ESTIMATED OF LOCAL MAXIMUM FLOW IN WATERSHED OF ZAMBEZE

F. E. Banze[1] and H. D. O. A. Fill[1&2]

1. Estudante de Mestrado da UFPR, Programa PPGERHA – email: felixbanze@gmail.com

2. Professor emérito da UFPR/DHS – email: heinzfill@yahoo.com - Curitiba-Paraná

ABSTRACT: information about maximum local flow is important to control and mitigate flooding in a given area or watershed or by hydraulic structures. The present study aimed to calculate the maximum flow for different return periods (2 to 10000 years) in the Zambeze watershed using probability distributions recommended in the literature. Annual natural maximum flows of a period of 58 years were used. The data were obtained at E-320 station, located in Mozambique (Tete province). To estimate the flood flows the following probability distributions were used: Log -Normal, Log - Pearson type III, generalized extreme value (GEV), Gumbel and Exponential. In order to test the fit of these distributions the Kolmogorov Simirnov test was used. The methods of L-Moments (MML) and Moments (MoM) were used to determine the parameters. The results allow to conclude that the maximum flow rate for the given return periods ranges from 5210 to 30200 $m^3 s^{-1}$ and all distributions used were suitable for the estimation of maximum flows, however the distribution that provided the best fit was Gumbel.

Key words: Maximum Flow, Parameter estimation, Probability distributions.

RESUMO: Informações sobre a vazão máxima local são importantes para controlar e atenuar cheias em uma determinada área ou bacia hidrográfica, ou no dimensionamento de obras hidráulicas. O presente trabalho teve como objetivo o cáculo da vazão máxima para diferentes períodos de retorno (2 a 10000 anos) na bacia do Zambeze utilizando distribuiçoes de probabilidade descritas na literatura. Para alcançar o objetivo foram utilizados vazões naturais máximas anuais de um período de 58 anos obtidos na estação E-320, localizada na província de Tete em Moçambique. Para a estimação das vazões maximas foram usadas as distribuições de probabilidade Log-Normal, Log-Pearson tipo III, Generalizada de Valores Extremos (GEV), Gumbel e Exponênsial. Para a verificação do ajuste das distribuições foi usado o teste Kolmogorov Simirnov e para a determinação dos parâmetros foram usados os métodos dos Momentos L (MML) e dos Momentos (MoM) . Os resultados encontrados permitiram concluir que, a vazão máxima para os períodos de retorno em estudo varia de 5210 $m^3\ s^{-1}$ a 30200 $m^3\ s^{-1}$ e todas as distribuições utilizadas são viáveis para a estimação da vazão máxima, contudo a distribuição de Gumbel forneceu o melhor ajuste.

1. INTRODUCTION

Floods are natural or artificial disasters causing major economic damage and often loss of human life. Usually for planning and design purposes of hydraulic structures the maximum river discharge associated to a given risk of being equaled or exceeded is used in flood forecasting and design hydraulic structures. Thus, information about maximum flow is important to control and mitigate flooding in a given area or watershed, and in hydraulic structures, rural or urban drainage, suchas dams, canals, irrigation perimeters, dykes, spillways and others. The correct estimate of this value has decisive importance in the cost and safety of engineering projects. (Tucci, 2001).

One of the most common problems in hydrology is the estimation of a flood associated with a given probability, from a short record of flow (Floods Frequency Analysis). For this purpose different types of distributions have been used in hydrology. One source of error is that it is unclear which distributions best represents the phenomenon, There is no sufficient firm theory for using one or other distribution; however, there are theoretical or practical justifications for the application of some families of distributions that could provide reasonable results. A second source of error is the Parameter estimation (Kite, 1978), where different methods are available each one with advantages and disadvantages.

However, according to (Fill, 2000), there are some criteria for choosing the family of distribution that could provide better fit, and they are:

(1) Knowledge of the nature of the underlying random experiment.

(2) Using of general theorems facing asymptotic results

(3) Best fit of the sample

(4) Robust distributions (distributions providing reasonable results).

(5) Customs and traditions

For the verification of best fit (criterion three) there are available statistic tests. Among the best the most known are: Chi- square, Kolmogorov-Smirnov and Filliben tests. In applying an statistical test it is necessary to establish a probability of type I error, chose as a rule between 1 to 5% (Naghettini and Pinto, 2007).

The Kolmogorov Smirnov test (KS) is a nonparametric statistical test, which is based on the maximum difference between the cumulative probability functions, empirical and theoretical, of continuous random variables. This difference is compared with tabulated values of Dmax for a given significance level, and sample size. If the calculated value of Dmax is greater than the tabulated value, the null hypothesis is rejected (Naghettini and Pinto, 2007).

According to (Naghettini end Pinto, 2007) the method of moment (MoM) is the simplest method of estimation parameter and for small samples it may display the same efficacy as other methods. When compared to the method of maximum likelihood, their it its less efficient. The Maximum Likelihood Method (MLM) produces parameters with less variance; however, this property is only asymptotical, which makes its application to small samples, similar with other methods. The L-moment method (ML-M) produce parameters comparable in quality to those produced by the MLM method and sometimes even better for small samples (Hosking, Wallis and Wood, 1985).

2. MATERIALS AND METHODS

2.1 Zambeze Watershed

The Zambeze River is the main river of Mozambique, 4th largest river in Africa and that has the greatest flow among African rivers that flow into the Indic Ocean. Its watershed covers an area of 1.390.000 km2, little less than half of the surface of the Nile watershed. It starts in Kalene Highland at the border between Congo and Zambia at about 1,500 m altitude, heads to Angola, and then south and east along the borders of Namibia and Botswana, Zambia and Zimbabwe, and finally enters Mozambique, emptying into the Indic Ocean, making a journey of approximately 2700 km (Jessen and Silva, 2008).

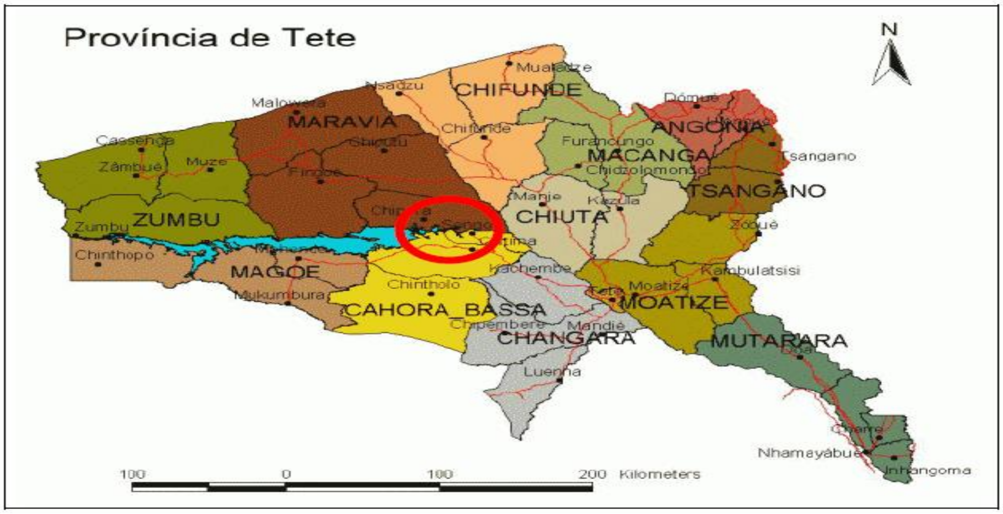

Figure 1 – Zambezi Watershed – (Jessen and Silva, 2008).

2.2 Data Collection and Failures

To determine the local maximum flow of the Zambezi River watershed natural series of daily flows were used at the gauging station E-320, belonging to Direcção Nacional de Águas (DNA), which is located in the province of Tete at coordinates 16º09' South, 33º35' East and the altitude of 118 m. The observation period covers 58 years (January of 1955 to December of 2012). At observation failures the correction of these was based on hydrological regionalization (Maidment, 1992). This transfer may include directly a series of flows, a precipitation, or even certain relevant statistical parameters such as mean, variance, maximum and minimum (Kaviski,1992).

According to (Tucci, 2002) hydrological regionalization is defined as the process of transferring information from one location to another within an area with similar hydrological behavior. This transfer can occur in the form of a variable or a parameter. Some methods of regionalization flows have been proposed, such as the traditional method described by (Electrobràs, 1985), using regional regression equations, applied to hydrologically homogeneous regions and linear interpolation which has been used in this study.

2.3 Probability Distribution Functions

A theoretical probability function is a function that relates values of a random variable to its probability of occurrence (Tucci, 2011). Usually a theoretical distribution (population) is adjusted to the observed values (sample) so that the extrapolation to values greater than the observed ones will be possible.

The theoretical distribution functions most used for hydrologic models in Flood Frequency Analysis are: Gumbel, Log-Pearson type III, Log-Normal, Exponential and Generalized of Extreme value (Maidment,1992). Which was used in this study.

The probability density function of the Gumbel distribution is:

$$f_x(x) = \frac{1}{\alpha}\exp\left[-\frac{x-\xi}{\alpha} - \exp\left(-\frac{x-\xi}{\alpha}\right)\right] \qquad \infty < x < \infty \qquad [1]$$

Where ξ e α are parameters.

The Log Normal distribution has the following probability density function:

$$f_x(x) = \frac{1}{(x) * \sigma_y * \sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{\ln(x) - \mu_y}{\sigma_y}\right)^2\right] \qquad 0 \leq x < \infty \qquad [2]$$

Where σ_y e μ_y are parameters.

The probability density function of the Exponential distribution is:

$$f_x(X) = \beta \exp[-\beta(x - \xi)] \qquad x \geq \xi \qquad [3]$$

Where ξ e β are parameters.

The Log Person III has the following function of probability density distribution:

$$f_x(X) = |\beta|\{\beta[\ln(x) - \xi]\}^{\alpha - 1\frac{\exp\{-\beta[\ln(x)-\xi]\}}{x\Gamma(\alpha)}} \qquad x \geq \xi \qquad [4]$$

Where α, ξ, β are parameters and Γ is the Gama function, given as:

$$\Gamma(\propto) = \int_0^{\infty} x^{\propto - 1} e^x \, dx \qquad [5]$$

The probability density function of the generalized extreme value distribution is:

$$f_x(X) = \frac{1}{\alpha}\left[1 - k\left(\frac{x-\xi}{\alpha}\right)\right]\frac{1}{k-1}\exp\left\{-\left[1 - k\left(\frac{x-\xi}{\alpha}\right)\right]\right\}$$ [6]

Where α e ξ are the distribution parameters.

2.4 Goodness of Fit Test

In order to check whether a theoretical probability distribution fits reasonable the sample data, the sample frequencies are compared to theoretical frequencies expected by the probabilistic model that was adapted to describe the observed data (Naghettini e Pinto, 2007).

In general to accept a distribution, a statistic based on the difference of the theorical frequency and the empirical frequency must be less than a critical value, according to the significance level. In this case, the model represents more accurately the sample data. Thus, all goodness of fit tests are testing the hypothesis of the sample to belong to the given theoretical distribution.

2.5 Kolmogorov-Smirnov

The goodness of fit test of Kolmogorov-Smirnov (KS) is a nonparametric test based on the maximum difference between the cumulative, empirical and theoretical distribution funtions of continuous random variables. This test is not applicable to discrete random variables (Martins et al., 2011).

2.6 Anderson-Darling

The Anderson-Darling test in principle follows the same procedure as the KS test, but its advantage is to be more sensitive for extreme values, as it gives more weight to the tails of the distribution points (Espinosa et al., 2004).

2.7 Chi-Square Test

The Chi-square test is the classical goodness of fit test most referred in the literature. This test groups the data within frequency classes and accumulates the relative squares of differences betweenthe observed and theoretical frequencies. The sum of the ratios of the squared differences of observed and expected frequencies with the observed frequency for each class generates the test statistic which is Chi-Square distributed (Elsebaie, 2011). One of its disadvantages is that this test is not feasible for small samples (Back, 2001). The chi-square test requires that the frequency of a class can not be less than five (Reis et al, 2011).

2.8 Return Period

According (Naghettini and Pinto, 2007) return period is the average period (in years) between successive occurrence of an event. For independent events it is equal to the inverse of the probability that such e event occurs at any year.

$$T_r = \frac{1}{1 - F_y(y)}$$ [7]

To achieve the objective of this work a distribution that offered the best fit was selected. To verify the fit the Kolmogorov-Smirnov test was used with at 5% of type I error probability. Parameters were estimated using the method of L moments or the Method of Moments. The excedence probability of the maximum flow was varied and thus, estimated maximum annual flows associated with different return periods were computed.

3. RESULTS AND DISCUSSION

3.1 Goodness fit

Using the statistical parameters of the maximum annual flow series extending from 1955 to 2012 and using the two estimation methods (L-moments and moments) gave the results shown in Table 1.

The best distributions are displayed in this table (1), and a high value of the test statistic, represented by Dmax, reveals large differences between observed and expected frequencies, being an indicator of poor fit of the distribution. Thus, the closer to zero the value of the test statistic, more representative will be the theoretical distribution. The critical values for Dmax (Kolmogorov-Smirnov) to a confidence level of 95%, and sample size 58 is equal to 0.178.

Table1. Distributions and its Value of the Kolmogorov Smirnov test.

location	Distributions	Method of parameters estimation	Value of Dmax in KS Test
Zambeze	Gumbel	ML-M	0.076
	Log Person III	MoM	0.08
	GEV	ML-M	0.077
	Log Normal	MoM	0.11
	Exponencial	ML-M	0.12

Comparing de computed Dmax values with this critical value no one of the analyzed distributions can be rejected at a 95% level. However table1 shows that the Gumbel distribution athained the best fit, followed closely by the GEV distribution. The parameters were estimated using the L-moment or classical moment methods. According to (Silva, 2011) the Gumbel, Log-Pearson III e GEV probability distributions are all suitable to estimate the maximum local flow or the maximum rainfall for different return periods.

3.2 Estimation of Flow

To estimate the flood flows the following probability distributions were used: Log -Normal, Log - Pearson type III, generalized extreme value (GEV), Gumbel and Exponential. In Table 2 are presented results of maximum flow for different return periods.

Table2. Results of the maximum flow ($m^3\ s^{-1}$) at Cahora Bassa estimated for different return periods and different models.

Distribuitions	Parameter Estimation	Return period							
		2	5	10	20	50	100	1000	10000
Gumbel	ML-M	5800	9000	11000	13000	15600	17500	23900	30200
Exponencial	ML-M	5200	9000	11400	14100	17600	20260	29100	38000
GEV	ML-M	7600	11000	12800	14600	17000	18780	24400	30000
Log -Person III	MoM	5400	10300	11100	12400	15100	17900	26400	34300
Log-Normal	MoM	5500	8900	11400	14000	17700	20690	31400	38000

The maximum flow ($m^3\ s^{-1}$) for the return periods studied varies from 5800 to 30200 for Gumbel distribution from 5400 to 34300 for Log Person III, from 7600 to 30000 for GEV, from 5500 to 38000 , for Log Normal and finally to the exponential distribution varies from 5210 to 38000$m^3\ s^{-1}$.

According to (Jessen and Silva, 2008) Q10.000 estimated at Cahora Bassa, is approximately 30.000 $m^3\ s^{-1}$ which agrees with the value obtained in this study for the same return period using Gumbel or GEV distribution which were those with the best fit.

(Silva,Silva and Guale, 2005) using statistical treatment of the flow at Cahora Bassa since 1961 reached the following results: maximum flow equal to 4500$m^3\ s^{-1}$ in February, with estimated return period of 2 years, 7000$m^3\ s^{-1}$ in March, with a return period estimated 4 years, 14000 and 17000$m^3\ s^{-1}$ in March, with a return period estimated of 50 and 100 years, including flood flows of 10000 years which may reach values on the order of 28000$m^3\ s^{-1}$. These results are similar to those obtained in this research for the Gumbel distribution.

4. CONCLUSIONS

According to the results obtained in this study, it is concluded that:

1) The maximum natural flow in the Zambezi watershed at Cahorra Bassa for different return periods (2 to 10000 years) for analysis in this research ranges from 5800m^3s^{-1} to 30200 m^3s^{-1}. This result is for the best fitting distributions.

2) The Gumbel distribution followed closely by the GEV distribution presented the best fit using the Kolmogorov Smirnov test.

3) Neither of the Gumbel, Log-Normal III, Log-Pearson III, Exponential and GEV distribution may be rejected at 95% level for estimation of maximum local flow at Cahora Bassa within the Zambeze watershed.

4) Considering all of the five distributions analyzed flows for de T_r =10000 may reach 38000 m^3s^{-1}. This may be considered a very conservative estimate.

5. **ACKNOLEDGEMENTS**:

The authors thank CNP/MCT for granting Scholarships for Master degree to the first author and the post-Graduate Program in environmental and Water Resources Engineering at Federal University of Parana (PPGEAHR / UFPR) for technical and scientific support.

6. REFERENCES

Back, 2001: Seleção de distribuição de probabilidade para chuvas diárias extremas do Estado de Santa Catarina. *Revista Brasileira de Meteorologia.*

ELETROBRÁS,1985: Metodologia para Regionalização de Vazões. Rio de Janeiro.

Elsebaie, 2001: Developing rainfall intensity–duration– frequency relationship for two regions in Saudi Arabia. *Journal of King Saud University – Engineering Sciences*.

Espinosa, Calil and Lahr,2004: Métodos paramétricos e não - paramétricos para determinar o valor característico em resultados de ensaio de madeira. *Scientia Forestalis*, Piracicaba.

Fill, 2000: Notas de Aula, Curitiba-Brasil.

Hosking, Wallis and Wood, 1997: Estimating on of the Generalized Extreme Value distribution by the method of probability – weithted moments.*Tech-home trics*

Jessen and Silva, 2008: “Gestão Hidrológica da Albufeira de Cahora Bassa em Periodos críticos (Cheias e Secas) *5º congresso luso-Moçambique de Henginharia 2º Congresso de Engenharia de Moçambique*, songo-Mocambique,setembro 2008.

KAVISKI,1992: Métodos de Regionalização de Eventos e Parâmetros Hidrológicos. Dissertação de Mestrado, UFPR.

Kite,1978: Frequency and risk analyses in hydrology. Fort Collins, *Water Resources Publications*.

Maidment, 1992: Handbook of Hydrology. *McGraw-Hill*.New York, cap.18

Martins and Reis,2011: estimativa da vazão e da precipitação máxima utilizando modelos probabilísticos na bacia hidrográfica do rio benevente. *Enciclopédia Biosfera, Centro Científico Conhecer* – Goiânia.

Naghettini and Pinto, 2007: Hidrologia estatística. Belo Horizonte: CPRM,

Reis, Beijo, and Liska,2010: “Comparação dos métodos de verificação do ajuste da distribuição Gumbel a dados extremos” In*: XIX Congresso de pós-graduação da UFLA*, 2010.

Silva,2003: “O Impacto da Gestão da Albufeira de Cahora Bassa nos Domínios Social, Económico e Ambiental” *3rd Congresso Luso-Moçambicano de Engenharia*, Maputo 19-21 Agosto 2003.

Silva, Silva and Guale, 2005: “Monitoramento Ambiental da Albufeira de Cahora Bassa”. *7º SILUSBA - Simpósio de Hidráulica e Recursos Hídricos dos Países de Língua Oficial Portuguesa* - Évora, Portugal – 30 de Maio a 2 de Junho de 2005

Tucci,2002: Regionalização de vazões. *Editora da Universidade*. UFRGS.

TUCCI,2011: Hidrologia: ciência e aplicação. 2. *ed. Porto Alegre*, RG: UFRGS.