

UM ESTIMADOR DE QUANTIS DE CHEIAS BASEADO EM ASSIMETRIA REGIONAL

Cláudio Marchand Krüger
Heinz Dieter Fill

Centro de Hidráulica e Hidrologia Prof. Pangot de Souza - CEHPAR
Caixa Postal 1309 - CEP 80001-970 - Curitiba - PR
kruger@cch.copel.br - fill@cch.copel.br

RESUMO

Apresenta-se um novo estimador de quantis de cheias, o qual utiliza a distribuição de probabilidades generalizada de extremos (GEV), e o método dos momentos com pesos probabilísticos (PWM). Os parâmetros de locação e escala são estimados com dados locais e uma regressão regional com características fisiográficas é usada para estimar o parâmetro de forma. O desempenho deste estimador é comparado com o de outros estimadores disponíveis na literatura, através de simulações Monte Carlo. As comparações mostram a superioridade do estimador proposto, no caso de períodos de observação moderados e erros de modelo da regressão regional de moderados a grandes. Apresenta-se também um estimador bayesiano empírico, o qual combina os resultados do estimador proposto com um estimador de regressão de quantis normalizados que usa 2 parâmetros regionais e 1 parâmetro local. Este estimador bayesiano mostrou-se preciso e robusto sobre uma ampla faixa de períodos de observação, erros de modelo da regressão regional e valores do coeficiente de variação linear (L-Cv) local.

DESCRIÇÃO DO PROBLEMA

Em engenharia de recursos hídricos é de grande importância e responsabilidade a tarefa de estimar o valor da vazão associado a uma dada probabilidade de excedência em um certo local de um rio. Este problema tipicamente envolve a estimativa de uma função que relacione valores extremos de vazão com a sua probabilidade de

excedência, a chamada curva ou distribuição de frequência de vazões máximas. Um dos fatores que mais dificulta a solução deste problema é o fato de que normalmente os períodos de observações de vazão disponíveis são muito menores que os períodos de retorno desejados, conduzindo a estimativas com pouca precisão, se não forem utilizados métodos que compensem a insuficiência de dados observados locais. O uso de métodos regionais para a análise de frequência de cheias é uma das maneiras de compensar a insuficiente caracterização temporal, utilizando informações de outros locais de observação que tenham alguma similaridade com o local de interesse.

Um método que utiliza informações regionais para melhorar a precisão de estimativas de cheias é conhecida como o método "Index Flood" (Dalrymple, 1960), o qual utiliza as séries de vazões máximas observadas de vários locais para estimar, a menos de um fator de escala, uma única distribuição de frequências regional. O Natural Environment Research Council (1975) propôs uma abordagem sistemática para estudos de cheias no Reino Unido, no qual é recomendado o uso do método "Index Flood" combinado com a distribuição generalizada de extremos ("GEV"), para análise regional de frequência de cheias. Mais recentemente, Wallis (1981), baseado nas pesquisas de Greenwood et al. (1979) e Landwehr et al. (1979, 1980), reintroduz o método Index Flood usando uma distribuição Wakeby e momentos com pesos probabilísticos (no original, "probabilistic weighted moments", ou PWMs) para estimativa de parâmetros. Hosking et al. (1985) derivaram expressões para estimativa de parâmetros da distribuição GEV com o

método dos PWMs, concluindo que o novo método se compara favoravelmente com outros métodos tradicionais de estimativa de parâmetros. A chamada abordagem "GEV/PWM/Index Flood" constitui-se um método de análise regional de frequência de cheias competitivo e recomendado em muitos estudos (Wallis e Wood, 1985; Lettenmaier et al., 1987; Cunnamane, 1988; Potter e Lettenmaier, 1990; Lu, 1991; Stedinger et al., 1993; Fill, 1994). Este método, que utiliza uma distribuição GEV regionalizada para representar a distribuição de probabilidades das vazões máximas anuais normalizadas (i.e., divididas pela média) e estima os seus parâmetros por PWMs, será neste trabalho denotado por IF.

Recentemente, Fill (1994) propôs a chamada "regressão de quantis normalizados" (no original, "Normalized Quantile Regression - NQR") que relaciona os quantis normalizados com características fisiográficas das bacias, através de equações de regressão, conseguindo um desempenho superior ao do método IF para regiões hidrologicamente heterogêneas. No caso de regiões homogêneas, o método NQR se reduz ao IF (Fill, 1994). Os estimadores IF e NQR são recomendados para regiões com poucos anos de observações (10 anos, como ordem de grandeza). Para regiões com séries de extensão moderada (25-50 anos) pode ser recomendável o uso de algum método que estime os parâmetros de locação e escala (média e coeficiente de variação) da distribuição GEV através de dados locais e apenas o parâmetro de forma (assimetria) através de um método regional. Estimadores que utilizam esta abordagem são chamados estimadores 2P.

Estimadores que estimam todos os 3 parâmetros da distribuição GEV a partir de dados locais só são atraentes para registros muito longos, cerca de 100 anos (Lu, 1991; Fill, 1994) e portanto têm pouco interesse na maioria dos casos práticos. Entretanto, como foi demonstrado por Fill (1994), a questão de qual estimador (NQR/IF ou 2P) seja mais preciso não depende apenas da extensão da

série de observações disponível mas também da precisão do modelo de regressão e do verdadeiro coeficiente de variação das vazões máximas normalizadas. Essas duas últimas grandezas são de difícil estimativa na prática, o que pode levar a dificuldades em decidir qual estimador (NQR/IF ou 2P) usar.

Fill (1994) propôs o uso de um método bayesiano para a solução desse dilema, estabelecendo um estimador de quantis de cheias que realiza uma ponderação contínua entre os estimadores NQR/IF e 2P, conseguindo-se assim um desempenho quase sempre superior a ambos. O estimador 2P utilizado nessa ponderação usa, para obter o parâmetro de forma, uma média regional dos estimadores dos coeficientes de assimetria lineares (L-Ca). O objetivo principal deste trabalho é implementar um estimador 2P com regressão regional para o parâmetro de forma da distribuição GEV e compará-lo com o estimador 2P convencional utilizado por Lu (1991) e Fill (1994). Foi ainda construído um estimador bayesiano correspondente usando a mesma técnica de ponderação contínua proposta por Fill (1994).

A DISTRIBUIÇÃO GEV E OS PWMs

A distribuição generalizada de extremos (GEV), foi proposta por Jenkinson (1955) e é a distribuição recomendada para análise de frequência de cheias no Reino Unido (Natural Environment Research Council, 1975). A função de distribuição acumulada para uma distribuição GEV é a seguinte:

$$F_{\kappa}(x) = \exp\left[-\left(1 - \kappa \frac{x - \xi}{\alpha}\right)^{1/\kappa}\right] \quad \text{para } \kappa \neq 0$$

$$F_{\kappa}(x) = \exp\left[-\exp\left(-\frac{x - \xi}{\alpha}\right)\right] \quad \text{para } \kappa = 0$$
(1)

onde ξ é um parâmetro de locação, α é um parâmetro de escala, e κ é o parâmetro de forma. Se $\kappa = 0$ a distribuição GEV corresponde à distribuição Gumbel ou extrema tipo 1.

O p -ésimo quantil de uma distribuição GEV pode ser calculado por:

$$\begin{aligned} x_p &= \xi + \frac{\alpha}{\kappa} \{1 - [-\ln(p)]^\kappa\} & \text{para } \kappa \neq 0 \\ x_p &= \xi - \alpha \ln[-\ln(p)] & \text{para } \kappa = 0 \end{aligned} \quad (2)$$

Os parâmetros de distribuições GEV têm sido estimados por diversos métodos. Segundo Hosking et al. (1985), Lu (1991) e Krüger (1996), o desempenho de estimadores obtidos pelo método dos momentos com pesos probabilísticos (PWMs), para distribuições Gumbel e GEV com três parâmetros se compara favoravelmente em relação a estimadores por momentos ordinários e máxima verossimilhança em amostras pequenas.

Estimadores não-tendenciosos para os PWMs foram derivados por Landwehr et al., (1979). Da amostra ordenada $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ e para $r < n$, define-se h_r como um estimador não-tendencioso do PWM de ordem r , β_r :

$$h_r = \left[n \binom{n-1}{r} \right]^{-1} \sum_{i=r+1}^n \binom{i-1}{r} x_{(i)} \quad (3)$$

Os momentos com pesos probabilísticos são capazes de caracterizar teoricamente uma distribuição, no entanto, os valores dos PWMs não possuem um significado prático para identificar e comparar a locação, escala e forma de distribuições. Para este fim, é mais conveniente utilizar combinações lineares dos PWMs, denominados momentos lineares (ou momentos-L). As relações entre os três primeiros momentos-L λ_r , e os PWMs β_r são (Hosking, 1986):

$$\begin{aligned} \lambda_1 &= \beta_0 \\ \lambda_2 &= 2\beta_1 - \beta_0 \\ \lambda_3 &= 6\beta_2 - 6\beta_1 + \beta_0 \end{aligned} \quad (4)$$

O primeiro momento-L, λ_1 , coincide com a média ou valor esperado da variável aleatória. O segundo momento, λ_2 , é uma medida da escala ou dispersão da distribuição e é definido como a semi-diferença entre os valores esperados do maior e menor elemento de uma amostra de dois elementos. É conveniente padronizar os momentos de ordem mais elevada, de modo a obter valores independentes das unidades de medida da variável aleatória, como os tradicionais coeficiente de variação (Cv) e coeficiente de assimetria (Ca). Definem-se então as chamadas *razões de momentos-L*, τ_2 (L-Cv) e τ_3 (L-Ca) (Hosking, 1986; Stedinger et al., 1993):

$$\begin{aligned} \tau_2 &= \lambda_2 / \lambda_1 \\ \tau_3 &= \lambda_3 / \lambda_2 \end{aligned} \quad (5)$$

Hosking et al. (1985) derivaram as expressões dos estimadores dos parâmetros da distribuição GEV, $\hat{\xi}$, $\hat{\alpha}$ e $\hat{\kappa}$. Como κ não pode ser escrito como função explícita dos β_r , utiliza-se uma solução aproximada para estimativa deste parâmetro, válida para $-1/2 < \kappa < 1/2$:

$$\hat{\kappa} = 7,8590c + 2,9554c^2 \quad (6)$$

onde

$$c = \frac{2}{\hat{\tau}_3 + 3} - \frac{\ln(2)}{\ln(3)} = \frac{2}{\hat{\tau}_3 + 3} - 0,6309 \quad (7)$$

e

$$\hat{\alpha} = \frac{\hat{\lambda}_2 \hat{\kappa}}{\Gamma(1 + \hat{\kappa})(1 - 2^{-\hat{\kappa}})} \quad (8)$$

$$\hat{\xi} = \hat{\lambda}_1 + \frac{\hat{\alpha}[\Gamma(1 + \hat{\kappa}) - 1]}{\hat{\kappa}} \quad (9)$$

Para $\kappa = 0$, os parâmetros da distribuição Gumbel podem ser estimados por:

$$\hat{\alpha} = \frac{\hat{\lambda}_2}{\ln 2} \quad (10)$$

$$\hat{\xi} = \hat{\lambda}_1 - \gamma \hat{\alpha} \quad (11)$$

onde γ é a constante de Euler ($\gamma = 0,5772\dots$)

O quantil $100p$ % da distribuição GEV pode ser expresso como função dos dois primeiros momentos-L e pelo parâmetro de forma (Fill, 1994):

$$q_p = \lambda_1 + \frac{\lambda_2}{1-2^{-\kappa}} \left[1 - \frac{(-\ln p)^\kappa}{\Gamma(1+\kappa)} \right] \quad \text{se } \kappa \neq 0; \quad (12)$$

$$q_p = \lambda_1 + \frac{\lambda_2}{\ln 2} [-\gamma - \ln(\ln p)] \quad \text{se } \kappa = 0.$$

Dividindo (12) por λ_1 , obtém-se o quantil de vazões normalizadas $x_p = q_p / \mu_q$, descritas por uma distribuição GEV, expresso em função do fator de frequência $K(\kappa, p)$:

$$x_p = 1 + K(\kappa, p) \tau_2 \quad (13)$$

onde:

$$K(\kappa, p) = \frac{1}{1-2^{-\kappa}} \left[1 - \frac{(-\ln p)^\kappa}{\Gamma(1+\kappa)} \right] \quad \text{se } \kappa \neq 0; \quad (14)$$

$$K(\kappa, p) = \frac{1}{\ln 2} [-\gamma - \ln(\ln p)] \quad \text{se } \kappa = 0.$$

ESTIMADORES REGIONAIS DE QUANTIS DE CHEIAS

Estimadores de quantis de cheias têm sido propostos com variados graus de regionalização, sendo que para os propósitos deste trabalho podem ser agrupados em 3 grupos principais conforme o número de parâmetros a serem estimados localmente e regionalmente e segundo o método de regionalização empregado. A Tabela 1 ilustra essa classificação.

Tabela 1. Estimadores regionais de cheias.

Método	Num. de parâm. locais	Num. de parâm. regionais	Regionalização
3P	3	0	-
2P	2	1	média
2P/AR*	2	1	regressão
1P	1	2	média
NQR	1	2	regressão
Regressão regional	0	3	regressão

*Desenvolvido neste trabalho

Sem perda de generalidade, todos os estimadores analisados neste trabalho, a menos dos estimadores por regressão regional, utilizam um procedimento em duas etapas, primeiro estimando um quantil normalizado e depois calculando o valor da vazão multiplicando este quantil por um fator apropriado. Considerando um quantil normalizado adimensional $100p$ % igual a $x_p^i = q_p^i / \lambda_1$ para o local i , um estimador \hat{x}_p^i pode ser obtido por:

$$\hat{x}_p^i = 1 + K(\hat{\kappa}, p) \hat{\tau}_2 \quad (15)$$

Este quantil normalizado é multiplicado pela vazão média $\hat{\lambda}_1 = \bar{q}$ estimada com os dados locais, resultando o quantil de cheia no local i :

$$\hat{q}_p^i = \hat{x}_p^i \hat{\lambda}_1 \quad (16)$$

O Método Index Flood

O método Index Flood (Dalrymple, 1960) foi desenvolvido com a finalidade de diminuir o erro amostral na estimativa de quantis de vazões extremas, explorando a similaridade hidrológica existente entre bacias em uma região considerada homogênea. O método consiste essencialmente em combinar os registros históricos de estações com características similares para produzir uma única distribuição de frequências acumuladas regional adimensional. Dividindo-se as séries de vazões máximas por um valor índice ("index flood"), as curvas que representam a

distribuição de freqüências podem ser superpostas para o estabelecimento de uma única curva média regional.

Estudos mais recentes (Wallis, 1981; Wallis e Wood, 1985), apresentaram uma nova versão do método Index Flood, incorporando o método dos momentos com pesos probabilísticos (PWMs), em conjunto com distribuições GEV ou Wakeby.

Segundo Stedinger et al. (1993), o algoritmo IF para estimativa de quantis de cheias pode ser descrito como a seguir. Supondo que existam K locais em uma região com registros de vazão q_t^k , $t = 1, \dots, n_k$ e $k = 1, \dots, K$.

1. Em cada local k calcular os três estimadores de momentos-L: $\hat{\lambda}_1^k, \hat{\lambda}_2^k, \hat{\lambda}_3^k$, utilizando estimadores PWM não-tendenciosos b_r ($r = 0, 1, 2$).
2. Para obter uma distribuição normalizada de freqüências para a região, calcular a média regional dos momentos-L normalizados de ordem $r = 2$ e 3:

$$\hat{\lambda}_r^R = \frac{\sum_{k=1}^K w_k (\hat{\lambda}_r^k / \hat{\lambda}_1^k)}{\sum_{k=1}^K w_k} \quad \text{para } r = 2, 3 \quad (17)$$

Para $r = 1$, $\hat{\lambda}_1^R = 1$. Em (17), w_k são pesos, uma escolha simples pode ser $w_k = n_k$, onde n_k é o tamanho da amostra para o local k .

3. Utilizando os momentos-L normalizados médios $\hat{\lambda}_1^R, \hat{\lambda}_2^R$ e $\hat{\lambda}_3^R$, determinar os parâmetros e quantis \hat{x}_p^R da distribuição GEV regional normalizada.
4. O estimador do percentil $100p$ % da distribuição de vazões em um local k é:

$$\hat{q}_p(k) = \hat{\lambda}_1^k \hat{x}_p^R \quad (18)$$

onde $\hat{\lambda}_1^k$ é a média amostral calculada com os dados do local k :

$$\hat{\lambda}_1^k = \frac{1}{n_k} \sum_{t=1}^{n_k} q_t^k \quad (19)$$

Regressão regional pelo Método dos Mínimos Quadrados ponderado (WLS)

Seja θ o vetor ($m \times 1$) contendo os verdadeiros valores de uma variável hidrológica (ou o seu logaritmo) observada nos m locais em uma região e seja $\hat{\theta}$ um estimador não-tendencioso de θ . Seja Σ uma matriz diagonal contendo as variâncias amostrais $\text{var}(\hat{\theta}_i)$ de $\hat{\theta}_i$ em relação a θ_i na sua diagonal. O procedimento WLS assume que Σ é uma aproximação razoável da matriz de covariâncias amostrais do vetor estimador $\hat{\theta}$. Este é o caso se for possível negligenciar as correlações entre $\hat{\theta}_i$ e $\hat{\theta}_j$, para $i \neq j$. Outra hipótese básica é a de que θ_i , para um local i escolhido ao acaso seja gerada por uma função linear de um conjunto de características da bacia (como área de drenagem, declividade, entre outros, ou, como é mais comum, os seus logaritmos) mais um erro ε_i , de modo que:

$$\hat{\theta} = \mathbf{A}\beta + \mathbf{e} \quad (20)$$

onde \mathbf{A} é uma matriz ($m \times k$) contendo as ($k - 1$) características das m bacias aumentada com uma coluna contendo elementos iguais a um; β é um vetor ($k \times 1$) com os coeficientes desconhecidos a serem estimados e \mathbf{e} é um vetor ($m \times 1$) contendo erros de modelo mais os erros amostrais, com $E(\mathbf{e}_i) = 0$ e variâncias dadas por:

$$\text{var}(\mathbf{e}_i) = \text{var}(\varepsilon_i) + \text{var}(\hat{\theta}_i) = \sigma_\varepsilon^2 + \text{var}(\hat{\theta}_i) \quad (21)$$

onde $\text{var}(\hat{\theta}_i)$ é a variância amostral de $\hat{\theta}_i$ em relação a sua média θ_i , no local i e σ_ε^2 é a

variância do erro de modelo. Em notação matricial resulta:

$$E(\mathbf{ee}^t) = \mathbf{I} + \Sigma = \Lambda \quad (22)$$

onde \mathbf{I} é a matriz identidade ($m \times m$). Levando em conta as hipóteses apresentadas anteriormente, os estimadores não-tendenciosos de mínima variância para os coeficientes β podem ser determinados minimizando-se a seguinte soma de quadrados (Tasker e Stedinger, 1986):

$$\sum_{i=1}^m \frac{(\hat{\theta}_i - \hat{\beta}_0 - \sum_{j=1}^{k-1} \hat{\beta}_j a_j^i)^2}{\sigma_{\epsilon}^2 + \text{var}(\hat{\theta}_i)} = \min \quad (23)$$

A solução deste problema de otimização resulta, em notação vetorial (Johnson, 1984):

$$\hat{\beta} = (\mathbf{A}^t \mathbf{A}^{-1} \mathbf{A})^{-1} \mathbf{A}^t \mathbf{A}^{-1} \hat{\theta} \quad (24)$$

O estimador $\hat{\beta}$ possui mínima variância entre todos os estimadores lineares não-tendenciosos de β quando Λ , definida na equação (22) é conhecida. Infelizmente, em geral Λ não é conhecida e deve ser estimada através dos dados.

Segundo Stedinger e Tasker (1985), uma estimativa da variância do erro de modelo σ_{ϵ}^2 , é obtida pela solução iterativa de:

$$(\hat{\theta}_p - \mathbf{A}\hat{\beta})^t \mathbf{A}^{-1} (\hat{\theta}_p - \mathbf{A}\hat{\beta}) = m - k \quad (25)$$

A matriz de covariâncias amostrais Σ pode ser estimada por expressões assintóticas ou via simulação Monte Carlo em função da distribuição de probabilidades das vazões máximas.

As expressões (24) e (25) formam um sistema de $k+1$ equações com $k+1$ incógnitas, que são $\hat{\sigma}_{\epsilon}^2$ e os k elementos de $\hat{\beta}$.

Regressão de Quantis Normalizados (NQR)

Fill (1994) combinou os conceitos do método Index Flood e regressão regional para desenvolver o chamado método de regressão de quantis normalizados (NQR). No método NQR, ao invés de se utilizar um quantil de cheias normalizado constante na região, é feita a regressão dos quantis observados com características fisiográficas de cada local.

Considere-se uma região com m locais, cada um com um registro de vazões de cheias $\{q_i(1), q_i(2), \dots, q_i(n_i)\}$, $i = 1, 2, \dots, m$. As vazões em cada local são consideradas independentes e identicamente distribuídas, podendo ou não haver correlação entre as vazões em diferentes locais. Seja μ_i o valor esperado, σ_i a variância, e q_p^i o "verdadeiro" quantil 100p% da variável aleatória $q_i(t)$. Então $x_i(t) = q_i(t) / \mu_i$ são as vazões normalizadas, cujo quantil 100p% é denotado por $x_p^i = q_p^i / \mu_i$. Assumindo-se que um modelo log-linear relaciona os quantis x_p^i com $k-1$ características fisiográficas das bacias, como área de drenagem, declividade, uso do solo, resulta:

$$\ln x_p^i = \beta_0 + \sum_{j=1}^{k-1} \beta_j \ln a_j^i + \epsilon_i \quad i = 1, 2, \dots, m \quad (26)$$

onde β_j são parâmetros, a_j^i são características fisiográficas e ϵ_i são erros independentes, normais e identicamente distribuídos. Denotando por $\theta_p = \ln \mathbf{x}_p$, onde o negrito indica notação vetorial, a equação anterior pode ser escrita como:

$$\theta_p = \mathbf{A}\beta + \epsilon \quad (27)$$

onde θ_p e ϵ são vetores de dimensão m , β é um vetor de parâmetros de dimensão k , e \mathbf{A} é uma matriz $m \times k$, contendo os logaritmos das características fisiográficas e na primeira coluna elementos iguais a um.

No caso do procedimento WLS, a matriz de covariâncias Λ possui os elementos fora da diagonal principal iguais a zero, correspondendo à hipótese de independência dos resíduos da regressão. O vetor de parâmetros β e o erro de modelo $\hat{\sigma}_i^2$ são estimados usando as equações (24) e (25).

Finalmente, se $\ln \hat{x}_p$ é normalmente distribuído, o estimador NQR é log-normalmente distribuído, com média e variância iguais a:

$$E(\hat{x}_p) = \exp\left[\left(\hat{\beta}_0 + \sum_{i=1}^k \hat{\beta}_i \ln a_i\right) + \sigma_p^2 / 2\right] \quad (28)$$

$$\text{var}(\hat{x}_p) = \exp\left[2\left(\hat{\beta}_0 + \sum_{i=1}^k \hat{\beta}_i \ln a_i\right) + \sigma_p^2\right] [\exp(\sigma_p^2) - 1] \quad (29)$$

Usando-se a média como uma estimativa do quantil normalizado, a vazão de cheia estimada no local i é obtida multiplicando-se (28) pela vazão média \bar{q}_i :

$$\hat{q}_p^i = \bar{q}_i E(\hat{x}_p) \quad (30)$$

Estimador 2P

Uma maneira de reduzir a grande variância amostral dos estimadores locais (aqui denotados por estimadores 3P) é usar um parâmetro de forma constante para todos os locais em uma região. Lu (1991) e Lu e Stedinger (1992) mostraram que, para distribuições GEV realistas (probabilidade de vazões negativas menores que 2%), estimadores usando um parâmetro de forma fixo apresentaram desempenho consistentemente melhor que o estimador 3P. Segundo os mesmos autores, o estimador 3P só torna-se atrativo para registros muito longos e valores do parâmetro de forma κ diferentes de zero.

No caso da distribuição GEV, o estimador 2P corresponde ao caso onde um dos parâmetros é estimado independentemente dos dados locais. Isto pode ser feito obtendo-se o coeficiente de

assimetria de outras fontes (mapas, por exemplo) ou como média regional das estimativas locais. Lu (1991) e Fill (1994) utilizaram um coeficiente de forma κ único, estimado com base no valor médio do L - Ca da região. O valor médio de L - Ca utilizado para o cálculo de $\hat{\kappa}^R$, onde o índice R indica "valor regional" é:

$$\hat{\kappa}^R = \frac{\sum_{i=1}^m n_i \hat{\kappa}_i}{\sum_{i=1}^m n_i} \quad (31)$$

O valor de $\hat{\kappa}^R$ é utilizado na equação (15) para calcular o valor do quantil 2P adimensional. Em contraste com o estimador 3P, estimadores 2P podem apresentar grande tendenciosidade, especialmente em locais com valores de L - Ca bem diferentes do valor médio regional (Fill, 1994, Krüger, 1996). Por outro lado, apresentam menores erros médios quadráticos que o estimador 3P (Lu e Stedinger, 1992).

Estimador 2P com Assimetria Regional (2P/AR)

O parâmetro de forma κ de uma distribuição GEV é difícil de estimar confiavelmente com registros curtos ou mesmo de tamanho moderado (Lu, 1991 e Lu e Stedinger, 1992). Estimativas confiáveis do parâmetro de forma podem ser obtidas através de médias regionais dos PWMs (ou equivalentes momentos-L), como foi utilizado no estimador 2P apresentado anteriormente, ou através de regressão regional do parâmetro de forma ou coeficiente de assimetria com parâmetros fisiográficos de bacias. No presente trabalho, o algoritmo sugerido por Tasker e Stedinger (1986) foi utilizado, relacionando, em uma equação de regressão, valores estimados do parâmetro κ com características fisiográficas das bacias, através de expressão similar à do estimador NQR.

$$\kappa = \beta_0 + \sum_{j=1}^{k-1} \beta_j \ln a_j + \varepsilon_i, \quad i = 1, 2, \dots, m \quad (32)$$

onde β_j são parâmetros, a_j são características fisiográficas e ε_i são erros independentes, normais e identicamente distribuídos. O algoritmo WLS (equações (24) e (25)) é empregado para estimar os parâmetros β_j e o erro de modelo $\hat{\sigma}_\varepsilon^2$.

A estimativa regional do coeficiente de forma para o local i , denotada por $\hat{\kappa}_p^i$ é obtida por:

$$\hat{\kappa}_p^i = \hat{\beta}_0 + \sum_{j=1}^{k-1} \hat{\beta}_j \ln a_j \quad (33)$$

Considerando um quantil normalizado adimensional 100p % igual a $x_p^i = q_p^i / \hat{\lambda}_1^i$ para o local i , um estimador \hat{x}_p^i pode ser obtido por:

$$\hat{x}_p^i = 1 + K(\hat{\kappa}_p^i, p) \hat{\sigma}_\varepsilon^i \quad (34)$$

com $K(\kappa, p)$ dado por (14).

O quantil normalizado é então multiplicado pela vazão média $\hat{\lambda}_1^i = \bar{q}_i$, estimada com dados locais, resultando o quantil de cheia no local i :

$$\hat{q}_p^i = \hat{x}_p^i \hat{\lambda}_1^i \quad (35)$$

ESTIMADOR BAYESIANO EMPÍRICO

A inferência bayesiana difere fundamentalmente da abordagem clássica, que considera os parâmetros de uma distribuição como quantidades fixas embora desconhecidas. Admitindo-se uma distribuição paramétrica, os parâmetros desconhecidos são estimados exclusivamente a partir da amostra. Na abordagem bayesiana supõe-se que, além da amostra, existem outras fontes de informação independentes. Essa informação pré-existente sobre os valores dos parâmetros é representada por uma

distribuição de probabilidades, chamada de distribuição *a priori*. Esta distribuição deve ser formulada independentemente do conhecimento da amostra, através de hipóteses baseadas em outros dados empíricos, considerações teóricas ou julgamentos subjetivos. Uma vez formulada a distribuição *a priori* do parâmetro em análise, esta distribuição pode ser atualizada utilizando-se dados amostrais, através do teorema de Bayes, resultando em uma nova distribuição, condicionada ao vetor de valores observados, chamada distribuição posterior (Berger, 1985):

$$p(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)d\theta} \quad (36)$$

onde $\pi(\theta)$ é a distribuição *a priori* do parâmetro θ , $f(x|\theta)$ é a função de verossimilhança, representada pela função densidade de probabilidades conjunta das observações, dado o parâmetro θ e $p(\theta|x)$ é a distribuição posterior do parâmetro θ , dado o vetor de observações x .

O estimador bayesiano empírico (BAY) utilizado no presente estudo é similar ao estimador bayesiano implementado por Fill (1994). O objetivo seria o de, dado um conjunto de locais com suas características fisiográficas e registros de vazões, obter um estimador mais robusto que o estimador NQR ou o estimador 2P.

Utilizando conceitos de estatística clássica apenas, uma expressão para estimativa de quantis por um estimador "ótimo", i.e., de mínima variância (denotado por $\hat{x}_p^{(o)}$), resultado de uma combinação linear dos estimadores NQR e 2P pode ser escrita como:

$$\hat{x}_p^{(o)} = a + b\hat{x}_p^{NQR} + (1-b)\hat{x}_p^{2P} \quad (37)$$

A expressão acima indica que o estimador ótimo é uma ponderação entre os resultados dos estimadores NQR e 2P mais

uma correção para tendenciosidade a resultante dos dois estimadores. Os valores ótimos de a e b foram deduzidos por Fill (1994) e são dados por:

$$a^* = -b \text{ bias}(\hat{x}_p^{NQR}) - (1-b) \text{ bias}(\hat{x}_p^{2P}) \quad (38)$$

$$b^* = \frac{\text{var}(\hat{x}_p^{2P}) - \text{cov}(\hat{x}_p^{NQR}, \hat{x}_p^{2P})}{\text{var}(\hat{x}_p^{NQR}) + \text{var}(\hat{x}_p^{2P}) - 2 \text{cov}(\hat{x}_p^{NQR}, \hat{x}_p^{2P})} \quad (39)$$

onde:

$$\text{bias}(\hat{x}_p) = E(\hat{x}_p - x_p)$$

$$\text{var}(\hat{x}_p) = E[\hat{x}_p - x_p + \text{bias}(\hat{x}_p)]^2$$

$$\text{cov}(\hat{x}_p^{NQR}, \hat{x}_p^{2P}) = E\left\{ [\hat{x}_p^{NQR} - x_p + \text{bias}(\hat{x}_p^{NQR})] [\hat{x}_p^{2P} - x_p + \text{bias}(\hat{x}_p^{2P})] \right\}$$

Dada a dificuldade de estimar a tendenciosidade (bias) e a covariância dos dois estimadores, Fill recomenda o uso de um estimador "sub-ótimo", denominado estimador "bayesiano empírico", que despreza essas grandezas. O resultado é:

$$\hat{x}_p^{BY} = b \hat{x}_p^{NQR} + (1-b) \hat{x}_p^{2P} \quad (40)$$

com:

$$b = \frac{\text{var}(\hat{x}_p^{2P})}{\text{var}(\hat{x}_p^{NQR}) + \text{var}(\hat{x}_p^{2P})} \quad (41)$$

O estimador dado pela expressão (40) foi utilizado no presente trabalho, empregando os resultados obtidos pelos estimadores NQR e 2P/AR, descritos anteriormente.

COMPARAÇÃO DOS ESTIMADORES POR SIMULAÇÃO MONTE CARLO

Modelo de geração de vazões

Neste trabalho, o método Monte Carlo foi utilizado para comparar o desempenho do estimador proposto (2P/AR) e do estimador

bayesiano (BAY) em relação a estimadores previamente estudados: estimador 2P, estimador 3P, estimador pelo método Index Flood (IF) e estimador pelo método da regressão de quantis normalizados (NQR).

A distribuição utilizada para gerar as vazões foi uma distribuição GEV, fazendo-se análises de sensibilidade com a distribuição log-normal a dois parâmetros. Para um dado local i , os verdadeiros parâmetros que definem a distribuição de vazões no local são representados pelo vetor $[\lambda_1, \tau_2, \kappa_1]^t$. Sem perda de generalidade, uma média unitária $\lambda_1 = 1$ foi adotada em todos os locais. Assume-se com frequência que o verdadeiro $L-Cv$ (τ_2) se relaciona com as características fisiográficas do local por um modelo log-linear:

$$\ln \tau_2 = \beta_0 + \sum \beta_i \ln a_i + \varepsilon \quad (42)$$

onde a_i são características fisiográfica e β_i são parâmetros. Neste trabalho, a área de drenagem (A) e declividade do talvegue principal da bacia (S) foram as características fisiográficas utilizadas.

Lu (1991) analisou dados reais coletados em um grande número de locais ao redor do mundo, publicados por Farquharson et al (1987) e concluiu que, para dados reais, o logaritmo do Cv e o parâmetro de forma κ da distribuição GEV são correlacionados. Fill (1994), analisando os mesmos dados, estabeleceu uma equação de regressão semelhante à obtida por Lu, relacionando κ e $L-Cv$ e que foi utilizada neste trabalho. Portanto, os modelos assumidos para a geração do verdadeiro $L-Cv$ e parâmetro de forma no local i são os seguintes:

$$\ln \tau_2 = \alpha_0 + \alpha_1 \ln A_i + \alpha_2 \ln S_i + \varepsilon_i \quad (43)$$

$$\kappa_i = \beta_0 + \beta_1 \ln \tau_2 + \delta_i \quad (44)$$

onde α_i e β_i são parâmetros e ε_i e δ_i são variáveis aleatórias normais, independentes e identicamente distribuídas com média zero e variâncias (erros de modelo) iguais a σ_ε^2 e σ_δ^2 .

respectivamente. As variâncias σ_ε^2 e σ_δ^2 , juntamente com valores dos parâmetros da regressão α_i e β_i , definem as características "ambientais". Estas características, aplicadas em um conjunto fixo de locais, definem um "cenário". Um "local" é definido por valores fixos da área de drenagem, declividade e comprimento do registro de vazões. Dado um cenário, geram-se valores aleatórios do verdadeiro $L-Cv$ e parâmetro de forma κ para cada local na região. Estes verdadeiros $L-Cv$ e κ permitem a geração das séries de vazões locais. Em seguida, com as amostras de vazões de cada local, o método dos momentos-L pode ser aplicado para calcular estimadores locais de τ_2 e $\hat{\kappa}$. Com estes valores e as características fisiográficas de cada bacia (A_i e S_i), estimativas dos quantis normalizados \hat{x}_p são calculadas para os estimadores 3P, 2P, 2P/AR, IF, NQR e BAY. Multiplicando-se os quantis normalizados pela vazão média amostral local \bar{q} , são obtidas as estimativas dos quantis de cheia locais, para cada estimador e local:

$$\hat{q}_p = \hat{x}_p \bar{q}, \quad (45)$$

Finalmente, estas estimativas podem ser comparadas com os verdadeiros valores q_p , calculados com os valores verdadeiros de τ_2 e κ . Para cada local, as tendenciosidades e erros médios quadráticos são calculados. Foram utilizadas 5000 repetições nas simulações.

Cenários analisados

Para obter resultados que fossem representativos de problemas de estimativa de cheias reais, os parâmetros do modelo de geração de vazões foram escolhidos de modo a reproduzir o comportamento de bacias hidrográficas reais. Assim, a seleção dos cenários consiste na escolha de valores realistas de σ_ε^2 e σ_δ^2 e dos coeficientes de regressão α_i e β_i das equações (43) e (44),

baseados na análise de dados fisiográficos e hidrológicos de bacias reais

Para a equação (43), foram coletados os dados de 162 bacias localizadas na região Sul do Brasil, em sua grande maioria situadas no Paraná e Santa Catarina. Para cada bacia, determinou-se uma estimativa do $L-Cv$ das vazões máximas anuais. Foi realizada então a regressão dos logaritmos dos $L-Cv$ estimados contra os logaritmos das áreas de drenagem (A) e declividades médias (S) das bacias.

A Figura 1 mostra a distribuição de áreas e declividades das bacias analisadas. Os locais analisados cobrem uma extensa faixa de variação de áreas de drenagem (39 a 34432 km²) e declividades (0,6 a 60,1 m km⁻¹). A regressão foi realizada pelo método dos mínimos quadrados ponderado (WLS). A Tabela 2 mostra os resultados dessa regressão. Com base no erro de modelo obtido (0,0452), adotou-se um erro de modelo nas simulações entre 0,02 e 0,08. Para a escolha de valores realistas dos coeficientes de regressão β_i e do erro σ_ε^2 do modelo representado pela equação (44) foram utilizados os resultados de Farquharson et al (1987). Logo, o modelo para a geração dos cenários das simulações é o seguinte:

$$\ln \tau_2 = -0,7564 - 0,0818 \ln A_i + 0,0237 \ln S_i + \varepsilon_i$$

$$\text{com } \sigma_\varepsilon^2 = 0,02(0,02)0,08 \quad (46)$$

$$\kappa = -0,668 - 0,425(\ln \tau_2) + \delta$$

$$\text{com } \sigma_\delta^2 = 0,0059 \quad (47)$$

A comparação dos estimadores analisados foi realizada em um total de 16 combinações de comprimentos de registro histórico e erros de modelo na geração do $L-Cv$, para o caso base, representado por 50 locais. Todas as simulações foram realizadas com vazões independentes e distribuídas segundo uma distribuição GEV. A Tabela 3 mostra a nomenclatura utilizada para cada uma das combinações.

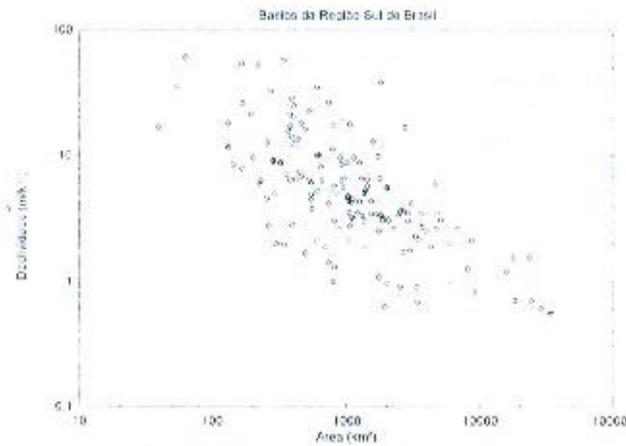


Figura 1. Bacias da região sul do Brasil. Áreas e declividades.

Tabela 2. Resultado da regressão de L-Cv para bacias do sul do Brasil.

Erro modelo σ^2_{ϵ}	Coeficientes α_j			prob ($\alpha_j = 0$)		
	Interc.	1n(A)	1n(S)	Interc.	1n(A)	1n(S)
0,0452	-0,7564	-0,0818	0,0237	$<10^{-5}$	$<10^{-5}$	10^{-5}

Tabela 3. Nomenclatura para identificação dos cenários das simulações Monte Carlo.

número de locais	tamanho do registro	Erro de modelo para o L-Cv			
		0,02	0,04	0,06	0,08
50	10	1C2	1C4	1C6	1C8
50	25	2C2	2C4	2C6	2C8
50	50	3C2	3C4	3C6	3C8
50	100	4C2	4C4	4C6	4C8

Análises de sensibilidade realizadas com regiões compostas por 10, 25 e 100 locais, cujos resultados não são apresentados aqui, conduziram às mesmas conclusões quanto ao desempenho dos estimadores. Resultados detalhados de uma faixa mais ampla de cenários são relatados por Krüger (1996).

Os estimadores baseados em regressão regional (NQR e 2P/AR) utilizaram como variáveis independentes nas regressões os valores de área de drenagem e declividade apresentados na seção anterior. Todas as comparações são baseadas no quantil de uma cheia com tempo de recorrência igual a 100 anos ($p = 0,99$), com estimadores calculados a partir das amostras de vazões independentes geradas segundo uma distribuição GEV. Todos os estimadores utilizam o método dos

momentos-L para estimativa dos parâmetros da distribuição GEV. Análises de sensibilidade realizadas com a substituição da distribuição geradora das vazões por uma log-normal a dois parâmetros não alteraram as conclusões apresentadas para o caso base analisado (Krüger, 1996).

O conjunto de figuras 2 mostra as raízes quadradas dos erros médios quadráticos (REMQ) calculadas no espaço logarítmico para os seis estimadores estudados, como função do L-Cv local, para regiões compostas por 50 locais, comprimentos de registro iguais a 10, 25, 50 e 100 anos e erros de modelo iguais a 0,02 e 0,08. Cada caso constitui uma combinação dessas três variáveis. O valor de L-Cv médio no eixo horizontal constitui-se uma média dos L-Cv gerados para cada local em todas as 5000 simulações. A Tabela 4 apresenta um resumo dos resultados de todas as simulações do caso base, em termos das médias aritméticas do REMQ e do valor absoluto da tendenciosidade (BIAS) para os 50 locais.

ANÁLISE DOS RESULTADOS E CONCLUSÕES

Este trabalho apresenta alguns métodos recentes de estimativa de quantis de cheias, dando ênfase à evolução das abordagens do tipo Index Flood, combinadas com o uso de momentos com pesos probabilísticos (PWMs) e a distribuição de probabilidades generalizada de extremos (GEV). Um estimador de quantis normalizados que utiliza dois parâmetros (locação e escala) da distribuição GEV estimados com dados locais e o parâmetro de forma obtido através de regressão regional com características fisiográficas das bacias foi desenvolvido. Este estimador foi denominado 2P/AR. Foi implementado um estimador bayesiano empírico (BAY), que utiliza uma combinação linear do estimador por regressão de quantis normalizados (NQR) e do estimador 2P/AR. O desempenho dos estimadores 2P/AR e BAY

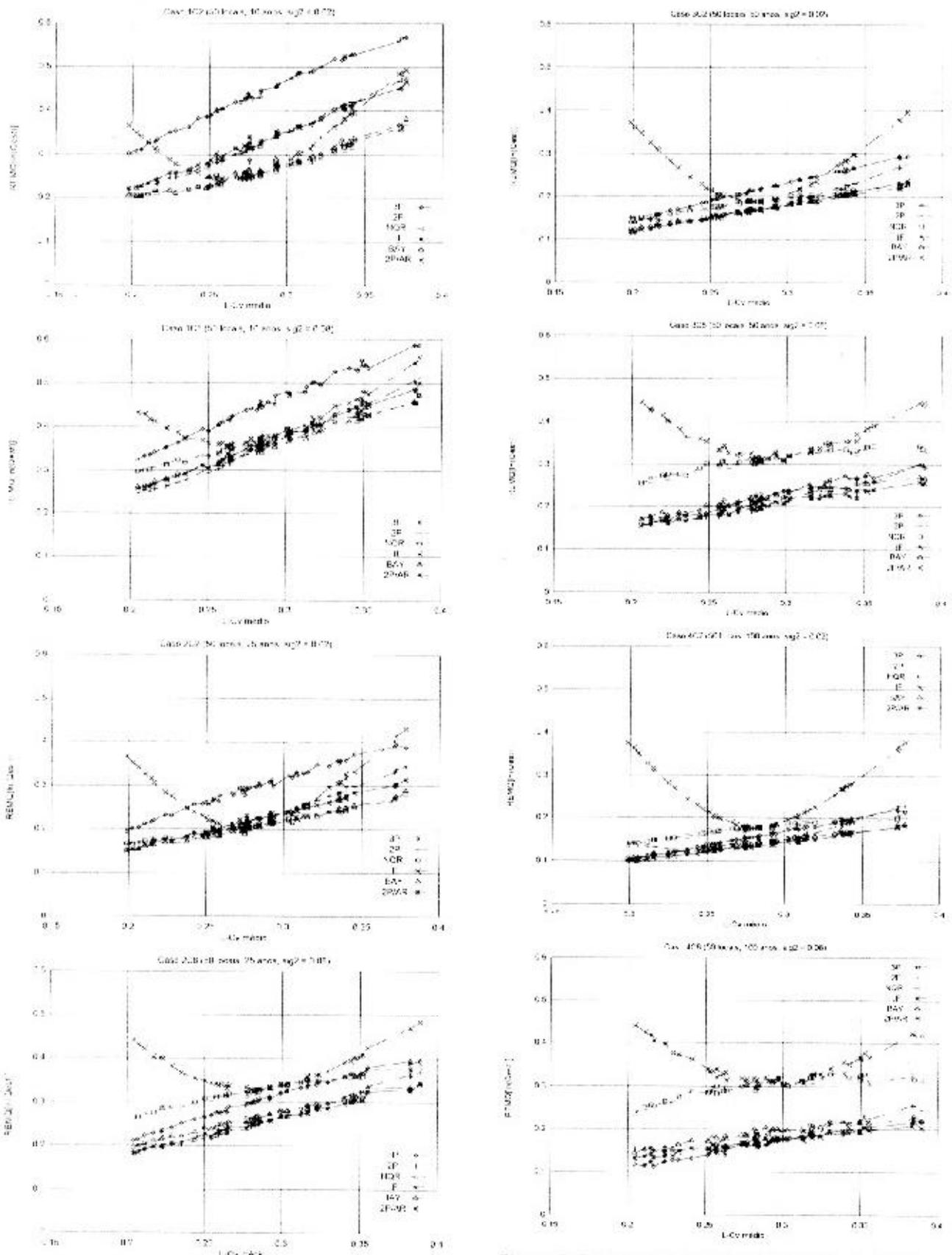


Figura 2. Resultados das simulações Monte Carlo.

Tabela 4. Resumo dos resultados (REMQ médio e BIAS médio).

estim.	Erros de modelo para o L-Cv							
	0,02 REMQ	0,02 BIAS	0,04 REMQ	0,04 BIAS	0,06 REMQ	0,06 BIAS	0,08 REMQ	0,08 BIAS
	<i>n</i> = 10							
2P	0,327	0,242	0,340	0,270	0,351	0,291	0,362	0,313
3P	0,438	0,042	0,444	0,043	0,449	0,041	0,455	0,040
IF	0,306	0,415	0,344	0,421	0,377	0,426	0,407	0,438
NQR	0,267	0,094	0,308	0,108	0,343	0,120	0,375	0,138
2P/AR	0,331	0,233	0,344	0,268	0,355	0,298	0,368	0,325
BAY	0,269	0,114	0,306	0,166	0,334	0,226	0,357	0,283
	<i>n</i> = 25							
2P	0,227	0,194	0,238	0,209	0,249	0,229	0,259	0,251
3P	0,296	0,027	0,299	0,022	0,304	0,023	0,307	0,027
IF	0,263	0,413	0,303	0,419	0,338	0,424	0,368	0,430
NQR	0,211	0,053	0,257	0,057	0,296	0,067	0,328	0,078
2P/AR	0,223	0,131	0,235	0,159	0,247	0,190	0,257	0,219
BAY	0,206	0,096	0,233	0,149	0,254	0,199	0,269	0,242
	<i>n</i> = 50							
2P	0,179	0,179	0,192	0,193	0,202	0,210	0,210	0,223
3P	0,216	0,011	0,221	0,014	0,223	0,021	0,226	0,033
IF	0,246	0,412	0,289	0,421	0,324	0,427	0,353	0,426
NQR	0,188	0,032	0,238	0,038	0,277	0,047	0,310	0,044
2P/AR	0,171	0,084	0,184	0,115	0,195	0,145	0,205	0,166
BAY	0,170	0,077	0,193	0,123	0,209	0,165	0,222	0,196
	<i>n</i> = 100							
2P	0,149	0,173	0,161	0,187	0,172	0,200	0,181	0,216
3P	0,157	0,007	0,160	0,011	0,163	0,013	0,166	0,022
IF	0,237	0,416	0,281	0,423	0,316	0,427	0,346	0,436
NQR	0,175	0,025	0,226	0,030	0,267	0,032	0,301	0,044
2P/AR	0,137	0,060	0,149	0,089	0,162	0,119	0,172	0,147
BAY	0,144	0,062	0,164	0,100	0,179	0,139	0,190	0,176

foram comparados através de simulações Monte Carlo com o desempenho de estimadores previamente estudados. As principais conclusões obtidas no trabalho são as seguintes:

- i. Na comparação dos estimadores, o estimador IF apresentou bom desempenho para locais com séries curtas de observações e *L-Cv* próximos da média da região, ou seja, em regiões homogêneas. A qualidade das estimativas decai acentuadamente para os locais com *L-Cv* diferenciado da média regional.
- ii. O desempenho do estimador NQR se aproxima do IF à medida em que a homogeneidade da região aumenta,

apresentando desempenho superior em regiões heterogêneas e com séries curtas (10 anos). Teoricamente, para regiões perfeitamente homogêneas, o estimador NQR se reduz ao estimador IF. A superioridade do estimador NQR é mais visível para os locais com elevado *L-Cv*. Nos casos de má adaptação do modelo regional utilizado pelo estimador NQR (erro de modelo alto), ocorre uma diminuição na qualidade do estimador NQR em termos médios, mantendo-se porém a superioridade relativa para os locais com *L-Cv* altos;

- iii. Para períodos de observação maiores (iguais ou maiores que 25 anos) ou no

caso de um mau ajuste do modelo regional os estimadores que utilizam dois parâmetros estimados com dados locais (2P e 2P/AR) são recomendáveis, geralmente superando o estimador NQR. Em locais com baixo $L-Cv$, os estimadores 2P e 2P/AR podem superar o estimador NQR mesmo para períodos curtos. O novo estimador 2P/AR apresentou quase sempre resultados melhores que o estimador 2P convencional. As maiores diferenças relativas favorecendo o estimador 2P/AR aparecem para períodos de observações entre 25 e 50 anos. Mesmo para períodos de observação muito longos (100 anos) o estimador 2P/AR superou os outros estimadores quando o modelo regional era bem adaptado à realidade ($\sigma^2 = 0,02$);

- iv. Somente para períodos de observações muito longos (maiores ou da ordem de 100 anos) o estimador com três parâmetros estimados localmente (3P) mostrou-se competitivo; ainda assim, superando o estimador 2P/AR apenas nos casos de um mau ajuste do modelo regional ($\sigma^2 = 0,08$), e em locais com $L-Cv$ abaixo da média da região;
- v. Os estimadores IF e 2P apresentam grandes tendenciosidades. O estimador 3P é praticamente não-tendencioso. Os estimadores NQR e 2P/AR apresentam pouca tendenciosidade, podendo ser considerados como "aproximadamente não-tendenciosos".
- vi. Em algumas situações pode ser difícil decidir qual o melhor estimador a ser empregado, NQR ou 2P/AR, sobretudo porque, na prática, o verdadeiro erro de modelo do qual depende esta decisão é desconhecido. Para resolver este dilema e melhorar ainda mais a precisão das estimativas de quantis, foi implementado um estimador bayesiano

empírico (BAY), resultante de uma combinação linear destes dois estimadores. O estimador BAY apresentou grande robustez pois, na quase totalidade dos casos analisados, apresentou os menores erros médios quadráticos. Apenas no caso de longos períodos de observação e mau ajuste do modelo regional, os estimadores 2P/AR ou 3P apresentam resultados ligeiramente superiores:

- vii. Assim como os dois estimadores que o compõe, NQR e 2P/AR, o estimador BAY pode ser considerado como aproximadamente não-tendencioso. Embora apresente, em alguns casos, desempenho ligeiramente inferior aos estimadores NQR, 2P ou 2P/AR, como em geral não se dispõem de informações sobre a adequação do modelo regional, o uso do estimador BAY parece ser uma escolha razoável e robusta para estimar vazões de cheias em regiões com registros inferiores a 100 anos.
- viii. Dentro de um enfoque de métodos robustos, as conclusões apresentadas sugerem que o uso de distribuições de três parâmetros com parâmetro de forma regional pode ser uma alternativa competitiva e recomendável para a estimativa de quantis de cheias.

REFERÊNCIAS

- BERGER, J. O. (1985) *Statistical decision theory and Bayesian analysis*. New York : Springer. 617 p. (Springer Series in Statistics).
- CUNNANE, C. (1988) Methods and merits of regional flood frequency analysis, *J. of Hydrology*, v. 100, n. 1/3, p. 269-290.
- DALRYMPLE, T. (1960) Flood-frequency analyses. *Geological Survey Water-Supply Paper*. Washington DC, paper 1543-A.
- FARQUHARSON, F. A. K., GREEN, C. S., MEIGH, J. R., SUTCLIFFE, J. V. (1987). Comparison of flood frequency curves for many different regions of the world : In: Singh, Vijay P. *Regional flood frequency Analysis*. Dordrecht : D Reidel. p. 223-256.

- FILL, H. D. (1994). *Improving flood quantile estimates using regional information*. Ithaca: Cornell Univ. 256 p. Tese (Doutorado) - Cornell Univ.
- GREENWOOD, J. A., LANDWEHR, J. M., MATALAS, N. C., WALLIS, J. R. (1979). Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Water Resources Res.*, v. 15, n. 5, p. 1049-1064, Oct.
- HOSKING, J. R. M. (1986). *The theory of probability weighted moments*. Yorkton Heights: IBM Research Division, 160 p. Research Report RC 12210.
- HOSKING, J. R. M., WALLIS, J. R., WOOD, E. F. (1985). Estimation of the generalized extreme value distribution by the method of probability weighted moments. *Technometrics*, v. 27, n. 3, p. 251-261, Aug.
- JENKINSON, A. F. (1955). The Frequency Distribution of the Annual Maximum (or Minimum) Values of Meteorological Elements. *Journal of the Royal Meteorological Society* 81, pp. 158-171.
- JOHNSTON, J. (1984). *Econometric Methods*. 3. ed. New York: McGraw-Hill. 568 p.
- KRUGER, C. M. (1996). *Estimador de quantis de cheias baseado em assimetria regional*. Curitiba: UFPR. 132 p. Dissertação (mestrado).
- LANDWEHR, J. M., MATALAS, N. C., WALLIS, J. R. (1979). Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. *Water Resources Res.*, v.15, n. 5, p. 1055-1064, Oct.
- LANDWEHR, J. M., MATALAS, N. C., WALLIS, J. R. (1980). Quantile estimation with more or less floodlike distributions. *Water Resources Res.*, v. 16, n. 3, p. 547-555, June.
- LETTENMAIER, D. P., WALLIS, J. R., WOOD, E. F. (1987). Effect of regional heterogeneity on flood frequency estimation. *Water Resources Res.* v. 23, n. 2, p. 313-323, Feb.
- LU, L. H. (1991). Statistical methods for regional flood frequency investigations. Ithaca: Cornell University. Tese (Doutorado) - Cornell University.
- LU, L. H., STEDINGER, J. R. (1992). Variance of two and three parameter GEV/PWM quantile estimators: formulae, confidence intervals, and a comparison. *J. of Hydrology*, v. 138, n. 1/2, p. 247-267, Sept.
- NATURAL ENVIRONMENT RESEARCH COUNCIL (1975). *Flood studies report: in five volumes*. London: NERC, v. 1.
- POTTER, K. W., LETTENMAIER, D. P. (1990). A comparison of regional flood frequency estimation methods using a resampling method. *Water Resources Res.*, v. 26, n. 3, p. 415-424, Mar.
- STEDINGER, J. R., TASKER, G. D. (1985). Regional hydrologic analysis. 1. Ordinary, weighted and generalized least squares compared. *Water Resources Res.*, v. 21, n. 9, p. 1421-1432, Sept.
- STEDINGER, J. R., VOGEL, R. M., FOUFOULA-GLORGIU, E. (1993). Frequency analysis of extreme events. In: MAIDMENT, D. R. ed. *Handbook of Hydrology*, New York: McGraw-Hill, cap. 18.
- TASKER, G. D., STEDINGER, J. R. (1986). Regional skew with weighted LS regression. *J. of Water Resources Planning and Management*, v. 112, n. 2, p.225-237, Apr.
- WALLIS, J. R. (1981). Risk and uncertainties in the evaluation of flood events for the design of hydraulic structures. In: Guggino, E. et al. *Piene e siccità*. Catania: Fondazione Politecnica del Mediterraneo. p. 3-36.
- WALLIS, J. R., WOOD, E. F. (1985). Relative accuracy of log Pearson III Procedures. *J. of Hydraulic Eng.*, v. 111, n. 7, p. 1043-1056, July.

A Flood Quantile Estimator Based on Regional Skew

ABSTRACT

A new regionalized flood quantile estimator using the generalized extreme value distribution (GEV) and the probabilistic weighted moments method (PWM) is presented. At-site data are used for the estimation of the location and scale parameters and a regional regression with physiographic characteristics is employed for the shape parameter of the flood flow distribution. The performance of the new estimator is compared with similar estimators already available in the literature by Monte Carlo simulation. These comparisons show the superiority of the proposed estimator in the case of moderate records and model errors of

the regional regression from moderate to large. An empirical Bayesian estimator, which combines the results of the proposed estimator and a normalized quantile regression estimator using two regional and one at-site parameters is also presented. This Bayesian estimator is an accurate and robust choice over a fair range of record sizes, model errors in the regional regression and values of the at-site L-coefficient of variation (L-Cv).