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 1: Desenvolvimento Teórico

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 primeiro 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 emprego de distribuições de probabilidades com cauda superior finita. Este artigo aborda o desenvolvimento teórico do método proposto, enquanto o segundo artigo apresenta três aplicações a bacias com diferentes conjuntos de dados disponíveis. Depois de contextualizar o problema de estimação de enchentes extremas, introduz-se aqui a viabilidade de agregação da Cheia Máxima Provável, ou simplesmente PMF (de Probable Maximum Flood), à análise de frequência de variáveis hidrológicas, mediante emprego do paradigma Bayesiano para inferência estatística e de distribuições de probabilidade limitadas superiormente. O limite superior da distribuição em questão está sujeito a incertezas, as quais podem ser sumariadas a partir de distribuições a priori eliciadas, em alguns casos estudados, a partir de um grande banco de estimativas de PMF, conforme aqui demonstrado. Na sequência, o sumário a priori das incertezas pode ser atualizado pela inclusão de informações exatas, obtidas dos registros sistemáticos, e censuradas, de fontes não sistemáticas históricas ou paleohidrológicas, por meio de uma função de verossimilhança suficientemente genérica. A possibilidade do emprego desse grande conjunto de informações em uma estrutura de inferência Bayesiana abre perspectivas para a obtenção de estimativas mais confiáveis de quantis de enchentes extremas.

Palavras Chave: *cheia máxima provável, análise Bayesiana, enchentes extremas, eventos hidrológicos raros.*

INTRODUÇÃO

As estratégias tecnológicas de coexistência com o risco de inundações dependem da quantificação de algumas características das enchentes e da frequência com que são superadas. Embora possa existir uma notável discrepância entre as percepções popular e especializada do que seja uma enchente, os hidrólogos, em geral, a definem como sendo a vazão máxima observada em certa seção fluvial, capaz de extravasar ou não os limites do leito menor, e que pode provocar ou não prejuízos materiais. Em geral, a essa vazão se associa uma probabilidade anual de igualdade ou superação, cujo inverso, expresso em anos, recebe a denominação de tempo de retorno. Em alguns contextos, outras características de uma enchente, tais como sua duração e seu volume, são tão importantes quanto a vazão máxima, e passíveis de serem analisadas sob igual lógica probabilística.

*Universidade Federal de Minas Gerais

Nathan e Weinmann (2001) categorizam as enchentes como grandes (probabilidades de excedência maiores que 10^{-2}), raras (probabilidades de excedência entre 10^{-2} e $2x10^{-3}$) e extremas (probabilidades de excedência menores que $2x10^{-3}$). Em geral, as enchentes consideradas grandes ainda se situam no domínio das medições e observações diretas, enquanto as enchentes raras localizam-se entre estas e o chamado "limite crível de extrapolação" da curva de probabilidades anuais de superação. As enchentes extremas nomeadamente possuem ínfimas probabilidades anuais de superação, bem além do limite crível de extrapolação. Apesar disso, as estimativas das enchentes extremas e quantificação das incertezas associadas a estas estimativas são absolutamente necessárias para a quantificação dos riscos associados ao colapso de estruturas hidráulicas, entre as quais se destacam as grandes barragens.

As incertezas inerentes às estimativas das vazões extremas são sabidamente muito grandes, e não são quantificáveis, de modo preciso, por procedimentos convencionais de inferência estatística. Uma vez que as amostras de vazões máximas anuais, de tamanho usual na faixa de 25 a 80 anos, não oferecem subsídios importantes para a estimação de enchentes muito raras, no intervalo de probabilidades entre $2x10^{-3}$ e 10^{-6}, os engenheiros hidrólogos, em geral, são forçados a lançar mão de expedientes de fixação de um limite superior para as enchentes, com base no entendimento corrente dos processos hidrológicos sob condições extremas. Neste contexto, a cheia máxima provável (PMF, do acrônimo em inglês *Probable Maximum Flood*) é o critério mais utilizado no Brasil e em outros países para o dimensionamento de grandes estruturas hidráulicas (ICOLD, 1987; FEMA, 2004; USNRC, 1977).

De modo sumário, a PMF representa o limite superior da enchente potencial em uma dada seção fluvial, resultante de uma tempestade hipotética, de duração e altura críticas, denominada "Precipitação Máxima Provável" (PMP), antecedida por severíssimas, porém, fisicamente plausíveis, condições hidrológicas e hidrometeorológicas. A PMP, por sua vez, é definida pela Organização Meteorológica Mundial (WMO, 1986) como a maior altura de chuva, para uma dada duração, cuja ocorrência sobre certa área, em certa região geográfica, em uma determinada época do ano, seja meteorologicamente possível.

Usualmente, a transformação da PMP em PMF é realizada por meio de modelos de simulação hidrológica, cujos parâmetros devem ser previamente calibrados para a bacia em estudo. Nestes modelos, as variáveis de estado devem refletir as condições críticas que antecedem um evento que, hipoteticamente, representa o limite superior da enchente potencial local. De fato, a suposições implícitas no conceito de PMF são a existência de limites físicos para o suprimento de umidade atmosférica e precipitação sobre uma bacia, assim como para a resposta hidrológica a este evento. No entanto, a determinação da PMF não é inequívoca e depende fortemente da quantidade e da qualidade de observações históricas de variáveis hidrometeorológicas. Esse fato faz com que a PMF seja suscetível às incertezas impostas pela disponibilidade de dados e pelas ferramentas de modelagem, podendo, assim, ser vista como uma variável aleatória. Com base nestes argumentos Dawdy and Lettenmaier (1987) consideram a PMF "quase-determinística" e afirmam que, sob o ponto de vista prático, a sua probabilidade de excedência, embora pequena, não é zero.

Embora haja uma larga aceitação de que cheias "quase-determinísticas", tal como a PMF, provêem um dimensionamento seguro de grandes estruturas hidráulicas, há a necessidade de se estimar a probabilidade de excedência destas cheias extremas, no sentido de incorporar este critério aos estudos de quantificação de risco (STEDINGER et al., 1989; DAWDY e LETTENMAIER, 1987; DUBLER e GRIGG, 1996; USBR, 2004). Esta situação paradoxal tem revelado as dificuldades de se reconciliar dois conceitos aparentemente opostos em hidrologia de enchentes (IACWD, 1985). Neste sentido, conforme mencionado por Nathan and Weinmann (2001), as soluções adotadas têm sido pouco rigorosas, como, por exemplo, arbitrar à PMP, e indiretamente à PMF, probabilidades de excedência na faixa de 10^{-4} a 10^{-7}.

Por outro lado, a despeito de ser uma idéia aparentemente oposta à análise de frequência de cheias, a PMF, quando assimilada em uma estrutura lógica, pode ser usada como um valor de referência para a estimação de cheias extremas e suas respectivas probabilidades de excedência. De acordo com Klemeš (1993), "*… a PMP e a PMF podem ser vistas como tentativas racionais de síntese de eventos muito raros a partir de seus componentes formadores conhecidos, tais como documentados nos registros históricos, e que, quando combinados da forma mais crítica possível, embora muito pouco provável, produzem estimativas consistentes com o entendimento corrente sobre o clima, a meteorologia e a hidrologia de eventos extremos*". No entanto, de forma a levar em conta a PMF como uma síntese das cheias extremas em uma dada bacia, algumas mudanças conceituais são necessárias no modo como a análise de frequência é usualmente feita.

Sob o ponto de vista da incorporação do conceito de PMF aos modelos probabilísticos, a análise convencional de frequência de cheias apresenta três grandes obstáculos. Primeiramente, as amostras relativamente curtas de registros fluviométricos, via de regra, não excedem a uma centena de anos e não fornecem informações substanciais sobre eventos extremos. Tais fatos tornam forçosa a extrapolação das curvas de frequência até quantis extremamente raros, muito além do limite crível de extrapolação, com todas as incertezas associadas à escolha do modelo distributivo e às estimativas de seus parâmetros. O segundo obstáculo refere-se ao fato de que todas as distribuições de probabilidades, usuais na análise de frequências de eventos hidrológicos extremos, são, por construção, ilimitadas superiormente e, portanto, não acomodam a idéia intrínseca ao conceito da PMF. Finalmente, o terceiro obstáculo relaciona-se à deficiência das metodologias existentes em avaliar e assimilar, em um contexto lógico, as incertezas associadas às estimativas de PMF. A estratégia aqui proposta para contornar estes problemas será descrita nas seções que se seguem.

ESBOÇO METODOLÓGICO

Nas últimas décadas, como uma tentativa de superar o obstáculo da escassez de dados sobre enchentes extremas nos registros fluviométricos, tem-se procurado incluir informações auxiliares, denominadas não sistemáticas, na análise de frequência de cheias (STEDINGER e COHN, 1986; FRANCÉS et al., 1994; FRANCÉS, 2001; NAULET, 2002). Existem duas fontes de informações não sistemáticas: as históricas, referentes a eventos de cheia diretamente observados e, de algum modo, registrados por seres humanos; e as reconstruções das denominadas "paleocheias" (BAKER, 1987; BAKER et al., 2002), as quais correspondem às enchentes que atingiram, em algum momento do período pré-histórico, contido na época holocênica, níveis d'água extremos, os quais podem ser reconstituídos a partir de evidências físicas, geológicas e/ou botânicas ainda existentes.

As informações históricas e paleohidrológicas relativas às enchentes são, em geral, do tipo censurada, podendo ser limitadas superiormente (UB), inferiormente (LB) ou em ambos os sentidos mencionados (DB). Estes níveis de censura ou de percepção estão associados aos limites atingidos por uma particular cheia, em um determinado período de tempo. Em todas as situações, o valor exato do nível de água (ou o valor exato da descarga máxima) alcançado durante uma cheia não é conhecido. Para informações do tipo UB, sabe-se que nenhuma vazão superou um determinado nível de percepção durante o intervalo de tempo considerado. Para informações do tipo LB, sabe-se que as vazões superaram um determinado nível. Finalmente, para informações do tipo DB, sabe-se que as vazões ficaram compreendidas dentro de um determinado intervalo dado por (LR; UR). A partir de algumas hipóteses, os dados não sistemáticos de vazões podem ser avaliados por meio de métodos estatísticos adequados e, então, incorporados à análise de frequência de cheias, possuindo, assim, o potencial de aumentar a confiabilidade das estimativas de quantis extremos de vazão (FRANCÉS, 2001).

A viabilidade de usarem-se dados paleohidrológicos, abrangendo centenas ou milhares de anos, para melhorar as estimativas de cheias extremas tem sido objeto de debates, os quais contrastam a variabilidade natural do clima em grandes escalas temporais e as premissas de estacionariedade, independência e distributivas das variáveis aleatórias empregadas na análise convencional de frequência de cheias (NRC, 1999; BAKER, 2003). Reconhecendo, por um lado, a complexidade inerente a esta questão, mas verificando, por outro, a frequente situação de completa ausência de informação sistemática sobre cheias extremas, compartilha-se, aqui, do ponto de vista, segundo o qual, "... *parece mais pragmático não desconsiderar ou ignorar os dados paleohidrológicos, mas, ao contrário, pelo menos avaliar o que tais dados têm a dizer sobre o fenômeno de cheias extremas*", tal como expressado por Baker (2003). Neste mesmo sentido, Hirschboeck (2003) argumenta que os registros de paleocheias fornecem um dos melhores indicadores de eventos extremos históricos e que as marcas de cheias preservam e registram somente os eventos mais extremos ocorridos na bacia, os quais são, precisamente, as informações faltantes nas pequenas amostras de dados sistematicamente observados.

A existência de um limite superior para as magnitudes de cheias e a capacidade de determiná-lo em uma dada bacia são fontes de uma longa controvérsia na literatura sobre a hidrologia de enchentes (HORTON, 1936; YEVJEVICH, 1968; KLEMES, 1987; YEVJEVICH e HARMANCIOGLU, 1987). Essas controvérsias têm dividido a análise de cheias em duas linhas: a estocástica, na qual uma distribuição de probabilidades ilimitada superiormente é ajustada aos dados, admitindo-se implicitamente a hipótese de que qualquer vazão futura, por maior que seja, tem probabilidade de ocorrência diferente de zero; e a determinística, a qual é bem exemplificada pela análise hidrológica, com base na PMP/PMF, já mencionada.

A completa predominância de distribuições de probabilidades ilimitadas superiormente na análise de frequência convencional deve-se principalmente às dificuldades em se estimar o limite superior, com base em poucos dados observados. A este respeito, Hosking e Wallis (1997) argumentam que é irrelevante considerar fisicamente impossível a ocorrência de uma cheia, com probabilidade de excedência de 10^{-5}, se o objetivo principal da análise de frequência é estimar o quantil com probabilidade de 10^{-2}. Entretanto, quando as enchentes extremas são o objeto de estudo, afirmar que as vazões em uma seção fluvial podem ou não crescer até o infinito, ou afirmar que o modelo probabilístico para as vazões possui cauda pesada, leve ou limitada superiormente, tal como estudado por Bernardara et al. (2008), tem um profundo impacto na análise de frequência. Neste contexto, Enzel et al. (1993) avaliaram um total de 32.120 dados de vazão máxima anual na bacia do rio Colorado, nos Estados Unidos, e encontraram evidências da existência de limites superiores que não foram ultrapassados nos últimos milênios. Esta e outra recente pesquisa (Jacoby et al., 2008), apesar de possivelmente não erradicarem a controvérsia sobre o assunto, oferecem argumentos empíricos em favor da existência de um limite superior para as vazões.

A ideia da utilização de informações sistemáticas e não sistemáticas, conjuntamente com distribuições de probabilidades limitadas superiormente, para melhorar a estimação de cheias extremas foi introduzida por Francés e Botero (2003). As informações exata e censurada sobre as cheias do rio Júcar, na Espanha, foram reunidas por meio de uma função de verossimilhança, a qual foi a base para a inferência dos parâmetros das seguintes distribuições de probabilidades limitadas superiormente: (i) o modelo EV4, introduzindo por Kanda (1981), para descrever a frequência de eventos sísmicos extremos no Japão; (ii) o modelo TDF, proposto por Elíasson (1997), para estudar precipitações extremas na Islândia; e (iii) o modelo LN4, empregado por Takara e Loebis (1996), para modelar chuvas extremas no Japão e Indonésia. Botero (2006) aplicou com sucesso esta idéia a uma grande quantidade de bacias espanholas, as quais possuíam diferentes coleções de informações sistemáticas e não sistemáticas sobre cheias. Botero (2006), entretanto, não buscou o objetivo paralelo de atribuir uma probabilidade de excedência à PMF local, ou mesmo revisar sua estimativa, porque esta permanecia, na formulação de seu modelo, como uma quantidade definitivamente determinística.

Este artigo descreve um procedimento para incluir o conceito da PMF na análise de frequência, mediante o uso conjunto da distribuição de probabilidades limitada superiormente LN4 e de informações sistemáticas e não sistemáticas sobre as cheias, preservando algumas características do método proposto por Francés e Botero (2003). A análise é feita, entretanto, segundo o paradigma Bayesiano, o qual oferece a estrutura lógica apropriada para incluir as incertezas referentes ao limite superior das vazões. Com efeito, a informação relacionada à PMF, é incluída no modelo por meio de uma distribuição *a priori* para o limite superior da LN4, o que permite uma melhor descrição das incertezas relacionadas aos eventos extremos.

As seções que se seguem são referentes (i) à completa descrição do modelo estatístico, (ii) à construção de uma possível distribuição *a priori* do limite superior, a partir de um grande conjunto de estimativas de PMF's compiladas para diferentes bacias dos Estados Unidos, e (iii) às conclusões preliminares do presente estudo. Em artigo complementar, neste mesmo número da Revista Brasileira de Recursos Hídricos, encontram-se descritas as aplicações do método aqui apresentado, a três estudos de caso, com diferentes conjuntos de informações disponíveis.

MODELO ESTATÍSTICO

Análise Bayesiana

Do ponto de vista Bayesiano, uma grandeza aleatória é uma quantidade desconhecida que pode variar e assumir diferentes valores (tal como uma variável aleatória, por exemplo) ou simplesmente ser uma quantidade fixa sobre a qual há alguma ou nenhuma informação disponível (tal como um parâmetro, por exemplo). As incertezas sobre estas grandezas aleatórias são descritas por distribuições de probabilidades, as quais refletem a maneira subjetiva de como o especialista ou a pessoa que está analisando o problema avalia a chance de ocorrência de um determinado evento.

Além da informação obtida a partir dos dados observados, que também é considerada pela Escola Clássica, a Escola Bayesiana considera outras fontes de informação para resolver os problemas de inferência. A Escola Bayesiana segue a postura subjetivista, a qual estabelece que medidas de probabilidade descrevem a incerteza (pessoal) sobre objetos que são desconhecidos (variáveis aleatórias ou não). Assim como em estatística clássica, o parâmetro populacional é uma quantidade fixa, e não é uma variável aleatória. No entanto, como é desconhecido, em estatística Bayesiana a incerteza que se tem sobre ele deve ser descrita por uma distribuição de probabilidades. Em termos formais, seja θ um parâmetro de interesse e que pode assumir diferentes valores no espaço paramétrico Θ. Denote-se por $\mathcal{H}$ a informação prévia que um especialista detém sobre θ. Baseado em $\mathcal{H}$, a incerteza sobre θ é resumida pela distribuição a priori $\pi(\theta \mid \mathcal{H})$ que descreve o estado de conhecimento sobre a quantidade aleatória, antes que qualquer dado tenha sido observado. Se Θ é um conjunto finito, esta inferência a priori representa a chance de ocorrência atribuída a cada valor de θ, antes de agregar-se a informação contida na amostra.

Em geral, $\mathcal{H}$ não contém toda a informação relevante sobre o parâmetro, e neste caso, a distribuição *a priori* não é uma boa inferência para θ, a menos, obviamente, que o analista tenha total conhecimento sobre o parâmetro e, em assim sendo, não há razão para se fazer inferência, o que, de fato, não ocorre em situações reais. Se a informação contida em $\mathcal{H}$ não é suficiente, um experimento deve se realizado para obter informações adicionais sobre o referido parâmetro. Suponha-se que uma variável aleatória *X*, a qual é relacionada a θ, possa ser amostrada ou observada. Antes de se observar *X* e supondo-se que o atual valor de θ seja conhecido, a incerteza sobre a quantidade *X* pode ser resumida pela função de verossimilhança $f(X \mid \theta, \mathcal{H})$. Esta função fornece a probabilidade de ocorrência de cada amostra particular *x* de *X*, caso θ seja o valor verdadeiro do parâmetro e o espaço amostral seja discreto. Após realizar-se o experimento, o conhecimento *a priori* sobre θ deve ser atualizado, usando-se a nova informação *x*. O contexto formal para atualizar a distribuição *a priori* é o teorema de Bayes. De acordo com este teorema, a distribuição *a posteriori*, que agrega o conhecimento atualizado sobre θ, é dada por:

$$\pi(\theta \mid x, \mathcal{H}) = \frac{f(x \mid \theta, \mathcal{H})\pi(\theta \mid \mathcal{H})}{f(x \mid \mathcal{H})}, \tag{1}$$

onde a distribuição preditiva *a priori* $f(x \mid \mathcal{H})$ é calculada usando-se a seguinte expressão:

$$f(x \mid \mathcal{H}) = \int_{\Theta} f(X \mid \theta, \mathcal{H})\pi(\theta \mid \mathcal{H})d\theta . \tag{2}$$

A distribuição *a posteriori*, dada pela equação (1), descreve a incerteza sobre θ depois de observarem-se os dados, ou seja, $\pi(\theta \mid X, \mathcal{H})$ é a inferência *a posteriori* sobre θ, a partir da qual é possível estabelecer-se qualquer característica relativa a θ.

No que refere-se à análise de eventos extremos em engenharia e em outras disciplinas, Coles e Powell (1996) enfatizam que o objetivo principal da inferência estatística é, na verdade, a predição de valores futuros da variável em análise, sublinhando que a abordagem Bayesiana oferece a solução mais coerente para tal. Ainda neste contexto, McRobbie (2004) afirma que a análise Bayesiana é, sob o ponto de vista da predição, uma abordagem mais racional e prudente e, assim, não parece fazer sentido perguntar se a abordagem oferece estimativas mais precisas que a frequentista. Em outras palavras, a análise de eventos extremos em engenharia é um problema de predição e não de estimação paramétrica. McRobbie (2004) prossegue argumentando que antecipar ou avaliar a probabilidade de ocorrência de eventos futuros, cujas magnitudes estão muito além dos dados observados, é essencialmente uma tarefa subjetiva. Assim, o "mito da objetividade", apontado como uma deficiência da abordagem Bayesiana, embora também exista na abordagem frequentista, não faz sentido em problemas correntes de engenharia.

No caso de cheias extremas, engenheiros e hidrólogos necessitam tomar decisões que envolvem a avaliação de eventos extraordinários, tais como a cheia de 10.000 anos de período de retorno, com base em informações incompletas e abstratas. Assim, há uma subjetividade intrínseca ao evento que, diferentemente da Escola Clássica de estatística, pode ser coerentemente analisada em uma abordagem Bayesiana. Na Escola Bayesiana, a subjetividade é avaliada a partir do conhecimento do especialista sobre as características probabilísticas da natureza do evento. Sob esta ótica, a correta descrição da subjetividade, inerente ao processo natural de ocorrência de um evento extremo, depende da habilidade do especialista em selecionar, criticar, interpretar e julgar o conjunto de informações disponíveis (VICK, 2002).

De forma a atender aos objetivos a serem aqui alcançados, algumas etapas sequenciais são necessárias. Primeiramente, admite-se que os picos de cheia possam ser modelados por uma distribuição de probabilidades limitada superiormente e que a PMF local possa ser utilizada para estimar o parâmetro que identifica o limite superior da referida distribuição. Em segundo lugar, apresenta-se a etapa crucial de eliciação de uma distribuição *a priori* genuína para o limite superior. Esta distribuição *a priori* pode ser obtida mediante a quantificação das incertezas sobre as estimativas regionais de PMF. A terceira etapa é a construção da função de verossimilhança, com base nas informações exata e censurada, a partir dos dados sistemáticos e não sistemáticos disponíveis sobre as cheias locais. Finalmente, todos os termos da equação 1 podem ser então reunidos e resolvidos de modo a obter a distribuição *a posteriori* dos parâmetros. Estas etapas serão detalhadas nas subseções que se seguem.

A distribuição de probabilidades LN4 de variável aleatória limitada superiormente

Entre as distribuições de probabilidades que são frequentemente utilizadas pelos hidrólogos, para modelar variáveis relacionadas a cheias, algumas podem apresentar limite superior, a depender de particulares combinações dos valores numéricos de seus parâmetros. Entre estas, estão distribuições de 3, 4 e 5 parâmetros. Por exemplo, a distribuição

Generalizada de Valores Extremos (GEV) será limitada superiormente se o seu parâmetro de forma for positivo, o que ocorre quando o coeficiente de assimetria for menor que 1,1396. A distribuição LogPearson tipo III (LP3) também possui limite superior se o valor de seus parâmetros resultar em um coeficiente de assimetria negativo. Tais propriedades, próprias dessas distribuições, restringem seu emprego como modelo geral para cheias, aqui tomadas como variáveis aleatórias com limite superior. Com efeito, é consenso entre os hidrólogos que as distribuições de probabilidades, consideradas apropriadas ao estudo de cheias, são assimétricas à direita (assimetria positiva) e, na maioria dos casos, com coeficiente de assimetria superior a 1,1396.

Outros modelos que devem ser mencionados são as distribuições de probabilidades Kapa (HOSKING e WALLIS, 1997) e Wakeby (HOUGHTON, 1978). O modelo Kapa de 4 parâmetros pode ter limites superior e/ou inferior para valores particulares de seus dois parâmetros de forma. Analogamente, o modelo Wakeby de 5 parâmetros também pode ter limite superior para uma particular combinação de seus três parâmetros de forma. Estas são formas distributivas gerais que podem acomodar uma série de outros modelos e que incluem aqueles com limites e ilimitados. Para a modelagem proposta neste artigo, é necessário que o limite superior tenha uma forma paramétrica explícita, o que impede o uso de modelos generalizados para a distribuição dos máximos anuais de vazão. As distribuições TDF, EV4 e LN4 têm a referida característica e, assim, são modelos candidatos para a abordagem Bayesiana da análise de frequência de cheias extremas.

Elíasson (1997) propôs a distribuição TDF para modelar as alturas máximas anuais de precipitação de 24 horas de duração na Islândia e no Estado americano de Washington. O modelo TDF, para variáveis limitadas superiormente, é derivado da transformação de uma variável ilimitada de Gumbel e sua forma analítica é descrita por um parâmetro de escala, um de forma e dois de posição, a saber, os limites superior e inferior. Kanda (1981) combinou as expressões analíticas das distribuições de valores extremos do tipo II e III, e encontrou uma nova forma paramétrica com limites superior e inferior, a qual denominou distribuição de valores extremos do tipo IV, ou simplesmente EV4, usando-a para modelar a velocidade do vento e a intensidade de terremotos no Japão. A distribuição Lognormal de 4 parâmetros, com limites superior e inferior, foi proposta por Slade (1936) como a forma limite de um modelo descrito em termos de seu desvio padrão e pela flutuação máxima e mínima em torno do valor médio. Takara e Loebis (1996) referem-se a este modelo como LN4, o qual foi usado como uma possível alternativa para a análise de frequência de chuvas extremas da Indonésia e Japão. Mais tarde, Takara e Tosa (1999) utilizaram as distribuições EV4 e LN4 para modelar eventos de cheias e precipitação do Japão, fixando os respectivos limites superiores nas correspondentes estimativas de PMP ou PMF.

Conforme mostrado por Fernandes et al. (2010), o modelo EV4 se aproxima do limite superior mais rapidamente do que se observa nos modelos TDF e LN4. Este comportamento também foi relatado por Botero (2006). Dependendo das características dos dados de cheias, tais como a assimetria e da presença de *outliers* na amostra, esse comportamento pode reduzir a capacidade descritiva quando comparada aos outros dois modelos. Esta e outras propriedades matemáticas das distribuições LN4, EV4 e TDF devem ser objetos de pesquisas futuras, que fornecerão um melhor guia na escolha de modelos probabilísticos para variáveis limitadas superiormente.

De modo análogo aos procedimentos da análise de frequência convencional, selecionar uma distribuição de probabilidade limitada superiormente, de um conjunto de modelos candidatos, depende fortemente dos dados disponíveis. Fernandes et al. (2010), aplicou a metodologia aqui descrita aos dados do rio American, localizado nos Estados Unidos, e concluíram que o modelo LN4 apresentou os melhores resultados. No artigo complementar a este, serão descritas outras aplicações às bacias do rio Llobregat e do rio Pará, localizadas, respectivamente, na Espanha e no Brasil. Com o objetivo de escolher somente um dentre os três modelos paramétricos candidatos, foi efetuada uma avaliação visual do ajuste dos quantis observados e estimados. Em todas as aplicações estudadas, o modelo LN4 apresentou os melhores resultados, seguido pela TDF e pela EV4, em ordem decrescente de desempenho.

Estes argumentos, associados ao objetivo principal deste artigo, que é o de descrever o método Bayesiano para a análise de frequência, e ao da construção de uma distribuição *a priori* informativa para o limite superior das vazões, são as principais justificativas para o uso do modelo LN4. Ressalva-se, entretanto, que em outras aplicações, caso uma diferente distribuição limitada superiormente se mostrar mais adequada, o procedimento aqui descrito pode ser facilmente adaptado ao modelo escolhido.

O modelo LN4 é construído a partir da seguinte transformação:

$$Y = \ln\left(\frac{X-\varepsilon}{\alpha - X}\right) \tag{3}$$

onde $\varepsilon \in \Re_+$ denota o limite inferior de *X*, $\alpha \in \Re_+$ representa o limite superior de *X* e $Y \sim \text{Normal}(\mu_Y; \sigma_Y)$.

Takara e Tosa (1999) mostram que a LN4 não é fortemente afetada ao admitir-se que o limite inferior é igual a zero. Assim, tem-se um modelo mais parcimonioso, com um parâmetro a menos para ser estimado, e uma distribuição *a priori* a menos a ser eliciada.

Admitindo-se que o limite inferior seja zero, a função densidade de probabilidade de uma variável aleatória LN4, denotada por $X \sim \text{LN4}(\mu_Y; \sigma_Y; \alpha)$, com parâmetro de posição $\mu_Y \in \Re$, de escala $\sigma_Y \in \Re_+^*$ e limite superior $\alpha \in \Re_+$, é dada por:

$$f_X(x \mid \mathbf{\Theta}) = \frac{\alpha}{x(\alpha - x)\sigma_Y\sqrt{2\pi}} \times \exp\left\{-\frac{1}{2\sigma_Y^2}\left[\ln\left(\frac{x}{\alpha - x}\right) - \mu_Y\right]^2\right\} \tag{4}$$

com $0 \le x < \alpha$. A correspondente função acumulada de probabilidades é dada por:

$$F_X(x \mid \mathbf{\Theta}) = \Phi\left[\frac{1}{\sigma_Y}\ln\left(\frac{x}{\alpha - x}\right) - \frac{\mu_Y}{\sigma_Y}\right] \tag{5}$$

onde Φ denota a distribuição normal padrão acumulada.

A função de verossimilhança considerando dados sistemáticos e não sistemáticos de cheias

Durante o período sistemático de observações os picos anuais de cheia são tidos como exatos, por suposição, ou seja, não estão sujeitos a erros decorrentes de curva-chave mal definida ou de a qualquer outro de natureza parecida; tais dados serão denotados por EX no restante deste artigo. As informações sobre cheias, sistemáticas e não sistemáticas, podem conter dados censurados, os quais incluem registros limitados superiormente, inferiormente ou duplamente limitados, de acordo com os níveis atingidos por uma determinada cheia.

De acordo com a estrutura conceitual introduzida por Stedinger e Cohn (1986), a função de verossimilhança para dados exatos e censurados, compreendendo informações sistemáticas e não sistemáticas, pode ser construída da forma descrita a seguir. Sejam $X_1, \cdots, X_{N_{EX}}$ as cheias máximas anuais registradas em um período sistemático de N_{EX} anos. As informações censuradas são denotadas por $UB_1, \cdots, UB_{N_{UB}}$, as quais representam as cheias limitadas superiormente, $LB_1, \cdots, LB_{N_{LB}}$, que são as cheias limitadas inferiormente, e $DB_1, \cdots, DB_{N_{DB}}$, que referem-se às cheias duplamente limitadas, contidas em um intervalo (UR; LR). Admitindo-se que os dados referidos sejam estacionários e identicamente distribuídos, a função de verossimilhança correspondente pode ser escrita como:

$$\begin{aligned} f(x \mid \mathbf{\Theta}) = & \prod_{i=1}^{N_{EX}} f_X(EX_i \mid \mathbf{\Theta}) \times \\ & \prod_{i=1}^{N_{UB}} F_X(UB_i \mid \mathbf{\Theta}) \times \\ & \prod_{i=1}^{N_{LB}} [1 - F_X(LB_i \mid \mathbf{\Theta})] \times \\ & \prod_{i=1}^{N_{DB}} [F_X(UR_i \mid \mathbf{\Theta}) - F_X(LR_i \mid \mathbf{\Theta})] \end{aligned} \tag{6}$$

onde f_X e F_X são, respectivamente, a função densidade e acumulada da variável aleatória X e $\mathbf{\Theta}$ denota o vetor de parâmetros. Se $X \sim \text{LN4}(\mu_Y; \sigma_Y; \alpha)$, então f_X e F_X são dadas pelas equações (4) e (5) e $\mathbf{\Theta} = (\mu_Y; \sigma_Y; \alpha)$.

A equação (6) e todo o processo de inferência, compreendido pelas equações (1) e (2), podem ser modificados para acomodar cada tipo de informação sobre cheias, incluindo períodos de não estacionariedades. Entretanto, conforme observado por Hirschboeck (2003), avaliações deste tipo requerem a identificação de uma distribuição mista para as cheias, capaz de incluir a variação temporal da circulação atmosférica e das respostas hidrológicas. Tal distribuição é de construção bastante complexa e representa um desafio para futuras pesquisas.

Para completar a especificação do modelo dado pela equação (1) é necessário especificar as distribuições *a priori* para o vetor de parâmetros $\mathbf{\Theta} = (\mu_Y; \sigma_Y; \alpha)$. Admite-se aqui que α, μ_Y e σ_Y são independentes, com as seguintes distribuições *a priori* marginais: $\sigma_Y \sim \text{Gama}(\rho_\sigma;\beta_\sigma)$, $\alpha \sim \text{Gama}(\rho_\alpha;\beta_\alpha)$ e $\mu_Y \sim \text{Normal}\left(\mu_\mu; 1/\sigma_\mu\right)$, onde μ_μ, σ_μ, ρ_σ e β_σ denotam os respectivos hiperparâmetros. Na subseção a seguir, apresentam-se alguns argumentos em favor da escolha dos aludidos modelos para as distribuições *a priori* dos parâmetros da LN4.

Distribuições *a priori* para os parâmetros da LN4

A distribuição *a priori* é, por definição, a síntese matemática do grau de conhecimento, ou de certeza, que um certo especialista tem a respeito da grandeza de interesse. A especificação desta distribuição é feita com base em todo o conhecimento adquirido pelo especialista em toda a sua vida, através de observações, experiências relacionadas à situações similares, pesquisas feitas por outros pesquisadores sobre o tema, etc. A distribuição *a priori* é construída antes que os dados amostrais sejam observados. Entre os parâmetros da LN4, o limite superior α é o único que, sob condições hidrometeorológicas extremas, mostra uma clara conexão com as características climáticas e hidrológicas das bacias hidrográficas. A interpretação dos parâmetros μ_Y e σ_Y – respectivamente, média e desvio padrão da variável transformada LN4 – em termos hidrológicos ou hidrometeorológicos, é muito mais complicada. Na verdade, ao se avaliar a distribuição LN4, é difícil encontrar alguma relação evidente entre os parâmetros μ_Y e σ_Y, e as características físicas das bacias em estudo. Assim, construir uma distribuição *a priori* informativa para os parâmetros μ_Y e σ_Y é uma tarefa, no mínimo, complexa. Portanto, uma vez que o parâmetro μ_Y pode assumir qualquer valor real e o parâmetro σ_Y é sempre positivo, admitiu-se uma distribuição Normal, pouco informativa, para o primeiro e uma distribuição Gama, pouco informativa, para o segundo, ou seja, consideramos $\mu_Y \sim \text{Normal}\left(1{,}0; 10^{-6}\right)$ e $\sigma_Y \sim \text{Gama}\left(1{,}0; 10^{-8}\right)$. Salienta-se que tais distribuições *a priori* são próprias, garantindo que as distribuições *a posteriori* sejam próprias, e se aproximam, em termos de informação sobre o respectivo parâmetro trazida para o problema de inferência, das distribuições *a priori* não informativas no sentido Bayes-Laplace (BERGER, 1985). Para estes parâmetros as distribuições não informativas no sentido Bayes-Laplace são impróprias, pois são distribuições uniformes em $\Re$ e $\Re^*$ respectivamente.

No caso das vazões máximas de cheia, o conhecimento *a priori* sobre o seu limite decorre, por exemplo, da análise de eventos extremos em bacias similares, das características geomorfológicas do local de interesse e das características hidráulicas do trecho fluvial em estudo.

As características geomorfológicas, quando vistas sob o ponto de vista do passado geológico da bacia, poderiam fornecer indícios de cheias extremas (paleocheias), nunca superadas em um longo intervalo de tempo, possibilitando, assim, alguma caracterização das incertezas relacionadas ao desconhecimento do verdadeiro valor do limite superior. Por outro lado, tais vazões fazem parte da função de verossimilhança (vide equação 6) e seu uso na construção da distribuição *a priori* para o limite superior acarretaria em uma indevida duplicidade de informações empregadas no modelo estatístico.

As características hidráulicas do trecho fluvial, sobretudo a forma da sua seção fluvial, também poderiam prover informações sobre a cheia máxima fisicamente possível naquele trecho. Estas informações poderiam provir de simulações hidráulicas para vários níveis de água na planície de inundação. No entanto, a prescrição de um limite superior para as vazões por meio de simulações hidráulicas é tarefa muito complexa e, na maioria dos casos, impossível de ser realizada. Com efeito, a despeito da complexidade do modelo hidráulico em si, do ponto de vista geométrico (forma da seção transversal) apenas, a vazão pode ser tão grande quanto se queira, já que a área da seção transversal sempre aumenta com o nível da água, ou seja, não poderia haver limite superior. Obviamente, esta possibilidade é fisicamente um absurdo, uma vez que imaginar uma seção com profundidade tão grande quanto se queria está bastante longe da realidade. Assim, a prescrição de um limite superior para as vazões, por meio de simulação hidráulica, passa pela definição do limite máximo da altura da lâmina d'água na seção transversal. Esta definição, em última análise, conduziria a um problema tão ou mais complexo quanto o problema inicial.

Deste modo, resta a definição da distribuição *a priori* por meio da análise de eventos extremos, observados ou estimados, em bacias similares à bacia de interesse. Uma estimativa pontual razoável para o limite superior da LN4 é o atual valor da PMF para a bacia de interesse. Entretanto, não se pode afirmar categoricamente que este limite superior é a PMF e as incertezas sobre esta grandeza devem ser cuidadosamente sintetizadas por uma distribuição *a priori*. Se dispuséssemos de um grande banco de dados de valores de PMF para uma bacia em particular, estimadas com métodos idênticos, em diferentes datas, usando em cada estimativa todo o conjunto de dados hidrológicos e hidroclimatológicos disponível, então, provavelmente, teríamos meios de eliciar uma distribuição *a priori* altamente informativa para o limite superior. Obviamente, esta situação hipotética não existe, o que força o uso de outras abordagens, baseadas nas informações disponíveis.

A fim de eliciar a distribuição *a priori* do limite superior, podem ser empregados alguns procedimentos que se apoiam nos bancos de dados de estimativas de PMF. Como exemplos de tais procedimentos, descrevem-se, a seguir, aqueles dois preconizados por Fernandes et al. (2010), a partir de um grande conjunto de estimativas de PMF para distintas bacias norte-americanas. Outros procedimentos, semelhantes aos que aqui se descrevem, porém adequados aos conjuntos de informações disponíveis em outros países, como a Espanha e o Brasil, serão objeto de descrição pormenorizada em artigo complementar, publicado neste mesmo número da Revista Brasileira de Recursos Hídricos.

Procedimento A

A *United States Nuclear Regulatory Commision* compilou um conjunto de 561 estimativas de PMF (USNRC, 1977), válidas para 561 bacias hidrográficas de variadas dimensões e características, disseminadas por todo o território dos Estados Unidos da América. De um modo que guarda alguma analogia ao usado por Vogel et al. (2007), no contexto de atribuição de probabilidades a curvas envoltórias de cheias, o banco da USNRC de estimativas de PMF pode também ser usado para a construção da distribuição *a priori* do limite superior de cheias em uma dada bacia.

Segundo o procedimento A, as 561 estimativas de PMF foram agrupadas em 10 categorias, em conformidade às magnitudes da áreas de drenagem das bacias a que se referiam. Na sequência, para cada categoria, foi elaborado o histograma de frequências para as estimativas de PMF, dentro daquela particular faixa de variação das áreas de drenagem. Os histogramas das 10 categorias apresentam-se todos assimétricos à direita e se ajustaram bem a densidades Gama de 2 parâmetros. A figura 1 ilustra os ajustamentos à distribuição Gama para duas das dez categorias estudadas.

Esta simples análise indica que a distribuição Gama é uma forma paramétrica, pelo menos razoável, no que concerne à modelagem da repartição de frequências das estimativas de PMF, para um conjunto de bacias de áreas de drenagem semelhantes. Certamente, a distribuição *a priori* do parâmetro referente ao limite superior da distribuição LN4 não será idêntica à distribuição das estimativas de PMF de uma dada categoria de áreas de drenagem. Entretanto, a imaginar-se que uma certa categoria, limitada por duas áreas de drenagem, as quais são permitidas aproximarem-se uma da outra a ponto de confundirem-se, espera-se que o histograma correspondente seja também assimétrico à direita e uma razoável síntese da repartição de frequências das estimativas de PMF de bacias com igual área de drenagem, embora situadas em distintas regiões geográficas. Se ainda for feita a suposição adicional de que as referidas bacias situam-se em regiões similares, dos pontos de vista climático e hidrológico, seria legítimo postular que o histograma correspondente teria forma semelhante à anterior, provavelmente com variância muito inferior.

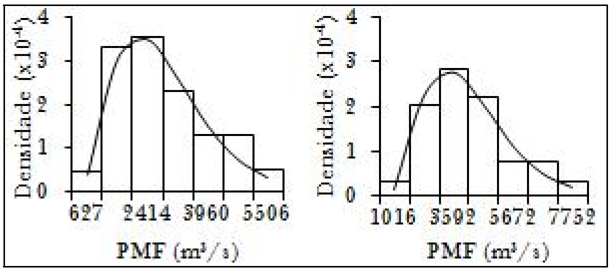

Figura 1 – Histogramas escalonados e densidades Gama das estimativas de PMF, para duas categorias de bacias

Com base nos anteriores argumentos, e admitindo-se a PMF local como um estimador do parâmetro *a*, parece plausível admitir a distribuição Gama, de assimetria positiva, como uma forma paramétrica razoável para a modelagem da distribuição *a priori* do limite superior do modelo LN4. Em uma primeira impressão, pode parecer contraditório empregar uma distribuição ilimitada à direita, como a Gama, como um modelo da distribuição *a priori* da limite superior da LN4. Pelo contrário, uma distribuição genérica ilimitada à direita, como a Gama, conforma-se melhor ao objetivo de sintetizar as incertezas sobre o limite superior da LN4, antes mesmo de levar em conta os dados observados, incluindo as chances de haver incrementos de qualquer magnitude, relativamente a um certo valor do limite superior. Depois de incluídas as observações no processo de inferência, o anterior conhecimento a respeito da variação do limite superior da LN4 pode ser atualizado pela distribuição *a posteriori*, com prováveis alterações de escala e forma, de modo a traduzir as incertezas quanto à fixação da máxima descarga possível em uma dada bacia.

Pode-se caracterizar completamente uma distribuição Gama mediante o conhecimento de dois de seus pontos: aqui consideraremos uma estimativa do coeficiente de variação do limite superior α e uma afirmação conjectural a respeito de probabilidade de não-excedência de uma estimativa conhecida de α. Com efeito, para uma distribuição Gama, com parâmetros ρ_α e β_α, a combinação das duas equações do método dos momentos resulta em $\hat{\rho}_\alpha = CV_\alpha^{-2}$. O hiperparâmetro β_α pode ser estimado associando-se uma probabilidade de não-excedência *p* à estimativa corrente da PMF local ou, em outros termos, β_α será tal que satisfaça a equação $P(\alpha \le PMF \mid \rho_\alpha, \beta_\alpha) = p$.

A distribuição *a priori* de α deve refletir o estado de conhecimento a respeito da máxima descarga possível em uma certa seção fluvial. Admite-se aqui que a estimativa corrente da PMF na referida seção fluvial representa a melhor informação possível de que se dispõe, tendo em conta o conhecimento mais atualizado da hidrometeorologia e hidrologia locais, e a aplicação das melhores práticas de modelagem. Entretanto, independentemente de quão precisa é a estimativa corrente da PMF local, não se conhece a probabilidade do limite superior α de igualá-la ou superá-la. De modo a obviar esta dificuldade, sem criar viés, arbitrou-se $p = 0,5$, ou seja, admitiu-se uma complete ignorância a respeito do evento, segundo o qual, o limite superior da distribuição LN4 poderá superar ou não superar a estimativa corrente da PMF local.

No que concerne a CV_α, também necessário à completa caracterização da distribuição *a priori* Gama, a análise focalizou o coeficiente de variação de cada uma das 10 categorias de estimativas de PMF, para as 561 bacias norte-americanas. Estas estimativas encontram-se grafadas segundo suas respectivas áreas de drenagem na figura 2, com ambos os eixos em coordenadas logarítmicas. Entre as 10 categorias, o máximo coeficiente de variação foi de 0,63, com um mínimo de 0,33 e um valor médio de 0,43. Uma vez que estes valores incluem as dispersões tanto horizontais, entre as diferentes áreas de drenagem em cada categoria, quanto verticais, entre as estimativas de PMF de diferentes bacias situadas em distintas áreas do território dos Estados Unidos, parece claro que eles encontram-se algo inflados. A despeito deste fato e de modo a avaliar a variabilidade de α, mesmo que a grosso modo, admitiram-se as seguintes estimativas para CV_α: 0,3, 0,5 e 0,7, respectivamente em ordem crescente de incerteza, em torno do valor corrente da PMF local. Em sumário, o procedimento A, para estimarem-se os parâmetros da distribuição *a priori* Gama para o limite superior *α*, consiste em assegurar que $\hat{\rho}_\alpha = CV_\alpha^{-2}$, com CV_α = 0,3 ou 0,5 ou 0,7, de modo que $\hat{\beta}_\alpha$ seja tal que $P(\alpha \le PMF_{local} \mid \rho_\alpha, \hat{\beta}_\alpha) = 0,5$.

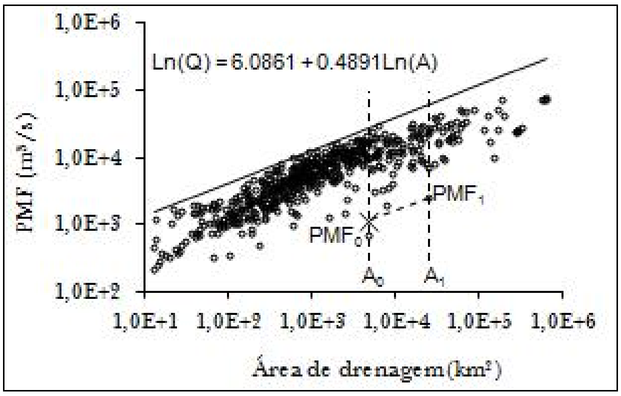

Figura 2 – Estimativas de PMF *versus* áreas de drenagem de 561 bacias dos Estados Unidos

Procedimento B

O procedimento B baseia-se na transposição de estimativas de PMF de outras bacias para aquela de interesse. Mais uma vez, emprega-se o banco de dados de 561 estimativas de PMF para as bacias norte-americanas, tal como compilado pela USNRC (1977). Como ilustrado na figura 2, as 561 áreas de drenagem estendem-se de 8 a 600.000 km² e correspondem a bacias com características climáticas e hidrológicas completamente distintas entre si. Ao invés de transpor as estimativas de PMF de outros sítios para a bacia de interesse, por meio de uma simples relação linear entre áreas de drenagem, emprega-se aqui a forma da curva envoltória das estimativas, tal como proposta por USNRC (1977), que também indica-se na figura 2. Formalmente, se A_1 e PMF_1 denotam, respectivamente, a área de drenagem e a estimativa da PMF de qualquer uma dentre as 561 bacias, e A_0 é a área de drenagem da bacia de interesse, então a estimativa da PMF transposta, representada por PMF_0, é dada por:

$$PMF_0 = PMF_1 \left(\frac{A_0}{A_1} \right)^{0.4891} \tag{7}$$

Se a equação (7) for aplicada às 561 bacias compiladas no estudo USNRC (1977), ao final, ter-se-ão 561 estimativas de PMF para uma dada área de drenagem, todas abaixo da curva envoltória ilustrada na figura 2. Isto é equivalente a afirmar que, se uma particular área de drenagem, for sujeita a todas as condições climáticas e hidrológicas, representadas pelos rebatimentos das 561 estimativas até a abscissa de interesse, então as 561 respostas hipotéticas poderão ser objeto de análise de frequência para a estimação da distribuição *a priori* do limite superior α, da distribuição LN4. As aplicações dos procedimentos A e B, à bacia do rio American na localidade de Folsom, situada no estado americano da Califórnia, serão descritas em artigo complementar a este, publicado neste número da Revista Brasileira de Recursos Hídricos. No referido artigo complementar, descrever-se-ão também aplicações para outras bacias, localizadas na Espanha e no Brasil, com menor disponibilidade de informações sobre enchentes extremos.

CONCLUSÕES

Este é um de dois artigos complementares sobre um método de estimação de quantis de enchentes extremas, com o emprego conjunto de análise Bayesiana, de informações não sistemáticas e de distribuições limitadas superiormente. Enquanto este artigo estabelece os fundamentos teóricos da proposta metodológica, o segundo descreve os estudos de caso para três bacias com diferentes conjuntos de informações acerca de enchentes extremas.

No presente artigo, fez-se uma contextualização do problema de estimação de cheias extremas, com destaque para a aparente oposição entre os conceitos de Precipitação Máxima Provável (PMP) e Cheia Máxima Provável (PMF), e a análise de frequência de variáveis hidrológicas. Verificou-se que, longe de serem conceitos opostos ou contraditórios, as estimativas de PMF (e/ou PMP) são valores de referência que podem ser incorporados à análise de frequência, desde que os arcabouços lógicos sejam convenientemente adaptados para acomodá-los.

Viu-se que um dos pontos a serem adaptados refere-se ao emprego de distribuições de probabilidade com limite superior para a variável aleatória. Dentre um conjunto não muito amplo de tais distribuições, este artigo avançou com a recomendação da distribuição transformada de Slade, ou simplesmente LN4, como modelo de cauda superior finita, em decorrência não só de suas aptidões descritivas, como também de sua forma analítica apresentar explicitamente o limite superior, como um parâmetro.

Contrariamente a outros trabalhos, nos quais o valor corrente da PMF local é tomado como a estimativa do limite superior, e em conformidade com a constatação de que as estimativas de PMF estão sujeitas a incertezas de diversas origens, utilizou-se aqui o paradigma Bayesiano como estrutura lógica apropriada para a inclusão das incertezas referentes ao limite superior das vazões. Mostrou-se que estas incertezas podem ser sumariadas por distribuições *a priori* eliciadas a partir de um grande banco de estimativas de PMF compilado para distintas bacias situadas nos Estados Unidos. Outros procedimentos semelhantes serão descritos no artigo complementar, quando as distribuições *a priori* forem eliciadas para os diversos estudos de caso.

Ainda como aspecto metodológico a ser destacado, menciona-se que a inferência estatística beneficiou-se ainda da possibilidade de inclusão de informações não-sistemáticas sobre enchentes extremas, mediante uso de uma função de verossimilhança suficientemente genérica para acomodar dados sistemáticos exatos e dados censurados provenientes de fontes históricas e/ou paleohidrológicas. Os estudos de caso, objeto do segundo dos dois artigos complementar aqui publicados, ressaltarão os pontos positivos e as limitações da metodologia aqui proposta

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.

REFERÊNCIAS

BAKER, V. R. *A bright future for old flows: origin, status and future of paleoflood hydrology.* In: Proceedings of the 2002 PHEFRA Workshop – Palaeofloods, Historical Floods and Climatic Variability: Applications in Flood Risk Assessment, Chapter II, p. 13-18, Barcelona, 2003.

BAKER, V. R. *Paleoflood hydrology and extraordinary flood events.* Journal of Hydrology, v. 96, p. 79-99, 1987.

BAKER, V. R.; WEBB, R. H.; HOUSE, P. K. *The scientific and societal value of paleoflood hydrology.* In: House, P. K.; Webb, R. H.; Baker, V. R.; Levish, D. R. (Ed.), Ancient floods, modern hazards: Principles and applications of paleoflood hydrology, Water Science and Application 5, American Geophysical Union, Washington, p. 1-19, 2002.

BERGER, J. O. *Statistical Decision Theory and Bayesian Analysis.* New York: Springer-Verlag, 1985.

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

COLES, S. G.; POWELL, E. A. *Bayesian methods in extreme value modelling: a review and new developments.* Intern. Statis. Review, v. 64, n. 1, p. 119-136, 1996.

DAWDY, D. R.; LETTENMAIER, D. P. *Initiative for risk-based flood design.* Journal of Hydrology Engineering, 113(8), 1041-1051, 1987.

DUBLER, J. R.; GRIGG, N. S. *Dam safety policy for spillway desugn floods.* Journal of Professional Issues in Engineering Education and Practice, 122 (4), 163-169, 1996.

ELÍASSON, J. *A statistical model for extreme precipitation.* Water Resources Research, v. 33(3), p. 449-455, 1997.

ENZEL. Y.; ELY, L.L.; HOUSE, P.K.; BABER, V.R.; WEBB, R.H. *Paleoflood Evidence for a Natural Upper Bound to Flood Magnitudes in the Colorado River Basin.* Water Resources Research, v. 29(7), p. 2287–2297, 1993.

FEMA – FEDERAL EMERGENCY MANAGEMENT AGENCY. *Federal Guidelines for Dam Safety: Selecting and Accommodating Inflow Design Floods for Dams.* The Interagency Committee of the U. S. Department of Homeland Security. Washington, DC, 2004.

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.

FRANCÉS, F. *Incorporating non-systematic information to flood frequency analysis using the maximum likelihood estimation method.* The use of historical data in natural hazard assessments, Glade, T., Albini, P. and Francés, F., eds., Chapter B: Flooding, Kluwer Academic Publishers, Doordrecht, 89-99, 2001.

FRANCÉS, F.; BOTERO, B. A. *Probable Maximum Flood estimation using systematic and non-systematic information.* Proceedings of 2002 PHEFRA Workshop – Palaefloods, Historical Floods and Climatic Variability: Applications in Flood Risk Assessment (eds. Thorndycraft, V.R.; Benito, G.; Barriendos, M.; Lllasat, M.C.), Chapter 34, pp. 223-229, 2003.

FRANCÉS, F.; SALAS, J. D.; BOES, D. C. *Flood frequency analysis with systematic and historical or paleoflood data based on the two parameter general extreme values models.* Water Resources Research, 30(6), 1653-1664, 1994.

HIRSCHBOECK, K. *Floods, paleofloods, and droughts: insights from the upper tails.* In: Proceedings of the CLIVAR/PAGES/IPCC Workshop - A Multi-Millennia Perspective on Droughts and Implications for the Future, Tucson, United States, 2003.

HORTON, R. E. *Hydrologic conditions as affecting the results of the application of methods of frequency analysis to flood records.* U.S. Geological Survey Water-Supply Papers, n. 771, p. 433–449, 1936.

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

HOUGHTON, J. Birth of a parent: the Wakeby distribution for modeling flood flows. Water Resources Research, 15(6), p. 1361-1372, 1978.

IACWD – INTERAGENCY ADVISORY COMMITTE ON WATER DATA. *Feasibility of Assigning a Probability to the Probable Maximum Flood.* Office of Water Coordination, Washington, DC, 1985.

ICOLD – INTERNATIONAL COMMISSION OF LARGE DAMS. *Bulletin 59 - Dam safety guidelines.* Paris, France, 1987.

JACOBY, Y.; GRODEK, T.; ENZEL, Y.; PORAT, N.; McDONALD, E. V.; DAHAN, O. *Late holocene upper bounds of flood magnitudes and twentieth century large floods in the ungauged, hyperarid alluvial Nahal Arava,* Israel. *Geomorphology*, v. 95, n. 3-4, p. 274-294, 2008.

KANDA, J. *A new value distribution with lower and upper limits for earthquake motions and wind speeds.* Theoretical and Applied Mechanics, University of Tokyo Press, v. 31, p. 351-360, 1981.

KLEMEŠ, V. *Hydrological and engineering relevance of flood frequency analysis.* In: Proceedings of the International Symposium on Flood Frequency and Risk Analysis – Regional Flood Frequency Analysis, D. Reidel Publishing Company, Baton Rouge, USA, p. 1-18, 1987.

McROBBIE, A. *The bayesian view of extreme events.* In: Proceedings of the Henderson colloquium – designing for the consequences of hazards, The Institution of Structural Engineers, London, 2004.

NATHAN, R. J.; WEINMANN, P. E. *Estimation of large and extreme floods for medium and large catchments: book VI.* In: Australian rainfall and runoff – a guide to flood estimation, The Institution of Engineers, Australia, 4° ed., 2001.

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.

NRC – NATIONAL RESEARCH COUNCIL. *Improving American river flood frequency analyses,* Committee on American River Flood Frequencies, Washington: National Academy Press, 132 p., 1999.

SLADE, J. J. *An asymmetric probability function.* Transactions: American Society of Civil Engineers, n. 62, p. 35-104, 1936.

STEDINGER, J. R.; COHN, T. A. *Flood frequency analysis with historical and paleoflood information.* Water Resources Research, v. 22(5), p. 785-794, 1986.
Stedinger, J. R.; Cohn, T. A. *Flood frequency analysis with historical and paleoflood information.* Water Resources Research, v. 22(5), p. 785-794, 1986.

STEDINGER, J. R.; HEATH, D. C.; THOMPSON, K. *Risk analysis for dam safety evaluation.* Report TCN 87-653 prep. for the U. S. Army Corps of Engineers, Ft. Belvoir, VA, 1989.

TAKARA, K.; LOEBISs, J. *Frequency analysis introducing probable maximum hydrologic events: Preliminary studies in Japan and in Indonesia.* In: LOEBIS, J. (Ed.), International Symposium on Comparative Research on Hydrology and Water Resources in Southeast Asia and the Pacific, 1996, Indonesian National Committee for International Hydrological Programme, p. 67-76, 1996.

TAKARA, K.; TOSA, K. *Storm and flood frequency analysis using PMP/PMF estimates.* In: International Symposium on Floods and Droughts, Nanjing, 1999, China, pp. 7-17, 1999.

USBR - United States Bureau of Reclamation. *Hydrologic hazard curve estimating procedures,* Research Report DSO-04-08, Denver, USA, 2004.

USNRC – UNITED STATES NUCLEAR REGULATORY COMMISSION. *Design basis floods for nuclear power plants.* Regulatory Guide 1.59, Washington, USA, 1977.

VICK, S. G. *Degrees of belief – subjective probability and engineering judgment.* Reston: ASCE Press, USA, 455 p., 2002.

VOGEL, R. M.; MATALAS, N. C.; ENGLAND, J. F.; Castellarin, A. *An assessment of exceedance probabilities of envelope curves.* Water Resources Research, v. 43, doi:10.1029/2006WR005586, 2007.

WMO – WORLD METEOROLOGICAL ORGANIZATION. *Manual for estimation of probable maximum precipitation.* Operational Hydrologic Report No. 1, WMO No. 332, 2° ed., Geneva, 270 p., 1986.

YEVJEVICH, V. *Misconceptions in hydrology and their consequences.* Water Resources Research, v. 4(2), p. 225-232, 1968.

YEVJEVICH, V.; HARMANCIOGLU, N. B. *Some reflections on the future of hydrology.* In: Proceedings of the Rome Symposium – Water for the Future: Hydrology in Perspective, IAHS Publ. 164, International Association of Hydrological Sciences, Wallingford, UK, p. 405-414, 1987.

Method To Estimate Extreme Flood Quantiles With The Joint Use Of A Bayesian Analysis, Non-Systematic Information And Upper Bounded Probability Distributions – Part 1 : Theoretical Development

ABSTRACT

This is the first 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 presents the theoretical grounds of the proposed method while the second paper describes its applications to three case studies with different sets of available data. After putting the problem of estimating extreme floods into context, the paper shows the feasibility of aggregating the at-site Probable Maximum Flood (PMF) estimate into flood frequency analysis, though the Bayesian paradigm for statistical inference and using upper-bounded probability distributions. The distribution upper- bound is uncertain and its uncertainties can be summarized through prior distributions elicited from a large data set of PMF estimates, as shown here. Then the prior distribution can be updated by including exact information on floods, from the existing systematic records, and also censored information, from historical sources and/or paleofloods, by means of a generic likelihood function. The possibility of employing such a large set of information on floods within a Bayesian framework opens up a reasonable prospect of more reliable estimates of extreme flood quantiles.

Key-words: *maximum probable flood, Bayesian analysis, extreme floods, rare hydrologic events.*