PICO DE CHEIA: DEVE–SE MODELAR VAZÕES DIÁRIAS OU MÁXIMAS ANUAIS?

por

Fernanda da Serra Costa[1], Jerson Kelman[1,2], Jorge M.Damázio[1,3]

Resumo— Este artigo aborda o problema da escolha do modelo para descrição de um fenômeno hidrológico considerando o grau de complexidade dos modelos em contra posição a quantidade de informação utilizada na etapa de estimação dos parâmetros do mesmo.

O caso específico estudado é a determinação do pico da cheia de 100 anos de recorrência, y_{100}. Para investigação deste problema é adotada a técnica estatística de reamostragem conhecida como Método Bootstrap.

INTRODUÇÃO

Qualquer estudo hidrológico passa sempre pela definição do universo das variáveis intervenientes no processo que se queira modelar. Quanto mais abrangente for este universo, mais complexo terá que ser o modelo e mais difícil será a estimação dos correspondentes parâmetros. Em compensação, a quantidade de informações das variáveis intervenientes tenderá a ser maior.

Neste artigo estuda–se um dilema situado neste contexto; o que é preferível: (a) determinar o pico da cheia de T anos de recorrência a partir exclusivamente dos máximos anuais ou (b) utilizar a totalidade dos registros diários de vazão?

Na alternativa (a) ajusta–se uma distribuição de probabilidade F(.) ao conjunto de m máximos $y = \{y_1, y_2,...,y_m\}$, obtidos de um registro de vazões diárias, y_T é obtido pela solução de $P(Y \geq y_T) = 1/T$. Nesta abordagem despreza–se uma grande parte das informações contidas no registro fluviométrico.

Na alternativa (b) utiliza–se um modelo estocástico para a série de vazões diárias. Como as vazões diárias, ao contrário dos máximos anuais, são variáveis aleatórias dependentes e apresentam sazonalidade, o modelo estocástico é bem mais complexos do que a distribuição de probabilidades adotada para modelar os máximos anuais. A desvantagem causada pela complexidade do modelo estocástico em termos de dificuldade para estimação dos parâmetros, pode ser compensada pelo grande número de informações que contribuirão na estimação do pico da cheia, conduzindo a uma possivelmente melhor estimativa de y_T.

[1]Pesquisador do Centro de Pesquisas de Energia Elétrica – CEPEL, Cidade Universitária, Ilha do Fundão, C.P. 2754 – CEP 20001, Rio de Janeiro. RJ. Tele: (021) 598.2244.
[2]Professor da COPPE/UFRJ, Rio de Janeiro, RJ.
[3]Professor colaborador da COPPE/UFRJ, Rio de Janeiro, RJ

A seguir apresenta–se uma avaliação do comportamento de cada uma destas abordagens através de experiências controladas tomando–se como base a técnica estatística de reamostragem conhecida como Bootstrap.

BOOTSTRAP

Seja $\underset{\sim}{x} = \{x_1, x_2,..., x_n\}$ uma *amostra* contendo n sorteios independentes da variável aleatória X. O método Bootstrap, desenvolvido por Efron (Efron, 1982), para determinação da distribuição de probabilidades da estimativa de um parâmetro qualquer de $f_x(.)$, assim como para estimação de seu desvio padrão, pode ser resumido no seguinte algoritmo:

1– Faz–se uma reamostragem com reposição das observações da amostra $\underset{\sim}{x}$ formando uma "pseudo–amostra" (ou amostra Bootstrap), $\underset{\sim}{x}^b_{boot}$

$$\underset{\sim}{x}^b_{boot} = \{x^*_1, x^*_2, ..., x^*_n\} \tag{1}$$

2– A partir da "pseudo–amostra" $\underset{\sim}{x}^b_{boot}$, pode–se calcular a estimativa do parâmetro Θ de interesse,

$$\hat{\Theta}^b_{boot} = g(x^*_1, x^*_2, ..., x^*_n) \tag{2}$$

Repetições independentes dos passos 1 e 2, fornecem $\hat{\Theta}^1_{boot}, \hat{\Theta}^2_{boot}, ..., \hat{\Theta}^b_{boot}$, estimativas do parâmetro Θ com as quais é possível determinar a distribuição empírica de probabilidades de $\hat{\Theta}^b_{boot}$.

O estimador Bootstrap do desvio padrão de $\hat{\Theta}$ é:

$$\hat{S}_{boot} = \{\frac{1}{B-1} \Sigma^B_{b=1} (\hat{\Theta}^b_{boot} - \hat{\Theta}_{boot})^2\}^{1/2} \tag{3}$$

onde:

$$\hat{\Theta}_{boot} = \frac{1}{B} \Sigma^B_{b=1} \Theta^b_{boot} \tag{4}$$

e B é o número de "pseudo amostras" utilizadas.

Bootstrap nos máximos anuais

Consideremos uma série constituída de m segmentos de n vazões diárias:

$x_{i,j}$= vazão diária $\quad$ i = 1, m (segmento) $\quad$ j = 1, n (dia)

$$\underset{\sim}{x}_1 = \{x_{1,1}, x_{1,2}, \ldots, x_{1,n}\}$$

.

.

.

$$\underset{\sim}{x}_i = \{x_{i,1}, x_{i,2}, \ldots, x_{i,n}\}$$

.

.

.

$$\underset{\sim}{x}_m = \{x_{m,1}, x_{m,2}, \ldots, x_{m,n}\} \qquad (5)$$

De cada segmento é retirada a vazão máxima, obtendo–se uma série de m vazões máximas:

$$y_1 = \max_{j=1,n} \{x_{1,j}\}$$

.

.

.

$$y_i = \max_{j=1,n} \{x_{i,j}\}$$

.

.

.

$$y_m = \max_{j=1,n} \{x_{m,j}\} \qquad (6)$$

Logo:

$$\underset{\sim}{y} = \{y_1, y_2, \ldots, y_m\} \qquad (7)$$

Como os máximos anuais são variáveis aleatórias independentes, podemos aplicar o Bootstrap diretamente aos m máximos anuais (y), obtendo–se B "pseudo" séries de máximos anuais.

A cada "pseudo" série de máximos anuais ajusta–se uma distribuição, a qual supõe–se que seja a distribuição de probabilidade da população a partir do qual a série $\{y_1, y_2, \ldots, y_m\}$ tenha sido amostrada. Através da extrapolação da distribuição dos y's, *obtem–se uma estimativa da vazão de tempo de recorrência* $\hat{y}_T$.

Desta forma cada "pseudo" série de máximos anuais produz uma estimativa $\hat{y}_T$. Como são obtidas B "pseudo" séries, são produzidas B estimativas $\hat{y}_T$, que formam a distribuição empírica de probabilidades de $\hat{y}_T$.

Podemos observar que nesta metodologia, nenhuma "pseudo" série de máximos anuais apresentará observações diferentes das observadas na série de máximos original. O que pode ocorrer é a repetição de valores, já que o resorteio no Bootstrap é com repetição.

Bootstrap nas vazões diárias

Aplicar o Bootstrap às vazões diárias requer alguns cuidados, pois em geral as vazões diárias são variáveis aleatórias dependentes, modeladas por um processo estocástico e neste caso o Bootstrap não pode ser diretamente aplicado as vazões. Uma forma de contornar este problema, é aplicar o Bootstrap aos ruídos do processo *estocástico, que são variáveis aleatórias independentes e identicamente distribuídas*, (Unny, 1985), como será visto a seguir.

Consideremos uma série de vazões diárias constituída de m segmentos de n vazões, equação (5).

Através da série formada pelas m.n vazões diárias são estimados os parâmetros $\hat{\underset{\sim}{\Theta}} = \{\Theta_1, \Theta_2, \ldots, \Theta_L\}$ e $\underset{\sim}{\Phi} = \{\Phi_1, \Phi_2, \ldots, \Phi_K\}$ de um processo estocásticos ARMA (L,K):

$$x_{i,j} = \sum_{\ell=1}^{L} \Theta_\ell x_{i,j-\ell} + \sum_{k=0}^{K} \Phi_k a_{i,j-k} \qquad K < n \qquad L < n \qquad (8)$$

Com os parâmetros estimados do modelo e os segmentos $\{x_{i,1}, \ldots, x_{i,n}\}$ de vazões diárias estimam–se os segmentos de "ruidos".

$$\underset{\sim}{a}_1 = \{a_{1,K+1}, \ldots, a_{1,n}\}$$

.

.

.

$$\underset{\sim}{a}_i = \{a_{i,K+1}, \ldots, a_{i,n}\}$$

.

.

.

$$\underset{\sim}{a}_m = \{a_{m,K+1}, \ldots, a_{m,n}\} \qquad (9)$$

Aplica–se então, o Bootstrap à série de m(n–K) ruídos $(a_{i,j-K})$.

Em cada reamostragem é obtido uma "pseudo" série de m(n–K) ruídos, que são separados em m segmentos, os quais juntamente com os parâmetros estimados darão origem a *m* "pseudo" segmentos de n vazões diárias:

$$x_1^b = \{x_{1,1}^b, \ldots, x_{1,n}^b\}$$

.
.
.

$$x_m^b = \{x_{m,1}^b, \ldots, x_{m,n}^b\} \tag{10}$$

De cada segmento $\{x_{i,1}^b, \ldots, x_{i,n}^b\}$ é obtida a vazão máxima y_i.

$$y_1^b = \max_{j=1,n} \{x_{1,j}^b\}$$

.
.
.

$$y_m^b = \max_{j=1,n} \{x_{m,j}^b\} \tag{11}$$

Obtém-se assim B "pseudo" séries de máximos anuais $y^b = \{y_1^b, \ldots, y_m^b\}$.

Como no caso dos máximos anuais, a cada "pseudo" série de máximos anuais ajusta-se uma distribuição de probabilidades a partir da qual a série de máximos foi amostrada. Por extrapolação desta distribuição obtem-se uma estimativa da vazão de T anos de recorrência $\hat{y}_T$.

No caso particular em que as vazões diárias são independentes (L = K = 0) esta metodologia se torna mais simples, pois o Bootstrap pode ser aplicado diretamente às vazões diárias.

Nesta metodologia, ao contrário da anterior novos valores de máximos anuais podem aparecer. No caso particular de independência os "pseudos " valores de y pertencem ao conjunto dos x's (registro de vazões diárias). No caso geral, as "pseudos" observações de y (máximos anuais) podem não pertencer ao registro de vazões diárias. Este é um dos fatores que nos levam a cogitar a hipótese de que a consideração das vazões diárias pode conter maior quantidade de informações.

ESTUDOS REALIZADOS

Os estudos realizados para comparação das metodologias, utilizaram um "cenário" artificial de 100 séries, cada uma constituída de 10 (m = 10) segmentos de 30 (n = 30) vazões diárias.

Para a obtenção desta séries adotou-se um modelo estocástico, autoregressivo de ordem um (AR(1)) para decrever o comportamento das vazões diárias. É importante ressaltar que não podemos esperar que este simples processo estocástico realmente represente as vazões diárias, mas ele é útil para lançar luz no tópico específico deste estudo.

Assim:

$$x_j = \rho\, x_{j-1} + \sqrt{1-\rho^2}\; a_j \tag{12}$$

onde:

x_j = vazões diária padronizada ($\mu_x = 0$, $\sigma_x = 1$) no tempo j.
ρ = coeficiente de correlação Lag 1, isto é, correlação entre x_j e x_{j+1}
a_j = ruido normal (0, 1)

Os coeficientes de correlação adotados neste estudo foram:

ρ = 0,95 vazões altamente correlacionadas
ρ = 0,50 vazões medianamente correlacionadas
ρ = 0,00 vazões independentes

No caso de ρ = 0, as vazões diárias passam a ser variáveis aleatórias independentes, identicamente distribuídas com distribuição N(0,1), e a equação (12) reduz-se a: $x_j = a_j$.

A distribuição dos máximos anuais do processo estocástico (12) pode ser obtida, quando ρ = 0 elevando-se à potência 30 a distribuição acumulada normal-padrão ($F_y(y) = \Phi^{30}(y)$). Nos casos em que $\rho \neq 0$ usou-se como aproximação a distribuição "GUMBEL", uma vez que, um dos principais resultados da "Teoria dos Extremos" é que se as variáveis aleatórias x_i são independentes, com distribuição tipo exponencial, então o máximo definido como $y = \max\{x_1, \ldots, x_n\}$ *tem distribuição de probabilidades, tal que.* (Gumbel, 1958),

$$\lim_{n \to \infty} F(x) = \exp[-\exp(-\psi(x-\mu))] \tag{13}$$

Para n = 30 esta conclusão não é correta, e a extrapolação para valores cujo tempo de recorrência é grande conduz a erros. Kelman, 1985, verificou que neste caso há uma superestimação do valor de y_T, quando a distribuição dos x_j é normal, e uma subestimação quando a distribuição é lognormal. Além disto, no caso de $\rho \neq 0$, as vazões diárias não são independentes. No entanto a distribuição GUMBEL é largamente utilizada, e por isso foi a adotada neste estudo.

Assim, na aplicação das metodologias, a cada pseudo-amostra (7) ou (11) ajustou-se a distribuição GUMBEL, cujos parâmetros são: (Benjamim, 1970)

$$\alpha = \frac{\Pi}{\sqrt{6}\;\sigma_y} \qquad \mu = m_y - \frac{0{,}577}{\alpha} \tag{14}$$

onde:

m_y é a média dos máximos anuais

σ_y é o desvio padrão dos máximos anuais

A extrapolação da distribuição para obter $\hat{y}_T$ é obtida por:

$$y_T = \mu + (z/\alpha)$$

O quantil em estudo foi o correspondente ao tempo de recorrência de 100 anos, isto é, vazão centenário (y_{100}). Neste caso, z = 4,60. A tabela 1 apresenta os valores de população para y_{100} para os diferentes valores de correlação.

ρ	0.00	0.50	0.95
y_{100}	3.40	3.39*	3.07*

Tabela 1 – "Valores da Vazão Centenária para y_{100}"
(* obtido por simulação)

RESULTADOS

Para a avaliação da eficiência do uso das metodologias para a estimação da vazão centenária foram calculadas tendenciosidades e erros médios quadráticos. Calculou-se ainda os mesmos índices para estimativas clássicas de vazão centenária obtidas, para cada segmento, pelo ajuste da distribuição de Gumbel aos seus 10 máximos. A tabela 2 apresenta estes valores.

ρ	y_{100}	Tendência			Em. Méd. Quad.		
		M_1	M_2	$C\ell$	M_1	M_2	$C\ell$
0,00	3,40	0,05	0,06	0,01	0,20	0,20	0,24
0,50	3,39	0,09	0,16	0,01	0,36	0,21	0,34
0,95	3,07	0,23	0,38	0,08	0,67	0,39	0,66

Tabela 2– "Resultados de aplicação dos Métodos de Estimação de y_{100}
M_1 : Aplicação de Bootstrap aos máximos anuais.
M_2 : Aplicação de Bootstrap às vazões diárias
$C\ell$: Estimativa clássica".

A análise dos resultados apresentados na tabela 2 indica que:

1– A utilização de vazões diárias no cálculo de y_{100} resultou em estimativas cujos erro médio quadráticos foram sensivelmente menores que aqueles obtidos quando considerou-se apenas os máximos anuais principalmente nos casos em que a correlação entre as vazões diárias era maior.

2– Quanto a tendenciosidade e a variância das estimtivs e y_{100}, comparando as duas modelagens podemos dizer que a adoção de vazões diárias, conduz a estimativas de maior tendenciosidade porém de menor variância que a utilização apenas dos máximos anuais.

CONCLUSÕES

A aplicação do Bootstrap aos máximos anuais e as vazões diárias, como forma de analisar o problema da quantidade de informação a considerar na modelagem de picos de cheias, indicou que a utlização de vazões diárias conduz a resultados levemente superiores. Recomenda-se que estudo similar seja realizado com séries reais para verificar esta tendencia.

O problema de aplicar o Bootstrap quando as variáveis aleatórias não são independentes (vazões diárias) e são modeladas por processos estocásticos foi contornado aplicando-se o Bootstrap aos "ruídos" independentes do processo.

REFERÊNCIAS

BENJAMIN, J.R., CORNELL, G.A., Probability Statistics and Decision for Civil Engineers, Mc Grall-Hill Book Company, USA, 1970.

COSTA, F.S., Aplicação de Técnicas Estatísticas de Reamostragem em Hidrologia, Tese de Mestrado, COPPE/UFRJ, Rio de Janeiro, Brasil, 1988.

EFRON, B., The Jackknife, The Bootstrap and others Resampling Plans, Society for Industrial and Applied Mathematicsd, Philadelphia, Pensylvania, USA, 1982.

GUMBEL, E.J., "Statistics of Extremes", Columbia University Press, USA, 1958.

KELMAN, J., "Statistical Approach to Flood", Simposia on Statistics in Honours of Professor V.W. Joshi's 70th Birthday, University of Western Ontario, Canadá, 1985.

UNNY, T.E., COVER, R.A., "Application of Computer Intensive Statistics to Parameter Uncertainty in Streamflow Synthesis", Simposia on Statistics in Honours of Professor V.W.Joshi's 70th Berthday, University of Westerm. Ontário, Canadá, 1985.