Theory and Methods of Drought System Analysis

Zhenmin Zhou, Zhiliang Wang

North China College of WRHP, Zhengzhou, Henan 450045, China, zzm1022@yahoo.com.cn

Abstract: Based on analysis on statistic parameters including mean, coefficient of variation and serial correlation of drought, the foresaid parameters on drought duration and severity were studied. The extreme drought duration and severity expressions were developed. Taking Dongzhou Reservoir as an example, the related design parameters were calculated. [Nature and Science 2003;1(1):62-66].

Key Words: drought frequency; drought duration; drought severity; truncated level

1. Introduction

Droughts have attracted the much attention in different areas at the present. The sequences of the rainfall or stream flow that are used to characterize droughts are known as drought variables. In general, droughts are measured in terms of deficiency in the rainfalls or streams flows below a predefined reference level. The cumulative deficit during a drought spell is known as the severity, the reference level for identification of the droughts, or for quantifying the severity. It is termed the truncation level and is taken to be equal to the long-term mean of the sequence of the drought variable.

From the point of view of creating provisions for meeting the exigencies during a drought spell, the most important parameters of concern are the longest duration and the largest severity for a desired return period of T years, and are termed as extreme duration and severity in this paper. The sequence representing the drought could be distributed normally or may have a skewed distribution. The simple type of pdfs used for modeling the annual rainfall and runoff sequences are normal, log-normal and gamma, all with two parameters. Likewise, the simplest kind of dependence is Markovian, represented by the lag-one serial correlation coefficient ρ. For a given place or a basin the reliable statistics available for annual rainfall and stream flow are mean, coefficient of variation and lag-one serial correlation coefficient. One might wish to know the answers to the following questions based on the above readily available information. (1) How does the coefficient of variation affect the extreme duration and severity characteristics? (2) How do the skewness (pdf of the drought variable sequence) and the dependence (ρ) influence the extreme values of the aforesaid parameters? (3) Is there any formula terms of the largest severity and the return period that may appear parallel to the flood frequency formula, commonly cited in the hydrological texts? The flood frequency formula is used for design of the runoff handling structure. The drought frequency formula should be usable for the design of water storage structures for the drought periods. Therefore, based on statistic theory and analysis methods, the research was carried out on above problems.

2 Stochastic Analysis of Extreme Drought Duration and Severity

2.1 Evaluation of Drought Probability Quantile Q

If a time-series x_i (modular form, mean=1 and standard deviation= σ_χ) is truncated at the mean level x_m =1, then the event of excesses ($x_i > x_m$) and deficits ($x_i < x_m$) would emerge along the time axis. The truncation level can be assigned a probability quantile as q=P ($x_i \leqslant x_m$ =1), where q is the probability of drought corresponding to the truncation level x_m =1 and P(…) stands for the notation of probability. For normal pdf, the relationship between q and the truncation level can be obtained by standardizing the sequence x_i into z_i (z_i has mean=0 and σ_z =1 and a standard normal distribution). Thus, for the normal pdf (coefficient of skewness, C_s=0) the value of z_i (corresponding to x_m=1) is 0, and q=p($x_i \leqslant 1$)=P($z_i \leqslant 0$)=0.5 is obtained straightforwardly from the standard normal probability tables or through the following standard normal probability integral

$$q = p\left(z \leq z_m\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_m} \exp(-0.5 z^2) dz \tag{1}$$

The value of z_i corresponding to x_m=1 for normal pdf is referred to as z_m(=0) in the text.

For the log-normal distribution, the value of the truncation level z_{ml}, equivalent to the mean of the sequence x_m=1, can be worked out using the following explanation. If x_i is log-normal variate with a mean x_m and standard deviation σ_χ (coefficient of variation=C_v), ln(x_i) is normal with a mean μ_1 and standard deviation σ_1 . The x_i sequence can be standardized as $u_i = \left(x_i - x_m\right)/\sigma_x$, where the U_i sequence has mean=0 and standard deviation=1.

Therefore x_i can be expressed in terms of mean (x_m) and coefficient of variation (C_v) as follows

$$x_i = x_m (1 + C_v \cdot u_i) \tag{2}$$

As the variate $\ln(x_i)$ can also be standardized, the new normal standardized variate (designated as Z_i) can be expressed as

$$z_i = \frac{\ln(x_i) - \mu_1}{\sigma_1} = \frac{\ln[x_m(1 + c_v \cdot u_i)] - \mu_1}{\sigma_1} \tag{3}$$

At the mean level of x_i sequence, u_i= 0, hence z_i, designated as z_{ml}, can be expressed using equation (3) as

$$z_{ml} = \frac{\ln(x_m) - \mu_1}{\sigma_1} \tag{4}$$

It is known that moments of x_i and $\ln(x_i)$ are related by the following relationships

$$\mu_1 = \ln(x_m) - 0.5\sigma_1^2 \tag{5}$$

$$\sigma_1^2 = \ln(1 + C_v^2) \tag{6}$$

Using equations (5) and (6), equation (4) can be simplified as

$$Z_{ml} = \frac{\ln x_m - (\ln x_m - 0.5\sigma_1^2)}{\sigma_1} = 0.5\sigma_1 = 0.5[\ln(1 + C_v^2)]^{0.5} \tag{7}$$

Note the modular form, hence x_m=1 in all the above relationships. Since z_{ml} is a standard normal variate, the value of the probability quantile of the equation $q = P(z_i \le z_{ml})$ can be obtained from the standard normal probability tables or computed by the numerical integration of the standard normal function shown in equation (1) with z_m replaced by z_{ml}. Looking at the structure of equation (7), it is obvious that z_{ml} is great than 0 for the mean of the sequence x_i (i.e. x_m=1). The value of q for the log-normally distributed sequence will always be greater than 0.5 and is dependent on the C_v of the sequence. The effect of skewness is implicitly included in C_v. As for the log-normal pdf, C_s and C_v are functionally related through the following relationship

$$C_s = 3C_v + C_v^3 \tag{8}$$

The same analysis can be extended to the two-parameter gamma distribution. If γ_i is the standardized gamma variate [γ_i = (x_i - 1)/C_v], then z_i, i.e. the standardized normal variate, is related approximately to γ_i through the following linkage relationship

$$z_i = \frac{6}{c_s}\left[(0.5C_s \cdot \gamma_i + 1)^{0.333} - 1\right] + \frac{C_s}{6} \tag{9}$$

For a two-parameter gamma variable $C_s = 2C_v$, equation (9) can be reduced to

$$z_i = 3C_v^{-1}\left[x_i^{0.333} - 1\right] + 0.333C_v \tag{10}$$

Therefore, a modular gamma variate can be transformed into a standard normal variate using equation (10) and at the mean level, $x_m = 1$, $z_{m\gamma} = 0.333\,C_v$. Note the value of z_i is designated as $Zm\gamma$ for x_i=1. That is, if the drought variable is gamma distributed, at the truncation level equal to the mean of the sequence the equivalent standard normal variates $Zm\gamma$ will be 0.0666, 0.1332, 0.2098, 0.2664 and 0.333 for C_v values of 0.2, 0.4, 0.6, 0.8 and 1, respectively. For a normally distributed drought variable all of them will be equal to 0, regardless of the C_v. Therefore, the probability quantile for a gamma-distributed drought variable could be worked out for the desired value of C_v by consulting the standard normal probability tables or through integration of equation (1), replacing z_m by $Zm\gamma$. Note here again that $Zm\gamma$ >0 for x_m=1. For example, for a gamma-distributed drought variable with C_v=0.4, $Zm\gamma$ = 0.13 and q = 0.55 at the level of truncation equal to the mean level of the sequence, against 0.5 for a normally distributed drought variable, and 0.58 for a log-normally distributed variable. Values of q for normal, two-parameter gamma and log-normal pdfs for various C_v values are shown in Table 1.

2.2 Modeling Extreme Drought Duration

Any uninterrupted sequence of deficits can be regarded as a drought length (duration) equal to the number of deficits in the sequence, designated by L (L=1, 2, 3, ... , j). Each drought duration is associated with deficit sum D, i.e. the sum of the individual deficits, d_1, d_2, d_3, ... , in the successive epochs of the spell. This deficit sum is termed as drought standardized sequence of the drought variable. The actual severity (D) is functionally related to the standardized severity (S) through the relationship $D = \sigma_x \cdot S$ (Guven, 1983). The term S/L is termed the drought intensity or magnitude. One can expect n (n=0, 1, 2, 3, ... , i) drought spells (runs) over a period of "T" years and correspondingly there will be n values of severities designated as D_1, D_2, D_3 .... or S_1, S_2, S_3 .... , in the standardized terms.

A designer is interested in the longest values of L, designated as L_T, and the largest value of S, designated as S_T. The period "T" means a sample size of T (T = 10, 20, ..., 100) years and can be regarded as equivalent to a return period of "T" years. The probabilistic relationship for L_T can be obtained by applying the theorem of extreme of random number of random variables. The computations for drought durations and form does not affect the magnitude of the duration (L)

Table 1. Values of Q and R at the Mean Level for Different Probability Distributions

Model	For all c_v	c_v=0.2		c_v=0.4		c_v=0.6		c_v=0.8		c_v=1.0		Remarks
Distribution	n	γ	ln	γ	ln	γ	ln	γ	ln	γ	ln	
q ρ=0	0.5	0.53	0.54	0.55	0.58	0.58	0.61	0.61	0.64	0.64	0.66	n :normal γ :gamma ln : log
Different ρ 、c_v and r value												
ρ=0.1	0.53	0.56	0.57	0.58	0.61	0.61	0.64	0.63	0.67	0.66	0.71	
ρ=0.3	0.60	0.62	0.63	0.64	0.67	0.67	0.70	0.69	0.74	0.71	0.78	
ρ=0.5	0.67	0.69	0.70	0.71	0.73	0.73	0.76	0.75	0.80	0.77	0.84	
ρ=0.7	0.75	0.77	0.77	0.78	0.80	0.80	0.82	0.81	0.86	0.83	0.89	

but does affect the magnitude of the severity (D). The mean has no effect on the severity as it is the difference between the magnitude of the drought variable and the mean itself. So, in the course of mathematical derivations for severity, the mean cancels out, leaving standardized severity, which can be transformed into actual severity D by the relationship mentioned above. The following relationships can therefore be deduced for L_T following the work of Guven (1983).

$$P(L_T \le j) = p(n=0) + \sum_{i=1}^{\infty} [P(L \le j)]^{i} \cdot p(n=i) \quad (11)$$

Note j takes on discrete values, P(...) stands for the notation of cumulative probability and p(...) for the probability of discrete events. Spells may evolve randomly or in a Markovian fashion and can be regarded to follow the Poisson law of probability. The geometric law of probability adequately simulates the drought length L. The following relationships can be written for probabilities of number of drought runs and length of runs.

$$p(n=i) = \frac{\exp[-Tq(1-r)]\ [Tq(1-r)]^{i}}{i!} \quad (12)$$

$$P(L_T \le j) = 1 - r^{j-1};\ P(L > j) = r^{j-1} \quad (13)$$

where q is the probability quantile defined earlier and r is the conditional probability of any year being a drought year, given that the past year is also a drought year. Substituting equations (12) and (13) into equation (11) and simplifying, one can get the following relationship

$$P(L_T \le j) = \exp\{-Tq(1-r)[1-P(L \le j)]\} \quad (14)$$

The expected value of L_T can be obtained by using the formula

$$E(L_T) = \sum_{j=1}^{\infty} jp(L_T = j) \quad (15)$$

Since $p(L_T = j) = P(L_T \le j+1) - P(L_T \le j)$, one can therefore derive an expression for $p(L_T = j)$ by involving equations (13) and (14), and the resulting equation is as follows.

$$p(L_T = j) = \exp[-T(1-r)r^{j-1}]\{\exp\lfloor Tq(1-r)^2 r^{j-1} \rfloor - 1\} \quad (16)$$

The conditional probability, r, is related to the lag-one serial correlation coefficient ρ, through the following equation.

$$r = q + \frac{1}{2\pi q}\int_0^{\rho} \exp[-0.5 z_0^2/(1+\tau)](1-\tau^2)^{-1/2} d\tau \quad (17)$$

where ρ is the lag-one serial correlation coefficient of the Markov process and τ is the dummy variable of integration. Equation (17) can be evaluated by a numerical integration procedure and values of r for given ρ at truncation level z_0 ($z_0 = z_m$ for normal pdf, $Z_0 = Z_{m\gamma}$ for gamma pdf, and $Z_0 = Z_{ml}$ for log-normal pdf) can be computed. It is evident from equation (17) that for independent processes r = q. It should be noted that if the x_i sequence is normal, ρ can be substituted as it is, but for log-normal pdf ρ should be converted to ρ_1, which is the right value in the normalized domain and $Z_0 = Z_{ml}$. The linkage relationship between ρ and ρ_1 is of the following form

$$\rho_1 = \frac{\ln\lfloor \rho C_V^2 + 1 \rfloor}{\ln(1 + C_V^2)} \quad (18)$$

For a gamme pdf, ρ remains unchanged but $Z_{m\gamma}$ replaces Z_0. Values of q and r for various pdfs, C_V and ρ values are shown in Table 1. It is obvious from Table 1 that q for the normal pdf is constant and is equal to 0.5 regardless of the coefficient of variation, whereas for log-normal pdf it is dependent on the C_V

and in turn on C_s. Values of r are dependent on ρ for normal pdf and on C_v and ρ for gamma and log-normal pdfs. In all cases log-normal distribution tends to have a greater value of q and r for identical values of ρ and C_v in the sequences of the drought variable, indicating that extreme values of the drought duration are going to be larger under the log-normal structure of data representing the drought. With the above procedural details equation (15) can be solved numerically in order to evaluate $E(L_T)$ for the desired return period "T" in years.

2.3 Modeling Extreme Drought Severities

The probabilistic relationship for S_T can be obtained parallel to that for L_T, i.e. equation (11), and can be expressed as follows

$$P(S_T \le Y) = p(n=0) + \sum_{i=1}^{\infty} P(S \le Y)^i \, p(n=i) \tag{19}$$

in which Y can take on values such as 0, 0.1, 0.2, ... , 60 (Y=60 represents an extremely large value) and these values are dimensionless. Plugging the expression for p(n=i), indicated by equation (12), into equation (19), one obtains the following

$$P(S_T \le Y) = \exp\{-Tq(1-r)[1-P(S \le Y)]\} \tag{20}$$

The term S denotes severity (based on the standardized sequences) or the sum of the deficits in each epoch of the drought spell, which can be approximated as normally distributed in view of the central limit theorem. The pdf of S can therefore be written in the following form

$$P(S \le Y) = \frac{1}{\sqrt{2\pi\sigma_s}} \int_0^Y \exp\left[-0.5\left(\frac{S-\mu_s}{\sigma_s}\right)^2\right] ds \tag{21}$$

In a drought spell consisting of k consecutive years (L=k), the expressions for the mean, μ_s, and standard deviation, σ_s, of severity can be written as follows

$$\mu_s = K(\mu_t) \tag{22}$$

$$\sigma_s^2 = k \cdot \sigma_t^2 \left(\frac{1+\rho}{1-\rho} - \frac{2\rho(1-\rho^k)}{k(1-\rho)^2}\right) \tag{23}$$

where μ_t and σ_t are the mean and standard deviation of the individual deficits, and ρ is as defined earlier.

Parallel to the expression for $E(L_T)$ in equation (15), the expression for $E(S_T)$ can be written as

$$E(S_T) = \int_0^{Y=\infty} S_T f(S_T) dS_T \tag{24}$$

Although the pdf of S_T is not known, equation (24) can still be solved numerically. In the numerical procedure, first, equation (21) is integrated numerically in order to evaluate P(S ≤ Y), and plugging this value into equation (20) yields the estimate of $P(S_T \le Y)$. The value of Y can be allowed to range from 0 to 60 [infinity in equation (24) is approximated by 60] with an increment of 0.1. Let these values of Y be designated as Y_0=0, Yi=0.1, Y_2=0.2, ..., etc., equation (24) can be expressed in the numerical integral form as follows

$$E(S_T) = \sum_{j=0}^{n_1} \left((Y_j + y_{j+1}\right) / 2\left[P(S_T \le Y_{j+1}) - (S_T \le Y_j)\right] \tag{25}$$

In which n_1=60/0.1, and Y_{600}=60.

Since values of $E(S_T)$ are standardized and in non-dimensional form, the actual drought severity designated as D_T can be expressed as

$$D_T = E(S_T)\cdot\sigma_X = E(S_T)\cdot C_V \cdot x_m = x_m - x_m + E(S_T)\sigma_x \tag{26}$$

the last portion of the above equation can be manipulated as

$$D_T = x_m - C_V^{-1}\cdot\sigma_X + E(S_T)\sigma_X = x_m + F_T\cdot\sigma_X \tag{27}$$

where F_T can be called the drought frequency factor and be written in equation form as

$$F_T = \lfloor E(S_T) - C_V^{-1} \rfloor \tag{28}$$

It can be seen that equation (27) is analogous to the flood frequency formula $Q_T = Q_m + K_T\sigma_Q$ commonly cited in hydrological texts.

3. Application

The Dongzhou reservoir is located at the Chaiwen River of the Dawen Basin within the downstream of the Yellow River, which controls the area of 189 km^2. The main water supply is for irrigation as well as for drinking water. The design irrigation area is 8667 hm^2. The average annual runoff within the controlled area amounts to 39×10^6 m^3, C_v=0.40. The flows tend to be log-normally distributed with a negligible level of carryover (ρ). The reservoir should meet the demand 1 in 100 years drought. Therefore there is a need to estimate the volume of water to be stored for designing the reservoir. The solution of the problem starts by estimation q for C_v=0.40 $(\sigma_x = 16\times10^6 m^3)$ in equation (1) and using the values of q and r (r=q for independent flows) in equations (15) and (16). The value of q equals to 0.58, and $E(L_T)$=7.15, then $F_T = [7.15 - 0.40^{-1}] = 4.65$, $D_T = 0.39 + 4.65 \times 0.16 = 1.134 \times 10^8$ m^3.

Had the discharge is normal distribution, then $q = 0.5$, $E(L_T) = 5.78$, $D_T = 92 \times10^6$ m^3. If the discharge is gamma distribution, the calculated results are $q = 0.55$, $E(L_T) = 6.74$, and $D_T = 1.07 \times 10^8$ m^3. When $\rho \neq 0$, if $\rho = 0.5$, then, for log-normal discharge, $q = 0.58$, $r = 0.73$, $E(L_T) = 10.63$, $D_T = 1.69 \times 10^8$ m^3, for normal distribution, $q = 0.5$, $r = 0.67$, $E(L_T) = 8.57$, $D_T = 1.36 \times 10^8$ m^3, for gamma distribution of discharge, $q = 0.55$, $r = 0.71$, $E(L_T) = 9.92$, $D_T = 1.43 \times 10^8$ m^3. The above calculation mainly demonstrates the role of skewness and the carryover effects in the storage needs during drought periods.

Therefore, when it is difficult to identify the pdf of the drought variable, the assumption of log-normality allows a conservative design, which is a desirable feature. The assumption of normality leads to the requirement for least storage under identical conditions of C_v and of the dependence structure of the rainfall or the runoff sequences.

4. Conclusions

The results of the present analysis can be stated as follows:

1. The non-normal pdf of the drought variable influences the behavior of $E(L_T)$ and $E(S_T)$ significantly. The log-normal pdf tends to lengthen the values of $E(L_T)$ and enlarge the values of $E(S_T)$ for identical conditions of C_v, C_s, ρ and T, followed by the gamma and normal pdfs. The effect of skewness is accordingly reflected in the increased values of the above parameters in relation to the normal pdf.

2. For the purpose of designing storage structures for drought periods, one can safely assume $E(S_T) \approx E(L_T)$, or the drought intensity ≈ 1 (in the standardized terms) for C_v=0.2～1, ρ=0～7, and a return period of 25～1000 years. The above approximation leads to a conservative design for water storage facilities during drought periods.

3. The skewness and the persistence in the sequences of the drought variable enhance the extremal drought durations and the severities. In particular, the effect of skewness on the extreme drought durations cannot be regarded as insignificant, contrary to the existing belief in the hydrological literatures.

4. A drought frequency formula can be derived to estimate the drought severity in relation to the return period, analogous to the flood frequency formula. The drought frequency factor, F_T, is equal to $[E(S_T)-C_v^{-1}]$ and can easily be computed by knowing pdf, C_v and ρ of the drought variable, and the desired return period "T".

This research has been supported by IDSS developing for water resources optimum allocation and management {20025700003} and by Innovatory found in Henan Province, China.

Correspondence to: Zhenmin Zhou
20 Zhenghua Road
Zhengzhou, Henan 450045, China
Telephone: 0086-0371-5790658
Cellular phone: 13015542843
E-mail: zzm1022@yahoo.com.cn

References

Bogardi JJ, et al. Practical generations of synthetic rainfall event time series in a semi-arid climatic zone. J Hydrol 1998;103:357-73.

Guven O. A simplified semiempirical approach to probabilities of extreme hydrologic droughts. WRR 1983;19:441-53.

Song H, Zhang S. Basic Analysis on Regional Drought Statistic Features. Advances in Water Science J. China. 1994;3.