

# CADERNOS DO IME – Série Estatística

Universidade do Estado do Rio de Janeiro - UERJ  
ISSN on-line 2317-4536 / ISSN impresso 1413-9022 - v.52, p.20-33, 2022  
DOI: 10.12957/cadest.2022.71381

## INTERVALOS DE CONFIANÇA PARA QUANTIS DE MÁXIMOS ANUAIS DE ACUMULADOS DE DADOS ACIMA DE UM LIMIAR GERADOS POR MODELO ESTOCÁSTICO NÃO-PARAMÉTRICO DE VAZÕES DIÁRIAS

Jorge Machado Damázio  
UERJ – Universidade do Estado do Rio de Janeiro  
CEPEL- Centro de Pesquisas de Energia Elétrica  
[damazio@ime.uerj.br](mailto:damazio@ime.uerj.br)

Daniela de Souza Kyrillos  
CEPEL- Centro de Pesquisas de Energia Elétrica  
[dani@cepel.br](mailto:dani@cepel.br)

### Resumo

*Apresenta-se resultados de testes com diferentes métodos de cálculo de intervalos de confiança para quantis de máximos anuais de acumulados de dados acima de um limiar gerados por modelo estocástico não-paramétrico de vazões diárias cuja formulação implica em limite superior para vazões e para máximos anuais dos acumulados. O modelo estocástico foi calibrado com registros fluviométricos de afluentes diários ao aproveitamento hidroelétrico de Salto Santiago no rio Iguaçu, PR, Brasil, e utilizado para gerar amostras dos máximos anuais com  $N=12.000$ ,  $20.000$ ,  $100.000$  e  $120.000$ . Para cada método foram calculados intervalos de confiança a 90% de quantis com recorrência de 30 anos e 400 anos. Foram testados dois métodos paramétricos: a aproximação gaussiana válida para amostras grandes e uso da transformação da integral de probabilidades (método TIP). Como distribuição populacional dos máximos anuais acumulados considerou-se distribuições mistas, Gama ou de Kamaraswamy, com massa de probabilidade no valor nulo. Foram testados também três métodos não-paramétricos: uma formulação não-paramétrica para o método TIP, o bootstrap por percentis e o bootstrap básico. Os resultados indicaram diversidade nos intervalos das diferentes metodologias para tempo de recorrência de 400 anos e  $N=12.000$  ou  $20.000$ . Para as amostras maiores os intervalos das diferentes metodologias são, na prática, muito similares para qualquer das duas recorrências testadas. Os testes permitiram recomendar o uso dos métodos não paramétricos.*

**Palavras-chave:** Intervalos de Confiança para Quantis, Métodos Paramétricos e Não Paramétricos.

## 1. Introdução

O ONS, operador nacional do sistema elétrico brasileiro, em suas atividades de planejamento da operação, calcula anualmente volumes a serem alocados no ano seguinte nos reservatórios de algumas hidrelétricas para o amortecimento das possíveis cheias do próximo ano considerando um determinado tempo de recorrência de inundações nos vales de jusante das barragens. Para o cálculo destes volumes, denominados “volumes de espera”, o ONS utiliza ferramentas computacionais específicas inseridas dentro do Sistema para Estudo de Prevenção de Cheias – SPEC, desenvolvido pelo CEPEL. A boa qualidade dos cálculos de volumes de espera permite precisão na redução dos riscos de inundações, garantindo uma operação durante a estação chuvosa sem desperdícios de água, evitando-se prejudicar desnecessariamente o atendimento à demanda de energia elétrica. Em geral os tempos de recorrência utilizados para a proteção das áreas de jusante se situam entre 30 a 50 anos enquanto que as séries históricas de aflúências aos aproveitamentos não passam de 70 anos. Visando a maior precisão das estimativas, a metodologia adotada no sistema SPEC utiliza um modelo estocástico de séries temporais não-paramétrico, modelo DIANA-SPEC (KELMAN, DAMÁZIO, & COSTA, 1983), calibrado a partir do histórico para geração de um número grande de cenários de sequências de aflúências diárias para a estação chuvosa do próximo ano. Para cada cenário gerado o SPEC calcula curvas de volumes de espera ao longo da estação chuvosa com base em envoltórias das chamadas trajetórias críticas, definidas pelos acumulados acima de um limiar de defluência, sendo o limiar de defluência definido exogenamente à metodologia pela maior defluência que não traz prejuízos (restrição de vazão máxima defluente do reservatório). Para as recorrências entre 30 e 50 anos os cálculos de volume de espera utilizam 12.000 cenários denominados séries sintéticas de vazões diárias afluentes aos reservatórios. No entanto em alguns pontos do sistema hidráulico do SIN vem sendo adotado tempos de recorrência maiores, no entorno dos 400 anos, visando a proteção das próprias estruturas do aproveitamento em seu canal de fuga, com é o caso do reservatório de Salto Santiago. Como a precisão do cálculo de volume de espera decresce com o tempo de recorrência adotado, nestes casos de recorrência alta, a precisão dos cálculos é menor, podendo, no entanto, ser compensada com o aumento do número de séries sintéticas. Uma abordagem que pode ser utilizada na avaliação da necessidade de aumento no número de séries sintéticas é o uso de intervalos de confiança de quantis.

Este artigo utiliza o máximo dos acumulados acima do limiar de cada trajetória crítica como métrica univariada da necessidade de volume de espera associado a cada série sintética e como parâmetro de volume de espera a estimativa do quantil  $1-(1/TR)$  desta métrica obtida com o conjunto de séries sintéticas geradas, onde TR é a recorrência desejada. Baseando-se nesta estimativa, o artigo apresenta uma comparação entre diversos métodos de cálculo de intervalos de confiança de quantis em termos de estabilidade dos intervalos de confiança calculados para o aproveitamento de Salto Santiago para diferentes recorrência e número de séries sintéticas.

## 2. Intervalos de Confiança para Volumes de Espera

Para cada série sintética  $i$  de vazões diárias, a correspondente trajetória crítica definem os volumes de espera necessários para cada dia  $t$  da estação chuvosa e são calculadas pela recursão (1a) e (1b):

$$v(T, i) = 0 \quad (1a)$$

$$v(t - 1, i) = \max(q(t, i) - QR + v(t, i), 0), \quad t = T, T - 1, \dots, 0 \quad (1b)$$

onde,  $v(t, i)$  é o volume de espera no dia  $t$ , QR é a restrição de vazão defluente máxima,  $q(t, i)$  é a vazão afluente do dia  $t$ ,  $t=0$ , marca o início da estação chuvosa e  $t=T$  o fim da estação chuvosa.

Desprezando-se a sazonalidade dos regimes de cheias, o volume de espera a ser alocado em um reservatório pode ser tratado como uma métrica univariada calculada para cada cenário de afluências diárias pelo máximo da correspondente trajetória crítica (2):

$$ve(i) = \max(v(t, i), \quad t = T, \dots, 0) \quad (2)$$

Na vida real, o volume de espera da próxima estação chuvosa é modelado como uma variável aleatória contínua VE cuja distribuição de probabilidades acumulada é dada por  $F_{VE}(\cdot)$ . Supõe-se que  $F_{VE}(\cdot)$  é totalmente desconhecida (métodos não-paramétricos) ou conhecida a menos de seus parâmetros (métodos paramétricos).

Considere disponibilizado um conjunto de volumes de espera  $\{ve_1, ve_2, \dots, ve_N\}$  calculados pelo modelo DECA-SPEC para cada cenário de afluências diárias gerado pelo modelo DIANA-SPEC. No SPEC este conjunto é visto como uma amostra aleatória de tamanho  $N$  de VE. A partir desta amostra procura-se estimar o quantil  $p$  de VE, denotado  $\zeta_p$  e definido como  $\zeta_p = F_{VE}^{-1}(p)$ , onde  $p$  é a probabilidade correspondente a TR, tempo de recorrência desejado, através de  $p=1-(1/TR)$ .

O modelo DECA-SPEC adota uma abordagem não-paramétrica para estimar  $\zeta_p$ . A amostra  $\{ve_1, ve_2, \dots, ve_N\}$  é ordenada no vetor  $y=(y_i, i=1, \dots, N)$  com  $y_1 \leq y_2 \leq \dots \leq y_N$  e a estimativa é obtida pelo  $N^*p$ -ésimo valor da amostra ordenada (3).

$$\hat{\zeta}_p = y_{N^*p} \quad (3)$$

A estimativa em (3) aplicada à amostra fornece o valor requerido de volume de espera, no entanto, como toda estimativa, está sujeita a variações amostrais e, portanto, a diferenças em relação ao valor preciso  $\zeta_p$ . Qualitativamente sabe-se que precisão desta estimativa depende da recorrência adotada (TR) e do número de cenários de afluência utilizados (N), porém uma métrica de precisão ajudaria a calibrar o número de cenários a ser utilizado no cálculo de volumes de espera. Na literatura clássica de inferência costuma-se acompanhar estimativas de um intervalo no seu entorno com uma medida de confiança de que o desconhecido valor preciso esteja dentro do intervalo. Estes intervalos são chamados de “intervalos de confiança  $(1-2\alpha)\%$ ” ou “IC $(1-2\alpha)\%$ ”, onde  $(1-2\alpha)\%$ , comumente chamado de coeficiente de confiança do intervalo, é a probabilidade de o intervalo conter o valor preciso.

A derivação de uma fórmula de cálculo de IC's para um parâmetro populacional de uma forma geral segue os procedimentos analíticos explicados a seguir, tomando-se como exemplo de parâmetro populacional o quantil  $\zeta_p$ . O primeiro passo para a derivação de IC é encontrar uma função da amostra e de  $\zeta_p$ ,  $Z = g(ve_1, ve_2, \dots, ve_N, \zeta_p)$ , tal que a distribuição de Z não dependa de parâmetros desconhecidos, inclusive de  $\zeta_p$  (MOOD, GRAYBILL & BOES, 1974). Tendo sido encontrada uma função com esta característica, a distribuição de Z permite escrever declarações de probabilidades para intervalos (a,b) em Z do tipo (4):

$$\Pr[a < Z < b] = (1-2\alpha)\% \quad a, b \text{ e } \alpha \text{ conhecidos} \quad (4)$$

Substituindo Z em (4) pela função  $g(ve_1, ve_2, \dots, ve_N, \zeta_p)$  obtêm-se (5):

$$\Pr[a < g(ve_1, ve_2, \dots, ve_N, \zeta_p) < b] = (1-2\alpha)\% \quad (5)$$

Por uma manipulação de (5) denotada pivoteamento, isola-se  $\zeta_p$  no interior no intervalo, obtendo-se a fórmula para o desejado intervalo para  $\zeta_p$  (6):

$$\Pr[T_1 < \zeta_p < T_2] = (1-2\alpha)\% \quad T_1 < T_2 \quad (6)$$

onde  $T_1$  e  $T_2$  são estatísticas calculadas na amostra (7a) e (7b):

$$T_1 = l_1(ve_1, ve_2, \dots, ve_N) \quad (7a)$$

$$T_2 = l_2(ve_1, ve_2, \dots, ve_N) \quad (7b)$$

Observe-se que a declaração de probabilidade em (6) se refere conceitualmente à coleta sequencial de inúmeras amostras tamanho  $N$  com cálculo para cada uma de intervalos  $(T_1, T_2)$  conforme (7a) e (7b). Em  $(1-\alpha)\%$  das amostras os intervalos calculados conterão o valor exato de  $\zeta_p$

Chamamos atenção a dois aspectos da discussão acima. Considerando que o coeficiente de confiança  $(1-\alpha)\%$  que se deseja esteja pré-definido, existirá sempre uma infinidade de intervalos  $(a, b)$  capazes de atender a declaração de probabilidades em (4). Uma forma clássica de resolver esta questão é procurar otimizar o intervalo de confiança minimizando o tamanho do intervalo em (6). O segundo aspecto refere-se à confiabilidade da declaração em (6). Como diversas aproximações são utilizadas para se chegar à diferentes formulações (7a) e (7b), o coeficiente de confiança nominal da formulação utilizada é sempre uma aproximação. Algumas aproximações funcionam bem para amostras grandes (intervalos de confiança assintóticos) ou quando as distribuições populacionais são especificadas, por exemplo gaussianas. No caso prático o tamanho de amostra é finito e ou não se conhece a distribuição populacional, ou já se sabe que esta não necessariamente é a população presumida pela formulação. O coeficiente de confiança na prática obtido por uma dada formulação para uma dada população e para um dado tamanho de amostra, denominado cobertura, em geral é menor que o nominal. Estudos de simulação em computadores são usados para estimar curvas de cobertura para as formulações (LAU, GONZALEZ, & NOLA, 2020).

Os métodos de cálculo de intervalos de confiança da literatura são classificados em métodos paramétricos (válidos para distribuições paramétricas específicas) e não-paramétricos (válidos para qualquer distribuição paramétrica). A seguir apresenta-se alguns métodos paramétricos e não paramétricos para cálculo de  $IC(1-2\alpha)\%$  para  $\zeta_p$ .

## 2.1 Intervalos de Confiança Paramétricos

A estimativa em (4) é não tendenciosa,  $E(\hat{\zeta}_p) = \zeta_p$ , e tem variância dada por:

$$\text{var}(\hat{\zeta}_p) = (p*(1-p))/(N*f_{VE}(\zeta_p)*f_{VE}(\zeta_p)) \quad (8)$$

Um intervalo de confiança  $IC(1-\alpha)\%$  aproximado para amostras grandes (MOOD, GRAYBILL & BOES, 1974, p. 394) pode ser obtido considerando a aproximação

assintótica normal para a distribuição amostral da estatística de ordem  $y_{N^*p}$  e escrito como (9):

$$\hat{\zeta}_p \pm u_{1-\alpha} \sqrt{\text{var}(\hat{\zeta}_p)} \quad ; u \sim N(0,1) \quad (9)$$

O cálculo do intervalo de confiança conforme (9), por adotar a aproximação normal, é um intervalo simétrico em torno de  $\hat{\zeta}_p$ , o que pode não ser uma boa aproximação quando a distribuição amostral de  $Y_{N^*p}$  é assimétrica, o que pode acontecer com  $N$  pequeno. Mais além, (9) exige obtenção das densidades de probabilidades  $f_{VE}(\zeta_p)$  para o cálculo da variância das estimativas em (8) o que pode ser feito supondo uma distribuição paramétrica e estimando seus parâmetros com a amostra. Para as situações em que não se deseja supor uma distribuição paramétrica, Venables e Ripley (1999) sugerem obter as densidades de probabilidades usando estimadores de densidade não-paramétricos mas reportam graus significativos de tendência e variabilidade nestas estimativas.

Loucks e Van Beek (2017) apresentam uma fórmula para cálculo de IC paramétrico para quantis fazendo uso da transformação da integral de probabilidades  $U_i = F_{VE}(y_i)$ , que será denominada aqui como transformação TIP. Os autores mostram que  $U_i$  tem distribuição Beta( $a=i$ ,  $b=N-i+1$ ) e propõem o cálculo IC(1-2 $\alpha$ )% para  $\zeta_p$  a partir do IC(1-2 $\alpha$ )% para  $U_{N^*p}$  conforme (10):

$$(F_{VE}^{-1}(b_{N^*p}^{(\alpha)}), F_{VE}^{-1}(b_{N^*p}^{(1-\alpha)})) \quad (10)$$

onde  $b_{N^*p}^{(\alpha)}$  e  $b_{N^*p}^{(1-\alpha)}$  são obtidos integrando a distribuição Beta( $a=N^*p$ ,  $b=N-N^*p+1$ ) e  $F_{VE}^{-1}(\cdot)$  é a inversa da distribuição de probabilidades acumuladas de VE. Na abordagem paramétrica, pressupõe-se uma distribuição paramétrica para  $F_{VE}(\cdot)$ , e calibra-se seus parâmetros com a amostra disponível.

O intervalo em (10) é construído para se situar no entorno de estimativas de  $\zeta_p$  obtidas da distribuição paramétrica ajustada  $\hat{\zeta}_p = F_{VE}^{-1}(p)$ . Uma forma de acomodar este intervalo com a estimativa amostral em (3) é ajustar a distribuição paramétrica de forma a garantir (11):

$$F_{VE}^{-1}(p) = \hat{\zeta}_p \quad (11)$$

A modelagem paramétrica das distribuições de volumes de espera foi feita adotando distribuições mistas com massa de probabilidade  $p_0$  para volume de espera nulo

e distribuição contínua para os volumes espera positivos. Para a distribuição contínua verificou-se duas alternativas de modelagem: a distribuição gama mista e a distribuição Kumaraswamy (KUMARASWAMY, 1980; BAYER, DÉBORA, & PUMI, 2017). A distribuição gama admite volumes de espera por todo o intervalo de valores positivos  $(0, +\infty)$  enquanto a distribuição Kumaraswamy utiliza num intervalo limitado superiormente  $(0, b)$  onde  $b$  é um parâmetro para calibração. O intervalo limitado superiormente teoricamente é compatível com o modelo DIANA-SPEC.

## 2.2 Intervalos de Confiança Não-Paramétricos

O método da transformação TIP torna-se um método não-paramétrico se os limites do IC em (10) forem obtidos por interpolação na distribuição de probabilidades acumuladas amostral dada pelo mapeamento dos volumes de espera ordenados  $y_i$  em posições de plotagem  $pp_i$ ,  $i=1, \dots, N$  usando (12):

$$pp_i = \frac{i-a}{N+1-2*a} \quad i=1, \dots, N \quad (12)$$

onde  $a=0$  corresponde ao uso como posição de plotagem as estimativas de probabilidades de não-excedência não tendenciosas.

Outra abordagem para cálculo de IC não-paramétrico é o uso de métodos de Reamostragem como o Jackknife e o Bootstrap (EFRON, 1982).

O bootstrap, de forma geral, utiliza cálculos da estatística em diferentes reamostragens com repetição da amostra disponível. Seja  $B$  o número de reamostragens e  $\hat{\zeta}_{p,i}^*$   $i=1, \dots, B$  as estimativas bootstrap obtidas em cada reamostragem de bootstrap. A concepção do bootstrap é de que é possível inferir a variabilidade de  $\hat{\zeta}_p$  entorno de  $\zeta_p$  a partir da variabilidade das estimativas  $\hat{\zeta}_{p,i}^*$  entorno de  $\hat{\zeta}_p$ . A literatura apresenta diversas alternativas para o cálculo de  $IC(1-\alpha)\%$  a partir das reamostragens bootstrap. As duas principais são:

bootstrap por percentis (13):

$$(k_{(\alpha)}, k_{(1-\alpha)}) \quad (13)$$

onde  $k_{(\alpha)}$  e  $k_{(1-\alpha)}$  são percentis de  $\hat{\zeta}_p^*$ .

bootstrap básico (14):

$$(2\hat{\zeta}_p - k_{(1-\alpha)}, 2\hat{\zeta}_p - k_{(\alpha)}) \quad (14)$$

Observa-se que o intervalo (14) é calculado como a reflexão do intervalo em (13) no entorno de  $\hat{\zeta}_p$ .

### 3. Testes

Os métodos de cálculo de intervalos de confiança para volumes de espera discutidos acima foram testados utilizando às séries históricas de afluições diárias ao reservatório de Salto Santiago utilizadas nos Ciclos de Planejamento do ONS de 21-22 e 22-23 respectivamente. A série histórica usada no ciclo de 22-23, além adicionar as medições fluviométricas de outubro de 2019 a setembro de 2020, traz ainda uma mudança no dia inicial do ano operacional. No caso do aproveitamento de Salto Santiago esta mudança de dia inicial do ano tem um potencial de impacto no cálculo de volume de espera devido à algumas vazões extremas que passaram a ser incorporadas no histórico do início do ano operacional.

O modelo DIANA-SPEC foi calibrado separadamente para cada uma das duas séries e utilizado para gerar conjuntos de N cenários anuais de afluições diárias para N=12.000, 20.000, 100.000 e 120.000.

Os estudos oficiais do ONS para cálculo de volume de espera em Salto Santiago visam garantir a recorrência de 400 anos para a defluência máxima de 19.000 m<sup>3</sup>/s, valor este correspondente em condições naturais à decorrência de cerca de 200 anos. Além do caso oficial, as análises neste trabalho consideraram também um caso hipotético de cálculo de volume de espera em Salto Santiago visando garantir a recorrência de 30 anos para a defluência máxima de 8.000 m<sup>3</sup>/s, correspondente em condições naturais à decorrência de 15 anos. O modelo DECA-SPEC foi utilizado para calcular os volumes de espera do caso oficial para os conjuntos de cenários com N=12.000, 100.000 e 120.000, e do caso hipotético para os conjuntos de cenários com N=12.000, 20.000 e 100.000.

Para cada conjunto de cenários foram calculados IC90% paramétricos (Normal e TIP) e não-paramétricos (TIP, bootstrap por percentis e bootstrap básico). Os IC90% obtidos foram analisados em termos do comprimento e em termos de simetria. Usou-se coeficientes adimensionais de comprimento e de simetria centrados na estimativa  $\hat{\zeta}_p$ , denotados respectivamente por LAD e SK, e calculados por (15) e (16):

$$LAD = 100 \sqrt{\frac{\frac{1}{2}((LINF - \hat{\zeta}_p)^2 + (LSUP - \hat{\zeta}_p)^2)}{\hat{\zeta}_p}} \quad (15)$$



$$SK = \frac{\frac{1}{2}((LINF - \hat{\zeta}_p)^3 + (LSUP - \hat{\zeta}_p)^3)}{\left(\frac{1}{2}((LINF - \hat{\zeta}_p)^2 + (LSUP - \hat{\zeta}_p)^2)\right)^{1.5}} \quad (16)$$

onde LINF e LSUP são respectivamente os limites inferior e superior do intervalo.

## 4. Resultados

### 4.1 Intervalos de Vonfiança a 90%. Caso Oficial

As tabelas 1 a 6 apresentam os resultados dos cálculos de IC90% no caso oficial pelos diversos métodos para  $N=12.000$ ,  $N=100.000$  e  $N=120.000$  respectivamente. Uma observação interessante é que para qualquer método e  $N$ , para qualquer par de intervalos existe intercessão indicando uma certa compatibilidade entre os cenários gerados e os intervalos de confiança calculados no estudo. Pode-se também notar que conforme esperado, em todos os métodos ocorre a redução do comprimento dos IC90% quando se passa de  $N=12.000$  para  $N=100.000$ . A redução no comprimento também ocorre quando se passa de  $N=100.000$  para  $N=120.000$ , embora com menor intensidade. Também pode-se notar que para qualquer método cresce com  $N$  as proximidades entre limites dos IC90% e seus indicadores SK e LAD dos dois ciclos mostrando o esperado aumento da estabilidade dos resultados com o aumento do número de séries sintéticas.

Na análise comparativa entre métodos pode-se observar uma diferença importante para  $N=12.000$  nos comprimentos dos IC90% paramétricos e não paramétricos. Para  $N=12.000$ , nos dois ciclos os comprimentos dos IC90% paramétricos foram bem inferiores aos tamanhos dos IC90% não-paramétricos, embora no ciclo 22-23 esta diferença ocorra de forma menos intensa. Observa-se também que o uso da distribuição de Kumaraswamy fornece IC90%'s maiores que os IC90%'s obtidos como uso da distribuição gama.

De forma geral para  $N=100.000$  e para  $N=120.000$  os intervalos dos diferentes métodos são muito próximos entre si.

Em termos de simetria dos IC90%, o único destaque se refere aos IC90% com assimetria negativa postulados pelo método Bootstrap Básico.

Tabela 1. IC90% de Volume de Espera. Recorrência 400 anos. Salto Santiago. N=12.000. Caso Oficial.  
Ciclo 21-22.  $\hat{\zeta}_{0,9975} = 191 \text{ Hm}^3$

	Paramétrico Gama		Paramétrico Kumaraswamy		Não-Paramétrico		
	Normal	TIP	Normal	TIP	TIP	Boot Percentil	Boot Básico
intervalo	(-39, 421)	(25, 461)	(-407, 789)	(0, 1148)	(58, 945)	(56, 1025)	(-643, 326)
SK	0	0,67	0	1,32	1,34	1,35	-1,35
LAD	120%	114%	313%	361%	283%	313%	313%

Tabela 2. IC90% de Volume de Espera. Recorrência 400 anos. Salto Santiago. N=12.000. Caso Oficial.  
Ciclo 22-23.  $\hat{\zeta}_{0,9975} = 580 \text{ Hm}^3$

	Paramétrico Gama		Paramétrico Kumaraswamy		Não-Paramétrico		
	Normal	TIP	Normal	TIP	TIP	Boot Percentil	Boot Básico
intervalo	(273, 886)	(293, 893)	(70, 1090)	(164, 1115)	(274, 1191)	(255, 1232)	(-72, 905)
SK	0	0,13	0	0,37	0,89	0,89	-0,89
LAD	53%	52%	88%	83%	84%	89%	89%

Tabela 3. IC90% de Volume de Espera. Recorrência 400 anos. Salto Santiago. N=100.000. Caso Oficial.  
Ciclo 21-22.  $\hat{\zeta}_{0,9975} = 371 \text{ Hm}^3$

	Paramétrico Gama		Paramétrico Kumaraswamy		Não-Paramétrico		
	Normal	TIP	Normal	TIP	TIP	Boot Percentil	Boot Básico
intervalo	(288, 454)	(289, 497)	(262, 480)	(266, 483)	(281, 497)	(295, 500)	(242, 447)
SK	0	0,04	0	0,10	0,48	0,71	-0,71
LAD	22%	22%	29%	29%	29%	29%	29%

Tabela 4. IC90% de Volume de Espera. Recorrência 400 anos. Salto Santiago. N=100.000. Caso Oficial.  
Ciclo 22-23.  $\hat{\zeta}_{0,9975} = 394 \text{ Hm}^3$

	Paramétrico Gama		Paramétrico Kumaraswamy		Não-Paramétrico		
	Normal	TIP	Normal	TIP	TIP	Boot Percentil	Boot Básico
intervalo	(307, 481)	(309, 482)	(295, 493)	(298, 495)	(313, 478)	(321, 480)	(308, 467)
SK	0	0,06	0	0,08	0,06	0,24	-0,24
LAD	22%	22%	25%	25%	21%	20%	20%

Tabela 5. IC90% de Volume de Espera. Recorrência 400 anos. Salto Santiago. N=120.000. Caso Oficial.  
Ciclo 21-22.  $\hat{\zeta}_{0,9975} = 398 \text{ Hm}^3$

	Paramétrico Gama		Paramétrico Kumaraswamy		Não-Paramétrico		
	Normal	TIP	Normal	TIP	TIP	Boot Percentil	Boot Básico
intervalo	(321, 475)	(322, 475)	(301, 495)	(303, 497)	(315, 493)	(314, 494)	(302, 482)
SK	0	0,03	0	0,06	0,20	0,20	-0,20
LAD	19%	19%	24%	18%	22%	23%	23%

Tabela 6. IC90% de Volume de Espera. Recorrência 400 anos. Salto Santiago. N=120.000. Caso Oficial.  
Ciclo 22-23.  $\hat{\zeta}_{0,9975} = 394 \text{ Hm}^3$

	Paramétrico Gama		Paramétrico Kumaraswamy		Não-Paramétrico		
	Normal	TIP	Normal	TIP	TIP	Boot Percentil	Boot Básico
intervalo	(314,474)	(316,475)	(301,487)	(304,489)	(313,480)	(313,480)	(308,475)
SK	0	0,06	0	0,08	0,09	0,09	-0,09
LAD	20%	20%	24%	24%	21%	21%	21%

#### 4.2 Intervalos de confiança a 90%. Caso Hipotético

As tabelas 7 a 12 apresentam os resultados dos cálculos de IC90% no caso hipotético pelos diversos métodos para N=12.000, N=20.000 e N=100.000 respectivamente. Da mesma forma que no caso oficial manteve-se as intercessões entre intervalos de confiança dos diferentes ciclos. De forma geral os comprimentos dos IC90% do caso hipotético são menores que os correspondentes comprimentos dos IC90% do caso oficial, refletindo a maior incerteza das estimativas para recorrência de 400 anos. Na análise comparativa entre métodos pode-se observar para N=12.000 e N=20.000 comprimentos maiores dos IC90% dos métodos paramétricos. Para N=100.000, as diferenças de comprimentos dos IC90% dos diferentes métodos são muito pequenas.

Tabela 7. IC90% de Volume de Espera. Recorrência 30 anos. Salto Santiago. N=12.000. Caso Hipotético.  
Ciclo 21-22.  $\hat{\zeta}_{0,9667} = 617 \text{ Hm}^3$

	Paramétrico Gama		Paramétrico Kumaraswamy		Não-Paramétrico		
	Normal	TIP	Normal	TIP	TIP	Boot Percentil	Boot Básico
intervalo	(457, 778)	(465, 786)	(488,745)	(494,750)	(527,766)	(527,762)	(472,707)
SK	0	0,16	0	0,11	0,69	0,66	-0,66
LAD	26%	26%	21%	21%	20%	20%	20%

Tabela 8. IC90% de Volume de Espera. Recorrência 30 anos. Salto Santiago. N=12.000. Caso Hipotético.  
Ciclo 22-23.  $\hat{\zeta}_{0,9667} = 571 \text{ Hm}^3$

	Paramétrico Gama		Paramétrico Kumaraswamy		Não-Paramétrico		
	Normal	TIP	Normal	TIP	TIP	Boot Percentil	Boot Básico
intervalo	(412,730)	(423, 740)	(444,699)	(450,704)	(468,680)	(469,677)	(465,673)
SK	0	0,19	0	0,14	0,08	0,05	-0,07
LAD	28%	28%	22%	22%	19%	18%	18%

Tabela 9. IC90% de Volume de Espera. Recorrência 30 anos. Salto Santiago. N=20.000. Caso Hipotético. Ciclo 21-22.  $\hat{\zeta}_{0,9667} = 696 \text{ Hm}^3$ 

	Paramétrico Gama	Paramétrico Kumaraswamy		Não-Paramétrico			
	Normal	TIP	Normal	TIP	TIP	Boot Percentil	Boot Básico
intervalo	(566,827)	(571,831)	(591,801)	(594,804)	(599,784)	(596,775)	(615,794)
SK	0	0,11	0	0,08	-0,15	-0,36	0,27
LAD	19%	19%	15%	15%	13%	13%	13%

Tabela 10. IC90% de Volume de Espera. Recorrência 30 anos. Salto Santiago. N=20.000. Caso Hipotético. Ciclo 22-23.  $\hat{\zeta}_{0,9667} = 610 \text{ Hm}^3$ 

	Paramétrico Gama	Paramétrico Kumaraswamy		Não-Paramétrico			
	Normal	TIP	Normal	TIP	TIP	Boot Percentil	Boot Básico
intervalo	(483,738)	(489,743)	(502,718)	(506,722)	(544,715)	(534,706)	(514,684)
SK	0	0,14	0	0,11	0,64	0,37	-0,39
LAD	21%	21%	18%	18%	14%	14%	14%

Tabela 11. IC90% de Volume de Espera. Recorrência 30 anos. Salto Santiago. N=100.000. Caso Hipotético. Ciclo 21-22.  $\hat{\zeta}_{0,9667} = 748 \text{ Hm}^3$ 

	Paramétrico Gama	Paramétrico Kumaraswamy		Não-Paramétrico			
	Normal	TIP	Normal	TIP	TIP	Boot Percentil	Boot Básico
intervalo	(688,808)	(689,809)	(697,799)	(698,800)	(703,793)	(703,792)	(704,793)
SK	0	0,05	0	0,04	-0,01	-0,03	0,03
LAD	8%	8%	7%	7%	6%	6%	6%

Tabela 12. IC90% de Volume de Espera. Recorrência 30 anos. Salto Santiago. N=100.000. Caso Hipotético. Ciclo 22-23.  $\hat{\zeta}_{0,9667} = 683 \text{ Hm}^3$ 

	Paramétrico Gama	Paramétrico Kumaraswamy		Não-Paramétrico			
	Normal	TIP	Normal	TIP	TIP	Boot Percentil	Boot Básico
intervalo	(625,741)	(626,742)	(632,734)	(633,734)	(641,730)	(642,730)	(636,724)
SK	0	0,05	0	0,05	0,17	0,20	-0,20
LAD	9%	9%	7%	7%	7%	6%	6%

## 5. Conclusões

A análise dos resultados dos teste para o caso oficial de Salto Santiago dos Ciclos de Planejamento do ONS de 21-22 e 22-23 mostrou que a precisão dos calculos de volume de espera melhora quando se utiliza 100.000 séries sintéticas. A análise do caso hipotético de Salto Santiago para uma recorrência de 30 anos mostrou que a melhora da precisão quando se utiliza mais séries sintéticas é bem menor, podendo-se considerar adequado o uso de 12.000 séries sintéticas.

Os testes realizados com os métodos de cálculo de intervalos de confiança selecionados indicaram dois métodos não-paramétricos, o algoritmo TIP ou o bootstrap por percentil, como adequados para serem implementados no sistema SPEC. Recomenda-se utilizar o indicador adimensional de comprimento relativo do intervalo como critério para definição do número de séries sintéticas de vazões diárias adequado nos cálculos de volumes de espera.

## Referências

- BAYER, F. M.; DÉBORA, B. M.; PUMI, G. Kumaraswamy autoregressive moving average models for double bounded environmental data. **Journal of Hydrology**, 555, pp. 385-396, 2017.
- CEPEL. Sistema SPEC - Sistema para Estudos de Prevenção de Cheias - *DIANA 5.04, SIMRESC 1.0, CAEV 3.9, VESPO 4.0 e AUXARISCO 1.02* - **Manual de Metodologia**. Rio de Janeiro: CEPEL, 2012.
- COSTA, F. D.; DAMÁZIO, J. M.; NEVES, F. P. Sistema SPEC - Sistema para Estudos de Prevenção de Cheias em Sistemas Hidroelétricos. **Anais XV Simpósio Brasileiro de Recursos Hídricos**. Belo Horizonte: ABRH, 1999.
- DAMAZIO, J. M.; MARIEN, J. L.; COSTA, F. D. Building Flood Control Rules Curves for Multipurpose Multi-reservoirs System Using Controllability Conditions. **Water Resources Research**, 30, pp. 1134-1144, 1994.
- EFRON, B. **The Jackknife, the Bootstrap and Other Resampling Plans**. Philadelphia: Society for Industrial and Applied Mathematics, 1982.
- KELMAN, J.; DAMÁZIO, J. M.; COSTA, J. P. Geração de Séries Sintéticas de Vazões Diárias - Modelo Diana. **Revista Brasileira de Engenharia**, 1, pp. 5-22, 1983.
- KUMARASWAMY, P. A generalized probability density function for double-bounded random process. **Journal of Hydrology**, 46, pp. 79-88, 1980.
- KYRILLOS, D. Otimização da Programação Diária em Situação Normal de Cheias e em Situação de Emergência em Reservatórios de Sistemas Hidroelétricos. **Tese de Doutorado**, COPPE-UFRJ, Rio de Janeiro, 2006.
- LAU, S.; GONZALEZ, J.; NOLA, D. (19 de outubro de 2020). **Principles and Techniques of Data Science**. Fonte: <https://www.samlau.me/test-textbook/prereqs.html>
- LOUCKS, D. P.; VAN BEEK, E. **Water Resources Systems Planning and Management. An Introduction to Methods, Models and Applications**. Cham, Switzerland: Springer, 2017.
- MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. **Introduction to the Theory of Statistics**. Singapore: McGraw-Hill Book, 1974.
- ONS. **Diretrizes para as Regras de Operação de Controle de Cheias (Ciclo 2012-2013)**. Rio de Janeiro: ONS, 2012.
- ONS. **Plano Anual de Prevenção de Cheias (Ciclo 2015-2016)**. Rio de Janeiro: ONS.
- VENABLES, W.; RIPLEY, B. (1999). **Modern Applied Statistics with S-PLUS**. New York: Springer Verlag, 2015.

## CONFIDENCE INTERVALS FOR QUANTILES OF MAXIMUM ANNUAL DATA ACCUMULATIONS ABOVE A THRESHOLD GENERATED BY NON-PARAMETRIC STOCHASTIC MODEL OF DAILY FLOWS

### **Abstract**

*The paper presents results of tests with different methods for calculations of confidence intervals for quantiles of annual maximums of accumulated data above a threshold generated by a non-parametric stochastic model of daily flows whose formulation implies an upper limit for flows and for annual maximums of accumulated. The stochastic model was calibrated with fluviometric records of daily inflows to the Salto Santiago hydroelectric project on the Iguaçu River, PR, Brazil, and used to generate samples of annual maximums with  $N=12,000$ ,  $20,000$ ,  $100,000$  and  $120,000$ . For each method, 90% confidence intervals of 30-year and 400-year recurrence quantiles were calculated. Two parametric methods were tested: the Gaussian approximation valid for large samples and the use of the probability integral transformation (TIP method). As a population distribution of the accumulated annual maximums, mixed, Gamma or Kamaraswamy distributions were considered, with probability mass at zero value. Three non-parametric methods were also tested: a non-parametric formulation for the TIP method, the percentile bootstrap and the basic bootstrap. The results indicated diversity in the intervals of the different methodologies for recurrence time of 400 years and  $N=12,000$  or  $20,000$ . For larger samples, the intervals of the different methodologies are, in practice, very similar for any of the two recurrences tested. The tests allowed us to recommend the use of non-parametric methods.*

**Key-words:** *Confidence Intervals for Quantiles, Parametric and Non-parametric Methods.*