

ERROS ENVOLVIDOS NA ESTIMATIVA DA VAZÃO MÁXIMA UTILIZANDO CURVA-CHAVE. CASO DE ESTUDO: BACIA DO RIO IBICUI, RS

Luis Carlos Brusa e Robin Thomas Clarke

*Instituto de Pesquisas Hidráulicas – Universidade Federal do Rio Grande do Sul
Caixa Postal 15029 – CEP 91501-970 Porto Alegre, RS
brusa@if1.if.ufrgs.br*

RESUMO

A utilização de métodos diretos para determinar a vazão numa seção fluvial é uma tarefa demorada e, freqüentemente, envolve elevados custos. Por isso, geralmente, procede-se a estimativa da descarga de forma indireta mediante o uso da curva-chave, a qual na maioria das vezes é representada na forma exponencial $Q = C(h+a)^b$. Esta equação, freqüentemente, apresenta bons resultados para interpolar valores entre as cotas máximas e mínimas para a qual foi definida. Porém, quando é necessário extrapolar a curva-chave para cotas maiores (menores) aos máximos (mínimos) observados, é freqüente o surgimento de erros na estimativa.

Este trabalho foi desenvolvido utilizando registros de sete postos fluviométricos localizados na bacia hidrográfica do rio Ibicuí-RS. Os objetivos foram: i) quantificação das incertezas em vazões máximas anuais, obtidas pelo uso da curva-chave e; ii) avaliação de como essas incertezas aumentam ao incluir na análise a incerteza no parâmetro a da curva-chave exponencial. Os resultados obtidos mostram que as incertezas, expressas em termos de um intervalo de confiança de 95% para as vazões estimadas, não são pequenas, e as incertezas são maiores ainda quando é levada em consideração a incerteza na estimativa do parâmetro a.

INTRODUÇÃO

A medição direta de vazão é um procedimento caro e demorado. Geralmente, a estimativa é realizada de uma forma indireta através do uso da curva cota-vazão ou, como é mais conhecida, a curva-chave. Essa permite estimar a descarga numa determinada seção fluvial realizando uma simples leitura de nível na régua linimétrica. Segundo Lambie (1978) e Mosley e Mc Kerchar (1993), na maioria das vezes a curva-chave pode ser determinada mediante uma relação exponenci-

al da forma $Q = C(h+a)^b$ onde Q é a vazão, h é o nível da régua correspondente à vazão, C e b são coeficientes próprios de uma estação e a é a diferença em altura entre o zero da régua e o nível que representa a descarga nula; freqüentemente, duas ou mais curvas podem ser necessárias para representar a relação h-Q, cada uma referente a uma faixa particular de cotas. Na prática, os coeficientes C e b são calculados mediante regressão linear entre $\log_e Q$ e $\log_e(h+a)$, o coeficiente a é aquele que forneça a melhor regressão linear representado através do coeficiente de determinação r^2 . Nos casos em que a representação gráfica indique que são necessárias várias curvas para definir corretamente a relação h-Q para as diferentes faixas de cotas, a adaptação do método é direta e simples, levando em consideração que a cota onde ocorre a interseção das curvas representa o ponto de mudança de controle, podendo ser utilizados os procedimentos apresentados por Williams (1959).

O valor obtido de r^2 , coeficiente de determinação que indica a percentagem da variância explicada pelo ajuste, pode ser muito elevado. Lambie (1978), destaca que a curva-chave é definida, inicialmente, para interpolação, e quando é requerida a extrapolação da curva para cotas maiores (menores) que as cotas máximas (mínimas) registradas, a estimativa pode estar sujeita a erros. Porém, na prática a extrapolação da curva-chave é inevitável. Portanto, é importante conhecer quais são as incertezas que resultam do uso da curva-chave. O primeiro objetivo deste trabalho é quantificar essas incertezas.

Quando a curva-chave é definida por mínimos quadrados mediante o ajuste da expressão $\log_e Q = \log_e C + b \cdot \log_e(h+a) + \varepsilon$ (sendo o parâmetro a conhecido, obtido por um método de tentativa e erro) é um procedimento simples e direto calcular a precisão de estimativa de $\log_e Q$ para um determinado valor de h, bem como o intervalo de confiança $100(1-\alpha)\%$ para Q mediante o processo de transformação inversa de $\log_e Q$ para o valor original de Q. Entretanto, este procedimento não leva em consideração a incerteza na estimativa de a. O segun-

do objetivo deste trabalho é a avaliação do aumento na incerteza na estimação de uma vazão máxima anual, que resulta, por não levar em conta a incerteza, na estimação do parâmetro a da curva-chave.

DADOS DISPONÍVEIS

Para o desenvolvimento deste trabalho, foram utilizados dados de sete estações fluviométricas localizadas na bacia hidrográfica do rio Ibicuí. O rio Ibicuí é tributário do rio Uruguai e encontra-se totalmente inserido dentro do Estado do Rio Grande do Sul. Para cada estação, existem pares de valores cota-descarga obtidos mediante medidas de vazão em campo. Devido à existência de numerosos erros nos dados originais, foi necessário realizar uma filtragem dos mesmos para proceder a sua eliminação. Alguns desses erros foram facilmente detectáveis mediante representação gráfica, outros por uma análise numérica dos dados. Finalizada esta primeira filtragem, procedeu-se a uma segunda filtragem, com a finalidade de eliminar observações que foram realizadas num mesmo dia ou muito próximos. Na Tabela 1 são identificadas as estações utilizadas neste estudo, o número de pares (h, Q) disponível em cada estação, e o valor de r^2 obtido no ajuste por mínimos quadrados da relação $\log_e Q$ e $\log_e(h+a)$.

Com a exceção do rio Toropí em Cachoeira Cinco Veados, todos os valores de r^2 mostrados na Tabela 1, excedem 90% e sugerem que as curvas-chave foram bem determinadas; e mesmo esta estação excepcional mostra um valor de r^2 de 87,7%. Porém, estes valores altos de r^2 dão uma falsa impressão da precisão com que as descargas foram estimadas. Para justificar isso, podemos considerar que a variável de interesse é a vazão máxima anual; para cada ano de registro, em cada uma das sete estações listadas na Tabela 1, é

retirado um único valor de descarga, talvez com a finalidade de ajustar uma distribuição de extremos a partir da qual, a máxima descarga anual com período de retorno de T -anos será estimada. Para as sete estações apresentadas na Tabela 1, a Tabela 2 mostra: i) o número de anos de registro, N , e; ii) o número de anos, G , no qual o nível de água máximo anual h excedeu h_{\max} , o máximo nível de água medido.

Da Tabela 2 pode-se observar que, em termos médios, em dois de cada três anos de registro, a cota máxima h_t do ano excede a máxima entre todas as cotas utilizadas para definir a curva-chave. Assim, se queremos determinar a descarga máxima anual Q_i associada à cota máxima anual h_t em cada um destes anos, devemos proceder à extrapolação da curva-chave para realizar o cálculo. Portanto, nesses anos de extrapolação, as incertezas nas descargas máximas anuais Q_i serão maiores do que a máxima incerteza medida pelo intervalo de confiança da descarga estimada Q_{\max} . Mosley e McKerchar (1993) descrevem alguns métodos de extrapolação baseados nas curvas cota-velocidade e cota-área, mas não se encontra na literatura hidrológica indicações das incertezas implícitas nestes métodos.

Intervalo de confiança no nível 95% para Q_{\max}

Consideramos inicialmente o problema de determinação do intervalo de confiança (95%) para a estimativa do Q_{\max} , a descarga máxima que corresponde à cota h_{\max} , a máxima do conjunto $\{h_t\}$ dos pares $\{h_t, Q_t\}$ utilizados para definir a curva-chave. Supõe-se que o coeficiente a na expressão $\log_e Q = \log_e C + b \cdot \log_e(h+a) + \varepsilon$ é conhecido através de um cálculo iterativo: assim, o “modelo” da curva-chave é, essencialmente, uma regressão linear

Tabela 1. Estações fluviométricas na bacia hidrográfica do rio Ibicuí, RS.

Posto	Rio	Nome da estação	Área de drenagem (km ²)	Período de observação	r ²	
					(1)	(2)
76500000	Ibicuí	Jacaqua	27771	1942/1996	159	96,1
76560000	Ibicuí	Manoel Vianna	29321	1967/1995	185	95,7
76600000	Ibicuí	Passo do Itaum	31008	1967/1983	123	99,5
76800000	Ibicuí	Passo Mariano Pinto	42498	1967/1996	146	94,7
76085000	Toropí	Cachoeira 5 Veados	1635	1976/1996	51	87,7
76120000	Toropí	Ponte rio Toropí	3310	1955/1983	287	92,5
76470000	Jaguarizinho	Passo de Jaguarizinho	1345	1955/1971	194	94,4

(1) Número de pares de valores que definem a curva-chave; (2) Valor de r^2 obtido considerando que o parâmetro a é conhecido em $Q = C \cdot (h+a)^b$.

Tabela 2. Número de anos de registro (N) e número de anos G no qual o nível de água máximo anual supera h_{max} .

Posto	N	G	% G/N
76500000	17	15	88
76560000	23	3	13
76600000	15	3	20
76800000	36	14	39
76085000	17	17	100
76120000	19	14	74
76470000	14	14	100
Média	20	11	62

com dois parâmetros, $\log_e C$ e b . As suposições são: i) os resíduos são normalmente distribuídos; ii) os resíduos possuem variância constante e; iii) os resíduos são estatisticamente independentes. Estas suposições devem ser verificadas; os resultados da verificação não serão apresentadas neste trabalho porque os métodos são apresentados em muitos textos de estatística aplicada.

Neste "modelo" de dois parâmetros devem ser determinados $\log_e C$ e b , sendo que o coeficiente a é constante e estimado previamente por procedimentos de tentativa e erro. Estes parâmetros podem ser estimados utilizando o método de mínimos quadrados mediante a minimização da função $F2$ com relação a $\log_e C$, e b .

$$F2 = \sum [\log_e Q - \log_e C - b \cdot \log_e (h + a)]^2 \quad (1)$$

O erro padrão da estimativa de $\log_e Q$ correspondente a h_{max} é $SE2$, a raiz quadrada de $V2$ na seguinte expressão:

$$V2 = s^2 + \left[\frac{\partial F}{\partial (\log_e C)} \right]^2 \cdot \text{var}(\log_e C) + \left[\frac{\partial F}{\partial (b)} \right]^2 \cdot \text{var}(b) + 2 \cdot \left[\frac{\partial F}{\partial (\log_e C)} \cdot \frac{\partial F}{\partial (b)} \right] \cdot \text{cov}(\log_e C; b) \quad (2)$$

sendo s^2 o valor mínimo da função $F2$ dividido por $(N-2)$, e N o número de pares de pontos (h_i, Q_i) , $i = 1 \dots N$ usados para definir a curva-chave.

A Expressão (2) é equivalente à seguinte equação:

$$V2 = s^2 \cdot \left[1 + \frac{1}{N} + \frac{(x_o - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right]$$

para a variância da estimativa da variável dependente y , dado o valor da variável independente x_o , numa regressão linear da forma $E[y] = \beta_0 + \beta_1 \cdot x$ com $x = \log_e (h+a)$. As duas expressões têm o mesmo valor numérico.

Aproximadamente, o intervalo de confiança 95% para a estimativa de $\log_e Q_{max}$ é $\{\log_e Q_{max} \pm 2 \cdot SE2\}$ e para Q_{max} $\{\exp(\log_e Q_{max} \pm 2 \cdot SE2)\}$, respectivamente. Porém, estes limites não consideram as incertezas na estimativa do parâmetro a . A matriz de variância-covariância das estimativas destes parâmetros é dada, aproximadamente, pelo negativo da matriz inversa das derivadas parciais segundas $(\partial^2 F / (\partial \theta_i \partial \theta_j))$ ($\theta_i = \log_e C$, b) no ponto onde a função $F2$ possui um mínimo global.

Os valores obtidos de $\log_e C$, b e a (este último determinado por procedimento de tentativa e erro) na utilização de um modelo com dois parâmetros são apresentados na Tabela 3.

Tabela 3. Modelo de dois parâmetros - valores dos parâmetros $\log_e C$ e b , com seus erros padrão (s.e.).

Posto	a	$\log_e C$	s.e.	b	s.e.
76500000	-32	-6,1340	0,1820	2,1203	0,0331
76560000	34	-8,6140	0,1750	2,4259	0,0293
76600000	-12	-4,3847	0,0810	1,8219	0,0143
76800000	70	-5,1930	0,1160	1,9858	0,0205
76085000	34	-12,9260	0,8220	3,2790	0,1750
76120000	17	-12,1340	0,4240	2,7711	0,0745
76470000	-70	-6,6570	0,2230	2,0393	0,0486

Quando é utilizado um modelo de três parâmetros, os coeficientes $\log_e C$, b e a podem ser estimados conjuntamente mediante o uso de mínimos quadrados não-lineares, minimizando a seguinte expressão:

$$F3 = \sum [\log_e Q - \log_e C - b \cdot \log_e (h + a)]^2 \quad (3)$$

com relação a $\log_e C$, b e a . A matriz variância-covariância tem o mesmo significado que o definido no modelo de dois parâmetros. Neste modelo de três parâmetros, o erro padrão da estimativa de $\log_e Q_{max}$ é $SE3$, a raiz quadrada de $V3$ na seguinte expressão:

$$V3 = s^2 + \left[\frac{\partial F}{\partial (\log_e C)} \right]^2 \cdot \text{var}(\log_e C) + \left[\frac{\partial F}{\partial (b)} \right]^2 \cdot \text{var}(b) + \left[\frac{\partial F}{\partial (a)} \right]^2 \cdot \text{var}(a) +$$

$$\begin{aligned}
 & 2 \cdot \left[\frac{\partial F}{\partial(\log_e C)} \cdot \frac{\partial F}{\partial(b)} \right] \cdot \text{cov}(\log_e C; b) + \\
 & 2 \cdot \left[\frac{\partial F}{\partial(\log_e C)} \cdot \frac{\partial F}{\partial(a)} \right] \cdot \text{cov}(\log_e C; a) + \\
 & 2 \cdot \left[\frac{\partial F}{\partial(b)} \cdot \frac{\partial F}{\partial(a)} \right] \cdot \text{cov}(b; a) \quad (4)
 \end{aligned}$$

Nesta expressão, s^2 é o valor mínimo da função F3 dividido por $(N-3)$, sendo N o número de pares de pontos (h_i, Q_i) , $i = 1 \dots N$. Aproximadamente, o intervalo de confiança ao nível 95% da estimativa de $\log_e Q_{\max}$ é $\{\log_e Q_{\max} \pm 2 \cdot SE3\}$ e para Q_{\max} , $\{\exp(\log_e Q_{\max} \pm 2 \cdot SE3)\}$.

Os valores obtidos de $\log_e C$, b e a, utilizando o modelo de três parâmetros, são mostrados na Tabela 4.

Tabela 4. Modelo de três parâmetros – valores dos coeficientes $\log_e C$, b e a, com seus erros padrão (s.e.).

Posto	$\log_e C$	s.e.	b	s.e.	a	s.e.
76500000	-6,11	3,08	2,117	0,486	-32,30	42,8
76560000	-8,71	4,25	2,439	0,618	35,90	83,1
76600000	-4,36	3,59	1,818	0,552	-12,50	67,2
76800000	-5,20	4,63	1,986	0,702	70,10	88,5
76085000	-12,96	7,03	3,290	1,250	34,20	39,2
76120000	-12,11	1,40	2,767	0,235	16,78	8,2
76470000	-6,66	3,80	2,040	0,690	-70,00	27,7

A Tabela 4 mostra as magnitudes dos erros padrão das estimativas do parâmetro a; nota-se também que a incerteza nessas estimativas é incluída na Expressão (4), mas não incluída na Equação (2).

Limites aproximados dos intervalos de confiança 95%, obtidos mediante a utilização de modelos de dois e três parâmetros

Os valores apresentados na Tabela 5 mostram os limites de confiança (aproximados) inferior (L.I.) e superior (L.S.) num nível de 95% na estimativa de Q_{\max} ; e o valor de Q_{\max} estimado (\hat{Q}) para a cota máxima h_{\max} , respectivamente.

Da análise da Tabela 5 podem ser extraídas as seguintes observações: i) a amplitude do intervalo de confiança 95% não é pequena, mesmo para o modelo de dois parâmetros; ii) o intervalo de confiança 95% fica muito maior quando é incluída a

incerteza na estimativa do parâmetro a na determinação da curva-chave. Deve ser lembrado que a descarga estimada (\hat{Q}) apresentada na Tabela 5, não requer a extrapolação da curva-chave, sendo esta a descarga que corresponde a cota máxima h_{\max} no conjunto $\{h_i, Q_i\}$ $i = 1, 2, \dots, N$ dos pontos que determinam a curva-chave. Nos casos em que a extrapolação é a única forma de obter uma estimativa da vazão (e a Tabela 2 mostra a frequência com que essa extrapolação pode ser requerida), a incerteza na estimativa das descargas ainda será maior.

Tabela 5. Limites aproximados do intervalo de confiança 95% para modelos de curva-chave de dois e três parâmetros.

H_{\max} (cm)	Curva-chave (2 par.) (a conhecido; $\log_e C$, b estimados)		\hat{Q}	Curva-chave (3 par.) ($\log_e C$, a, b estimados)	
	L.I.	L.S.		L.I.	L.S.
660	1724	1998	1856	1136	3042
1054	3950	4513	4222	2359	7556
1029	3608	3911	3757	1684	8325
833	3908	4328	4113	1961	8532
260	210	434	302	96	1015
758	413	721	546	356	829
385	141	182	160	72	353

Obs.: L.I. – limite inferior; L.S. – limite superior. Unidade de vazão em m^3/s .

A incerteza no valor estimado de Q_{\max} , mesmo que os valores de r^2 mostrados na Tabela 1 sejam altos, pode ser explicada da seguinte forma: considere-se o modelo mais simples de dois parâmetros $\log_e C$ e b (o parâmetro a é suposto sem erro); o problema é simplesmente o ajuste de uma regressão linear nessas condições. É bem conhecido que a amplitude do intervalo de confiança é menor ao ponto médio da variável independente ($\log_e(h+a)$ neste contexto) e que a amplitude é maior, quando for maior a distância deste ponto. No caso desses sete postos, o valor de $\log_e(h_{\max}+a)$ está longe da média dos N valores de $\log_e(h+a)$, portanto a amplitude do intervalo 95% é grande. A amplitude diminuiria se houvessem incluídos na curva-chave, pontos adicionais $\{h_i, Q_i\}$ que aumentassem a média dos $\log_e(h+a)$. Mas isso necessitaria mais trabalho no campo.

DISCUSSÃO

Embora as sete estações usadas na análise não terem sido escolhidas de uma forma estritamente aleatória a partir de um conjunto de estações fluviométricas, provavelmente as caracte-

rísticas das incertezas mostradas na Tabela 5 também sejam encontradas em outras estações localizadas dentro da região. É impossível dizer se as incertezas também serão encontradas em outras regiões, mas é possível supor que elas sejam semelhantes naquelas regiões que possuam condições de acesso, clima e limitações de infraestrutura, similares às existentes na região Sul do Brasil.

As incertezas nos valores de Q_{\max} mostrados na Tabela 5 têm implicações importantes sobre outros métodos hidrológicos que utilizam dados de descargas, por exemplo: a estimativa da cheia com período de retorno T-anos mediante o ajuste de uma distribuição de probabilidades a uma série de vazões máximas anuais. Muito tem sido escrito acerca das características de diferentes distribuições geralmente usadas para estes fins, onde é assumido que a descarga máxima anual está livre de erros. Outra hipótese assumida é que os erros são independentes (isto é, que a vazão máxima anual em qualquer ano é estatisticamente independente da descarga máxima anual de outro ano); porém, como todas as descargas máximas anuais obtidas de uma seqüência foram estimadas da mesma curva-chave, a série em sua totalidade é função dos dados usados no ajuste da curva, e portanto, serão correlacionados.

Dos parâmetros C, a e b das sete estações apresentadas nas Tabelas 3 e 4, o de maior interesse é o expoente b. O valor médio das sete estimativas de b foi $2,350 \pm 0,374$. Isto é comparável com o valor de $b = 2,17$ dado por Mosley e McKerchar (1993) para canais de seção parabólica, assumindo canais com largura maior que 20 vezes a profundidade.

Quando existem dois ou mais controles naturais na seção do escoamento, surge o inconveniente da existência de distintos valores de C, a e b para cada uma das faixas. Considerando que a cota h_0 (cota onde acontece a mudança na seção de controle) é conhecida, a curva-chave de dois parâmetros (a conhecido) pode ser definida conjuntamente nas duas seções (h_i, Q_i), $h_i < h_0$ e (h_i, Q_i), $h_i \geq h_0$ por mínimos quadrados usando o método definido por Tocher (1952) e Williams (1959). Em caso de se utilizar uma curva-chave de três parâmetros, a soma combinada dos desvios quadrados pode ser minimizada, sujeito à seguinte restrição:

$$\log_e C_1 + b_1 \cdot \log_e (h_0 + a_1) = \log_e C_2 + b_2 \cdot \log_e (h_0 + a_2) \quad (5)$$

onde os subfixos 1 e 2 indicam as duas faixas de cotas.

REFERÊNCIAS

- LAMBIE, J. C. (1978). Measurement of flow-velocity-area methods. Chap. 1 of *Hydrometry: Principles and Practices* (Ed. R. W. Herschy). John Wiley & Sons. Chichester. England.
- MOSLEY, M. P., MCKERCHAR, A. I. (1993). Streamflow: Chap. 8. *Handbook of Hydrology* (Ed. in Chief: David R Maidment). McGraw-Hill Inc: New York.
- TOCHER, K. D. (1952). *On the concurrence of a set of regression lines*. Biometrika: (39), p109-117.
- WILLIAMS, E. J. (1959). *Regression Analysis*. John Wiley & Sons Ltd: London.

Errors Involved in Estimating Maximum Flows by Means of Rating Curves. A Case Study from the Basin of the River Ibicuí, Southern Brazil

ABSTRACT

Direct methods for measuring flow in a river section are time-consuming and often costly. For these reasons, river flow is generally estimated indirectly by means of a stage-discharge curve, frequently of exponential form $Q = C \cdot (h+a)^b$. Such equations commonly give good estimates of flow when used to interpolate at a stage lying between the maximum and minimum stages used to derive them, but when extrapolation to stages lying outside this range is required, large errors of estimation may arise.

The work described in this paper used records from seven gauging stations in the basin of the River Ibicuí, in the southern Brazilian state of Rio Grande do Sul. The objectives were: i) to quantify the uncertainties in maximum annual discharges that arise from uncertainties in rating-curve definition; ii) to explore how this uncertainty is increased when allowance is made for uncertainty in estimating the parameter a of the exponential rating-curve. Results show that the uncertainty, expressed as 95% confidence intervals for the estimated flows, was considerable, and was even greater when uncertainty in the estimate of the parameter a is allowed for.