京都大学防災研究所年報　第 52 号 B　平成 21 年 6 月
Annuals of Disas. Prev. Res. Inst., Kyoto Univ., No. 52 B, 2009

Selection of Regional Frequency Distribution using Simulated Flood Data

Binaya Kumar MISHRA*, Kaoru TAKARA, Yosuke YAMASHIKI
and Yasuto TACHIKAWA*

*Graduate school of Urban and Environmental Engineering, Kyoto University

Synopsis

Regional flood frequency analysis (RFFA), which utilizes regional hydrologic characteristics, is used for estimating design floods. Selection of regional frequency distribution is one of the important elements of the RFFA. Presence of adequate hydrometric stations is essential in each of the hydrologic regions for reliable selection of regional frequency distributions. However, there are only two hydrometric stations in two of the five hydrologic regions. This study deals with selection of regional frequency distributions in each of the five hydrologic regions, using the L-moment ratio diagrams with emphasis on the regions that have inadequate hydrometric stations. Simulated floods have been employed to address the observation data inadequacy in the two hydrologic regions. A rainfall-runoff model SimHyd enabled the generation of simulated flood data. Generalized extreme value and lognormal (3-parameter) distributions were found to be fitting well in all of the hydrologic regions.

Keywords: regional distribution, SimHyd, rainfall-runoff model, L-moment

1. Introduction

Estimation of maximum discharge of different return periods is required in design and planning of various hydraulic structures. The maximum discharge is also called as design flood. Usually, at-site flood frequency analysis is used for estimation of design floods at the locations which have long periods of flood record. However, such locations rarely coincide with the locations at which design flood estimation is required. In such situation, regional hydrologic characteristics can be used for estimation of design flood. Regional flood frequency analysis methods use regional hydrologic characteristics for estimating the design floods.

At present, direct regression method and index flood method are widely used regional flood frequency procedures. While the former was extensively used in the U.S., the latter seems to be gaining interest among researchers. Research on index flood method has been increased noticeably after the development of L-moments based tests.

Numerous studies (for example, Parida et al., 1998; Kumar et al., 1999) are found on design flood estimation for different countries using the regional flood-frequency analysis methods. In Nepal case, study on design flood estimation is extremely limited. Water and Energy Commission Secretariat of Nepal (WECS) is a frequently used method for estimating design flood in ungauged basins of Nepal (Sharma and Adhikari, 2004). The WECS method is based on the direct-regression based regional flood frequency analysis. All the Nepalese river basins were considered in one hydrologic region in the WECS method. However, all the Nepalese basins should not be considered in one hydrologic homogeneous region since there is large physiographic/climatic variability among the rivers basins. Hence, the Nepalese river basins were divided into five hydrologic regions in the previous study (Mishra et al., 2009). The hydrologic regions were proposed by superimposing NRCS runoff curve number incorporated sample river basins map over a long-term monsoon rainfall map. The proposed hydrologic regions were tested using the L-moment based method of regional hydrologic homogeneity test.

This study is aimed for selecting regional frequency distributions in each of the five hydrologic regions using the L-moment ratio diagrams. The selection of regional frequency distribution is an important element of index flood-based regional flood frequency analysis method. In the index flood method, the basic relationship for estimating design flood (Q_T^i) of return period T at site i is expressed as Eq.(1) (Dalrymple, 1960):

$$Q_T^i = q_T \mu_i \tag{1}$$

Where q_T and μ_i are regional frequency distribution factor and index flood respectively.

The regional frequency distribution factor (growth factor) is a dimensionless quantile which remains same throughout a hydrologic region for a specific return period T. The regional frequency distribution is governed by the values of flood moment coefficients such as coefficient of variance, skewness and kurtosis of various stations in the region.

There should be enough hydrometric stations in a hydrologic region for reliable selection of regional frequency distributions. Although numerous authors (for example, Parida et al., 1998 and Kumar et al., 1999) have investigated on selection of regional frequency distributions, the in-depth analysis in relation to number of hydrometric stations in a hydrologic region is limited to Hosking and Wallis (1997). They found that the rate of decrease in the RMSE (Root Mean Square Error) is close to $n^{-1/2}$ where n represents number of sites in a hydrologic region. Their analyses showed that the RMSE in growth factor gets largely decreased when the numbers of hydrometric stations are increased from 2 to 5. According to 5T guideline (Burn and Goel, 2000), a minimum of 5 times T flood values should be available in the concerned hydrologic region for reliable estimation of T-year return period design flood.

In Nepal case, there are only two hydrometric stations in two of the five hydrologic regions. The availability of only two hydrometric stations in a hydrologic region is inadequate for reliable selection of regional frequency distributions. The limitations of observation hydrometric data were addressed by generating synthetic flood data in some ungauged basins.

The synthetic flood data were generated using the climatic (rainfall, evapotranspiration) data of the ungauged basins. A conceptual rainfall runoff model SimHyd (Podger, 2004) was used to generate synthetic flood data after the calibration process. The L-moment ratio diagrams were used to select regional frequency distributions (Hosking and Wallis 1997; Hosking 2005).

2. Study area and data analysis

Selection of regional frequency distribution is intended for hydrologic homogeneous regions of Nepal which 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 to 248 km north–south. Within this relatively small latitudinal extent, altitude rises from 60 m in south to 8848 m (Mount Everest, the world’s highest peak) in North. Physiographically, the country is divided into five regions: Terai (Plain), Siwalik Hills, Middle Mountains, High Mountains and High Himalayas (Fig.1).

Annual rainfall varies from 250 mm in the rain-shadow areas of north-west to about 5000 mm on the windward slopes. Western parts receive less rainfall in the monsoon than that of central and eastern parts. During the winter, however, rainfall is more reliable in the west than in the east. More rainfall occurs on the south-eastern slopes which act as windward side to monsoon winds during the summer. The hilly areas of western and north-western slopes as well as those behind the high mountains receive little rainfall. Isohyets of 1500 to 2500 mm cover most of the eastern and central hilly regions while those in the western region are between 1000 to 1500 mm (Sharma, 2005).

Mahakali, Karnali, Narayani and Koshi are the major river systems of Nepal. Although the geographical area of Nepal is 147,181 km^2, Nepal drains the discharge from more than 194,000 km^2; the additional area being mainly in Tibet (Sharma and Adhikari, 2004). Floods in Nepal are primarily dominated by the monsoons characterized by high precipitation during the summer monsoon from June to September. The remaining period receives only about 20% of the annual precipitation.

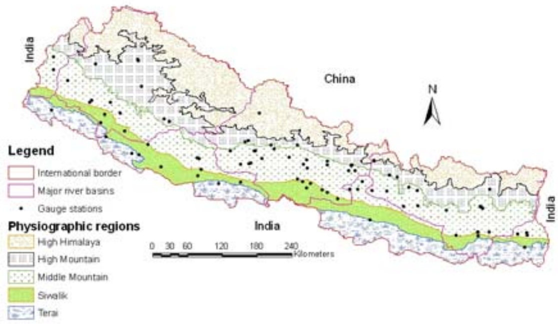

Fig. 1 Physiographic regions map of Nepal.

Mishra et al. (2009) has identified five hydrologic homogeneous regions inside the Nepalese territory (Fig.2). The hydrologic regions 1, 2, 3, 4 and 5 cover 16.24%, 14.67%, 33.20%, 24.63% and 11.26% land area of the total Nepalese territory area respectively.

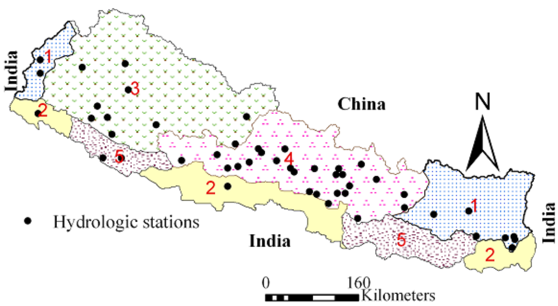

Fig. 2 Hydrologic regions of Nepal with regions 1, 2 and 5 spreading over different area.

There are 7, 10 and 24 hydrometric stations in regions 1, 3 and 4 respectively whereas there are only two stations in each of the region 2 and 5. Department of Hydrology and Meteorology (DHM), Nepal is responsible for distributing hydrologic/climatic data in Nepal. Annual maximum streamflow data were collected for 49 stream gauging stations of Nepal (Fig.2). Out of 49 stations, 46 stations possess more than 10 years of annual maximum discharge series. The daily discharge data of hydrometric station 296 was collected for calibrating the conceptual rainfall runoff model SimHyd. The precipitation stations (Table 2): 208, 214, 412, 1119, 1120, 1226 and 1319, were used for collecting daily rainfall data whereas the agrometrological stations: 405, 409, 419, 1215 and 1319 were used for collecting monthly evapotranspiration data.

Screening of flood data was started by visual inspection. The flood data series being homogeneous and stationary are the basic assumptions in flood frequency analysis. The flood data series was tested for homogeneity and stationarity using the method of Mann and Whitney (1947). All the data series were found homogeneous and stationary at 5% significance level.

Presence of outliers in the data causes difficulties when fitting a distribution to the data; hence outliers should be removed before performing flood frequency analysis. The GB test (Grubbs and Beck 1972; Rao and Hamed 2000) was used to detect outliers. Approximate relationship proposed at 10% significance level by Pilon and Harvey (1993) was used in calculating the GB statistic. The study found one outlier at 14 stations (station indices: 267, 404.6, 438, 439.8, 445.3, 465, 530, 570, 589, 602, 620, 627.5, 650 & 660), 2 outliers at station index 447.9 and three outliers at station index 241 (Fig.3). Except one (station index 241), all these outlier stations were found situated in high rainfall region when their spatial positions were observed.

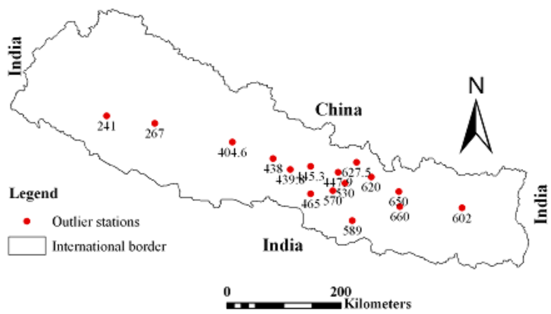

Fig. 3 Outlier hydrometric stations of Nepal.

3. L-moment ratio diagram

In this study, the L-moment ratio diagram has been employed to select regional frequency distribution.

$$\tau_4 = (1+5\tau_3^2)/6 \tag{2}$$

$$\tau_4 = 0.10701 + 0.11090\tau_3 + 0.84838\tau_3^2 - 0.06669\tau_3^3 + 0.00567\tau_3^4 - 0.04208\tau_3^5 + 0.03763\tau_3^6 \tag{3}$$

$$\tau_4 = 0.12282 + 0.77518\tau_3^2 + 0.12279\tau_3^4 - 0.13638\tau_3^6 + 0.11368\tau_3^8 \tag{4}$$

$$\tau_4 = 0.1224 + 0.30115\tau_3^2 + 0.95812\tau_3^4 - 0.57488\tau_3^6 + 0.19383\tau_3^8 \tag{5}$$

$$\tau_4 = \tau_3(1+5\tau_3)/(5+\tau_3) \tag{6}$$

The L-moment ratio diagrams are plot of L-moment ratios (L-coefficient of variance (t), skewness (t_3) and kurtosis (t_4)) of various hydrometric stations in a region on the theoretical relationships of the frequency distributions. A diagram based on L-skewness (τ_3) versus L-kurtosis (τ_4) can be used to identify appropriate regional frequency distributions.

For a given region, the sample L-moment ratios t_3 and t_4 for each station as well as their regional average are plotted on the L-moment ratio diagram. A suitable parent distribution is that which averages the scattered points closely. Relationships (Rao and Hameed, 2000) for constructing the L-moment diagram for the generalized logistic (GLO), generalized extreme value (GEV), lognormal (LN3), Pearson type III (PT3) and generalized Pareto (GPa) distributions, which have been used in this study as probable regional frequency distributions, are given respectively by Eqs. 2-6.

4. Hydrologic simulation

The regional frequency distribution is selected from the values of moment ratios: coefficient of variance, skewness and kurtosis at various stations in the region. Hence, there should be enough hydrometric stations in each of the hydrologic regions for reliable selection of regional frequency distributions.

There are five numbers of hydrologic homogeneous regions inside the Nepalese territory. Out of these five hydrologic regions, there are 7, 10 and 24 hydrometric stations in regions 1, 3 and 4 respectively; whereas there are only two stations in each of the region 2 and 5 (Fig.3).

Use of simulated flood data is intended to address the gauged data limitation in the two hydrologic regions. Simulated flood data is the annual maximum daily discharge obtained from hydrologic simulation. A rainfall runoff model can enable generation of synthetic flood data after suitable calibration. Rainfall-runoff models are used for various applications, ranging from the estimation of basin water yield to the estimation of land use and climate change impacts on runoff characteristics. Most rainfall-runoff models can be calibrated successfully to reproduce the recorded runoff; however it is difficult to determine appropriate parameter values to use for modeling runoff in an ungauged basin. This section describes the application of the lumped conceptual daily rainfall-runoff model SimHyd in generation of synthetic flood data.

4.1 Rainfall runoff model: SimHyd

SimHyd, a lumped conceptual rainfall runoff model, simulates daily runoff using daily rainfall and areal potential evapotranspiration (PET). SimHyd is one of the most commonly used rainfall-runoff models in Australia (Chiew and Siriwardena, 2005). The model is a component of the rainfall runoff library (RRL) produced by Cooperative Research Centre for Catchment Hydrology (CRCCH) (Podger, 2004). This is a mass balance model that is based on conceptual relationships. The structure of SimHyd and the algorithms describing water movement into and out of the storages are shown in Fig. 4. The nine parameters of SimHyd are shown in red color.

In this model, daily rainfall first fills the interception store which is emptied each day by evaporation. The excess rainfall is then subjected to an infiltration function that determines the infiltration capacity. The excess rainfall that exceeds the infiltration capacity becomes infiltration excess runoff.

Moisture that infiltrates is subjected to a soil moisture function which diverts the water to the stream (interflow), groundwater store (recharge) and soil moisture store. Interflow is first estimated as a linear function of the soil wetness (soil moisture level divided by soil moisture capacity). Groundwater recharge is estimated as a linear function of the soil wetness.

Evapotranspiration from the soil moisture store is estimated as a linear function of the soil wetness, but can not exceed the atmospherically-controlled rate of areal potential evapotranspiration. The soil moisture store has a finite capacity and overflows into the groundwater store. Baseflow from the groundwater store is simulated as a linear recession from the store.

The model therefore estimates runoff generation from three sources: infiltration excess runoff, interflow (and saturation excess runoff) and base flow.

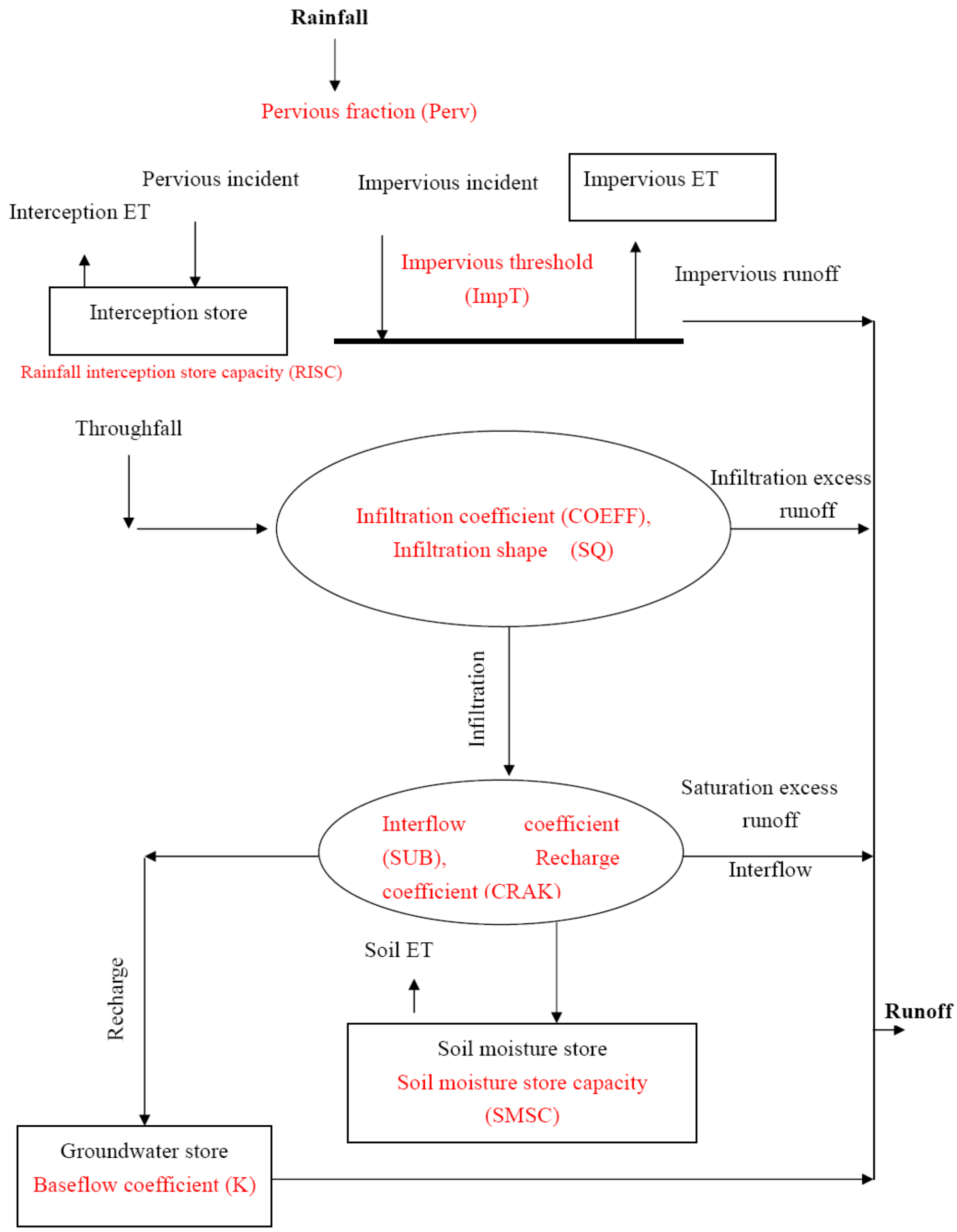

Fig. 4 Structure of rainfall runoff model SimHyd.

4.2 Calibration and validation

Generation of synthetic flood data is intended in regions 2 and 5. Hence, model calibration should be performed in each of these regions using relevant gauged basins data. However, daily discharge could be collected in only one of the hydrologic regions. The distinction between these two regions is mainly in term of rainfall pattern. Therefore, at the moment, the same model parameters were used for both of the regions.

The rainfall runoff model was calibrated using daily discharge data of the Kiran-khola river basin (outlet index: 296) (Fig.5) situated in hydrologic region 5. The calibration basin has drainage area of 132.28 km^2 with outlet location at 81°33’ 19”, 28°05’ 57” (latitude, longitude). Calibration of the model was performed over the period 2000/01/01 to 2005/12/31. The model possesses manual as well as automatic optimization facilities for parameter calibration. In this study, SCE-UA (shuffled complex evolution-university of Arizona) option was selected as automatic optimization of the model parameters. The Nash-Sutcliffe efficiency was selected as objective function while calibrating the model. The Nash-Sutcliffe coefficient was found to be 0.44. The derived calibration parameters are shown in Table 1.

The model performance was checked by comparing the observed annual maximum flood values and simulated values for six years in the Jhanjhari river basin (station index: 363) which is located in the same hydrologic region. The performance correlation between modelled and measured flood flow is shown in Fig.6.

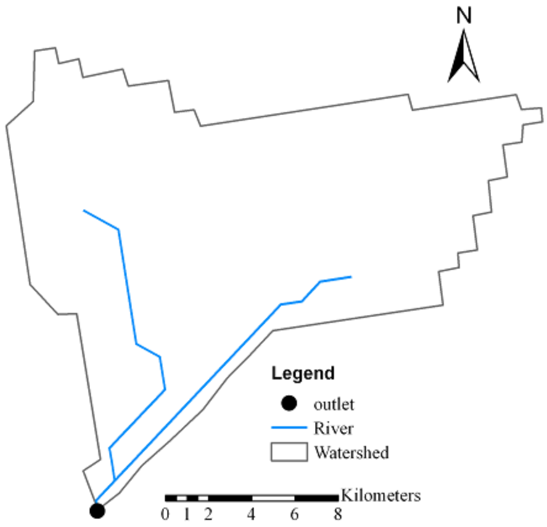

Fig. 5 Kiran-khola River basin for calibration of the model.

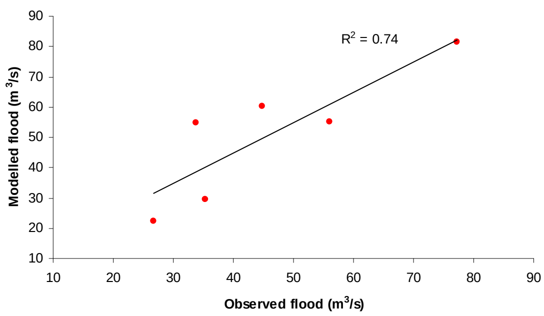

Fig. 6 Comparison of annual maximum discharges in the Jhanjhari River.

Table 1 Calibrated model parameter values at the Kiran-khola River.

Parameter	Description	Unit	Value
K	Baseflow linear recession parameter	none	0.66
ImpT	Impervious threshold (threshold for runoff from impervious area)	mm	1.52
COEFF	Infiltration coefficient	day^{-1}	250.83
SQ	Infiltration shape (part of the infiltration exponent)	none	0.03
SUB	Constant of proportionality in interflow equation (Interflow coefficient)	day^{-1}	0.14
Perv	Pervious fraction	none	0.90
RISC	Rainfall interception store capacity	mm	3.50
CRAK	Recharge coefficient	day^{-1}	0.76
SMSC	Soil moisture store capacity	mm	319.50

Table 2 Description of synthetic flood data basins and relevant climatic stations.

Region	Basin's outlet		Drainage area, km^2	climatic station's index		Sample size	Remarks
	latitude	longitude		rainfall	PET		
2	28 50 00	80 31 16	12.50	214	405	15	Mohana river (281)
	27 42 10	83 27 50	554.00	only obs. flow data		17	Tinau (390)
	26 35 16	87 39 21	162.84	1408	1319	13	Near Damak
	28 33 10	80 48 15	232.85	208	405	13	
	26 25 07	87 14 52	315.90	1319	1319	15	Near Biratnagar
5	28 05 57	81 33 19	132.28	412	419	15	c-Boring (296)
	28 09 22	81 45 13	72.74	412	409	15	Jhanjhari, Dhakeri (363)
	26 32 20	86 54 00	112.86	1226	1215	13	
	26 48 41	85 46 45	145.87	1119	1215	13	Near Gaushala, Mahottari
	26 50 05	85 34 15	219.54	1120	1215	13	Near Malangwa

4.3 Simulated floods

The suitably calibrated rainfall runoff model: SymHyd (Podger, 2004) enabled generation of synthetic flood data basins in the regions 2 and 5. At first, daily discharge was generated using the daily rainfall data and monthly PET of closer stations to the concerned basin. Maximum of the annual daily discharge series was considered as flood data of that year for estimation of regional frequency distribution. With the generation of synthetic annual maximum discharge data series, there are five data basins in each of the regions 2 and 5 respectively (Table 2).

5. Results and discussion

The L-moment ratio diagrams were employed in each of the five hydrologic regions for selecting suitable regional frequency distribution. The L-skewness and L-kurtosis of all sites in the regions as well as corresponding regional averages is shown in Figs. 7-11, along with the theoretical lines of the five probable distributions.

From the Figs. 7-11, it can be concluded that the GPa distribution is inappropriate in Nepalese basins in relation to design flood prediction. The other four distributions, which do not show significant differences, may be selected as reasonable distribution in all of the hydrologic regions. However, the GEV and LN3 fit closely in all of the hydrologic regions

After selection of regional frequency distribution, growth factors can be computed for the selected regional frequency distributions. Fig.12 shows the variation of growth factor with the return period.

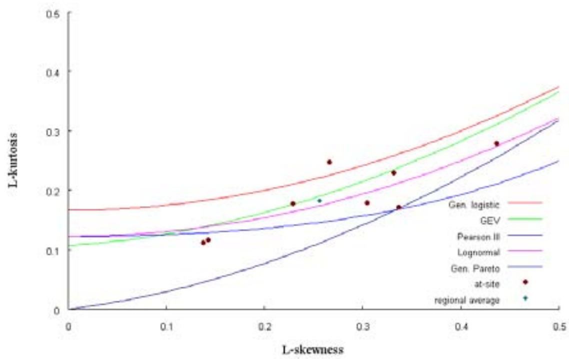

Fig. 7 L-moment ratio diagram for region 1.

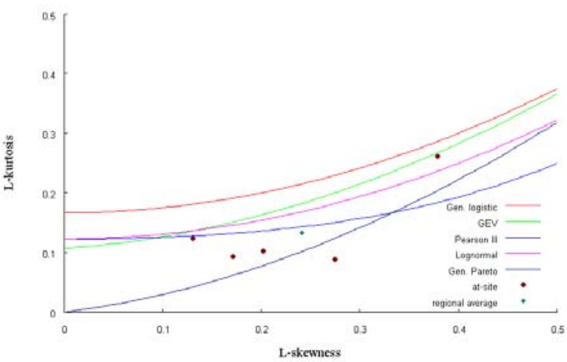

Fig. 8 L-moment ratio diagram for region 2.

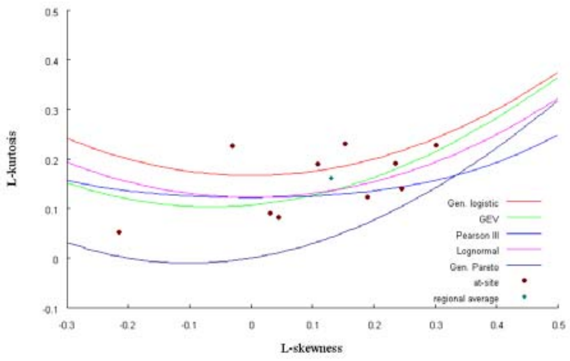

Fig. 9 L-moment ratio diagram for region 3.

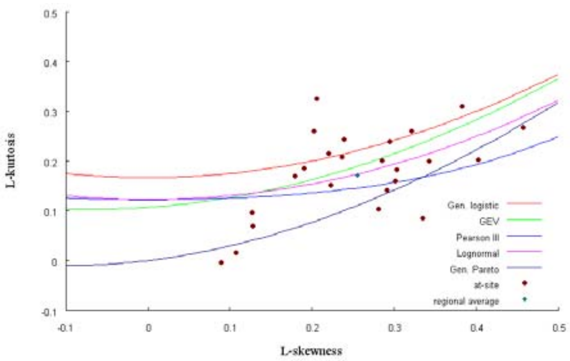

Fig. 10 L-moment ratio diagram for region 4.

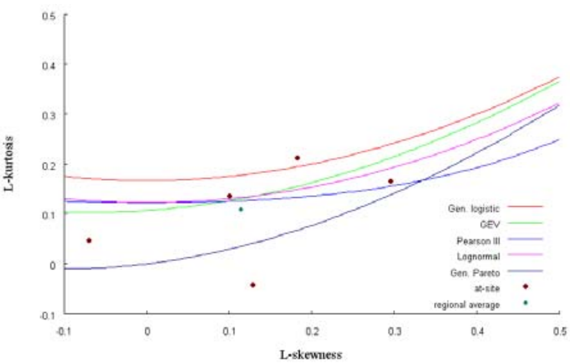

Fig. 11 L-moment ratio diagram for region 5.

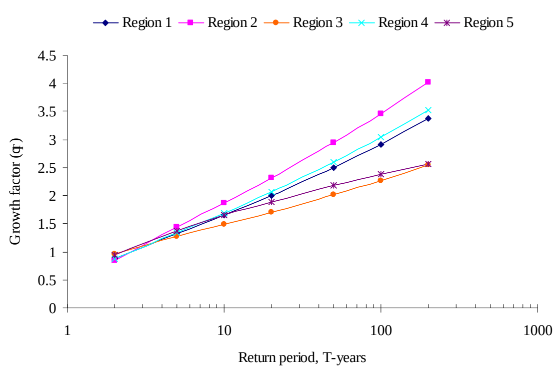

Fig. 12 Variation of growth factors with return periods.

Although the use of L-moment ratio diagram is supposed to be better method for selection of regional frequency distribution, the method needs to be assessed using some well established alternative methods for validation of the selected distributions. The conventional method of regional frequency distribution selection may be employed to justify the method of L-moment ratio diagram. In the conventional method, the probable frequency distributions are applied to fit the data of each station in the region. The distribution which fit majority of the stations in the region closely is selected as regional frequency distribution.

The conventional method of regional frequency distribution selection has been tested in hydrologic region 1. Out of 7 test stations in the region, the distributions: LN3, GEV and GLO were found to be closely fitting at 3, 3 and 1 stations respectively. These results justify the adopted L-moment ratio diagram method of distributions selection.

6. Summary and Conclusions

In this study, regional frequency distribution was selected for each of the five hydrologic regions. The selection of regional frequency distribution in hydrologic regions 1, 3 and 4 were made using observation flood data of the various hydrometric stations in the regions. However, in hydrologic regions 2 and 5, regional frequency distributions were selected with the use of supplemented synthetic flood data. The commonly used 3-parameter distributions: generalized logistic, GEV, Pearson type III, lognormal and generalized Pareto were used as possible regional frequency distributions.

Method of L-moment ratio diagram was used for selection of appropriate regional frequency distributions. The conventional method of regional frequency distribution selection was tested in hydrologic region 1 for validating the methods of L-moment ratio diagram.

Generation of synthetic flood data is intended to address the limitation of inadequate numbers of hydrometric stations in the hydrologic regions 2 and 5. A rainfall runoff model SimHyd enabled the generation of synthetic flood data. Structure and principle of the rainfall runoff model SimHyd were introduced. The model was calibrated using the daily discharge data of the Kiran-khola River basin. The model performance was checked by comparing the observed annual maximum flood value and simulated value of the Jhanjhari river basin. The results showed that the SimHyd model can perform well. The suitably calibrated rainfall runoff model SymHyd enabled generation of five numbers of flood data basins in each of the regions 2 and 5.

Acknowledgements

The authors would like thank to the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan for providing financial support in carrying out this research work. The authors also wish to thank all the members of Innovative Disaster Prevention Technology and Policy Research Lab, Disaster Prevention Research Institute, Kyoto University, Japan for their comments/help at various instances.

References

Burn, D.H. and Goel, N.K. (2000): The formation of groups for regional flood frequency analysis, Hydrological Sciences Journal, Vol. 45, No. 1, pp. 97-112.

Chiew, F.H.S. and Siriwardena, L. (2005): Estimation of SimHyd parameter values for application in ungauged catchments (http://www.mssanz.org.au/modsim05/papers/chiew_2.pdf).

Dalrymple, T. (1960): Flood frequency analysis, US Geological Survey's Water Supply Paper, No. 1543A, pp.11-51.

Grubbs, F. and Beck, G. (1972): Extension of sample sizes and percentage points for significance tests of outlying observations, Technometrics, Vol. 14, No. 4, pp. 847-854.

Hosking, J.R.M. and Wallis, J.R. (1997): Regional Frequency Analysis: An approach based on L-moments, Cambridge University Press, Cambridge, UK.

Hosking, J.R.M. (2005): FORTRAN routines for use with the method of Lmoments, Version 3.04.

Kumar, R., Singh, R.D. and Seth, S.M. (1999): Regional flood formulas for seven subzones of zone 3 of India, Journal of hydrologic engineering, Vol. 4, pp. 240-244.

Mann, H.B. and Whitney, D.R. (1947): 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.

Mishra, B.K., Takara, K. and Tachikawa, Y. (2009): Integrating the NRCS-runoff curve number in delineation of hydrologic homogeneous regions, Journal of hydrologic engineering, ASCE, DOI: 10.1061/(ASCE)HE.1943-5584.0000101.

Parida, B.P., Kachroo, R.K. and Shrestha, D.B. (1998): Regional flood frequency analysis of Mahi-Sabarmati Basin (Subzone 3-a) using Index flood procedure with L-moments, Water Resources Management, Vol. 12, pp. 1-12.

Pilon, P.J. and Harvey, K.D. (1993): Consolidated frequency analysis- reference manual, Version 3.1, Ecosystem Services and Evaluation Directorate, Canada.

Podger, G. (2004): Rainfall runoff library Manual,CRC for Catchment Hydrology, Australia (http://www.toolkit.net.au/cgi-bin/WebObjects/toolkit).

Rao, A. R. and Hamed, K.H. (2000): Flood Frequency Analysis, CRC Press LLC, Florida.

Sharma, K.P. and Adhikari, N.R. (2004): Hydrological Estimations in Nepal, Department of Hydrology and Meteorology, Kathmandu. Nepal.

Sharma, B.P. (2005): Gateway to Land and Water Information, Nepal National Report, MENRIS, ICIMOD, Kathmandu, Nepal.

洪水解析データを用いた地域（洪水）頻度分布の選定

Binaya Kumar MISHRA*・宝　馨・山敷庸亮・立川康人*

* 京都大学大学院工学研究科都市環境工学専攻

要　旨

地域的な水文特性を用いた地域洪水頻度解析(RFFA)を計画洪水の推定に用いた。地域頻度確率分布の選定はRFFAにおいて非常に重要な要素の一つである。それぞれの水文流域において十分な水文観測所の存在が、地域頻度確率分布の信頼性の高い選定のために非常に重要である。しかしながら本研究の対象とした五つの水文流域のうち二つは，水文観測所が二つしか存在しない。本研究においては、水文観測所の数が不十分な水文流域に重きをおいたL積率比関係図を用いて，五つの水文流域それぞれの地域頻度確率分布の選択を行なった。水文観測所の数の不足した二つの水文流域については観測データの代わりに洪水解析流量を用いた。流量の模擬発生には降雨流出モデルであるSimHydによる結果を用いた。一般化極地分布や３母数対数正規分布は全ての流域において良い適合性を示すことが分かった。

キーワード：地域確率分布, SimHyd, 降雨流出モデル, L積率