Vivekanandan, N.; Mathew, F. T.

A Comparative Study of Flood Estimation Using Data Transformation Methods

Proceedings of ninth International Conference on Hydro-Science and Engineering (ICHE 2010), IIT Madras, Chennai, India.
2 - 5 August 2010

A COMPARATIVE STUDY OF FLOOD ESTIMATION USING DATA TRANSFORMATION METHODS

N. Vivekanandan[1] F.T. Mathew[2]

***Abstract*:** Estimation of flood for a desired return period is often a requirement for planning, design and operation of hydraulic structures such as dams, barrages, culverts and urban drainage systems. This paper details a study involving the use of five transformation methods of Square Root (SRT), Cube Root (CRT), Power (BCT), SMEMAX and modified SMEMAX for estimation of flood. Kolmogorov-Smirnov test statistic is employed for checking the applicability of the transformation methods to the recorded annual flood data. Diagnostic tests involving root mean square error, root mean absolute error and coefficient of determination are used for selection of suitable transformation method for flood estimation for the data under study. The methodology is applied to river data relating to Tapi and three of its tributaries. The results of the study show that BCT is more suitable amongst five transformation methods for Tapi at Burhanpur and Purna at Lakhpuri while CRT for Girna at Dapuri and Bori at Malkheda.

***Keywords*:** *Cube root; Diagnostic test; Kolmogorov-Smirnov; Power; SMEMAX; Square root*

INTRODUCTION

Estimation of flood for a desired return period is of utmost importance for the safe and economic design of hydraulic structures such as dams, barrages, road/ railway bridges and culverts. Approaches to flood estimation include deterministic and probabilistic methods. Deterministic approach to flood modelling generally adopts a procedure wherein the river system is modeled by a set of variables, often representing the catchment characteristics, wherein each combination of variable values represents a unique state of condition of the system, with the manipulation of the variable simulating the movement of the system from state to state. Probabilistic approach includes fitting of standard probability distributions to recorded discharge data, and transformation methods wherein the techniques attempt to find a unique distribution in the transformed range regardless of the distribution of the original sample.

Transformation methods attempt to transform the original data set, which has an arbitrary distribution, into a set of data with a known distribution. The advantages of transformations methods are as follows: (i) These methods do not fit a distribution to the sample itself, no assumptions about the underlying distribution are required; (ii) There is no need to attempt to find sometime tenuous theoretical justification upon which many of the distributions are based; and (iii) These methods can be realized if a relatively simple distribution is used in the transformed range.

Number of transformation methods such as Square Root (SRT), Cube Root (CRT), Power (BCT), SMEMAX (SMX) and Modified SMEMAX (MSX) are widely used for fitting series of annual

1 Assistant Research Officer, Central Water and Power Research Station, Pune 411024, Maharashtra, India; E-mail: anandaan@rediffmail.com
2 Joint Director, Central Water and Power Research Station, Pune 411024; E-mail:ftmathew@gmail.com

flood data (AFD) for estimation of flood. Research reports indicate that there is no general agreement for adopting a standard transformation method for a region or a country for flood estimation though different hydrologists have recommended different transformation methods for modelling of flood data (Hameed et al, 1982; Jam and Singh, 1986; Panigrahi et al, 2002; Loganathan et al, 1985).

This paper details a study for estimating flood for different return periods using five transformation methods, with the objective to identify the most suitable transformation for modelling of AFD for Tapi at Burhanpur, Bori at Malkheda, Girna at Dapuri and Purna at Lakhpuri. Kolmogorov-Smirnov (KS) test statistic was employed for checking the applicability of the transformation methods to the recorded AFD. Diagnostic tests involving root mean square error (RMSE), root mean absolute error (RMAE) and coefficient of determination (R^2) were used for selection of suitable transformation method for flood estimation for a particular region.

METHODOLOGY

Transformation methods are used to transform the sample data to a near normal distribution, thus underlying it tractable by bringing into its purview a whole gamut of developments, irrespective of the underlying distribution of the sample data. The methodology adopted in estimating the flood by five transformation methods, KS and diagnostic tests are detailed below:

Square Root and Cube Root Transformations

The parameters of SRT and CRT are determined by maximum likelihood method and used to estimate the flood (Q_T) using back transformations of SRT and CRT and are:

$$Q_T = \sqrt{\overline{Y} + K_T \sigma_y} \text{ and } Q_T = \sqrt[3]{\overline{Y} + K_T \sigma_y} \qquad (1)$$

Here, $\overline{Y}$ and σ_Y are mean and standard deviation of the transformed data series (Islam and Kumar, 2003). K_T is the frequency factor corresponding to C_s=0.0 and probability of exceedance for a given return period (T).

Power Transformation

The generalized form of power transformation is defined as Box-Cox transformation (BCT) and expressed by:

$$Y_i = \begin{cases} \dfrac{Q_i^{\lambda} - 1}{\lambda}, \lambda \neq 0 \\ \text{Log}(Q_i), \lambda = 0 \end{cases} \qquad (2)$$

where λ is a constant of transformation (Kumar and Devi, 1982). The value of λ is determined by trial and error method so that the coefficient of skewness (C_s) of the transformed series becomes nearly zero. The flood estimate for a given return period can be obtained using back transformation as follows:

$$Q_T = \begin{cases} (\lambda Y_T + 1)^{1/\lambda}, \lambda \neq 0 \\ \mathrm{Exp}(Y_T), \lambda = 0 \end{cases} \quad \text{Here, } Y_T = \overline{Y} + K_T \sigma_Y \tag{3}$$

SMEMAX Transformation

According to the SMEMAX (Small, MEdian, MAXimum) transformation, the transformed variate (Y_i) is given as:

$$Y_i = \begin{cases} \dfrac{Q_i - Q_S}{2\mathrm{Cos}A}, Q_i \langle Q_M \\ \dfrac{(Q_M - Q_S) + (Q_i - Q_M)\mathrm{Cot}A}{2\mathrm{Cos}A}, Q_i \rangle Q_M \end{cases} \tag{4}$$

The transformed angle (A) is defined by $A = \arctan((Q_1 - Q_M)/(Q_M - Q_S))$. The flood estimate for a given return period can be obtained using back transformation $Q_T = 2Y_T \mathrm{Cos}A + Q_S$ (Prakash, 1981). Here, Q_S, Q_M and Q_1 are the smallest, median and maximum sample values of the original series of AFD.

Modified SMEMAX Transformation

MSX is the modified version of SMEMAX transformation (Rasheed et al, 1982). According to this method, the transformed variate (Y_i) is given as:

$$Y_i = \begin{cases} Y_M\left(1 - \dfrac{(Q_M - Q_i)}{(Q_M - Q_S)}\right), Q_i \langle Q_M \\ Y_M\left(1 - \dfrac{(Q_M - Q_i)}{(Q_1 - Q_M)}\right), Q_i \rangle Q_M \end{cases} \tag{5}$$

The flood estimate for a given return period can be obtained using back transformation $Q_T = Q_M + (Q_M - Q_S)[(Y_T / Y_M) - 1]$, where $Y_M = ((Q_M - Q_S)/2\mathrm{Cos}A)$.

Model Performance

KS test was used for evaluating the applicability of transformation methods for flood estimation. The test statistic is defined by:

$$K_C = \underset{i=1}{\overset{N}{\mathrm{Max}}}\left(F_e(Q_i) - F_D(Q_i)\right) \tag{6}$$

where $F_e(Q_i) = (i - 0.35/N)$ is the empirical cumulative distribution function of Q_i, $F_D(Q_i)$ is the computed cumulative distribution function of Q_i and N is the number of observations. The rejection region of KS test at the desired significance level 'η' is $K_C > K_{N,1-\eta}$. If the computed values (K_C) of KS test of the transformation is less than that of theoretical value at the desired significance level 'η' then the selected transformation method is accepted to be adequate than any other transformations (Zhang, 2002).

Diagnostic tests, involving RMSE, RMAE and R^2 were used for selection of suitable transformation method for estimation of flood for different return periods (Singh, 1987). A qualitative assessment of the goodness of fit is ascertainable from the probability plot of the observed and corresponding estimated AFD. Theoretical descriptions of diagnostic tests are given by:

$$RMSE = SQRT\left(\frac{1}{N}\sum_{i=1}^{N}\left(Q_I - Q_I^*\right)^2\right) \tag{7}$$

$$RMAE = SQRT\left(\frac{1}{N}\sum_{I=1}^{N}\left|Q_I - Q_I^*\right|\right) \tag{8}$$

$$R^2 = \left(\frac{\sum_{i=1}^{N}\left(Q_i - \overline{Q}\right)\left(Q_i^* - \overline{Q^*}\right)}{\sqrt{\left(\sum_{i=1}^{N}\left(Q_i - \overline{Q}\right)^2\right)\left(\sum_{i=1}^{N}\left(Q_i^* - \overline{Q^*}\right)^2\right)}}\right)^2 \tag{9}$$

where, Q_i is the observed AFD, Q_i^* is the estimated AFD by five transformation methods, $\overline{Q}$ is the series mean of observed AFD and $\overline{Q^*}$ is the series mean of estimated AFD.

APPLICATION

Study Area and Data Used

Tapi river rises near Multai, Betul district of Madhya Pradesh at an elevation of 700 m. The river flows 724 km through Madhya Pradesh, Maharashtra and Gujarat states before falling into the Arabian sea. Bori, Girna and Purna are the tributaries of river Tapi. In the present study, an attempt has been made to estimate the flood using five transformation methods at the selected sites of river Tapi and its tributaries. AFD in respect of river Tapi at Burhanpur, Bori at Malkheda, Girna at Dapuri and Purna at Lakhpuri for the period 1977-2004 have been used. Table 1 gives the summary statistics of recorded AFD.

Table 1. Summary statistics of AFD of river Tapi and its tributaries

River	Tributary	Gauging site	Catchment area (km^2)	Statistical parameters			
				$\overline{Q}$ (10^2 m^3/s)	σ_Q (10^2 m^3/s)	C_s	C_k
Tapi	-	Burhanpur	9,170	80.791	67.318	1.491	1.867
	Bori	Malkheda	1,830	1.281	1.316	1.416	1.257
	Girna	Dapuri	8,901	6.312	8.415	2.499	7.399
	Purna	Lakhpuri	3,560	7.885	5.152	0.417	-1.012

$\overline{Q}$, σ_Q, C_s and C_k are the mean, standard deviation, coefficient of skewness and coefficient of kurtosis of the recorded AFD.

Estimation of Flood Discharge using Transformation Methods

Statistical software, namely, RAINBOW and HYDRO was used to fit the AFD adopting five transformation methods. The software computes the parameters of the transformed data by five transformation methods, and derived flood estimates for different return periods from 2-yr to100-yr. For BCT, transformation constant (λ) of 0.11, 0.06, 0.07 and 0.66 were used for transform the data for Tapi, Bori, Girna and Purna respectively. For SMX and MSX, the transformation angles (A) were considered as 1.373, 1.407, 1.459 and 1.028 and used to transform the data under study. The flood estimates for different return periods obtained using five transformation methods for river Tapi and three of its tributaries at the respective sites is used to develop flood frequency curves and is delineated in Figs. 1-4.

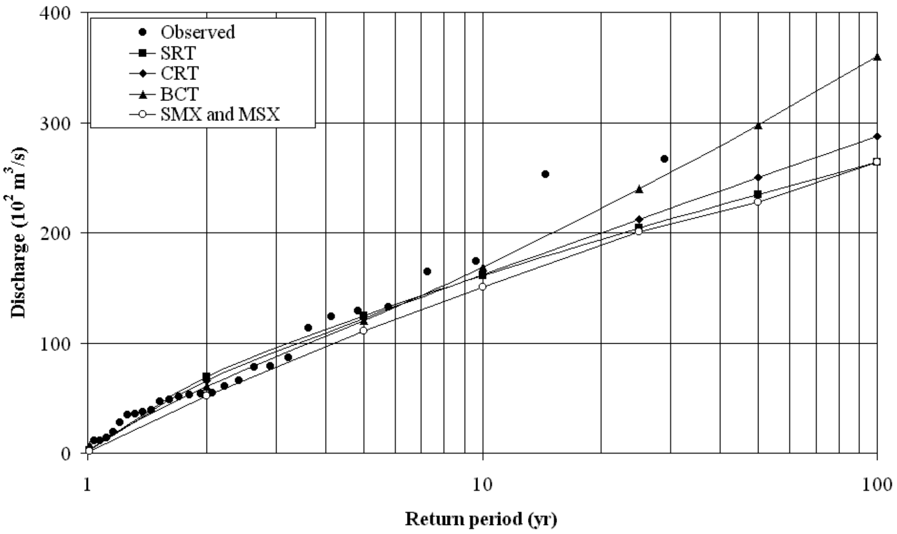

Fig. 1. Probability plot of observed and estimated AFD using five transformation methods for river Tapi at Burhanpur

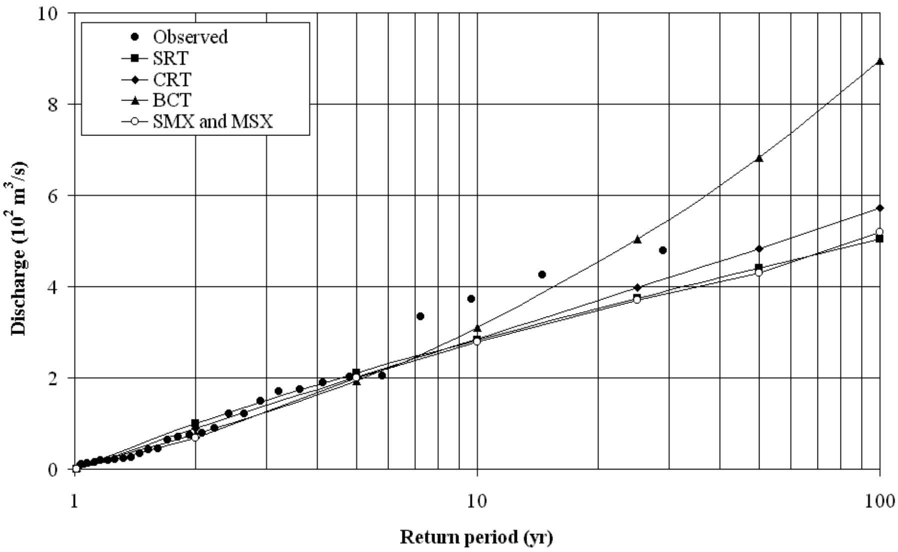

Fig. 2. Probability plot of observed and estimated AFD using five transformation methods for river Bori at Mallkheda

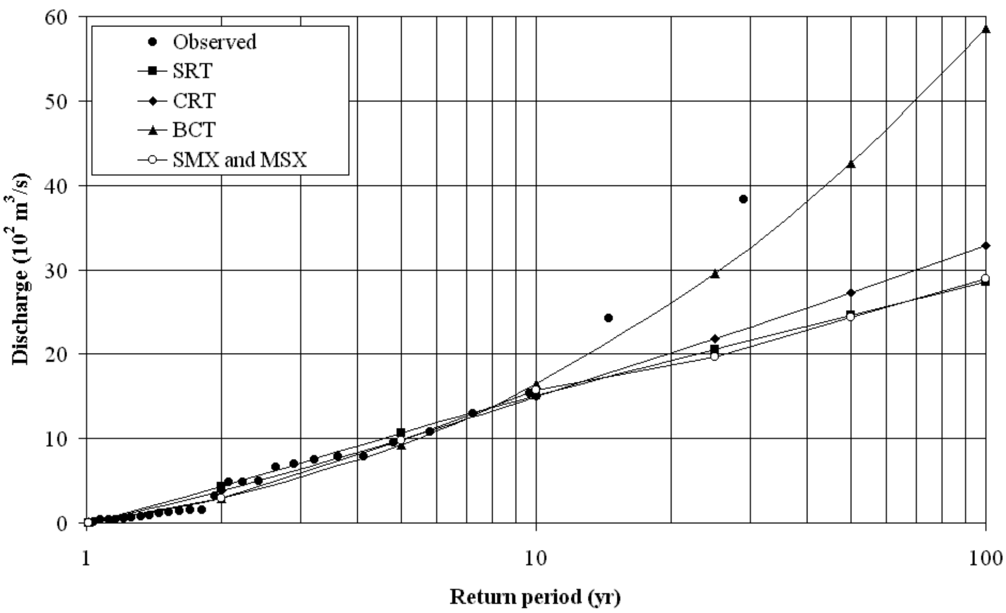

Fig. 3. Probability plot of observed and estimated AFD using five transformation methods for river Girna at Dapuri

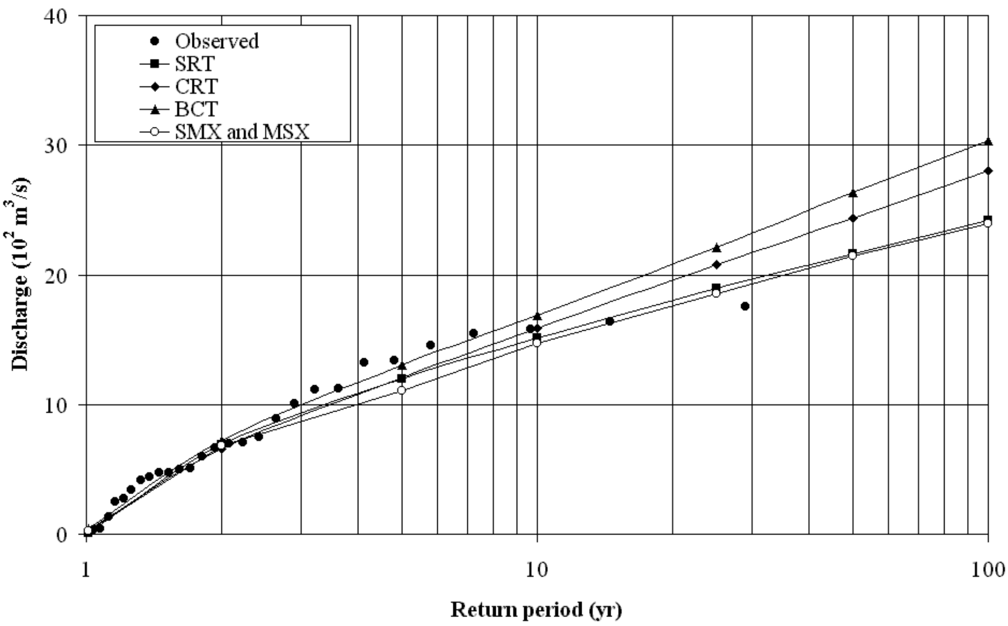

Fig. 4. Probability plot of observed and estimated AFD using five transformation methods for river Purna at Lakhpuri

From Figs.1-4, it can be seen that: (i) The estimated flood using BCT is comparatively higher for return period above 5-yr when compared with the corresponding values of other four transformation methods; (ii) SRT and CRT provides higher estimates for return periods less than 10-yr for Tapi, Bori and Girna at the respective sites; (iii) The fitted curve using BCT shows divergence for Bori at

Malkheda and Girna at Dapuri, and convergence for Tapi at Burhanpur and Purna at Lakhpuri for return period above 10-yr; and (iv) The estimated flood by SMX and MSX is comparatively less than the corresponding values of other transformation methods for all the four sites.

Performance Analysis

KS test statistic of five transformation methods for the data were computed by using Eq. (6) and is given in Table 2.

Table 2. Index values of KS test for five transformation methods

River	Tributary	Gauging site	Indices of K_C for			
			SRT	CRT	BCT	SMX and MSX
Tapi	-	Burhanpur	0.129	0.106	0.139	0.153
	Bori	Malkheda	0.080	0.074	0.112	0.135
	Girna	Dapuri	0.107	0.097	0.123	0.143
	Purna	Lakhpuri	0.174	0.160	0.183	0.157

SRT: Square root; CRT: Cube root; BCT: Power (Box-Cox); MSX: Modified SMEMAX (SMX)

From Table 2, it may be noted that the computed values of KS test for five transformation methods are less than the theoretical value ($K_{29,0.05} = 0.253$) at five percent level. At this level, all five transformation methods considered in the study are acceptable for modelling of AFD for river Tapi and three of its tributaries at the respective sites.

Diagnostic analysis involving RMSE, RMAE and R^2 was used for selection of most suitable transformation method for modelling of AFD, though KS test gave sufficient information on fitting of transformation methods to the data under study. Index values of diagnostic tests for five transformation methods were computed through Eqs. (7-9) and is given in Table 3.

Table 3. Index values of diagnostic tests for five transformation methods

Site	SRT			CRT			BCT			SMX and MSX		
	A	B	C	A	B	C	A	B	C	A	B	C
Burhanpur	6.563	5.735	0.952	6.495	5.684	0.971	5.324	4.823	0.982	6.355	6.108	0.935
Malkheda	3.513	3.277	0.940	3.496	3.260	0.963	5.134	4.685	0.924	5.235	5.020	0.938
Dapuri	3.519	3.013	0.944	3.433	2.895	0.964	5.234	4.342	0.961	6.123	5.296	0.958
Lakhpuri	3.046	2.777	0.903	3.092	2.877	0.950	2.995	2.672	0.961	4.123	3.693	0.935

A: Root mean square error (RMSE); B: Root mean absolute error (RMAE); C: Coefficient of determination (R^2)

From R^2 values, it is observed that there is generally a good correlation between the observed and estimated discharges using five transformation methods for all the four sites; with the values varying from 0.903 to 0.982. From Table 3, it is noted that RMSE and RMAE of BCT are minimum for rivers Tapi and Purna, when compared with the corresponding indices of other transformation methods. Likewise, the indices of RMSE and RMAE of CRT are minimum when compared with other transformation methods for rivers Bori and Girna. From the results of GoF test and diagnostic analysis, it is suggested that BCT could be employed for modelling of flood data for river Tapi at Burhanpur and Purna at Lakhpuri, while CRT is better suited for Bori at Malkheda and Girna at Dapuri.

CONCLUSIONS

The paper details a comparative study of flood estimation using five transformation methods and its applicability to river Tapi and three of its tributaries. The paper shows the BCT is the best suited among five transformation methods for modelling of flood data for Tapi at Burhanpur and Purna at Lakhpuri while CRT for Bori at Malkheda and Girna at Dapuri. The study compares superiority of BCT and CRT over other transformations for the respective river basins through KS and diagnostic tests. The results also show that no uniform transformation method is applicable to fit the AFD recorded at the respective sites of rivers Tapi, Bori, Girna and Purna. The paper gives the results of the study reported herein and the methodology is expected to be of assistance to stakeholders. The probability plots of observed and estimated discharge using five transformation methods for different return periods are developed and presented in the paper.

ACKNOWLEDGEMENT

The authors are grateful to Dr. I.D. Gupta, Director, CWPRS, Pune, for according permission to publish this paper.

REFERENCES

Hameed R. Rasheed, Marur V, Ramamoorthy, A. S., and Al-Dabbagh. 1982. Modified SMEMAX transformation for frequency analysis. *J. American Water Resources Association*, 18(3), 509 – 512.

Islam, A., and Kumar, A. 2003. HYDRO: A program for frequency analysis of rainfall data. *IIE J. Agricultural Engineering*, 84, 1-5.

Jam, D., and Singh V. P. 1986. A comparison of transformation methods for flood frequency analysis. *J. American Water Resources Association*, 22(6), 903 – 910.

Kumar Arun and Devi Rama. 1982. Statistical determination of design low flows – A comment. *J. Hydrology*, 58, 175-177.

Loganathan, G.V., Kuo, C.Y., and McCormick, T.C. 1985. Frequency analysis of low-flows. *Nordic Hydrology*, 16, 105-128.

Panigrahi B, Sudhindra Nath Panda, and Mull R. 2002. Prediction of hydrological events for planning rainfed rice. *J. Hydrological Sciences*, 47(3), 435-448.

Prakash Anand. 1981.Statistical determination of design low flows. *J. Hydrology*, 51, 109-118.

Rasheed Hameed, R., Marur V. Ramamoorthy and Al-Dabbagh, S. 1982. Modified SMEMAX transformation for frequency analysis. *Water Resources Bulletin*, 18(3), 509-511.

Singh, V.P. 1987. On Statistical Intercomparison of EV I Estimators by Monte Carlo Simulation. *Advanced Water Resources Publications*, 10, 87-105.

Zhang, J. 2002. Powerful goodness-of-fit tests based on the likelihood ratio. *J. Royal Statistical Society* Series B, 64(2), 281-294.