Annual Journal of Hydraulic Engineering, JSCE, Vol.53, 2009, February

ESTIMATION OF INDEX FLOOD IN HYDROLOGIC REGIONS WITH LIMITED FLOOD DATA AVAILABILITY

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

[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, 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)

Delineation of hydrologic homogeneous regions, selection of regional frequency distribution and estimation of index flood are three major steps of regional flood frequency analysis. This study deals with the estimation of index flood relationship in each of the previously identified five hydrologic regions of Nepal. In three regions which have enough hydrologic stations, regression technique was employed on observed annual maximum discharges. The remaining regions possess only two streamflow gauge stations each. Hence, the index flood relationship derived using regression technique with such limited stations may not be reliable. To address this problem, the authors, at first, used a conceptual daily rainfall-runoff model, SimHyd, to generate synthetic annual maximum discharge series and then, derived relationship for index flood estimation using regression technique. Nash-Sutcliffe efficiency coefficient was used to assess the predictive power of the derived relationships.

Key Words **:** *Index flood, rainfall-runoff model, regression, SimHyd, synthetic*

1. INTRODUCTION

Information on maximum discharge of different return periods is required in design and planning of various hydraulic structures. Such maximum discharge is popularly known as design flood. Regional flood frequency analysis[1)] is one of the widely used design flood estimation techniques at ungauged sites or the sites with short length of observation data.

This work is related with index flood based regional flood frequency analysis approach. It assumes that the coefficient of variation (C_v) of floods is constant and does not vary with drainage area[2)]. The previous work[3)] which identified five hydrologic homogeneous regions inside the Nepalese territory was tested for similarity of C_v among the basins in a region. The basic relationship for estimating design flood (Q_T^i) of return period T at site i in index flood[4)] based regional flood frequency analysis method is expressed as Eq.(1):

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

The first right-side term q_T in Eq.(1) represents regional frequency distribution parameter which remains the same throughout a hydrologic region for a specific return period. The second term μ_i is called index flood. It is defined as the average likely flood. For the gauged site, it is usually taken as mean or median of the annual maximum discharge series[5)]. The mean of the maximum discharge series corresponds to 2.33-years flood in Gumbel distribution. Index flood is related to physiographic/climatic factors like catchment area, land cover, soil type, rainfall, etc for ungauged basins. Index flood is popularly known as scale factor because it is mostly governed by the corresponding catchment area. For each hydrologic region, a particular relationship needs to be established for estimation of index flood. The present study intends to establish relationship for estimating index flood in each of the identified hydrologic homogeneous region.

Index flood estimation may be constrained by the data availability for the application of a particular procedure. For gauged sites, the estimation of index flood is generally made by taking mean6) of the observed annual maximum discharge series. In the case of ungauged sites, most of the previous works7) used simple regression for deriving index flood relationship. A very limited studies8) have used multi-regression technique by employing the attributes: main stream slope, basin's elevation, etc. in addition to drainage area. No specific study is available on estimating index flood for Nepalese context.

The regression technique can be considered reliable only when the number of observation stations is large enough. According to Reddy9), number of observation stations should be at least 3 to 4 times the number of regression parameters for reliable regression relationship. In addition, according to 5*T* guideline10), a minimum of 5 times *T* observed annual maximum discharge values should be available in the concerned hydrologic region for reliable estimation of T-year return period design flood. If the 50 year event is selected, then 5T guideline implies that at least 250 station-years flood values are required for an effective quantile estimate.

The hydrologic regions 2 and 5 (**Fig.1**) have only two streamflow gauging stations, hence inadequate for performing regression technique. To address this problem, the study generated synthetic annual maximum discharge series for various ungauged basins using climatic (rainfall, evaporation etc.) data available in the regions. Then, regression technique was applied on synthetic-observed values. But, the remaining regions 1, 3 and 4 were found with large number of streamflow gauging stations. Therefore, for these regions, index flood relationships were derived using observed annual maximum discharge series.

The present work addresses the problem associated with estimation of index flood, particularly in the region having limited observed streamflow data. An overview on study area and data availability has been provided in the next section. The methodology section deals the general approaches used in deriving index flood relationship. Description on the rainfall-runoff model, SimHyd11), which is intended for generating synthetic annual maximum discharge series, has been discussed in Section 4 and 5. The adopted procedure in establishing index flood relationship is discussed in Section 6. The results and discussion have been dealt in second last section. The final section concludes the achievement of the study.

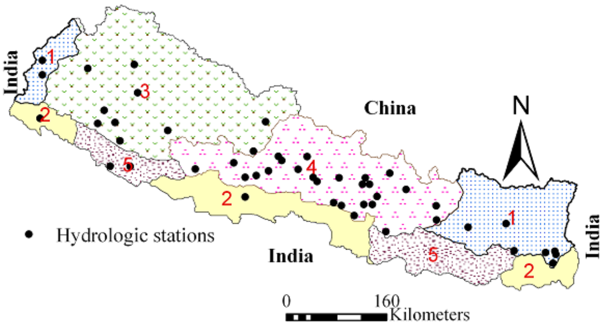

Fig.1 Hydrologic homogeneous regions of Nepal with regions 1, 2 and 5 are spread over different area.

2. STUDY AREA AND DATA AVAILABILITY

Estimation of index flood relationships is intended for previously identified five hydrologic homogeneous regions (**Fig.1**) of Nepal. In general, Nepal gets more than 1500 mm of rainfall annually. About 80% of total annual rainfall is concentrated in the months of July, August and September. Consequently, it brings many large floods and landslides. The regions 1, 2, 3, 4 and 5 covers 16.24%, 14.67%, 33.20%, 24.63% and 11.26% land area of the total Nepalese territory area respectively. The total area of Nepal is 147,181 km^2.

About two-third of total hydrological stations in regions 1, 3 and 4 were used for deriving index flood relationship. Because the statistical parameters (mean, coefficient of variance, skewness etc.) largely fluctuate for smaller sample size, the stations which consisted at-least 10 years of annual maximum discharge series were only considered. The remaining one-third stations are intended for validation of derived relationship. These three regions cover mostly steep terrain. The regions 1 and 4 get more rainfall with increasing elevation whereas region 3 gets less rainfall with increasing elevation.

The hydrologic regions 2 and 5 have only two river discharge gauging stations. The daily discharge data of one hydrological station was used for calibration of the rainfall-runoff model. These daily discharges are the mean of streamflows corresponding to stage readings at 08:00, 12:00 and 16:00 hours. The discharge data of another station was used for validation of the model. Enough climatological stations are available in these two regions for generating synthetic discharge data using the calibrated model in ungauged basins. Region 2 gets relatively more rainfall than region 5.

Hydrological (discharge) and climatological (rainfall, evapotranspiration) data were collected from Department of Hydrology and Meteorology, Kathmandu, Nepal. The information like basin's elevation and river slope were derived from SRTM 90 m DEM whereas land cover and hydrological soil group were derived from AVHRR satellite imagery and FAO global soil data respectively. These physiographic data can be freely downloaded from the concerned websites.

3. METHODOLOGY

The index flood estimation methods (**Fig. 2**) depend on whether the site of interest is gauged or ungauged. If the site of interest is gauged, estimation methods use information on annual maximum discharge series at the site of interest. This method may be called as direct method. The observed sample mean is mostly used as estimator of index flood.

For the ungauged site in a hydrologic homogeneous region, the index flood estimation methods use information on regional basins' flood values and the corresponding physiographic/climatic properties. This method may be called as indirect method. This work is related with indirect method of index flood estimation. Different index flood estimation techniques are discussed hereunder.

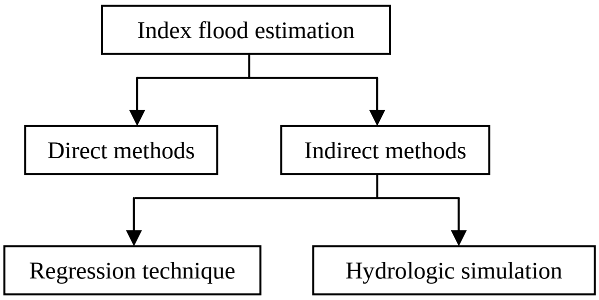

Fig.2 Index flood estimation methods.

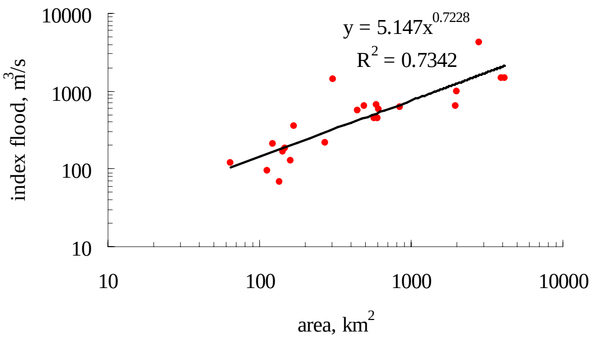

Fig.3 Index flood versus catchment area for different gauged basins in Region 4.

(1) Using simple regression technique

Index flood at any location is mainly governed by its drainage area, *A*. Another advantage of using only drainage area is its easy availability. To establish relationship for index flood estimation in term of drainage area, there should be enough numbers of gauged stations in the concerned hydrologic region. At first, index flood at each gauged station is computed as mean of the annual maximum discharge series. These mean values are called observed index flood values. Then, a plot of these observed index flood values is made against the corresponding drainage area. As illustration, **Fig.3** shows the plot of index flood against drainage area for region 4. Equation of the line which fits most closely to these plot points is defined as index flood relationship. The present study intends to perform simple regression for all the hydrologic regions.

(2) Using multiple regression technique

For better estimation of index flood, the relationship may be expressed in term of more than one independent variable. In addition to drainage area, the information like slope, land cover, soil type, rainfall, etc. can be integrated in estimation of index flood. Such integration may be performed by applying multiple regression technique. The multi-regression model expresses the estimate of index flood, $\boldsymbol{\mu_i}$ as:

$$\mu_i = C_0 A_1^{c_1} A_2^{c_2} \ldots\ldots A_n^{c_n} \qquad (2)$$

In Eq.(2), A_1, A_2,….,A_n represents an appropriate set of physiographic/climatic attributes of the basin. The constant C_0 and exponent c_1, c_2, …,c_n are estimated from observed index flood of each basin and corresponding physiographic/climatic information. In the present study, the authors intend to perform multiple regressions for different possible combinations of physiographic/climatic attributes to establish better index flood relationship.

(3) Using hydrological simulation

As already mentioned in Section 2, there are only two river gauging stations in each of the region 2 and 5. Index flood relationship derived using regression technique with the data of such limited gauged stations can not be considered reliable for estimation of index flood. A rainfall-runoff model, after calibration and validation with the help of available streamflow data in the region, enables generation of annual maximum discharge series at site of interest. Then, index flood is computed by taking average of these synthetic annual maximum discharge series.

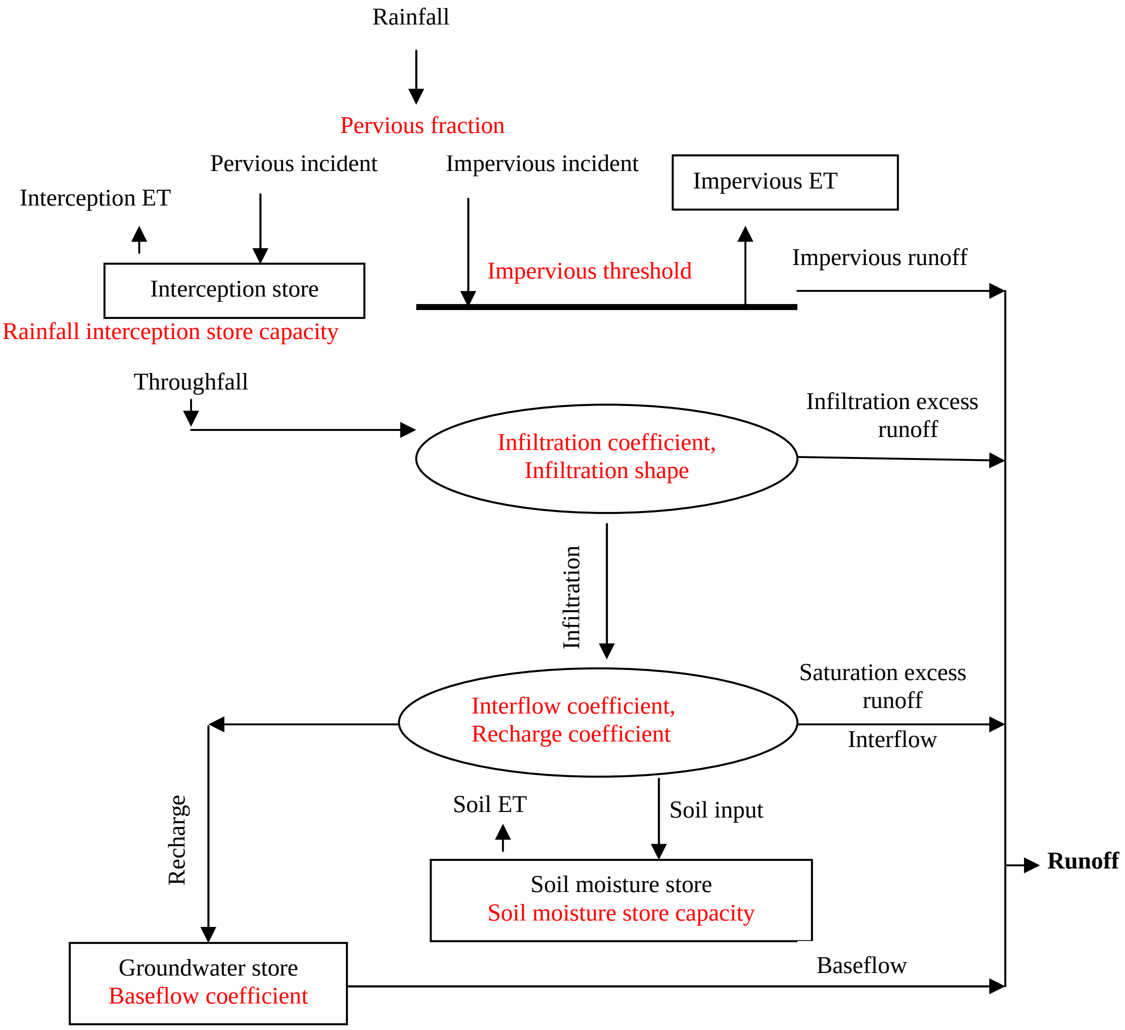

Fig.4 Structure of SimHyd model.

4. SIMHYD: A RAINFALL-RUNOFF MODEL

In order to generate the synthetic annual maximum discharge series at different sites in the region 2 and 5, the study uses hydrological simulation with SimHyd[11)], a lumped conceptual daily rainfall-runoff model. SimHyd simulates daily runoff using daily precipitation and potential evapotranspiration (PET) as input data. SimHyd is one of the commonly used rainfall-runoff models and has been extensively tested using data from various countries. SimHyd is one of the rainfall-runoff models in RRL (Rainfall-Runoff Library), a software product in the Catchment Modelling Toolkit.

The structure of SimHyd and the algorithms describing water movement into and out of the storages are shown in **Fig.4**. In SimHyd, 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.

5. CALIBRATION AND VALIDATION

Calibration of the model was performed using daily stream flow of the Kiran River basin (station index 296) located in region 5 over the period 2000/01/01 to 2005/12/31. The model possesses both manual as well as well-established automatic optimization facilities for parameter calibration. This study used both: manual method for matching annual maximum discharge values and automatic optimization for other periods. The SCE-UA (shuffled complex evolution - university of Arizona) option was selected for carrying automatic optimization. The Nash-Sutcliffe efficiency was selected as objective function while calibrating the model. The derived calibration parameters with Nash coefficient 0.61 is shown in **Table 1**.

Table 1 Calibrated model parameter values at the Kiran River.

Parameter	Value
Base flow coefficient	0.45
Impervious threshold (mm)	0.51
Infiltration coefficient (mm)	252.00
Infiltration shape	2.50
Interflow coefficient	0.14
Pervious fraction	0.40
Rainfall interception store capacity (mm)	3.50
Recharge coefficient	0.80
Soil moisture store capacity (mm)	370.00

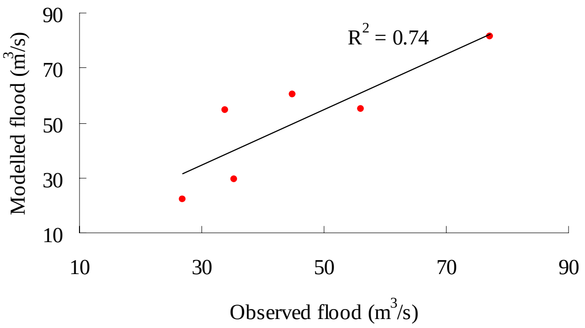

Fig.5 Model performance in the Jhanjhari River basin.

The model performance was checked by comparing the observed annual maximum discharge values and simulated values of the Jhanjhari River basin (station index 363) located in the same hydrologic region. The performance correlation between modeled and measured flow is shown in **Fig.5**. The plotted point corresponds to six years of annual maximum discharge values at the Jhanjhari River.

The physiographic characteristics of region 2 and 5 are very similar. Different regions were made only because of different rainfall pattern. At the time of carrying out the present study, streamflow data could not be made available for the stations in region 2. Therefore, for the time being, the same model parameter derived for region 5 is was used for synthetic annual maximum discharge generation purpose in region 2.

6. RELATIONSHIP ESTABLISHMENT

With the generation of synthetic annual maximum discharge data, there are finally six and five data basins in region 2 and 5 respectively. There are 7, 10 and 24 observed data basins in region 1, 3 and 4 respectively. Establishment of index flood relationship got started with computation of mean annual maximum discharge values at all the stations. The average value, called as observed index flood, was plotted against the corresponding drainage area (*A*) to carry out simple regression.

For better estimation of index flood, multiple regression technique needs to be applied. The additional attributes which were proposed in carrying multiple regressions are: average annual daily maximum rainfall (*P*), soil type, river slope (*S*) and land cover information; since these attributes may strongly govern index flood values. But, the annual daily maximum rainfall in different basins of regions 1, 3 and 4 could not be collected. Therefore, basin elevation (*H*) was considered instead of average annual daily maximum rainfall (*P*) because rainfall is largely governed by elevation in Nepalese context as mentioned in Section 2. Impact of land cover and soil type was considered in term of NRCS runoff curve number.

ArcHydro delineate basins and streams from digital elevation models, extract limited set of hydrologic modeling parameters from the DEMs in ArcGIS environment. HEC-geoHMS finds use in preparing additional parameters from the DEMs and other supported data layers, and write parameter files. For computing basin's average elevation (*H*), runoff curve number (*CN*) and river slope (*S*), the software ArcHydro and HEC-GeoHMS were used.

7. RESULTS AND DISCUSSION

Table 2 shows the index flood (μ_i in m^3/s) relationships for different hydrologic regions. It was obtained after analyzing the results of simple and multiple regressions. Except for Region 3, no other parameters than the drainage area (*A* in km^2) was found effective in governing index flood although the study performed multiple regressions for all the possible combinations of drainage area (*A*), river slope (*S*), average annual daily maximum rainfall (*P*), runoff curve number (*CN*) and average basin elevation (*H*) . The index flood relationships were found weakly correlated (smaller R^2) when considered the parameters: *CN*, *S* and *P*. This is because of similar values of *CN*, *S* and *P* throughout a particular hydrologic region as the previous work$^{3)}$ has delineated hydrological homogeneous regions based on similarity in *CN*, *S* and average monsoon rainfall pattern among the river basins,.

To assess predictive power of the derived relationship, the estimated index flood values were compared with observed index flood values. Nash efficiency coefficient (Eq.3) has been used to evaluate predictive power of the derived relationship.

Table 2 Index flood relationship for five hydrologic regions of Nepal.

Region	Regression relationship of index flood	R^2-value	Nash coefficient
1	$\mu_i = 0.71A^{0.99}$	0.75	0.56
2	$\mu_i = 0.99A^{1.06}$	0.97	0.99
3	$\mu_i = 0.18A - 5.94$	0.83	0.29
	$\mu_i = 0.18A - 0.22H + 732$	0.89	0.49
4	$\mu_i = 5.15A^{0.72}$	0.73	0.64
5	$\mu_i = 1.68A - 110.37$	0.90	0.91

The index flood (μ_i) relationship involving drainage area (A) has been presented for all the five hydrologic regions. In case of Region 3, an additional relationship involving elevation (H) and drainage area (A) is shown because the inclusion of H found improving the predictability considerably. The Nash coefficient (E) for N observation basins with average index flood ($\overline{Q}_o$) is (Eq. 3):

$$E = 1 - \frac{\sum_i^N (Q_o^i - Q_m^i)^2}{\sum_i^N (Q_o^i - \overline{Q}_o)^2} \qquad (3)$$

Here, Q^i_o and Q^i_m are observed index flood and modelled index flood in basin i respectively.

8. CONCLUSIONS

This study analyzed several indirect estimation methods of index flood. The first method was simple regression which is suitable for the basins having no other information than the drainage area. For the basins with additional information available, multiple regression techniques were applied for establishing better index flood relationships.

For the hydrologic regions with inadequate discharge data, hydrologic simulation was proposed for generating synthetic annual maximum discharge series. For this purpose, a conceptual rainfall-runoff model: SimHyd was successfully calibrated and validated. This model was used for estimating index flood by taking mean of the synthetic annual maximum discharge series. Nash efficiency coefficient was used to check their predictive accuracy.

In this way, the authors successfully developed index flood estimation tools for the regions with different conditions of data availability. The authors intend to use these index flood relationships in accomplishing the regional flood frequency analysis of Nepalese river basins in next step.

ACKNOWLEDGMENT: The authors would like to thank MEXT, Japan for providing financial support in carrying research work. The authors also wish to thank Visiting Prof. Ismail Abustan (Universiti Sains Malaysia) and all lab members for their helpful suggestions and comments at different instances.

REFERENCES

1) Rao A.R. and Hamed, K.H.: Flood Frequency Analysis, CRC press LLC, Florida, USA, 2000.

2) Kalma, J.D. and Sivapalan, M.: Scale issues in hydrological modeling, *Advances in Hydrological Processes*, John Wiley & Sons, 1995.

3) Mishra, B.K., Takara, K. and Tachikawa, Y.: NRCS curve number employed hydrologic homogeneous regionalization in regional flood frequency analysis, *Proceedings on International Conference on Hydro-science and Engineering*, Nagoya, Japan, 2008.

4) Dalrymple, T.: Flood frequency analysis, *US Geol. Survey Water Supply Paper 1543A*, pp. 11-51, 1960.

5) Bocchiola, D., De Michele, C. and Rosso, R.: Review of recent advances in index flood estimation, *Hydrology and Earth System Sciences*, Vol.7 (3), pp. 283-296, 2003.

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

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

8) Brath, A., Castellarin, A., Franchini, M. and Galeati, G.: Estimating the index flood using indirect methods, *Hydrological sciences-Journal*, Vol. 46(3), pp. 399-418, 2001.

9) Reddy, P.J.: Stochastic hydrology, Laxmi Publications (P) LTD, New Delhi, 1997.

10) 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.

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

(Received September 30, 2008)