

## **Método para a Estimação de Quantis de Enchentes Extremas com o Emprego Conjunto de Análise Bayesiana, de Informações não Sistemáticas e de Distribuições Limitadas Superiormente – Parte 2: Aplicação**

**Wilson Fernandes\***, **Mauro Naghettini\***, **Rosângela Loschi\***

*wilson@ehr.ufmg.br; naghet@netuno.lcc.ufmg.br; loschi@est.ufmg.br*

*Recebido: 17/01/12 - revisado: 27/08/12 - aceito: 24/01/13*

### **RESUMO**

*Este é o segundo de dois artigos complementares que versam sobre um método para a estimação dos quantis de enchentes extremas, a partir de uma abordagem Bayesiana, com a possibilidade de inclusão de informações históricas e paleohidrológicas e o emprego de distribuições de probabilidades com cauda superior finita. Este artigo apresenta três aplicações a bacias com diferentes conjuntos de dados. Os estudos de caso foram selecionados de forma a avaliar diferentes aspectos do arcabouço metodológico. As aplicações mostraram que o método proposto melhora as estimativas de quantis extremos de vazão em dois aspectos: do ponto de vista descritivo, exemplificado pelos bons ajustes das distribuições de probabilidades limitadas superiormente aos dados sistemáticos e não sistemáticos; e do ponto de vista da diminuição das incertezas na predição de vazões, exemplificado pelos intervalos de credibilidade mais curtos, obtidos ao se utilizarem dados não sistemáticos. Além disso, o método proporcionou uma maneira racional de se incluir a estimativa local da PMF na análise de frequência. Por fim, verificou-se que o conhecimento subjetivo do especialista, além de poder ser incorporado à análise por meio da correta definição da distribuição a priori do limite superior, melhora significativamente os resultados a posteriori tanto dos parâmetros como dos quantis com pequena probabilidade de excedência.*

**Palavras Chave:** *enchentes extremas, análise de frequência de vazões, eventos hidrológicos raros.*

### **INTRODUÇÃO**

O primeiro de dois artigos complementares, publicados neste número da Revista Brasileira de Recursos Hídricos, tratou dos fundamentos teóricos associados à abordagem Bayesiana para a análise de frequência de cheias extremas com o uso de distribuições de probabilidades limitadas superiormente. Este segundo artigo procura avaliar os benefícios da referida abordagem por meio da sua aplicação a três casos distintos, a saber: bacias do rio American, do rio Llobregat e do rio Pará, localizadas, respectivamente, nos Estados Unidos, Espanha e Brasil.

A escolha da bacia do rio American foi devida à abundância de dados disponíveis e à existência de estudos envolvendo outros métodos de análise, o que permitiu compará-los ao método descrito no primeiro de dois artigos complementares aqui publicados. O rio Llobregat

foi escolhido pelo fato de ter sido objeto de estudo, sob a ótica frequentista, de Botero (2006), empregando distribuições limitadas. Assim, nesse caso, é possível comparar os métodos bayesianos e os métodos convencionais. A bacia do rio Pará foi escolhida pelo fato de não possuir muita informação hidrológica, o que permitirá a avaliação do método proposto em um cenário de escassez de dados. Além disso, a bacia do rio Pará foi objeto de estudo de Fernandes e Naghettini (2008), que aplicaram diversos métodos de análise de cheias, o que permitiu a comparação dos resultados do método aqui proposto com aqueles obtidos por meio de uma ampla gama de métodos de estimação de cheias máximas anuais.

### **APLICAÇÃO PARA A BACIA DO RIO AMERICAN EM FOLSOM**

Aplicou-se o método proposto à bacia do rio American, no reservatório de Folsom, localizada no estado americano da Califórnia. Para uma descrição

\*Universidade Federal de Minas Gerais

mais pormenorizada, dessa aplicação recomenda-se ao leitor a consulta de Fernandes et al. (2010).

De acordo com USBR (2002), a bacia do rio American no reservatório de Folsom possui uma área de drenagem de 4.820 km<sup>2</sup> e é monitorada pelo Serviço Geológico dos Estados Unidos desde 1905 e possui 52 anos de dados completos. A série de dados sistemáticos foi considerada aleatória, independentemente, homogênea e estacionária, a um nível de significância de 5%, conforme os testes convencionais dessas hipóteses sugeridos em Naghettini e Pinto (2007).

No que se refere aos dados não sistemáticos, estudos conduzidos por USBR (2002) identificaram a presença de 4 níveis paleohidrológicos distintos, todos do tipo UB, estendendo, assim, a série de dados em até 2000 anos anteriores ao presente, ou, de modo sintético, em até 2000 AP. Além disso, o local possui 5 estimativas de cheias do tipo DB. A figura 1 mostra a estrutura do conjunto completo de dados disponíveis para a bacia.

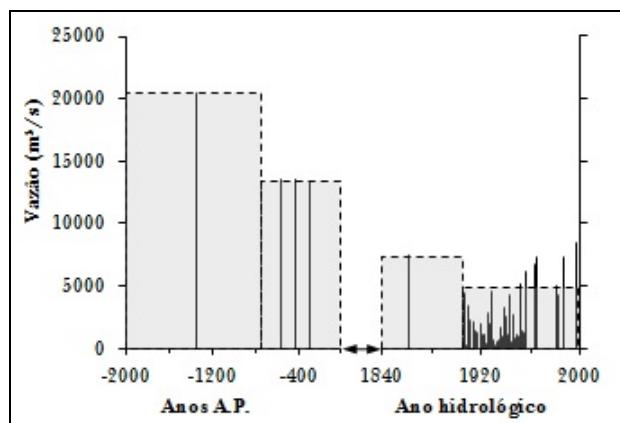


Figura 1 – Dados hidrológicos disponíveis para o rio American (A.P. – Antes do Presente)

#### Distribuições *a priori* para o limite superior das vazões

Os métodos utilizados para eliciar a distribuição *a priori* do limite superior das vazões máximas anuais foram discutidos no primeiro destes dois artigos complementares. De acordo com os procedimentos A e B e considerando a estimativa da vazão de pico da PMF, elaborada por USACE (2001), de 25.655 m<sup>3</sup>/s, foram definidas 4 distribuições *a priori*. Além dessas, considerou-se ainda uma quinta distribuição, pouco informativa, denotada por (V). A tabela 1 mostra as características destas distribuições, enquanto a

figura 2 mostra suas funções densidade de probabilidades.

Tabela 1 – Características das distribuições *a priori* para o rio American

Mod.	$\rho_\alpha$	$\beta_\alpha$	Média	CV <sub>α</sub>	P( $\alpha > \text{PMF}$ )
I	11,11	4,17x10 <sup>-4</sup>	26633	0,3	0,500
II	4,00	1,43x10 <sup>-4</sup>	27952	0,5	0,500
III	2,04	6,64x10 <sup>-5</sup>	30717	0,7	0,500
IV	5,20	4,33x10 <sup>-4</sup>	12018	0,4	0,017
V	1,00	10 <sup>-8</sup>	10 <sup>8</sup>	1,0	~1,0

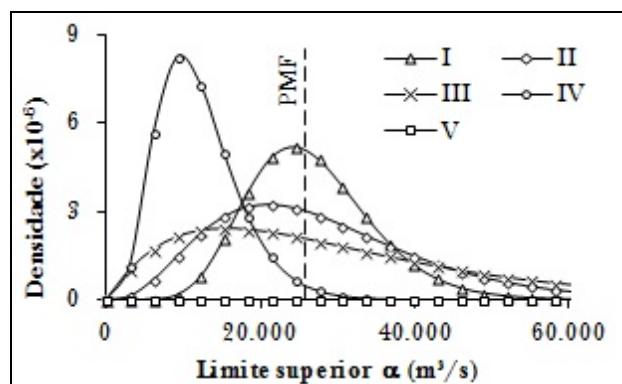


Figura 2 – Distribuições *a priori* para o limite superior do rio American

#### Estatísticas *a posteriori*

As estatísticas *a posteriori* das quantidades de interesse são obtidas por meio da equação 1, referida no primeiro artigo complementar. O cálculo analítico de integrais do tipo mostrado na equação 1 é impossível na maioria das aplicações práticas, especialmente nos casos multidimensionais. O cálculo dessas integrais multidimensionais pode ser evitado com a utilização algoritmos de amostragem, tais como aqueles que empregam a integração de Monte Carlo via cadeias de Markov – MCMC, da sigla em inglês Markov Chain Monte Carlo. Com efeito, de acordo com Gilks et al. (1996), essa classe de algoritmos permite obter uma amostra de uma distribuição de probabilidades, tal como  $\pi(\theta|X)$  na equação 1, através de uma cadeia de Markov construída de forma tal que, após um grande número de realizações, sua distribuição de equilíbrio seja  $\pi(\theta|X)$ . Uma vez tendo uma amostra da distribuição *a posteriori* de  $\theta$ , ou seja,  $\pi(\theta|X)$ , a estimativa de qualquer valor de interesse pode ser obtida por integração de Monte Carlo.

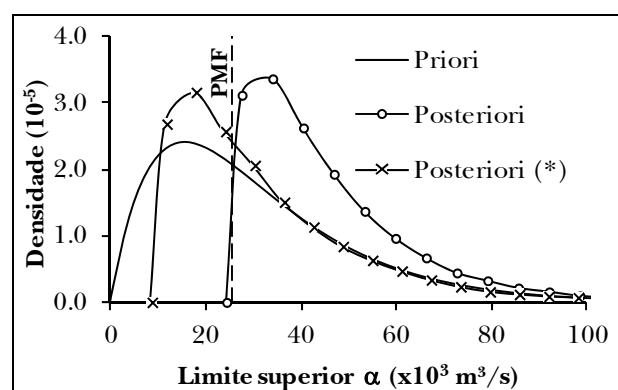
Amostras da distribuição conjunta dos parâmetros da LN4 foram obtidas por meio do software WinBUGS (LUNN, 2000). Para cada caso, gerou-se uma cadeia de tamanho 600.000, após a convergência ter sido atingida foram descartadas as 100.000 primeiras realizações como período de *burn-in* e tomou-se um *lag* de 10 iterações para evitar a autocorrelação da cadeia. Assim, foi obtida uma amostra de tamanho 50.000 da distribuição *a posteriori* dos parâmetros.

A tabela 2 mostra um resumo dos resultados encontrados para o limite superior  $\alpha$ . A figura 3 mostra as distribuições *a priori* e *a posteriori* para o modelo III. Os modelos marcados com “\*” referem-se àqueles que utilizaram somente dados sistemáticos em sua formulação.

**Tabela 2 – Estatísticas *a posteriori* do limite superior para bacia do rio American**

Mod.	Média	CV $_{\alpha}$	P( $\alpha$ >PMF)	95% IC (HPD)
I	32.538	0,19	0,91	(24.070; 44.140)
II	38.109	0,29	0,94	(24.080; 59.810)
III	44.735	0,40	0,96	(24.070; 80.290)
IV	28.104	0,12	0,72	(24.070; 34.980)
V	9,5x10 <sup>7</sup>	1,03	1,0	(5,9x10 <sup>4</sup> ; 2,9x10 <sup>8</sup> )
I*	25.947	0,30	0,47	(11.710; 41.370)
II*	27.296	0,47	0,47	(8.785; 51.770)
III*	31.478	0,62	0,51	(8.664; 70.190)
IV*	14.628	0,29	0,022	(8.609; 22.960)
V*	10 <sup>8</sup>	0,99	1,0	(5,7x10 <sup>4</sup> ; 3,0x10 <sup>8</sup> )

CV – coeficiente de variação, DP – desvio padrão, IC – Intervalo de credibilidade



**Figura 3 – Distribuições *a priori* e *a posteriori* para o limite superior do modelo III, rio American.**

A tabela 2 mostra uma notável redução do CV quando os dados não sistemáticos são incluídos na análise. Um exemplo disto pode ser visto nos modelos I e I\*, ambos baseados na mesma distribuição *a priori*. Uma vez que os dados sistemáticos são significativamente inferiores à estimativa da PMF, não se espera que os mesmos contribuam para aumentar o conhecimento sobre o limite superior. Por outro lado, a inclusão dos dados não sistemáticos produz dois efeitos desejáveis: o primeiro é o aumento do conhecimento sobre o limite superior, representado pela diminuição do CV, e o segundo é o aumento da certeza *a posteriori* sobre ele, representado pela menor amplitude do intervalo de credibilidade.

Análises semelhantes podem ser conduzidas para os demais modelos, com exceção dos modelos V e V\*. De fato, os modelos V e V\* mostram que, independentemente do tipo de dado incluído na função de verossimilhança, o uso de uma distribuição *a priori* não informativa para o limite superior não é recomendado neste caso. A estimativa de máxima verossimilhança para o limite superior da distribuição LN4 depende fortemente do tipo de dado que compõe a função de verossimilhança (BOTERO, 2006). Na presença de dados EX, UB e DB, como é o caso em questão, a função de verossimilhança é bastante plana. Portanto, a informação trazida pela verossimilhança sobre o limite superior não é bastante forte para reduzir de forma satisfatória a incerteza *a posteriori* sobre  $\alpha$  e, assim, “qualquer valor maior que o máximo observado” pode ser considerado uma estimativa para o limite superior.

Dentre os modelos analisados, aquele que apresentou o melhor ajuste foi o III, embora os modelos I, II e IV (não mostrados aqui) tenham demonstrado capacidade descritiva semelhante. A figura 4 mostra a curva de quantis para os modelos III e V. A posição de plotagem para os dados empíricos foi calculada de acordo com o procedimento introduzido por Hirsch e Stedinger (1987) e modificado por Naulet (2002).

Novamente, nesta análise, é possível verificar o ganho, em termos descritivos, decorrente da inclusão dos dados não sistemáticos e de uma distribuição *a priori* informativa para o limite superior. Conforme menção anterior, maiores detalhes da aplicação do método proposto na bacia do rio American podem ser visto em Fernandes et al. (2010).

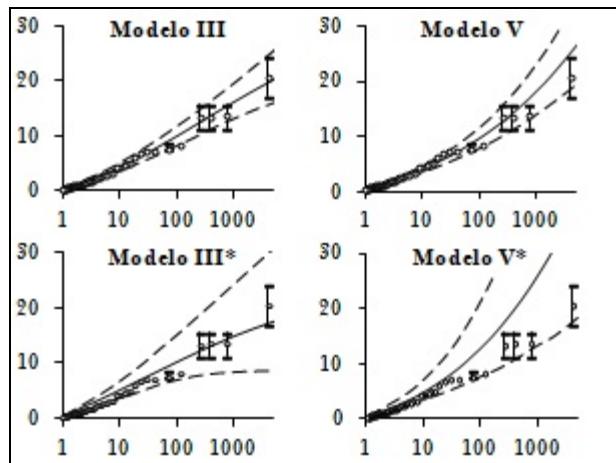


Figura 4 – Curva de quantis para as vazões máximas anuais do rio American de acordo com os modelos III e V. O eixo das abscissas é o período de retorno em anos, e o das ordenadas é o quantil em  $10^3 \text{ m}^3/\text{s}$ . As barras verticais representam os dados do tipo DB.

### APLICAÇÃO PARA A BACIA DO RIO LLOBREGAT, EM PONT DU VILOMARA

Outra aplicação do método proposto foi feita à bacia do rio Llobregat, em Pont Du Vilomara, localizada na região nordeste da Espanha. Essa bacia foi escolhida pelo fato de ter sido objeto de estudo de Botero (2006), que empregou distribuições limitadas superiormente na modelagem de vazões máximas anuais e utilizou o método clássico de máxima verossimilhança para estimar os parâmetros das distribuições. Com isso, foi possível avaliar as principais características do método bayesiano frente aos métodos frequentistas. A figura 5 mostra a localização da bacia do rio Llobregat.

A bacia do rio Llobregat drena uma área de 1.882  $\text{km}^2$  e é monitorada pelo CEDEX (*Centro de Estudios y Experimentación de Obras Públicas*). A bacia conta com registros sistemáticos diários de vazões no período de 1916 a 1989 e 48 anos de dados completos. Também neste caso a série foi considerada aleatória, independente, homogênea e estacionária, a um nível de significância de 5%.

Os dados não sistemáticos para o rio Llobregat correspondem aos dados históricos compilados pelo projeto europeu SPHERE (*Systematic, Paleoflood and Historical Data for the Improvement of Flood Risk Estimation*) e obtidos por Thorndycraft et al. (2005) a partir da datação, por meio de  $^{14}\text{C}$ , dos perfis estratigráficos de Pont Du

Vilomara. Assim, foram identificadas quatro cheias do tipo LB e quatro níveis UB no período histórico. Além disso, foram identificadas seis cheias do tipo DB no período sistemático. A figura 6 mostra a estrutura do conjunto de dados disponíveis para o rio Llobregat.

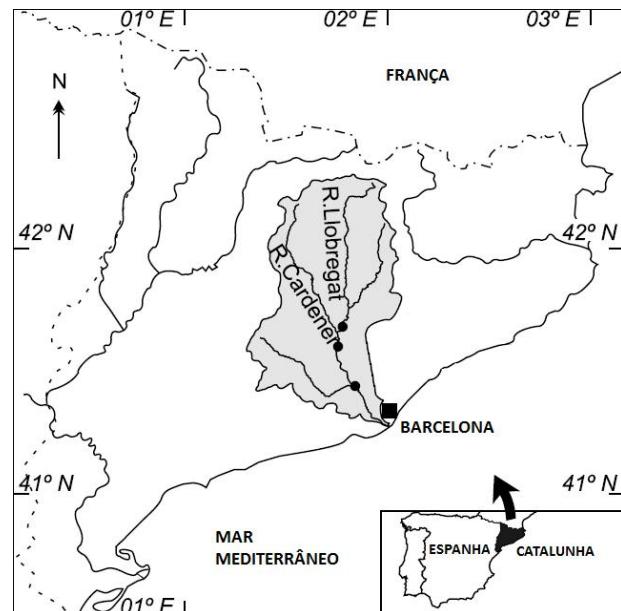


Figura 5 – Localização da bacia do rio Llobregat (adapt. Thorndycraft et al., 2005)

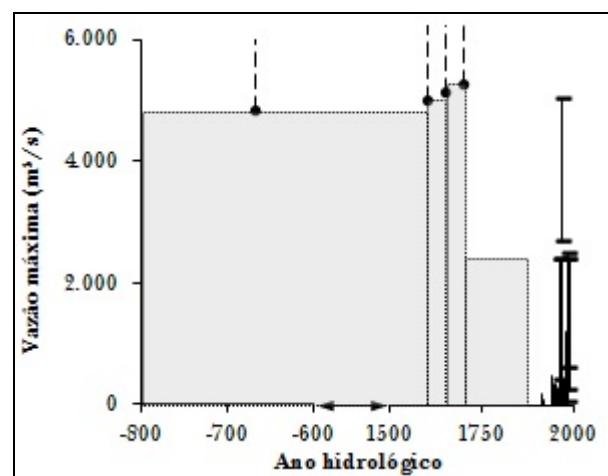


Figura 6 – Dados hidrológicos disponíveis para o rio Llobregat

*Distribuições a priori para o limite superior das vazões*

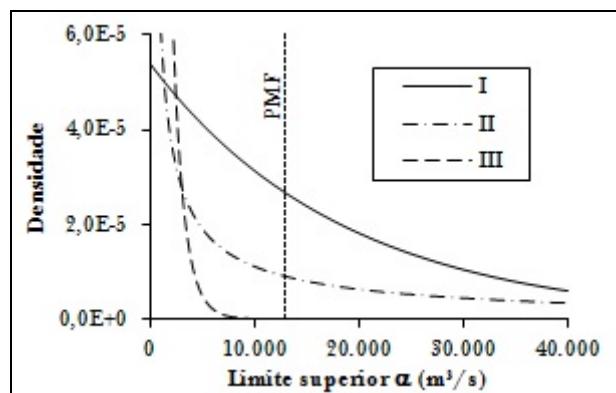
Adotando as recomendações de Botero (2006), a estimativa de cheia recorde para Pont Du Vilomara foi obtida por meio da curva envoltória

definida por Rodier e Roche (1984). Com isso tem-se uma estimativa de  $12.887 \text{ m}^3/\text{s}$ . Analogamente ao estudo para os Estados Unidos, os 242 recordes de vazões compilados pelo CEDEX na Espanha foram divididos em quatro grupos, de acordo com a similaridade da área de drenagem, e forneceram como resultado coeficientes de variação entre 1,0 e 2,0.

Seguindo o procedimento A, com  $\rho = 0,5$ , PMF igual a  $12.887 \text{ m}^3/\text{s}$  e  $CV = 1,0$  e 2,0, foram definidas duas distribuições *a priori* para o limite superior. Seguindo o procedimento B, com 242 cheias recordes compiladas pelo CEDEX, transpostas para a bacia do rio Llobregat, em Pont Du Vilomara a partir da curva envoltória de vazões definida por Rodier e Roche (1984) foi definida a distribuição *a priori* III. Além disso, foi considerada uma distribuição *a priori* pouco informativa, tal como para a aplicação do rio American. A tabela 3 mostra as características destas distribuições, enquanto a figura 7 mostra suas respectivas funções densidade de probabilidades.

**Tabela 3 – Características das distribuições *a priori* para o rio Llobregat**

Mod.	$\rho_\alpha$	$\beta_\alpha$	Média	$CV_\alpha$	$P(\alpha > \text{PMF})$
I	1,00	$5,38 \times 10^{-5}$	18592	1,0	0,500
II	0,25	$2,27 \times 10^{-6}$	76371	2,0	0,500
III	2,04	$6,64 \times 10^{-5}$	668	1,4	~0,0
IV	1,00	$10^{-8}$	$10^8$	1,0	~1,0



**Figura 7 – Distribuições *a priori* para o limite superior do rio Llobregat**

#### Estatísticas *a posteriori*

As estatísticas *a posteriori* foram obtidas seguindo os mesmos critérios numéricos utilizados

para o rio American. A tabela 4 resume os resultados encontrados.

**Tabela 4 – Estatísticas *a posteriori* do limite superior para o rio Llobregat**

Mod.	Média	CV	$P(\alpha > \text{PMF})$	95% IC (HPD)
I	38.116	0,57	0,97	(8.704; 81.390)
II	191.151	1,13	0,99	(7.806; 610.600)
III	9.991	0,21	0,09	(6.574; 14.120)
IV	$10^8$	0,99	1,0	(2,1x10 <sup>5</sup> ; 3,0x10 <sup>8</sup> )
I*	19.080	0,95	0,51	(1.219; 55.420)
II*	99.257	1,65	0,68	(1.241; 419.600)
III*	2.546	0,44	~0,0	(1.212; 4.760)
IV*	$10^8$	1,00	1,0	(3,4x10 <sup>4</sup> ; 3,0x10 <sup>8</sup> )

A avaliação dos resultados para o rio Llobregat é semelhante àquela feita para o rio American. No entanto, neste caso, houve uma redução na amplitude do intervalo de credibilidade ao se utilizarem somente dados sistemáticos. Esse resultado não esperado pode ser explicado pelo fato de os dados hidrológicos do rio American (PMF e paleovazões) serem muito mais informativos, no que tange ao limite superior, do que os dados do rio Llobregat (cheias recordes e paleovazões).

O principal objetivo da aplicação para a bacia do rio Llobregat é comparar os resultados obtidos pelo método bayesiano com os resultados obtidos pelos métodos clássicos. Botero (2006) ajustou a distribuição LN4 aos dados de Pont Du Vilomara a partir de dois métodos: estimativa de todos os parâmetros, incluindo o limite superior, pelo método da máxima verossimilhança (ML) e o da fixação do limite superior na PMF e estimativa dos demais parâmetros pelo máximo de verossimilhança (PMF). A figura 8 ilustra a comparação, a partir da curva de quantis, entre o método aqui proposto (modelo I) e os resultados obtidos por Botero (2006).

O método bayesiano, representado pelo modelo I na figura 8, apresentou um ajuste superior ao obtido usando os métodos clássicos, discutidos em Botero (2006). Estes não conseguiram se ajustar ao maior dado não sistemático, fornecendo uma estimativa menor que o limite inferior do último dado DB.

A comparação do método clássico e bayesiano, apenas pela curva de quantis, como aqui realizada, não mostra todas as vantagens de um sobre o outro. Deve ser considerado, neste contexto, que o método bayesiano, além de fornecer uma

estrutura mais racional para a incorporação de um limite superior no estudo de cheias extremas, possibilita uma análise mais detalhada dos parâmetros e quantis das distribuições de probabilidade.

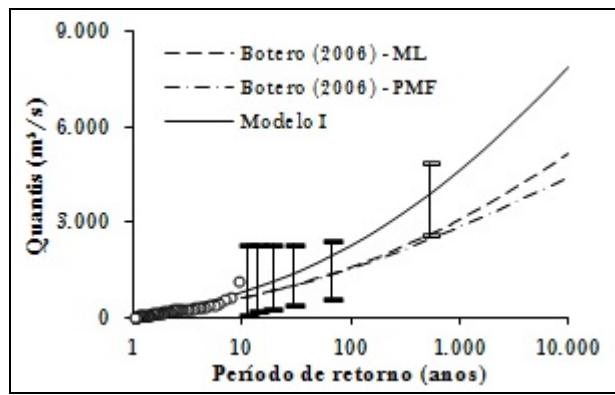


Figura 8 – Comparação entre os métodos bayesiano e clássico para a LN4. As barras verticais representam os dados do tipo DB.

#### APLICAÇÃO PARA A BACIA DO RIO PARÁ, EM PONTE DO VILELA

Uma última aplicação do método proposto foi feita na bacia do rio Pará, em Ponte do Vilela, Minas Gerais, Brasil. Essa bacia foi escolhida pelo fato de não haver qualquer estudo paleohidrológico na mesma, o que possibilitou a avaliação do método proposto sob condições de escassez de informações. Além disso, essa bacia foi objeto de estudo de Fernandes e Naghettini (2008), que aplicaram uma série de métodos na estimativa de cheias extremas. Assim, foi possível comparar os resultados do método proposto com os resultados obtidos por outras abordagens. A figura 9 mostra a localização da bacia objeto de estudo.

A bacia do rio Pará possui uma área de drenagem de 1.620 km<sup>2</sup> e é monitorada pela Companhia Energética de Minas Gerais – CEMIG, entre outras entidades. Os dados de vazão se estendem por um período que vai de 1938 a 2000 e 63 anos de dados completos. Essa série foi considerada aleatória, independente, homogênea e estacionária, a um nível de significância de 5%.

A bacia não possui estudos paleohidrológicos, restringindo a disponibilidade de dados àqueles sistemáticos. A figura 10 mostra a estrutura do conjunto de dados disponíveis para o local.

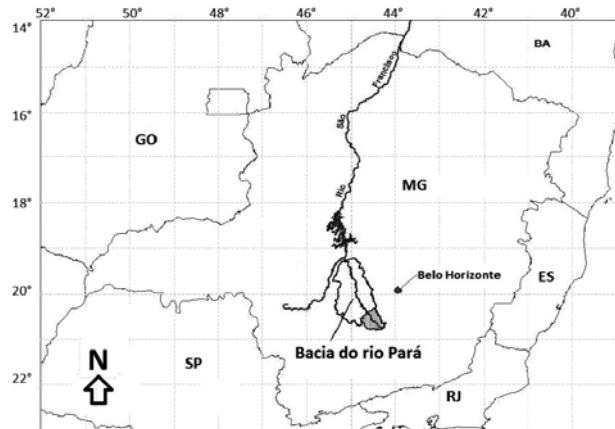


Figura 9 – Localização da bacia do rio Pará

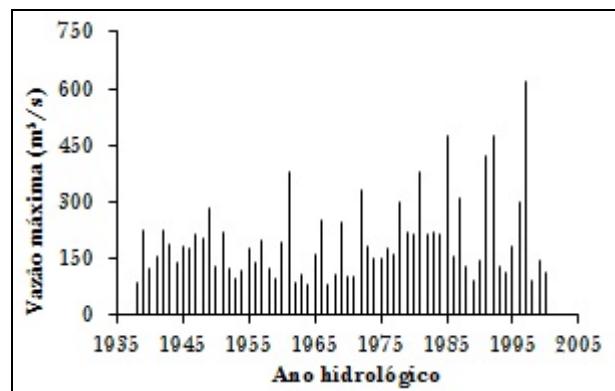


Figura 10 – Dados hidrológicos disponíveis para o rio Pará

#### Distribuições *a priori* para o limite superior das vazões

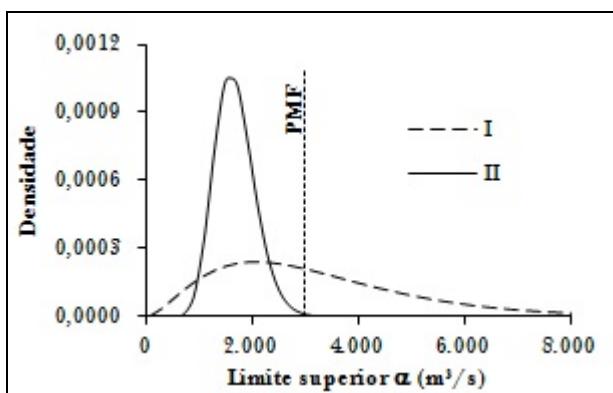
O Comitê Brasileiro de Barragens (CBDB, 2002) publicou uma compilação de dados sobre 61 aproveitamentos energéticos do Brasil. Destes, apenas 34 contam com estimativas de PMF, constituindo a base para a especificação do limite superior para Ponte do Vilela. Uma vez que a amostra é composta de apenas 34 valores, não foi feita uma divisão dos dados em grupos, tal como nos outros casos. Assim, a análise se estende a bacias com áreas de drenagem bastante distintas, perfazendo um coeficiente de variação da ordem de 0,6.

Uma vez que não há estimativa de PMF para Ponte do Vilela, a mesma foi estimada pela curva envoltória estimada a partir dos dados do CBDB (2002). Com isso, tem-se uma PMF de 2.988 m<sup>3</sup>/s. Seguindo o procedimento A, com  $p = 0,5$  e  $CV = 0,6$ , foi definida uma distribuição *a priori* para o limite superior. Seguindo o procedimento B, com 34

estimativas de PMF compiladas pelo CBDB (2002) transpostas para a bacia do rio Pará a partir da curva envoltória estimada para a região, foi definida uma segunda distribuição *a priori*. Além disso, foi considerada uma distribuição *a priori* pouco informativa, tal como para a aplicação do rio American. A tabela 5 mostra as características destas distribuições, enquanto a figura 11 mostra suas respectivas funções densidade de probabilidades.

**Tabela 5 – Características das distribuições *a priori* para o rio Pará**

Mod.	$\rho_\alpha$	$\beta_\alpha$	Média	CV $_\alpha$	P( $\alpha > \text{PMF}$ )
I	2,78	$8,54 \times 10^{-4}$	3255	0,6	0,500
II	19,91	$1,18 \times 10^{-2}$	1693	0,2	~0,0



**Figura 11 – Distribuições *a priori* para o limite superior na bacia do rio Pará**

#### Estatísticas *a posteriori*

Amostras da distribuição conjunta dos parâmetros da LN4 foram obtidas por meio do software WinBUGS (LUNN, 2000). Para cada caso, gerou-se uma cadeia de tamanho 1.200.000, após a convergência ter sido atingida foram descartadas as 200.000 primeiras realizações como período de *burn-in* e tomou-se um *lag* de 20 iterações para evitar a autocorrelação da cadeia. Assim, foi obtida uma amostra de tamanho 50.000 da distribuição *a posteriori* dos parâmetros. A tabela 6 mostra os resultados obtidos.

Os resultados *a posteriori* para o limite superior são bastante semelhantes aos encontrados nos outros casos. Isso sugere as mesmas análises feitas anteriormente e não serão aqui repetidas.

**Tabela 6 – Estatísticas *a posteriori* do limite superior para o rio Pará**

Mod.	Média	CV	P( $\alpha > \text{PMF}$ )	95% IC (HPD)
I	3922	0,48	0,63	(1065; 7715)
II	1806	0,21	0,004	(1120; 2544)
III	$10^8$	1,00	1,0	(1,3x10 <sup>5</sup> ; 3,0x10 <sup>8</sup> )

O principal objetivo da aplicação do rio Pará é a comparação da abordagem proposta com outros métodos estatísticos tradicionalmente utilizados para estimativa de quantis extremos de cheias. Fernandes e Naghettini (2008) listam os quantis máximos de vazão para Ponte do Vilela obtidos pela aplicação dos seguintes métodos:

- Modelo chuva-vazão (SIM): Lima (2004) desenvolveu um método que emprega um modelo estocástico de geração de chuva diária, baseado em uma cadeia de Markov de primeira ordem, e o modelo chuva-vazão Rio Grande (Naghettini et al., 2002) para gerar longas séries sintéticas de vazão. Lima (2004) aplicou o método à bacia do rio Pará, gerando uma série contínua de 10.000 anos de vazão, da qual foi feita uma análise de frequência simples com as vazões máximas anuais;
- Análise de frequência convencional (AFC): neste caso, foi feito o ajuste da distribuição GEV aos dados observados de Ponte do Vilela utilizando o método dos momentos-L;
- Análise de frequência regional (REG): CPRM (2001) fez um estudo de regionalização de vazões máximas anuais numa área que abrange a bacia do rio Pará. O cálculo dos quantis máximos de vazão foi feito empregando-se a metodologia proposta por Hosking e Wallis (1997) para análise regional com momentos-L; e
- Pico-Volume-Precipitação (PVP): aplicação do método proposto por Fernandes e Naghettini (2008).

A figura 12 mostra a comparação entre as curvas de quantis obtidas pelos métodos listados acima e com o modelo LN4-I.

Na faixa dos dados observados, até um período de retorno em torno de 100 anos, todos os métodos forneceram resultados semelhantes. A partir desse valor, ou seja, na faixa de extração, as diferenças entre os métodos começam a aparecer. As questões envolvendo a engenharia de recursos

hídricos residem precisamente nessa faixa, onde é necessário fazer previsões para vazões extremas ainda não observadas. Assim, a adequação do modelo não deve ser vista somente sob o ponto de vista da aderência, mas também sob o ponto de vista estrutural.

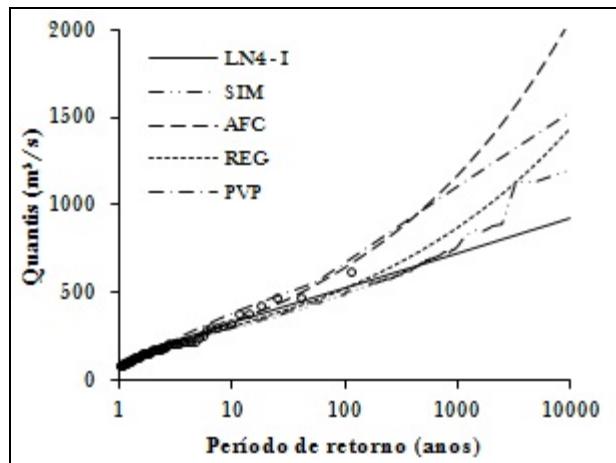


Figura 12 – Comparação entre diversos métodos para os dados de Ponte do Vilela

Neste sentido, destacam-se os modelos REG e o PVP, os quais agregam dados de diferentes fontes em suas respectivas estruturas. No que se refere à previsão, o modelo que mais destoou dos demais foi a análise de frequência convencional (AFC), com os maiores valores para quantis com altos períodos de retorno. De fato, todos os métodos, com exceção da abordagem proposta, predizem quantis bastante altos para a faixa de valores não observados. Uma vez que não há nenhuma informação sobre o comportamento das vazões nessa faixa de valores, pressupõe-se que os métodos tradicionais superestimem as vazões com período de retorno alto.

Do que foi exposto, espera-se que o método proposto estime quantis extremos de uma forma mais verossímil, já que supor quantis que aumentam indefinidamente parece contradizer os princípios físicos que atuam na formação das vazões de cheia. Um estudo paleohidrológico na bacia poderia fornecer indícios mais esclarecedores no que se refere às suposições adotadas nessa análise.

## CONCLUSÃO

A hipótese de que as vazões máximas anuais possam ser consideradas como variáveis aleatórias

limitadas superiormente motivou o uso de distribuições limitadas para descrever seu comportamento probabilístico. Dentre as distribuições limitadas, foi utilizada nesta pesquisa o modelo LN4, segundo a proposta de Botero (2006).

O método desenvolvido permitiu incluir o conceito de PMF na análise de frequência através do uso de distribuições limitadas superiormente. As estimativas, tanto para os quantis de vazão quanto para o limite superior, foram aprimoradas pela inclusão de informações não sistemáticas. Diferentemente do método proposto por Botero (2006), a inferência sobre os componentes do modelo (vazões observadas, paleocheias e parâmetros) foi feita seguindo o paradigma bayesiano. Essa abordagem permitiu incluir na análise informações sobre a PMF, de forma científica, uma vez que seu valor é visto como uma informação *a priori* sobre o limite superior e não como o próprio limite superior.

No que se refere à inclusão de dados não sistemáticos, verificou-se que os mesmos exercem papel determinante na análise de eventos extremos de vazão. Conforme visto no primeiro destes dois artigos complementares aqui publicados, vários pesquisadores demonstraram que a incorporação de dados não sistemáticos à análise de frequência pode estender significativamente o período de observação, reduzindo o grau de extrapolação e melhorando o nível de confiança na estimação do risco associado a eventos raros. Esse fato também foi verificado para o método proposto. A redução significativa da amplitude do intervalo de credibilidade, com a inclusão de informações não sistemáticas, é um indicativo de que esse tipo de dado contribui não só para a melhoria das estimativas dos parâmetros, mas também para uma melhor previsão de valores futuros de vazão, sobretudo aqueles com baixa probabilidade de excedência.

A inclusão da PMF na análise foi feita através da especificação da distribuição *a priori* para o limite superior. Dois procedimentos foram propostos: o primeiro baseou-se na distribuição espacial da PMF (Procedimento A) e o outro baseou-se na transposição da PMF de outras bacias para o local de interesse por meio da curva envoltória regional (Procedimento B). De uma forma geral, ambos os procedimentos levaram a distribuições *a priori* plausíveis para o limite superior. No entanto, o procedimento A é mais intuitivo, no sentido de permitir ao especialista introduzir na análise seu conhecimento acerca da PMF local, ou seja, compete a ele definir quão

precisa é a estimativa atual da PMF e quanto próxima a estimativa da PMF está do limite superior das vazões máximas anuais. Assim, o procedimento A está mais condizente com a definição de distribuição *a priori*. Adicionalmente, verificou-se, a partir da aplicação para o rio Llobregat, que o uso de vazões recordes, em contraposição a um banco de dados de PMF's, também permite uma boa descrição para o limite superior.

Os resultados permitiram concluir a favor do uso de distribuições *a priori* informativas para o limite superior. De fato, o uso de distribuições *a priori* não informativas para o limite superior fornece estimativas *a posteriori* além da plausibilidade física para as bacias analisadas. As curvas de quantis para esses casos foram as que apresentaram os piores ajustes, além de terem os maiores intervalos de credibilidade. Assim, não parece haver vantagem em utilizar o método proposto quando não for possível estabelecer uma distribuição informativa para o limite superior.

De uma forma geral, os objetivos pretendidos foram alcançados. As principais conclusões podem ser resumidas da seguinte forma:

- Pelo menos para as aplicações aqui realizadas, as distribuições de probabilidade limitadas superiormente, além de serem mais plausíveis para a modelagem de vazões máximas anuais, permitem um bom ajuste aos dados sistemáticos e não sistemáticos;
- A abordagem bayesiana permite avaliar mais profundamente as incertezas relativas à estimação de parâmetros e à predição de valores futuros de vazão. Neste particular, a inferência fornecida pela distribuição *a posteriori* dos parâmetros é muito mais informativa do que a estimação pontual da análise frequentista. Além disso, a possibilidade da inclusão de informações subjetivas e de outras fontes que não a amostral, que agregam o conhecimento do especialista a respeito de uma determinada quantidade, por meio da correta especificação das distribuições *a priori*, torna a análise mais intuitiva e mais semelhante ao processo de tomada de decisão em projetos de engenharia;
- A PMF, sobretudo em caráter regional, fornece informações essenciais para a construção de uma distribuição *a priori* informativa para o limite superior. Além disso, os procedimentos para a especificação

da distribuição *a priori* são flexíveis o bastante para acomodar o caráter subjetivo do conhecimento do especialista sobre a PMF;

- Os dados não sistemáticos melhoram significativamente as estimativas, tanto dos parâmetros quanto das distribuições preditivas de vazões. Na aplicação em Ponte do Vilela, onde não há informação não sistemática, os resultados foram adequados no que diz respeito à aderência aos dados observados. No entanto, a avaliação da capacidade preditiva do modelo fica comprometida, uma vez que a amostra é pequena. Assim, recomenda-se que, sempre que possível, sejam realizados estudos exploratórios com o objetivo de obter os dados não sistemáticos.

Embora as aplicações aqui descritas sejam bastante distintas, no que diz respeito à quantidade e à qualidade das informações, elas não são suficientes para garantir a generalização dos resultados obtidos. Neste sentido, Hosking e Wallis (1997) afirmam que a avaliação de métodos estatísticos deve ser feita por meio de experimentos de Monte Carlo.

Martins e Stedinger (2001) conceberam um experimento de Monte Carlo que mostra a viabilidade dessa técnica para gerar populações com dados sistemáticos e não sistemáticos. Assim, futuras pesquisas poderiam abordar o tema e generalizar os resultados aqui obtidos. Nesse mesmo sentido, poderia ser avaliada a influência da possível correlação espacial entre as estimativas de PMF nas distribuições *a posteriori*. Esses experimentos poderiam, por fim, guiar o analista na construção da distribuição *a priori* para o limite superior.

Por fim, vale salientar que a pesquisa não foi realizada com a intenção de se esgotar o assunto. Pelo contrário, as etapas metodológicas estão sujeitas a diferentes abordagens, podendo ser exploradas sob diversos contextos. A este respeito, a forma como foi construída a distribuição *a priori* para o limite superior reflete uma análise subjetiva, porém fisicamente embasada, sobre como descrever a variabilidade desse parâmetro, sendo perfeitamente possível conceber outras abordagens. Assim, espera-se que o método proposto constitua a base para desenvolvimentos futuros.

Alguns outros pontos não explorados podem ser abordados em futuros desenvolvimentos, quais sejam:

- Na construção das funções de verossimilhança (vide primeiro artigo complementar), estão implícitas as hipóteses de homogeneidade e estacionariedade dos dados sistemáticos e não sistemáticos. Essas hipóteses, sobretudo quando analisadas sob a ótica de dados paleohidrológicos, podem não se verificar na prática. Embora a função de verossimilhança seja flexível o suficiente para acomodar dados que não respeitem essas hipóteses, a especificação de distribuições dependentes do tipo de dado, e/ou com um conjunto paramétrico variável no tempo, é uma tarefa complexa, que ainda não foi abordada no contexto de análise de frequência de dados sistemáticos e não sistemáticos.
- Outra hipótese admitida na análise é a ausência de incertezas referentes às estimativas das paleocheias, bem como a data de ocorrência desses eventos, e os erros de medição dos dados sistemáticos. Mais uma vez, a função de verossimilhança pode ser alterada de forma a acomodar essas diferentes formas de erro. De fato, O'Connell et al. (2002) incluíram essas fontes de erro na função de verossimilhança dentro de uma abordagem bayesiana. No entanto, essa opção ainda não foi avaliada para uma função de verossimilhança construída com distribuições limitadas superiormente, devendo ser objeto de futuras pesquisas.

Por fim, o leitor interessado em mais detalhes metodológicos, bem como outros resultados diferentes daqueles aqui apresentados, pode consultar o trabalho de Fernandes (2009).

## AGRADECIMENTOS

Os autores agradecem à FAPEMIG (APQ 4683-5.04/07) e ao CNPq (305870/2006-8, 304505/2006-4 e 472877/2006-2) pelos apoios concedidos a esta pesquisa. Agradecem também aos revisores anônimos por suas valiosas sugestões.

## REFERÊNCIAS

BOTERO, B. A. *Estimación de crecidas de alto período de retorno mediante funciones de distribución com límite*

*superior e información no sistemática.* 223 f. Tese (Doutorado em Ingeniería Hidráulica y Medio Ambiente) – Departamento de Ingeniería Hidráulica y Medio Ambiente da Universidad Politécnica de Valencia, Espanha, 2006.

CBDB – COMITÉ BRASILEIRO DE BARRAGENS. *Large Brazilian spillways: an overview of Brazilian practice and experience in designing and building apillways for large dams.* CBDB/ICOLD, Rio de Janeiro, Brasil, 205 p., 2002.

CPRM – COMPANHIA DE PESQUISA DE RECURSOS MINERAIS. *Regionalização de vazões. Sub-bacias 40 e 41: Relatório Final - Vazões Máximas.* v. 4, CPRM/ANEEL. Belo Horizonte, 2001.

FERNANDES, W. *Método para a estimação de quantis de enchentes extremas com o emprego conjunto de análise bayesiana, de informações não sistemáticas e de distribuições limitadas superiormente.* Tese (Doutorado em Saneamento, Meio Ambiente e Recursos Hídricos) – Escola de Engenharia, Universidade Federal de Minas Gerais, Belo Horizonte, 2009.

FERNANDES, W.; NAGHETTINI, M. *Integrated frequency analysis of extreme flood peaks and flood volumes using the regionalized quantiles of rainfall depths as auxiliary variables.* Journal of Hydrologic Engineering, v. 13(3), p. 171-179, doi: 10.1061/(ASCE)1084-0699(2008)13:3(171), 2008.

FERNANDES, W.; NAGHETTINI, M.; LOSCHI, R. A *bayesian approach for estimating extreme flood probabilities with upper-bounded distribution functions.* Stoch. Environ. Res. Risk Assess., v. 24, p.1127-1143, doi:10.1007/s00477-010-0365-4, 2010.

GILKS, W. R.; RICHARDSON, S.; SPIEGELHALTER, D. J. *Markov Chain Monte Carlo in Practice.* London: Chapman & Hall/CRC, UK, 286 p., 1996.

HIRSCH, R. M.; STEDINGER, J. R. *Plotting positions for historical floods and their precision.* Water Resources Research, v. 23(4), p. 715-727, 1987.

HOSKING, J. R. M.; WALLIS, J. R. *Regional frequency analysis - an approach based on L moments.* Cambridge: Cambridge University Press, 240 p., 1997.

LIMA, A. A. *Metodologia integrada para determinação da enchente de projeto de estruturas hidráulicas por meio de séries sintéticas de precipitação e modelos chuva-vazão.*

Dissertação (Mestrado em Saneamento, Meio Ambiente e Recursos Hídricos) – Escola de Engenharia, Universidade Federal de Minas Gerais, Belo Horizonte, 2004.

LUNN, D. J.; THOMAS, A.; BEST, N.; SPIEGELHALTER, D. *WinBUGS - a bayesian modelling framework: concepts, structure, and extensibility*. Statistics and Computing, v. 10, p. 325-337, 2000.

MARTINS, E. S.; STEDINGER, J. R. *Historical information in a generalized maximum likelihood framework with partial duration and annual maximum series*. Water Resources Research, v. 37(10), p. 2559-2567, 2001.

NAGHETTINI, M. C.; NASCIMENTO, N. O.; THIMOTTI, T.; LIMA, A. A.; SILVA, F. E. O. *Modelo Rio Grande de Simulação Hidrológica para Previsão de Vazões de Curto Prazo: Formulação Teórica*. Departamento de Engenharia Hidráulica e Recursos Hídricos da UFMG, Belo Horizonte, 2002.

NAGHETTINI, M. C. e PINTO, E. J. A. *Hidrologia estatística*, CPRM – Serviço Geológico do Brasil, Belo Horizonte, 561 p., 2007.

NAULET, R. *Utilisation de l'information des crues historiques pour une meilleure prédétermination du risque d'inondation. Application au bassin de l'Ardèche à Vallon Pont-d'Arc et St-Martin d'Ardèche*. Thèse UJF, PhD INRS-ETE, Grenoble, França, 2002.

O'CONNELL, D. R. H.; OSTENAA, D. A.; LEVISH, D. R.; KLINGER, R. E. *Bayesian flood frequency analysis with paleohydrologic bound data*. Water Resources Research, v. 38(5), p. 16.1-16.13, 2002.

RODIER, J. A.; ROCHE, M. *World catalogue of maximum observed floods*. International Association of Hydrological Sciences, IAHS, Publ. n. 143, 1984.

THORNDYCRAFT, V. R.; BENITO, G.; RICO, M.; SOPEÑA, A.; SÁNCHEZ-MOYA, Y.; CASAS, A. *A long-term flood discharge record derived from slackwater flood deposits of the Llobregat River, (NE) Spain*. Journal of Hydrology, n. 313, p. 16-31, 2005.

USACE – UNITED STATES ARMY CORPS OF ENGINEERS. *American river basin, California, Folsom dam and lake - Revised PMF study*. Hydrology Office Report, USACE, Sacramento District, Sacramento, USA, 2001.

USBR – UNITED STATES BUREAU OF RECLAMATION. *Flood hazard analysis, Folsom dam - Central Valley Project*. Flood Hydrology Group, Denver, USA, 2002.

***Method To Estimate Extreme Flood Quantiles With The Joint Use Of A Bayesian Analysis With Non-Systematic Information And Upper Bounded Probability Distributions – Part 2 : Application***

***ABSTRACT***

*This is the second of two companion papers dealing with a method to estimate extreme flood quantiles, through a Bayesian approach and finite upper-tail probability distributions, with an allowance to include historical and paleoflood information. This paper describes three examples of applications to catchments with different sets of available flood data. The three case studies have been selected so as to evaluate the different aspects of the methodological framework. They showed that the proposed approach improved two aspects of extreme flood estimation: the better description of flood probabilities, as exemplified by the good fits of the upper-bounded probability models to systematic and non-systematic data; and the uncertainty reduction associated with flood predictions, as exemplified by the shorter credibility intervals as estimated with non-systematic flood data. In addition, the approach provides a rational framework to include the local Probable Maximum Flood (PMF) estimate in frequency analysis. As a final remark, it is worthwhile mentioning that the expert subjective knowledge, as introduced into the analysis by the appropriate definition of a prior distribution for the upper bound, can potentially improve not only the later results but also the parameters and rare quantile estimates.*

***Key-words:*** *extreme floods, flow frequency analysis , rare hydrologic events.*