Stochastic Hydrol. Hydraul. 6 (1992) 239–254

Stochastic Hydrology and Hydraulics

The use of geometric and gamma-related distributions for frequency analysis of water deficit

L. Mathier, L. Perreault and **B. Bobée**

INRS-Eau, Université du Québec, Institut national de la recherche scientifique, Carrefour Molson, 2800 rue Einstein, Québec (Québec), Canada, G1X 4N8

F. Ashkar

Département de mathématiques, Université de Moncton, Moncton (Nouveau-Brunswick), Canada, E1A 3E9

Abstract: This paper presents an approach to perform statistical frequency analysis of water deficit duration and severity using respectively the geometric and exponential distributions. Monthly mean water discharges are compared to a given threshold and classified in two mutually exclusive ways. This leads to a two state random variable such that: a success represents the absence of a water deficit event (mean monthly discharge exceeds threshold), and a failure, a water deficit event (mean monthly discharge is below threshold). If we suppose that this random variable gives rise to a Markov process of order 1, then the duration of a water deficit event X (consecutive months in deficit) will have a geometric distribution. In turn, the summation of discharges in deficit will give the severity of a water deficit event which can be represented by a one-parameter exponential distribution. The threshold or base level is taken as a percentile of the observed mean discharges of a given month. This base level, which varies from month to month, can be viewed as the limit of an acceptable deficit (or energetic failure) associated to a given empirical probability of being in deficit. The second step of the approach is to estimate the value of the parameter for each distribution using the maximum likelihood method. Expressions for the estimator of a given percentile, $\hat{x}_q$, as well as its variance are deduced. Finally, the presented models are applied to observed data.

Key words: Water deficit, Geometric distribution, Exponential distribution, Deficit duration, Deficit severity

1 Introduction

Management of water resources is generally based on the analysis of extreme phenomena like floods and droughts. Due to a recent decrease in water availability, some hydroelectric power companies have expressed the need of a better knowledge of water deficits to prevent production drops and energetic failures. They are particularly interested in the statistical modelling of some hydrological characteristics of low flow events like duration and severity (cumulative volume in deficit), which will be defined more precisely later.

In the traditional approach of drought hazard assessment, low flows are defined as the smallest annual values of the mean daily discharges of a river (Gumbel 1963), or as the minimum average discharge for a given period (Matalas 1963), generally taken to be 1 day or 7 days. The lowest average flow for 7 consecutive days is the most widely used definition in the United States (TCLFE 1980). However, for the study of water deficit, this traditional approach is not useful because it does not consider the drought duration (Sen 1980). If one is interested in drought in terms of magnitude and duration simultaneously, the use of a threshold discharge as defined by Yevjevich (1967) is more suitable (Sen 1980,1982). In this approach a drought is defined as a period where discharges are less than a certain threshold discharge (Yevjevich 1967) usually taken as a constant base level. Table 1 displays some works related to the previously mentioned drought definitions, which are frequently used in the literature.

Because threshold discharges are purely conventional notions and there is no universal definition of them, the base level should be evaluated using hydrological justification or economic requirements (TCLFE 1980, Ozga-Zielinska 1989). A new definition of water deficit for the management of a power plant based on economic requirements has been presented by Mathier et al. (1990, 1991) according to the specific needs of Alcan company in Saguenay-Lac-Saint-Jean, Québec, Canada, for the management and planning of their hydroelectric system. Instead of defining droughts in a strict sense using the minimum discharge for n consecutive days (Gumbel 1963, Matalas 1963) or a constant threshold level over the whole period of observation (Yevjevich 1967), water deficit events were defined according to a threshold which varied from month to month using a fixed criterion representing the limit of an acceptable deficit (Mathier et al. 1990, 1991). An extension of this methodology using a more general definition of the base level is presented in this paper.

Several fitting techniques and parameter estimation methods have been used in the literature to analyze low flows. It will be beyond the scope of this paper to review all the available techniques. In the case of drought studies, it is recommended that the theoretical probability distributions have lower limits equal to or greater than zero and no more than three parameters (Matalas 1963). Good results of frequency analysis of drought magnitude variables (discharge, volume, severity, ...) have been obtained using the Gumbel (Gumbel 1963, Matalas 1963, Condie & Nix 1975, Condie & Cheng 1983), Gamma (Joseph 1970), Weibull (Joseph 1971, Loganathan et al. 1985, Pilon 1990), Pearson type 3 (Matalas 1963) and log Pearson type 3 (Hoang 1978, Loganathan et al. 1985, Tasker 1987, Vogel & Kroll 1989) distributions. For drought duration and severity defined by a constant base level on the streamflow hydrograph of instantaneous discharges (continuous scale), the exponential distribution gave good results (Zelenhasic & Salvai 1987).

As mentioned by Acreman (1990), a discrete distribution is required for the modelling of event duration, because this variable can only take positive integer values. It will be shown theoretically and verified empirically in this paper that water deficit duration has a geometric distribution. On the other hand, the one-parameter exponential distribution should be examined to describe water deficit severity as it is equivalent to the geometric distribution in the continuous domain (Johnson & Kotz 1969, Feller 1957) (both distributions have in common a constant hazard function). These two distributions with parameters estimated by the method of maximum likelihood have the advantage of fulfilling all previously quoted requirements in the frequency analysis of low flows (Matalas 1963). In the case of water deficit severity all the above mentioned distributions can be used at the expense of more parameters to be estimated. In a recent study Mathier et al. (1990) compared the Gamma, the generalized Gamma and the Peason type 3 distributions using the method of moments and maximum likelihood, and concluded that best results on severity, using these multi-parameter distributions, are obtained with the Gamma distribution with parameters estimated by the method of maximum likelihood.

The objective of this study is to estimate events x_q of a given duration or a given severity corresponding to a probability of exceedance q and to construct asymptotic confidence intervals for these events using respectively the geometric and exponential distributions.

For the purpose of frequency analysis of water deficit duration and severity, the maximum likelihood estimator $\hat{x}_q$ of the true but unknown value x_q and its variance will be given for each distribution. Finally, in a case study, distributions are fitted to observed data in order

Table 1. Drought definitions frequently used in the literature

Drought definition	References
Smallest annual values of the mean daily discharges.	(Gumbel 1963, Matalas, 1963, Condie & Nix 1975, Hoang 1978)
Smallest mean discharge for 7 consecutive days.	(Matalas 1963, Yonts 1971, Condie & Cheng 1983, Loganathan et al. 1985, Tasker 1987, Vogel & Kroll 1989, Pilon 1990)
Smallest mean discharge for 14 consecutive days.	(Joseph 1970, Joseph 1971)
Smallest mean discharge for n consecutive days.	(Lara 1970, Riggs 1972, CTGREF 1978, Prakash 1981)
Constant base level.	(Yevjevich 1967, Miquel & Phien Bou Pha 1978, Sen 1980, Sen 1982, Güven 1983, Paulson et al. 1985, Sadeghipour & Dracup 1985, Zelenhasic & Salvai 1987, Ozga-Zielinska 1989)

to estimate events corresponding to a given probability of exceedance q along with related asymptotic confidence intervals. The results of chi-square goodness-of-fit test for the geometric distribution and Kolmogorov-Smirnov test for the exponential are also presented.

2 Data base, definition of the base level and water deficit variables

The data base is composed of daily discharges from 1943 to 1989 for a representative drainage basin of the Alcan company hydroelectric system in Saguenay-Lac-Saint-Jean, Québec, Canada.

Given the daily discharges q_{ijk} (i = day, j = month and k = year) and the mean discharge $\bar{Q}_j$ for each month j over the whole period, water deficit events on a monthly base have been defined by Mathier et al. (1990, 1991) according to a given base level Q_{0j} such that:

$$Q_{0j} = c\,\bar{Q}_j \qquad j = 1, ..., 12 \tag{1}$$

i.e. a fixed proportion c of the overall monthly mean discharge $\bar{Q}_j$, computed from historical data. This base level can be viewed as the limit of an acceptable deficit or energetic failure in comparison with the expected mean discharge ($c = 1$) or a given proportion of this mean discharge ($0 < c < 1$). This definition is useful to determine water deficit duration and severity when specific information on the evaluation of threshold discharges is available. Examples of the use of equation (1) are presented in Mathier et al. (1990, 1991) using a base level, Q_{0j} taken as 85% of the monthly mean discharges, $\bar{Q}_j$, corresponding to $c = 0{,}85$ in equation (1). This base level represents a 15% risk of energetic failure (Ta Trung 1989), which was given as an a priori acceptable risk of deficit for Alcan company according to the experience of their hydrologists.

For a more general use, this definition of the base level has two disadvantages. Firstly, it is not interpretable a priori in terms of exceedance probability, so the choice of the proportion c is difficult when only little past experience is available. Secondly, this definition depends on the mean which is known to be affected by outliers. This estimator of central tendency is efficient only when the sample observations are drawn from a normal population.

We present here a new definition of the base level, Q'_{0j}. We firstly notice that the base level Q_{0j} as defined in equation (1) is in fact a percentile of unknown order p (p = probability of non-exceedance) of the distribution of the monthly mean discharge. The only case where we can associate an a priori probability of exceedance to the base level Q_{0j} is when c = 1 and the distribution of the monthly mean discharges is symmetrical (coefficient of skewness, Cs = 0). The base level is then equal to the median (probability of non-exceedance = 0.5). Since we have no information on the shape of the distribution and we want to attach an a priori probability of non-exceedance to the base level, it is natural to define it as a given percentile of order p of the observed distribution of mean water discharges $\bar{Q}_{jk}$ for each month j over the N years of historical data,

$$Q'_{0j} = (\bar{Q}_{jk})_p \tag{2}$$

where $(\bar{Q}_{jk})_p$ is defined as:

$$P(\bar{Q}_{jk} < (\bar{Q}_{jk})_p) = p \tag{3}$$

In practice, for a given month j we associate to each of the N monthly mean discharges, arranged in increasing order, an empirical probability of non-exceedance p_{jk} (Weibull plotting position formula),

$$p_{jk} = \frac{k}{N + 1} \tag{4}$$

k being the rank. Any other plotting positions formula could also be used (Cunnane 1978). Hence, Q'_{0j} corresponds to the observation $\bar{Q}_{jk}$ having a plotting position or an empirical probability p_{jk} equal to the fixed value of p, or one that is obtained by linear interpolation between two successive observations with probability of non-exceedance p_{jk} and $p_{j(k+1)}$.

For symmetrical distributions of observed $\bar{Q}_{jk}$ values (Cs = 0), the use of equation (1) with c = 1 or the use of equation (2) with p = 0.5 will give the same base level since the mean and median will be equal. In order to determine water deficit events, every monthly mean water discharge $\bar{Q}_{jk}$ of a given year k, is compared to its respective base level Q'_{0j}. The outcome of this comparison can be classified in two mutually exclusive ways such that: a success represents the absence of water deficit event (mean monthly discharge exceeds threshold), and a failure represents a water deficit event (mean monthly discharge is below threshold). As a deficit is detected every time $\bar{Q}_{jk} < Q'_{0j}$, the base level provides a probabilistic interpretation of the occurrences of deficits in month j (equation 3). The total number of consecutive failures before a success, represents the duration of a water deficit event. In turn, the summation of discharges in deficit over the duration of a water deficit event yields the cumulative water discharge in deficit or "severity" of this event. Thus, if the median is used as a criterion to define the monthly base levels in equation (2) (p = 0.5), the distribution of water deficit duration will correspond to a 50% probability of being in deficit for each month.

Figure 1 illustrates water deficit events and variables. Since a water deficit event is detected every time $\bar{Q}_{jk} < Q'_{0j}$, the duration (DUR) of a water deficit is the total number of

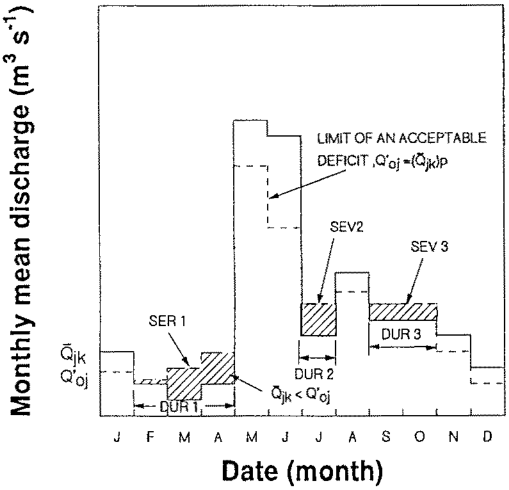

Figure 1. Definition of water deficit events and variables

months in deficit in the same sequence, and the severity (SEV, in m^3/s) is the summation of water discharges in deficit within a water deficit period. A computer program has been developed to extract automatically water deficit characteristics from the data base (q_{ijk}) for one or several specified values of p (equation 2). To illustrate the method, analysis of water deficit will be performed for one representative p value on a basin used in management by Alcan. For the selected basin, computation of water deficit events over the data base provided the data sets of water deficit duration (DUR) and severity (SEV) needed for the analysis.

3 Theoretical considerations

In this section, we present expressions for estimators, $\hat{x}_q$, of quantiles in the geometric and exponential distributions along with their asymptotic variance. The method of maximum likelihood is used for parameter estimation. These distributions are members of the exponential class of probability distributions, and it can be shown that the method of maximum likelihood leads to estimators that are sufficient statistics for the parameters (Lehmann 1983). Therefore, $\hat{x}_q$ will have good asymptotic properties (consistency, unbiaisness, efficiency). Detailed derivations of the expression for $\hat{x}_q$ and its variance are given in appendix A for the geometric distribution and in appendix B for the exponential distribution.

3.1 Fitting water deficit duration using geometric distribution

In the case of water deficit duration, monthly mean discharges $\overline{Q}_{jk}$ are compared to their respective base levels, Q'_{0j}. Each observation can be classified in two mutually exclusive ways (positive deviation or negative deviation in comparison with the base level) (Yevjevich 1967), which gives rise to a two state random variable, Y, defined as follows:

$Y = 1$, represents the absence of water deficit event (positive deviation, $\overline{Q}_{jk} - Q'_{0j} \geq 0$), and;

$Y = 0$, represents a water deficit (negative deviation, $\overline{Q}_{jk} - Q'_{0j} < 0$).

This variable can be observed for each month in the sequence of monthly mean discharges. We consider therefore the sequence formed by the $T = 12 \times K$ monthly realization of the random variable Y, where K stands for the number of years:

$$Y_1, Y_2, ..., Y_t, ..., Y_T$$

Now, let us suppose that the state of month t depends only on the state of month $t - 1$. Moreover, suppose that the transition probabilities are homogeneous, i.e.

$$\begin{aligned} &\text{Prob } (Y_t = 1/Y_{t-1} = 0) = \pi \ \forall t \\ &\text{Prob } (Y_t = 0/Y_{t-1} = 1) = \beta \ \forall t \\ &\text{Prob } (Y_t = 0/Y_{t-1} = 0) = 1-\pi \ \forall t \\ &\text{Prob } (Y_t = 1/Y_{t-1} = 1) = 1-\beta \ \forall t \end{aligned} \tag{5}$$

This structure of dependency, which is reasonably consistent with common hydrological observations is seen to constitute a Markov process of order 1 (Ross, 1980).

Let the random variable X denote the duration of a sequence of consecutive months where the state of Y did not change, for instance:

X = the number of consecutive months where $Y = 0$.

Then this variate X has a geometric distribution (Ross, 1980) with p.d.f. given by

$$P(X=x) = \pi(1-\pi)^{x-1}, \ x = 1, 2, 3, ..., \ 0 \leq \pi \leq 1 \tag{6}$$

where π is the probability defined above (equation 5).

Estimation of the parameter π is done by the method of maximum likelihood which gives

$$\hat{\pi}_q = \frac{1}{\bar{x}} \tag{7}$$

The maximum likelihood estimator of x_q and the asymptotic variance of $\hat{x}_q$ are respectively expressed by

$$\hat{x}_q = \frac{\ln (q)}{\ln (1 - (1/\bar{x}))} \tag{8}$$

and

$$\text{Var}(\hat{x}_q) = \ln^2(q) \cdot \ln^{-4}(1-(1/\bar{x})) \cdot \frac{1}{n\bar{x}(\bar{x}-1)} \tag{9}$$

Details of the derivation of equations (8) and (9) are given in Appendix A.

Using the central limit theorem (Lehmann 1983), an asymptotic confidence interval for x_q at an arbitrary $(1-\alpha)$ % level of confidence can be computed as:

$$\hat{x}_q - u(\alpha/2)[\text{Var}(\hat{x}_q)]^{1/2} \leq x_q \leq \hat{x}_q + u(\alpha/2)[\text{Var}(\hat{x}_q)]^{1/2} \tag{10}$$

where $u(\alpha/2)$ is the standard normal variate with probability of exceedance $\alpha/2$.

3.2 Fitting water deficit severity using exponential distribution

For water duration severity the random variable of interest, X, is equal to the summation of discharges in deficit over the duration of a water deficit event. As mentioned previously, Mathier et al. (1990) obtained good results using the Gamma distribution for fitting severity data obtained according to definition (1). The coefficient of skewness (C_s) of severity data obtained by definition (2) is near 2 (see Table 5), we may assume that X follows a one-parameter exponential distribution which is a particular case of the Gamma distribution corresponding theoretically to $C_s = 2$. Thus we assume that severity (X) is exponentially distributed with p.d.f.,

$$f(x) = \frac{1}{\lambda} e^{-x/\lambda} \quad , x > 0 \tag{11}$$

The maximum likelihood estimator of λ is expressed by

$$\hat{\lambda} = \bar{x} \tag{12}$$

Therefore x_q is estimated by

$$\hat{x}_q = \bar{x} \ln (1/q) \tag{13}$$

and the asymptotic variance of $\hat{x}_q$ by

$$\text{Var}(\hat{x}_q) = \ln^2(1/q) \cdot \frac{\bar{x}^2}{n} \tag{14}$$

An asymptotic confidence interval for x_q at a given $(1-\alpha)$ % level of confidence can again be obtained using equation (10). The derivations of equations (13) and (14) are given in Appendix B.

The program DEFICIT developed at INRS-Eau (Mathier et al. 1990) allows the automatic fitting of the geometric and exponential distributions and gives values of $\hat{x}_q$ (equations 8 and 13) and $\text{Var}(\hat{x}_q)$ (equations 9 and 14) for 17 probabilities of exceedance. This program is applied in the next section to illustrate the fitting on observed data (DUR and SEV data sets) using the geometric and exponential distributions. The results of goodness-of-fit tests are presented as well.

4 Applications

For illustrative purpose, the value of p in equation (2) needed to compute the base level has been set to 0.35, which represents a 35% probability of being in deficit. DUR and SEV data sets have been extracted on this basis.

Prior to the fitting of the distributions, basic assumptions of independence, homogeneity and absence of outliers in the data sets (X_1, X_2, ..., X_N) should be tested for (Bobée & Ashkar 1991). This was done using the Wald-Wolfowitz test for independence (Wald & Wolfowitz 1943) and the Grubbs and Beck test for the detection of outliers (Grubbs & Beck 1972). Because no heterogeneity was suspected in the data sets, no homogeneity test was performed.

The result of the test of independence for the two data sets (water deficit duration (DUR) and severity (SEV)) indicates at a 5% level of significance that in both cases observations are independent.

The Grubbs and Beck test, at a 5% level of significance, indicates that no outliers are present in the two data sets.

4.1 Fitted distributions and goodness-of-fit tests

This section presents an example of distributions fitted to observed data of water deficit duration and severity (DUR, SEV). As an indicative tool we also present the results of two goodness-of-fit tests. Because the chi-square test is a natural choice for discrete distributions (D'Agostino & Stephens 1986), it will be used for the geometric distribution. This test is adjusted to reflect the estimation of the parameter π from the sample by subtracting one degree of freedom (Sokal & Rohlf 1969).

In the case of continuous distributions the Kolmogorov-Smirnov test is often more powerful than tests of the chi-square type (D'Agostino & Stephens 1986). Since this test has been adapted by Durbin (Durbin 1975) for the particular case of the exponential distribution with parameter λ unknown, it will be used here. Finally, events corresponding to a given probability of exceedance are estimated and the related asymptotic confidence intervals are deduced.

4.1.1 Geometric distribution and the chi-square test

Table 2 presents the sample moments and the estimated parameter (equation 7) of the geometric distribution.

Figure 2 presents the histogram of observed water deficit duration (DUR) together with the fitted geometric distribution. Visual examination of this figure shows that the adequacy of the fitted distribution is good. Goodness-of-fit is evaluated using the chi-square test (Table 3).

Since the calculated value (1.539) of the chi-square test (Table 3) is less than the critical value (7.82) at a 5% level of significance, the test indicates that the observations are adequately represented by a geometric distribution.

Using equations (8), (9) and (10) we obtain for the geometric distribution, $\hat{x}_q$, $[\mathrm{Var}(\hat{x}_q)]^{1/2}$ and the asymptotic confidence interval (95%) of the true value of x_q (Table 4).

The geometric distribution has the advantage of taking into account the discrete character of water deficit duration. Furthermore, the proposed method is based on a definition of deficit duration allowing an a priori identification of the statistical distribution of interest (geometric).

4.1.2 Exponential distribution and the Kolmogorov-Smirnov test

The sample moments of water deficit severity (SEV), and the estimate of the parameter λ of the exponential distribution are given in Table 5.

Figure 3 presents the observed water deficit severity (SEV) together with the fitted one-parameter exponential distribution (EX). The visual fit of the exponential distribution to observed data is reasonably good. For a quantitative appraisal of the fit we have used the

Table 2. Sample moments and estimation of the parameter for the geometric distribution

Number of observations	N = 99
Arithmetic mean (months):	$\bar{x}$ = 1.9394
Standard deviation (months):	S = 1.2357
Coeff. of skewness:	C_s = 1.4750
Estimate of π (month^{-1}):	$\hat{\pi}$ = 0.5156

Table 3. Chi-square goodness-of-fit test for the geometric distribution

Class	Observed Frequency	Expected Frequency
1	50	51.0
2	23	24.7
3	15	12.0
4	7	5.8
5	4	5.4

Number of classes	= 5
Degrees of freedom	= 5 (-1) -1 = 3
Chi-square	= 1.539
Critical value (α = 5%)	= 7.820

Table 4. Estimated value, $\hat{x}_q$, of duration events, standard deviation of $\hat{x}_q$ and asymptotic confidence intervals for x_q at a 95% level for different probabilities of exceedance (q) using the geometric distribution

Probability of exceedance	$\hat{x}_q$ event	Standard deviation of $\hat{x}_q$	Confidence intervals 95%	
.001	9.529	0.979	7.611	11.448
.005	7.309	0.751	5.838	8.781
.010	6.353	0.653	5.074	7.632
.020	5.397	0.554	4.310	6.483
.050	4.133	0.424	3.301	4.965
.100	3.176	0.326	2.537	3.816
.200	2.220	0.228	1.773	2.667
.300	1.661	0.171	1.327	1.995
.500	0.956	0.098	0.764	1.149
.700	0.492	0.051	0.393	0.591
.800	0.308	0.032	0.246	0.370
.900	0.145	0.015	0.116	0.175
.950	0.071	0.007	0.057	0.085
.980	0.028	0.003	0.022	0.033
.990	0.014	0.001	0.011	0.017
.995	0.007	0.001	0.005	0.008
.999	0.001	0.0001	0.001	0.002

Table 5. Sample moments and estimate of the parameter of the exponential distribution

Number of observations	$\underline{N} = 99$
Arithmetic mean ($m^3\ sec^{-1}$):	$\bar{x} = 105.7495$
Standard deviation ($m^3\ sec^{-1}$):	$S = 120.1997$
Coeff. of skewness:	$C_s = 1.8531$
Estimate of λ ($m^3\ sec^{-1}$):	$\hat{\lambda} = 105.7495$

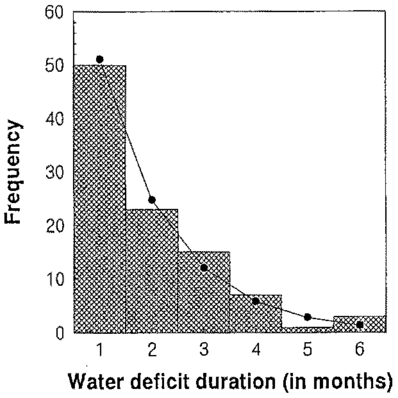

Figure 2. Histogram of observed water deficit duration (DUR) together with fitted geometric distribution

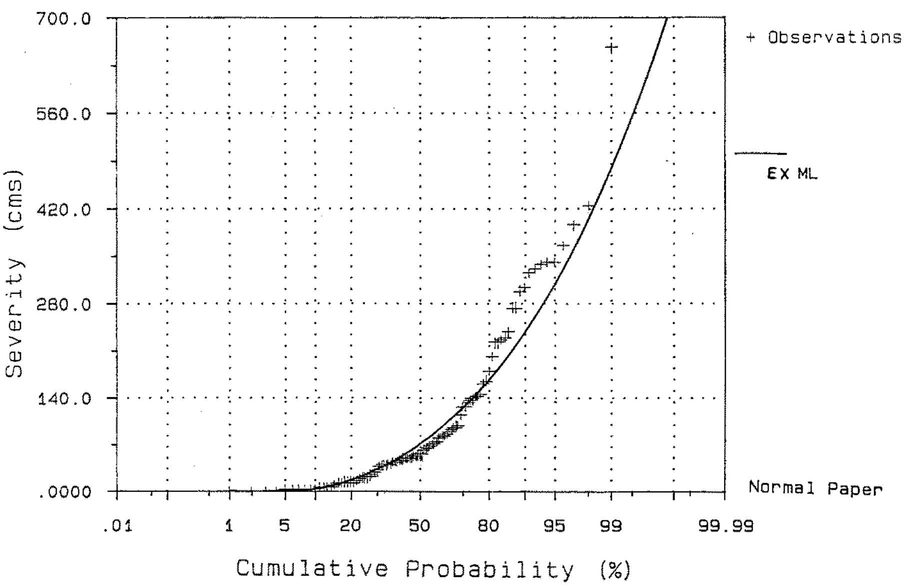

Figure 3. The observed water deficit severity (SEV) together with fitted one-parameter exponential distribution (EX) using the maximum likelihood method (ML)

Table 6. Kolmogorov-Smirnov goodness-of-fit test for the exponential distribution

D statistic, Kolmogorov-Smirnov test
$D = 1.0405$ Critical value ($\alpha = 5\%$) = 1.0753

Table 7. Estimated value, $\hat{x}_q$, of severity events, standard deviation of $\hat{x}_q$ and asymptotic confidence intervals for x_q at a 95% level for different probabilities of exceedance (q) using the exponential distribution

Probability of exceedance	$\hat{x}_q$ event	Standard deviation of $\hat{x}_q$	Confidence intervals 95%	
.001	730.492	73.417	586.594	874.389
.005	560.294	56.312	449.923	670.665
.010	486.994	48.945	391.063	582.926
.020	413.694	41.578	332.202	495.187
.050	316.797	31.839	254.392	379.202
.100	243.497	24.472	195.531	291.463
.200	170.197	17.105	136.671	203.724
.300	127.320	12.796	102.239	152.400
.500	73.300	7.367	58.861	87.739
.700	37.718	3.791	30.288	45.148
.800	23.597	2.372	18.949	28.246
.900	11.142	1.120	8.947	13.337
.950	5.424	0.545	4.356	6.493
.980	2.136	0.215	1.716	2.557
.990	1.063	0.107	0.853	1.272
.995	0.530	0.053	0.426	0.634
.999	0.106	0.011	0.085	0.127

Kolmogorov-Smirnov test (D statistic) adapted by Durbin (1975) for the exponential distribution with unknown population mean. Table 6 shows the result of this test.

The calculated value (1.0405) of the Kolmogorov-Smirnov test statistic is less than the critical value (1.0753) at a 5% level of significance (as defined by Durbin 1975), thus indicating that the observations are adequately represented by an exponential distribution.

Using equations (13), (14) and (10), $\hat{x}_q$, $[\mathrm{Var}(\hat{x}_q)]^{1/2}$ and the asymptotic confidence interval (95%) for the true value of x_q have been computed and are displayed in Table 7.

Result of the goodness of fit test (Table 6) indicates that the fit is just barely acceptable. If a more precise fit is needed for the water deficit severity, the use of a distribution with 2 or 3 parameters may be appropriate. Since the one-parameter exponential distribution is a special case of the 2-parameter gamma distribution (G2 (α, λ)) with shape parameter $\lambda = 1$ or equivalently with $C_s = 2$, and of the 3-parameter generalized gamma distribution (GG3 (α, λ, s)) with shape and power parameters equal to unity ($\lambda = s = 1$) (Bobée & Ashkar 1991), more adequate fitting could be obtained using these distributions. This, however, would be at the expense of more parameters to be estimated, implying an increased variance of the estimator of x_q. As an example, Figure 4 shows the comparison of the fittings of the

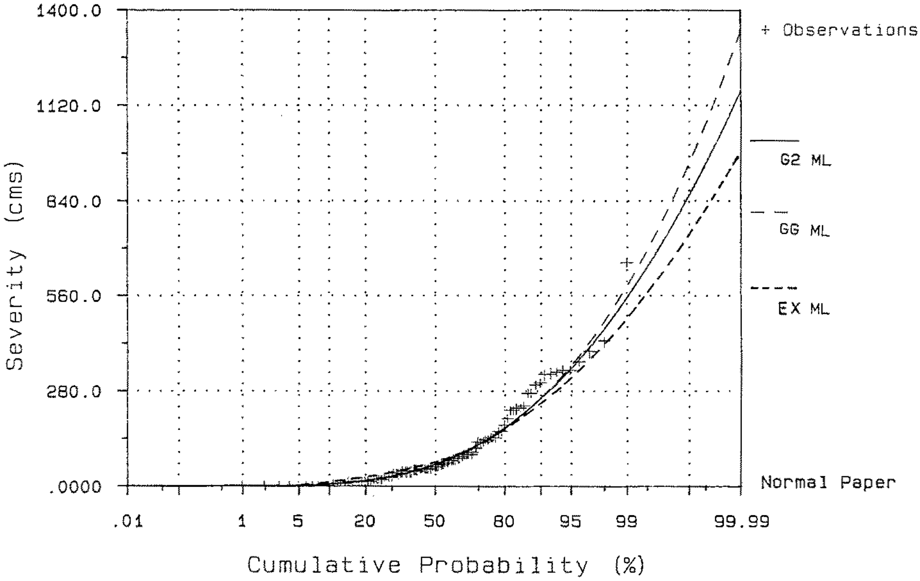

Figure 4. Comparison of the fitting for the G2, GG3 and one-parameter exponential (EX) distributions using the maximum likelihood method (ML) (SEV data set)

Table 8. Sample skewness and estimates of the parameters of the G2 and GG3 distributions using the maximum likelihood method

Station	Sample skewness	G2 $\hat{\alpha}$	G2 $\hat{\lambda}$	GG3 $\hat{\alpha}$	GG3 $\hat{\lambda}$	GG3 $\hat{s}$
SEV data	1.853	0.0073	0.7763	0.0105	1.0114	0.8428
1	2.435	0.0248	0.8463	0.0619	1.5007	0.7018
2	1.819	0.0094	0.7606	0.0096	0.7757	0.9893
3	2.061	0.0094	0.7851	0.0104	0.8482	0.9519
4	1.711	0.0421	0.9281	0.0645	1.2505	0.8296
5	2.255	0.0022	0.5795	0.0077	1.2642	0.6018

G2, GG3 and one-parameter exponential (EX) distributions. The fittings obtained for these distributions are quite similar, especially for cumulative probabilities less than 98%. Table 8 gives the sample skewness and the estimated parameters of the G2 and GG3 distributions by the method of maximum likelihood using the HFA software (Bobée & Ashkar 1991) for the SEV data set and five other stations of the Alcan hydroelectric system.

In all cases the sample skewness is close to 2.0. Except for stations 1 and 5, the estimated λ and s parameters are close to unity. Therefore, it is reasonable to use the one parameter exponential distribution in this example.

5 Conclusion

This paper presents an original and easy-to-use approach to study water deficit events defined according to economic requirements. The use of a variable threshold allows the identification of water deficit reflecting the risk of energetic failure and drop in power production on a monthly base. The approach gives a theoretical base for the choice of the geometric distribution to represent water deficit duration. This distribution takes into account the discrete character of the duration variable. Using the analogy between the geometric and the exponential distributions (Johnson & Kotz 1969, Feller 1957), this latter distribution has given reasonably good results to describe water discharges in deficit. In both cases, only one parameter needs to be estimated. The maximum likelihood method, which gives estimators with optimal asymptotical properties (Lehmann 1983) can be used for both distributions. Moreover, parameters π and λ are directly estimated from the arithmetic mean. Therefore, the procedure and the interpretation of the results are straightforward. If more flexibility is required for the frequency analysis of water deficit severity, a multi-parameter distribution such as the 2-parameter Gamma distribution or the generalized Gamma distribution may be appropriate.

The methodology presented in this study can be used to gain a better knowledge of the links between droughts and energetic failure in an economic context. Future research will be undertaken to study the influence of different base levels on water deficit duration and severity variables.

Acknowledgements

We thank Dave Monaghan, Daniel Morissette and Vincent Fortin for the programming of DEFICIT. The comments of Dr. Peter Funder Rasmussen are greatly appreciated. Participation of Hau Ta Trung and Claude Gignac from Alcan company is acknowledged. Financial support for this research was provided by the Natural Sciences and Engineering Research Council of Canada (NSERC) and Alcan company (University-Industry Program, Grant No. CRD 0039064).

References

Acreman, M. C. 1990: A simple stochastic model of hourly rainfall for Farnborough, England. Hydrological Sciences Journal 35, 119-148

Bobée, B.; Ashkar, F. 1991: The gamma family and derived distributions applied in hydrology. Water Resources Publications, Yevjevich Edt., Littleton, Co. 217 p.

Condie, R.; Cheng, L. G. 1983: Low flow frequency analysis. Program Loflow, Water Resources Branch, Inland Waters Directorate, Env, Canada

Condie, R.; Nix, G. A. 1975: Modelling of low flow frequency distributions and parameter estimation. International Water Resources Symposium, Water for Arid Lands, Iran

C.T.G.R.E.F. 1978: Distribution de fréquence adaptée aux étiages. Méthodes manuelles et graphiques, Groupement d'Antony, Div. Hydrologie hydraulique fluviale, No. 6, Note technique

Cunnane, C. 1978: Unbiased plotting positions - a review. J. Hydrol. 37, 205-222

D'Agostino, R. B.; Stephens, M. A. 1986: Goodness-of-fit techniques. Marcel Dekker, inc., New York

Durbin, J. 1975: Kolmogorov-Smirnov tests when parameters are estimated with applications to tests of exponentiality and tests on spacing. Biometrika 62, 5-22

Feller, W. 1957: An introduction to probability theory and its applications. Vol. 1, Second Edition, John Wiley, New York, 461 p.

Grubbs, F. E.; Beck, G. 1972: Extension of sample sizes and percentage points for significance tests of outlying observations. Technometrics 14, 847-854

Gumbel, E. J. 1963: Statistical forecast of droughts. Bulletin of the International Association of Scientific Hydrology VIII e année 1, 5-23

Güven, 0.1983: A simplified semiempirical approach to probabilities of extreme hydrologic droughts. Water Resources Research 19, 441-453

Hoang, V. D. 1978: Meilleur ajustement statistique aux débits journaliers de crue et d'étiage des riviéres du Québec, Ministère des Richesses naturelles. Dir. Gén. des Eaux, H. P. 47

Johnson, N. L.; Kotz, S. 1969: Discrete distributions. John Wiley & Sons, New York

Joseph, E. S. 1970: Probability distribution of annual droughts. In: Journal of the Irrigation and Drainage Division, 461-474, Proceedings of the American Society of Civil Engineers, 96

Joseph, E. S. 1971: Recurrence intervals of annual minimum streamflows. In: Proceedings Symposium on Statistical Hydrology, 366-370, Tucson, Arizona

Lara, O. G. 1970: Low flow frequencies of Illinois streams. United States Dept. of the Interior, Geological Survey

Lehmann, E. L. 1983: Theory of point estimation. John Wiley & sons, New York

Loganathan, G. V., Kuo, C. Y.; McCormick, T. C. 1985: Frequency Analysis of Low Flows. Nordic Hydrology 16, 105-128

Matalas, N. C. 1963: Probability distributions of low flows, U. S. G. S. Prof. Rep., 437-A

Mathier, L., Ta Trung, H., Bobée, B., Gignac, C., Ashkar, F.; Boucher, S. 1990: Etude des débits déficitaires (étiages). INRS-Eau, rapport interne, 118

Mathier, L., Perreault, L., Bobée, B.; Ashkar, F. 1991: Frequency analysis of water deficit duration and severity. Second International Conference on Computer Methods and Water Resources (CMWR 91), Marrakech, Morocco

Miquel, J.; Phien Bou Pha, B. 1978: Tempétiage: un modéle d'estimation des risques d'étiage. Société hydrotechnique de France, XVe journée de l'hydraulique

Ozga-Zielinska, M. 1989: Droughts and floods - their definition and modelling. In: New Directions for Surface Water Modelling, 313-322, Proceedings of the Baltimore Symposium, May 1989, IAHS Publ., 181

Paulson, Jr. E. G., Sadeghipour, J.; Dracup, J. A. 1985: Regional frequency analysis of multiyear droughts using watershed and climatic information. Journal of Hydrology 77, 57-76

Pilon, P. J. 1990: The Weibull distribution applied to regional low flow frequency analysis. In: Regionalization in Hydrology, 227-237, Proceedings of the Ljubljana Symposium, April 1990, IAHS Publ., 191

Prakash, A. 1981: Statistical determination of design low flows. Journal of Hydrology 51, 109-118

Riggs, H. C. 1972: Low flow investigations. Tech. of Water Resources Investigations of the United States Geological Survey, Book 4, Chap. B. 1

Ross, S. M. 1980: Introduction to probability models. Academic Press, Second Edition, New York

Sadeghipour, J.; Dracup, J. A. 1985: Regional frequency analysis of hydrologic multiyear droughts. Water Resources Bulletin 21, 481-487

Sen, Z. 1980: Statistical analysis of hydrologic critical droughts. Journal of the Hydraulics Division, January, 99-115

Sen, Z. 1982: Stochastic analysis of droughts. In: Statistical Analysis of Rainfall and Runoff, (Ed. Singh, V. P.), 637-650, Wat. Res. Publ.

Sokal, R. R.; Rohlf, F. J. 1969: Biometry. W. H. Freeman and company, San Francisco

Ta Trung, H. 1989: Capacité énergétique à long terme du systéme hydroélectrique Alcan au Saguenay-Lac-Saint-Jean. Groupe Ressources hydrauliques, Énergie électrique Québec, Alcan, Rapport No. 89-15

Tasker, G. D. 1987: A comparison of methods for estimating low flow characteristics of streams. Water Resources Bulletin 23, 1077-1083

TCLFE (Task Committee on Low Flow Evaluation) 1980: Characteristics of low flows. Journal of the Hydraulics Division, May, 717-731

Vogel, R. M.; Kroll, C. N. 1989: Low-flow frequency analysis using probability-plot correlation coefficients. Journal of Water Resources Planning and Management 117, 338-357

Wald, A.; Wolfowitz, J. 1943: An exact test for randomness in the nonparametric case based on serial correlation. Ann. Math. Stat. 14, 378-388

Yevjevich, V. 1967: An objective approach to definitions and investigations of continental hydrologic droughts. Hydrology Papers, Colorado State University, Fort Collins, Colorado, 23

Yonts, W. L. 1971: Low flow measurements of North Carolina streams. State of North Carolina, Board of Water and Air Resources, Dept. of Water and Air Resources, July

Zelenhasic, E.; Salvai, A. 1987: A method of streamflow drought analysis. Water Resources Research 23, 156-168

Appendix A: Determination of the estimator $\hat{x}_q$ of x_q and the estimator of the asymptotic variance of $\hat{x}_q$ for the geometric distribution.

Let $x_1,...,x_j,...,x_N$, be a sample of size N drawn from a geometric distribution with parameter π (GEO(π)). The probability density function (p.d.f.) of X is given by:

$$P(X=x) = \pi(1-\pi)^{x-1}, \qquad x = 1, 2, 3, \ldots \quad \text{and } 0 \leq \pi \leq 1$$

The maximum likelihood estimator of π is

$$\hat{\pi} = \frac{1}{\bar{x}}$$

where $\bar{x}$ is the sample mean. Since $0 \leq (1-\pi) \leq 1$, x_q can be expressed as follows by putting $y_q = x_q - 1$:

$$q = P(X>x_q) = \sum_{j=x_q}^{\infty} P(X=j) = \pi \left[\frac{1}{1-(1-\pi)} - \frac{1-(1-\pi)^{y_q+1}}{1-(1-\pi)} \right]$$

This equation is easily solved for x_q to give

$$x_q = \frac{\ln(q)}{\ln(1-\pi)}$$

Replacing π by its estimator $1/\bar{x}$, we obtain the maximum likelihood estimator of x_q:

$$\hat{x}_q = \frac{\ln(q)}{\ln(1 - 1/\bar{x})}$$

The determination of the asymptotic variance of $\hat{x}_q$ is based on the central limit theorem (Lehmann 1983, p. 336) and is expressed as

$$V(\hat{x}_q) = (d\hat{x}_q/d\pi)^2 \, \mathrm{Var}(\bar{x})$$

Since $\mathrm{Var}(\bar{x}) = (1 - \pi)/n\pi^2$, we can directly deduce that:

$$V(\hat{x}_q) = \ln^2(q) \ln^{-4}(1 - \pi) \frac{\pi^2}{n(1 - \pi)}$$

Replacing π by its estimator $1/\bar{x}$, the estimator of $V(\hat{x}_q)$ is given by

$$\mathrm{Var}(\hat{x}_q) = \ln^2(q) \ln^{-4}(1 - 1/\bar{x}) \frac{1}{n\bar{x}(\bar{x} - 1)}$$

Appendix B: Determination of the estimator $\hat{x}_q$ of x_q and the estimator of the asymptotic variance of $\hat{x}_q$ for the exponential distribution.

Let $x_1,...,x_j,...,x_N$, be a sample of size N drawn from an exponential distribution EX(λ) of parameter λ. The p.d.f. of X is given by:

$$f(x) = \frac{1}{\lambda} e^{-x/\lambda}, \quad x \geq 0.$$

The maximum likelihood estimator of the parameter λ

$$\hat{\lambda} = \bar{x}.$$

The quantity, x_q is defined by

$$P(X > x_q) = \int_{x_q}^{\infty} f(x;\lambda)dx = q$$

which in the case of the exponential distribution is easily solved for x_q to give:

$$x_q = - \lambda \ln (q)$$

The maximum likelihood estimator of x_q is obtained by replacing λ by its estimator $\bar{x}$:

$$\hat{x}_q = - \hat{\lambda} \ln(q) = \bar{x} \ln(1/q)$$

and the variance of this estimator can be expressed as

$$V(\hat{x}_q) = \ln^2 (1/q) V(\bar{x}) = \ln^2(1/q) \frac{1}{n^2} \sum_{1}^{n} V(x_i)$$

$$= \ln^2(1/q) \frac{\lambda^2}{n}.$$

Replacing λ by its estimator $\bar{x}$, the maximum likelihood estimator of $V(\hat{x}_q)$ becomes:

$$Var(\hat{x}_q) = \frac{[\bar{x}\ln(1/q)]^2}{n}$$