $n=10000$
500	11.28	15.77	20.17	Predicted for $n=10000$

One should compare these results for minimum, maximum, and expected values to the 90% confidence limits for x_T that are generally given. For example Example 12.6.1. in Chow et al. (1988) calculates these to be 16.2, 28.9, and 74.8, respectively for Pearson Type 3 distribution for just the 100-year return period which in any case depends heavily on the sample size.

10.6 Conclusions

A state-of-the-art programming tool for consolidated frequency analysis of floods and low flows was introduced. The program includes the necessary preprocessing steps. The disadvantages of currently available methodologies for floods was briefly described. An alternative technique, possibly complementary to existing methodologies for analyzing flows, was described. The advantage of this alternative technique is the universal relationship between flows and the sample size for most common distributions and the simplicity in estimating the parameters of this equation using common regression. But much research yet has to be done in order to understand the implications of this new technique.

Acknowledgments

The research on alternative methods to frequency analysis has been supported in part by the first author's NSERC research grant.

References

- Chow, V. T., D. R. Maidment, and L. W. Mays, 1988. *Applied Hydrology*, McGraw-Hill Book Comp., New York.
- Gumbel, E.J., 1958. *Statistics of Extremes*, Columbia University Press, p. 116.
- Kite, G. W., 1977. *Frequency and Risk Analysis in Hydrology*, Water Resources Publ., Fort Collins.
- Logan, L. A., 1980. *Low Flow Frequency Analysis for Waste Assimilation Studies*, Ontario Ministry of the Environment, Toronto.
- Phatarfod, R. M., 1995. Working Paper, Department of Mathematics, Monash University.