Stochastic model of flow duration curves for selected rivers in Bangladesh

M. F. BARI[1] & KH. MD. SHAFIUL ISLAM[2]

1 *Department of Water Resources Engineering, Bangladesh University of Engineering & Technology, Dhaka 1000, Bangladesh*
bari@wre.buet.ac.bd

2 *Department of Civil Engineering, Khulna University of Engineering & Technology, Khulna 920300, Bangladesh*

Abstract To overcome the difficulty of a traditional flow duration curve (FDC) in which the chronological sequence of occurrence of flows is masked, a stochastic approach has been applied to obtain a calendar year FDC. Estimates of mean flows by rank based on order statistics enable the construction of such a FDC once the underlying frequency characteristics and distribution of the population is known. This paper discusses the theoretical development of a stochastic FDC and the choice of a suitable probability distribution for mean daily discharges. The flow duration curve model is then applied to records from four selected rivers in north-western Bangladesh. Based on visual as well as χ^2 and Kolmogorov-Smirnov goodness-of-fit tests, the mixture lognormal distribution exhibited a better fit. Expected flow duration curves were generated for each gauging station using the chosen distribution and visually compared with observed curves. These flow duration curves represent, respectively, the computed and observed mean discharges by rank and showed a close agreement. Reliability of the FDC was investigated by comparison of computed standard deviations for the chosen distribution with those observed for all ranks, and the mixture lognormal was found to be a flexible distribution for the highly seasonal regimes of the selected rivers.

Key words flow duration curve; goodness-of-fit; mean daily discharge; mixed lognormal distribution; northwest Bangladesh; stochastic model

INTRODUCTION

Knowledge of the frequency of various flow rates in a river is essential for assessment of water available for consumptive uses, such as municipal and industrial supply, irrigation as well in-stream needs for navigation, hydropower, dilution of wastes, fish and aquatic ecology. Traditionally, the entire flow frequency regime has been summarized by a flow duration curve, which describes the average characteristics of all flow quantiles in a year by representing the proportion of time that a particular discharge in a river is equalled or exceeded during the period of observation. One drawback of this type of flow duration curve is that the chronological sequence of occurrence of the flows is masked. To overcome this difficulty, LeBoutillier & Waylen (1993) used a stochastic approach and obtained a calendar year flow duration curve. In this approach, observations of discharge in each year are ranked and averages by rank over all years of record are calculated. Then the flow duration curve represents the average ranked flow and preserves such flow frequency measures as the mean annual flood. In calculating the calendar year flow duration curve, both the mean discharge of each rank and an associated distribution about the mean may be retained. The nature of this distribution reflects the physical processes within the basin. The estimation of this curve has a strong theoretical parallel to order statistics that permit estimation of the mean discharge at any rank and its variance. LeBoutillier & Waylen (*1993*) discussed the theoretical development of a stochastic flow duration curve and applied the model to records from several rivers in British Columbia, Canada. The objective of this paper is to apply this methodology to discharge records from selected rivers in Bangladesh and obtain stochastic flow duration curves.

STOCHASTIC MODEL OF FLOW DURATION

Stochastic model of flow duration curve is based on the principle of order statistics. An order statistic is a ranked observation from a sample (Galambos, 1984). Let X_1, X_2, X_3, … X_n denote a random sample from a parent population with continuous cumulative distribution function F_x. Arranged in decreasing order, $X_{(1)}$, > $X_{(2)}$, > $X_{(3)}$, >…,..> $X_{(n)}$ are collectively termed the order statistics of the random sample, such that $X_{(r)}$, for $1 < r \le n$, is the rth order statistic (Gibbons, 1985). Then annual maximum flow in a year will be denoted by $X_{(1)}$, the annual minimum value by $X_{(365)}$, and the annual median by $X_{(183)}$ in a non-leap year. Estimates of $\mu_{(r)}$, the mean by rank, from the order statistics theory should enable the construction of a flow duration curve, once the underlying frequency distribution of the population is known. Gibbons (1985) presents the general results of the large sample approximation to the mean and variance of the *r*th order statistic $X_{(r)}$ of a sample of size *n* from the continuous distribution F_x. The approximation to the mean by rank, $\mu_{(r)}$, is:

$$\mu_{(r)} = F_x^{-1}[h(r)] \tag{1}$$

The approximation to the variance, by rank is:

$$\sigma^2_{(r)} = g(r)\{f_x[F_x^{-1}(h(r))]\}^{-2} \tag{2}$$

where $h(r) = 1 - [\frac{r}{n+1}]$, and $g(r) = \frac{r(n-r+1)}{(n+1)^2(n+2)}$

F_x and F_x^{-1} are the cumulative distribution function of the parent population and its inverse, f_x is the density function. Generally the distribution of daily discharge is similar to those of many seasonal series and is both highly skewed and peaked. This shape is suggestive of a lognormal, gamma or a generalized extreme value distribution (LeBoutillier & Waylen, 1993).

DATA

Mean daily discharge from four gauging stations on four rivers, the Baral, Dharla, Punarbhaba, and Teesta in the northwest region of Bangladesh, were selected for this study. The available daily discharge record lengths used were 29 years for the Baral River at Malanchi railway crossing (Station 16.1), 18 years for the Dharla at Kurigram (Station 76), 29 years for the Punarbhaba at Phulhat (Station 236), and 34 years for the Teesta at Kaunia (Station 294). The rivers in this region are dominated by two main seasons, the monsoon and dry season. The monsoon lasts from June to September and is characterized by heavy rainfall, high humidity and high flows in rivers. In the dry season river discharges reduce significantly, exhibiting skewness and peakedness over the annual cycle. The steps involved in application of the stochastic flow duration curve model are: (a) rank each year's data from largest to smallest, (b) calculate means by rank across all years, (c) pool all observations and fit probability distributions to the aggregate of all daily discharge observations, (d) select a suitable distribution on the basis of goodness-of-fit, (e) generate the means by rank, and (f) assess the performance of the chosen distribution's ability to reproduce the flow duration curve.

ANALYSIS AND RESULTS

Analysis began by ranking the daily discharge data of each year in descending order for each station. The mean and variance of rank across all years of data were calculated. Then the 365 values of calculated means by rank were used to determine a suitable probability distribution to model the stochastic flow duration curve. The model stochastic flow duration curve was then obtained and reliability assessed.

Estimation of parameters of selected distributions

Four probability distributions: lognormal, gamma, GEV, and mixture lognormal, possessing the ability to match the combination of skewness and peakedness were chosen. The parameters of the lognormal, gamma, and mixture lognormal distributions were estimated by the maximum likelihood method and that of the GEV distribution by the probability weighted method (Greenwood *et al.*, 1979). Parameter estimates are summarized in Table 1.

Goodness-of-fit tests

For visual goodness-of-fit tests, theoretical distributions were superimposed on empirical distributions. For deriving empirical distributions, observations of mean daily discharge for each of the four selected rivers were assembled into 20 equal groups, relative frequencies computed and plotted

Table 1 Parameter estimates and goodness-of-fit test statistics for selected distributions.

Distribution	Parameter	River Baral	River Dharla	River Punarbhaba	River Teesta
Two-parameter Lognormal	μ_y	2.300	5.412	3.073	6.230
	σ_y	2.182	0.992	1.495	1.083
	K-S	0.0812	0.0988	0.0374	0.0398
	χ^2	48.6	70.3	25.8	19.5
Three-parameter Lognormal	μ_y	2.190	4.801	2.672	5.664
	σ_y	2.284	1.583	1.900	1.617
	a	0.251	55.16	2.671	102.9
	K-S	0.0788	0.0877	0.0358	0.0366
	χ^2	42.1	52.3	24.1	18.7
Two-parameter Gamma	η	0.341	1.108	0.501	1.005
	λ	212.6	338.3	152.8	874.0
	K-S	0.1533	0.2064	0.1126	0.1088
	χ^2	77.3	142.5	251.9	35.5
Three-parameter Gamma	α	1.284	1.284	1.284	1.284
	β	84.78	309.8	87.4	750.1
	γ	–36.5	–23.0	–35.7	–85.5
	K-S	0.1677	0.1729	0.2662	0.1125
	χ^2	81.9	137.3	373.0	38.9
Generalized extreme value	K	–0.51	–0.40	–0.66	–0.34
	α	35.95	170.1	25.90	448.8
	ξ	16.09	170.4	16.74	398.7
	K-S	0.2066	0.2009	0.0596	0.1443
	χ^2	93.7	155.6	39.1	42.7
Mixture Lognormal	μ_{y1}	0.830	4.649	2.006	5.345
	σ_{y1}	0.857	0.355	0.560	0.420
	ρ	0.652	0.570	0.586	0.562
	μ_{y2}	5.054	6.422	4.585	7.302
	σ_{y2}	0.505	0.568	1.026	0.540
	K-S	0.0536	0.0766	0.0290	0.0299
	χ^2	29.9	42.2	23.6	15.5

against class intervals. Four candidate distributions were then superimposed on the empirical distributions. As expected, the mixture lognormal distribution exhibited a better fit because the daily discharge data series at hand were skewed and peaked with accentuated upper tails. Two quantitative tests, the χ^2 and K-S tests, were also applied to further ascertain the goodness-of-fit. The values of the χ^2 and K-S statistics are included in Table 1 and these also suggest that the mixture lognormal distribution exhibit a better fit.

Stochastic flow duration curve

The mean daily flow values by rank for each of four selected rivers were computed using the mixture lognormal distribution with the aid of equation (1). Model stochastic flow duration curves based on the mixture lognormal distribution were then obtained for each river plotting the computed rank flow values against respective ranks. Similarly, observed stochastic flow duration curves were obtained by plotting measured mean flows by rank, which were computed earlier, against corresponding ranks. Figure 1 shows a visual comparison of the model stochastic and observed stochastic flow duration curves for the Baral River at Malanchi railway crossing. Figure 1 and similar comparisons for other rivers showed that the mixture lognormal distribution is suitable for deriving stochastic flow duration curves for the four selected rivers.

To assess the structure and reliability of the flow duration curve, rank variances computed by the mixture lognormal distribution were compared with the rank variances obtained from observed discharges. In Fig. 2 the computed and observed standard deviations are compared for the Baral River. It is seen that the shapes are similar but the magnitudes are different and similar differences were found for other rivers. The consistent underestimation of the variance for all rivers is probably due to the inflation of the total variance in the time series resulting from seasonality. The year to year consistency of the seasonality function controls variance in the intermediate ranks.

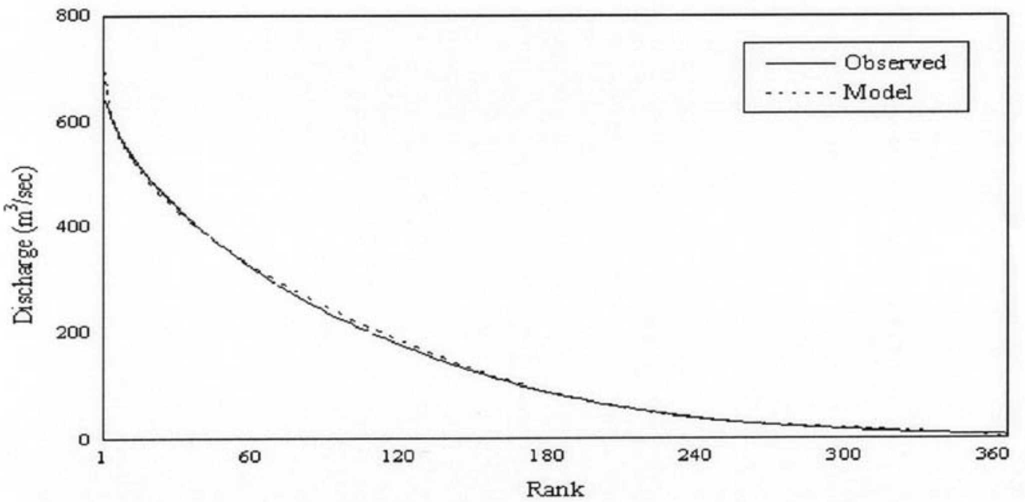

Fig. 1 Observed and model stochastic flow duration curve based on mixture lognormal distribution for Baral river at Malanchi railway crossing.

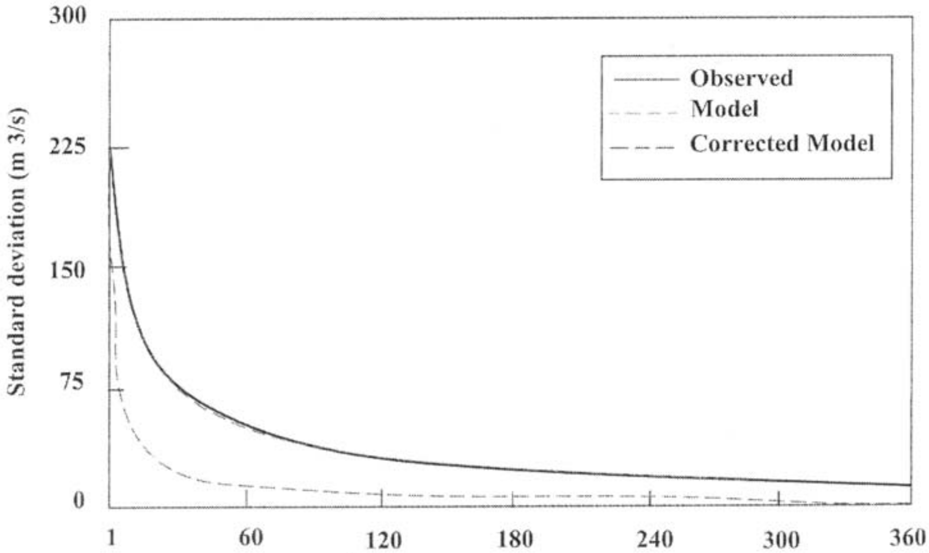

Fig. 2 Observed, model and corrected standard deviation by rank for Baral River at Malanchi railway crossing.

The more consistent the seasonality function, the higher would be the variance across all ranks that may be expected. The Baral River appeared to possess greater seasonal consistency in both the daily discharge and its variability, whereas the Dhrala, Punarbhaba and Teesta rivers were found to be the less consistent. Application of the following nonlinear regression correction (LeBoutillier & Waylen, 1993; Islam, 2000) may improve fit:

$$\hat{\sigma}^2_{(r)} = a(\sigma^2_{(r)})^b r^c \tag{3}$$

The correction contains a simple arithmetic scaling factor, *a* to inflate the variances and applies a proportional scaling by rank, *c*, and an inverse scaling for the model variances, *b*, such that the greater scaling is placed at higher ranks where the large sample approximations perform the poorest. Selection of the form of the nonlinear regression correction equation was made because of its intrinsic linearity, ease of estimation and its ability to reproduce the lack-of-fit characteristics of the variance model (LeBoutillier & Waylen, 1993). As shown in Fig. 2, the corrected model standard deviation by rank are superimposed on the observed and the model standard deviation for the Baral River and this was done for each river. Finally, the envelope of stochastic flow duration curves about ±1 standard deviation, corrected by nonlinear regression equation by rank, was generated and Fig. 3 shows one such curve for Baral River.

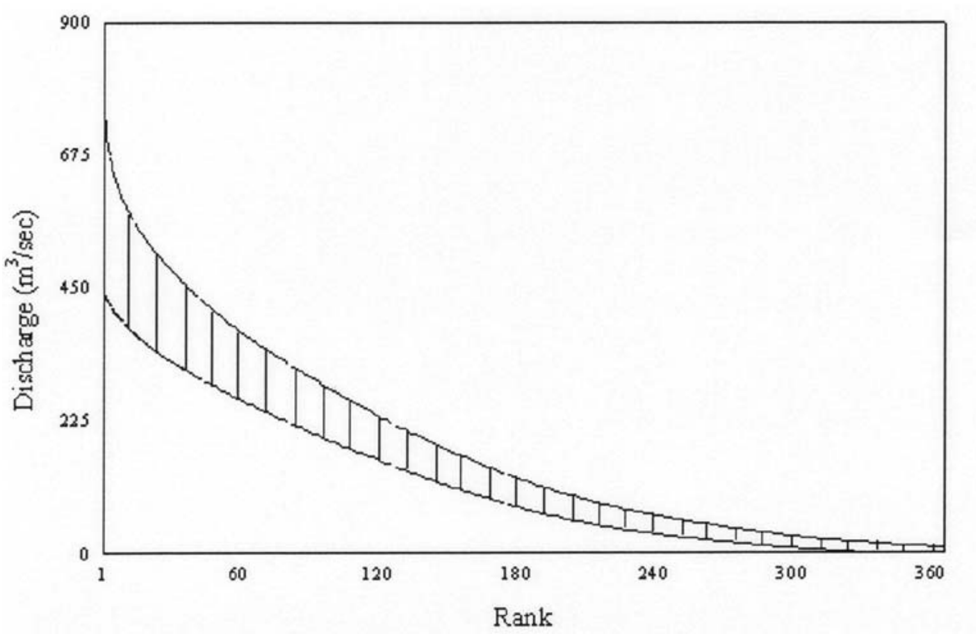

Fig. 3 Envelope of flow duration curves by ranks about ±1 standard deviation for the Baral River at Malanchi railway crossing.

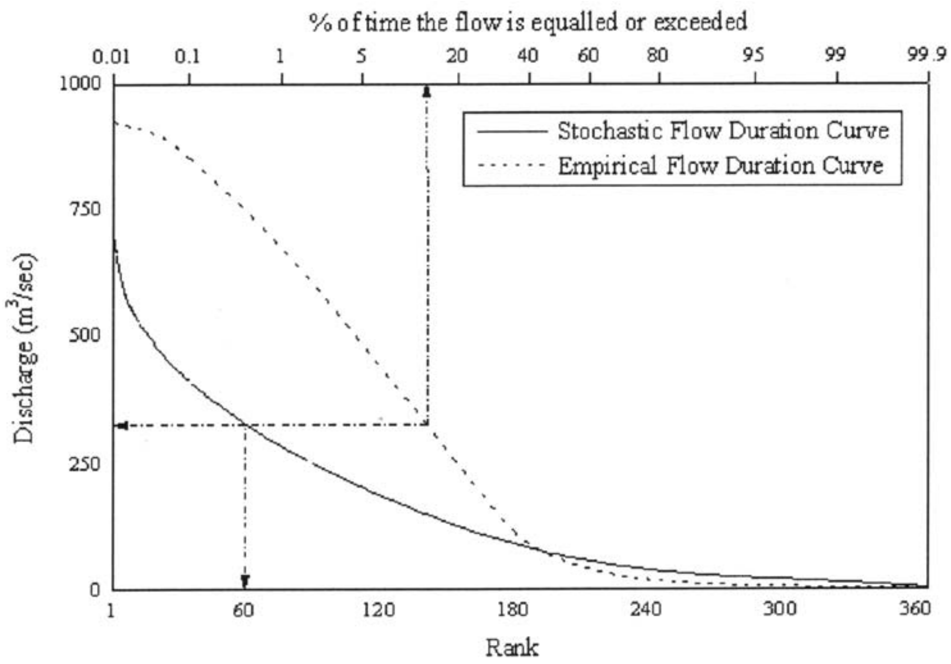

Fig. 4 Comparison of stochastic and empirical flow duration curve for Baral River at Malanchi railway crossing.

Comparison of stochastic and empirical flow duration curve

Empirical flow duration curves for each of four rivers at selected stations were calculated, using available daily discharge data, for comparison with respective stochastic flow duration curves. For this purpose, the available discharge record of each station was divided into equal class intervals and the exceedence probabilities for each class interval were calculated. The empirical flow duration curve for the Baral River is presented in Fig. 4 together with the stochastic flow duration curve. An inspection shows that the flow obtained from the stochastic flow duration curve for a given rank compares reasonably well with the value obtained from the empirical flow duration curve at the corresponding exceedence probability.

CONCLUSIONS

Identification of a parent distribution of daily discharges allows the estimation of rank means and variances for flow duration curves. The inclusion of the estimation of variances by rank adds information about the structure and reliability of the flow duration curve. Both mean and variance by rank may be derived using the form and parameters of the parent distribution instead of by calculating these from the ranked observations of the daily discharges. Performance on the variance can be improved through the use of nonlinear regression.

It can be concluded from this study that the mixture lognormal distribution can be used for deriving stochastic flow duration curves for the four selected rivers. The stochastic model, which is numerically based on large sample approximations of order statistics, duplicates the otherwise empirical construction technique of the calendar year flow duration curve. Moreover, the use of a mixed lognormal distribution permits the representation of a variety of flow regimes. It requires the estimation of parameters of the distribution for obtaining the 365 calculated means for construction of the stochastic flow duration curve. The basic modelling procedure may be used in any hydrological environment, and it can be expected that the parameters of the parent distribution will vary along physiographic and hydrometeorological variations which corresponds to climate and topography.

REFERENCES

Galambos, J. (1984) Order statistics. In: *Handbook of Statistics, Nonparametric Methods* (ed. by P. R. Krishnaiah & P. K. Sen), vol. 4, 359–382. Elsevier, New York, USA.

Gibbons, J. D. (1985) *Non Parametric Statistical Inference.* Marcel Decker, New York, USA.

Greenwood, J. A., Lendwher, J. M., Matalas, N. C. & Wallis, J. R. (1979) Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Water Resour. Res.* **15**(5), 1049–1054.

Islam, K. M. H. (2000) A stochastic model of flow duration curves for selected rivers in northwestern Bangladesh. MSc Engineering Thesis, Department of Water Resources Engineering, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh.

LeBoutillier, D. W. & Waylen, P. R. (1993) A stochastic model of flow duration curves. *Water Resour Res*. **29**(10), 3535–3541.