Flood Frequency Analysis of Annual Maximum Stream Flows for Luvuvhu River Catchment, Limpopo Province, South Africa

L.R. Singo[1], P.M. Kundu[1], J.O. Odiyo[1], F.I. Mathivha[1] and T.R. Nkuna[1]
[1]*University of Venda, Department of Hydrology and Water Resources,*
Thohoyandou, South Africa

Abstract Luvuvhu River Catchment (LRC) is one of the regions in South Africa where floods have caused enormous damage and impacted negatively on fauna and flora. Over the years, the catchment has experienced floods resulting from heavy rainfall associated with the ITCZ. Annual maximum flow data from 8 stations with 50 years hydrological data were used to analyze flood frequencies in the catchment. To derive the probability of occurrence of flood events, the frequency distributions which could best describe the past characteristics and magnitudes of such floods were tested. This involved the determination of the best flood frequency models, which could be fitted to the available historical recorded data. The distribution models used included the Generalized Extreme Value, Gumbel or Extreme Value type 1, Log-Normal and the Log Pearson type III distributions. The extreme value analysis showed that the Gumbel and Log Pearson type III distributions provided the best fit.

Key words: Catchment, Flood frequency, ITCZ, Peak discharge

INTRODUCTION

Luvuvhu River Catchment is one of the regions in South Africa where floods have caused enormous damage to both property and life and impacted negatively on fauna and flora. Over the years, the catchment has experienced floods resulting from heavy rainfall associated with the Intertropical Convergence Zone (ITCZ). Risk of floods has been felt in the past, with the major occurring during the 1999/2000 rainy season. Meteorologists and weather observers from the South African Weather Services (SAWS) have managed to develop some mitigation strategies such as tracking the paths of ITCZ and providing early warnings in time when heavy rainfalls are anticipated in the study area using radar and satellite imageries.

Attempts to estimate flood peaks using frequency models have been made in Luvuvhu River Catchment. This study estimated flood peaks using annual maximum discharge data. A study by Van Zyl (2006) noted that climatic extremes, especially excessive rainfalls, were the most common cause of floods in South Africa. They vary from semi-predictable seasonal rains over wide geographic areas, which give rise to the annual wet-season floods in tropical areas, to almost random convectional storms over small basins. Prolonged rainfall over large drainage basins is also associated with tropical cyclones or the intense depressions of mid-latitudes. Duaibe (2006) noted that a number of human induced factors contribute to floods in a catchment. These include factors such as land degradation; deforestation of catchment areas; increased population density along river banks; inadequate land use planning; zoning and control of floodplain development; inadequate drainage, particularly in cities, and inadequate management of discharge from river reservoirs.

The South African National Disaster Management Center estimated that at least 50 000 people, and possibly more than 100 000 are living along rivers and streams in South Africa below levels reached by previous floods. Because of their proximity to the river banks, they are at risk of flooding in years of above average floods. The need to accurately predict extreme flood events is imperative in designing for not only the safety of infrastructure, but also people's lives.This helps provide warnings to riparian users of Luvuvhu River Catchment when heavy rains are anticipated to occur to avert damage.

A number of frequency distribution models have been used in the past for hydrologic frequency determination. Though several probability models have been developed to describe the frequency distribution of extreme hydrologic events, major problems arise when selecting the best method to use since there is no general agreement as to which distribution, or distributions, that should be used for the frequency analysis of extreme hydrologic events. Therefore, the selection of an appropriate model depends mainly on the characteristics of available data at the particular site (Olofintoye *et al*., 2009). In flood frequency analysis, statistical and deterministic models are commonly used to estimate future floods. The statistical approach uses probabilistic methods to model flood events, while the deterministic approach requires the use of physically-based approach. Of the two approaches, statistical probability methods remain the most widely used in the scope of hydrology (Helsel and Hirsch, 2010). A statistical estimate based on a flood record for the stream of interest is typically the preferred method since it is most closely based on floods that have actually occurred on the stream (NOAA, 2011).

The study area

Luvuvhu River Catchment shown in Figure 1, located between latitudes 22°17'33.57''S and 23°17'57.31''S and longitudes 29°49'46.16''E and 31°23'32.02''E in the extreme northeastern corner of South Africa. The river rises as a steep mountain stream in the southeasterly slopes of the Soutpansberg Mountain range, flows through Kruger National Park, and empties into the Limpopo River at the border with Mozambique and Zimbabwe. The catchment covers an area of approximately 5941 km^2 and is situated on a plateau of 1 312 meters above sea level (DWAF, 2002). The area receives one cycle of rainfall that extends from October of the previous year and ends in April of the following year; wherein the dry season runs from May to September.

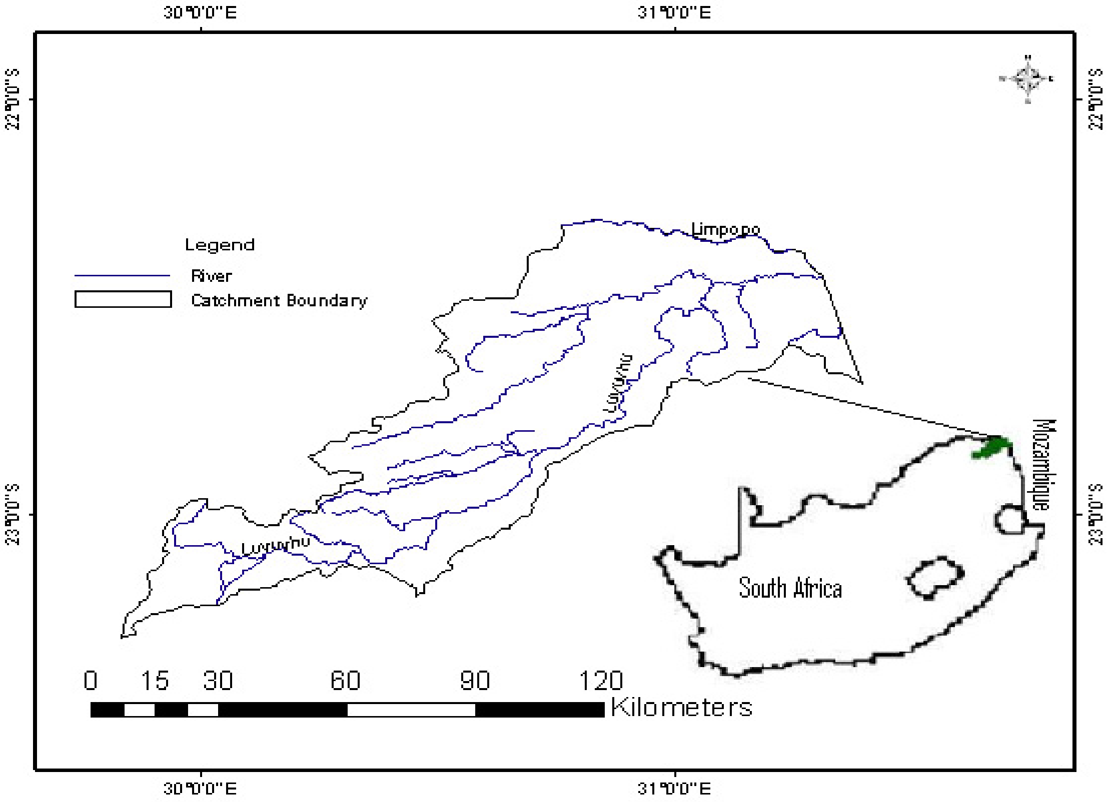

Figure 1: The study area

MATERIALS AND METHODS

The governing equations within the EasyFit software version 5.5 (2011) included equations 1 to 7. The data for all 8 streamflow stations was combined to produce a single data set representing the study area. The distribution with the highest frequency was selected as the best distribution for a particular region. In this study, the LP3 and EV1 probability distributions were selected as the best valid distributions to model flood probability in the catchment. The Weibull's formula was applied to streamflow data to calculate the exceedance probability and the return period. The frequency modeling for seven different flood levels, the

2-, 5-, 10-, 25-, 50-, 100- and 200-year floods, was performed for flood prevention and protection in the catchment.

Flood Frequency Models

EasyFit software version 5.5 (2011) was used to analyse flood frequency in the catchment. Sequential steps used to estimate the floods of given recurrence intervals involved selecting the highest peak flood in a hydrological year. The highest peaks in each hydrological year were arranged in descending order of magnitudes and ranks were then assigned. Within a single data series, some discharge values occurred more than once and were given the same rank. The return period of each of the ordered value and the probability of each event being equaled to or exceeded were computed using the Weibull's method. Four frequency distributions including the Generalized Extreme Value distributions (GEV), the Gumbel's or Extreme Value type 1 distribution (EV1), the Log-normal distribution and the Log-Pearson type **III** distribution (LP3) were compared for flood estimation.

Generalized Extreme Value distributions (GEV) Model

The probability distribution function for the GEV is given by equation 1:

$$F(x) = \exp\left[-\left(1-k\frac{x-u}{\alpha}\right)^{\frac{1}{k}}\right] \qquad (1)$$

where, *u* is the location parameter, *α* is the scale parameter, and *k* is the shape parameter. Hosking (1977) approximated that the design flood at a desired return period is given by equation 2:

$$Q_t = u + \left(\frac{\alpha}{\kappa}\right)\left\{1-(-\log\left(\frac{T-1}{T}\right))^{k}\right\} \qquad (2)$$

where: *T*= desired return period

$$k = 7.8590 + 2.9554c^2$$

$$c = \frac{2}{3+\tau_3} - \frac{\ln 2}{\ln 3}$$

$$\alpha = \frac{\lambda_2 k}{(1-2^{-k})\Gamma(1+k)}$$

$$u = \lambda_1 - \alpha\{1-\Gamma(1+k)\}/k$$

where: $\Gamma =$ gamma function, τ_3 = L-skewness coefficient and λ_1 and λ_2 = first and second L-moment approaches. The τ_3, λ_1 and λ_2 were derived from the Probability Weighted Moments (PWMs) equations following Cunnane (1989).

The location parameter (*u*) describes the shift of a distribution in a given direction on the horizontal axis; the scale parameter (*α*) describes how spread out the distribution is, and defines where the bulk of the distribution lies. As the scale parameter increases, the distribution will become more spread out. The shape parameter (*κ*) strictly affects the shape of the distribution and governs the tail of each distribution.

The shape parameter is derived from skewness, as it represents where the majority of the data lies, which creates the tail(s) of the distribution.

Gumbel's or Extreme Value type 1 distribution (EV1) Model

Gumbel (1941) introduced EV1 as a concept of extreme value distribution and as a model for prediction of hydrologic events such as flood peaks. The probability of occurrence of an extreme event, equal or larger than a value, is given by equation 3:

$$P(X \geq x_0) = 1 - e^{-e-y} \qquad (3)$$

where: P = the probability of occurrence, X = the event of the hydrologic series, x_0 = the desired value of the event and y = the reduced variate: $y = \alpha(x - a)$

where: x = the variate value, $a = \bar{x} - 0.45005\sigma_x$ and $\alpha = 1.2825/\sigma_x$

σ_x = standard deviation of variate X and $\bar{x}$ = mean of the variate X

The reduced variate is calculated as:

$$y = \frac{1.2825\left(x - \bar{x}\right)}{\sigma_x} + 0.577$$

Where values 1.2825 and 0.577 are the constants for reduced mean and reduced variate, respectively. The estimated peak discharges in Gumbel's distribution at each return period are computed as following equation 3. The frequency factor (*K*) of the Gumbel's distribution is computed using the following formula:

$$K = \frac{y_T - \bar{y}_n}{\sigma_n}$$

where: K = the frequency factor, y_T = the value of peak discharge for a given recurrence interval, y_n = Gumbel's reduced mean and σ_n = Gumbel's reduced standard deviation.

$$y_T = -In\left[In\left(\frac{T_r}{T_r - 1}\right)\right]$$

Log-Normal distribution Model

In log-normal distribution, the data is said to be normally distributed with the mean and standard deviation. A standard normal distribution table is then used to select normal deviate (Z) to be used in a log-normal frequency analysis. Graphical estimates of either flood magnitudes or probabilities can also be computed using the following formulas:

$$Z = \frac{Y - \bar{Y}}{S_y} \qquad (4)$$

where Z = standard normal distribution, Y = value of random variable, $\bar{Y}$ = the mean and S_y = standard deviation

Log-Pearson type III distribution (LP3) Model

The LP3 distribution is a member of the family of Pearson Type 3 distributions, and is also referred to as the Gamma distribution. In LP3 method, the rearranged annual maximum flow data is first transformed to logarithms discharges of base 10. The series mean, standard deviation and the skewness co-efficient are then computed to determine estimated discharges for a given recurrence interval or exceedance probability for a specific event. The skewness coefficient is an important hydrological characteristic which gives a measure of shape of a sampling distribution, given by equation 5:

$$G = \left[\frac{n^2\left(\sum x^3\right) - 3n\left(\sum x^2\right)\left(\sum x\right) + 2\left(\sum x\right)^3}{n(n-1)(n-2)\sigma_x{}^3} \right] \qquad (5)$$

where: G = skewness coefficient σ_x, X_{av}, $\sum x$ and n are as above

The estimated discharge values for a given period can be evaluated using the logarithm of the design flood given by the following formula:

$$X_T = \log Q_T = X_{av} + K\sigma_x \qquad (6)$$

Where: Q_T = the discharge for the estimated T-year return period, K = the probability factor based on n-years recurrence interval which can be determined using a frequency factor table for Gamma and log-Pearson Type III Distributions (Haan, 1977), X_{av} = the mean of the logarithms of annual peak flows at the streamflow-gaging station and X_T, σ_x = the standard deviation about the mean of the logarithms of annual peak flows.

The design flood itself is given by:

$$X_T = anti\log Q_T = 10^{Q_T} \qquad (7)$$

Goodness of fit measures

The goodness fit measure involves identifying a distribution that fits the observed data. When computing the magnitudes of extreme events, such as flood flows, it is required that the probability distribution function be invertible, so that a given value of recurrence interval (T_r) and the corresponding value of frequency factor (K) can be determined. Four plotting position formulae were adopted and compared to select the best flood frequency distribution that best fitted the annual maximum flood flow of Luvuvhu River catchment.

RESULTS AND DISCUSSION

The highest measured flow of 106.72 m^3/s was recorded during the 1976/77 hydrological year, while the lowest flood flow of 0.859 m^3/s was recorded in 1991/92. A 5-year moving average was smoothed in the data to highlight significant changes in the trend. The smoothed trend showed significant hydrological conditions which may suggest that erratic rainfall and catchment characteristics such as land use may have caused flood peaks to increase in the area. The 50-year mean instantaneous flood flow area was 20.83 m^3/s with a standard variate of 20.20 m^3/s and a mean coefficient of variability (*Cv)* of 0.969. The coefficient of variation was applied in the data to measure the consistency and the steepness for the frequency curves in streamflow data. The *Cv* value obtained indicated that the distribution of flood flows was highly variable. The variability in flood flow may be attributed to varying hydrological phenomena responsible for generating the flood events in LRC. Figure 2 shows the 50- year annual maximum flow for the area.

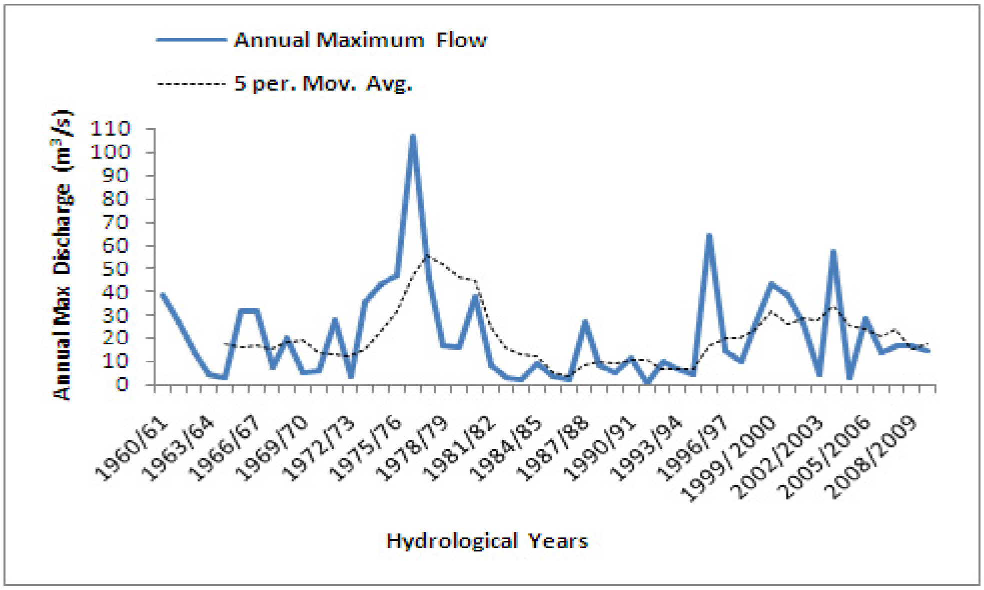

Figure 2: Annual maximum flow for the study area

The probability density functions (PDF) for the four tested distributions shown in Figure 3 showed that GEV, EV1, Log-Normal and LP3 distributions were most likely to best fit the data. The histogram of annual maximum flood data which revealed a positive skewed distribution showed a unimodal distribution which was skewed to the right. Clearly the PDF showed that the Log normal and LP3 exhibited similar probability densities which were different from that of the GEV and EV1 distributions. The cumulative distribution function (CDF) (Figure 3b) showed the non-exceedence probability for a given magnitude.

The probability-probability plot (Figure 3c), which is a graph of the empirical CDF values plotted against the theoretical (fitted) CDF values was used to determine how well a specific distribution fitted to the observed data. It is recommended that if the maximum absolute difference is less than 0.05 (or 5%) the fit can then be considered good.

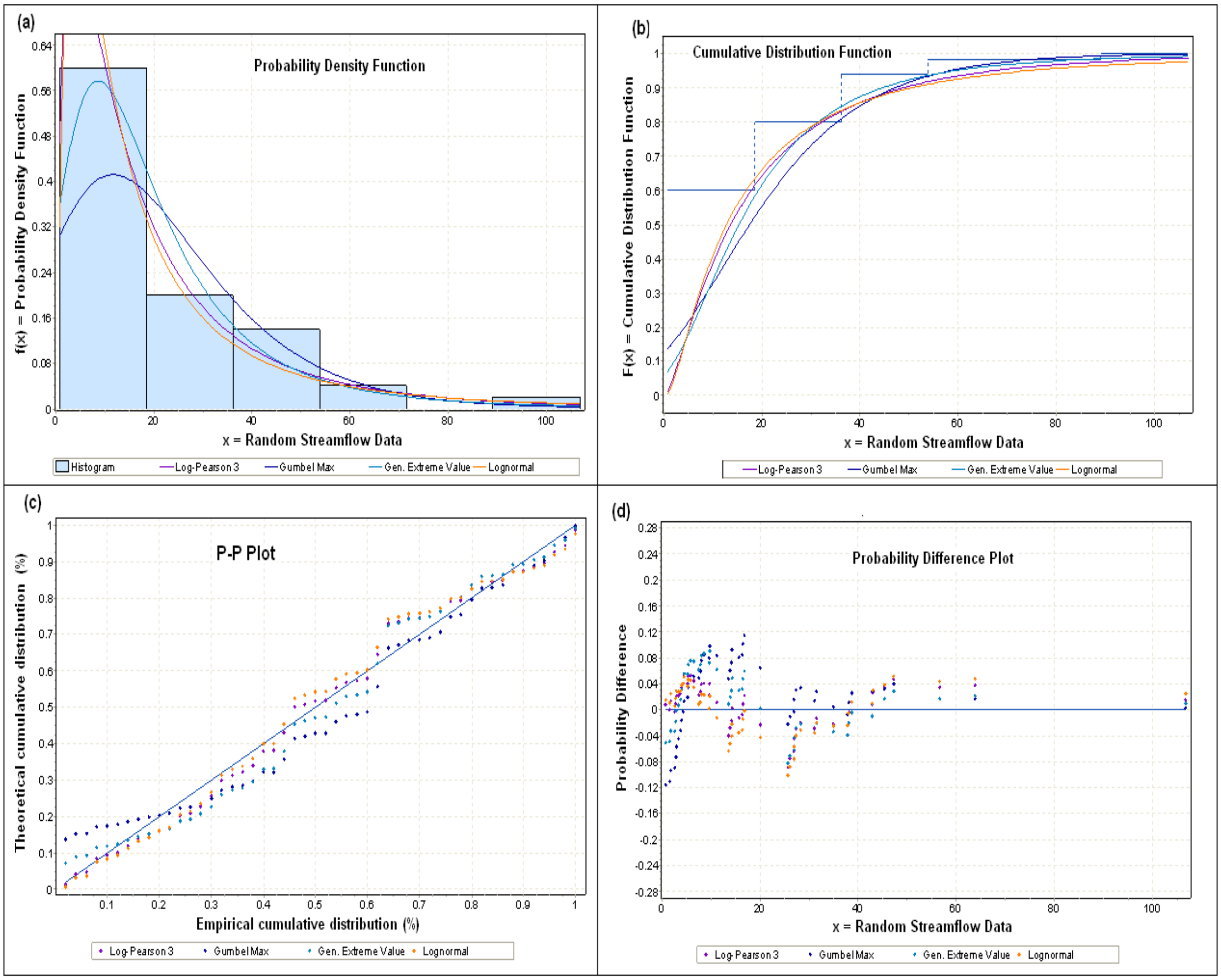

Figure 3: (a) Probability density functions, (b) cumulative distribution functions, (c) probability-probability plots and (d) probability difference for the four compared frequency distributions.

Estimated Peak flows

Flood peaks corresponding to return periods of 2, 5, 10, 25, 50, 100 and 200 years were estimated for flood prevention and protection of risks in the catchment. Each computed flood magnitude was determined at 95-percent confidence interval. This interval was the range that contained the true flood magnitude for a particular exceedence probability. The parameters for the four distributions are shown in Table 1. In general, a distribution with a larger number of flexible parameters, would be able to model the input data more accurately than a distribution with a lesser number of parameters (Cunnane, 1989). In this case, the LP3 and EV1 showed larger numbers of flexible parameters, and were recommended for estimating flood risks in the catchment. The estimated discharges using GEV, EV1, Log-Normal LP3and distributions are shown in Tables 2, 3, 4 and 5 respectively.

Table 1: Estimated parameters for the four distributions

	Distribution	Parameters
1	Gen. Extreme Value	k=0.22411 σ =11.52 μ =10.939
2	Gumbel	σ =15.753 μ =11.739
3	Log-Pearson 3	α =30.839 β =-0.19511 γ =8.5669
4	Lognormal	σ =1.0726 μ =2.5497

Table 2: Computed Discharges for GEV distribution

Return Period (Years)	**Exceedence Probability (%)**	**Non-Exceedence probability (%)**	**GEV Parameters**			**Estimated Discharge (m^3/s)**
			k	σ	μ	
2	50	50	0.22411	11.52	10.939	62.275
5	20	80	0.22411	11.52	10.939	62.321
10	10	90	0.22411	11.52	10.939	62.332
25	4	96	0.22411	11.52	10.939	62.338
50	2	98	0.22411	11.52	10.939	62.340
100	1	99	0.22411	11.52	10.939	62.341
200	0.5	99.5	0.22411	11.52	10.939	62.342

Table 3: Computed discharges for EV1 distribution

Return Period (Years)	**Exceedence Probability (%)**	**Non-Exceedence probability (%)**	**Mean of Series**	**STDEV of Series**	**Gumbel's Reduced Mean**	**Gumbel's Reduced STDEV**	**k-value**	**Estimated Discharge (m^3/s)**
2	50	50	20.83	20.20	0.5485	1.161	-0.156	17.679
5	20	80	20.83	20.20	0.5485	1.161	0.819	37.374
10	10	90	20.83	20.20	0.5485	1.161	1.466	50.443
25	4	96	20.83	20.20	0.5485	1.161	2.276	66.805
50	2	98	20.83	20.20	0.5485	1.161	2.888	79.168
100	1	99	20.83	20.20	0.5485	1.161	3.498	91.489
200	0.5	99.5	20.83	20.20	0.5485	1.161	4.093	103.509

Table 4: Computed discharges for Log-Normal distribution

Return Period (Years)	**Exceedence Probability (%)**	**Non-Exceedence probability (%)**	**Mean of Logarithms**	**STDEV of Logarithms**	**Z-value**	**Estimated Discharge (Q) (m^3/s)**	**Estimated Discharge (10^Q) (m^3/s)**
2	50	50	1.107	0.471	0.0000	1.107	12.794
5	20	80	1.107	0.471	0.8416	1.50339	31.871
10	10	90	1.107	0.471	1.2816	1.71063	51.361
25	4	96	1.107	0.471	1.7507	1.93158	85.424
50	2	98	1.107	0.471	2.0538	2.07434	118.669
100	1	99	1.107	0.471	2.3264	2.20273	159.490
200	0.5	99.5	1.107	0.471	2.576	2.3203	209.072

Table 5: Computed Discharges for LP3 distribution

Return Period (Years)	Exceedence Probability (%)	Non-Exceedence probability (%)	Mean of Logarithms	STDEV of Logarithms	Skew	K-value	Estimated Discharge (Q) (m^3/s)	Estimated Discharge (10^Q) (m^3/s)
2	50	50	1.107	0.471	-0.4	0.066	1.1384	13.75
5	20	80	1.107	0.471	-0.4	0.855	1.5097	32.33
10	10	90	1.107	0.471	-0.4	1.231	1.6866	48.60
25	4	96	1.107	0.471	-0.4	1.606	1.8631	72.96
50	2	98	1.107	0.471	-0.4	1.834	1.9704	93.41
100	1	99	1.107	0.471	-0.4	2.029	2.0621	115.37
200	0.5	99.5	1.107	0.471	-0.4	2.201	2.1431	139.03

The results showed that the GEV distribution performed poorly whereas EV1, Log-Normal and LP3 distributions provided good estimates for floods in the LRC. The coefficient of determinations at each distribution showed statistically significant relation between estimated flood peaks and assigned return periods. The estimated flood values are useful in the engineering design of hydraulic structures in the catchment.

CONCLUSION

Flood-peak discharge magnitudes and frequencies at gauging stations were developed to give the annual maximum series (AMS). The 50-year annual peak-flood data from 8 streamflow and rainfall stations in LRC were used to define the magnitude and frequency of floods. Four probability distribution models, including, the GEV, EV1, Log-Normal and LP3 distributions were compared for modeling annual maximum series for Luvuvhu River catchment using EasyFit software. Peak flood magnitudes for the 2-, 5-, 10-, 25-, 50-, 100- and 200-year recurrence intervals were estimated using the four distributions which were then compared to select the best estimate for the study area.

REFERENCES

Cunnane, C. (1989) Statistical Distributions For Flood Frequency Analysis. *Operational Hydrology Report* no. 33, World Meteorological Organization.

DWAF (2002) A proposed National Water Resources Strategy for South Africa. Department of Water Affairs and Forestry, South Africa

Duaibe, K. (2009) Human activities and flood hazards and risks in the south west Pacific: a case study of the Navua catchment area, Fiji Islands. Published MSc thesis, University of Wellington, New Zealand.

Gumbel, E.J. (1941) The return period of flood flows. Ann. Math. Statist, 12(2), 163-190.

Haan, C.T. (1977) Statistical methods in Hydrology, Iowa State University Press, Ames, Iowa, USA.

Helsel, D.R., Hirsch, R.M. (2010) Statistical Methods in Water Resources. U.S. Geological Survey, Investigations Book 4, Chapter A3. U.S. Geological Survey.

Hosking, J.R.M. (1990) L-moments: Analysis and estimation of distributions using linear combinations of order statistics, *J. R. Stat. Soc.*, *Ser. B*, 52, 105–124.

MathWave Technologies (2011) EasyFit Software Version 5.5, USA

NOAA (2011) Flood Frequency Estimates for New England River Restoration Projects: Considering Climate Change in Project Design. US.DOC/NOAA fisheries services, USA.

Olofintoye, O.O., Sule, B.F., Salami, A.W. (2009) Best–fit Probability distribution model for peak daily rainfall of selected Cities in Nigeria**.** New York Science Journal, 2(3), 1-12.

Van Zyl, K. (2006) A study on a Disaster Risk Management plan for the South African Agricultural Sector. AGRI SA, NAFU SA, TAU SA, Pretoria.