

PROBABILISTIC MODELS FOR ESTIMATION EXTREME VALUES OF WATERSHED FLOW RATES

Leandro Campos Pinto¹
 Pedro Terra Lima²
 João José da S. Junior³

Abstract

The applicability of probability models to time series representing minimum and maximum daily flow rates were tested for Itabapoana River in Brazil. The series represents data from 1968 to 2007, 40 year period. The distribution adequacy was verified using Kolmogorov-Smirnov and chi-square tests. For maximum flow rates, the statistical distribution best fitted were 2 and 3 parameters log-normal. For minimum flow rates, the best fitting were obtained through Weibull and Gama functions. Using these fittings, the maximum and minimum flow rates could be estimated with high confidence for Itabapoana river.

Key-words: Probability models. Flow. Specific flow rate indexes.

Introduction

High water needs by world population has required larger exploitation of rivers causing demand conflicts and environmental impacts due to water availability shortage. Such pressure on water resources imposes higher needs for accuracy estimating flow rates. On the other hand, flooding has caused economical damages due to irregular occupancy of river beds. In this case, correct flood level estimation is necessary.

Besides that the maximum flow of a river which is understood as the value associated with a risk of being equaled or exceeded in use in flood forecasting and design hydraulic structures such as culverts, channels, conduits, among others. Thus, knowledge of the maximum flow is needed to control and mitigate flooding in a given area or basin in hydraulic works and rural urban drainage in irrigation schemes, dams and spillways of dams, among others.

The estimation of this value is of decisive importance in the cost and safety of engineering projects, to be the result of precipitation occurring or likely to occur (TUCCI, 2001).

According Silvino et al. (2007), the water, due to its scarcity in quantity and quality, is no longer a free resource and economic value gained from the law of the National Water Resources Policy (Law No. 9433) of 8 January 1997. Due to this reason, design of water structures to supply water, irrigate, generate electricity and others, has demanded very precise estimates of maximum and minimum flow rates in order to reduce costs of constructions.

Tucci (1997) proposes that high flow rates, at short or long terms, can be estimated through probability models, based on risk analysis. The short term flow rate estimate is based on present precipitation values. According to Euclides (2002), maximum flow rate estimates can be associated to recurrence and it is very important to predict flooding levels, being very important for adoption of structural constructions to control or attenuate flooding effects.

The minimum flow rate is characterized by two factors: its magnitude and duration. At practice, the instantaneous minimum values are not that important since it approaches the minimum daily value. Minimum values of flow rate with larger duration are, however, more important to be determined, since it can represent an unfavorable situation for water use or environmental preservation (TUCCI, 2002).

When flow rate time series are available, it is necessary to work them out adequately through probabilistic analysis and stochastic studies (HAAN, 2002; DOUGLAS; VOGEL, 2006).

¹Universidade Federal de Lavras - Engenheiro Agrícola, Doutorando em Ciência do Solo, DCS - UFLA/Lavras-MG. leandcampos@yahoo.com.br

²Universidade Federal de Lavras - Engenheiro Agrícola, Doutorando em Ciência do Solo, DCS - UFLA/Lavras-MG. pedroterralima@yahoo.com.br

³Universidade Federal de Lavras - Engenheiro Agrônomo, Professor substituto departamento de Engenharia Agrícola, Doutorando em Recursos Hídricos em Sistemas Agrícolas - UFLA/Lavras-MG. Bolsista da Fapemig. jjjunior@deg.ufla.br

Maione (1977) applied Gumbel probabilistic models, as well as Log-normal (two and three parameters), Fuller-Coutagne and Pearson with two and three parameters, at high flow rate time series at Dora Baltea river in Italy and demonstrate that the Log-Normal with two parameters provided the best fitting. (EUCLYDES: PICOLLO, 1985) applied these same models at a watershed at Water Circuitry Region of Minas Gerais state in Brazil and observed that values predicted by Gumbel and Log-Normal two parameters were very close.

Kaviski (1983) applied probabilistic models at minimum flow rates at small watersheds at Santa Catarina state in Brazil and concluded that the Weibull distribution was the most precise for estimating values.

Based on the diversity of probabilistic models for flow rate studies (maximum and minimum values), as well as the need for accuracy on water structures design, this research decided to investigate the time series analysis for Itabapoana river in Brazil.

Methodology

Watershed

Itabapoana watershed (Figure 1) is 4875 km² large. The river is 264 km long, starting at Caparaó ranges between Espírito Santo and Rio de Janeiro state, discharging its water at Atlantic Ocean. This river is important due to its hydro-eletric potential with several waterfalls along it. Its tributaries are Preto and São João rivers. Then it collects water from Muqui do Sul, Guaçui, Calçado and other small rivers. According to Reis et al. (2008), the main activities at Itabapoana watershed are services at primary sector, mainly those linked to pasture, coffee, sugarcane and tropical fruit growth.

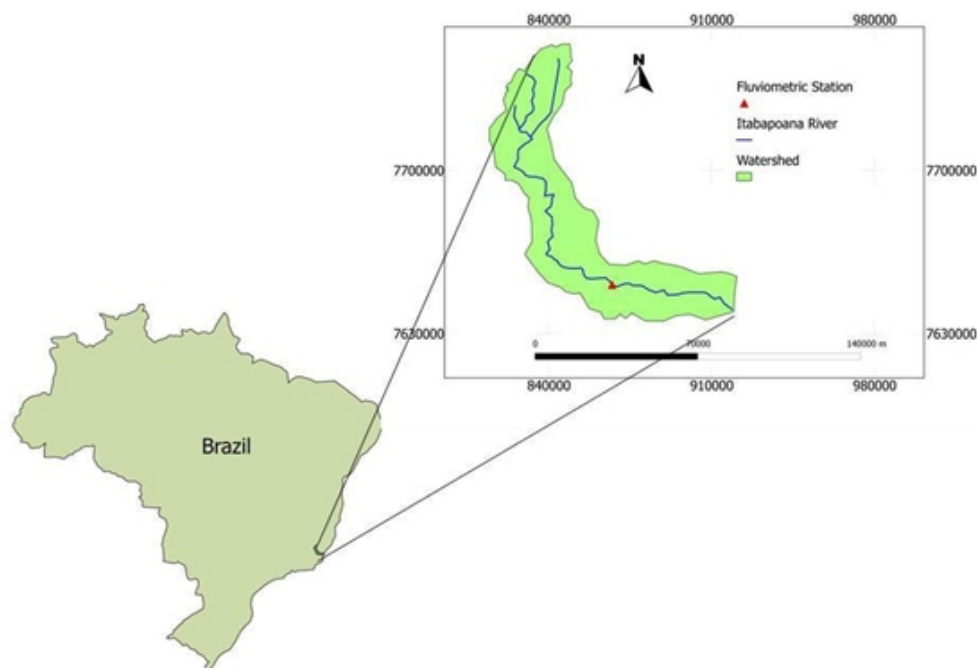


Figura 1: Itabapoana Watershed location at Brazilian territory.

Daily flow rates were obtained at Hidroweb, a Brazilian water information database, managed by ANA – National Water Agency. The flow rate station is based at Itabapoana bridge, at coordinates 21° 12' 22" S e 41° 27' 46" W, 40 m altitude and draining area of 2854 km². The time series investigated has maximum and minimum flow rates from 1968 to 2007, for 40 years long.

Probability distribution functions and adherence tests

We used the Gumbel probability distributions, which is applied to the time series of extreme values. According to Tucci (2002), its Probability Density Function can be obtained from equation (1), Log-normal on 2 and 3 parameters given by equations (4) and (8) and the range equation (16) to estimate the maximum flows and the same distributions plus the Weibull distribution equation (13) to estimate the minimum flows for various return periods. The distributions were adjusted based on the method of moments (NAGHETTINI: PINTO, 2007).

To assess the adequacy statistical distributions for all periods studied, we used the Kolmogorov-Smirnov (KS) equation (21) and Chi-square (λ^2) equation (22) at 5% significance level, following guidelines Ferreira (2005) and Naghettini e Pinto (2007). These tests were chosen because they are the most commonly used in the literature and considered as the most appropriate for continuous random variables such as flow rates (SILVINO et al., 2007).

In the test of KS comparison is made between the maximum deviation in magnitude due to the difference between the values observed and theoretical frequency, with the tabulated value based on the sample size and level of significance. In the test of λ^2 comparison is made between the sum of squared deviations between the observed frequencies and theoretical (λ^2 calculated) and the value in the table (λ^2 tabulated), depending on the number of degrees of freedom (number of classes - number of parameters - 1) and level of significance. Obtained the number of classes grouping the square root of the number of data. For the model to probabilities is deemed appropriate, the calculated values should be equal to or less than tabulated for each test.

Gumbel probability distribution

Gumbel, in 1945, suggested that this distribution of extreme values would be appropriate for frequency analysis of flooding, since the series was annual, that is, each flow rate range of the extreme values were the highest throughput of a sample of 365 possibilities (higher flow year).

The distribution also known as Gumbel extreme value distribution of type I or type II distribution Fisher-Tippet, according to Gumbel (1958), has the function of cumulative probabilities given by equation 1.

$$FDP: f(x) = \alpha \cdot e^{-\alpha \cdot (x-\mu) - e^{-\alpha \cdot (x-\mu)}} \quad (1)$$

Wherein x is the variable (maximum flow or minimum annual daily in $m^3 s^{-1}$), $\hat{\alpha}$ and $\hat{\mu}$ are the parameters of the probability distribution which are given by equations (2) and (3).

$$\hat{\alpha} = \frac{1,2826}{s} \quad (2)$$

$$\hat{\mu} = \bar{x} - 0,45 \cdot s \quad (3)$$

Wherein \bar{x} and s will correspond to the average and standard deviation of the series.

Probability Distribution of Log-normal parameters 2.

The annual number of extreme events consisting of the maximum daily flow averages each year because they are not completely random flows do not follow a Gaussian distribution. However, if instead of flows are considered the logarithms of their values, the latter approximate relatively well the normal distribution. Thus, denoting by x to hydrological variable (in this case, x representing the flow), and making $y = \log x$ is the FDP given by equation 4.

$$FDP: f(x) = \frac{1}{x \cdot \sigma_n \cdot \sqrt{2\pi}} e^{-0,5 \left(\frac{\ln(x) - \mu_n}{\sigma_n} \right)^2} \quad (4)$$

Wherein μ_n and σ_n are the parameters of the distribution and the average and standard deviation of the logarithms of the information.

The values of the parameters of this distribution can be estimated based on the mean and standard deviation data without logarithmic transformation. And are given by equations 5 and 6.

$$\mu_n = \frac{1}{2} \ln \left(\frac{x}{\bar{x}^2 + s^2} \right) \quad (5)$$

$$\sigma_n = \sqrt{\ln \left(\frac{\bar{x}^2 + s^2}{\bar{x}^2} \right)} \quad (6)$$

The basic equation for hydrological studies, using a log-normal distribution second parameter is adapted from general equation Ven Te Chow equation (7) (HAAN, 2002):

$$X_{tr} = e^{\mu n + \sigma n \cdot K_{tr}} \quad (7)$$

Probability Distribution of Log-normal 3 parameters

Log-Normal distribution of parameters is similar to 3-Log Normal Tipo II distribution, except for the fact that the variable x deducts the amount n represents a lower limit. In this case, the variable $y = (x - n)$ is distributed according to a Normal with mean μn and standard deviation σn . The corresponding density function is given by equation 8.

$$FDP: f(x) = \frac{1}{(x-\beta) \cdot \sigma n \cdot \sqrt{2\pi}} e^{-0,5 \left(\frac{\ln(x-\beta) - \mu n}{\sigma n} \right)^2} \quad (8)$$

$$\beta = \bar{X} - \frac{s}{\eta y} \quad \eta y = \frac{(1-\phi^{2/3})}{\phi^{1/3}} \quad (9)$$

$$\phi = \frac{[-y + (y^2 + 4)]^{0,5}}{2} \quad (10)$$

$$y = \frac{n}{(n-1) \cdot (n-2)} \cdot \frac{\sum_i^n (x_i - \bar{x})^3}{s^3} \quad (11)$$

$$X_{TR} = e^{\mu n + \sigma n \cdot \sigma n} + \beta \quad (12)$$

Weibull Probability Distribution

This distribution has applications to historical series of minimum values, normally worked for minimum flows or similar series. The Weibull distribution is a derivation of asymptotic Extreme Value distribution. His FDP is given by equation 13.

$$FDP: f(x) = \lambda \cdot \beta \cdot x^{\beta-1} \cdot e^{-\lambda \cdot x^\beta} \quad (13)$$

Wherein λ and β are parameters of the distribution and are associated with average (μ) and variance (σ^2), respectively, by:

$$\mu = \left(\frac{1}{\lambda} \right)^{\frac{1}{\beta}} \cdot \Gamma \left(1 + \frac{1}{\beta} \right) \quad (14)$$

$$\sigma^2 = \left(\frac{1}{\lambda} \right)^{\frac{2}{\beta}} \cdot \left[\Gamma \left(1 + \frac{2}{\beta} \right) - \left(\Gamma \left(1 + \frac{1}{\beta} \right) \right)^2 \right] \quad (15)$$

The value of the variable x associated with the turnaround time (TR) is given by:

$$x = \left[\frac{\ln \left(1 - \frac{1}{TR} \right)}{-\lambda} \right]^{\frac{1}{\beta}} \quad (16)$$

Probability Distribution Gamma

According to Clarke (1993) that distribution has been widely used in hydrology in order to model full year. According to the author the Gamma distribution has a long upper tail long this feature that allows the distribution represent annual floods and other hydrological extremes wherein this asymmetry invariably present. And probability density function is given by equation (17).

$$FDP: f(x) = \frac{1}{\beta^v \times \Gamma(v)} \cdot X^{v-1} \cdot e^{-\frac{x}{\beta}} \quad (17)$$

The parameters of this distribution is β and v , which can be obtained by:

$$\beta = \frac{s^2}{\bar{x}} \quad (18)$$

$$v = \frac{\bar{x}^2}{s^2} \quad (19)$$

The function Gamma any number can be approximated by:

$$\Gamma(n) = \left[\frac{\sqrt{2\pi}}{n} \cdot \left(p_o + \sum_{i=1}^5 \frac{p_i}{n+i} \right) \right] \cdot (n + 5.5)^{n+0.5} \cdot e^{-(n+5.5)} \quad (20)$$

To consider this estimate:

$$p_o = 1,000000000190015; p_1 = 76,180091729471460; p_2 = -86,505320329416770; p_3 = 24,014098240830910; p_4 = -1,231739572450155; p_5 = 1,208650973866179 \times 10^{-3}$$

Adhesion Kolmogorov-Smirnov (KS) and Chi-square test (χ^2)

Kolmogorov-Smirnov (KS)

In this test, promotes the calculation of the difference between the observed frequencies (sample) and the frequencies expected based on the probability distribution, comparing the largest difference obtained a value corresponding to the test statistic.

This statistic is obtained as a function of sample size (n) and significance level (α) to be adopted (5% in most cases). The null hypothesis to be tested is the hypothesis H_o that the observed frequency can be estimated by pobability distribution, ie, as the tabulated value is statistically zero, it can be concluded that less than or equal to this value will also be statistically zero. This test is given by equation (21).

$$D_{obs} = \text{Max}|F(x) - S(x)| \quad (21)$$

Testing χ^2

This test is more stringent than the previous one by grouping the data series as classes frequently and accumulate errors between the observed and theoretical frequencies, reimbursement of all classes and not just the maximum difference. The sum of these errors (obtained by summing the errors of all classes) generates the second calculated value χ^2 .

The test statistic is obtained by means of χ^2 table, adopting the tabulated value based on degrees of freedom of the distribution and level of significance. For probability distribution has adherence to the data, the calculated value of χ^2 should be smaller than the table. This test is given by equation (22)

$$\chi^2 = \sum_{i=1}^k \left(\frac{(F_{0I} - Fe_i)^2}{Fe_i} \right) \quad (22)$$

Results and discussion

Values of maximum daily flow rates and maximum specific flow rate indexes for 2, 5, 10, 20, 50 and 100 years recurrence are presented at Table 1. It is possible to observe that Log-normal 2 parameters estimated the highest values (flow rate and specific flow rate indexes) for periods of return larger than 10

years. Use of this function might represent highest security on estimating flow values and, consequently, over design of water structures.

Adherence tests can be evaluated through table 2, wherein it is possible to observe data adherence for maximum flow rates for all probability functions, at both tests (KS and χ^2). Log-normal and 3 parameters models generated the smaller calculated values of χ^2 , being followed by Log-normal 2 parameters and Gama functions. It is possible that estimates using these models can be more precise. The probabilistic models that presented larger adherence for maximum flow rates were those that best fit flooding flow rates (log-normal 2 and 3 parameters and Gama) as mentioned by Clarke (1993). It is important, however, to point that all models can be considered adequate to estimate maximum flow rates for all return time values used.

Silvino et al. (2007), evaluating maximum flow rates at Paraguay River, also got good adherence for Log-normal probability function using KS tests, as well as Mello et al (2010) for the upper Rio Grande watershed.

Table 1 – Maximum flow rates (Q) and maximum specific flow rate indexes (RE) for different periods of return at Itabapoana flow rate station.

Model	Return period (years)	Q (m ³ /s)	RE (L/s km ²)
Gumbel	2	278,03	97,42
	5	398,11	139,49
	10	477,61	167,35
	20	553,87	194,07
	50	652,58	228,65
	100	726,55	254,57
Log 2P	2	272,82	95,59
	5	396,34	138,87
	10	481,98	168,88
	20	568,16	199,07
	50	678,74	237,82
	100	768,71	269,34
Log 3P	2	282,33	98,92
	5	403,59	141,41
	10	479,21	167,91
	20	550,27	192,81
	50	635,69	222,74
	100	701,33	245,74
Gama	2	270,4	94,74
	5	403,27	141,30
	10	485,88	170,25
	20	545,48	191,13
	50	635,79	222,77
	100	636,12	222,89

Table 2 – Adherence test results for probability functions used to estimate the maximum flow rates.

Distribution	KS		χ^2	
	$\Delta f_{calc.m\acute{a}x}$	$\Delta f_{tab.}$	$\Sigma \chi^2_{calc.m\acute{a}x}$	$\chi^2_{tab.}$
Gumbel	0,1101 ^A	0,21	4,509 ^A	9,488
Log-normal 2P	0,0949 ^A	0,21	0,153 ^A	7,815
Log-normal 3P	0,1148 ^A	0,21	0,144 ^A	5,991
Gama	0,1111 ^A	0,21	0,163 ^A	7,815

A – Test adequacy

The minimum flow rates (m³/s) and respective indexes for different periods of return are presented at Table 3. Analyses of the adherence values for all distributions listed at table 4, resulted in good fitting for all distributions, being the Gama function the one that provided the best quality of fitting due to its smaller χ^2 calculated value. Following this function, Log-normal 2 and 3 parameters as well as Weibull also provided good fittings according to the χ^2 statistics.

Silvino et al (2007) also obtained good fitting using Gama function through χ^2 tests for Paraguay River flow rate data analysis.

Table 3 – Minimum flow rates (Q) and minimum specific flow rate indexes (RE) for different periods of return at Itabapoana flow rate station.

Probability function	Returnperiod (years)	Q (m ³ /s)	RE (L/s km ²)
Gumbel	2	14,72	5,16
	5	11,35	3,98
	10	9,11	3,19
	20	6,97	2,44
	50	4,2	1,47
	100	2,12	0,74
Weibull	2	14,21	4,98
	5	10,82	3,79
	10	9,03	3,16
	20	7,6	2,66
	50	6,07	2,13
	100	5,13	1,80
Log 2P	2	14,54	5,09
	5	11,5	4,03
	10	10,18	3,57
	20	9,18	3,22
	50	8,21	2,88
	100	7,59	2,66
Log 3P	2	14,82	5,19
	5	11,53	4,04
	10	9,95	3,49
	20	8,68	3,04
	50	7,39	2,59
	100	6,52	2,28
Gama	2	14,43	5,06
	5	12,03	4,22
	10	10,79	3,78
	20	9,37	3,28
	50	9,37	3,28
	100	9,08	3,18

Table 4 – Adherence test results for probability functions used to estimate the minimum flow rates.

Distribution	KS		χ^2	
	$\Delta f_{calc.m\acute{a}x}$	$\Delta f_{tab.}$	$\Sigma \lambda^2_{calc.m\acute{a}x}$	$\lambda^2_{tab.}$
Gumbel	0,1548 ^A	0,21	4,245 ^A	9,488
Log-normal 2P	0,1402 ^A	0,21	0,125 ^A	7,815
Log-normal 3P	0,1520 ^A	0,21	0,090 ^A	7,815
Gama	0,1753 ^A	0,21	0,075 ^A	9,488
Weibull	0,1080 ^A	0,21	0,102 ^A	7,815

A – Test adequacy

Conclusions

- Investigation of best fitting models to estimate flow rate extreme values is very important to gain safety for water structure design as well as management of watersheds.
- The Log-normal 2 parameters represented highest security on estimating flow values for periods of return larger than 10 years.
- Gumbel and Gama probability functions were adequate to fit the maximum daily flow rates. Log-normal 2 parameters presented, however, the best fitting according to Kolmogorov-Smirnov tests. Log-normal 3 parameters best fitted the values according to Chi-square test.
- Gumbel and Log-normal 2 and 3 parameters were adequate to fit the minimum daily values. Weibull function generated best fitting based on Kolmogorov-Smirnov test, while Gama function was the best according to chi-square test.

Modelos de probabilidade na determinação de vazões de extremos

Resumo

Analisou-se, neste trabalho, a aplicação de modelos de probabilidade às séries históricas de vazões máximas e mínimas diárias anuais para diversos períodos de retorno no rio Itabapoana, localizado entre a divisa

dos estados do Espírito Santo e Rio de Janeiro. A série histórica compreende as vazões máximas e mínimas anuais entre 1968 e 2007 resultando em 40 anos de dados. Na verificação da adequação das distribuições, foram utilizados os testes de Kolmogorov-Smirnov e Qui-quadrado. Observou-se que, para as vazões máximas, as distribuições que obtiveram maior aderência foram a Log-normal 2e 3 parâmetros e para as vazões mínimas, as distribuições Weibull e Gama. Com isto, puderam-se estimar com maior grau de confiabilidade, as vazões máximas e mínimas para o rio Itabapoana.

Palavras-chave: Modelos de probabilidade. Vazão. Índices de vazão específica

Literature

ANA (Agência Nacional das Águas). Sistema Nacional de Informações hidrológicas. Disponível em: <<http://www.ana.gov.br/>> Acesso em: abr. 2011.

CLARKE, R. T. Hidrologia Estatística. in **Hidrologia Ciência e Aplicação**. Porto Alegre: Editora da UFRGS. 1993. P. 659 - 698.

DOUGLAS, E. M.; VOGEL, R. M. Probabilistic behavior of floods of record in the United States. **Journal of Hydrologic Engineering**, Reston, v.11, n.5, p.482-488, 2006.

EUCLYDES, H.P. Regionalização de vazão máxima, mínima e média de longo período e da curva de permanência para a bacia do Alto Médio São Francisco, In: **Tecnologia para elaboração de projetos hidroagrícolas em bacias**. Viçosa: FV/RURALMINAS, 2002 p. 1-142. (Boletim Técnico, n.6)

EUCLYDES, H. P.; PICCOLO, M.J. **Aplicação de um critério de semelhança hidrológica pela distribuição de probabilidade de vazão de pico de cheia da Microregião do Circuito das Águas - Sul de Minas Gerais**. 1985. (Monografia). Istituto di Idraulica e Costruzioni Idrauliche del Politecnico di Milano, Itália.

FERREIRA, D. F. **Estatística básica**. Lavras: UFLA, 2005. 654 p.

KAVISKI, E. Vazões de estiagens em pequenas bacias hidrográficas do estado de Santa Catarina. In: Simpósio Brasileiro de Hidrologia e Recursos Hídricos, 5, 1983. Florianópolis. **Anais...** Florianópolis: ABRH, 1983, p.43-67.

HAAN, C. T. **Statistical methods in hydrology**. 2 ed. Ames: The Iowa State University, 2002. 377p.

MAIONE, U. **Apontamenti di idrologia – Le piene fluviali**. Pavia, La Goliardica Pavese. 1977, 224p.

MELLO, C. R.; VIOLA, M. R.; BESKOW, S. Vazões máximas e mínimas para bacias hidrográficas da região Alto Rio Grande, MG. **Ciência e Agrotecnologia**, Lavras, v. 34, n.2, p. 494-502, mar/abr., 2010.

NAGHETTINI, M.; PINTO, E. J. A. **Hidrologia Estatística**. Belo Horizonte: CPRM, 2007. 552p.

POLÍTICA NACIONAL DE RECURSOS HÍDRICOS. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/l9433.htm> Acesso em: 06 jun. 2014.

REIS, J. A. T.; GUIMARÃES, M. A.; BARRETO NETO, A. A.; BRINGHENTI, J. Indicadores regionais aplicáveis à avaliação do regime de vazão dos cursos d'água da bacia hidrográfica do Rio Itabapoana. **Geociências**, v. 27, n.4, p. 509-516, São Paulo, UNESP. 2008.

SILVINO, A. N. O.; SILVEIRA, A.; MUSIS, C. R.; WYREPKOWSKI, C. C. Determinação de vazões extremas para diversos períodos de retorno para o Rio Paraguai utilizando métodos estatísticos. **Geociências**, v. 26, n.4, p. 369-378, São Paulo, UNESP. 2007.

TUCCI, C. E. M. **Regionalização de vazões**. Ed. Universidade/UFRGS, 2002, 256 p.

_____. **Hidrologia: ciência e aplicação**. 2. ed. Porto Alegre, RG: UFRGS, 2001. 943 p.

_____. **Hidrologia: ciência e aplicação**. Porto Alegre: Ed. da Universidade: ABRH: EDUSP, 1997.

Histórico

Submetido em: 29/07/2013

Aceito em: 16/07/2014