



Journal of Urban and Environmental
Engineering

E-ISSN: 1982-3932

celso@ct.ufpb.br

Universidade Federal da Paraíba
Brasil

Onyutha, Charles

STATISTICAL MODELLING OF FDC AND RETURN PERIODS TO CHARACTERISE QDF AND
DESIGN THRESHOLD OF HYDROLOGICAL EXTREMES

Journal of Urban and Environmental Engineering, vol. 6, núm. 2, 2012, pp. 132-148

Universidade Federal da Paraíba
Paraíba, Brasil

Available in: <http://www.redalyc.org/articulo.oa?id=283224989013>

- How to cite
- Complete issue
- More information about this article
- Journal's homepage in redalyc.org

redalyc.org

Scientific Information System

Network of Scientific Journals from Latin America, the Caribbean, Spain and Portugal

Non-profit academic project, developed under the open access initiative

STATISTICAL MODELLING OF FDC AND RETURN PERIODS TO CHARACTERISE QDF AND DESIGN THRESHOLD OF HYDROLOGICAL EXTREMES

Charles Onyutha*

Department of Civil and Environmental Engineering, Makerere University, Uganda

Received 25 April 2012; received in revised form 10 December 2012; accepted 20 December 2012

Abstract:

In this paper, firstly, flow duration curves (FDCs) for hydrological extremes were calibrated for a range of aggregation levels and seasons to provide compressed statistical information for water resources management at selected temporal scales and seasons. Secondly, instead of the common approach of using return periods, T (years) for deriving discharge duration frequency (QDF) relationships, the method of using exceedance frequencies, E (%) was introduced so as to provide answer to important question like, what is the streamflow at a given aggregation level and selected E (%)? Thirdly, the concept of estimated design threshold (EDT) was introduced and proposed for consideration in the risk analysis for design of water resources structures. This study was based on the long daily discharge record for the period 1950 – 2008 at station 1EF01 in Kenya, on the Nzoia river with watershed area of 12,676 km² located in the North Eastern quadrant of Lake Victoria Nile Sub Basin. In the statistical modelling of FDCs and T (years), suitable extreme value distributions (EVD) were selected and calibrated to fit nearly independent high flows and low flows. The FDCs and T -curves were used to determine the EDT. The FDCs were used to model the QDF relationships. To derive QDF relationships of hydrological extremes, for a given range of aggregation levels, extreme value analysis (EVA) was carried out and suitable EVD selected. Next was the calibration of parameters of the EVD and analysis of relationship between the model parameters and aggregation levels. Finally, smooth mathematical relationships were derived using little but acceptable modifications to the model parameters. Such constructed QDF relationships can be used for various applications to estimate cumulative volumes of water available during droughts or floods at various aggregation levels or E (%) of hydrological extremes. The EDT when obtained for a range of aggregation levels can also be used to understand climate variability, and the modeled FDCs can be useful in enhancing forecasts of hydrologic variables under climate variability and/or change.

Keywords: Hydrological extremes; EVA; FDC; QDF; design threshold; return periods; climate.

© 2012 Journal of Urban and Environmental Engineering (JUEE). All rights reserved.

* Correspondence to: Charles Onyutha, Tel.: +256 752 711163; +256 712 711163. E-mail: conyutha@gmail.com, conyutha@tech.mak.ac.ug

INTRODUCTION

A flow-duration curve (FDC) provides the percentage of time (duration) a daily or monthly (or some other time interval) streamflow is exceeded over a historical period for a particular river basin (see Vogel & Fennessey, 1994). The credit for the first usage of FDC goes to Clemens Herschel in about 1880 (Foster, 1934). Applications of FDC are of interest for many hydrological problems related to river and reservoir sedimentation, hydropower generation, water quality assessment, water-use assessment, water allocation and habitat suitability (Castellarin *et al.*, 2004). According to WMO (2008) in the Operational Hydrology Report No. 50, the FDC as a key tool for the sustainable management of water resources can be estimated from gauged data or, ungauged sites, from statistical or simulation models of the natural regime, historical regime and the target regime (**Fig. 1**). Determining these FDCs enables estimation of “abstractable volume” (the maximum volume of water that can be abstracted from the river without resulting in an unacceptable deterioration in instream ecology or an adverse impact on the downstream water users) and a “hands-off flow” (the discharge at which abstraction must cease).

A number of studies and reviews on FDC have been conducted and/or presented in literature (see among others, Vogel & Fennessey (1994, 1995), Studley (2001), Croker *et al.* (2003), Fennessey & Vogel (1990), LeBoutillier & Waylen (1993), Smakhtin (2000), Nathan & McMahon (1992), Hironobu *et al.* (2003)). Mandal & Cunnane (2009) developed a simple regression based generalized model for the FDC for Irish rivers which can be used for predicting FDC for any ungauged catchment from the known catchment physiographic and climatological characteristics. Alternative parameterizations of FDCs have been examined in a number of studies.

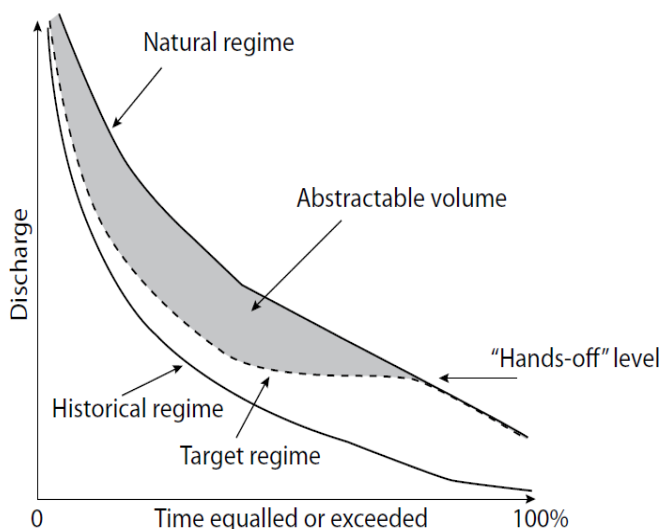


Fig. 1 Abstraction of streamflow based on FDC (adopted from WMO, 2008).

Best *et al.* (2003) found that an adequate fit for examining the impact of land use change on hydrology using the FDC can be given by a 5-parameter model. Cigizoglu & Bayazit (2000) also utilized a 5-parameter harmonic model in their study on FDC. Castellarin *et al.* (2004) introduced an index flow approach to modelling the relationship between an FDC and Annual Flow Duration Curves (AFDCs) of daily streamflow series, which can reproduce the FDC, as well as the mean, median, and variance of the AFDCs without resorting to assumptions regarding the seasonal or persistence structure of daily streamflow series. Nathan & McMahon (1992) adopted a systems approach in which they used multivariate techniques to develop relationships between low flow parameters and climatic and land information data. Cole *et al.* (2003) proposed the use of long-term flow duration curves as an indicator of data quality. They argued that their proposed method visually highlights irregularities in river flow data and enables the type and location of the error to be readily located.

Proper management of water resources under global climate change and/or anthropogenic influence is an important key to development. While the increase in the occurrence of hydrological extremes including floods, droughts and water quality problems continue to unavoidably hamper development, the characterization of such hazards lies in an important concept of their predictability, the 'return period'. Return period is the recurrence interval between one event and the next, either of equal or larger magnitude. Since it jointly describes the probability of occurrence of extreme events and the temporal dimension, return period can be taken as a measure of safety (inverse proportional to the 'risk'); and thus, flood protection, drought preparedness and water quality standards can be set out or put in place. It's important for design of hydraulic structures such as dams, bridges, culverts, sewer systems etc which can be gears of development. Although return periods are often fixed by regulations or follow from a risk or cost-benefit analysis, this paper presents the concept of the estimated design threshold (EDT) to provide the basis for the selection of return periods to be considered for the design of water resources structures.

First and foremost, there is need to note that, most water resources engineers use FDCs in their empirical forms. This is because what the engineers deem momentous are the frequency and the mean of the flows e.g. to produce electricity. However, important is also the extreme flows regime which can only be captured through the extreme value analysis (EVA) and calibration of a suitable distribution for the data. Nathan & McMahon (1992) were able to fit low flow frequency curves of a number of catchments in the south-eastern Australia. They used their methodology presented in Nathan & McMahon (1990b) to fit the frequency curves. However, in fitting of the frequency

distributions, Nathan & McMahon (1992) assumed the low flow minima to follow a Weibull distribution. Such an assumption is subject to the systematic over/underestimation in the tail of the distribution; to avoid such a problem, in this study, quantile plots or $Q-Q$ plots as suggested by Willems *et al.* (2007) was adopted. According to Willems *et al.* (2007), using $Q-Q$ plots, analysis can be made of the shapes of the distribution's tail and discrimination can be made between heavy tail and normal tail behavior. In an exponential $Q-Q$ plot, a normal tail shows a straight line at the upper tail points while a heavy tail is characterized by a continuous bending up of the upper tail points. In a Pareto $Q-Q$ plot, a normal tail continuously bends down while the heavy tail is typified by a straight line at the upper tail points.

Secondly, the practice in most cases is to use a single time scale e.g. monthly, annually etc to derive an FDC; however, consideration of time scales of various resolutions like in this study, makes it possible to obtain relationships with more compressed statistical information that can be availed to a water resource manager. Nathan & McMahon (1992) were able to obtain curves corresponding to the 1, 7, 15, 30, 60, 120, 183, and 284 day annual minima. In this study, such principle of using a range of aggregation levels was extended to model, using independent identically distributed (iid) streamflow data of sufficient record length of about 60 years to model FDCs of hydrological extremes for the various seasons for Nzoia river (station 1EF01) at Rwamba Bridge in Kenya.

EVA carried out for a range of aggregation levels, constitute amplitude/duration/frequency relationships which summarize the most important statistical information available in a hydrological time series. This relationship can be called QDF (for discharges) or IDF (intensity for rainfall). Such relationships were first established as early as 1930s (Bernard, 1932).

Computed relationships over a range of relevant frequencies provide substantial compressive information. Aggregation levels are simply durational intervals over which the hydrological values are averaged. Premised on such durations, the conditional relationships are essentially cumulative functions of the amplitude values in the time series (Chow *et al.*, 1988). Although it is common in practice to consider aggregation levels of 7, 10 or 30 days, according to Institute of Hydrology (1980), values as high as 365 days have been reported.

Amplitude/duration/frequency relationships in the form of QDF or IDF have been presented in a number of studies (see among others, Taye & Willems, (2011), Juraj & Taha, (2007), Borga *et al.* (2005), Maurino, (2004), Zaidman *et al.* (2003), Javelle *et al.* (2002), Willems, (2000), Javelle *et al.* (1999)). Importantly, the note to be taken into account is that, all such amplitude/duration/frequency relationships are

dependent on T (years) and may not answer an important question like, what is the streamflow at aggregation level of one week with say, 20% E (%)? Such a question can only be answered by transformation of the T (years) into E (%) in the QDF relationships as shown in this paper. The FDC and T-curve for the same sample were jointly important to select what is, in study, termed as an EDT. The EDT is characterized by a high T (years) and low E (%). Hydrologic design is the process of assessing the impact of hydrologic events on a water resource system and choosing values for the key variables of the system so that it will perform adequately. There are many factors besides hydrology that bear on the design of water resource systems; these include public welfare and safety, economic, aesthetics, legal issues, and engineering factors such as geotechnical and structural design (Chow *et al.*, 1988). Water resources planning and management are two-fold; firstly in the hydropower generation, usage and management of domestic and industrial water supply, fish and wildlife improvement, recreation, low-flow augmentation for water quality management, and watershed management; secondly in the water control such as sediment control, drainage, salinity control, and pollution abatement. According to Chow *et al.* (1988), the Estimated Limiting Value (ELV) defined as the largest magnitude possible for a hydrologic event at a given location based on the best available hydrologic information; is implicit in the commonly used Probable Maximum Precipitation (PMP) and the corresponding Probable Maximum Flood (PMF). The PMP is defined by the World Meteorological Organization (1983) as the "quantity of precipitation that is close to physical upper limit for a given duration over a particular basin". Based on the worldwide records, the PMP can have a return period of as long as 500,000,000 years, corresponding approximately to a frequency factor 15. However, this return period varies geographically. Some would arbitrarily assign a return period, say 10,000 years, to the PMP or PMF, but this suggestion has no physical basis (Chow *et al.*, 1988). However, in this paper a related concept of EDT defined as the critical magnitude of hydrologic event above which significant damage to public life and property is highly likely to result, is introduced. The EDT is statistically obtainable as the intersection point of the FDC and Return Period curve when they are presented on the same axes. The EDT should be obtained from an adequately long hydrological data record to limit the uncertainty that might result from the EVA yielding FDC and the corresponding return periods. Based on the fact that the water on earth is generally constant (i.e. the global hydrological process is a closed system), the upper limit of the hydrological design scale is not infinite, although usually unknown and the lower limit of the design scale is zero. Given the knowledge of the feasible range of the hydrologic events and the corresponding damage that

can result, the optimum design T (years) can be determined by hydroeconomic analysis. The capital cost of a hydrologic structure increases as the design T (years) increases, yet the expected damage reduces due to enhanced protection catered for. Whereas the ELV may be adopted as the approximate upper limit of the design scale, the resultant design for the engineering structures or projects is normally uneconomical. Eventually, further adjustments are often undertaken to arrive at the final design value based on cost, risk analysis, and hydroeconomic analysis. What remains unclear in the risk analysis is the lower limit for the range over which the analysis can be done. Normally, there is the critical point above which design consideration for safety should be carefully undertaken as it marks somewhat the onset of the region over which remarkable damage to public life and property most likely results; in this paper, therefore the EDT is proposed to be the lower limit for the range over which risk analysis can be carried out.

THEORY

Extreme Value Analysis (EVA)

Considering α , and k as scale and shape parameters respectively, the probability distribution, $G(x)$ of the extremes above threshold x_t calibrated to t observations from a sample of size s for generalized Pareto distribution (GPD) of Pickands (1975) can be given by:

$$G(x) = 1 - \left\{ 1 + k \left(\frac{x - x_t}{\alpha} \right) \right\}^{-\frac{1}{k}} \quad (1)$$

The cumulative distribution function (CDF) of GPD with $k = 0$ is an exponential distribution given by:

$$G(x) = 1 - \exp \left\{ - \left(\frac{x - x_t}{\alpha} \right) \right\} \quad (2)$$

In the studies on low flows, mostly Weibull or Fréchet distributions are used. Melesse *et al.* (2010) is an example of such studies. The more general Fréchet distribution is found for X . For m independent low flow minima,

$$F(x) = P(X \leq x) = \exp \left(- \frac{x^{-\tau}}{\alpha} \right) \quad (3)$$

$$\alpha = \frac{x_t^{-\tau}}{c} \quad \text{and} \quad c = -\ln \left(\frac{t}{m} \right) \quad (4)$$

However, application of a transformation ($1/Q$) to streamflows enables low flows to follow GPD or

exponential instead of Weibull or Fréchet distribution. This allows for the EVA to be performed normally like for high flows. Eventually, the validity of the Fréchet distribution can be checked in the Weibull $Q-Q$ plot after transformation $1/Q$. When an exponential tail is observed for Y , then an exponential EVD can be calibrated above a specific threshold y_t :

$$G(y) = P\{Y \leq y | Y \geq y_t\} = 1 - \exp \left(- \frac{y - y_t}{\alpha} \right) \quad (5)$$

This equation can be transferred as follows towards a distribution for X : (using $x_t = 1/y_t$):

$$G(y) = P\{X \leq x | X \leq x_t\} = P\{Y \geq y | Y \geq y_t\} \quad (6)$$

$$G(y) = 1 - P\{Y \leq y | Y \geq y_t\} \quad (7)$$

$$G(y) = \exp \left(- \frac{y - y_t}{\alpha} \right) = \exp \left(- \frac{x^{-1} - x_t^{-1}}{\alpha} \right) \quad (8)$$

$$G(x) = P\{X \leq x | X \leq x_t\} \quad (9)$$

$$G(x) = \exp \left(- \frac{x^{-1} - x_t^{-1}}{\alpha} \right) = \frac{\exp \left(- \frac{x^{-1}}{\alpha} \right)}{\exp \left(- \frac{x_t^{-1}}{\alpha} \right)} \quad (10)$$

Eq. (10) matches the Fréchet distribution for $\tau = 1$ (conditional distribution for values lower than x_t).

Quantile Plots, (Q-Q plots)

In a Q-Q plot, empirical quantiles which match the observed extremes x_i , $i=1, \dots, m$ ($x_1 \leq \dots \leq x_m$), are shown against theoretical quantiles. Weibull plotting position of a quantile plot p_i with a score ($0 \leq c \leq 1$) as expressed below as their corresponding probabilities of exceedance is used (Willems, 1998a,b).

$$p_i = \frac{i}{m + c} \quad (11)$$

Considering $G(x)$ to denote CDF being tested in a $Q-Q$ plot, for each empirical quantile x_i , the theoretical quantile can be defined as $G^{-1}(1-p_i)$. Since in practice, the important matter would be to test the validity of the distribution $G(x)$ without knowledge of the parameter values, the adapted $Q-Q$ plots with the so-called quantile function $U(p)$ plotted instead of the inverse distribution $G^{-1}(1-p)$ are therefore used. The function

$U(p)$ can be taken to refer to the simplest function of p that is linearly dependent on $G^{-1}(1-p_i)$ and independent of the parameter values of $G(x)$. The same quantile function (reduced variate) as below, is found for the case of exponential, Pareto and Weibull quantile plots although it does not exist, for all types of distributions:

$$U(p) = -\ln(p) = -\ln(1 - G(x)) \quad (12)$$

With $U(p)$ or $\ln(U(p))$ on the horizontal axis and x or $\ln(x)$ on the vertical axis, exponential, Pareto and Weibull QQ -plots can respectively be drawn from the expressions:

$$\{-\ln(p_i); x_i\} \quad (13)$$

$$\{-\ln(p_i); \ln(x_i)\} \quad (14)$$

$$\{\ln\{-\ln(p_i)\}; \ln(x_i)\} \quad (15)$$

The optimal threshold, x_t above which the weighted regression has to be performed in an optimal estimation of k , is the threshold that minimizes the mean squared error (MSE) of the regression. Let j be the rank of events, and t , number of observations above the threshold; the MSE of weighted linear regression in an exponential $Q-Q$ plot is:

$$MSE_t = \frac{1}{t-1} \left\{ \sum_{j=1}^{t-1} w_j \left(x_j - x_t - \hat{\alpha}_t \ln\left(\frac{t}{j}\right) \right)^2 \right\} \quad (16)$$

$$\text{Using Hill weights } w_j = \left\{ -\ln\left(\frac{j}{t}\right) \right\}^{-1} \quad (17)$$

The estimate of parameter α of the exponential distribution can be obtained using:

$$\hat{\alpha}_t = \frac{\sum_{j=1}^{t-1} w_j \ln\left(\frac{j}{t}\right) (x_t - x_j)}{\sum_{j=1}^{t-1} w_j \ln^2\left(\frac{j}{t}\right)} \quad (18)$$

The Hill-type estimator for parameter α of the exponential distribution can be obtained using Hill-weights and is given by:

$$\hat{\alpha}_t = \frac{1}{t-1} \left\{ \sum_{j=1}^{t-1} x_j \right\} - x_t \quad (19)$$

The Concept of Temporal Aggregation

Importantly, to mimic the delayed response of a watershed, temporal aggregation can be applied to an

observed hydrological time series. This is done by passing n -day averaging window to smoothen the time series. Temporal aggregation transforms the hydrological time series e.g. from hourly to daily, daily to weekly, daily to monthly, daily to annual etc. Hydrological time series in aggregated form can be derived in two ways; one is by an overlapping n -day averaging window, and the other is by non-overlapping n -day averaging window. The former is mainly applicable to time series of continuous hydrological events (variables) e.g. discharge, temperature of water, water levels etc, while the latter can be used for discontinuous hydrological variables e.g. precipitation. The latter method is important in the extraction of partial duration series (PDS) such as averaged weekly, monthly, yearly, seasonal etc. The treatments accorded to the two methods vary as well. The overlapping method which can be applied e.g. to discharge constitutes averaging window while for the non-overlapping method of aggregation, the sum may be used e.g. for rainfall. Although a general formula covering both methods of averaging temporal aggregation is presented in **Eq. (20)**, only the overlapping method was used in the study.

$$Q_j = \frac{1}{a} \sum_{i=x}^{i=y} x_i \quad \text{for } 1 \leq i \leq a \text{ and } 1 \leq j \leq m \quad (20)$$

where m = the sample size for the x variable, j = the block of the time series under consideration, i = the position of the variable x in j , a = the aggregation level, x_i = the value of the variable at position i , Q_j = is the resulting averaged value of each block, j . For the general formula presented in **Eq. (20)**:

a) For overlapping method,

$$x = j \quad (21)$$

$$y = j + (a-1) \quad (22)$$

b) For non-overlapping method,

$$x = a(j-1) + 1 \quad (23)$$

$$y = aj \quad (24)$$

The choice of aggregation levels can be made by the management of water resources of a particular watershed. For hydrological time series, the largest aggregation level relevant for an application corresponds to the recession time of the river catchment under study (for the relationships being constructed). The recession time can be taken to refer to the maximum period of time during which the joint

influence from the time series on an indicator variable can be realized. For hydrological extremes, i.e. flooding and droughts, the indicator variable equals the magnitude of stream flows (discharges) or the water levels. For pollution, the indicator variables may be in the equal to the concentrations of physico-chemical water quality variables.

Flow Duration Curves (FDC) and Return Period

The concept of return period is an important one because it enables the determination of risk (economic or otherwise) associated with a given flood or drought magnitude. Let n denote the record length in years of the observed time series, t represent the streamflow rank; the empirical return period, T_e is:

$$T_e[\text{years}] = \frac{n}{t} \quad (25)$$

Taking x_T to denote T-year event, the computations of theoretical return period, T for high flows (**Eq. (26)**) and low flows (**Eq. (27)**) are based on the exceedance probability $[1 - G(x)]$ and non-exceedance probability $G(x)$ respectively as given by:

$$T[\text{years}] = \left(\frac{n}{t}\right) \frac{1}{1 - G(x_T)} \quad (26)$$

$$T[\text{years}] = P[X \leq x | X \leq x_t] = \left(\frac{n}{t}\right) \frac{1}{G(x)} \quad (27)$$

Considering s as the sample size, and using Weibull plotting position of a quantile plot p_i with a score of zero and replacing i from **Eq. (11)** by t , the empirical exceedance frequency, E is given by;

$$E(\%) = \frac{100t}{s} \quad (28)$$

Just like for T (years), the exceedance frequency, E of the exceedance level x_E for high flows (**Eq. (29)**) and low flows (**Eq. (30)**) are computed based on the exceedance probability $[1 - G(x)]$ and non-exceedance probability $G(x)$, respectively.

$$E[\%] = \left(\frac{100t}{s}\right) \frac{1}{1 - G(x_E)} \quad (29)$$

$$E[\%] = P[X \leq x | X \leq x_t] = \left(\frac{100t}{s}\right) \frac{1}{G(x_E)} \quad (30)$$

The exceedance frequency, E of high flows with exponential distribution is therefore given by:

$$E[\%] = \left(\frac{100t}{s}\right) \left\{ \exp\left(\frac{x - x_t}{\alpha}\right) \right\}^{-1} \quad (31)$$

Eq. (31) indicates that, the slope of the calibrated EVD is negative. On the basis of the linear regressions in the Q-Q plot as explained in Beirlant *et al.* (1996) and used in Willems *et al.* (2007), E-percentage event (x_E) for the exponential case, therefore can be given by:

$$x_E = x_t + \hat{\alpha} \left[\ln\left(\frac{100t}{s}\right) - \ln(E) \right] \quad (32)$$

where t : number of observations above threshold, s : sample size, $\hat{\alpha}$: estimate of α , x_t : threshold. Making t the subject from **Eq. (28)** and substituting into **Eq. (25)** we obtain **Eq. (33)** below.

$$T_e = \frac{100n}{sE} \quad (33)$$

Using **Eq. (33)** we can transform an FDC into a return period curve if we have E (%); however, if we have T_e (years), we can make E the subject and transform a return period curve to an FDC.

The Concept of Estimated Design Threshold (EDT)

When a set of streamflow ranked from highest to lowest is plotted against a log-transformed return period, or log-transformed exceedance frequency, a line is obtained. The slope of the obtained line is positive (for return period) and negative (for FDC). In case, the same log-transformed abscissa is used for both return period and FDC, the two curves intersect. The discharge corresponding to the intersection of the FDC and return period curve can be called the EDT i.e. the discharge at which the values of its return period and the percentage of the times it is exceeded coincide. Since EDT is characterized by a high value of T (years) and low E (%), the discharge is termed so because it acts as the threshold above which considerable or extensive damage to life and property is likely to result from the hydrological event.

One important note to consider is that, the EDT is a single value from a streamflow sample irrespective of the sample size and record length (years) of the dataset. However, if we use **Eqs (25)** and **(28)** directly, the value we obtain for the EDT will vary with the sample size and the record length of the dataset (see **Fig. 5**). Eventually, there is the need to rescale E , from its expression out a total of 100% to a value equal to the record length, n (in years) of the dataset. The rescaling procedure enables imagery exactness of the T-curve and

the FDC about the point of their intersection. The rescaled exceedance frequency (E_r) is given by:

$$E_r = n \left(\frac{i}{s} \right) \quad (34)$$

With the rescaled exceedance frequency (E_r) and T_e , we can derive the exceedance frequency at the EDT, (E_c) or the return period at the EDT, (T_c) in three steps:

Step 1: Since for the EDT $T_e = E_r$, using **Eqs (25) and (34)**

$$i = \sqrt{s} \quad (35)$$

Step 2: Using i as the subject of **Eqs (25) and (34)**

$$T_e = \frac{n^2}{sE_r} \quad (36)$$

Step 3: Using n as the subject of **Eqs (25) and (34)**

$$T_e = \frac{sE_r}{i^2} \quad (37)$$

Finally, by equating **Eqs (36) and (37)** and using **(35)**:

$$E_r = T_e = \frac{n\sqrt{s}}{s} \quad (38)$$

To obtain the unrescaled exceedance frequency for the EDT (E_c) from the rescaled one, E_r , we replace n in **Eq. (38)** by 100% as below.

$$\text{Using } n, E_c = \frac{100\sqrt{s}}{s} \quad (39)$$

$$\text{Using } E_r, E_c = \frac{100E_r}{n} (\%) \quad (40)$$

Fig. 2 shows an illustrative concept of EDT. From **Fig. 2**, the location of EDT also denoted by Q_c on the curve EF is determined by point g . Point g is at the intersection of curve EF and the bisector of $x - y$ plane i.e. a line drawn from the origin, O to point g makes an angle of 45° with respect to x or y axis i.e. $\overline{Oh} = \overline{Oi}$. To determine Q_c , point g is located and extrapolated to h and i . Perpendiculars at points h and i meet curves AB and CD to determine E_c and T_c , respectively.

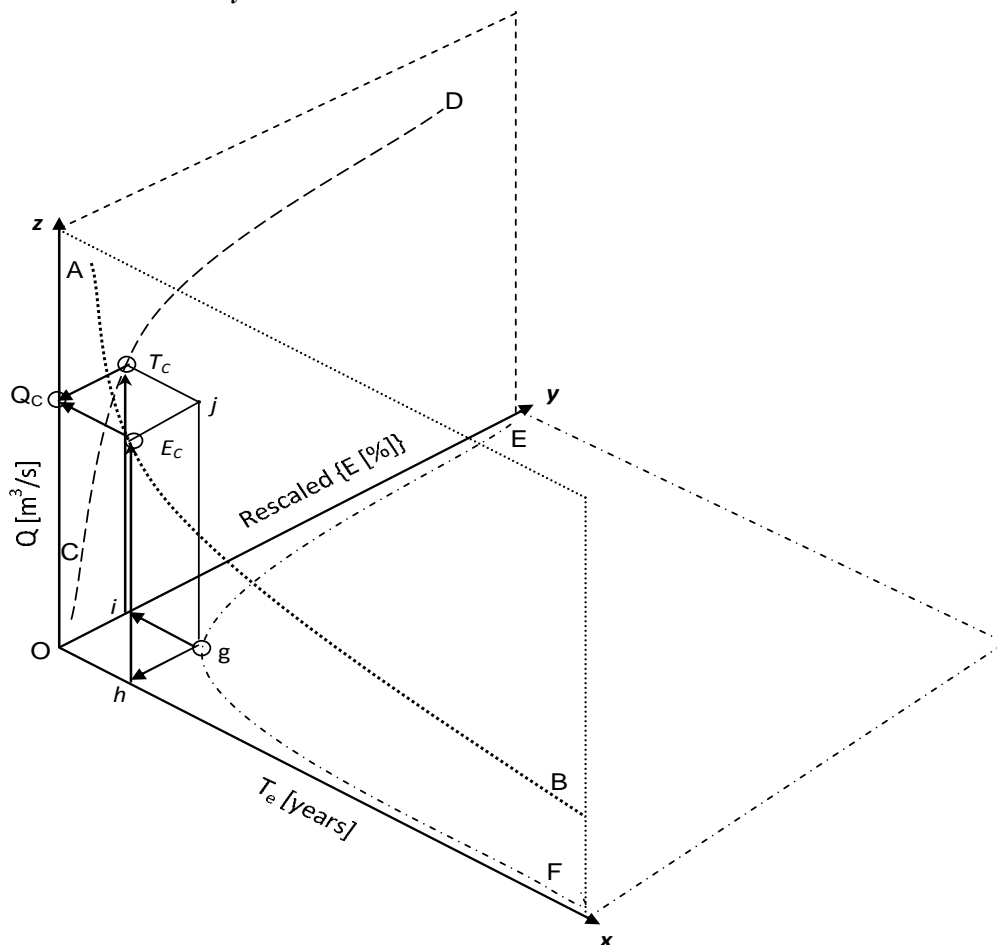


Fig. 2 The concept for defining EDT. The x , y and z axes are for rescaled exceedance frequency (years), return period (years) and discharge (m^3/s) respectively. The curves AB and CD in $x - z$ and $y - z$ planes are rescaled FDC and return period curves respectively. Curve EF is in plane $x - y$. The points E_c , T_c and g are on curves AB, CD and EF respectively. Q_c is EDT and is defined by T_c (critical return period and E_c (critical E_r)).

METHODOLOGY

Flow Duration Curve

First and foremost, to perform EVA, nearly iid extreme streamflow events were extracted from the full time series. The criteria for extraction of the iid events are based on the principle of USWRC (1976) and Lang *et al.* (1999) in which, subsequent rainfall-runoff peaks can be considered largely independent if the inter-event time exceeds the recession time and if the lowest inter-event discharge drops below a specific low flow level; see Willems (2009) for details on the method.

Empirically, FDC is constructed by first ranking the discharges from the highest to the lowest. The ranks are assigned in such a way that the highest ranking streamflow is allotted a value of 1 and the smallest discharge is given the highest rank. The E (%) for each streamflow is then computed using Eq. (28). The choice of the score c in the Weibull plotting position (Eq. (11)) is a crucial consideration in the EVA. In this study before proceeding, the choice of c is hereby verified. From a hydrological flow data of daily resolution and record length (n), 59 years; for the daily aggregation level, 145 nearly iid peaks over threshold (POTs) were sampled in total with the maximum and minimum flows being, 936.264 m³/s and 241.6 m³/s, respectively. Using $c = 1$ at $i = 108$, p_i was found to be 73.9726 %; but using $c = 0$, at $i = 108$, p_i was found to be 74.48276 %. The underestimation was minimal because the sample size was fairly large. With more stringent criteria for extraction of the iid POTs (see Willems (2009) for details on the method) purposefully, the sample size was reduced to 45 POTs in total; similarly at $i = 45$, p_i was found to be 97.8261 % (for $c = 1$) and 100 % (for $c = 0$) respectively. This shows that, when $c = 1$, underestimation of the exceedance probability is made depending on the sample size. Eventually, in this study we adopted $c = 0$.

In the Q - Q plot, the ranked discharge is plotted against the quantile function. In this study, with E as defined in Eq. (28), the $U(p)$ that was adopted for the FDC is:

$$U(p) = \ln(E) \quad (41)$$

Prior to the extraction of the extreme values from the time series for each of the selected stations, n -day moving averaging window was passed through the series. Aggregation levels of 1 day up to 1 year were considered. This is the range covered by multipurpose applications e.g. agricultural, irrigation, power plants, domestic supply, pollution etc. To come up with the amplitude/duration/frequency relationships, for the selected range of aggregation levels, EVA was carried

out and the suitable EVD selected. To enable an adequate selection of the most optimal threshold level and to avoid systematic over-/underestimation in the tail of the distribution, quantile plots or Q - Q plots were considered. The same principle of calibrating the EVDs by a weighted linear regression in the Q - Q plot suggested by Beirlant *et al.* (1996) and used by Willems *et al.* (2007) was adopted for this study. The weighting factors proposed by Hill (1975) were considered.

Fig. 3 shows examples of calibrated EVDs as linear regression lines in exponential Q - Q plots. As explained in Beirlant *et al.* (1996) and Willems *et al.* (2007), linear upper tail behavior as in Fig. 3 means that the tail can be described by an exponential EVD (which is a special case of the Generalized Pareto Distribution with zero shape parameter).

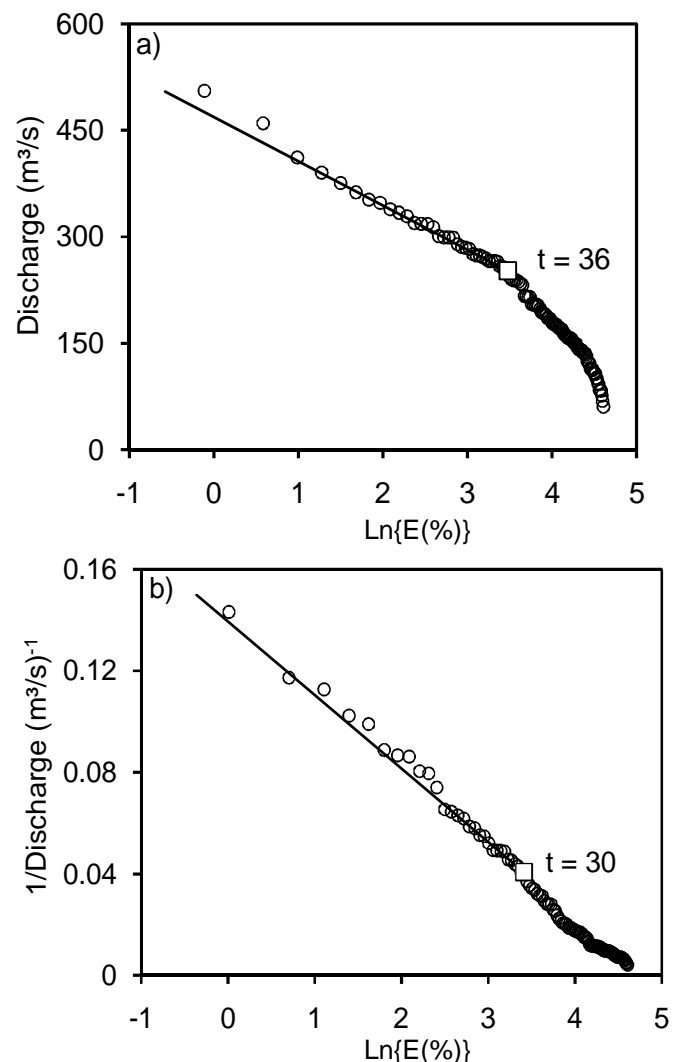


Fig. 3 Observations (o) in exponential Q - Q plots of discharges at aggregation level of 2 months; graph a) is for high flows and b) is for low flows. Symbol (□) denotes the rank for the optimal threshold; and the regression line is the calibrated EVD. $E(\%)$ is the exceedance frequency.

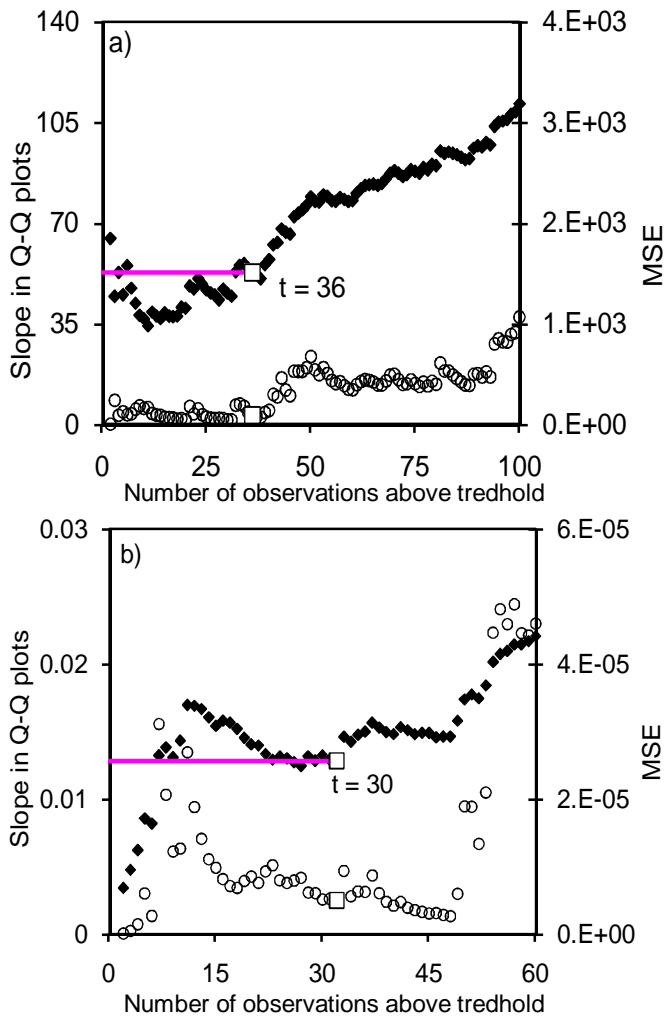


Fig. 4 Left vertical axis (\blacklozenge), Hill-type estimation of slope in the exponential Q-Q plot; right vertical axis (\circ), mean squared error (MSE) of Hill-type regression in the exponential Q-Q plot; (\square) selected optimal threshold; the graphs a) (for high flows) and b) (for low flows) are for aggregation level of 2 months.

The selection of optimal threshold values x_t above which the distributions are calibrated, were ensured to be at points above which the mean squared error (MSE) of the linear regression is minimal, i.e. within nearly horizontal sections in the plot of the slope versus the number of observations above threshold. This selection measure was undertaken to avoid high statistical uncertainty due to the high fluctuations which occur in the slope of the Q-Q plots for high thresholds due to randomness of the dataset.

In **Fig. 4**, the selected optimal thresholds corresponded to the flow values with threshold ranks $t = 36$ (i.e. 36th highest flow value) and 30 (i.e. 30th lowest flow values) for high and low (Q) streamflows respectively. A linear tail behavior in exponential Q-Q plot is obtained for $(1/Q)$ towards the higher $1/Q$ values (or lower Q values of low flows) whereas for high flows it's towards higher Q values (see **Fig. 4**).

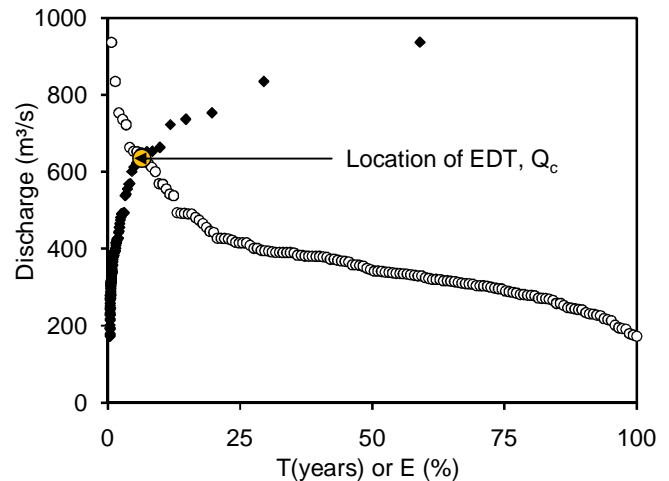


Fig. 5 FDC and return period curve for daily flows; the dark diamonds (\blacklozenge) are for Empirical Return period, T_e while Exceedance Frequency, E (%) is denoted by (\circ).

What followed next after the selection of x_t is fitting of the calibrated EVD in which **Eq. (31)** was used. The fitted calibration results for the FDCs of both high and low flows are shown in **Fig. 6**.

Empirically Estimated Design Threshold (EEDT)

With nearly iid extreme streamflow events ranked from the highest to the lowest, return periods were computed using **Eq. (25)**. **Equation (35)** was used to determine the location of the EEDT. For each aggregation level of sample size, s , the EEDT was computed using **Eq. (42)** below.

$$Q_c = \frac{1}{r} (Q_b (\sqrt{s} - i_a) + Q_a (i_b - \sqrt{s})) \quad (42)$$

where Q_c = the EEDT, $r = (i_b - i_a)$, s = the sample size, i_a = the rank of the discharge preceding EDT, i_b = the rank of the discharge after EDT, Q_a = the discharge at rank immediately after \sqrt{s} , and Q_b = the discharge at rank just preceding \sqrt{s} .

Fig. 5 shows, for the same sample, the plot return period curve alongside FDC with exceedance frequency (unrescaled). The abscissa was not log-transformed to allow clarity of the typical shapes of the return period curves and FDC to be depicted. Similar shapes are obtainable for $(1/Q)$ low flows.

Theoretically Estimated Design Threshold (TEDT)

Whereas the EEDT can be computed as an interpolant using **Eq. (42)**, depending on the distribution class of the data identified from the Q-Q plot, the TEDT can also be derived. Taking x as the TEDT, x_t as the threshold, α as the scale and t as the rank of x_t , using

Eqs (26) and (29) for high flows, at the EDT, T [years] equals E [%] and for the distribution of data with the tail that can be described by an exponential EVD (which is a special case of the GPD with zero shape parameter):

$$\left(\frac{n}{t}\right) \frac{1}{\exp\left(-\frac{(x-x_t)}{\alpha}\right)} = \left(\frac{100t}{s}\right) \frac{1}{\exp\left(\frac{x-x_t}{\alpha}\right)} \quad (43)$$

Simplifying and taking the natural logarithms, x can be made the subject as:

$$x = x_t - \frac{\alpha}{2} \ln\left(\frac{ns}{100t^2}\right) \quad (44)$$

The relationships between the EDTs and the corresponding aggregation levels were also examined. In the computation of the EDT, the discharges are sorted in descending order (from highest to lowest) for high flows; however, for low flows, the discharges are ranked in ascending order (from the lowest to the highest). This means that the largest discharge has the highest return period (for high flows) or the lowest return period (for low flows).

QDF relationships

With extracted iid streamflows, EVD is calibrated for each aggregation level under consideration. The relationships between the EVD parameters, θ and the aggregation levels, D were calibrated using the formula presented in Willems (note, 2003) as expressed below:

$$\theta(D) = cD^{-H/a} \left[1 + w \left(\frac{A}{D^{1/z}} \right)^a \right]^{-\beta} + \bar{q} \quad (45)$$

In this formula, A is the area of the catchment upstream of the discharge measuring station considered. The formula is based on scaling properties for the rainfall intensities and consequently of the river discharges. The scaling property indicates that the same EVD is valid for different aggregation levels after application of a scaling factor to the rainfall or discharges values. The scaling factor is different for different aggregation levels. The formula has five parameters: c , w , H , z and a . The last three parameters are called ‘scaling exponents’ in the scaling theory and a specific interpretation can be given to these parameters. The parameter H is called ‘Hurst-exponent’, while z represents the dynamic scaling exponent, and a , the scaling exponent applied for the aggregation level.

The QDF formula meets the following asymptotic properties for the parameters of EVDs.

- The parameters, θ of EVD (α , x_t and t for the exponential distribution) have a constant asymptotic value towards the lower aggregation levels ($D \rightarrow 0$).
- The scale parameter and the threshold rank (parameters α and t for the exponential distribution) have a zero asymptotic value towards the higher aggregation levels ($D \rightarrow \infty$).
- The threshold discharge (parameter x_t of the exponential distribution) approach asymptotically the mean long-term discharge value \bar{q} towards the higher aggregation levels ($D \rightarrow \infty$).

For the scale parameter and the rank threshold, \bar{q} equals 0. The threshold discharge \bar{q} can be considered equal to the mean discharge value to be calculated based on the complete time series.

In the computation of the E-percentage event (x_E), each parameter of the EVD was expressed in terms of aggregation level, D using **Eq. (45)** to obtain $\hat{\alpha}(D)$, $\hat{x}_t(D)$ and $\hat{t}(D)$; of course, for exponential case, $k = 0$. Values of E (%) are then selected. After calibrating θ of EVDs with aggregation levels (D), then the calibrated QDF relations were computed for a selected E (%) using **Eq. (46)**.

$$x_E = x_t(D) + \hat{\alpha}(D) \left[\ln\left(\frac{100 \times t(D)}{s}\right) - \ln(E) \right] \quad (46)$$

With the help of **Eq. (46)** extreme high and low streamflow quantiles then easily can be constructed as a simultaneous function of different values of E (%) and aggregation levels.

To derive smooth mathematical relationships, **Eq. (45)** for each parameter of the EVD was optimized using MSE minimization technique. Following carefully selection, in a consistent way, of the optimal threshold ranks for the different aggregation levels, smooth paths can be derived between parameters of the EVD and aggregation levels. It should be noted that, the smooth mathematical relationship between parameters θ of EVD and aggregation levels guarantees close matches between the empirical and theoretical T-year events in the QDFs.

Evaluation of calibrated FDCs

To check for bias by assessing the closeness between the empirical and theoretical FDCs, the distribution of the mean of residuals and the confidence intervals were determined. The differences between the empirical

extreme flows and the calibrated distributions were taken as the residuals. On assumptions that, a small sample size, s of observed discharges was obtained from a normally distributed population with mean, μ_q and variance, σ_q and also that, the errors on the return period models were random and followed normal distribution, a t -non cumulative probability distribution, $f_Q(q)$ and a t -mean statistic, \hat{t} were calculated for the hydrological extremes using Eqs (47) and (48)

$$f_Q(q) = \left\{ \Gamma\left(\frac{s}{2}\right) \sqrt{s\pi} \right\}^{-1} \Gamma\left(\frac{s+1}{2}\right) \left[1 + \frac{q^2}{s} \right]^{-\frac{s+1}{2}} \quad (47)$$

$$\hat{t} = \frac{(\bar{Q} - \mu_r)}{(S_r/\sqrt{s})} \quad (48)$$

where μ_r = mean of residuals, \bar{Q} = sample mean, and S_r = standard deviation of residuals. The null hypothesis for the double-sided sample mean t -test was that, 'H₀ = the mean of the residuals is equal to zero' and the computed 95% confidence intervals are shown in Fig. 12.

RESULTS AND DISCUSSION

Calibrated FDCs

Fig. 6 shows the calibrated results for the hydrological extremes. It can be seen that, the match between the empirical FDC and the fitted FDC for high flow (Fig. 6a) is better than that of low flow (Fig. 6b). This was also reflected in the magnitudes of the confidence intervals for the probability distributions of the residuals (see Fig. 12); this is due to the phenomenon termed as persistence in flows which complicates extraction of nearly iid low streamflow.

The exceedance frequency for $(1/Q)$ low flows is towards the higher $(1/Q)$ values whereas for high flows it is towards the higher Q values. The transformation $(1/Q)$ makes the smallest flow to become the maximum flow and hence the allotment of rank 1 when sorted in descending order. It means that, the $(1/Q)$ approach makes the smallest actual flow to achieve the smallest E (%). Ideally, when the approach is reversed i.e. back transformed, what we are calling the smallest actual flow will achieve the highest E (%). Therefore, with a back transformed or normal Q flow, whereas for high flows, E (%) would be applicable, for low flows we would talk of non-exceedance frequency, NE (%).

The validity of the calibrated EVD is at the tail of the POT events i.e. above x_t . Below x_t , uncertainty in the

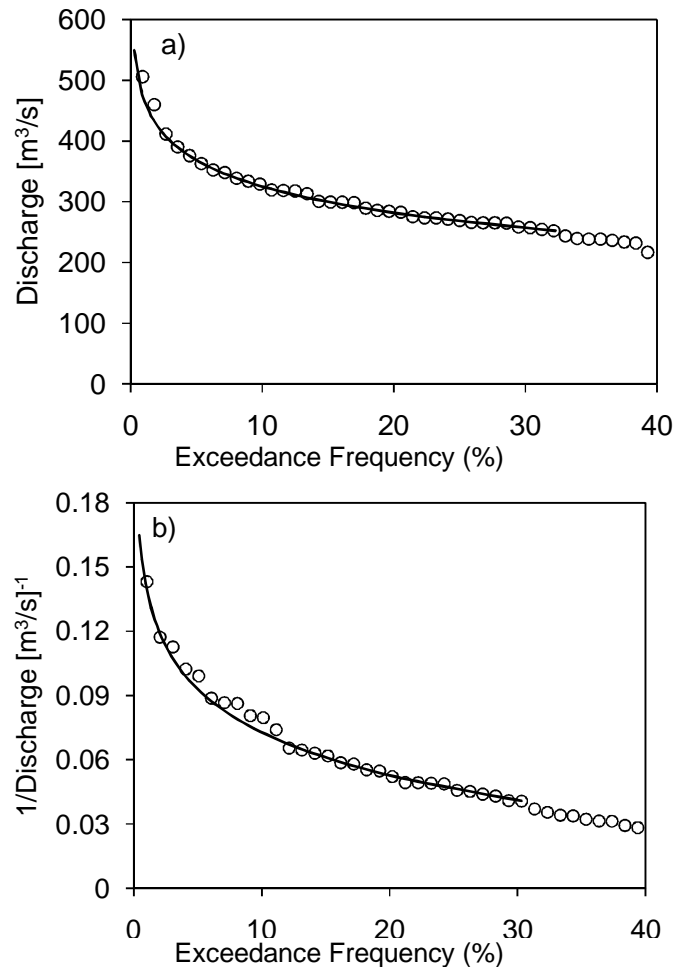


Fig. 6 Observed discharges (o) at aggregation level of 2 months with exponential distribution; the solid line is the calibrated EVD fitted to the tail of the data; graph a) is for high flows and b) is for low flows.

fitted distribution would arise due to the randomness of the lower ranking streamflows from the dataset. The hydrological extremes of Nzoia river at station 1EF01 were found to follow exponential distribution. Eventually, the empirical FDCs were calibrated with exponential distribution as the suitable EVD.

The shape of a typical FDC is such that its upper and lower regions are for analysis of high flows and low flows respectively (i.e. hydrological extremes) and thus significant in evaluating the stream and basin characteristics. What needs to be noted here is that, since we are dealing with the hydrological extremes, it means that Fig. 6a and b are respectively for the upper and lower regions of a typical FDC for a full dataset. The shape of the curve in the high-flow region indicates the type of flood regime the basin is likely to have, whereas, the shape of the low-flow region characterizes the ability of the basin to sustain low flows during dry seasons (Brown *et al.*, 2010). Fig. 6a and 6b are for POT high flows and $(1/Q)$ low flows respectively. It can

be seen that the FDC for $(1/Q)$ low flows is steeper than that for high flows. The steepness of the calibrated FDC for low flows (**Fig. 6b**) shows intermittence of the severe extreme droughts and for high flows (high flows for short periods) (**Fig. 6a**) indicates that the extreme flood events are rain-caused while. **Fig. 6a–b** are typical of a tropical watershed. For temperate watersheds which are sometimes characterized by snowmelts floods which last for several days, flatter upper limit of their FDCs would be expected. Generally, for low flows, flatness of an FDC is an indication of the phenomenon referred to as persistence which might be brought about by the artificial streamflow regulation or the presence of a large groundwater capacity or lake which sustains the base flow to the stream.

It can be seen from **Fig. 7** that, the FDC becomes flatter and occupies lower position as the aggregation level increases. This is due to the smoothing effect of the temporal aggregation which actually mimics the delayed response of a watershed. The difference between a given FDC and the next FDC presents the inverse of frequency of occurrence of streamflow events for the temporal scales under consideration.

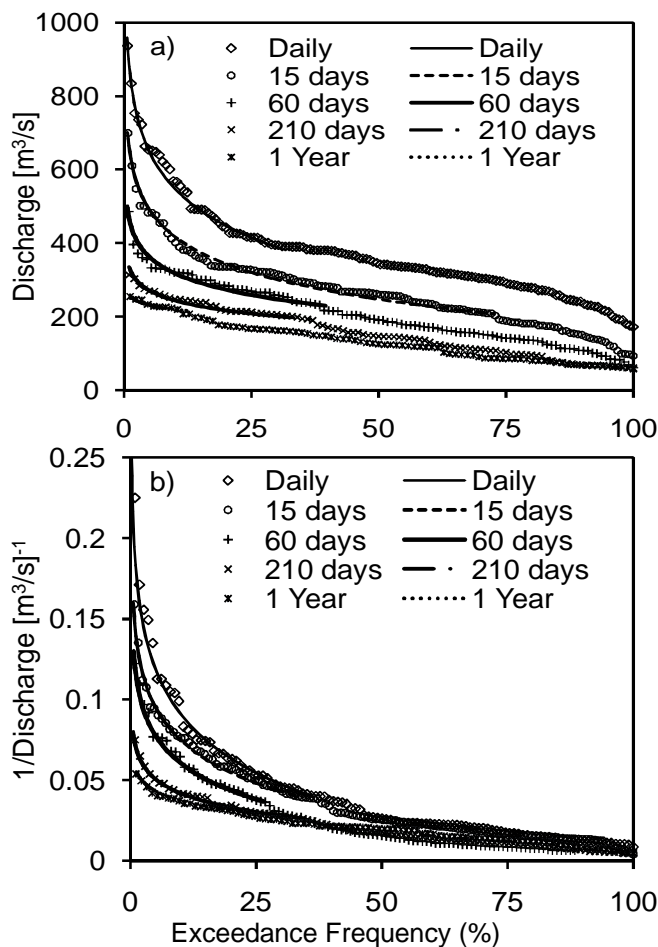


Fig. 7 Calibrated FDCs for various aggregation levels; graph a) is for high flows and b) is for low flows.

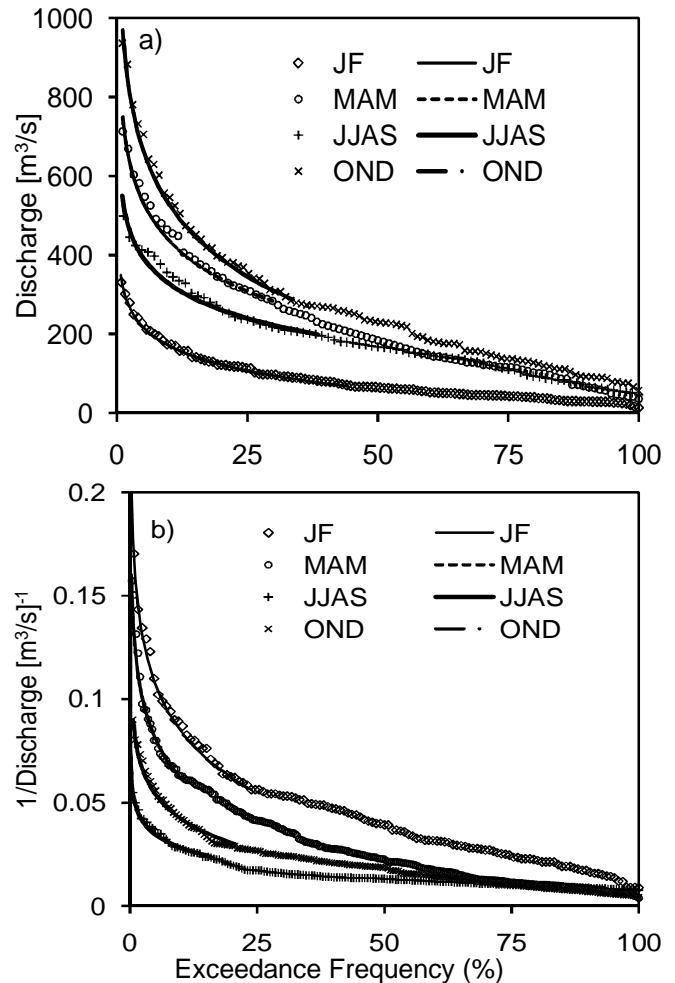


Fig. 8 Calibrated FDCs for different seasons; graph a) is for high flows and b) is for low flows; the letters in the legend represent the initials for the various months of the year.

Therefore, when several aggregation levels are used, the obtained relationships can be considered as a form of periodic (inverse frequency) domain analysis. Eventually just like we would obtain a QDF relationship when a range of relevant frequencies are used, periodic flow duration curves (PFDCs) relationships were established with various aggregation levels (**Fig. 7**). A PFDC curve describes the relationship between amplitudes and periods (duration) of occurrence of streamflows for various E (%). The difference between a given FDC and the next FDC can be seen to reduce with increase in aggregation levels, especially for high flows (**Fig. 7a**); however, for low flows, lack of clarity for the same is due to the phenomenon termed as persistence.

Instead of using aggregation levels as in **Fig. 7**, in **Fig. 8**, different seasons were used to obtain seasonal flow duration curves (SFDC). With the climate of Lake Victoria basin which may generally be described to vary from modified equatorial to semi-arid type, it is known to typically experience two wet seasons i.e. October, November and December, (OND i.e. short wet season)

and March, April and May (MAM, i.e. long wet season). As the wet seasons are responsible for periods of high flows; dry spells from June to September (JJAS) and January to February (JF) can be assumed to be responsible for the period of low flows in study area. The SFDCs for both high flows and low flows can be seen presented in **Fig. 8**.

Estimated Design Threshold (EDT)

The EDT also denoted by Q_c can be seen to reduce in high flows as the aggregation level increases (**Fig. 9a**); however, it increases as the aggregation level increases for low flows (**Fig. 10a**). For high flows, Q_c can be seen to reduce as T_c increases (**Fig. 9b**) while for low flows, Q_c increases as the T_c increases (**Fig. 10b**). Whereas **Fig. 9** presents the delays of the catchment response to the hydrological extremes with increase in temporal resolutions, **Fig. 10** emphasizes the persistence in which the streamflow is the direct contribution of groundwater. The slopes of the graphs in **Fig. 9a** and **Fig. 10a** can be taken to be an indicator of the dryness or duration of dry spells of the catchment under study. A steeper slope of the graph indicates higher intermittency in the daily flows i.e. the existence of longer dry spells. For this study, the slopes of the graphs can be seen to be modest;

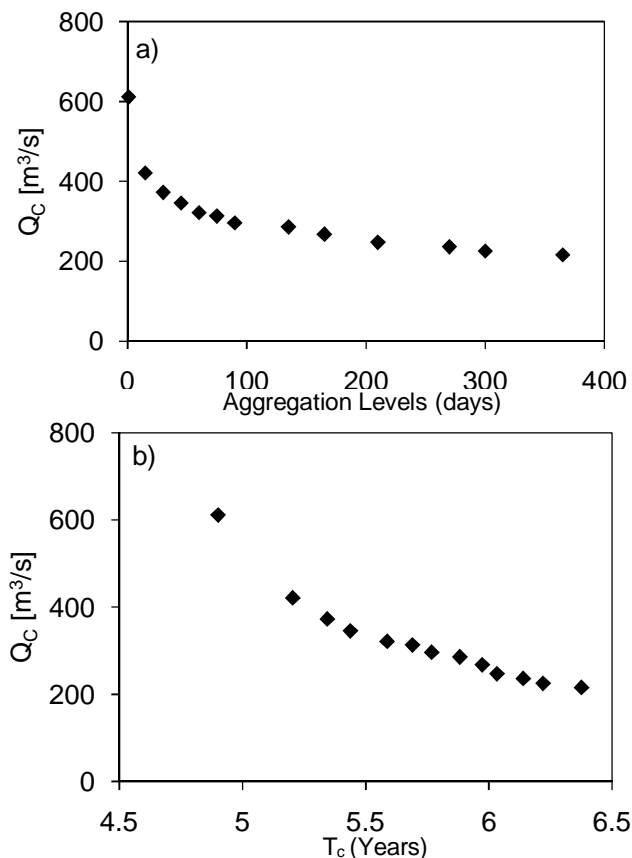


Fig. 9 Variation of the EDT i.e. discharge (Q_c) with critical return periods (T_c) and aggregation levels for high flows.

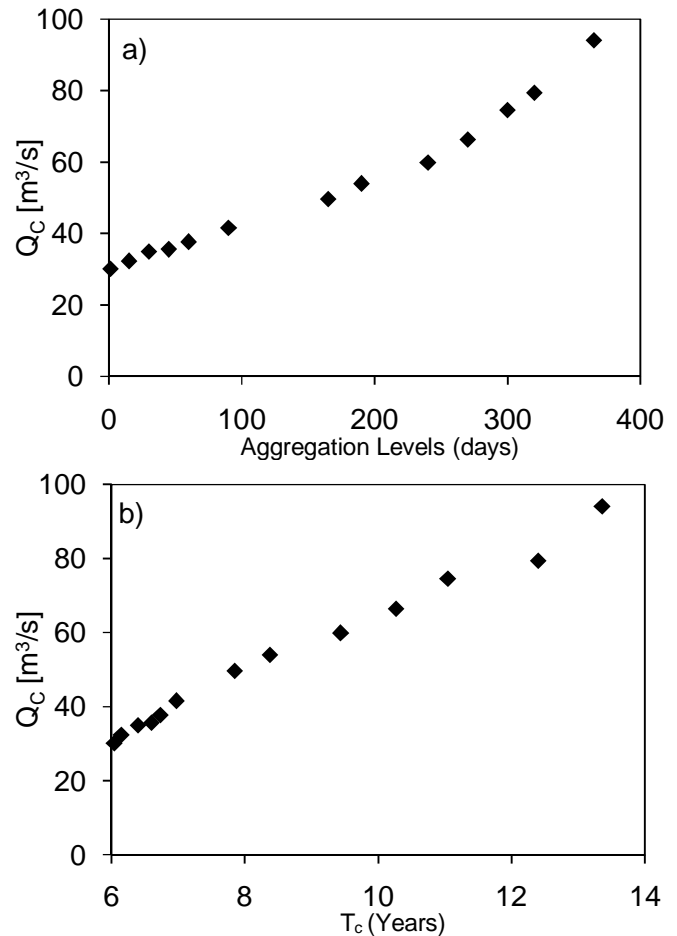


Fig. 10 Variation of critical discharge (Q_c) with critical return periods (T_c) and aggregation levels for low flows.

this is due to the moderate wet-dry variations of the area under consideration (Onyutha, 2011). It should be noted that for high flows, exceedance frequency (%) was used while for low flows, non-exceedance frequency (%) was used. Modelling of FDC for various aggregation levels and seasons can help to improve on understanding of climate variability the forecasting of hydrologic variables.

This concept of EDT can be important in number of ways:

- Climate change study:** it can be used to study the impact of climate change on hydrology of a watershed. This can be carried out by comparing the values of the EDT from the control and scenario streamflows. The variations of EDT with aggregation levels and/or T_c (**Fig. 9** and **Fig. 10**) can be constructed for the control and scenario extreme flows to deduce the impact of climate change on the hydrological extremes. In another method, for example, EDT can be obtained for the different phases of the El Niño-Sothern Oscillation (ENSO) phenomenon to reveal, in the climate conditions, the extent to which a system is impacted

upon by the ENSO-induced changes. A shift of the EDT to a higher position may be brought about by the increase in flows which might result from the wetter conditions caused by the cool phase of the ENSO over the equatorial Pacific (Brown *et al.*, 2010).

- b) **Design of hydraulic structures:** instead of the return periods of hydrological (water engineering) structures being fixed by regulations and/or following from a risk or cost-benefit analysis, T_c can be taken as a guide in the design of the structures under consideration. The design flow may be taken not less than the EDT i.e. the event corresponding to the T_c .
- c) **Risk assessment of hydraulic structures:** given the design life expectations, d , risk analysis (such as natural, inherent, or hydrologic risk of failure) with respect to T_c can be carried out. Ideally, a designer of a hydraulic structure would opt to incorporate if possible, the likelihood that no any events will exceed the design limits within the expected life of the structure; however, in engineering practice, we normally compliment the said idea of likelihood by considering the probability of at least an event (x) exceeding the design limits once during the expected life of the structure i.e. $P[x \geq EDT]$. We can assess the risk associated with the EDT (i.e. the proposed lower limit over which the risk analysis can be carried out) using the **Eq. (49)** expressed as:

$$R_c = 1 - \left(1 - \frac{1}{T_c}\right)^d = 1 - (1 - P\{x \geq EDT\})^d \quad (49)$$

where R_c : risk associated with the EDT, and d is the expected life.

QDF relationships

From **Fig. 11**, the irregularity resulting from the fact that, the calibrated relationships have smoother paths than the empirical points reflects the uncertainty in the EVA for the individual aggregation levels, especially for low flows (**Fig. 11b**). The mismatches between the empirical and the calibrated relationships for low flows, especially for lower aggregation levels are due to persistence phenomenon in streamflows. However, the uncertainty in the calibration of the EVDs for individual distributions can be reduced by fitting relationships between the parameters of the EVD and the aggregation levels. Since we are dealing with iid extreme high and low flows in this study, E (%) considered for the QDF relationships were 5, 10, 25, 35 and 45%. The main reason for the choice of the listed E (%) was that, if the

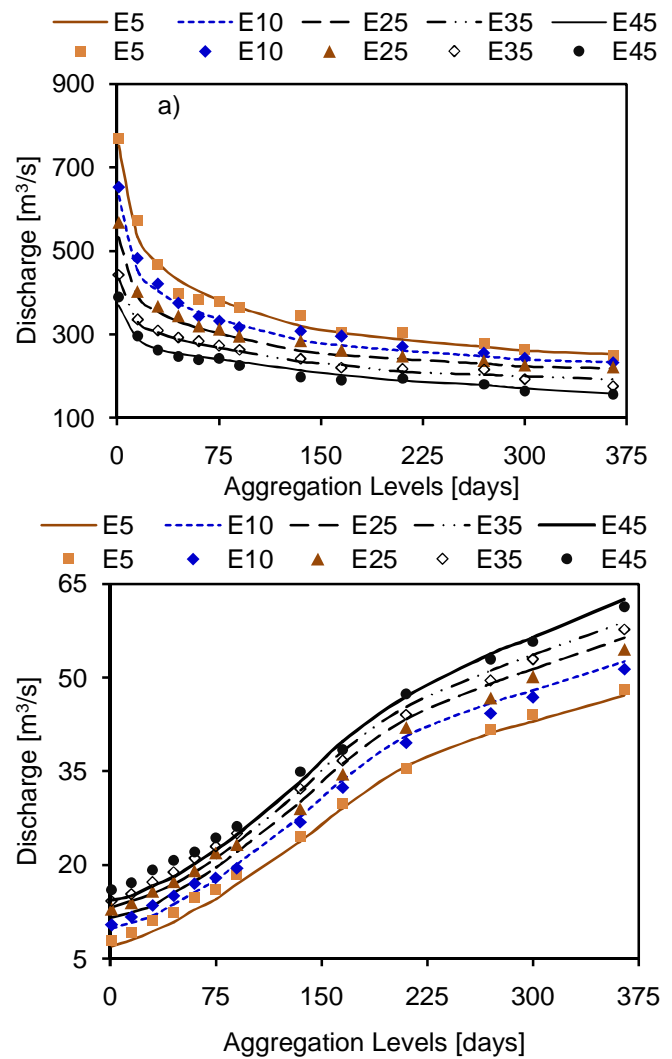


Fig. 11 Results for QDF relationships; graph a) is for high flows while b) is for low flows; the letter E denotes the exceedance frequency while the numbers after E are the selected percentages for E.

full hydrological time series were considered to construct an FDC, what we are calling high flows, depending on the selected threshold, would correspond to lower E (%) of up to about 45%; similarly, low flows would fall within higher frequencies from, say, 75% up to 100%. Whereas for a QDF in which frequency is defined in terms of T (years), the highest return period up to which empirical QDF-curves can be constructed is equal to the total length of the discharge time series in years, when we use E (%) instead of T (years), the upper limit for the curve for E (%) that would make sense is 100%.

For high flows it can be seen that, the higher the exceedance frequency, the lower the position occupied by the corresponding curve on the QDF relationships (**Fig. 11a**); however, for low flows, the reverse is true i.e. the higher the exceedance frequency, the higher the position occupied by the corresponding curve in the

QDF relationships (**Fig. 11b**). Furthermore, the slopes of the QDF curves for low and high flows are positive and negative respectively.

The aggregation levels used to establish the QDF relationships included daily, 15 days, 30 days, 45 days, 60 days, 75 days, 90 days, 135 days, 165 days, 210 days, 270 days, 300 days and 1 year. Many aggregation levels were chosen to enhance the accuracy of the QDF relationships. The selected range of aggregation levels in this study covered multipurpose applications e.g. agricultural, irrigation, power plants, domestic supply, pollution etc.

Evaluation of calibrated FDCs

The mean of the residuals from the calibrated EVD and the empirical exceedance frequency (%) for high flows and low flows are -0.28 and -0.11%. The confidence intervals at 5% level of significance for the probability distributions of residuals are -0.59 and 0.03% (for high flows) and -0.55 and 0.33% (for low flows). The fact that the mean values for the residuals are within the confidence limits confirms that the fitted (calibrated) EVD is unbiased. However, it can be noted that the confidence intervals for the probability distributions of residuals of low flows are wider than that of high flows (**Fig. 12**). This might have arisen from the serial correlation in the time series which occurs as a result of a memory in the hydrological system caused by large storages, such as extensive groundwater reservoirs or nearby Lake. The study area is very close to Lake Victoria which is the world's second largest freshwater lake. The high recession constants of base flows as compared to that of quick flows normally contribute to low variations in low flows (phenomenon referred to as persistence) which complicate the extraction of peak flow events and hence the uncertainty in the EVA.

CONCLUSION

The discharges of the study area were found to exhibit exponential distribution; and thus, empirical FDCs were calibrated with exponential distribution as the suitable EVD. Fitting of calibrated EVD to the FDCs was done for selected range of aggregation levels and also the different seasons. In the FDCs, the residuals from the calibrated EVD and the empirical exceedance frequency (%) were normally distributed for both high and low flows. The confidence intervals at 5% level of significance for the probability distributions of residuals were -0.59 and 0.03% (for high flows) and -0.55 and 0.33% (for low flows). The fitted (calibrated) EVDs for the FDC were unbiased for both high and low flows

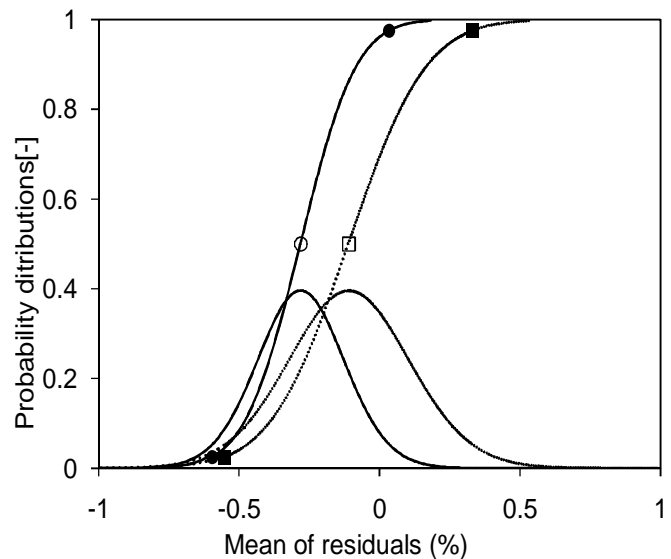


Fig. 12 Probability distributions for the mean of residuals obtained from the calibrated EVD and empirical exceedance frequency. The round dotted curves (on the right) are for low flows while thin solid curves (on the left) are for high flows. The bell-shaped curves are probability density distributions while the sigmoid-shaped curves are cumulative probability distributions. The limits for the 95 % confidence intervals for high flows and low flows are denoted by symbols (●) and (■) respectively, while (○) and (□) are for the mean of the distributions for the high flows and low flows respectively.

since the mean values of residuals were within the confidence limits. It can be concluded that, the FDC becomes flatter and occupies lower position as the aggregation level increases. Statistical modelling of FDC for various aggregation levels and seasons, like in this study can be used to enhance understanding of climate variability and forecasting of hydrologic variables for management of water resources.

QDF relationships were established for the hydrological extremes using exceedance frequencies instead of the conventional return periods; such QDF relationships to reveal more compressed statistical information that can be availed for water resources management. In the QDF relationships for high flows, curves of higher the exceedance frequency occupy lower positions; for low flows, however, curves of higher exceedance frequency, occupy higher positions. The QDF relationship of high flows has negative slope while that of low flows has positive slope. The constructed QDFs can be used for management in a number of applications including irrigation, hydropower supply, water supply etc, to assess the water management requirements in terms of cumulative volumes of water available during high and low flows. This paper has also presented the concept of EDT. The EDT is proposed in this paper as the lower limit for the range over which risk analysis for the design of

hydrologic structures can be carried out with respect to safety. For high flows, the EDT is typified by high return period and low exceedance frequency. However, for low flows, similar characterizations of the EDT are obtainable with transformed ($1/Q$) low flows. The EDT reduces in high flows as the aggregation level increases; however, it increases as the aggregation level increases for low flows. For high flows, the EDT reduces with increase in the return period at the EDT while for low flows, it increases as the return period at the EDT increases. The concept of EDT can be used in a number of ways including; climate change assessment, design of hydraulic structures, hydrological risk assessments etc.

Acknowledgment The Hydrological time series of daily resolution used for this research was obtained from the FRIEND/NILE project of UNESCO and the Flanders in Trust Fund.

REFERENCES

- Bernard, M. M. (1932) Formulas for rainfall intensities of long durations. *Trans. ASCE* **96**, 592–624.
- Best, A. E., Zhang, L., McMahon, T. A. & Western, A. W. (2003) Development of a model for predicting the changes in flow duration curves due to altered land use conditions. In Post, D. A. MODSIM 2003 International Congress on Modeling and Simulation; Townsville, Australia. Canberra, MSSANZ; 861–866.
- Borga, M., Vezzani, C. & Fontana, G. D. (2005) Regional Rainfall Depth–Duration–Frequency Equations for an Alpine Region, *Nat. Hazards* **36**, 221–235.
- Brown, C., Baroang, K. M., Conrad, E., Lyon, B., Watkins, D., Fiondella, F., Kaheil, Y., Robertson, A., Rodriguez, J., Sheremata, M. & Ward, M. N. (2010) Managing Climate Risk in Water Supply Systems. *IRI Technical Report 10-15*, International Research Institute for Climate and Society, Palisades, NY, 133 pp. Available in: <http://iri.columbia.edu/publications/id=1048>
- Castellarin, A., Galeati, G., Brandimarte, L., Montanari, A. & Brath, A. (2004) Regional flow-duration curves: reliability for ungauged basins. *Adv. Water Resour.* **27**, 953–96.
- Castellarin, A., Vogel, R.M. & Brath, A. (2004) A stochastic index flow model of flow duration curves. *Water Resour. Res.* **40**, W03104, doi:10.1029/2003WR002524.
- Chow, V.T., Maidment, D.R. & Mays, L.W. (1988) *Applied Hydrology*. McGraw-Hill Book Company, New York, U.S.A, pp 380–410.
- Cigizoglu, H. K. & Bayazit, M. (2003) A generalized seasonal model for flow duration curve. *Hydrol. Processes*. **14**(6), 1053–1067.
- Cole, R.A.J., Johnson, H.T. & Robinson, D.J. (2003) The use of flow duration curves as a data quality tool. *Hydrol. Sci. J.* **48**(6), 939–951.
- Crocker, K.M., Young, M.D.Z. & Rees, H.G. (2003) Flow duration curve estimation in ephemeral catchments in Portugal. *Hydrol. Sci. J.* **48**(3), 427–39.
- Fennessey, N.M. & Vogel, R.M. (1990) Regional flow-duration curves for ungauged sites in Massachusetts. *J. Water Resour. Plann. Managem. ASCE* **116**(4), 531–49.
- Foster, H. A. (1934). Duration curves. *ASCE Trans.* **99**, 1213–1267.
- Hill, B.M., 1975. A simple and general approach to inference about the tail of a distribution. *Ann. Statist.* **3**, 1163–1174.
- Hironobu, S., Vudhivanich, V., Whitaker, A.C. & Lorsirirat, K. (2003) Stochastic Flow Duration Curves for Evaluation of Flow Regimes in Rivers. *J. Am. Water Resour. Assoc.* **39**(1), 47–58.
- Institute of Hydrology (1980) *Low flows studies report*. Institute of Hydrology, Wallingford, United Kingdom.
- Javelle, P., Grésillon, J.M. & Galéa, G. (1999) Discharge-duration frequency curves modeling for floods and scale invariance. *Sci. Terre Planets* **329**, 39–44.
- Javelle, P., Ouarda, T.B.M.J., Lang, M., Bobée, B., Galéa, G. & Gré'sillon, J-M. (2002) Development of regional flood-duration frequency curves based on the index-flood method. *J. Hydrol.* **258**, 249–259.
- Juraj, M.C. & Taha, B.M.J.O. (2007) Regional flood–rainfall duration-frequency modeling at small ungaged sites, *J. Hydrol.* **345**, 61–69.
- Lang, M., Ouarda, T.B.M.J. & Bobée, B.B. (1999) Towards operational guidelines for over-threshold modeling, *J. Hydrol.* **225**, 103–117.
- LeBoutillier, D.V. & Waylen, P.R. (1993) A stochastic model of flow duration curves. *Water Resour. Res.* **29**(10), 3535–3541.
- LeBoutillier, D.V. & Waylen, P.R. (1993) Regional variations in flow-duration curves for rivers in British Columbia, Canada. *Phys. Geogr.* **14**(4), 359–78.
- Maurino, M.F. (2004) Generalized rainfall-duration-frequency relationships: applicability in different climatic regions of Argentina. *J. Hydrol. Engng.* **9**(4), 269–274.
- Mandal, U. & Cunnane, C. (2009) *Low-Flow Prediction For Ungauged River Catchments In Ireland*. Irish National Hydrology Seminar.
- Melesse, A., Abtew, W., Dessalegne, T. & Wang, X. (2010) Low and high flow analyses and wavelet application for characterization of the Blue Nile River system. *Hydrol. Process.* **24**(3), 241–252 (John Wiley & Sons).
- Nathan, R.J. & McMahon, T.A. (1990b) Practical aspects of low-flow frequency analysis. *Water Resour. Res.* **26**, 2135–2141.
- Nathan, R.J. & McMahon, T.A. (1992). Estimating Low Flow Characteristics in Ungauged Catchments. *Water Resour. Managem.* **6**(1), 85–100.
- Onyutha, C. (2011). *Regional Analysis of Hydrological Extremes in Lake Victoria Nile Sub Basin*. MSc Thesis, Katholieke Universiteit Leuven & Vrije Universiteit Brussel, Belgium, Europe.
- Pickands, J. (1975) Statistical inference using extreme order statistics. *Ann. Statist.* **3**, 119–131
- Smakhtin, V.U (2001). Low flow hydrology: a review. *J. Hydrol.* **240**, 147–186.
- Studley, S.E. (2001) *Estimated flow-duration curves for selected ungaged sites in Kansas*. US Geological Survey. Water-resources investigations report, 01–4142.
- Taye, M.T. & Willems, P. (2011) Influence of climate variability on representative QDF predictions of the upper Blue Nile Basin. *J. Hydrol.* **411**, 355–365.
- USWRC, (1976) Guidelines for determining flood flow frequency. *US Water Resour. Council, Bull. 17*. Hydrology Commission, Washington, DC, 73 p.
- Vogel, R.M. & Fennessey, N.M. (1995) Flow duration curves. II. A review of applications in water resource planning. *Water Resour. Bull.* **31**(6), 1029–1039.
- Vogel, R.M. & Fennessey, N.M. (1994) Flow-duration curves. I: New interpretation and confidence intervals. *J. Water Resour. Planning and Management, ASCE*. **120**(4), 485–504.
- World Meteorological Organization (WMO) (2008), Manual on Low-flow Estimation and Prediction, *Operational Hydrology Report No.50, WMO-No.1029*. CH 121- Geneva 2, Switzerland, 50–57.

- Willems, P. (2000) Compound intensity/duration/frequency-relationships of extreme precipitation for two seasons and two storm types. *J. Hydrol.* **233**, 189–205.
- Willems, P., Guillou, A. & Beirlant, J. (2007) Bias correction in hydrologic GPD based extreme value analysis by means of a slowly varying function. *J. Hydrol.* **338** (4), 221–236.
- Willems, P. (1998a). Hydrological applications of extreme value analysis, combining recently developed and traditional methods. *Report, Hydraulics Laboratory, Katholieke Universiteit Leuven*, Leuven, Belgium.
- Willems, P. (1998b). Hydrological applications of extreme value analysis. In: Wheeler, H., Kirby, C. (Eds.), *Hydrology in a Changing Environment*, vol. III. Wiley, New York, 15–25.
- Willems, P. (2003) Formula for the Calibration of QDF-Curves on the Basis of Scaling Properties and Correct Asymptotic Properties. *Internal Note, Hydraulics Division, Katholieke Universiteit Leuven*. Leuven, Belgium.
- Willems, P. (2009), A time series tool to support the Multi-criteria performance evaluation of rainfall-runoff models. *Environ. Modell. softw.* **24**, 311–321.
- Zaidman, M.D., Keller, V., Young, A.R. & Cadman, D. (2003) Flow-duration-frequency behavior of British rivers based on annual minima data. *J. Hydrol.* **277**, 195–213.