HYDROLOGICAL RISK ANALYSIS OF SOLIMÕES RIVER USING EXTREME VALUE THEORY

Alexandra R. Mendes de Almeida[(a)], Beatriz V. de Melo Mendes[(b)], Maria Célia S. Lopes[(c)], Gerson Gomes Cunha[(d)]

[(a), (c), (d)] LAMCE/COPPE-UFRJ
[(b)] IM/COPPEAD-UFRJ

[(a)] alexrma@lamce.coppe.ufrj.br, [(b)] bmendes@visualnet.com.br, [(c)] celia@lamce.coppe.ufrj.br, [(d)] gerson@lamce.coppe.ufrj.br

ABSTRACT
The Brazilian hydrography is made by naturally navigable rivers that, in most cases, are the only transport means between many villages located by their margins. It occurs in particular in Amazonas (Brazil), where large distances and adverse natural conditions, make difficult the use of other transportation types, whose implantation and use are too onerous.

Rivers are crucial not only for transporting large cargoes through large distances, they are also important to the local and international commerce, making viable the offer of products with competitive prices.

This work presents the study with extreme low levels of rivers at Amazon region. In this study, we consider specially Solimões River levels, because Solimões is very important in communicating regular lines of cargo and people fluvial transport between local villages.

The use of the Extreme Value Theory to model these extreme low levels of the Solimões river is due to its adequacy and accuracy in estimating probabilities of occurring rare events.

Keywords: fluvial transportation, extreme value theory, hydrology.

1. INTRODUCTION

The importance of Solimões river in Amazonas is because of its strategic geographic position, that makes it an attractive option to the products leakage, besides constitute a principal transportation way to the local people who lives at the margin of supplementary rivers.

The objective of this paper is to specify the occurrence of rare events that, in this context, are the low levels of the Solimões river.

In EVT we study the behavior of rare events, the less frequent events that in major cases have an unarguable impact, and on emphasizing events of less probability of occurrence we achieve a better accuracy in modelling the tail distribution of river levels.

We use Extreme Value Theory in modeling low levels of Solimões River, analyzing particularly six monitoring stations that are placed since the border of Brazilian territory until near its union with Negro River. They are: Tabatinga, São Paulo e Olivença, Santo Antônio do Içá, Fonte Boa, Itapeuá e Manacapuru.

First applications using EVT were in meteorological phenomena modeling area and involved maximum precipitation and annual level of inundation in United States. However, EVT approach is quite inclusive and can be applied in a variety of natural phenomena like: inundations, atmospherical pollution, engineering, actuary and financials.

2. EXTREME VALUE THEORY

Extreme Value Theory has emerged as one of the most important statistical disciplines for the applied sciences over the last 50 years, also becoming widely used in natural phenomenon like inundation, atmospheric pollution, engineering, level of hydroelectric pond (Almeida and Mendes 2005), actuary and finance.

The most important results was obtained in Fisher and Tippett (1928), Gnedenko (1943) and Galambos *et al* (1994).

The distinguishing feature of an extreme value analysis is the objective to quantify the stochastic behaviour of a process at unusually small levels. In particular, extreme value analyses usually require estimation of the probability of events that are more extreme than any that have already been observed.

The discussion about the better method in EVT to model the extremes comes from the quantity of data, considering that by definition, extreme observations are less frequency.

The EVT model that treat events that don't exceed a small threshold u is denoted by Generalized Pareto Distribution.

2.1. The Generalized Pareto Distribution

The Generalized Pareto Distribution (GPD) is an approximation to the limit distribution of excess smaller than a small threshold. We call *excesses* the values smaller than the threshold, and *excess* the difference between the excesses and the threshold.

Consider $X_1, \ldots, X_n$ a sequence of independent and identically distributed random variables, having marginal distribution function F. The extremes events are the X_i that exceed some high threshold u.

Denoting an arbitrary term in the X_i by X, the behavior of extremes events is given by the conditional probability:

(1)

If the parent distribution F is known, the distribution of threshold exceedances given above would also be known. Since, in practical applications, this is not the case, approximations that are broadly applicable for high values of the threshold are sought, Coles (2004).

The distribution function of the excess y is given by the GPD:

(2)

Where $\sigma>0$, $y\geq 0$ when $\xi\geq 0$, and $0\leq y\leq -\sigma/\xi$, when $\xi<0$.

The parameters ξ and σ is related to the shape and scale of the model.

This is a generalized distribution, including other distributions above the same parametric function. If $\xi>0$ then H(y) is a re-parametrized version of the Standard Pareto Distribution, widely used in actuarial modelling of big losses. If $\xi=0$ we have the exponential distribution and, finally if $\xi<0$ we have the Type II Pareto distribution.

3. THE CASE STUDY

3.1. The Data

The data analyzed was made available by the Agência Nacional de Águas (ANA-Brazil), and consists in daily level (in meters) of six monitor stations distributed across the Solimões river, and is considered by ANA as of most strategic interest: Tabatinga, São Paulo de Olivença, Santo Antônio do Içá, Fonte Boa, Itapeuá and Manacapuru.

These stations were chosen in order to cover all the extension of Solimões River since the Brazilian boundary (Tabatinga e São Paulo de Olivença) passing through points at medium riverbed (Santo Antônio do Içá, Fonte Boa e Itapeuá), until the proximities of its union with Negro River. Figure 1 shows all the extension of Solimões River as well as the location of all stations used.

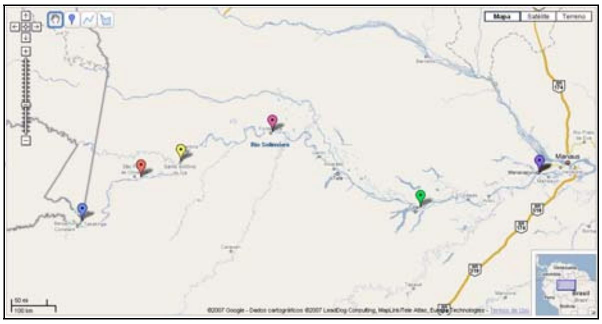

Figure 1: Monitor station's geographic localization in Solimões river.

The levels data series periods vary for each station. Tabatinga data series starts at july of 1982 and ends at November 2007, São Paulo de Olivença and Santo Antônio do Içá collected data from july 1973 until june 2007, Fonte Boa start at november of 1977 and ends at november of 2007, has data from april of 1971 until july of 2007, finally Manacapuru starts at june of 1972 and ends at January of 2007.

Environmental data are typically high correlated, and the daily levels presents short and long memory. The Ljung-Box test for serial correlation (Box and Pierce, 1970) and the R/S for long memory (Lo, 1991), reject their null hypothesis at 1% significance level. Figure 2 shows how strong is the daily levels autocorrelation function of Fonte Boa and Itapeuá monitor stations.

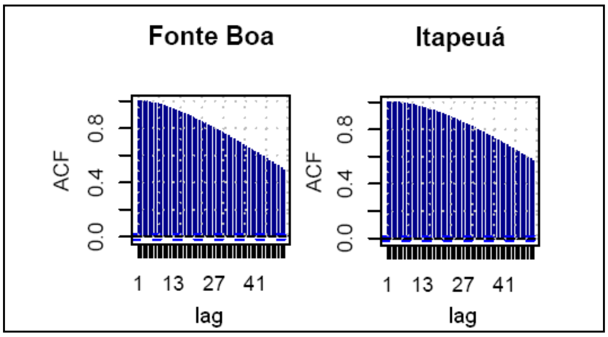

Figure 2: Autocorrelation function plot of Fonte Boa and Itapeuá monitor stations.

3.2. Monthly Minima

To remove the temporal dependency of the time series, we selected the monthly minimum of daily levels.

Another concern is about the trend and seasonality of the data. According to Figueroa and Nobre (1990) the spatial-temporal distribution of the precipitation, an important factor to determine the river level, is characteristic of each location. The Figure 3 shows the spatial variation of Amazonian precipitation:

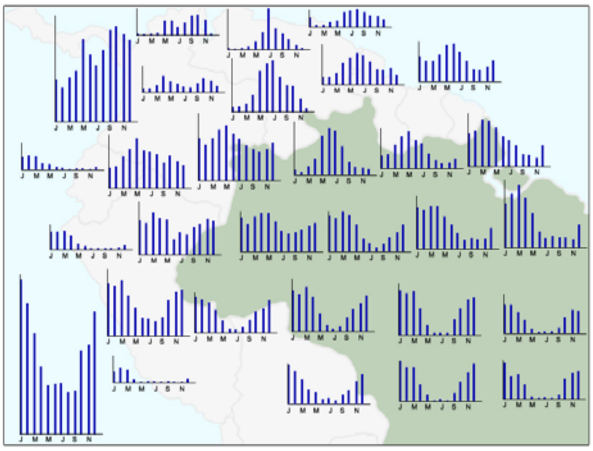

Figure 3: Amazonas spatial-temporal precipitation distribution.

We analyzed the levels profile of each monitoring station and considered spatial characteristics of Amazon region as shown in Figure 3. We verified a delay in water dam and decided to segment the data series in three distinct profiles: periods of dry, high and transition, enclosing different months for each station, once the Amazon region presents distinct precipitation indexes for locations not so far from each other.

Figure 4 shows box plot of levels separated by months of Tabatinga station, where blue color indicates the period considered as high, green color indicates months classified as transition period and red color indicates months of river dry.

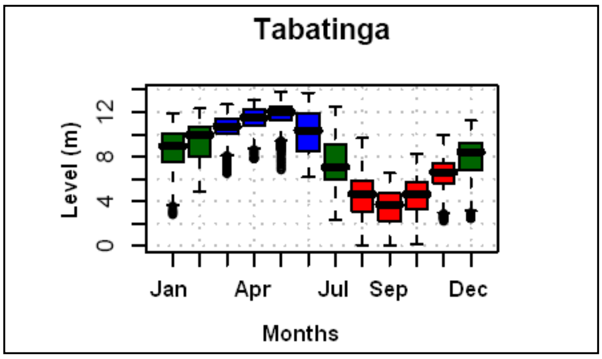

Figure 4: Box plot of monthly minima levels.

Although the monthly minima still have a weak dependency in short lags, like shown in figure 5, that can be explained by the kind of data; but we do not see long memory evidence anymore.

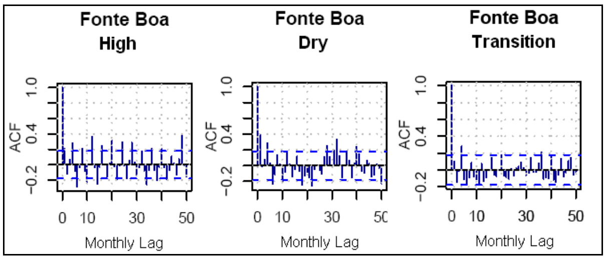

Figure 5: ACF for monthly minima levels by periods.

We conclude the data study verifying the series stationarity. The use the KPSS test, which the null hypothesis is stationariry (level and in tendency), and verify that for all monitor station and season, we could not reject the null hypothesis at in average, 5% significance level.

3.3. Modeling and Inferences

The choice of the threshold u is a critical step of the modelling. A too high threshold is likely to violate the asymptotic basis of the model, leading to bias; and a too low threshold will generate few excesses with which the model can be estimated, leading to high variance (Coles 2004).

We estimated the threshold u by a sensibility analysis, based on the fitting of models across a range of different threshold, resulting on the choice of the 21.5% quantile of each series/period.

The parameters of the GPD was estimated by a numerical method called *Probability Weighted Moments (Hosking 1985).*

Table 1 shows the threshold (u), the number of excesses N_u, and the estimated shape and scale, by period and monitor station.

Table 1: Estimates by period and monitor station.

Monitor Station	Period	u (cm)	N_u	ξ estimated	σ estimated
Tabatinga	High	841,84	22	-1,083	2,486
	Dry	216,01	23	-0,644	1,655
	Transition	492,00	21	-0,013	0,984
São Paulo de Olivença	High	1033,98	30	-0,374	1,849
	Dry	387,07	30	-0,119	0,944
	Transition	681,48	29	-0,040	0,993
Santo Antônio do Içá	High	1014,17	30	-0,041	1,520
	Dry	326,02	30	-0,226	1,294
	Transition	620,34	29	-0,026	1,052
Fonte Boa	High	1851,58	26	0,150	1,154
	Dry	1142,92	26	-0,197	0,946
	Transition	1462,58	26	-0,087	1,150
Itapeuá	High	1338,07	32	-0,079	1,199
	Dry	626,49	31	-0,278	1,856
	Transition	866,74	31	-0,491	3,179
Manacapuru	High	1594,92	30	-0,243	1,538
	Dry	833,96	30	-0,230	1,773
	Transition	1174,45	30	-0,320	2,524

The shape (ξ) parameter estimated of almost all station/period was negative, indicating the limitation of the series, as expected, and the scale (σ) parameter estimated was similar for the same periods of the stations.

To verify the goodness-of-fit of each model we used the Kolmogorov-Smirnov test, that basically compares the empirical distribution function with the cumulative distribution function specified by the null hypothesis. The results can be seen in table 2.

Table 2: Estimates by period and monitor station.

Monitor Station	Period	N_u	D_n	P-value
Tabatinga	High	22	0,13	87,38%
	Dry	23	0,11	92,18%
	Transition	21	0,12	86,17%
São Paulo de Olivença	High	30	0,07	99,69%
	Dry	30	0,13	65,77%
	Transition	29	0,12	77,53%
Santo Antônio do Içá	High	30	0,10	92,08%
	Dry	30	0,09	96,86%
	Transition	29	0,13	74,85%
Fonte Boa	High	26	0,14	71,61%
	Dry	26	0,09	98,11%
	Transition	26	0,11	93,54%
Itapeuá	High	32	0,07	99,55%
	Dry	31	0,09	96,04%
	Transition	31	0,15	44,70%
Manacapuru	High	30	0,08	99,38%
	Dry	30	0,16	40,12%
	Transition	30	0,11	85,31%

At the significance level of 5% we accepted all null hypothesis of goodness-of-fit.

4. RESULTS

Once the estimation are obtained and model adequacy is verified, we could use the model to obtain results that reflects estimation impact, which characterize the behavior of the model in a future scenery. An interesting analysis in this context of minimal levels of monitoring solutions of Solimões River consists in obtaining the events in adverse sceneries.

Frequently it is necessary to construct adverse scenarios to warn us of some extreme event. The best way to take care of this is to assign to this event an average time to occur.

In general, we impose a large period to study the value of the probability of its excess.

The return level of *t*-periods, as in Mendes (2004), is a concept that can be summarized basically by the association of a quantile of the excesses distribution and a expected period *t* between their occurrence.

The equation of return level of t-periods is given by:

(3)

Where x_t is the return level of *t*-periods, u is the threshold, ξ and σ are the shape and scale parameters, t is the number of periods, and ζ_u is the $P\{X<u\}$.

Table 3 shows the estimates of *t*-period return, for t=5, 10, 15 e 20.

Table 3: T-period return estimates.

Monitor Station	Period	*t*=5	*t*=10	*t*=15	*t*=20
Tabatinga	High	8,245	7,124	6,768	6,596
	Dry	2,043	1,159	0,798	0,594
	Transition	4,848	4,170	3,776	3,498
São Paulo de Olivença	High	10,207	9,108	9,585	8,260
	Dry	3,802	3,179	2,838	2,605
	Transition	6,743	6,066	5,678	5,407
Santo Antônio do Içá	High	10,031	8,995	8,403	7,988
	Dry	3,167	2,350	1,898	1,651
	Transition	6,127	5,405	4,989	4,697
Fonte Boa	High	18,431	17,579	17,038	16,634
	Dry	11,361	10,756	10,439	10,229
	Transition	14,542	13,773	13,344	13,049
Itapeuá	High	13,294	12,490	12,039	11,729
	Dry	6,131	4,984	4,409	4,039
	Transition	8,441	6,639	5,836	5,357
Manacapuru	High	15,838	14,874	14,381	14,060
	Dry	8,212	7,082	6,491	6,101
	Transition	11,564	10,030	9,279	8,802

The estimated return level of *t*-periods of each station shows that for 10 years period the high and transitions periods has a low estimates, enough to makes hydrological risk managers includes this statistics in their future decisions.

5. CONCLUSIONS

The Extreme Value Theory is a powerful methodology to model all nature of risks. In this paper we use this approach to estimate extreme low levels of the Solimões (AM-Brazil) river.

The calculus of more accurate estimates for the low levels of the river is important to choose the type (and size) of the ship for use in cargo transportation.

In fact the data segmentation in periods of high, dry and transition water results in robust statistics of return levels, that encourage us to study the relationship between the seasons of each monitor station, and mainly the lag of each extreme low level.

REFERENCES

Almeida, A. R. M. de, Mendes B.V.M., 2005. Modelagem de Eventos Extremos no Setor Hidrelétrico. *Iniciação Científica (UFRJ - CNPq).*

Coles, S. G., 2004. An Introduction to Statistical Modeling of Extreme Value. *Springer-Verlag.*

Embrechts, P., Klüppelberg, C. and Mikosch, T., 1997. Modelling Extremal Events. *Springer-Verlag.*

Figueroa, S. N., Nobre, C. A., 1990. Precipitation Distribution over Central and Western Tropical South America. *Climanálise, V. 5, N. 6.*

Fisher, R. A., Tippett, L. H. C., 1928. Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample. *Proceedings of the Cambridge Philosophical Society, V. 24, p. 180-190.*

Galambos, J., Leigh , S. , & Simiu, E., 1994. Extreme Value Theory and Applications. Kluwer, Amsterdam.

Gnedenko, B. V., 1943. Sur la Distribution Limite du Terme Maximum d´une Série Aléatoire. Annales des Mathématiques, v. 44, p. 423-453.

Hosking, J. R. M., Wallis, J. R. e Wood, E. F., 1985. Estimation of the Generalized Extreme Value Distribution by the Method of Probability-Weighted Moments. *Technometrics, 27, 251-161.*

Mendes, B.V.M., 2004. Introdução à Análise de Eventos Extremos. *Editora E-Papers.*