

A stochastic index flow model of flow duration curves

Attilio Castellarin

Faculty of Engineering, Università di Bologna, Bologna, Italy

Richard M. Vogel

WaterSHED Center, Department of Civil and Environmental Engineering, Tufts University, Medford, Massachusetts, USA

Armando Brath

Faculty of Engineering, Università di Bologna, Bologna, Italy

Received 26 July 2003; revised 8 January 2004; accepted 16 January 2004; published 10 March 2004.

[1] Annual flow duration curves (AFDCs) are used increasingly because unlike traditional period of record flow duration curves (FDCs), they provide confidence intervals for the median AFDC, they enable one to assign return periods to individual AFDCs, and they offer opportunities for developing a generalized stochastic model of daily streamflow. Previous stochastic models of FDCs and AFDCs were unable to reproduce the variance of AFDCs. We introduce an index flow approach to modeling the relationship between an FDC and AFDCs of daily streamflow series, which is able to 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. Our approach offers additional opportunities for the regionalization of flow duration curves and for the generation of time series of daily streamflows at ungauged sites. Our approach is tested on three river basins in eastern central

Italy. **INDEX TERMS:** 1860 Hydrology: Runoff and streamflow; 1869 Hydrology: Stochastic processes; 1812 Hydrology: Drought; 1821 Hydrology: Floods; 1884 Hydrology: Water supply; **KEYWORDS:** flow duration curve, stochastic modeling of daily flows, index flow, order statistics, generalized Pareto distribution, logistic distribution

Citation: Castellarin, A., R. M. Vogel, and A. Brath (2004), A stochastic index flow model of flow duration curves, *Water Resour. Res.*, 40, W03104, doi:10.1029/2003WR002524.

1. Introduction

[2] A correct representation of the streamflow frequency regime for a river basin is an essential component of many hydrologic applications including: reservoir and lake sedimentation studies, in-stream flow assessments, hydro-power feasibility analysis, water quality management, waste load allocation, water resource allocation, flood frequency analysis and flood damage assessment. The streamflow frequency regime is nicely summarized using a flow duration curve (FDC), which is simply the complement of the cumulative distribution function (cdf) of streamflow. An FDC provides the percentage of time (duration) a daily, or weekly, or monthly, or yearly (or some other time interval of) streamflow is exceeded over a historical period. The use of FDCs is widespread in hydrology and water resource engineering with the earliest use of an FDC attributed to Clemens Herschel, in about 1880 [Foster, 1934]. Vogel and Fennessey [1994, 1995] provide a brief history of the use of FDCs in hydrology.

[3] Until the introduction of annual-based FDCs by LeBoutillier and Waylen [1993] and Vogel and Fennessey [1994] the FDC has traditionally been computed using the complete period of record of streamflow, leading to steady

state or long-term probabilistic statements concerning streamflow exceedances. Searcy [1959] has summarized the properties of period of record FDCs and this topic remains fertile as evidenced from a recent review article [Smakhtin, 2001] as well as recent papers on the regionalization of FDCs [Crocker *et al.*, 2003], uncertainty analysis of FDCs [Yu *et al.*, 2002], the development of a stochastic model for FDCs [Cigizoglu and Bayazit, 2000; Sugiyama *et al.*, 2003], and use of FDCs for watershed management [Good and Jacobs, 2001].

[4] In practice, the period of record FDC is limited because with it, one can only make steady state probabilistic statements about streamflow exceedances. As in flood and low-flow frequency analysis, one often wishes to make probabilistic statements about a given calendar or water year. In an effort to make such probabilistic statements, LeBoutillier and Waylen [1993] and Vogel and Fennessey [1994] introduced the idea of an annual, water year or calendar year FDC, which we refer to here as an AFDC (annual flow duration curve) as opposed to the period of record FDC which we refer to as simply an FDC. AFDCs have been shown to be quite useful for making probabilistic statements about wet, typical and dry years, for computing confidence intervals associated with the AFDC representing the typical hydrologic condition and for assigning return periods to individual AFDCs [Vogel and Fennessey, 1994]. Since their introduction, a number of investigators have

found AFDCs to be quite useful for solving a wide range of problems [Claps and Fiorentino, 1997; Smakhtin and Toulouse, 1998; Good and Jacobs, 2001; Sugiyama et al., 2003]. The formal definition of FDC and AFDC is given in section 2.

[5] To enable a complete understanding of the streamflow frequency regime, a stochastic model is needed which relates the FDC to the AFDCs. Furthermore, the need for a stochastic model for daily streamflows is becoming increasingly important. As computational resources become more efficient and effective, there is a natural tendency for hydrologic models and water resource engineering applications to make use of streamflow data on finer and finer timescales. Daily or shorter time-scales are now commonplace in hydrologic models. Historically, stochastic models of streamflow have focused on monthly and annual time-scales, though there is an increasing interest in the development of stochastic models of daily streamflow. A complete stochastic model of daily streamflow must account for both the deterministic and stochastic components of daily flow series. The deterministic component must reproduce the seasonality associated with daily flow series and the stochastic component must reproduce both the persistence and frequency distribution of the daily flow series.

[6] LeBoutillier and Waylen [1993] introduced a five-parameter stochastic model of daily streamflows, which relates the FDC to the AFDC. The stochastic model developed by LeBoutillier and Waylen [1993] can reproduce the AFDC for a typical year however their model significantly underestimates the variability of observed AFDCs around the central AFDC. The authors argue that the underestimation of the variability of observed AFDCs is a consequence of neglecting the serial correlation of daily streamflows so they develop an empirical adjustment to enable preservation of the observed variance of the AFDC.

[7] Our primary goal, analogous to that of LeBoutillier and Waylen [1993], is to develop a mathematical model of the relationship between the FDC and the mean and variance of the AFDC. Achieving this goal is useful and necessary for subsequent studies which seek to (1) construct confidence intervals associated with AFDCs at ungauged sites, (2) assign return periods to individual AFDCs (3) develop regional models of flow duration curves, (4) generate daily streamflow series at ungauged sites and (5) develop a generalized stochastic model of daily streamflow.

[8] Our approach to relate the stochastic properties of the AFDC and FDC is to use an index flow method, analogous to the index flood method of regional flood frequency analysis [e.g., Dalrymple, 1960]. This approach assumes that daily streamflow is the product of two components, a long-term climatic component describing the alternation of wet and dry years, and a component that reflects the hydrological behavior of the river basin. The assumption enables the stochastic model to relate FDC and AFDC preserving the variability of observed AFDCs without resorting to empirical approximations regarding the serial structure of the daily streamflows.

2. Definition of Flow Duration Curves

[9] The FDC for a series of daily flows is the complement of the cumulative distribution function of the daily streamflows based on the complete record of flows. A nonpara-

metric approach to constructing an FDC is simple: (1) rank the observed streamflows in ascending order; (2) plot each ordered observation versus its corresponding duration or exceedance probability. The duration is often expressed as a percentage, and it coincides with an estimate of the exceedance probability, ε_i , of the i th observation in the ordered sample. If ε_i is estimated using a Weibull plotting position, the duration D_i is,

$$D_i = 100(\varepsilon_i) = 100 \left(1 - \frac{i}{n+1} \right), \text{ for } i = 1, 2, \dots, n, \quad (1)$$

where n is the length of the sample. A variety of nonparametric approaches for estimation of an FDC are introduced by Vogel and Fennessey [1994].

[10] Vogel and Fennessey [1994] suggested the computation of a series of y annual FDCs from the y -year record of daily streamflows. Each such flow duration curve based on a calendar year of data is termed an AFDC resulting in y different AFDCs, each one representing the exceedance probability of daily streamflow in a different year. Each AFDC is computed from $n = 365$ daily streamflows for the year using the same nonparametric approach given in equation (1) or one of the alternate nonparametric approaches given by Vogel and Fennessey [1994]. Then, measures of central tendency such as the mean or median AFDC can easily be derived from the set of y AFDCs, as well as the AFDC associated with a given nonexceedance probability, or, equivalently, a given recurrence interval [Vogel and Fennessey, 1994].

[11] The AFDC and the FDC are complementary rather than competitive concepts. FDCs display the complete range of observed river discharges and, therefore their interpretation depends on the period of record upon which they are based. Fennessey [1994], Hughes and Smakhtin [1996], and Smakhtin et al. [1997] showed that an FDC can be effectively used for filling gaps and for extending daily streamflow series, and, when a regional FDC model is available, for generating streamflow series at ungauged river basins. The mean and median AFDCs represent the exceedance probability of daily streamflows in a "typical", mean or median hypothetical year and are not affected by the observation of abnormally wet or dry periods during the period of record [Vogel and Fennessey, 1994].

[12] AFDCs are particularly useful because they provide confidence intervals for the median AFDC and they can be used to assign nonexceedance probabilities or, equivalently, return periods to individual AFDCs [Vogel and Fennessey, 1994]. Hence AFDCs can be effectively employed in deriving flood flow indexes, as well as low flow and water quality indexes, which are usually determined from the probabilistic structure of daily or weekly mean flows [Claps and Fiorentino, 1997]. In fact, the recurrence interval T of AFDCs can be easily expressed as a function of the nonexceedance probability, or equivalently, the percentile p of the AFDCs. If the goal is the frequency analysis of the streamflow regime for dry years, the recurrence interval T in years equals $100/p$; T equals $100/(100-p)$ otherwise. Consequently, the 2-year AFDC is the 50-percentile of AFDCs, while the 10-year AFDC is the 10-percentile of AFDCs if the interest is in droughts (e.g., derivation of low flow and water quality indexes), or the 90-percentile of AFDCs if the

frequency analysis focuses on the streamflow regime of wet years (e.g., derivation of flood flow indexes).

3. Stochastic Index Flow Model of Flow Duration Curves

[13] The index flood approach to regional flood frequency analysis [e.g., Dalrymple, 1960] is a good example of how the standardization of flows facilitates the interpretation of the statistical behavior of flood flow series. Using an analogy with the index flood approach we introduce an index flow approach to the stochastic modeling of daily streamflows. The approach assumes that the daily streamflow X is the product of two random variables, an index flow equal to the annual flow (AF) and a dimensionless daily streamflow X' ,

$$X = AF X' \quad (2)$$

AF describes the long-term climatic regime for a given basin, is mainly driven by annual precipitation, and models the alternation of dry and wet years. The standardized variable X' , or, more properly its probability density function (pdf), $f_{X'}$, is a key signature of the hydrologic behavior of the river basin. It describes the frequency of the standardized daily flows and is mainly controlled by the hydrologic regime, size and permeability of the basin.

3.1. Period of Record Flow Duration Curves (FDC)

[14] The FDC based on the complete period of record of flows is simply the complement of the cumulative distribution function (cdf) of X , F_X given by

$$\begin{aligned} F_X(x) &= P\{X \leq x\} = \int_{x_l}^x f_X(u) du = P\{AF X' \leq x\} \\ &= \int_{\Omega_{X'}} \int_{af_l}^{x/z} f_{AF,X'}(v, z) dv dz, \end{aligned} \quad (3)$$

where Ω_Y indicates the domain of a given random variable Y , f_X is the pdf of X , $f_{AF,X'}$ represents the joint probability distribution of AF and X' , and x_l and af_l are the lower bounds of $\Omega_{X'}$ and Ω_{AF} , respectively. If AF and X' are assumed to be independent, then $f_{AF,X'}$ equals the product of the two marginal distributions, and equation (3) becomes

$$F_X(x) = \int_{\Omega_{X'}} f_{X'}(z) \int_{af_l}^{x/z} f_{AF}(v) dv dz = \int_{\Omega_{X'}} f_{X'}(z) F_{AF}(x/z) dz. \quad (4)$$

where F_{AF} is the cdf of AF and $f_{X'}$ is the pdf of X' . Equation (4) can be solved analytically or numerically, provided expressions for F_{AF} and $f_{X'}$. The desired FDC can then be constructed by plotting the variable X versus the duration, equal to $100(1-F_X)$.

3.2. Annual Flow Duration Curves (AFDC)

[15] Let the daily streamflows for a given year be given by X_j , with $j = 1, 2, \dots, n$, for $n = 365$. To construct an AFDC the n flows are ordered in ascending order

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)} \quad (5)$$

where $X_{(r)}$, with rank $r = 1, 2, \dots, n$, is the r th-order statistic of the n random variables X_j [Balakrishnan and Rao, 1998].

Using equation (2), $X_{(r)}$ becomes the product of two random variables, the index flow AF and the r th-order statistic of the dimensionless daily streamflow, $X'_{(r)}$

$$X_{(r)} = AF X'_{(r)}. \quad (6)$$

We assume that the unordered dimensionless daily streamflows, X'_j , with $j = 1, 2, \dots, n$, are independent and identically distributed (iid), with cdf $F_{X'}$. Although it is well known that daily streamflows exhibit a high degree of serial correlation, the time dependence structure at the daily scale has no influence on either the AFDC or the FDC as long as the interannual variability of the annual flows is preserved.

[16] The construction of the mean AFDC requires the determination of the expected value of $X_{(r)}$, $E[X_{(r)}]$ for $r = 1, 2, \dots, n$, which, due to the supposed independence between AF and $X'_{(r)}$, is obtained from

$$E[X_{(r)}] = E[AF] E[X'_{(r)}], \text{ for } r = 1, 2, \dots, n. \quad (7)$$

$E[AF]$ and $E[X'_{(r)}]$ can be computed provided the pdf of AF and $X'_{(r)}$. If the underlying parent distribution of the dimensionless variable X' is known, then the pdf of $X'_{(r)}$, $f_{X'_{(r)}}$, or equivalently its cdf $F_{X'_{(r)}}$, can be derived using the theory of order statistics under the iid hypothesis [Balakrishnan and Rao, 1998],

$$f_{X'_{(r)}}(x) = r \binom{n}{r} [F_{X'}(x)]^{r-1} [1 - F_{X'}(x)]^{n-r} \frac{d}{dx} F_{X'}(x) \quad (8a)$$

$$F_{X'_{(r)}}(x) = \sum_{i=r}^n \binom{n}{i} [F_{X'}(x)]^i [1 - F_{X'}(x)]^{n-i} = I_{F_{X'}(x)}(r, n-r+1) \quad (8b)$$

where

$$I_p(a, b) = \frac{1}{B(a, b)} \int_0^p t^{a-1} (1-t)^{b-1} dt, \quad (9)$$

is the incomplete beta function and $B(a, b)$ is the Euler's beta function.

[17] Provided the expected values of AF and $X'_{(r)}$, for $r = 1, 2, \dots, n$, the mean AFDC can be constructed by plotting the n values of $E[X_{(r)}]$ against their corresponding duration, expressed in terms of the rank r .

[18] The interannual variability of the AFDCs can be described by the standard deviation of $X_{(r)}$, $\sigma_{(r)}$, which, due to the assumed independence of AF and $X'_{(r)}$, can be expressed as follows,

$$\sigma_{(r)} = \sqrt{E[AF^2] E[X_{(r)}^2] - E^2[X_{(r)}]}. \quad (10)$$

[19] Another approach to representing the interannual variability of flow duration curves is to compute percentiles of the AFDCs for a given duration. Percentiles of the AFDCs are useful for constructing confidence intervals for the median AFDC and for assigning return periods to AFDCs. To construct an analytical percentile AFDC it is

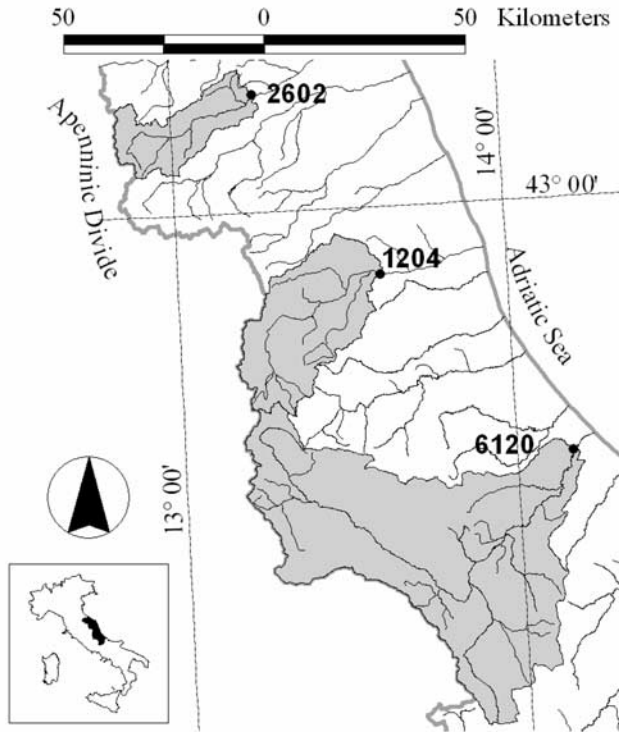


Figure 1. Location of river basins.

necessary to derive an analytical expression for the cdf of $X_{(r)}$, $FX_{(r)}$ for $r = 1, 2, \dots, n$. Analogously to the results for F_X in equation (4), we obtain,

$$F_{X_{(r)}}(x) = \int_{\Omega_{X'}} f_{X'_{(r)}}(v) F_{AF}(x/v) dv, \quad (11)$$

where $f_{X'_{(r)}}$ is given in equation (8a). The percentile of AFDCs corresponding to a given exceedance probability can be obtained from the quantile function of $X_{(r)}$, for $r = 1, 2, \dots, n$, by inverting equation (11), either analytically or numerically depending on the complexity of the expressions adopted for F_{AF} and F_X .

4. Case Study

4.1. Physiographic and Climatic Characteristics of the Basins

[20] An implementation of the index flow stochastic model of an FDC and an AFDC to three different river basins is illustrated herein. The river basins are located in the eastern central Italy (see Figure 1), are essentially unregulated, and possess more than 30 years of historic mean daily flows.

Daily streamflow series and climatic and physiographic characteristics of the river basins were obtained from the National Hydrographic Service of Italy (SIMN).

[21] The three river basins are (1) Tronto River at Tolignano (SIMN code: 1204), (2) Potenza River at Cannuccio (SIMN code: 2602) and (3) Pescara at St. Teresa (SIMN code: 6120). These three river basins have drainage areas of 900.5 km², 430.7 km² and 3082.0 km² with impervious percentages of 84%, 43% and 42%, respectively. The average annual precipitation for the three basins is 926.7 mm, 1103.7 mm and 872.2 mm respectively.

[22] Table 1 summarizes the length of record for each site, and Figure 2 illustrates the median and the 25th and 75th percentile annual hydrographs for the sites of interest, reporting daily streamflows as specific discharges in mm/h. The observations for 29 February were omitted while deriving the annual hydrographs in Figure 2 as well as during all phases of the study. Although all three basins show late winter dominated regimes (see Figure 2), this omission is not deemed to be significant due to the large number of observations available (see Table 1).

4.2. Statistical Characterization of Daily Streamflow Regime

[23] An underlying assumption of the index flow stochastic model is that the two components AF and X' can be assumed to be independent. The dependence between the observed AF and X' series was assessed by computing the correlation coefficient of the AF values and the corresponding X' values for the entire streamflow record of each site. The correlation coefficient was equal to zero in all three cases. As an example of the absence of correlation among the two series, Figure 3 plots the AF values against the corresponding X' values for site 1204, the remaining two sites present similar features. The dependence between the series was assessed further by computing the correlation coefficient between the series of AF and $X'(\delta)$ values, where δ indicates a particular day in a non-leap year. For all three sites, the series of 365 correlation coefficients shows limited values scattered around zero, and absolute values smaller than 0.15 in more than 50% of the cases. The hypothesis of independence of observed AF and X' values was assumed to hold. As section 6 will show, the results of the application of the index flow model to all three sites seem to confirm the validity of this assumption.

[24] The distribution of standardized daily streamflows results from a compound physical mechanism, and distributions with four or more parameters are generally necessary to obtain an accurate reproduction of the observed distribution of daily flows [see, e.g., *LeBoutillier and Waylen*, 1993]. Nevertheless, it is generally advisable to include in a model only a limited number of parameters, focusing attention on those parameters with a physical

Table 1. Considered River Basins: Sample Size, Parameter Estimates, and PPCC Test Statistics

Site	SIMN Code	Number of Years	Sample Size	$\hat{\xi}_L$	$\hat{\alpha}_L$	r_{LO}	r_{SLO}	$\hat{\alpha}_P$	\hat{k}_P	r_{GPA}	r_{SGPA}
Tronto at Tolignano	1204	34	12410	2.854	0.132	0.988	0.947	0.734	-0.123	0.953	0.995
Potenza at Cannuccio	2602	40	14600	2.016	0.148	0.989	0.952	0.675	-0.063	0.993	0.998
Pescara at St. Teresa	6120	59	21535	3.877	0.110	0.987	0.962	0.365	0.140	0.894	0.999

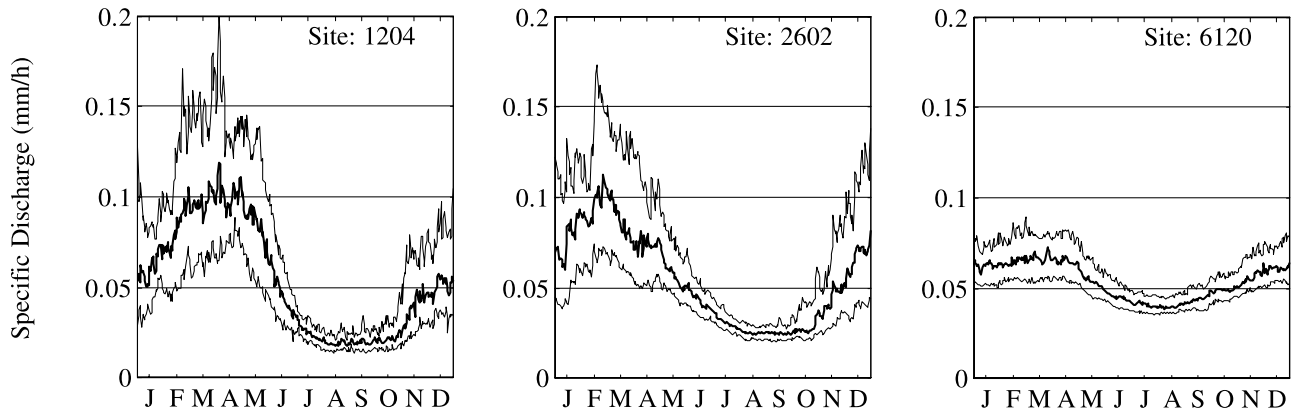


Figure 2. Annual hydrographs of mean daily flows: 25th, 50th (middle curve), and 75th percentiles.

meaning. Moreover, a reduced number of parameters is an important prerequisite for the regionalization of the FDC or AFDC stochastic model and for its application to ungauged sites.

[25] Accordingly, it was decided to represent the AF series with a two-parameter distribution and the X' series with a three-parameter distribution. The two-parameter logistic (LO) distribution was shown to provide an excellent representation of the log transformed AF series for all three sites. The cdf of the two-parameter LO distribution is,

$$F_Y(y) = \frac{1}{1 + \exp\left(-\frac{y - \xi_L}{\alpha_L}\right)}, \quad (12)$$

where ξ_L and α_L are the distribution parameters of position and scale, respectively, and $Y \in [-\infty, +\infty]$ [see *Hosking and Wallis*, 1997].

[26] Table 1 reports the estimates of parameters of position $\hat{\xi}_L$ and scale $\hat{\alpha}_L$ for the LO distribution fitted on the log-transformed AF series. The parameters were estimated using the method of L moments [Hosking and Wallis, 1997], which is often more efficient than the maximum likelihood when used with small to moderate length samples [Hosking et al., 1985; Hosking and Wallis, 1987]. A probability plot correlation coefficient (PPCC) test [Vogel, 1986] was used to test the goodness of fit at the 5% significance level. As Table 1 shows, the PPCC test statistic

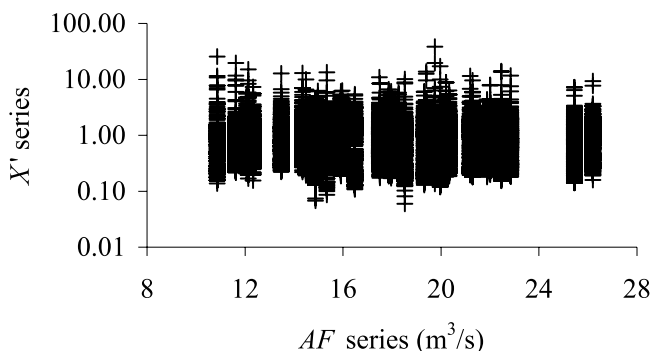


Figure 3. Dimensionless daily flow series (X') versus annual flow series (AF) for site 1204.

r_{LO} is always higher than the 5% level test statistic for the logistic distribution, r_{5LO} . The r_{5LO} test statistic was obtained by generating 10,000 sequences of logistic samples each with length equal to the original sample length, using the parameters estimates $\hat{\xi}_L$ and $\hat{\alpha}_L$ reported in Table 1 and a Weibull plotting position.

[27] Of the three-parameter distributions, the L moment ratios diagram [Hosking and Wallis, 1997] illustrated in Figure 4 indicates that the generalized Pareto (GPA) distribution is the most suitable choice for sites 1204 and 2602. The cdf of the three-parameter GPA distribution is,

$$F_Y(y) = 1 - \left[1 - k_P \frac{y - \xi_P}{\alpha_P}\right]^{\frac{1}{k_P}}, \quad (13)$$

where ξ_P , α_P and k_P are the distribution parameters of position, scale and shape, respectively, and $Y \in [\xi_P, \infty]$ if $k_P \leq 0$, and $Y \in [\xi_P, \xi_P + \alpha_P/k_P]$ if $k_P > 0$. Furthermore, Figure 4 shows that the generalized extreme value (GEV) distribution [Jenkinson, 1955] is the most appropriate distribution for representing the sample of dimensionless daily streamflows at site 6120.

[28] The GPA distribution was fit to the X' series using the method of L moments, which were obtained for all three

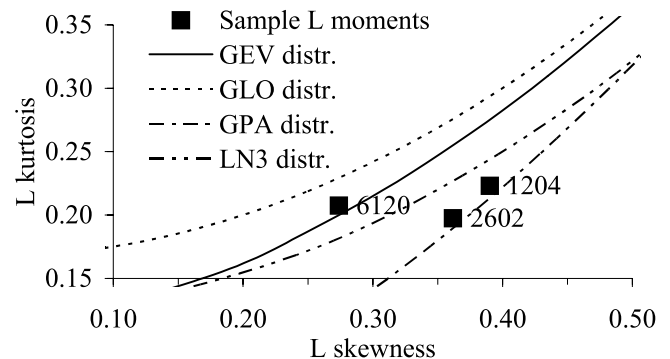


Figure 4. L moment ratio diagram for the standardized daily streamflows (three-parameter distributions: generalized extreme value (GEV), generalized logistic (GLO), generalized Pareto (GPA), and three-parameter lognormal (LN3) distributions).

sites by dividing each daily streamflow by the annual flow for the corresponding year. The estimates of the scale, $\hat{\alpha}_P$, and shape, \hat{k}_P , parameters are reported in Table 1. Table 1 also reports the r_{GPA} statistic and the 5% significance level PPCC test statistic, r_{5GPA} , which was obtained using the same procedure adopted for determining r_{5LO} . As Table 1 shows, the GPA hypothesis as the parent distribution of the standardized daily streamflows must be rejected at the 5% significance level for all three sites.

5. Log Logistic–Generalized Pareto Stochastic Model

[29] Limbrunner *et al.* [2000, Figure 7] and Vogel and Fennessey [1993, Figure 2] used L moment diagrams to show that the GPA and the three-parameter lognormal (LN3) pdfs provide a good approximation to the pdf of daily streamflow at hundreds of basins in the United States. Although the results of the PPCC test are not encouraging and despite what is illustrated in Figure 4 for site 6120, the GPA distribution was adopted for representing the observed X' series of all three sites. The present section presents a parameterization of the stochastic model that utilizes a two-parameter logistic (LO) distribution for modeling the log-transformed AF, and a three-parameter generalized Pareto (GPA) distribution to model X' , the combination of these two models is termed a LO-GPA model. The implementation of a LO-GPA parameterization of the model to the three considered sites enables us to evaluate (1) the suitability of the index flow hypothesis for a stochastic modeling of FDC and AFDC, (2) the sensitivity of the model to the selection of a particular distribution for representing the X' series, which is not necessarily the most suitable one.

[30] Although the total number of parameters for the LO-GPA model is 5, the implementation of the LO-GPA model requires estimation of 4 parameters, as $E[X'] = 1$, hence the following relationship between ξ_P , α_P and k_P holds,

$$\xi_P = E[X'] - \frac{\alpha_P}{1 + k_P} = 1 - \frac{\alpha_P}{1 + k_P}. \quad (14)$$

5.1. LO-GPA Stochastic Model of an FDC

[31] It can be shown that the cdf of daily flows in equation (4) becomes,

$$F_X(x) = \int_1^{u^*} \frac{-u^{\frac{1-k_P}{k_P}}}{k_P \left(1 + e^{g(u)} / x^{1/\alpha_L}\right)} du, \quad (15a)$$

where u^* equals ∞ if $k_P \leq 0$ and 0 if $k_P > 0$, and,

$$g(u) = \frac{\xi_L + \ln\left(\frac{\alpha_P}{k_P}(1-u) + 1 - \frac{\alpha_P}{1+k_P}\right)}{\alpha_L}. \quad (15b)$$

The FDC can be determined from equation (15) by plotting the variable X versus the duration, $100(1-F_X)$, provided estimates of the parameters ξ_L , α_L , α_P and k_P are available.

5.2. LO-GPA Stochastic Model of an AFDC

[32] The mean AFDC is given by equation (7), which depends on the expected values of AF and $X'_{(r)}$ for each

value of r . For the LO-GPA model, $E[AF]$ can be computed from

$$E[AF] = e^{\xi_L} \int_0^1 \left(\frac{F}{1-F}\right)^{\alpha_L} dF. \quad (16)$$

$E[X'_{(r)}]$ can be estimated using series approximations [Balakrishnan and Rao, 1998], or by numerical integration, provided an analytical expression for $f_{X'_{(r)}}$ is available. For the LO-GPA model we obtain,

$$f_{X'_{(r)}}(x) = \frac{r}{\alpha_P} \binom{n}{r} (1-h(x))^{r-1} h(x)^{n-r+1-k_P}, \quad (17a)$$

where,

$$h(x) = \left(1 - k_P \frac{x - \xi_P}{\alpha_P}\right)^{\frac{1}{k_P}}. \quad (17b)$$

Analogous results exist for $\sigma_{X'_{(r)}}$ through use of equation (10).

[33] The p^* percentile AFDC, can be obtained from,

$$F_{X'_{(r)}}(x^*) = \frac{p^*}{100}, \text{ for } r = 1, 2, \dots, n, \quad (18)$$

where $F_{X'_{(r)}}$ is given in equation (11). An alternative formulation for equation (11) is,

$$F_{X'_{(r)}}(x) = \left[F_{X'_{(r)}}(u)F_{MAF}(x/u)\right]_{x'_l}^{x'_u} - \int_{\Omega_{X'}} F_{X'_{(r)}}(v) \frac{d}{dv} F_{MAF}(x/v) dv, \quad (19)$$

where x'_l and x'_u indicate the lower and upper limits of $\Omega_{X'}$, respectively. Recalling $F_{X'_{(r)}}$ from equation (8b) and using the fact that $F_{X'_{(r)}}(x'_l) = 0$, and $F_{X'_{(r)}}(x'_u) = 1$, equation (19) can be written as

$$F_{X'_{(r)}}(x) = F_{MAF}(x/x'_u) - \int_{\Omega_{X'}} I_{F_{X'_{(r)}}(v)}(r, n-r+1) \frac{d}{dv} F_{MAF}(x/v) dv. \quad (20)$$

For the LO-GPA stochastic model, equation (20) becomes

$$F_{X'_{(r)}}(x) = \Theta + \int_{\frac{\alpha_P}{1+k_P}}^{v^*} I_{F_{X'_{(r)}}(v)}(r, n-r+1) \frac{l(v, x)}{\alpha_L v (1 + l(v, x))^2} dv, \quad (21a)$$

where $F_{X'}$ is the cdf of the GPA distribution, given in equation (13),

$$l(v, x) = \exp\left(\frac{1}{\alpha_L} \left(\xi_L - \ln\left(\frac{x}{v}\right)\right)\right), \quad (21b)$$

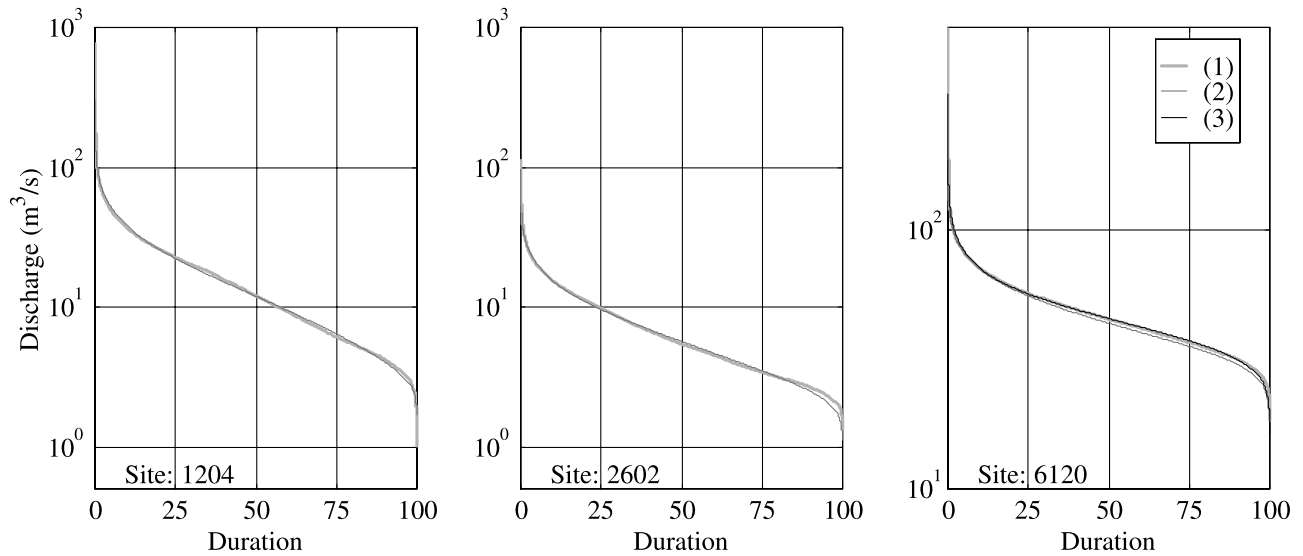


Figure 5. FDC: 1, empirical; 2, LO-GPA model, numerical integration of equation (15); 3, LO-GEV model, Monte Carlo experiment.

$\Theta = 0$ and $v^* = \infty$ if $k_p \leq 0$, and

$$\Theta = \left(1 + \exp(\xi_L/\alpha_L) \left((k_p(k_p + 1) + \alpha_p) / (x k_p(k_p + 1)) \right)^{\frac{1}{\alpha_p}} \right)^{-1} \quad (21c)$$

$$\text{and } v^* = 1 + \frac{\alpha_p}{k_p + k_p^2},$$

otherwise.

6. Results and Discussion

[34] Figure 5 compares the empirical FDC and FDC resulting from the application of the stochastic LO-GPA model. The LO-GPA FDC was obtained by numerically integrating equation (15) and Monte Carlo simulations were used to confirm that equation (15) is correct.

[35] Figures 6 and 7 compare the empirical estimates of $E[X_{(r)}]$ and $\sigma_{(r)}$ with the estimates of $E[X_{(r)}]$ and $\sigma_{(r)}$

obtained from LO-GPA stochastic model by numerically solving equations (7) and (10). Again, Monte Carlo experiments were also performed to assure that the theoretical expressions in equations (7) and (10) are correct.

[36] Figure 8 compares the empirical 25th, 50th, and 75th percentile of the AFDC with the 25th, 50th, and 75th percentiles obtained from the LO-GPA model using equation (18), for the 21 possible combinations resulting from $p^* = 25, 50, 75$ and $r = 1, 62, 123, 184, 245, 293, 365$. The 21 numerical solutions were derived using a combination of the Levenberg-Marquardt, Quasi-Newton and Conjugate Gradient algorithms [Polak, 1997], and then tested by means of Monte Carlo experiments.

[37] Figures 5 and 6 show that the LO-GPA model accurately reproduces the empirical FDC and mean AFDC for sites 1204 and 2602. Moreover, Figure 7 reveals that for these two sites the LO-GPA model provides an excellent

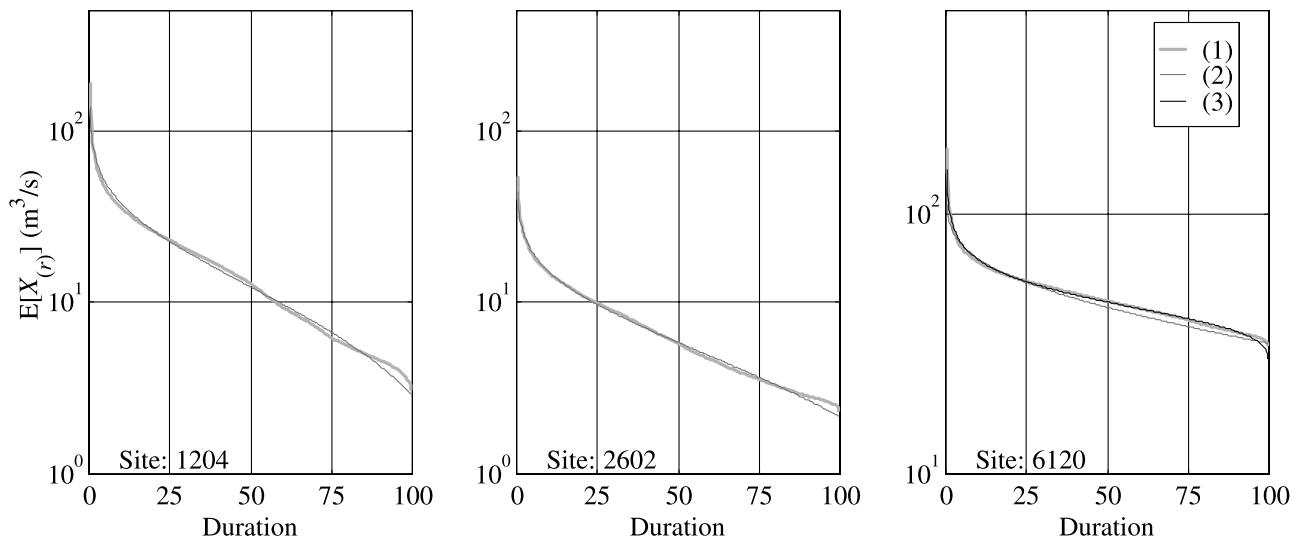


Figure 6. Mean AFDC: 1, empirical; 2, LO-GPA model, numerical solution of equation (7); 3, LO-GEV model, Monte Carlo experiment.

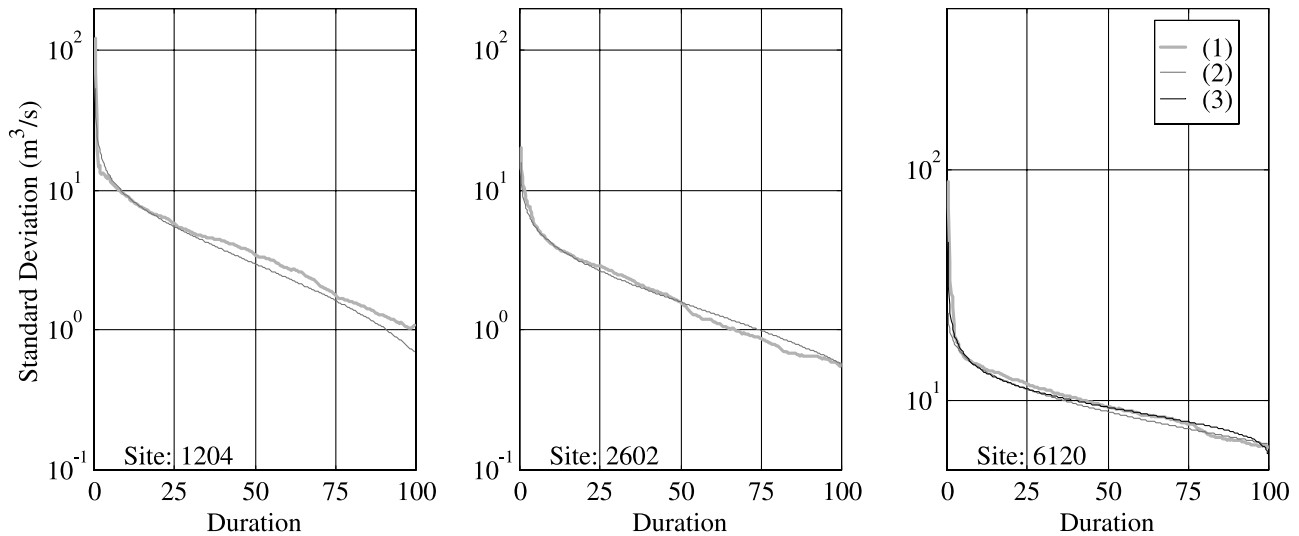


Figure 7. $\sigma(r)$ for $r = 1, 2, \dots, n$, with $n = 365$: 1, empirical; 2, LO-GPA model, numerical solution of equation (10); 3, LO-GEV model, Monte Carlo experiment.

approximation of the relationship between the standard deviation $\sigma(r)$ and the duration $D(r)$. As a result, the 25th, 50th, and 75th percentile AFDC derived by applying the stochastic model are consistent with the corresponding empirical counterparts, as shown in Figure 8. This is an innovation over the results of *LeBoutillier and Waylen* [1993], who were unable to reproduce $\sigma(r)$ without resorting to an empirical correction.

[38] The reproduction of the empirical FDC is still accurate for site 6120 (Figure 5), and Figure 6 shows that the mean AFDC obtained for site 6120 is also a reasonable approximation of the empirical curve. Furthermore, Figure 7 shows for the same site that the LO-GPA model is capable of capturing the main features of the empirical $\sigma(r)$. Consequently, Figure 8 shows for site 6120 a fair agreement between the empirical and modeled percentiles of AFDC.

[39] The slightly poorer results obtained for site 6120 could be a consequence of having represented the distribu-

tion of standardized daily flows with the GPA distribution instead of the GEV distribution (see Figure 4). The performance of a different parameterization of the index flow stochastic model of FDC was tested for site 6120. The parameterization still adopts a LO distribution to represent the log-transformed AF series, with parameters of position and scale reported in Table 1, but it uses a GEV distribution instead of a GPA distribution to represent the standardized daily flows for the considered site. The GEV distribution parameters were estimated by the method of L moments, setting to one the mean of the distribution. Figure 4 notwithstanding, the results of a PPCC goodness of fit indicate that the GEV hypothesis should not be accepted at the 5% significance level (i.e., $r_{GEV} = 0.983$ for site 6120, while $r_{5GEV} = 0.996$), analogously to what obtained for sites 1204 and 2602 and the GPA distribution.

[40] For the LO-GPA stochastic model we found perfect agreement between the results obtained by numerically

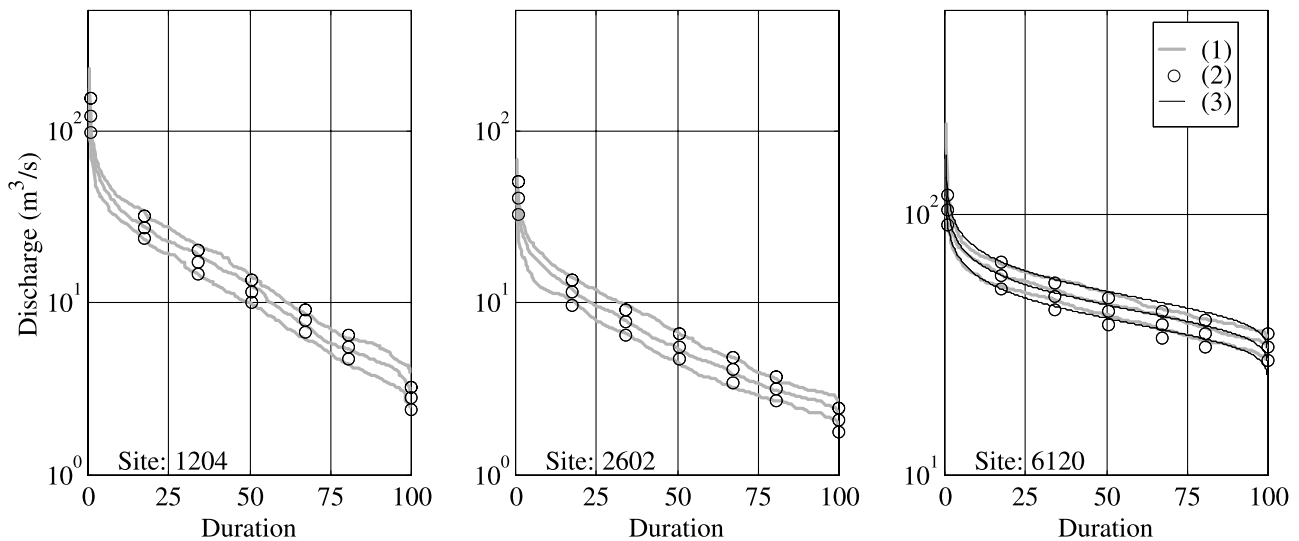


Figure 8. The 25th, 50th, and 75th percentile AFDC: 1, empirical; 2, LO-GPA model, numerical solution of equation (18); 3, LO-GEV model, Monte Carlo experiment.

integrating the theoretical equations and by performing repeated Monte Carlo simulations. Therefore the performance of the LO-GEV model for station 6120 are assessed here using Monte Carlo simulation only, by generating 100 random realizations from the LO-GEV model; each realization consists of a synthetic daily streamflow series with length equal to the original sample of site 6120 (see Table 1).

[41] The FDC, mean AFDC, the relationship between $\sigma(r)$ and $D(r)$ and the 25-, 50- and 75-percentile AFDC obtained for the LO-GEV model are reported in Figures 5, 6, 7, and 8. Compared to the LO-GPA model, the LO-GEV implementation of the stochastic index flow model of FDC provides for site 6560 a superior representation of the FDC (Figure 5), an improved approximation of the mean AFDC (Figure 6) and an enhanced representation of the relationship between $\sigma(r)$ and $D(r)$ (Figure 7). As a result, the LO-GEV estimates of the percentile AFDC present a better reproduction of the empirical percentile AFDCs (Figure 8).

[42] In summary, the applications of a four-parameter index flow FDC model were shown to provide high-quality representations of the FDC as well as the mean and median AFDC for the three sites. Furthermore, the models produced good, and in some cases excellent, approximations of interannual variability of FDC. The application showed that the index flow stochastic model is sensitive to the selection of a particular distribution for representing the series of X' values, even though, for the case study considered herein, the replacement of the parent distribution produced only slight modifications of the FDC and AFDCs.

7. Conclusions

[43] The primary goal of this study, analogous to that of LeBoutillier and Waylen [1993] was to develop a mathematical model of the relationship between the period of record flow duration curve (FDC) and the mean and variance of the annual flow duration curve (AFDC). Achieving this goal is useful and necessary for subsequent studies which seek to (1) construct confidence intervals associated with AFDCs at ungauged sites, (2) assign return periods to individual AFDCs (3) develop regional models of flow duration curves, (4) generate daily streamflow series at ungauged sites, and (5) develop a generalized stochastic model of daily streamflow.

[44] An index flow approach was introduced for modeling the relationship between an FDC and AFDCs and implemented using a four-parameter stochastic model on three river basins located in eastern-central Italy. Our index flow approach to modeling an FDC was able to provide an accurate description of the FDC as well as the mean and variance of the AFDCs. The primary innovation of this paper is that our index flow method is able to capture the interannual variability of AFDCs without representing the serial correlation and seasonality of daily flows. This is accomplished by standardizing the daily streamflow by dividing by the annual flow for the year in which the flow occurred. This simple step avoids the need for a much more complex theoretical analysis requiring assumptions regarding the stochastic (persistence and seasonality) structure of daily flow series. As a consequence, the proposed index flow model of FDC

represents an improvement to the stochastic modeling of an FDC over previous results reported by LeBoutillier and Waylen [1993].

[45] The index flow model should be tested further in different geographical areas and climatic contexts in order to gain a better understanding of the model's overall applicability and reliability. These initial results indicate that the index flow approach to the stochastic modeling of daily flows and its application to FDCs and AFDCs offer promising and interesting avenues for future research. The most valuable possibilities are likely to be the regionalization of the approach and its application for the generation of daily streamflow series at ungauged river basins.

[46] **Acknowledgments.** The thoughtful and very useful comments of two anonymous reviewers are gratefully acknowledged. The first author wishes to thank the University of Bologna for partially supporting this work with a research grant (Progetto Marco Polo – 2002, 3rd session).

References

- Balakrishnan, N., and C. R. Rao (1998), Order statistics: An introduction, in *Order Statistics: Theory and Methods, Handbk. Stat.*, vol. 16, edited by N. Balakrishnan and C. R. Rao, pp. 3–24, Elsevier Sci., New York.
- Cigizoglu, H. K., and M. Bayazit (2000), A generalized seasonal model for flow duration curve, *Hydrol. Processes*, 14, 1053–1067.
- Claps, P., and M. Fiorentino (1997), Probabilistic Flow duration curves for use in environmental planning and management, in *Integrated Approach to Environmental Data Management Systems, NATO ASI Ser.*, Ser. 2, vol. 31, edited by N. B. Harmancioglu et al., pp. 255–266, Kluwer Acad., Norwell, Mass.
- Crocker, K. M., M. D. Z. Young, and H. G. Rees (2003), Flow duration curve estimation in ephemeral catchments in Portugal, *Hydrol. Sci. J.*, 48(3), 427–439.
- Dalrymple, T. (1960), Flood frequency analyses, *U.S. Geol. Surv. Water Supply Pap.*, 1543-A.
- Fennessey, N. M. (1994), A hydro-climatological model of daily streamflows for the northeast United States, Ph.D. dissertation, Tufts Univ., Medford, Mass.
- Foster, H. A. (1934), Duration curves, *Trans. Am. Soc. Civ. Eng.*, 99, 1213–1267.
- Good, J. C., and J. M. Jacobs (2001), Ecologically sustainable watershed management using annualized flow duration curves, paper presented at World Water and Environmental Resources Congress, Am. Soc. of Civ. Eng., Orlando, Fla., May.
- Hosking, J. R. M., and J. R. Wallis (1987), Parameter and quantile estimation for the generalized Pareto distribution, *Technometrics*, 29, 339–349.
- Hosking, J. R. M., and J. R. Wallis (1997), *Regional Frequency Analysis: An Approach Based on L-Moments*, Cambridge Univ. Press, New York.
- Hosking, J. R. M., J. R. Wallis, and E. F. Wood (1985), Estimation of the generalized extreme value distribution by the method of probability-weighted moments, *Technometrics*, 27, 251–261.
- Hughes, D. A., and V. Y. Smakhtin (1996), Daily flow time series patching or extension: A spatial interpolation approach based on flow duration curves, *Hydrol. Sci. J.*, 41(6), 851–871.
- Jenkinson, A. F. (1955), The frequency distribution of the annual maximum (or minimum) of meteorological elements, *Q. J. R. Meteorol. Soc.*, 81, 158–171.
- LeBoutillier, D. V., and P. R. Waylen (1993), A stochastic model of flow duration curves, *Water Resour. Res.*, 29(10), 3535–3541.
- Limbrunner, J. F., R. M. Vogel, and L. C. Brown (2000), Estimation of harmonic mean of a lognormal variable, *J. Hydrol. Eng.*, 5(1), 59–66.
- Polak, E. (1997), *Optimization: Algorithms and Consistent Approximations*, Springer-Verlag, New York.
- Searcy, J. C. (1959), Manual of hydrology, 2, Low flow techniques, flow-duration curves, *U.S. Geol. Surv. Water Supply Pap.*, 1542-A.
- Smakhtin, V. U. (2001), Low flow hydrology: A review, *J. Hydrol.*, 240, 147–186.
- Smakhtin, V. Y., and M. Toulouse (1998), Relationships between low-flow characteristics of South African streams, *Water South Afr.*, 24(2), 107–112.

- Smakhtin, V. Y., D. A. Hughes, and E. Creuse-Naudine (1997), Regionalization of daily flow characteristics in part of the Eastern Cape, South Africa, *Hydrol. Sci. J.*, 42(6), 919–936.
- Sugiyama, H., V. Vudhivanich, A. C. Whitaker, and K. Lorsirirat (2003), Stochastic flow duration curves for evaluation of flow regimes of rivers, *J. Am. Water Resour. Assoc.*, 39(1), 47–58.
- Vogel, R. M. (1986), The probability plot correlation coefficient test for the normal, log-normal and Gumbel distributional hypothesis, *Water Resour. Res.*, 22(4), 587–590.
- Vogel, R. M., and N. M. Fennessey (1993), L moment diagrams should replace product moment diagrams, *Water Resour. Res.*, 29(6), 1745–1752.
- Vogel, R. M., and N. M. Fennessey (1994), Flow-duration curves. I: New interpretation and confidence intervals, *J. Water Resour. Plann. Manage.*, 120(4), 485–504.
- Vogel, R. M., and N. M. Fennessey (1995), Flow-duration curves. II: A review of applications in water resources planning, *Water Resour. Bull.*, 31(6), 1029–1039.
- Yu, P. S., T. C. Yang, and Y. C. Wang (2002), Uncertainty analysis of regional flow duration curves, *J. Water Resour. Plann. Manage.*, 128(6), 424–430.

A. Brath and A. Castellarin, Faculty of Engineering, Università di Bologna, Viale Risorgimento 2, 40136 Bologna, Italy. (armando.brath@mail.ing.unibo.it; attilio.castellarin@mail.ing.unibo.it)

R. M. Vogel, WaterSHED Center, Department of Civil and Environmental Engineering, Tufts University, Medford, MA 02155, USA. (richard.vogel@tufts.edu)