Annual Journal of Hydraulic Engineering, JSCE, Vol.52, 2008, February

# REGIONALIZATION OF NEPALESE RIVER BASINS FOR FLOOD FREQUENCY ANALYSIS

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

[1]Student Member of JSCE, Graduate student, 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, Dept. of Urban and Environmental Eng., Kyoto University
(Kyoto 615-8540, Japan)

Regional flood frequency analysis which is applicable for estimating design flood at both gauged and ungaged sites needs grouping of river basins into hydrological homogeneous regions. Specific flood, timing of flood events and spatial location are significant in reflecting physiogrpahy and climate of any river basins. These three physiographic/climatic characteristics of 46 independent Nepalese river basins have been used successfully to delineate the homogeneous regions. This study uses a clustering algorithm as a starting point for partitioning the independent catchments. The regions thus formed are subsequently revised to improve the regional homogeneity after performing the L-moment based homogeneity test. Spatial proximity consideration has been used to delineate region for mountainous dependent river basins whereas subjective considerations have been made to delineate region for river basins of alluvial plain region.

**Key Words**: *Regionalization, specific flood, clustering, L-moment, homogeneity test*

## 1. INTRODUCTION

Reliable estimation of design flood is required in design of hydraulic structures like reservoirs, levees, culverts, bridges, drainage works, irrigation diversion works *etc.* Regional flood frequency analysis is an effective method for estimating design flood. It can be used at both gauged and ungauged sites. Among the various methods of regional flood frequency analysis[1)], index flood method[2)] is the most popular. The index flood procedure includes three major steps: regionalization of river basins, estimation of regional distribution function and estimation of index flood. Regionalization refers to grouping of basins in hydrological homogeneous regions. Homogeneity implies that regions have similar flood generating mechanisms. A more specific definition of a homogeneous region is the region that consists of sites having identical frequency distribution apart from a site-specific scale factor. Regionalization is required to affect the spatial transfer of hydrologic information. For sites where flood data are not available, the analysis is based on regional data. For sites with available data, the joint use of measured data at a site, called at-site data, and regional data from a number of stations in a region provides sufficient information to enable a probability distribution to be used with greater reliability.

There are no specific guidelines for identifying homogeneous regions. This is due to the complexity in understanding precisely the factors that have effect on the generation of floods. Several attempts have been made by different authors to identify hydrologically homogeneous regions in different parts of the world based on physiographic characteristics or on flood data characteristics or a combination of both. Burn and Goel[3)] used three physiographic characteristics: catchment area, length of the main stream, and slope of the main stream of central Indian River basins. Regionalization using basin characteristics such as catchment area, rainfall and soil type was suggested as an attractive alternative to geographical regionalization by Wiltshire[4)]. Mosley[5)] used clustering of flood statistics in terms of specific flood (peak discharge generated by unit drainage area) and coefficient of variance ($C_v$) for delimitation of New Zealand river basins. Burn[6)] used seasonality statistics (average date of flood events) to avoid the dual use (region forming and homogeneity test) of statistics derived from flood magnitudes. Mkhandi & Kachroo[7] used mapping technique by superimposing the digitized information on catchment boundaries, topography map, mean annual rainfall map, flood statistics and location of gauging stations. Research on regionalization of Nepalese rivers is extremely limited. Water and Energy Commission Secretariat (WECS), Nepal developed relationships as WECS method[8] for estimating low, mean and peak flow duration statistics in ungauged river basins, considering Nepal as one homogeneous region. Medium Irrigation Project (MIP) design manuals[8] have classified the Nepalese basins into 7 regions for estimating mean and 80% reliable flow. David *et al.*[9] has studied flow regimes across the Nepalese Himalaya by classifying long-term average monthly runoff data for 28 independent river basins.

Flood generation is governed by many physiographic and climatic factors. Data for a sufficient number of physiographic characteristics are not always available, particularly of remote areas or developing countries. As a result, the groupings formed using the available physiographic variables may not be hydrologically homogeneous. The resulting regions are often hydrologically homogeneous when flood statistics are used as the variables in the similarity measure. However with this approach, the integrity of homogeneity test is compromised[10] as flood statistics are used both to form the regions and to subsequently evaluate the homogeneity. Also, regionalization is not possible for ungauged river basins by the mere use of flood statistics as similarity measure variables. Hence, delineation of homogeneous regions based on limited factors discussed above may be inadequate. Specific flood indirectly reflects all the physiographic and climatic factors governing the generation of flood. In addition to specific flood, the factors like date of flood events as seasonality measure and basins' centroid lat-long as spatial proximity will be helpful in better grouping of the river basins. Spatial proximity also acts as significant similarity measure variable for grouping of ungauged river basins. Unlike the WECS method of flood frequency analysis which considered the whole Nepal as one homogeneous region, it will be wise to divide whole Nepal into sub-regions for better prediction of flood because there is a vast physiographic and climatic difference among the river basins of Nepal.

The present work addresses the regionalization of the Nepalese river basins for flood frequency analysis that can be used for estimating design flood. The approach uses, as a starting point, a clustering technique that employs specific flood, average date of flood events and spatial location of the river basins. An overview has been provided on study area and data screening process, followed by presentation of proposed region forming processes. The L-moment-based regional homogeneity test has been described in the successive section to test the validity of initially formed regions. Results obtained after application of the proposed method herein and discussions on those obtained results have been presented simultaneously in the successive section. The final section summarizes the important research conclusions.

## 2. STUDY AREA AND DATA SCREENING

Nepal, roughly a rectangle in shape with an area of 147,181 square kilometer, is situated between China in north and India in remaining three sides. It has a length of 885 km east–west and width of 145–248 km north–south. Within this relatively small latitudinal extent, altitude rises from 60 m in south to the 8848 m (Mount Everest, the world's highest peak) in North. The country can be divided into five physiographic regions (**Fig. 1**) namely the Terai (Plain), Siwalik Hills, Middle Mountains, High Mountains and High Himalayas. The average annual precipitation is around 1600 mm of which almost 80% occurs during the period of June-September.

Nepal's hydrometric network comprises 170 gauging stations, of which 54 have been established since 1999[9]. Annual instantaneous maximum flow values of 46 independent watersheds have been used to carry out the present study. These basins were selected on the basis of availability of high quality data and to provide reasonable spatial coverage. These all river basins have a record length of at least 8 years to a maximum of 34 years. These flood data were collected from the hydrological data books published in 1998, 2003 and 2004 by Department of Hydrology and Meteorology, Nepal.

Data screening was started by visual inspection of the annual instantaneous maximum discharge of each gauge station. These maximum discharge data series being independent and homogeneous are the basic assumptions. Test for independence and homogeneity (assumption that the whole dataset come from same distribution) of the data was performed as discussed by Mann and Whitney[11]. Presence of outliers in the data causes difficulties when fitting a distribution to the data (Rao and Hamed[12]). Test for outlier was performed according to Grubbs[13] and Grubbs and Beck[14] and checked at 5% significance level. Out of total 46 gauge stations, there were 5 outliers at one station (station index 610), 2 outliers at seven stations and 1 outlier at 15 stations.

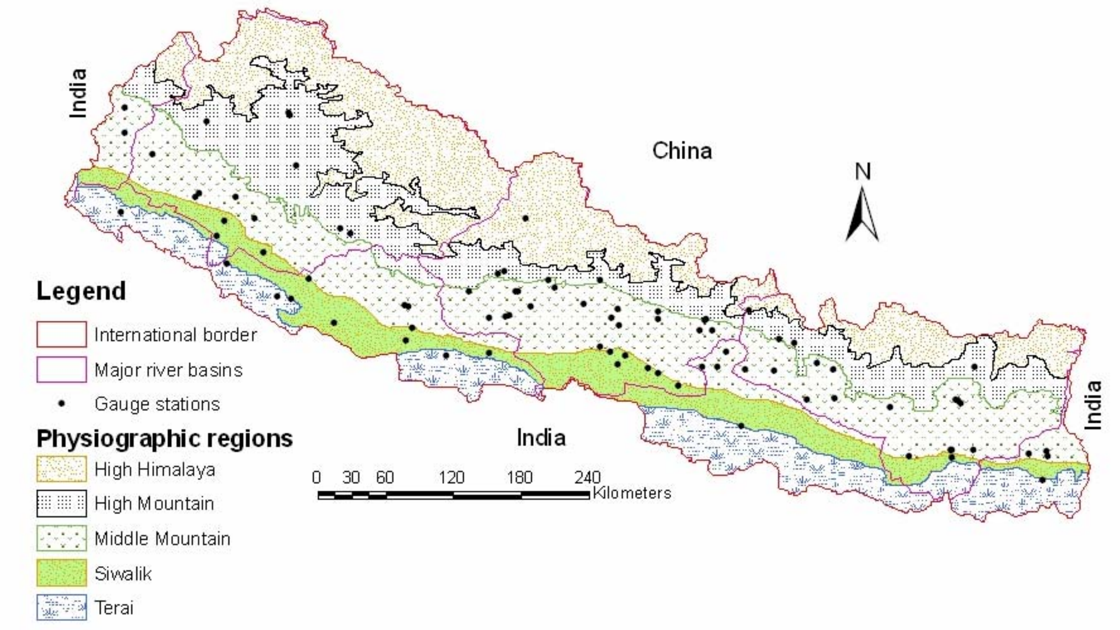

**Fig.1** Physiographic map of Nepal

## 3. PROPOSED APPROACH

This work employs a clustering technique as the starting point for the formation of regions that are homogeneous, identifiable, and sufficiently large. Clusters (groups) of catchments are initially formed using hierarchical, agglomerative cluster analysis according to their values of;

- Specific flood,
- Average date of flood occurrence, and
- Spatial location

The initially delineated groups are then modified, if necessary, to improve the homogeneity.

### (1) Specific flood

Specific flood, defined as peak discharges per unit drainage area, indirectly reflects the physiographic characteristics like soil type, average slope, drainage density, shape of the basins. In addition, it represents the intensities and amount of storm occurred to generate those series of floods. The magnitudes of specific flood are affected with the size of the basins. To reduce the effect of catchment area on the specific flood values, all these values were standardized to catchment area of 500 km$^2$ obtained after analyzing the plot of specific flood against the respective drainage area. The standardized specific flood was calculated as:

$$Q_{sp} = Q_{asp} A^{0.245} 500^{-0.245} \qquad (1)$$

where $Q_{sp}$ is the standardized specific flood and $Q_{asp}$ is the actual specific flood of a catchment of area $A$ (km$^2$). $Q_{asp}$ is obtained by dividing the observed average annual flood with the corresponding drainage area. The exponent 0.245 was found by authors using regression analysis between actual specific floods ($Q_{asp}$) and catchment area (A) for the 89 Nepalese catchments (**Fig. 2**).

### (2) Average date of flood occurrence

Timing of flood events was used in view of climatic similarity among the 46 independent river basins. It has been expressed in terms of average occurrence date **(Fig. 4)** of the annual instantaneous flood series by considering a flood cycle starting from mid-June to mid-September. The total number of days in one cycle ($2\pi$ radian) consists of 92 days by considering mid-June as 1st day and mid-September as 92nd day. Average date of flood ($\theta_f$) in angular unit radian was obtained after converting flood cycle dates ($d_f$) similar to Burn[6] and expressed as:

$$\theta_f = d_f 2\pi / 92 \qquad (2)$$

### (3) Spatial location

Merz and Bloschl[15] examined the predictive performance of various regionalization methods and concluded that the methods that only use catchment attributes perform significantly poorer than the methods based on spatial proximity. In addition, they concluded that the method that combines spatial proximity and catchment attributes yields the best predictive performance. Hence, in addition to specific flood and average date of flood, the latitudinal and longitudinal coordinates of river basins' centroid were used for spatial proximity consideration.

Clustering is a standard method of statistical multivariate analysis for dividing a data set into groups. A data vector is associated with each site, and sites are partitioned or aggregated into groups according to the similarity of their data vectors. The present study employs Ward's hierarchical agglomerative clustering method[16)] to form the regions according to basins' specific flood, average flood date and centroid lat-long values. In this method, initially there are $N$ clusters, each containing one basin's attributes, labeled 1 through $N$. At each stage of clustering, two clusters are merged. In the Ward's procedure, the clusters that are merged are chosen to minimize the within-cluster sum of squared deviations of each attribute about the cluster mean.

## 4. REGIONAL HOMOGENEITY TEST

Regional homogeneity can be evaluated using L-moments based homogeneity test[10)]. Basically, there are two ways to check regional homogeneity. The first one is discordancy measure, intended to identify those sites that are unusual with the group as a whole. The discordancy measure estimates how far a given site is from the centre of the group. If $u_i=[t^{(i)}\ \ t_3^{(i)}\ \ t_4^{(i)}]^T$ is a L-moment ratios ($LC_v$, $LC_s$ & $LC_k$) vector for the site $i$ and superscript $T$ as transpose of the vector, then discordancy ($D_i$) can be expressed as Eq. (3a).

$$D_i = \frac{1}{3}(u_i - \bar{u})^T A^{-1}(u_i - \bar{u}) \qquad (3a)$$

where $\bar{u}$ and $A$ for $N$ gauge locations in a region are defined as;

$$\bar{u} = \frac{1}{N}\sum_{i=1}^{N} u_i \qquad (3b)$$

$$A = \sum_{i}^{N} (u_i - \bar{u})(u_i - \bar{u})^T \qquad (3c)$$

A site ($i$) is declared to be discordant if $D_i$ is large[10)].

The second way of homogeneity test is heterogeneity measure, intended to estimate the degree of heterogeneity in a group of sites and to assess whether they might be reasonably treated as homogeneous. This test compares the variability of L-moment ratios for the catchments in a region with the expected variability, obtained from simulation, with the same record lengths as those in the region. The heterogeneity measure is defined as Eq. (4).

$$H = \frac{(V - \mu_v)}{\sigma_v} \qquad (4)$$

where $V$ is the weighted variance of the $LC_v$ or $LC_s$ or $LC_k$ for the region, $\mu_v$ is mean and $\sigma_v$ standard deviation of $V$ obtained after simulation. Heterogeneity measure is symbolized as $H_1$, $H_2$ and $H_3$ when considered $LC_v$, $LC_s$ and $LC_k$ respectively for calculating $V$. Heterogeneity measure based on $LC_v$ is suggested as the most reliable parameter for testing regional homogeneity. A region is homogeneous if $H<1$, possibly heterogeneous if $1 < H < 2$, and definitely heterogeneous if $H > 2$.

## 5. RESULTS AND DISCUSSION

Specific flood was calculated for average of annual instantaneous flood data of the available time series. **Fig. 2** shows a plot of average specific floods against corresponding drainage area. The plot suggests that the points are highly scattered as the specific flood is governed by many physiographic/climatic factors other than basin area and hence, there is poor correlation between specific flood and drainage area. Roughly it can be seen that specific flood is of high value in the range of drainage area 100-1000 km$^2$ and then starts decreasing with further increase in drainage area.

The authors carried out 15 point moving average in order to offset the influence of physiographic and climatic factors other than drainage area and to see the effect of drainage area more clearly. **Fig. 3** shows that the specific flood increases with drainage area up to a value of about 500 km$^2$ (break point) and then starts decreasing with the increase in drainage area. Hence, to avoid the effect of drainage area on specific flood, all the actual average specific flood values were standardized to a drainage area of 500 km$^2$ for carrying out regionalization. The values of standardized specific flood varied from a minimum 0.198 m$^3$/s/km$^2$ to a maximum 2.15 m$^3$/s/km$^2$.

Almost all the flood occurs in monsoon period which starts in June and ends in September. The mean date of occurrence of annual instantaneous flood was calculated for all the 46 independent river basins. All the mean dates were converted in radian unit by considering flood cycle of 92 days (equivalent to $2\pi$ radian) where June 16 is start of flood cycle and September 15 is end of flood cycle. On the inspection of mean flood date expressed in angular unit radian, it is found varying from 2.684 to 4.175. In other words, it can be said that average annual flood occurs between the fourth week of July and the second week of August. Considering the spatial location of river basins against average flood date values, timing of flood is found relatively earlier in the eastern part of Nepal (**Fig. 4**).

The difficult aspect of clustering is deciding suitable weight to the attributes and number of cluster. In this study, unit weight was applied to specific flood, average flood date and latitude values. Longitude was given double weight to that of latitude to make the very small basins not spreading over two regions. Starting from the number of clusters as two, while performing the clustering, all the 46 independent catchments were finally divided into five clusters (**Fig. 5**) until reasonable value of homogeneity was achieved. Ungauged and dependent river basins are shown as blank area.

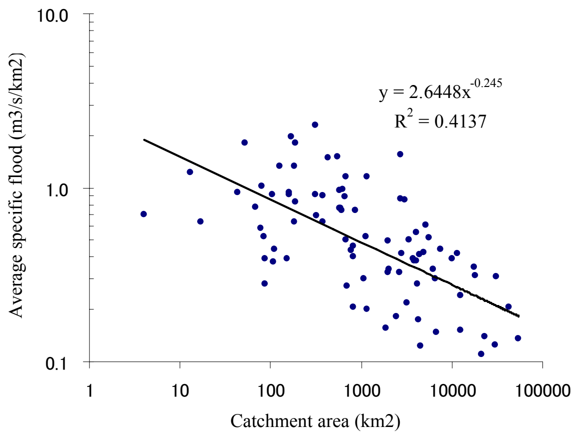

**Fig. 2** Average specific flood vs. drainage area plot for 89 Nepalese river basins.

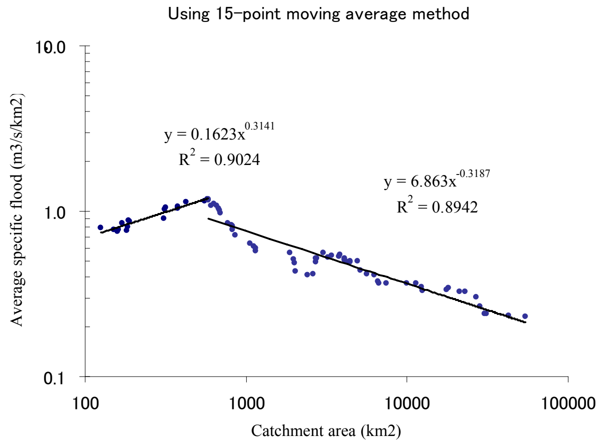

**Fig. 3** Moving average specific flood vs. drainage area plot for all Nepalese river basins.

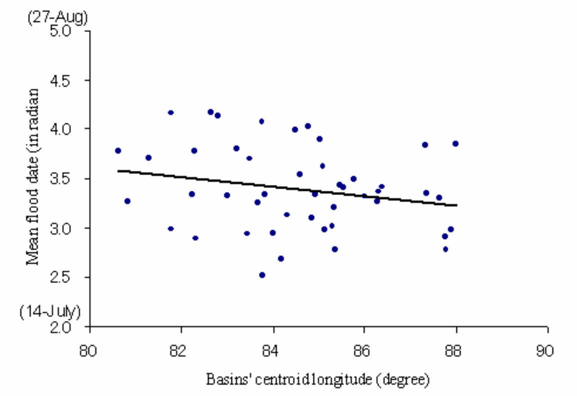

**Fig. 4** Mean flood date of 46 independent river basins located from west to east.

Values of the discordancy measure $D_i$ and the heterogeneity measure $H$ were computed using FORTRAN programme developed by Hosking$^{16)}$. **Table 1** shows heterogeneity measures $H_1$, $H_2$ and $H_3$ based on $LC_v$, $LC_s$ and $LC_k$ respectively for testing the homogeneity of those regions. The whole Nepal can not be considered as a hydrological homogeneous region since the value of heterogeneity measure is large. Heterogeneity measures got highly reduced after removing the discordant sites. However, the heterogeneity measure, $H_1$, for Regions **3** and **5** is larger than that suggested by Hosking and Wallis$^{10)}$. Therefore, estimates of frequency distribution parameters may be obtained by using at site mean and $LC_v$ in these two regions

Since spatial proximity is significant in regionalization as discussed in **Section 3**, the dependent river basins of mountainous regions were categorized in earlier defined regions based on spatial proximity as shown in **Fig. 6**. Concerned to river basins of south part of Nepal which consists of plain terrain and alluvial soil unlike the steep and non-alluvial mountainous river basins, these river basins were classified as a separate region (**Fig. 7**).

**Table 1** Heterogeneity measures for Nepalese river basins

| Region | All stations | | | | Discordant stations removed | | | |
|---|---|---|---|---|---|---|---|---|
| | Number of sites | Heterogeneity measure | | | Number of sites | Heterogeneity measure | | |
| | | $H_1$ | $H_2$ | $H_3$ | | $H_1$ | $H_2$ | $H_3$ |
| All | 46 | 12.39 | 2.39 | -1.68 | 44 | 12.19 | 2.13 | 1.70 |
| 1 | 11 | 2.93 | -0.36 | -1.23 | 10 | 1.00 | -0.83 | -1.39 |
| 2 | 10 | 3.46 | 0.84 | -0.23 | 9 | -0.14 | 0.15 | -0.31 |
| 3 | 7 | 5.98 | 0.49 | -1.28 | 5 | 2.90 | -0.54 | -1.47 |
| 4 | 12 | 8.72 | 2.56 | -0.03 | 7 | 1.07 | 0.48 | -0.20 |
| 5 | 6 | 6.00 | 2.16 | 0.33 | 5 | 2.57 | 1.92 | 1.06 |

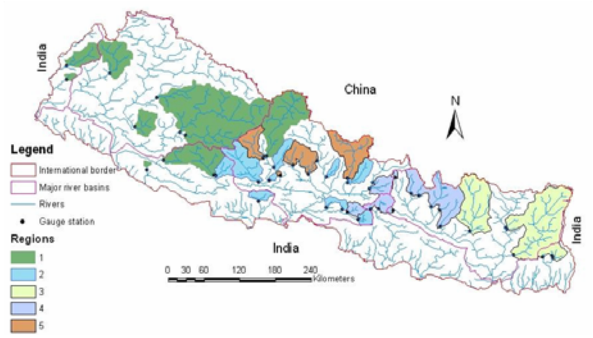

**Fig. 5** Regionalization of 46 independent river basins.

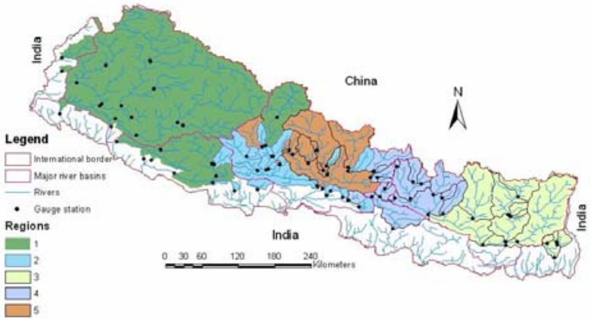

**Fig. 6** Regionalization of mountainous river basins.

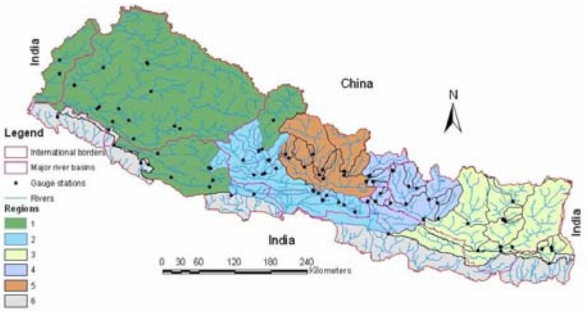

**Fig. 7** Regionalization of all Nepalese river basins.

The present work can be compared with the previous works to observe similarity or dissimilarity in the regions. David et al.9) classifies Himalanyan Nepalese basins into three clusters namely A, B &C while Medium Irrigation Project (MIP) method classifies the Nepalese basins into 7 regions. On comparing the results of David et al. with present study, we find the river basins of Cluster A are spread either over region 3 or region 5. Similarly, Cluster B covers a part of region 1 while the cluster C is partially spread over regions 1 and 4. On comparing the MIP classification with present classification, region 7 of MIP is very much similar to region 6 of the present study. Other regions are partially similar or quite different, may be, because of their different purposes.

## 6. CONCLUSIONS

The main objective to delineate hydrological homogeneous regions for flood frequency analysis was achieved despite the complexity in understanding precisely the factors that have effect on generation of floods. Since the heterogeneity measure values of the delineated regions are reasonable, this paper concludes effective attributes for delineation of regions as specific flood, average flood date and centroid lat-long of river basins. The authors found large heterogeneity measure value when considered all the 46 independent Nepalese river basins as one region. Hence, the present regionalization work will be able to predict design flood better as compared to the previous method developed in Nepal which considered whole Nepal as one homogeneous region. In the next stage, the authors will estimate regional distribution function and index flood value for each delineated region to accomplish the regional flood frequency analysis.

## REFERENCES

1) Cunnane, C.: Methods and merits of regional flood frequency analysis, Journal of Hydrology, Vol. 100, pp. 269-290, 1988.
2) Dalrymple, T.: Flood frequency analysis, US Geol. Survey Water Supply Paper 1543A, pp. 11-51, 1960.
3) Burn, D.H. and Goel, N.K.: The formation of groups for regional flood frequency analysis, Hydrological Sciences Journal, Vol. 45, No. 1, pp. 97-112, 2000.
4) Wiltshire, S.E.: Grouping basins for regional flood frequency analysis, Hydrological Sciences Journal, Vol. 30, No. 1, pp. 151-159, 1985.
5) Mosley, M.P.: Delimitation of New Zealand Hydrologic Regions, Journal of Hydrology, Vol. 49, pp. 173-192, 1981.
6) Burn, D.H.: Catchment similarity for regional flood frequency analysis using seasonality measures, Journal of Hydrology, Vol. 202, pp. 212-230, 1997.
7) Mkhandi, S.H. and Kachroo, R.K.: Regional flood frequency analysis for Southern Africa, XII EMES JOURNEES HYDROLOGIQUES DE L'ORSTOM, MONTEPELLIER, 10-11 Oct., 1996.
8) Hydrology and Ago-meteorology manual, Design manuals for irrigation projects in Nepal, M3, MacDonald and Partners Ltd., 1990.
9) David, M.H., Kansakar, S.R., Gerrard, A.J. and Rees, G.: Flow regimes of Himalayan rivers of Nepal: Nature and spatial patterns, Journal of Hydrology, Vol. 308, pp. 18-32, 2005.
10) Hosking, J.R.M. and Wallis, J.R.: Regional Frequency Analysis: An approach based on L-moments, Cambridge University Press, Cambridge, UK, 1997.
11) Mann, H.B. and Whitney, D.R.: On a test of whether one of two random variables is stochastically larger than the other, The Annuals of Mathematical Statistics, Vol. 18, pp. 50-60, 1947.
12) Rao A.R. and Hamed, K.H.: Flood Frequency Analysis, CRC press LLC, Florida, USA, 2000.
13) Grubbs, F.E.: Procedures for detecting outlying observations in samples, Technometrics, Vol. 11, No. 1, pp. 1-21, 1969.
14) Grubbs, F. and Beck, G.: Extension of sample sizes and percentage points for significance tests of outlying observations, Technometrics, Vol. 14, No. 4, pp. 847-854, 1972.
15) Merz, R. and Bloschl, G.: Flood frequency regionalization-spatial proximity vs. catchment attributes, Journal of Hydrology, Vol. 302, pp. 283-306, 2005.
16) Hosking, J.R.M.: FORTRAN routines for use with the method of L-moments, version 3.04, IBM research division, N.Y., USA, 2005.

**(Received September 30, 2007)**