AN ASSESSMENT OF PREDICTIVE ACCURACY FOR TWO REGIONAL FLOOD-FREQUENCY ESTIMATION METHODS

Binaya Kumar MISHRA[1], Kaoru TAKARA[2], Yosuke YAMASHIKI[3] and Yasuto TACHIKAWA[4]

[1]Student Member of JSCE, Dr. Eng., Dept. of Urban and Environmental Eng., Kyoto University (Kyoto 615-8540, Japan) E-mail:mishra@flood.dpri.kyoto-u.ac.jp
[2]Fellow of JSCE, Dr. Eng., Professor, DPRI, Kyoto University (Uji 611-0011, Japan)
[3]Member of JSCE, Dr. Eng., Associate Professor, DPRI, Kyoto University (Uji 611-0011, Japan)
[4]Member of JSCE, Dr. Eng., Associate Professor, Dept. of Urban and Environmental Engineering, Kyoto University (Kyoto 615-8540, Japan)

Direct-regression and index-flood methods are the two major types of regional flood-frequency estimation methods. While the former method is well-established for flood-frequency estimation in practice in many countries, the popularity of latter method is limited among the researchers i.e., universality of the latter method has not been established. In this regard, this study has attempted to assess the prediction accuracies in design floods for the two regional flood-frequency estimation methods. The design floods were assessed on 11 example Nepalese river basins using the Jackknife technique. The index-flood method was found to have slightly better prediction accuracies over the direct-regression method.

Key Words: *Direct-regression, frequency, index-flood, Nepalese river basins*

1. INTRODUCTION

Design flood (maximum discharge of a specific return period) estimations are required for various hydraulic works such as design of weir, barrage, dam, irrigation facilities, flood control measures etc. Over/under-estimates of design floods result losses like waste of resources, infrastructural damage, human life and many others. Research in design flood estimation is on the decline and there is a large gap between design flood research and practice[1)]. This needs redress if improvements to design flood estimation practice is to be made.

Several techniques are available for estimating design floods[2)]. The estimation methods can be broadly classified into two groups (**Fig. 1**): rainfall-based methods and streamflow-based methods.

Rainfall-based methods are more scientific and can account easily the changes of climate, landuse, etc. However, the rainfall-based methods are data/skill intensive. On the other hand, streamflow-based methods are relatively simpler. The streamflow-based equations are reliable for the regions with not many flow-control structures. In streamflow based method, no assumption is required regarding the relationship between the probabilities of rainfall and runoff**.**

This study is related with streamflow-based method of design flood estimation. The streamflow-methods are mainly based on the analysis of streamflow data. These methods include empirical equations, and at site or regional statistical analyses. Regional analysis methods may be used to estimate design floods at locations with inadequate streamflow data or no data.

Direct-regression and index-flood methods are the two major approaches of regional flood-frequency analysis. Delineation of hydrologic homogeneous regions is common major step of any regional flood-frequency analysis. Regionalization is performed to transfer the hydrologic characteristics from gauged basins to ungauged basins. In the previous study[3)], Nepalese river basins were grouped into five hydrologic regions (**Fig. 2**).

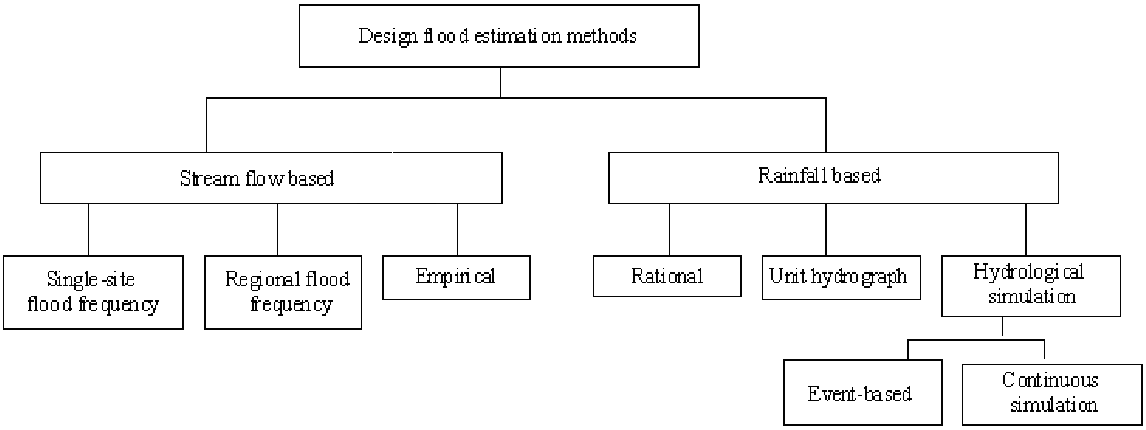

Fig. 1 Methods for estimating design flood.

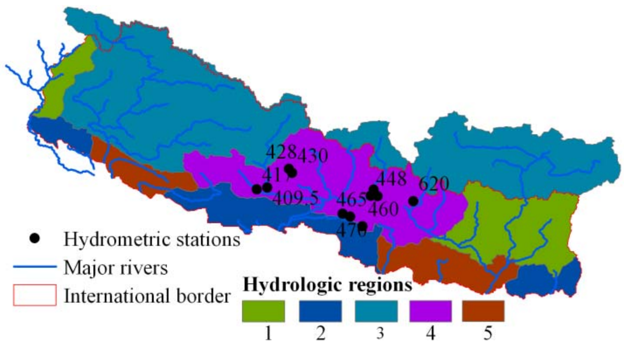

Fig. 2 Map showing the hydrologic homogeneous regions with 11 test basins outlet inside the hydrologic Region 4.

In developing the regional flood-frequency relationships, direct-regression based method has been commonly used in the previous works[4),5)]. Index-flood based method, with the use of L-moments, can result flood predictions as good as or better than those based on the direct-regression method of regional flood-frequency analysis[6),7),8)]. The index flood is expected to have better predictive because the index-flood method provides an appropriate procedure for statistical flood estimation of extreme events and also better represents the basin characteristics. Consequently, the index-flood based regional flood-frequency relationships were developed for the Nepalese using the flood data of 49 Nepalese river basins[9)].

In the previous work, the distributions: GEV, lognormal and Pearson type III were found to be reasonably fitting in all of the hydrologic regions. The drainage area was found to be mainly governing the value of index flood[10)]. The index-flood based regional flood-frequency relationships were found to have far-better predictive accuracy over the WECS (Water and Energy Commission Secretariat, Nepal) method[4)]. The WECS method is frequently used for estimation of return period floods in ungauged basins of Nepal and have been developed using the direct-regression method. However, the WECS method has considered all the Nepalese river basins as of one hydrologic region.

In the present study, the direct-regression method was used for deriving regional flood-frequency relationships in an example hydrologic Region 4. The hydrologic Region 4 was selected since this region has enough hydrometric (discharge) stations, and hence better regression relationships can be derived. The study investigated the predictive accuracies of the direct-regression and index-flood based regional flood-frequency equations. While investigating the predictive accuracies, the design floods predicted by the two regional equations were compared with that of at-site flood estimates (true estimates). The design flood predictive accuracies were tested at 11 river basins of Region 4. The predictive accuracies were assessed in term of mean and median errors in flood estimates.

2. REGIONAL FLOOD-FREQUENCY ANALYSIS

Regional flood-frequency analysis is an important method for estimating flood peaks within specified probabilities of exceedance at ungauged sites or enhancing estimation at gauged sites where historical records are short. It is a means of transferring flood-frequency information from gauged basins to ungauged basins on the basis of similarity in basin characteristics. Regional relationships can also mitigate the effect of outliers and can lead to more reliable extrapolation[2)]. Direct-regression and index-flood methods are the two major types of regional flood frequency analysis. A brief description on two major methods is given in the following sub-sections.

(1) Direct-regression method

In this method, the regression models may be used in the following form as **Eq. (1)**:

$$Q_T = aX_1^b X_2^c X_3^d \ldots \quad (1)$$

where Q_T is the peak discharge for the T-year return period, X_1, X_2, X_3…are physiographic/climatic characteristic variables of the gauged basins, and a, b, c, d... are regression parameters. Since it is unknown in advance what physiographic/climatic characteristics may have significant impact on flood-frequency estimates, a number of parameters are calculated and investigated as possible predictors of T-year discharges. Then the variables with the smallest significance are removed until the statistically significant terms remain. In the present study, use of drainage area was limited as regression variable since the drainage area was only found to be effective in governing the flood in the previous study[10)].

(2) Index-flood method

The index-flood method[8),11)] assumes that, within a hydrologic homogeneous region, the exceedance probability distribution of annual peak discharge is identical except for a site-specific scaling factor called the index flood. This index flood parameter reflects the important physiographic and meteorological characteristics of a basin. In this method, a relationship is established for estimating the flood quantile Q_T of return period T at site i as the product of index flood (average likely flood) μ_i, which is the function basin area, slope etc., and regional growth factor, q_T. The growth factor is a dimensionless frequency distribution quantity common to all sites within a hydrologic homogeneous region. The design flood estimation relationship may be expressed by **Eq. (2).**

$$Q_T = q_T \mu_i \quad (2)$$

For the estimation of index flood, a relationship in term of basin characteristics is established based on available information gathered from the gauged sites. Regional growth curves showing the relationship between q_T and T are derived once an appropriate frequency distribution has been found within a hydrologic region with N sites that fits all the gauged flood series.

In simple words, index flood based regional flood-frequency analysis method can be said of three major steps: hydrologic homogeneous regionalization, selection of regional frequency distribution and estimation of index flood relationship.

Unlike the direct-regression based regional flood- frequency analysis which is derived for a fixed values of return periods (e.g. T = 2, 5, 10, 20, 50, 100 ……years), the index flood-based regional relationships can be used for estimating design flood of any intermediate values of return periods.

3. METHODOLOGY

Evaluation of the developed regional flood-frequency relationships is an important aspect. The accuracy of two regional flood-frequency relationships has been assessed on 11 independent gauged stations (**Fig. 2**). All these test basins have their catchment boundary inside the hydrologic Region 4. The basins which possessed at-least 20 years of observation flood data series were selected for the assessment since reliable at-site estimates (assumed as true estimates here) can be expected only for the stations having longer observation flood series.

The regional flood-frequency relationships have been tested by comparing the return period floods of regional flood-frequency equations with that of at-site flood-frequency analysis method. Jackknife technique was employed for assessing the design flood estimates at each of the test stations. In this technique the station, at which assessment is to be performed, is excluded in deriving the regional flood-frequency relationships.

For illustration, let us consider the test station 409.5 of Region 4 for evaluating the predictive accuracy. The drainage area of this basin is 113.51 km^2. Out of total 24 stations in the region, the station 409.5 was excluded and the remaining 23 stations were considered for deriving the regional flood-frequency equations for the two methods. A brief detail on the development of flood-frequency relationships for evaluating the design flood at this station is presented below.

Index-flood method

Design flood, Q_T (m^3/s) of T year return period in basin i is (**Eq. (3)**):

$$Q_T = 6.23 A_i^{0.68}\left[0.726 - 2.73\left\{1-(-\ln\frac{T-1}{T})^{-0.137}\right\}\right] \quad (3)$$

where A_i is drainage area in km^2.

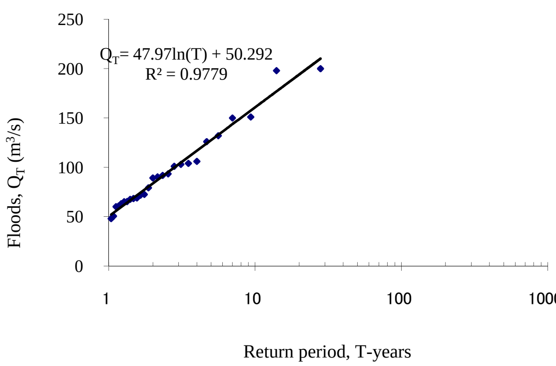

Fig. 3 Plot of floods against their return periods at station 409.5.

At-site method

The annual maximum flood series were arranged in descending order. Return periods were computed for the ordered values using the Weibul's plotting position formula (**Eq. (4)**):

$$T = \frac{n+1}{m} \tag{4}$$

where n is sample size and m is rank of the floods.

$$Q_T = 47.97\log(T) + 50.29 \tag{5}$$

The at-site flood-frequency equation (**Eq. (5)**) corresponds to the line fitting the plotted points of the floods versus return periods (**Fig. 3**).

Direct-regression method

Firstly, maximum discharges of 2, 5, 10, 20, 50, 100 and 200-years return period were computed using the method of at-site flood-frequency at the remaining 23 stations in the region. As mentioned earlier that the catchment area mainly govern the flood values in the delineated hydrologic homogeneous regions, a simple regression technique was applied for each of the return period floods as dependent variable and the drainage area as independent variable (**Fig. 4**). The general form of regional flood-frequency estimation relationship may be expressed by **Eq. (6)**:

$$Q_T = aA_i^{\,b} \tag{6}$$

where a and b are regression parameters (**Table 1**).

The similar process was repeated at each of the test stations. Assessment of the two regional flood-frequency methods has been made in terms of mean and median absolute error in design flood estimates of different return periods (**Eqs. (7-8)**):

$$\Delta Q_T^{IF} = \left|\frac{Q_T^{IF} - Q_T^{Atsite}}{Q_T^{Atsite}}\right| \tag{7}$$

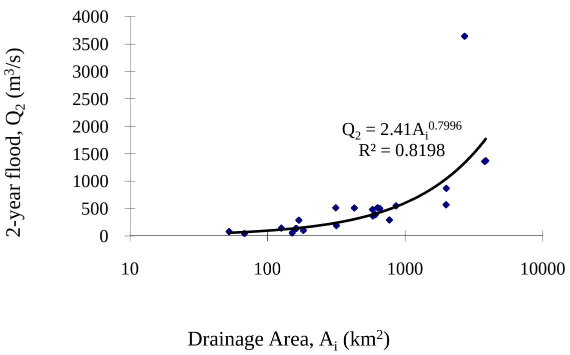

Fig. 4 Illustrative regression plot of 2-year flood against corresponding drainage area of 23 river basins.

Table 1 Values of regression parameters for T-years

T	a	b	R^2
2	2.41	0.80	0.81
5	5.09	0.76	0.75
10	7.19	0.74	0.72
20	9.31	0.73	0.70
50	12.15	0.73	0.68
100	14.30	0.72	0.67
200	16.47	0.72	0.66

$$\Delta Q_T^{DR} = \left|\frac{Q_T^{DR} - Q_T^{Atsite}}{Q_T^{Atsite}}\right| \tag{8}$$

where, ΔQ_T^{IF} is relative absolute error in index-flood based estimates for T-year return period;

ΔQ_T^{DR} is relative absolute error in direct-regression based estimates for T-year return period;

Q_T^{IF} is index-flood based estimates of T-year return period;

Q_T^{DR} is direct-regression based estimates of T-year return period; and

Q_T^{Atsite} is at-site flood-frequency analysis estimates for T-year return period.

4. RESULTS AND DISCUSSION

The assessment work started with the estimation of return period floods for the at-site flood-frequency analysis, direct-regression and index-flood methods. Considering the estimates of at-site flood-frequency analysis method as true estimates, the error in direct-regression and index-flood estimates were projected.

Figs. 5-11 show the comparative plot of flood estimates for 2, 5, 10, 20, 50, 100 and 200-years return periods respectively obtained using the at-site, direct-regression and index-flood methods. In these figures, some stations are found to have larger predictive discrepancies. These stations are situated at the boundary of the hydrologic regions, hence may be influenced by other region. This may result bigger discrepancies in the estimated values. From these plots, it is difficult to distinguish the predictive superiority of either method over another. The predicted floods seem closely similar at most of the stations for both direct-regression and index-flood methods.

To identify which regional method is better, relative absolute error in the estimates of direct-regression and index-flood methods were evaluated by considering the at-site flood estimates as true estimates. Using the **Eqs. (7-8)**, relative absolute error at each of the test stations were computed for both the regional methods.

The maximum absolute percentage error between the at-site flood-frequency analysis estimates and the index-flood based regional estimates at any stations was found to be 72.72%. In contrast, the maximum absolute percentage error in direct-regression regional estimates was found to be 68.58%.

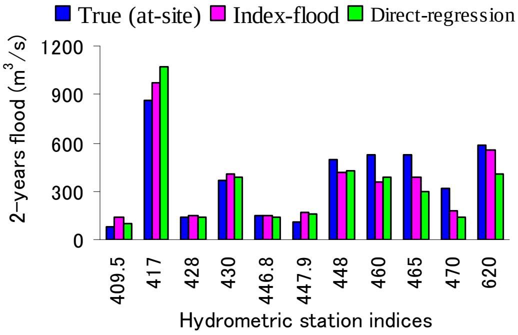

Fig. 5 Comparison of design floods for T = 2 years

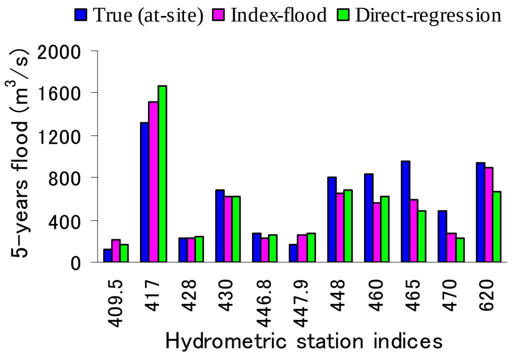

Fig. 6 Comparison of design floods for T = 5 years

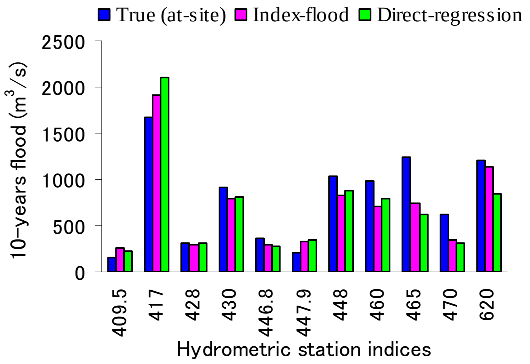

Fig. 7 Comparison of design floods for T = 10 years

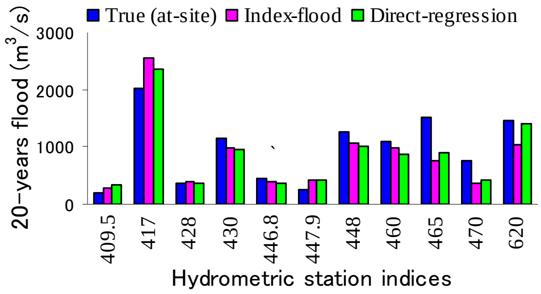

Fig. 8 Comparison of design floods for T = 20 years

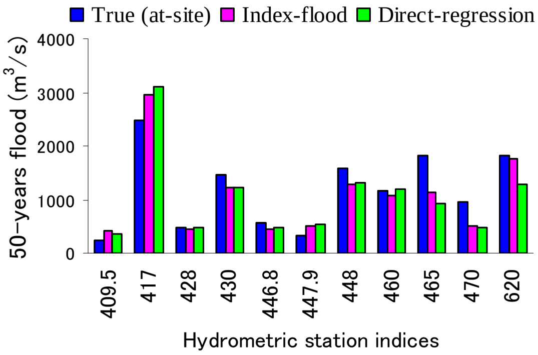

Fig. 9 Comparison of design floods for T = 50 years

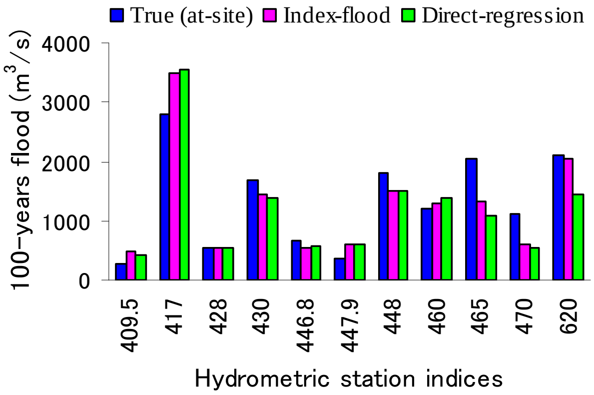

Fig. 10 Comparison of design floods for T = 100 years

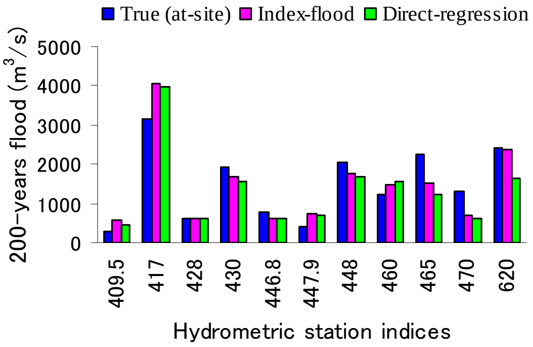

Fig. 11 Comparison of design floods for T = 200 years

Table 2 Average absolute mean and median error in design flood estimates.

Return period, T years	Mean absolute error (%)		Median absolute error (%)	
	Index	Direct	Index	Direct
2	24.75	24.78	16.15	23.73
5	26.90	27.49	19.19	26.02
10	28.14	30.15	20.32	26.28
20	30.80	29.43	20.89	26.37
50	28.25	29.81	20.44	26.39
100	29.06	31.45	29.06	31.45
200	31.28	33.06	22.55	27.47
Average	28.46	29.45	21.23	26.81

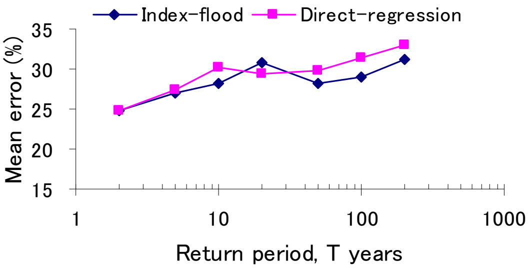

Fig.12 Plot of mean error for the direct-regression and index-flood methods.

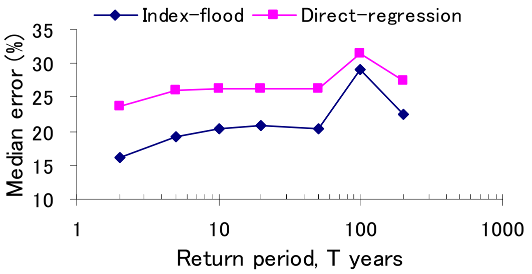

Fig. 13 Plot of median error for the direct-regression and index-flood methods.

Mean and median absolute error was used to show the trend in error for the two methods. **Table 2** shows the average mean and median absolute error (%) for the two regional methods. The respective graphical plot has been made in **Figs. 12-13**. These tables and figures show that the absolute percentage error in index-flood method is relatively smaller than that of direct-regression method.

5. CONCLUSIONS

The overall objective of this study was to assess the methods of index-flood and direct-regression based regional flood-frequency estimation techniques for better estimation of return period floods. The index flood-based regional flood frequency method was expected to have better predictive accuracies than the direct-regression method because the index-flood method provides an appropriate procedure for statistical flood estimation of extreme events and better represents the local characteristics. The objective was achieved, at first, by deriving the direct-regression based regional flood-frequency estimation relationships in one of the hydrologic homogeneous regions of Nepalese river basins and then comparing the estimated return period floods of direct-regression and index-flood methods with that of at-site method.

The plot of predicted floods for different return periods at the 11 test basins do not point out any clear-cut advantage/disadvantage of either regional flood-frequency methods. Comparative analysis on flood estimates in term of mean and median error for the index-flood and direct-regression methods point out that the index-flood method has slightly better predictive accuracy over the direct-regression method. These lead to conclude that index-flood based regional flood-frequency estimation method is better than the direct-regression based regional flood-frequency estimation method.

As the assessment of flood prediction accuracy is limited in only one hydrologic region at 11 test basins, the degree of assessment may not be considered well-enough. Therefore, the study recommends performing the flood predictive accuracy assessment in additional hydrologic regions to give more reliable conclusion.

REFERENCES

1) Chetty, K. and Smithers, J.: Continuous simulation modelling for design flood estimation in South Africa: Preliminary investigations in the Thukela catchment, Physics and Chemistry of the Earth, Vol. 30, pp. 634-638, 2005.
2) Charalambus, J.: Application of Monte Carlo simulation technique with URBS runoff-routing model for design flood estimation in large catchments, Masters of Engineering (Hons) Thesis, University of Western Sydney, Chapter 2 (http://arrow.uws.edu.au:8080/vital/access/manager/Repository/uws:769), 2004.
3) Mishra B.K., Takara K and Tachikawa Y.: Integrating the NRCS-runoff curve number in delineation of hydrologic homogeneous regions, Journal of hydrologic engineering, Vol. 14(10), pp. 1091-1097, 2009.
4) Sharma, K.P. and Adhikari, N.R.: Hydrological estimations in Nepal, Department of Hydrology and Meteorology, Kathmandu, 2004.
5) Perica, S. and Stayner, M.: Regional flood frequency analysis for selected basins in Utah, University of Utah, Report No. UT-4.12, 2004.
6) Rao, A.R.: Flood frequency relationships for Indiana, Purdue University, Report No. FHWA/IN/JTRP-2005/18 (http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1746&context=jtrp), 2006.
7) Rao, A.R. and Hamed, K.H.: Flood Frequency Analysis, CRC press LLC, Florida, USA, 2000.
8) Hosking, J.R.M. and Wallis, J.R.: Regional Frequency Analysis: An approach based on L-moments, Cambridge University Press, Cambridge, UK, 1997.
9) Mishra, B.K., Takara, K., Yamashiki, Y. and Tachikawa Y.: Hydrologic simulation aided regional flood frequency analysis of Nepalese river basins, Journal of flood risk management, DOI: 10.1111/j.1753-318X.2009.01041.x.
10) Mishra, B.K., Takara, K., Yamashiki, Y. and Tachikawa, Y.: Estimation of index flood in hydrologic regions with limited flood data availability, Annual Journal of Hydraulic Engineering, JSCE, Vol. 53, pp. 50-60, 2009.
11) Dalrymple, T.: Flood frequency analysis, US Geol. Survey Water Supply Paper 1543A, pp. 11-51, 1960.

(Received September 30, 2009)