FRIEND 2002—Regional Hydrology: Bridging the Gap between Research and Practice (Proceedings of the Fourth International FRIEND Conference held at Cape Town, South Africa, March 2002). IAHS Publ. no. 274, 2002.

Modelling of flood flows in Rufiji River basin, Tanzania

SIMON H. MKHANDI, RAJ KACHROO & FELIX MTALO

Water Resources Engineering Programme, University of Dar es Salaam, PO Box 35131, Dar es Salaam, Tanzania

e-mail: mkhandi@wrep.udsm.ac.tz

Abstract A simulation procedure applied to select a suitable statistical procedure to model flood flows in Rufiji River basin in Tanzania is presented. In this paper, a flood frequency procedure refers to the combination of a probability distribution with a suitable method of parameter estimation. A predictive ability test based on bias between observed and predicted flows is used to select the suitable procedure. Based on the results of the simulation, the most suitable statistical frequency procedures for modelling flood flows for the homogeneous regions of Rufiji basin seem to be log-Pearson type 3/method of moments (LP3/MOM) and Pearson type 3/method of probability weighted moments (P3/PWM). The two procedures gave the lowest bias values for different return periods and sample sizes considered in the study. The study is based on the data from 42 sites in the basin.

Key words Rufiji River; frequency analysis, homogeneous regions; regional homogeneity; graphical test; parent distribution; quantile estimate; statistical distribution

INTRODUCTION

The occurrence of extreme flow events has a direct impact on the socio-economic activities in a river basin. For example, when floods occur in the flood plain, human activities are often hampered. The interference may be quantified in terms of economic losses due to damage to the infrastructure, property and sometimes losses of life. As a result estimates of flood magnitudes have significant economic implications, besides being required for design of civil engineering works. Due to this link the methods used to estimate flood magnitudes are sensitive to the requirements of the decision-making process.

The use of regional information to estimate flood magnitudes at sites with little or no observed data has increasingly become important because many projects who require information on design floods are located in areas where actual observed flood data are either missing or inadequate. In this study, the regional flood frequency analysis (RFFA) approach is used to select suitable procedures to model flood flows for homogeneous regions in the Rufiji basin in Tanzania in order to improve the design and economic appraisal of civil engineering structures so as to minimize the risk of flood damage. The location of the Rufiji basin is shown in Fig. 1.

METHODOLOGY

Regionalization of the Rufiji basin

Regionalization refers to grouping of homogeneous sub-basins or regions in a basin. The purpose of grouping sub-basins in homogeneous regions is to achieve regional flood frequency relationships with lower standard errors as compared to those for the combined area (Kite, 1977). In this study, statistically homogeneous regions, i.e. those regions likely to have similar flood-producing characteristics, in the Rufiji basin were identified based on geographical information. The procedure involved superimposing the digital basin boundary map, the mean annual rainfall and the relief map using computer graphics software. The location of each gauging station was then plotted on the resulting composite map. From the composite map, regions that seemed to be homogeneous based on similar topography, elevation and mean annual rainfall were identified.

Once a homogeneous region was identified, a preliminary check was carried out to find out whether the proposed region was reasonably homogeneous. This involved computing a statistic C_c, known as the coefficient of variation of the at-site L coefficients of variation. A homogeneous region was expected to have a low value of C_c. The proposed region was further subdivided and the boundaries of the proposed region redefined if the value of C_c was found to be more than 0.3. This value was considered a reasonable upper limit to consider regional homogeneity. The next step was to test the region on homogeneity in a statistical sense. The graphical test proposed by Kachroo *et al.* (2000) was used to test the homogeneity of the identified regions.

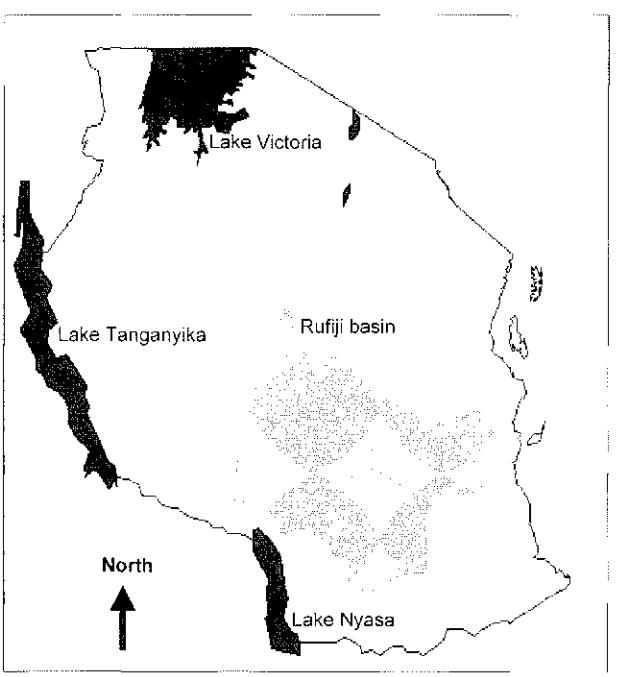


Fig. 1 Location of the Rufiji basin.

Selection of suitable statistical distribution to model flood flows

For each identified homogeneous region a suitable statistical distribution was selected to model the flood flows based on the predictive ability test (Cunnane, 1986; 1989). This test investigates how well a distribution and its associated method of parameter estimation can estimate the flood quantile (Q)–return period (T) relationship when the population distribution is not identical to that of the proposed distribution. A Monte-Carlo simulation procedure was used to identify a suitable flood frequency procedure for each of the delineated homogeneous regions. The simulation procedure comprised the following steps:

(a) Estimation of population statistics from weighted regional average standardized statistics ($\overline{Q}$, C_v and C_s).
(b) Generation of a random sample of specified length from parent population whose moments are the same as those calculated above (step (a)). Extreme value type 1 (EV1), Pearson type 3 (P3), gamma (G2), generalized extreme value (GEV) and lognormal (LN) distributions were chosen as parent distributions.
(c) Estimation of a quantile $\hat{Q}_T$ from the sample generated in step (b) for a given frequency procedure. A total of 14 commonly used flood estimation procedures were considered.
(d) Repeating steps (b) and (c) 1000 times to calculate the average value of $\hat{Q}_T$.
(e) Comparison of the population quantile Q_T with the average quantile $\hat{Q}_T$ estimated in step (d) for bias.
(f) Repeating steps (c) to (e) for the various frequency procedures considered.
(g) Repeating steps (b) to (f) for the various parent distributions.
(h) Calculation of average bias (difference between estimated and population flow peak) for various procedures. Averaging is done over a number of assumed parent populations.
(i) Ranking in terms of bias values to determine the performance of the different estimation procedures.

The procedure that gives the lowest bias is better as far as the prediction of flood quantiles is concerned.

DATA

In the study area, instantaneous annual maximum discharge data were available from 13 out of 42 recording gauging stations in the Rufiji basin. In order to incorporate more data in the analysis, available maximum mean daily discharges from non-recording gauging stations were used to extend the instantaneous data by establishing a linear relationship between instantaneous and maximum mean daily discharges from the automatic recording stations. The established relationship is given below:

$$Q_{\text{inst}} = 1.15 Q_{M\text{max}} \tag{1}$$

where Q_{inst} is instantaneous maximum discharge, and $Q_{M\text{max}}$ is maximum mean daily discharge.

Assuming that the above linear equation also applies to the corresponding discharges for the non-recording gauging stations, the maximum mean daily discharges were converted to instantaneous maximum discharges at the non-recording gauging stations. This procedure resulted in the increase of the number of instantaneous maximum discharge data series from 13 to 42 with an average length of 22 years. The data from the 42 stations was checked for quality using conventional statistical techniques. Those sites which showed inconsistency and dependence structure in the data series were excluded from further analysis. A total of 28 stations were finally selected for analysis.

RESULTS

Regionalization

The Rufiji basin was divided into four homogeneous regions considering the sub-basins where flood data were available. The sub-basins where data were not available

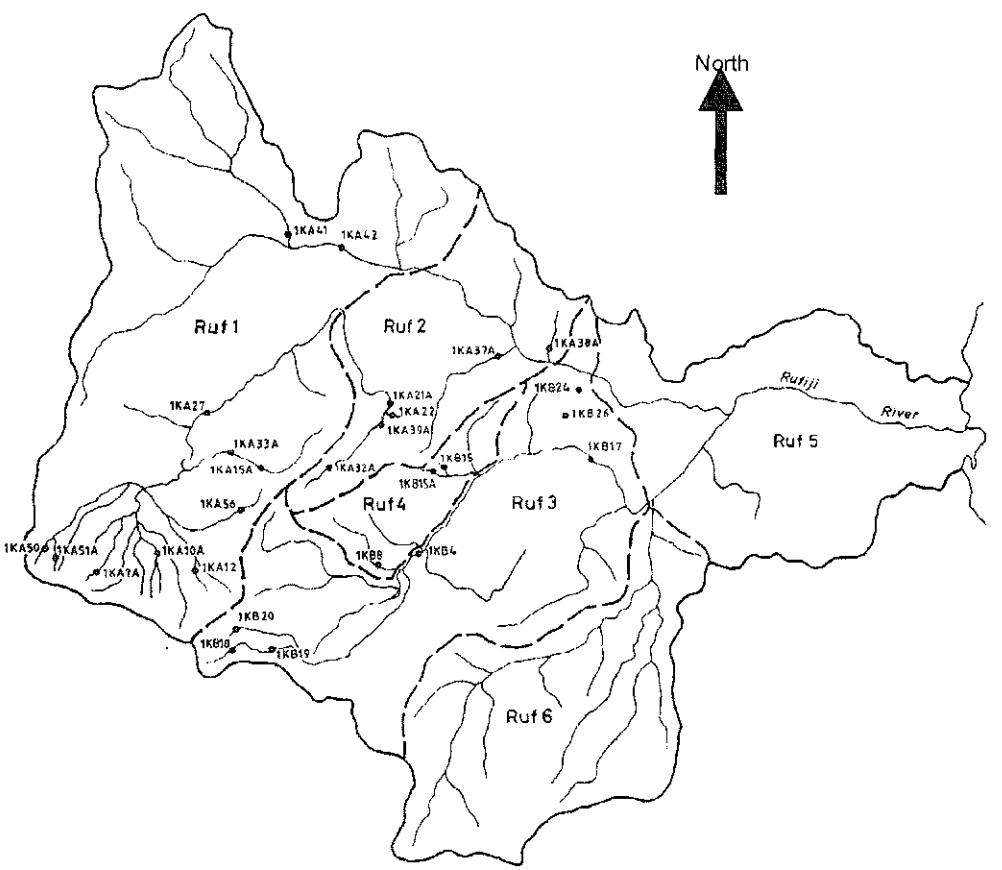


Fig. 2 Delineated homogeneous regions in the Rufiji River basin.

Table 1 Flood statistics for delineated regions.

Region	Sites	Years	C_v	C_s	C_k	C_c
Ruf1	11	22	0.56	1.08	4.05	0.23
Ruf2	6	27	0.40	1.04	5.50	0.23
Ruf3	7	16	0.44	0.53	3.13	0.15
Ruf4	4	15	0.20	−0.47	2.94	0.26

were excluded from the analysis. The delineated regions are presented in Fig. 2. The figure also shows the location of stations used in the analysis. The four delineated regions comprise of the following river sub-basins: Ruf1 (Great Ruaha); Ruf2 (Little Ruaha); Ruf3 (Ruhuji and Kilombero) and Ruf4 (Mpanga, Mngeta and neighbouring tributaries). Table 1 presents selected statistics of the data from the delineated regions, i.e. coefficient of variation (C_v), skewness (C_s), kurtosis (C_k) and coefficient of variation of C_v (C_c).

Selection of suitable flood frequency procedures

A total of 14 flood frequency procedures were considered in the simulation. The procedure which gave the lowest bias for a given return period and sample size was selected as the most suitable for quantile estimation for a given region. Selected probability distributions with the associated method of parameter estimation for the delineated regions are as follows: for the Ruf1 region the extreme value type 1 (EV1) with probability weighted method (PWM) of parameter estimation was found to be suitable for a small sample size (n = 15) and for larger samples (n = 30 and 50) the Pearson type 3 (P3) with the PWM was most appropriate. In the case of the Ruf2 region the EV1 with the method of moments (MOM) was found to be most adequate for small samples and the log-Pearson type 3 (LP3) with the MOM for larger samples. For the other two regions, Ruf3 and Ruf4, the LP3 with the MOM was suitable for the selected samples sizes (n = 15, 30 and 50).

The regional statistics used to fit the selected procedures in the four regions are presented in Table 2. The fitted procedures were used to compute standardized quantile estimates, i.e. ratios of quantile estimate (Q_T) divided by mean annual flood (MAF) for the return periods of 2, 5, 10, 20, 25, 50, 100, 200 and 500 years. The obtained results were plotted to produce the frequency curves for the four regions (Fig. 3). The slopes of the derived frequency curves differ from one region to the other depending on the variability of the flood regimes. A steep frequency curve reflects a highly variable

Table 2 Regional statistics estimated with the selected flood frequency procedures.

Region	Procedure	Parameter estimates:		
		Qmean	Qstd	Qskew
Ruf1	P3/PWM	1.00	0.59	1.17
Ruf2	LP3/MOM	−0.09	0.46	−0.58
Ruf3	LP3/MOM	−0.12	0.53	−0.69
Ruf4	LP3/MOM	−0.02	0.21	−0.54

Qmean = location parameter; Qstd = scale parameter; Qskew = shape parameter.

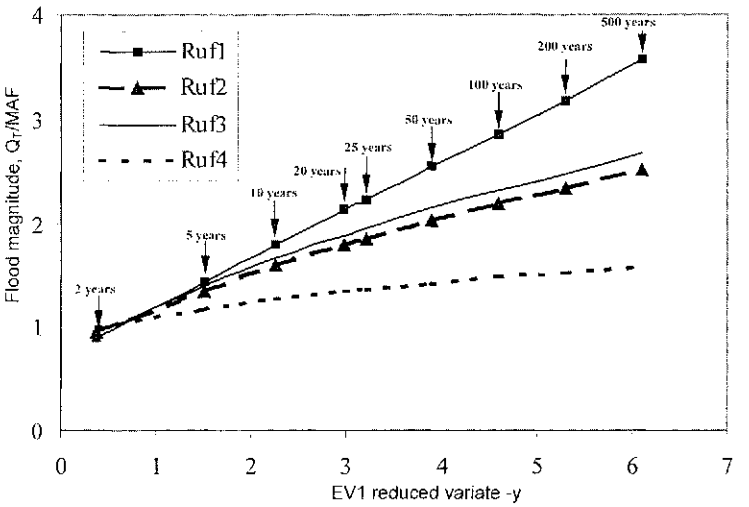


Fig. 3 Frequency curves for the four delineated homogeneous regions in the Rufiji basin.

flood regime in a region. The variability of the flood regime depends to a large extent on the meteorological phenomena generating the flood events.

CONCLUSION

The simulation study in the Rufiji basin showed that the LP3/MOM and the P3/PWM seem to be the most suitable flood frequency procedures to model flood flows in homogeneous regions. The two procedures gave the lowest bias values for different return periods and sample sizes considered in the study.

REFERENCES

Cunnane, C. (1986) Review of statistical models for flood frequency estimation. In: *Hydrologic Frequency Modelling* (ed. by V. P. Singh) (Int. Symp. on Flood Frequency and Risk Analysis, Louisiana, USA, May 1986), 49–95. Reidel, Dordrecht.

Cunnane, C. (1989) Statistical distribution for flood frequency analysis. *WMO Operational Hydrology Report no. 33. Geneva, Switzerland.*

Kachroo, R. K., Mkhandi, S. H. & Parida, B. P. (2000) Flood frequency analysis of southern Africa: I. Delineation of homogeneous regions. *Hydrol. Sci. J.* **45**(3), 437–448.

Kite, G. W. (1977) *Frequency and Risk Analysis in Hydrology*. Water Resources Publications, Colorado, USA.