Hydrological Extremes: Understanding, Predicting, Mitigating (Proceedings of IUGG 99 Symposium HS1, Birmingham, July 1999). IAHS Publ. no. 255, 1999.

Regional partial duration series modelling of hydrological droughts in Zimbabwean rivers using a two-component exponential distribution

THOMAS R. KJELDSEN, ALLAN LUNDORF & DAN ROSBJERG
Department of Hydrodynamics and Water Resources, Technical University of Denmark, Building 115, DK-2800 Lyngby, Denmark
e-mail: trk@email.isva.dtu.dk

Abstract A regional analysis of hydrological droughts is conducted using the partial duration series model with the two-component exponential (TCE) distribution as exceedence distribution. The index-flood method is applied in order to obtain more efficient estimates of the *T*-year events. Only one of the three parameters of the TCE distribution contributes significantly to the uncertainty of the *T*-year events and, therefore, only this parameter is regionalized. In total 25 gauging stations covering most of Zimbabwe were used to identify sufficiently homogeneous regions for use with the index-flood method. Separation of the stations according to mean annual precipitation indicated the existence of three or possibly more homogeneous regions, though the limited number of available stations did not allow for any rigorous conclusions.

INTRODUCTION

This study is concerned with frequency analysis of partial duration series (PDS) of the duration and deficit volume of historical streamflow drought events in selected Zimbabwean rivers. To improve the efficiency of the estimated *T*-year events, regional information is included in the analysis.

A definition of hydrological droughts based on the truncation level approach, as first described by Yevjevich (1967), is usually adopted. The basis for the drought analysis is a time series of river discharge (annual, monthly or daily discharge). The truncation level is an analytical interpretation of the water demand or the expected availability of water. When the flow in the river falls below the truncation level a drought has occurred, which ends when the flow once again exceeds the truncation level. The drought events are characterized by the two parameters, duration and deficit volume. The deficit volume is calculated as the cumulative difference between truncation level and flow in the drought period. Mathier *et al.* (1992) extended the truncation level model. They applied monthly runoff data and a truncation level which varied from month to month, see Fig. 1. Woo & Tarhule (1994) analysed PDS of drought events from four Nigerian rivers using daily discharge data and a constant truncation level, and they identified a large population of small events and a small population of large events.

The generally short flow records, combined with the fact that drought events can be very long, imply that drought samples, when insignificant events are disregarded, are usually very small, which makes the at-site estimation of *T*-year events more

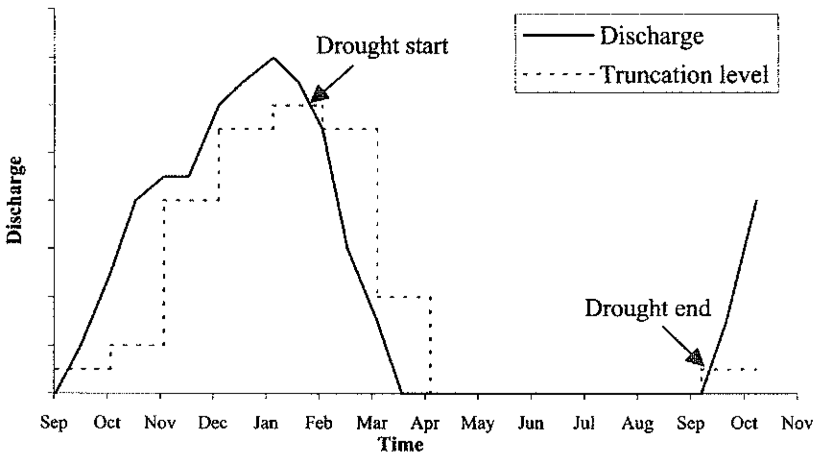

Fig. 1 Abstraction of drought events using a variable truncation level.

uncertain. The inclusion of regional information to assist the at-site estimation can, if sufficiently homogeneous regions can be identified, lead to more efficient estimates of the *T*-year events. Recently, Madsen & Rosbjerg (1997) carried out extensive work on index-flood modelling combined with PDS. This index-flood approach is applied in the following.

DROUGHT DEFINITION AND DATA

Because the expected flow in Zimbabwean rivers varies significantly from season to season, the variable truncation level approach has been adopted. A truncation level calculated as a fixed percentile of the flow duration curve for each month will reflect the general annual variability of the river flow. A flow-duration percentile of 75% has been chosen, i.e. the value exceeded 75% of the time in the distribution of the monthly flow. As ephemeral rivers were included in the investigation, a problem of defining the occurrence of a drought emerged. When considering an ephemeral river, the expected flow during the dry season is zero, which implies a truncation level equal to zero in that period. The problem emerges when a drought starts in the rainy season and continues into the following dry period. When the river runs dry in the period where it actually is expected to run dry, does the drought from the rainy season continue into this dry period? In this study, the following approach has been adopted: if a drought is present in the rainy season and continues into the dry season with zero truncation level, the duration of the drought will still increase, while the deficit volume remains constant as illustrated in Fig. 1. When the rainy season is expected to begin once again, the truncation level will be greater than zero. If the flow in the river is still below the truncation level the drought continues. If, on the other hand, the flow exceeds the truncation level, the drought has ended. Perennial rivers cause no problem, as they are not expected to run dry at any time. An important aspect to consider in drought modelling is the problem of dependency between successive drought events. The problem emerges when a prolonged drought period is interrupted by short insignificant wet periods, which divide the prolonged drought into two or more minor droughts. The presence of minor droughts instead of one major drought will distort the estimation of the parameters in the modelling of the drought events. Three methods of solving the problem were compared by Tallaksen *et al.* (1997). Based on the results from the comparison, the sequential peak algorithm (SPA) method has been adopted here. For an introduction to the SPA method in drought modelling, refer to Tallaksen *et al.* (1997).

The deficit volume events extracted from a particular station are normalized with the mean annual runoff at the considered station, which makes a comparison between different stations possible. The normalization procedure changes the unit of the deficit volume from m^3 to days.

The data analysed originate from Zimbabwe. In total 25 gauging stations categorized as measuring natural flow were selected. The length of the flow series generally varies between 20 and 40 years. The climate of Zimbabwe varies from a dry area in the southwestern part, with a mean annual precipitation (*MAP*) of 500–600 mm, to a more humid area in the Eastern Highlands (*MAP* > 800 mm). The different climatic conditions give rise to both perennial and ephemeral rivers. Both types of rivers are included in the analysis.

AT-SITE *T*-YEAR EVENT ESTIMATION

Based on the findings of Kjeldsen *et al.* (submitted) the two-component exponential (TCE) distribution is adopted as the exceedance distribution for both duration and deficit volume. The TCE distribution is a compound distribution describing one major population consisting of two exponential distributed sub-populations with mean values a_1 and a_2, where $a_1 > a_2$. The cumulative distribution function is given as:

$$\mathrm{F}(x) = p\left[1 - \exp(-x/a_1)\right] + (1-p)\left[1 - \exp(-x/a_2)\right] \quad x > 0 \tag{1}$$

The choice of threshold level is an important element in PDS modelling. In terms of a frequency factor k the selected threshold level x_0 in the basic series is given as:

$$x_0 = m + ks \tag{2}$$

where x_0 is the threshold level in PDS of drought events; m is the mean value of the uncensored series of deficit volume; s is the standard deviation of the uncensored series of deficit volume; and k is the frequency factor.

Kjeldsen *et al.* (submitted) showed that a reasonably chosen value of the frequency factor $k = k^*$ is approximately linearly dependent on the coefficient of variation C_v of the basic drought series. The relationship reads:

$$k^* = 0.1459C_v - 0.7417 \quad 1.45 \leq C_v \leq 4.45 \tag{3}$$

Equation (2) with $k = k^*$ was applied to censor the 25 drought series used in this study.

The parameters a_1, a_2 and p of the TCE distribution are estimated in the censored PDS. λ is the Poisson parameter describing the mean number of drought events per year in the major population.

Inserting $F(x) = 1 - 1/(\lambda T)$ and $x = x_T$ in equation (1) leads to:

$$p\exp(-x_T/a_2) - p\exp(-x_T/a_1) - \exp(-x_T/a_2) + 1/(\lambda T) = 0 \tag{4}$$

which must be solved by iteration to obtain x_T.

To evaluate the uncertainty of the estimated T-year events a second order Taylor approximation of the variance of x_T is applied:

$$\begin{aligned}\mathrm{var}(x_T) =& (\partial g/\partial a_1)_m^2\,\mathrm{var}(a_1) + (\partial g/\partial a_2)_m^2\,\mathrm{var}(a_2) + (\partial g/\partial p)_m^2\,\mathrm{var}(p) + \\ & (\partial g/\partial \lambda)_m^2\,\mathrm{var}(\lambda) + 2(\partial g/\partial a_1)_m(\partial g/\partial a_2)_m\,\mathrm{cov}(a_1, a_2)\end{aligned} \tag{5}$$

where $g(a_1, a_2, p, \lambda) = x_T$. Both the maximum likelihood parameter estimation procedure for the TCE distribution and the evaluation of the different parts of equation (5) are explained in depth by Kjeldsen *et al.* (submitted).

REGIONAL T-YEAR EVENT ESTIMATION

The basic assumption of the index-flood approach is that some parameters in the exceedance distribution describing the drought events can be considered identical inside a specific region. Correspondingly, if one or more of the parameters in the common distribution can be assumed constant, a homogeneous region has been identified. The parameters assumed constant (θ^R) are estimated from the at-site parameters (θ^{AS}) within the region as record weighted averages:

$$\theta^R = \sum_{i=1}^{m} w_i \theta_i^{AS} \Big/ \sum_{i=1}^{m} w_i \tag{6}$$

where m is the number of sites in the region and w_i is the number of events in the drought series from the ith station (Madsen & Rosbjerg, 1997). The concept of regionalization of a parameter is that the uncertainty of the regionally estimated parameter is smaller than the uncertainty of the at-site estimated parameter. By increasing the efficiency of the parameter estimates the uncertainty of the regionally estimated T-year events will become smaller than the purely at-site estimated. The introduction of regional information may, on the other hand, also lead to more biased estimates of the T-year events. Following the procedure presented by Kjeldsen *et al.* (submitted) for calculation of the different parts of equation (5) it is found that the parameter a_1, the mean value of the upper sub-population, generally accounts for more than 90% of the total uncertainty. Therefore, it is only attempted to regionalized a_1.

The total uncertainty of the T-year event is calculated according to equation (5). When the parameter a_1 is regionalized, the uncertainty of a_1 and the covariance between a_1 and a_2 have to be re-evaluated, leading to the revised equation for the variance of x_T:

$$\begin{aligned}\mathrm{var}(x_T) =& (\partial g/\partial a_1^R)_m^2\,\mathrm{var}(a_1^R) + (\partial g/\partial a_2^{AS})_m^2\,\mathrm{var}(a_2^{AS}) + (\partial g/\partial p^{AS})_m^2\,\mathrm{var}(p^{AS}) + \\ & (\partial g/\partial \lambda^{AS})_m^2\,\mathrm{var}(\lambda^{AS}) + 2(\partial g/\partial a_1^R)_m(\partial g/\partial a_2^{AS})_m\,\mathrm{cov}(a_1^R, a_2^{AS})\end{aligned} \tag{7}$$

where AS means at-site estimated and R means regionally estimated. The variance of the regionalized parameter a_1 is obtained by taking the variance of both sides of equation (6):

$$\mathrm{var}(a_1^R) = \left(\sum_{i=1}^{m} w_i^2 \,\mathrm{var}(a_{1,i}{}^{AS}) + 2\sum_{i=1}^{m-1} \sum_{j=i+1}^{m} w_i w_j \,\mathrm{var}(a_{1,i}{}^{AS})\,\mathrm{var}(a_{1,j}{}^{AS})\rho_{ij} \right) / \left(\sum_{i=1}^{m} w_i \right)^2 \quad (8)$$

where m is the number of sites and ρ_{ij} is the correlation between $a_{1,i}{}^{AS}$ and $a_{1,j}{}^{AS}$, the local estimates of upper population mean values at stations i and j. The covariance between $a_1{}^R$ (the regional estimate of the upper population mean value) and $a_{2,j}{}^{AS}$ (the at-site estimate of the lower population mean value at station j) is the weighted average of the covariance between the lower population mean value at station j and the upper population mean value of all the stations in the area:

$$\mathrm{cov}(a_1^R, a_{2,j}{}^{AS}) = \sum_{i=1}^{m} \mathrm{cov}(a_{1,i}{}^{AS}, a_{2,j}{}^{AS}) \,/ \sum_{i=1}^{m} w_i \qquad j = 1, \cdots, m \qquad (9)$$

The right hand side of equation (9) includes the covariance between $a_{2,j}{}^{AS}$ and $a_{1,j}{}^{AS}$, which is supposed to be greater than the covariance between all other combinations of $a_{2,j}{}^{AS}$ and $a_{1,i}{}^{AS}$ where $i \neq j$. Therefore, as a conservative estimate, the weighted average in equation (9) is approximated by:

$$\mathrm{cov}(a_1^R, a_{2,j}{}^{AS}) = \mathrm{cov}(a_{1,j}{}^{AS}, a_{2,j}{}^{AS}) \qquad j = 1, \cdots, m \qquad (10)$$

Kuczera's equation

When a homogeneous region has been identified it should be evaluated whether the introduction of regional information is beneficial or not. Kuczera's (1983) equation may serve as a guideline:

$$\Delta = \left\{ \left(a_1^{AS} - a_1^R\right)^2 / \left(\mathrm{var}(a_1^{AS}) + \mathrm{var}(a_1^R)\right) \right\} - 1 \qquad (11)$$

If $\Delta < 0$ regionalization is preferable, otherwise not.

Inter-site correlation

One very important aspect of the regional analysis is the correlation between observations from different stations. The existence of inter-site correlation between the drought series decreases the amount of information available from the regionalization, and the uncertainty of the regional parameter estimate is thereby increased. The calculation of the uncertainty of the regionally estimated parameter requires calculation of the correlation between estimates of upper population mean values at all the stations included in the regionalization. As a_1 is a mean value (of the upper population) the correlation ρ_{ij} between $a_{1,i}{}^{AS}$ and $a_{1,j}{}^{AS}$ is equal to the correlation between the series of drought events (Madsen & Rosbjerg, 1998). The calculation of the correlation between drought events in a PDS is complicated by the fact that drought events do not occur at the same time. Some stations have many small events and others have fewer but longer events. The method used to calculate the correlation structure has been adopted from Madsen & Rosbjerg (1998).

APPLICATION

Scope

In the following sub-sections it is attempted to identify homogeneous regions within Zimbabwe, corresponding to the deficit volume. The discussions and conclusions will be based on an overall judgement of how the 25 stations behave under different regionalization assumptions. However, because of limited space, only the results from station A61 and D42 will be shown as examples. Station A61 is situated in the dry southern part and station D42 in the more humid northern part.

All of Zimbabwe

First, all of Zimbabwe is considered as a homogeneous region, and the mean value of the upper population is estimated using all available 25 stations. The curve in Fig. 2 denoted "Reg. all" shows the TCE distribution fitted to all the observed deficit volume events. For both stations the regional estimation will substantially increase the bias of the estimated *T*-year events compared to the at-site estimated TCE distribution. Therefore, it is concluded that all of Zimbabwe cannot be regarded as a homogeneous region. The graphical investigation indicates different behaviour of ephemeral and perennial rivers. Thus, as the next step, and again in order to include as much information as possible in the regionalization, the rivers are divided into two groups: ephemeral and perennial rivers. It is investigated if a regionalization based on these two groups will render more acceptable results.

Ephemeral and perennial rivers

The separation resulted in 11 ephemeral rivers and 14 perennial rivers. The curves in Fig. 2 denoted "Reg. (per)" and "Reg. (eph)" show a comparison between the TCE

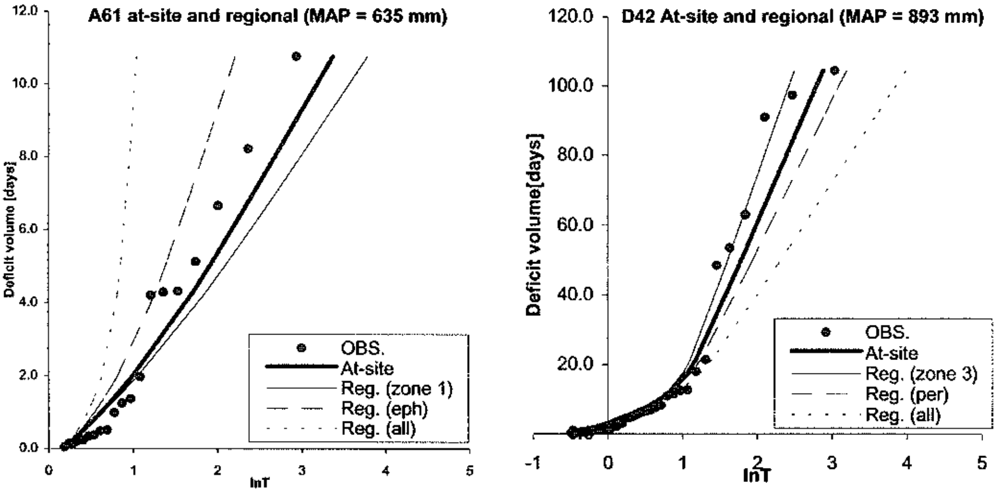

Fig. 2 Effect of various regionalizations of a_1 on deficit volume estimation.

distribution fitted to the observed deficit volumes. The sub-division of the rivers still introduces some bias as compared to at-site estimation, but a notably better fit is obtained. The ephemeral–perennial division indicates that the mean value of the upper population is correlated to *MAP*.

Mean annual precipitation

Figure 3 shows the mean value of the upper population plotted against *MAP*. From the figure the 25 stations may be sub-divided into three zones according to *MAP*:

Zone 1 (5 stations)	$MAP \leq 650$ mm
Zone 2 (9 stations)	$650 < MAP \leq 800$ mm
Zone 3 (11 stations)	$MAP > 800$ mm

Again a graphical comparison has been conducted to see if a regionalization based on MAP would result in the identification of a set of homogeneous regions. The curves in Fig. 2 denoted "At-site" and "Reg. zone" show the TCE distribution fitted to the observed deficit volume with both at-site and zonal regionalized parameters. Generally, for most stations a better fit to the observed events is obtained compared to the two previous regionalization attempts. Table 1 shows a comparison between the at-site estimated and zonal estimated 100-year events and the corresponding uncertainty for all 25 stations. The following points should be noticed referring to the three zones:

Zone 1 In the southern part of the country where persistent overestimation in the previous regionalization attempts was prevalent, the five stations are now subject to both under- and overestimation compared to the observed events (see Fig. 2 station A61). A further separation of the stations in the region could possibly identify less heterogeneous regions. Unfortunately, the lack of available data makes this impossible.

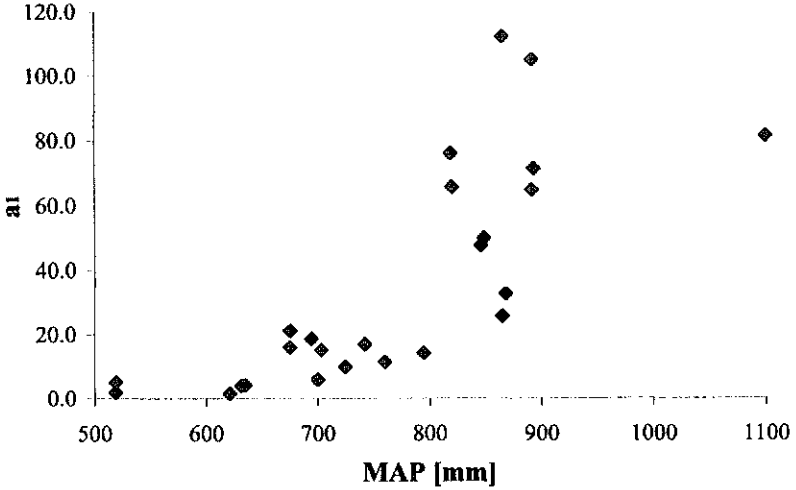

Fig. 3 Correlation between a_1 and *MAP*.

Table 1 At-site and regionalized estimates of 100-year events and corresponding variance.

Station	At-site		Regional (zones)		
	a_1 (days)	Var(a_1) (days^2)	a_1 (days)	Var(a_1) (days^2)	Δ
Zone 1					
A60	4.17	0.22	3.32	0.06	
A61	4.00	0.32	3.32	0.06	
B75	1.54	0.02	3.32	0.06	
B77	1.90	0.03	3.32	0.06	
B78	5.14	0.15	3.32	0.06	
Zone 2					
C6	5.31	0.14	12.81	1.03	47.1
C12	14.70	2.15	12.81	1.03	0.1
C18	57	0.88	12.81	1.03	4.2
C24	11.93	0.65	12.81	1.03	**-0.5**
C39	11.49	2.64	12.81	1.03	**-0.5**
C41	18.69	4.60	12.81	1.03	5.1
C70	12.99	1.93	12.81	1.03	**-1.0**
E2	13.64	1.20	12.81	1.03	**-0.7**
E49	18.55	5.90	12.81	1.03	3.8
Zone 3					
C22	27.21	5.63	49.17	14.78	
D24	61.47	41.53	49.17	14.78	
D27	64.84	17.18	49.17	14.78	
D28	41.80	7.81	49.17	14.78	
D42	43.84	38.12	49.17	14.78	
D50	29.21	6.53	49.17	14.78	
E47	76.14	37.58	49.17	14.78	
E72	74.34	56.13	49.17	14.78	
E108	25.32	7.79	49.17	14.78	
E114	42.25	51.30	49.17	14.78	
E136	123.21	544.21	49.17	14.78	

Zone 2 The stations in zone 2 are generally well described with the regionalized parameters. Kuzcera's (1983) equation indicates that four of the nine stations would benefit from the regionalization (see Table 1).

Zone 3 Zone 3 includes stations with values of *MAP* greater than 800 mm. Within zone 3 the same degree of correlation between a_1 and *MAP* as observed within zones 1 and 2 is no longer present. Furthermore, a large variability is observed between the estimated values of a_1. Despite the seemingly large variability of a_1 in the zone, no rigorous conclusions concerning the heterogeneity of the zone can be stated, as the observed variability of a_1 can be related to both heterogeneity and natural variability of the drought series. A more accurate picture of the variability of a_1 could be obtained if more stations became available, but this is not possible. A useful approach would be to employ catchment characteristics to investigate if other physical factors have significant influence on the drought series, e.g. catchment area, slope, soil type, etc. Thereby, less heterogeneous regions might be identified within zone 3.

Kuzcera's (1983) equation is not applied to zones 1 and 3 because of the homogeneity problems.

CONCLUSION

An investigation of hydrological droughts within Zimbabwe was carried out. The TCE distribution was used as exceedance distribution in PDS modelling of both duration and deficit volume. It was attempted to identify regions within Zimbabwe sufficiently homogeneous to regionalize parameters of the TCE distribution of the deficit volume. A separation of the available gauging stations according to *MAP* indicated the existence of such homogeneous regions. However, the limited number of available stations in the southern area (zone 1) made verification difficult. Zone 2 ($650 \text{ mm} < MAP < 800 \text{ mm}$) could be considered homogeneous and four of the nine stations benefited from the use of regional information. In the more humid area (zone 3) a positive correlation between a_1 (the mean value of the upper population) and *MAP* was not present. The large variability of a_1 within the zone makes it difficult, based on the amount of data at hand, to draw firm conclusions regarding the possible existence of a homogeneous region.

Generally, the gauging network available within Zimbabwe is not dense enough to allow for rigorous conclusions regarding the regional behaviour of droughts. Further analysis should extend the network by including additional data from neighbouring countries. Additional catchment characteristics should be included in the analysis to identify physical factors influencing the drought series.

Acknowledgement The authors express their thanks to Department of Water Resources—Hydrological Branch, Zimbabwe for kindly providing the data.

REFERENCES

Kjeldsen, T. R., Lundorf, A. & Rosbjerg, D. (submitted) Use of a two-component exponential distribution in partial duration modelling of hydrological droughts in Zimbabwean rivers. Submitted to *Hydrol. Sci. J.*

Kuczera, G. (1983) Effect of sampling uncertainty and spatial correlation on an empirical Bayes procedure for combining site and regional information. *J. Hydrol.* **65**, 373–398.

Madsen, H. & Rosbjerg, D. (1997) The partial duration method in regional index-flood modeling. *Wat. Resour. Res.* **33**(4), 737–746.

Madsen, H. & Rosbjerg, D. (1998) A regional Bayesian method for estimation of extreme streamflow droughts. In: *Statistical and Bayesian Methods in Hydrological Science* (ed. by E. Parent, B. Bobée, P. Hubert & J. Miguel), 327–340. UNESCO Studies and Reports in Hydrology, UNESCO, Paris.

Mathier, L., Perreault, L., Bobée, B. & Ashkar, F. (1992) The use of geometric and gamma-related distributions for frequency analysis of water deficit. *Stoch. Hydrol. Hydraul.* **6**(4), 239–254.

Tallaksen, L. M., Clausen, B. & Madsen, H. (1997) On the definition and modelling of streamflow drought duration and deficit volume. *Hydrol. Sci. J.* **42**(1), 15–33.

Woo, M.-K. & Tarhule, A. (1994) Streamflow droughts of northern Nigerian rivers. *Hydrol. Sci. J.* **39**, 19–34.

Yevjevich, V. (1967) An objective approach to definition and investigation of continental hydrological droughts. *Hydrol. Paper no. 23, Colorado State University, USA.*