

UNIVERSIDADE FEDERAL DE MINAS GERAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM SANEAMENTO,
MEIO AMBIENTE E RECURSOS HÍDRICOS

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

Wilson dos Santos Fernandes

Belo Horizonte

2009

**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**

Wilson dos Santos Fernandes

Wilson dos Santos Fernandes

**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**

Trabalho apresentado ao Programa de Pós-graduação em Saneamento, Meio Ambiente e Recursos Hídricos da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Doutor em Saneamento, Meio Ambiente e Recursos Hídricos.

Área de concentração: Recursos Hídricos

Linha de pesquisa: Modelagem de processos hidrológicos

Orientador: Mauro da Cunha Naghettini

Co-orientadora: Rosângela Helena Loschi

Belo Horizonte

Escola de Engenharia da UFMG

2009



UNIVERSIDADE FEDERAL DE MINAS GERAIS

Escola de Engenharia

Programa de Pós-Graduação em Saneamento, Meio Ambiente e Recursos Hídricos

Av. Contorno 842 – 7º andar 30110-060 Belo Horizonte – BRASIL

Tel: 55 (31) 3238-1882 Fax: 55 (31) 3238-1882 posgrad@desa.ufmg.br

www.smarh.eng.ufmg.br

FOLHA DE APROVAÇÃO

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

WILSON DOS SANTOS FERNANDES

Tese defendida e aprovada pela banca examinadora constituída pelos Senhores:

Prof. MAURO DA CUNHA NAGHETTINI

Prof. ROSÂNGELA HELENA LOSCHI

Prof. LUIZ RAFAEL PALMIER

Prof. MÁRIO CICARELI PINHEIRO

Prof. EDUARDO SÁVIO PASSOS RODRIGUES MARTINS

Prof. MIRIAM RITA MORO-MINE

Aprovada pelo Colegiado do PG SMARH

Versão Final aprovada por

Prof. Mauro da Cunha Naghettini
Coordenador

Prof. Mauro da Cunha Naghettini
Orientador

Belo Horizonte, 13 de fevereiro de 2009.

AGRADECIMENTOS

A terminar esta tese de doutorado resta-me registrar os meus sinceros agradecimentos às pessoas que de várias formas contribuíram para que se tornasse uma realidade.

À minha esposa Janine e à minha filha Carolina, pela paciência e compreensão nas noites que me dediquei à elaboração deste trabalho, sendo sempre meu porto seguro nas horas incertas e minha fonte de alegria e sossego nos momentos difíceis.

Aos meus pais pelo constante incentivo para o meu crescimento acadêmico e pela motivação que sempre despertaram em mim. Aos meus irmãos, que sempre me acharam melhor do que realmente sou, me motivando, assim, a continuamente melhorar.

Ao meu orientador, o Professor Mauro Naghettini, por toda a dedicação, compreensão e amizade características. Pelos desafios, cada vez mais complexos, que me impôs ao longo do trabalho e pelo estímulo e exigência crescentes à medida que a tese caminhava para a sua conclusão.

À minha co-orientadora, a Professora Rosângela, sem a qual, por certo, não teria concluído este trabalho. Pela incomensurável paciência ao tentar me fazer compreender os conceitos mais abstratos da teoria bayesiana. Com a sua ajuda, um frequentista *a priori* se tornou, ou pelo menos acha que sim, um bayesiano *a posteriori*.

Aos meus amigos do Departamento de Engenharia Hidráulica e Recursos Hídrico: Bob, Prof. Márcio, Prof. Márcia, Prof. Nilo e Prof. Palmier.

Aos membros da banca examinadora, que com seus comentários e sugestões tiveram um forte impacto na qualidade da versão final deste texto.

Agradeço o apoio financeiro do CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico – ao processo 140874/2006-2 e da FAPEMIG – Fundação de Amparo à Pesquisa do Estado de Minas Gerais – ao projeto CRA APQ 4683-5.04/07 (PPM), do qual a presente pesquisa é parte integrante.

Agradeço ao Dr. John England, do USBR, pelos dados de Folsom e pelas sugestões para aprimorar o método aqui desenvolvido. Ao Prof. Dr. Félix Francés, da Universidade Politécnica de Valência, pelos dados da Espanha. Ao Luiz César, da CEMIG, pelos dados de Ponte do Vilela.

RESUMO

Pesquisas recentes em processos fluviais sugerem a idéia de que algumas variáveis hidrológicas, tais como as cheias máximas anuais, são limitadas superiormente. No entanto, quase todas as distribuições de probabilidade que são atualmente empregadas em análise de frequência de cheias são ilimitadas. Isso se deve, em parte, à dificuldade de se estimar um limite superior para as vazões com base em pequenas amostras observadas. Esta tese descreve um estudo exploratório sobre o uso conjunto de distribuições de probabilidade limitadas superiormente e informações não sistemáticas sobre cheias dentro de uma estrutura de análise bayesiana. No contexto do método desenvolvido, o valor atual da cheia máxima provável (ou PMF do inglês *Probable Maximum Flood*) aparece como uma estimativa para o limite superior das cheias máximas anuais, a despeito do fato de a determinação da PMF não ser inequívoca e depender fortemente dos dados disponíveis.

No contexto bayesiano, as incertezas sobre a PMF são incluídas na análise pela correta especificação da distribuição *a priori* para o limite superior. Na seqüência, as informações sobre os registros sistemáticos, as cheias históricas e as paleocheias são agregadas através de uma função de verossimilhança composta, a qual é usada para atualizar a informação sobre o limite superior. Combinando as incertezas quanto à PMF com informações sobre cheias de várias fontes, a expectativa é a de melhorar a estimação do limite superior e melhor descrever as incertezas associadas às enchentes máximas anuais.

Um exemplo de aplicação do método proposto foi feito no rio American, próximo ao reservatório de Folsom, na Califórnia, EUA. Outra aplicação foi feita no rio Llobregat, em Pont Du Vilomara, localizado na região da Catalunha, Espanha. Uma última aplicação foi feita no rio Pará, em Ponte do Vilela, MG, Brasil. Os resultados mostraram que é possível agrupar conceitos aparentemente incompatíveis: a estimativa determinística da PMF, tomada como um limite teórico para as cheias, e a análise de frequência de cheias máximas, com a inclusão de dados não sistemáticos. Comparada à análise convencional, a conciliação desses conceitos dentro da lógica de análise bayesiana proporcionou estimativas mais confiáveis para as cheias de grandes períodos de retorno. Por outro lado, a adoção de distribuições de probabilidade limitadas, além de ser fisicamente mais plausível, proporcionou uma melhor descrição do comportamento probabilístico das cheias em comparação às distribuições ilimitadas superiormente.

ABSTRACT

Some recent researches on fluvial processes suggest the idea that some hydrological variables, such as flood flows, are upper-bounded. However, almost all probability distributions that are currently employed in flood frequency analysis are unbounded. The complete predominance of unbounded distributions in conventional flood frequency analysis is due mainly to the difficulties of estimating upper bounds from short data samples, with zero exceedance probabilities. This work describes an exploratory study on the joint use of an upper-bounded probability distribution and non-systematic flood information, within a bayesian framework. Accordingly, the current local estimate of the Probable Maximum Flood (PMF) appears as a natural estimate of the upper-bound for maximum flows, despite the fact that PMF determination is not unequivocal and depends strongly on the available data.

In the bayesian context, the uncertainty on the PMF can be included into the analysis by considering an appropriate prior distribution for the maximum flows. In the sequence, systematic flood records, historical floods, and paleofloods can be included into a compound likelihood function which is then used to update the prior information on the upper-bound. By combining a prior distribution describing the uncertainties of PMF estimates along with various sources of flood data into a unified bayesian approach, the expectation is to obtain improved estimates of the upper-bound and better describe the uncertainties associated with flood quantiles.

The application example was conducted with flood data from the American river basin, near the Folsom reservoir, in California, USA. Other application was conducted with flood data from the Llobregat river at Pont Du Vilomara, located in Cataluña region, Spain. A final application was conducted with flood data from the Pará river at Ponte do Vilela, in Minas Gerais, Brazil. The results show that it is possible to put together concepts that appear to be incompatible: the deterministic estimate of PMF, taken as a theoretical limit for floods, and the frequency analysis of maximum flows, with the inclusion of non-systematic data. As compared to conventional analysis, the conciliation of these two concepts within the logical context of bayesian theory advances towards more reliable estimates of extreme floods. On the other hand, upper-bounded probability distributions, besides being physically more plausible, better describe the probabilistic behavior of floods as compared to unbounded distributions.

SUMÁRIO

1	INTRODUÇÃO	1
2	OBJETIVOS	8
3	PALEOHIDROLOGIA E INFORMAÇÕES HISTÓRICAS	9
3.1	INTRODUÇÃO.....	9
3.2	INDICADORES PALEOHIDROLÓGICOS E MÉTODOS DE DETERMINAÇÃO.....	11
3.3	O USO DE DADOS PALEOHIDROLÓGICOS E A EXISTÊNCIA DE UM LIMITE SUPERIOR PARA AS VAZÕES.....	16
3.4	CONCLUSÃO.....	19
4	HIDROLOGIA DE ENCHENTES	20
4.1	INTRODUÇÃO.....	20
4.2	MÉTODOS ESTOCÁSTICOS	22
4.2.1	<i>Análise de frequência convencional</i>	22
4.2.2	<i>Análise de frequência regional</i>	23
4.2.3	<i>Análise de frequência com informações não sistemáticas</i>	25
4.3	MÉTODOS DETERMINÍSTICOS	30
4.3.1	<i>Modelos de transformação chuva-vazão</i>	30
4.3.2	<i>Curvas envoltórias de vazão</i>	32
4.3.3	<i>Precipitação máxima provável (PMP) e enchente máxima provável (PMF)</i>	35
4.4	MÉTODOS MISTOS	39
4.4.1	<i>Método PVP</i>	39
4.4.2	<i>Análise de frequência com o uso conjunto de distribuições limitadas superiormente, informações não sistemáticas e PMF</i>	43
4.5	CONCLUSÃO.....	47
5	ANÁLISE DE FREQUÊNCIA BAYESIANA	48
5.1	CONCEITOS GERAIS	48
5.2	ESTIMAÇÃO BAYESIANA E INTERVALOS DE CREDIBILIDADE.....	51
5.3	MÉTODOS DE CÁLCULO	54
5.4	CONCLUSÃO.....	58
6	ABORDAGEM BAYESIANA PARA ESTIMAÇÃO DE QUANTIS DE ENCHENTES EXTREMAS COM O USO DE DISTRIBUIÇÕES LIMITADAS SUPERIORMENTE E INFORMAÇÕES NÃO SISTEMÁTICAS	60
6.1	INTRODUÇÃO.....	60
6.2	DISTRIBUIÇÕES LIMITADAS SUPERIORMENTE.....	61
6.2.1	<i>Distribuições GEV e DGP</i>	61
6.2.2	<i>Distribuição EV4</i>	63
6.2.3	<i>Distribuição TDF</i>	67

6.2.4	<i>Distribuição Log-Normal de 4 parâmetros (LN4)</i>	69
6.3	FUNÇÃO DE VEROSSIMILHANÇA PARA DADOS SISTEMÁTICOS E NÃO SISTEMÁTICOS	71
6.3.1	<i>Cheias de intensidade conhecida, superior a um limiar fixo</i>	71
6.3.2	<i>Cheias de intensidade desconhecida, superior a um limiar fixo</i>	73
6.3.3	<i>Cheias de intensidade compreendida em um intervalo, superior a um limiar fixo</i>	73
6.3.4	<i>Generalização</i>	74
6.4	PROBABILIDADES DE EXCEDÊNCIA EMPÍRICAS	78
6.4.1	<i>Probabilidades de excedência empíricas na ausência de dados não sistemáticos</i>	78
6.4.2	<i>Probabilidades de excedência empíricas com dados não sistemáticos</i>	78
6.5	CONSTRUÇÃO DA DISTRIBUIÇÃO A PRIORI PARA O LIMITE SUPERIOR	80
6.5.1	<i>Distribuição a priori para o limite superior para os Estados Unidos</i>	81
6.5.2	<i>Distribuição a priori para o limite superior para a Espanha</i>	88
6.5.3	<i>Distribuição a priori para o limite superior para o Brasil</i>	92
6.6	CONCLUSÃO	96
7	APLICAÇÕES DO MÉTODO PROPOSTO: RESULTADOS E DISCUSSÃO	99
7.1	INTRODUÇÃO	99
7.2	APLICAÇÃO PARA A BACIA DO RIO AMERICAN EM FOLSOM	99
7.2.1	<i>A bacia do rio American</i>	99
7.2.2	<i>Dados hidrológicos sistemáticos e não sistemáticos</i>	101
7.2.3	<i>Distribuição a priori para o limite superior em Fair Oaks</i>	106
7.2.4	<i>Estatísticas a posteriori</i>	108
7.2.5	<i>Análise de sensibilidade da probabilidade de excedência da PMF</i>	126
7.2.6	<i>Comparação com modelos ilimitados superiormente</i>	128
7.3	APLICAÇÃO PARA A BACIA DO RIO LLOBREGAT, EM PONT DU VILOMARA	134
7.3.1	<i>A bacia do rio Llobregat</i>	134
7.3.2	<i>Dados hidrológicos sistemáticos e não sistemáticos</i>	135
7.3.3	<i>Distribuição a priori para o limite superior em Pont Du Vilomara</i>	137
7.3.4	<i>Estatísticas a posteriori</i>	139
7.3.5	<i>Comparação com os resultados da análise de frequência clássica</i>	151
7.4	APLICAÇÃO PARA A BACIA DO RIO PARÁ, EM PONTE DO VILELA	153
7.4.1	<i>A bacia do rio Pará</i>	153
7.4.2	<i>Dados hidrológicos sistemáticos</i>	154
7.4.3	<i>Distribuição a priori para o limite superior em Ponte do Vilela</i>	155
7.4.4	<i>Estatísticas a posteriori</i>	157
7.4.5	<i>Comparação com outros métodos de análise</i>	163
8	CONCLUSÕES E RECOMENDAÇÕES	166
9	REFERÊNCIAS BIBLIOGRÁFICAS	171

ANEXO 1 –	QUANTIS DE VAZÃO MÁXIMA ANUAL PARA O MODELO LN4 EM FOLSOM	178
ANEXO 2 –	QUANTIS DE VAZÃO MÁXIMA ANUAL PARA O MODELO EV4 EM FOLSOM	180
ANEXO 3 –	QUANTIS DE VAZÃO MÁXIMA ANUAL PARA O MODELO LN4 EM PONT DU VILOMARA.....	182
ANEXO 4 –	QUANTIS DE VAZÃO MÁXIMA ANUAL PARA O MODELO EV4 EM PONT DU VILOMARA.....	183
ANEXO 5 –	QUANTIS DE VAZÃO MÁXIMA ANUAL PARA O MODELO LN4 EM PONTE DO VILELA	184
ANEXO 6 –	QUANTIS DE VAZÃO MÁXIMA ANUAL PARA O MODELO EV4 EM PONTE DO VILELA	185

LISTA DE FIGURAS

Figura 1.1 –	Categorias das cheias (adap. de Nathan e Weinmann, 2001)	2
Figura 3.1 –	Classificação cronológica das informações relativas às cheias (Fonte: Lima, 2005)	11
Figura 3.2 –	Seção fluvial com os indicadores de paleohidrológicos (Adap.: Jarrett e England, 2002)	13
Figura 3.3 –	Unidades de sedimentos deixadas por cheias (Fonte: Benito et al., 2004)	13
Figura 3.4 –	Tipos de dados paleohidrológicos	16
Figura 4.1 –	Processo iterativo do método do algoritmo dos momentos esperados (Fonte: Lima, 2005)	29
Figura 4.2 –	Parte terrestre do ciclo hidrológico (adap. Tucci, 1998)	31
Figura 4.3 –	Exemplo de curva envoltória (linha contínua) para estimativas de PMF nos Estados Unidos; o gráfico (a) mostra as estimativas de PMF sem transformação e o gráfico (b) mostra as PMF's unitárias	32
Figura 4.4 –	Estimativa do fator k_m para o método estatístico de determinação da PMP (Fonte: Bertoni e Tucci, 1993)	38
Figura 5.1 –	Interpretação gráfica do intervalo de confiança freqüentista (Fonte: Naghettini e Pinto, 2007)	52
Figura 6.1 –	Efeito de cada parâmetro na forma da distribuição EV4	65
Figura 6.2 –	Efeito de cada parâmetro na forma da distribuição TDF	69
Figura 6.3 –	Efeito de cada parâmetro na distribuição LN4	71
Figura 6.4 –	Distribuição espacial das estimativas de PMF nos EUA	83
Figura 6.5 –	Histogramas para as PMF's de cada grupo e ajuste da distribuição Gama	85
Figura 6.6 –	Estimativas de PMF versus área de drenagem para 561 bacias americanas	88
Figura 6.7 –	Cheias recordes observadas em bacias da Espanha, em função de suas respectivas áreas de drenagem	89
Figura 6.8 –	Histogramas para as cheias recordes de cada grupo e ajuste da distribuição Gama	90
Figura 6.9 –	Cheias recordes observadas na Espanha e curva envoltória para as 41 cheias recordes mundiais	92
Figura 6.10 –	Histograma para as PMF's do Brasil e ajuste da distribuição Gama	94
Figura 6.11 –	Variação do CV das PMF's do Brasil com a área de drenagem	94
Figura 6.12 –	Curvas envoltórias de PMF para o Brasil e para a região sudeste	95
Figura 7.1 –	Vista de jusante da barragem de Folsom	100
Figura 7.2 –	Localização da bacia do rio American (Fonte: USBR, 2002)	100
Figura 7.3 –	Dados sistemáticos do rio American em Fair Oaks	103

Figura 7.4 –	Locais disponíveis para análise de paleocheias no rio American próximo a Fair Oaks (adapt. USBR, 2002)	104
Figura 7.5 –	Dados não sistemáticos do rio American em Fair Oaks (Ano de referência: 2000)	105
Figura 7.6 –	Histograma das estimativas de PMF transpostas para a bacia do rio American e ajuste da distribuição Gama	107
Figura 7.7 –	Distribuições a priori de I a V para o limite superior na bacia do rio American, em Folsom	108
Figura 7.8 –	Distribuições a posteriori do parâmetro μ para os modelos completos (linha tracejada) e para os modelos com somente dados sistemáticos (linha contínua)	110
Figura 7.9 –	Distribuições a posteriori do parâmetro σ para os modelos completos (linha tracejada) e para os modelos com somente dados sistemáticos (linha contínua)	112
Figura 7.10 –	Distribuição a priori (linha tracejada), distribuição a posteriori para os modelos com todos os dados (linha pontilhada-tracejada) e distribuição a posteriori para os modelos com dados sistemáticos (linha contínua)	113
Figura 7.11 –	Distribuição preditiva a posteriori (média) ou curva de quantis (linha contínua), intervalo de credibilidade de 95% (linha tracejada). Os círculos representam os dados sistemáticos e as barras com círculos representam os dados não sistemáticos. No eixo horizontal está o período de retorno em anos e no eixo vertical está a vazão em m^3/s	116
Figura 7.12 –	Comparação entre os quantis estimados (eixo vertical) e observados (eixo horizontal) para o modelo LN4	117
Figura 7.13 –	Distribuições a posteriori do parâmetro ξ para os modelos completos (linha tracejada) e para os modelos com somente dados sistemáticos (linha contínua)	119
Figura 7.14 –	Distribuições a posteriori do parâmetro σ para os modelos completos (linha tracejada) e para os modelos com somente dados sistemáticos (linha contínua)	120
Figura 7.15 –	Distribuição a priori (linha tracejada), distribuição a posteriori para os modelos com todos os dados (linha pontilhada-tracejada) e distribuição a posteriori para os modelos com dados sistemáticos (linha contínua)	122
Figura 7.16 –	Distribuição preditiva a posteriori (média) ou curva de quantis (linha contínua), intervalo de credibilidade de 95% (linha tracejada). Os círculos representam os dados sistemáticos e as barras com círculos representam os dados não sistemáticos. No eixo horizontal está o período de retorno em anos e no eixo vertical está a vazão em m^3/s	124
Figura 7.17 –	Comparação entre os quantis estimados (eixo vertical) e observados (eixo horizontal) para o modelo EV4	125
Figura 7.18 –	Distribuições a priori II, VI e VII para o limite superior na bacia do rio American, em Folsom	127
Figura 7.19 –	Distribuições a priori (linha tracejada) e a posteriori de α dos modelos VI (gráfico a) e VII (gráfico b) para as distribuições LN4 (linha tracejada-pontilhada) e EV4 (linha contínua)	127

Figura 7.20 –	Distribuição preditiva a posteriori (linha contínua), intervalo de credibilidade de 95% (linha tracejada) para os modelos VI e VII. No eixo horizontal está o período de retorno em anos e no eixo vertical está a vazão em m ³ /s.....	128
Figura 7.21 –	Curva de quantis para o modelo limitado superiormente LN4 (Modelo V) e para o modelo ilimitado LPIII; as linhas tracejadas correspondem aos intervalos de credibilidade de 95% de probabilidade	130
Figura 7.22 –	Curva de quantis para o modelo limitado superiormente LN4 (Modelo III) e para o modelo ilimitado LPIII; as linhas tracejadas correspondem aos intervalos de credibilidade de 95% de probabilidade	130
Figura 7.23 –	Curva de quantis para o modelo limitado superiormente LN4 (Modelo V) e para o modelo ilimitado GEV; as linhas tracejadas correspondem aos intervalos de credibilidade de 95% de probabilidade	132
Figura 7.24 –	Curva de quantis para o modelo limitado superiormente LN4 (Modelo III) e para o modelo ilimitado GEV; as linhas tracejadas correspondem aos intervalos de credibilidade de 95% de probabilidade	133
Figura 7.25 –	Localização da bacia do rio Llobregat (adapt. Thorndycraft et al., 2005)	134
Figura 7.26 –	Trechos da bacia do rio Llobregat (Fonte: ACA, 2001)	134
Figura 7.27 –	Dados sistemáticos do rio Llobregat, em Pont Du Vilomara.....	135
Figura 7.28 –	Dados não sistemáticos do rio Llobregat, em Pont Du Vilomara.....	137
Figura 7.29 –	Histograma das estimativas de cheias recordes transpostas para a bacia do rio Llobregat e ajuste da distribuição Gama.....	138
Figura 7.30 –	Distribuições a priori de I a III para o limite superior na bacia do rio Llobregat.....	139
Figura 7.31 –	Distribuições a posteriori do parâmetro μ para os modelos completos (linha tracejada) e para os modelos com somente dados sistemáticos (linha contínua)	140
Figura 7.32 –	Distribuições a posteriori do parâmetro σ para os modelos completos (linha tracejada) e para os modelos com somente dados sistemáticos (linha contínua).	141
Figura 7.33 –	Distribuições a priori (linha tracejada), a posteriori para os modelos com todos os dados (linha tracejada-pontilhada) e a posteriori para os modelos com dados sistemáticos (linha contínua)	142
Figura 7.34 –	Distribuição preditiva a posteriori (média) ou curva de quantis (linha contínua), intervalo de credibilidade de 95% (linha tracejada). Os círculos representam os dados sistemáticos e as barras com círculos representam os dados não sistemáticos. No eixo horizontal está o período de retorno em anos e no eixo vertical está a vazão em m ³ /s.....	144
Figura 7.35 –	Comparação entre os quantis estimados (eixo vertical) e observados (eixo horizontal) para o modelo LN4.....	145
Figura 7.36 –	Distribuições a posteriori do parâmetro ξ para os modelos completos (linha tracejada) e para os modelos com somente dados sistemáticos (linha contínua)	146

Figura 7.37 –	Distribuições a posteriori do parâmetro σ para os modelos completos (linha tracejada) e para os modelos com somente dados sistemáticos (linha contínua)	147
Figura 7.38 –	Distribuições a priori (linha tracejada) e a posteriori do parâmetro α para os modelos completos (linha tracejada-pontilhada) e para os modelos com somente dados sistemáticos (linha contínua).....	148
Figura 7.39 –	Distribuição preditiva a posteriori (média) ou curva de quantis (linha contínua), intervalo de credibilidade de 95% (linha tracejada). Os círculos representam os dados sistemáticos e as barras com círculos representam os dados não sistemáticos. No eixo horizontal está o período de retorno em anos e no eixo vertical está a vazão em m^3/s	150
Figura 7.35 –	Comparação entre os quantis estimados (eixo vertical) e observados (eixo horizontal) para o modelo EV4	151
Figura 7.36 –	Comparação entre os métodos bayesiano e clássico para a LN4	152
Figura 7.37 –	Comparação entre os métodos bayesiano e clássico para a EV4	152
Figura 7.38 –	Localização da bacia do rio Pará	154
Figura 7.39 –	Trechos do rio Pará (Fonte: www.cbhpara.org.br)	154
Figura 7.40 –	Dados sistemáticos do rio Pará em Ponte do Vilela	155
Figura 7.41 –	Histograma das PMF's transpostas para a bacia do rio Pará e ajuste da distribuição Gama.....	156
Figura 7.42 –	Distribuições a priori de I e II para o limite superior na bacia do rio Pará.....	157
Figura 7.43 –	Distribuições a posteriori para os parâmetros da LN4; a linha tracejada representa a distribuição a priori.....	158
Figura 7.44 –	Distribuição preditiva a posteriori (média) ou curva de quantis (linha contínua), intervalo de credibilidade de 95% (linha tracejada). Os círculos representam os dados sistemáticos e as barras com círculos representam os dados não sistemáticos. No eixo horizontal está o período de retorno em anos e no eixo vertical está a vazão em m^3/s	160
Figura 7.45 –	Comparação entre os quantis estimados (eixo vertical) e observados (eixo horizontal) para o modelo LN4.....	160
Figura 7.46 –	Distribuições a posteriori para os parâmetros da EV4; a linha tracejada representa a distribuição a priori.....	161
Figura 7.47 –	Distribuição preditiva a posteriori (média) ou curva de quantis (linha contínua), intervalo de credibilidade de 95% (linha tracejada). Os círculos representam os dados sistemáticos e as barras com círculos representam os dados não sistemáticos. No eixo horizontal está o período de retorno em anos e no eixo vertical está a vazão em m^3/s	162
Figura 7.48 –	Comparação entre os quantis estimados (eixo vertical) e observados (eixo horizontal) para o modelo EV4	163
Figura 7.49 –	Comparação entre diversos métodos para os dados de Ponte do Vilela	164

LISTA DE TABELAS

Tabela 6.1 –	Valores de a para as equações de probabilidade empírica (Adap. de Naghettini e Pinto, 2007)	78
Tabela 6.2 –	Características principais de cada grupo de PMF's analisado	84
Tabela 6.3 –	Características principais de cada grupo de cheias recordes	90
Tabela 6.4 –	Características principais das estimativas de PMF no Brasil (Fonte: CBDB, 2002)	93
Tabela 7.1 –	Paleocheias individuais em Fair Oaks	106
Tabela 7.2 –	Parâmetros e características das distribuições a priori para o limite superior na bacia do rio American, em Folsom	107
Tabela 7.3 –	Estatísticas a posteriori para o parâmetro μ da LN4	110
Tabela 7.4 –	Estatísticas a posteriori para o parâmetro σ da LN4	111
Tabela 7.5 –	Estatísticas a posteriori para o limite superior α da LN4	113
Tabela 7.6 –	Estatísticas a posteriori para o parâmetro ξ da EV4	119
Tabela 7.7 –	Estatísticas a posteriori para o parâmetro σ da EV4	120
Tabela 7.8 –	Estatísticas a posteriori para o limite superior α da EV4	121
Tabela 7.9 –	Parâmetros e características das distribuições a priori VI e VII para o limite superior na bacia do rio American, em Folsom	126
Tabela 7.10 –	Estimativas a posteriori para os parâmetros da GEV	132
Tabela 7.11 –	Cheias históricas em Pont Du Vilomara	136
Tabela 7.12 –	Parâmetros e características das distribuições a priori para o limite superior na bacia do rio Llobregat, em Pont Du Vilomara	138
Tabela 7.13 –	Estatísticas a posteriori para o parâmetro μ da LN4	140
Tabela 7.14 –	Estatísticas a posteriori para o parâmetro σ da LN4	141
Tabela 7.15 –	Estatísticas a posteriori para o limite superior α da LN4	142
Tabela 7.16 –	Estatísticas a posteriori para o parâmetro ξ da EV4	146
Tabela 7.17 –	Estatísticas a posteriori para o parâmetro σ da EV4	147
Tabela 7.18 –	Estatísticas a posteriori para o parâmetro α da EV4	148
Tabela 7.19 –	Estimativas para o limite superior em Pont Du Vilomara por diferentes métodos	151
Tabela 7.20 –	Parâmetros e características das distribuições a priori para o limite superior na bacia do rio Pará, em Ponte do Vilela	156
Tabela 7.21 –	Estatísticas a posteriori para os parâmetros da LN4	158
Tabela 7.22 –	Estatísticas a posteriori para os parâmetros da EV4	161

LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS

A, Ad	– Área de drenagem
CPRM	– Companhia de Pesquisa de Recursos Minerais – Serviço Geológico do Brasil
CV	– Coeficiente de variação
DB	– Vazões históricas limitadas inferior e superiormente
DGP	– Distribuição de Pareto generalizada
DP	– Desvio padrão
EMA	– Método do algoritmo dos momentos esperados
EV2	– Distribuição de Fréchet ou de extremos do tipo II
EV3	– Distribuição de Weibull ou de extremos tipo III
EV4	– Distribuição de valores extremos do tipo IV
EX	– Dados sistemáticos ou exatos
F, F_x	– Função acumulada de probabilidade
FDA	– Função acumulada de probabilidade
FDP	– Função densidade de probabilidade
f_m	– Fator de maximização da PMP
f_x, f	– Função densidade de probabilidade
GAM	– Distribuição Gama
GEV	– Distribuição Generalizada de Valores Extremos
GUM	– Distribuição de Gumbel
HPD	– Highest Probability Density ou intervalo de credibilidade
ICOLD	– Comitê Internacional de Grandes Barragens
L	– Função de verossimilhança
LB	– Vazões históricas limitadas inferiormente
LN4	– Distribuição Log-Normal de 4 parâmetros
LPIII	– Distribuição Log-Pearson III
MANIAC	– Mathematical Analyzer, Numerical Integrator and Computer
MCMC	– Markov Chain Monte Carlo
MLE	– Método da máxima verossimilhança
MML	– Método dos momentos-L
MOM	– Método dos momentos
MPH	– Momentos ponderados historicamente
NRC	– National Research Council
PMF	– Probable Maximum Flood (Cheia Máxima Provável)
PMP	– Probable Maximum Precipitation (Precipitação Máxima Provável)

PVP	– Método Pico-Volume-Precipitação
Q	– Vazão máxima anual
SWM	– Stanford Watershed Model
T	– Período de retorno em anos
TCEV	– Two-Component Extreme Value Distribution
TDF	– Distribuição de valores extremos transformada
UB	– Vazões históricas limitadas superiormente
USACE	– United States Army Corps of Engineers
USBR	– United States Bureau of Reclamation
USGS	– United States Geological Service
USNRC	– United States Nuclear Regulatory Commission
USWRC	– United States Water Resources Council
w	– Massa de água precipitável por unidade de área
WMO	– Organização Meteorológica Mundial
y_U	– Limiar de referência
α	– Limite superior das vazões máximas anuais
ε	– Limite inferior das vazões máximas anuais
μ	– Parâmetro de posição
μ_r	– Momento de ordem r
σ, β	– Parâmetro de escala
ξ, ρ	– Parâmetro de forma

1 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.

Nathan e Weinmann (2001) categorizam as enchentes como grandes (probabilidade de excedência maior que 10^{-2}), raras (probabilidade de excedência entre 10^{-2} e 2×10^{-3}) e extremas (probabilidade de excedência menor que 2×10^{-3}). A figura 1.1 mostra as categorias definidas pelos autores. Em geral, as enchentes consideradas grandes ainda se situam no domínio das medições e observações diretas, enquanto as enchentes raras localizam-se entre essas e o chamado "limite crível de extrapolação" da curva de probabilidades anuais de superação. As enchentes extremas sabidamente 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 essas estimativas são absolutamente necessárias para a quantificação dos riscos associados ao colapso de estruturas hidráulicas, entre as quais destacam-se as grandes barragens.

As enchentes extremas encontram-se nas fronteiras do desconhecimento. As incertezas inerentes às suas estimativas são sabidamente muito grandes, embora não sejam precisamente quantificáveis 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 2×10^{-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. Um desses limites superiores, talvez o mais usual na prática da engenharia hidrológica, é representado pela "Enchente Máxima Provável" (PMF, do acrônimo em inglês *Probable Maximum Flood*), empregada com muita frequência no Brasil e em outros países como critério de projeto de

vertedores de grandes barragens e preconizada para tal fim por alguns boletins do ICOLD - Comitê Internacional de Grandes Barragens (<http://www.icold-ciqb.org/>).

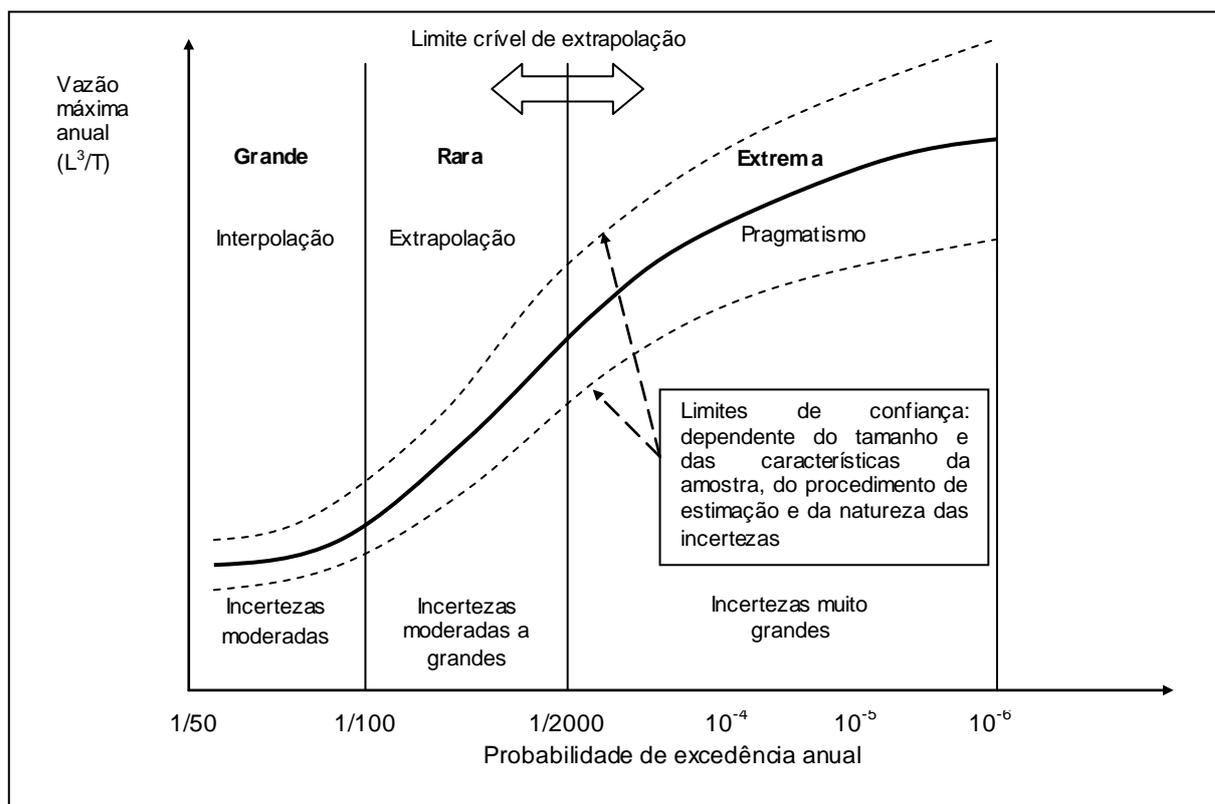


Figura 1.1 – Categorias das cheias (adap. de Nathan e Weinmann, 2001)

A estimação de enchentes raras e extremas foi objeto de constante interesse científico e tecnológico, ao longo do último século, motivada pela necessidade de prover diretrizes adequadas de projeto e avaliação de risco para um grande número de estruturas de aproveitamento de recursos hídricos e de mitigação de cheias. Antes de 1930, alguns métodos empíricos, como a curva envoltória de Creager e a fórmula de Fuller, eram de uso corrente, enquanto os fundamentos da teoria probabilística de valores extremos começavam a ser estabelecidos e aplicados ao estudo de enchentes. Durante a década de 1930, a impossibilidade dos métodos correntes em prever a ocorrência de algumas cheias catastróficas nos Estados Unidos motivou as autoridades desse país a estimular o desenvolvimento de abordagens determinísticas para a estimação de enchentes extremas.

A constatação por parte de influentes pesquisadores, tais como Horton (1936), de que deve existir um limite físico superior para a produção de tormentas e enchentes em uma certa bacia, simultaneamente à formulação da teoria do hidrograma unitário e ao desenvolvimento da meteorologia, impeliram grandemente o surgimento dos conceitos de precipitações e enchentes máximas possíveis. Ao final da década de 1930, esses conceitos já eram amplamente empregados nos Estados Unidos como elementos de projeto de grandes

estruturas hidráulicas. Nas décadas seguintes, vários refinamentos foram acrescentados à abordagem original, culminando com a publicação do "Manual de Estimação da PMP", por parte da Organização Meteorológica Mundial (WMO, 1986), e com o desenvolvimento paralelo de um grande número de modelos de simulação chuva-vazão (Singh e Woolhiser, 2002), necessários à transformação dos limites hidrometeorológicos em hidrológicos. Hoje, no Brasil e em diversos outros países, os conceitos de precipitação máxima provável (PMP) e cheia máxima provável (PMF) são correntemente empregados na prática da engenharia hidrológica como elementos essenciais para o projeto de vertedores de grandes barragens e de outras estruturas, tais como as centrais nucleares, cuja inundação é considerada ambiental e socialmente inaceitável.

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. Os métodos de cálculo da PMP podem ser o de maximização meteorológica local de eventos históricos com ou sem transposição de tormentas, ou o de estimação estatística, tal como o proposto por Hershfield (1961, 1965), posteriormente modificado por Koutsoyiannis (1999).

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. Nesses modelos, as variáveis de estado devem refletir as condições críticas que antecedem um evento hipoteticamente tido como o limite superior da enchente potencial local. Embora a PMF esteja associada a um limite superior, a sua determinação não é inequívoca e depende da disponibilidade de um conjunto de observações históricas. Isso faz com que a PMF seja suscetível às incertezas impostas pelas amostras disponíveis, sendo considerada por alguns órgãos governamentais, como o United States Bureau of Reclamation (USBR, 2004), como um limite prático superior para extrapolações das curvas de frequência de cheias. Outros, como mencionados por Nathan e Weinmann (2001), atribuem à PMP e, indiretamente à PMF, probabilidades de superação na faixa entre 10^{-4} e 10^{-7} , o que, de fato, é um mero expediente empírico, motivado pela necessidade de incorporação do componente hidrológico nos estudos de quantificação do risco global de falha de grandes barragens.

Sob o ponto de vista da incorporação do conceito de PMF aos modelos probabilísticos, a análise convencional de frequência de cheias apresenta dois grandes obstáculos. Primeiramente, as amostras relativamente curtas de registros fluviométricos, ou sistemáticos, via de regra não excedem a uma centena de anos e não fornecem informações 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 que todas as distribuições de probabilidades, usuais na análise de frequências de eventos hidrológicos extremos, são, por construção, ilimitadas superiormente e, portanto, não acomodam em sua formulação a idéia inerente ao conceito da PMF.

O predomínio de distribuições de probabilidades ilimitadas superiormente em hidrologia estatística se deve ao fato que as distribuições limitadas superiormente pressupõem a estimação do limite superior a partir de uma amostra de vazões de enchentes (ou de precipitações máximas), de tamanho relativamente pequeno. Além das dificuldades inerentes à estimação, há também o receio de se prescrever um modelo distributivo que pressupõe um valor da variável aleatória cuja probabilidade de excedência seja nula.

Por outro lado, indo além da análise convencional de frequência de cheias vista como um mero expediente de engenharia, os geomorfólogos argumentam que não há razão científica para afirmar que as cheias de um ribeirão ou um rio qualquer não tenham um limite físico e que, portanto, possam tender ao infinito. Essa é uma longa controvérsia que remonta ao início do século XX (Horton, 1936) e que persiste até os dias de hoje.

De acordo com Takara e Tosa (1999), variáveis hidrológicas, tais como a vazão ou a precipitação, podem assumir somente valores não-negativos, com um valor máximo fisicamente possível, prescrito por um limite superior finito, e devem ser analisadas sob esses aspectos. Nessa mesma linha de raciocínio, outros pesquisadores, como Boughton (1980) e Laursen (1983), recomendam que somente distribuições limitadas superiormente devam ser usadas para modelar as vazões de cheias ou as precipitações máximas.

Hosking e Wallis (1997) consideram errônea a recomendação de se usar somente distribuições limitadas e sustentam que, se o objetivo da análise de frequência é o de estimar o quantil de tempo de retorno de 100 anos, por exemplo, é irrelevante considerar como “fisicamente impossível” a ocorrência do quantil de 100.000 anos. Acrescentam que impor um limite superior ao modelo probabilístico pode comprometer a obtenção de estimativas

confiáveis de quantis para os tempos de retorno que de fato interessam. Contraponto a essa afirmação, Takara e Tosa (1999) analisaram o efeito da incorporação de limite superior aos dados de chuva e vazão de quatro bacias japonesas, com tamanhos diversos e coeficientes de assimetria variando entre 0,84 e 3,01, por meio do uso de distribuições limitadas, e apresentaram evidências de que, além do fato desses modelos serem fisicamente mais plausíveis, as estimativas dos quantis são mais acuradas, sobretudo para amostras pequenas.

Nas últimas décadas, como uma tentativa válida 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. De modo simples, as informações não sistemáticas são aquelas registradas anteriormente aos dados fluviométricos, direta ou indiretamente, seja por uma evidência natural ou por alguma forma de observação. 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), as quais correspondem às enchentes que atingiram, em algum momento do período pré-histórico, níveis d'água extremos, os quais podem ser reconstituídos a partir de evidências físicas, geológicas e/ou botânicas. Como será visto no decorrer desta tese, vários autores têm mostrado que a incorporação de dados não sistemáticos na análise de frequência pode melhorar a estimativa de quantis com alto período de retorno, além de se constituir uma alternativa válida para a melhor interpretação das incertezas das estimativas de parâmetros das distribuições de probabilidades.

No que concerne às distribuições limitadas superiormente, destacam-se: (i) a distribuição EV4, proposta por Kanda (1981) para analisar o comportamento probabilístico de abalos sísmicos e de ventos extremos no Japão; (ii) a distribuição de valores extremos transformada (TDF), proposta por Elíasson (1994), que a utilizou na análise de frequência de precipitações extremas na Islândia e nos Estados Unidos; e (iii) a distribuição Log-Normal de 4 parâmetros (LN4), proposta por Takara e Loebis (1996), com base na variável transformada de Slade (Slade, 1936 *apud* Botero, 2006), empregada na análise de frequência de precipitações extremas no Japão e Indonésia. Entretanto, embora os modelos distributivos propostos nos trabalhos citados sejam limitados superiormente, seus autores partiram da premissa de que o valor do limite superior é previamente conhecido e não é atribuído nenhum erro à estimação desse limite. Uma abordagem diferente do problema de estimação do limite superior dessas distribuições foi apresentada por Botero (2006).

O trabalho de Botero (2006) aponta uma direção promissora para a associação de probabilidades aos quantis de enchentes extremas. No trabalho citado, a PMF é, de fato, um dos parâmetros dos modelos distributivos limitados superiormente, a cujo valor é possível associar as incertezas inerentes à sua estimação. O núcleo do método de estimação da PMF, por meio de distribuições limitadas superiormente e da incorporação de informações não sistemáticas, é dado pela construção de uma função de verossimilhança cuja maximização fornece as estimativas desejadas. A seleção desse método de estimação se justifica por suas propriedades estatísticas, decorrentes de sua aplicação a amostras de tamanho majorado pela inclusão de novas informações, e pela maior facilidade de incorporar tais informações ao processo de inferência.

O método desenvolvido por Botero (2006) é o ponto de partida para o método proposto e implementado no decorrer desta pesquisa. Nesse sentido, esta tese de doutorado busca reavaliar, por meio da proposição de uma nova estrutura de análise, os métodos de incorporação de um limite superior à análise de frequência de vazões máximas anuais. Como será visto, o limite superior pode ser entendido, dentro de uma abordagem bayesiana, como uma quantidade fixa, porém desconhecida, cujas incertezas associadas à sua estimação podem ser modeladas por uma distribuição de probabilidades, construída subjetivamente, a partir do conhecimento do especialista sobre essa quantidade. Neste particular, será visto que as estimativas de PMF, que aqui são tomadas como um estimador natural para o limite superior, fornecem a base de conhecimento que permite avaliar as incertezas associadas ao limite superior.

A abordagem bayesiana será empregada visando contornar a difícil tarefa de se fixar ou estimar o valor do limite superior a partir de uma amostra reduzida de informações sobre eventos extremos. Neste contexto, o objetivo principal da pesquisa é utilizar a análise bayesiana para desenvolver um método que torne mais lógica e consistente a inclusão da PMF como estimador do limite superior, juntamente com os conceitos desenvolvidos por Botero (2006), que permitem a inclusão de informação não sistemática na análise de frequência, com o uso de distribuições de probabilidade limitadas superiormente.

A inovação introduzida por esta tese é metodológica. A oportunidade de se combinar informações sistemáticas e não sistemáticas sobre enchentes, dentro do contexto de análise bayesiana, com o emprego de modelos probabilísticos limitados superiormente, permite uma interpretação mais abrangente das diferentes fontes de incertezas presentes na estimação de cheias extremas. Além disso, o método aqui desenvolvido permite a reconciliação dos

conceitos da PMF e da análise de frequência de cheias, tão empregados na prática da engenharia de recursos hídricos, mas tão distantes e encapsulados em suas respectivas raízes determinísticas e estocásticas. Em consequência, o método aqui proposto é um importante passo adiante na busca de um instrumento confiável de avaliação das probabilidades associadas às denominadas cheias de projeto de estruturas de engenharia e, conseqüentemente, da correta inclusão do componente hidrológico na determinação do risco global de falha dessas estruturas.

Além deste capítulo introdutório, esta tese é organizada da seguinte forma: no capítulo 2, são apresentados os objetivos geral e específicos da pesquisa. No capítulo 3, é feita uma descrição dos conceitos referentes às informações não sistemáticas sobre as enchentes. Serão avaliados os tipos de dados não sistemáticos, bem como os métodos para sua obtenção e análise. No capítulo 4, é feita uma revisão da literatura sobre os principais métodos de análise de cheias. No capítulo 5, são apresentados os conceitos da abordagem bayesiana, destacando-se os métodos de cálculo das distribuições *a posteriori* e os métodos de estimação pontual. No capítulo 6, descreve-se a modelagem estatística utilizada nesta tese. Nesse capítulo são apresentadas as distribuições de probabilidade limitadas superiormente, as funções de verossimilhança que agregam os dados não sistemáticos e os métodos para a construção da distribuição *a priori* para o limite superior, ou seja, são descritos os componentes do método proposto. No capítulo 7, é feita uma aplicação completa do método à bacia do rio American, no estado americano da Califórnia. Em seguida, são descritas outras aplicações do método proposto às bacias do rio Llobregat, na Espanha, e do rio Pará, em Ponte do Vilela, no Brasil. O trabalho é concluído no capítulo 8, onde são feitas análises gerais sobre os resultados obtidos e são propostas as recomendações para futuros desenvolvimentos desta pesquisa. Finalmente, no capítulo 9, são apresentadas as referências bibliográficas utilizadas no trabalho.

2 OBJETIVOS

O objetivo principal desta pesquisa é desenvolver um método para a estimação de quantis de enchentes extremas com o emprego conjunto da análise bayesiana, de informações não sistemáticas e de distribuições limitadas superiormente.

Os objetivos específicos são listados a seguir:

- Desenvolver um método para a construção da distribuição *a priori* do limite superior com base na variabilidade das estimativas da PMF;
- Avaliar o impacto da inclusão da informação não sistemática na curva de quantis e na estimação dos parâmetros;
- Fazer uma análise de sensibilidade paramétrica e definir intervalos de credibilidade para os quantis estimados e para os parâmetros;
- Comparar os resultados obtidos para as distribuições limitadas superiormente com os obtidos para distribuições ilimitadas;
- Avaliar os resultados obtidos face a outros métodos de estimação de cheias extremas; e
- Avaliar as vantagens da abordagem bayesiana frente à abordagem convencional frequentista.

3 PALEOHIDROLOGIA E INFORMAÇÕES HISTÓRICAS

3.1 Introdução

A paleohidrologia é a ciência de reconstrução da magnitude e frequência de grandes cheias usando evidências geológicas e botânicas e uma variedade de técnicas interdisciplinares (Baker et al., 2002). De acordo com Benito et al. (2004), o termo “paleo” tem contribuído para a concepção errada de que as técnicas de paleohidrologia são usadas unicamente para estimar cheias muito antigas (escala geológica). De fato, a maioria dos estudos de paleohidrologia envolve a estimativa de cheias pré-históricas (últimos 5.000 anos), históricas (últimos 1.000 anos), modernas (últimos 50 anos) e até mesmo cheias recentes em locais não monitorados. Numa classificação climática, os estudos paleohidrológicos se limitam ao Holoceno, que teve início há cerca de 10.000 anos, quando terminou a última grande glaciação na Terra.

Segundo Saint-Laurent (2004), a pesquisa científica nas últimas duas décadas tem se caracterizado pelo crescente número de estudos sobre aquecimento global e seus impactos nos vários ambientes terrestres. A pesquisa em hidrologia tem tentado estabelecer uma ligação entre mudanças climáticas e as mudanças nos sistemas fluviais. No entanto, a escala cronológica usada é frequentemente muito curta para avaliar corretamente os impactos das mudanças climáticas nos fenômenos hidrológicos. Além disso, é difícil estabelecer quais fenômenos são resultados da atual mudança climática e quais são resultados de mudanças antropogênicas, especialmente nos últimos 100 anos. Essas dificuldades têm levado os pesquisadores a considerar escalas cronológicas maiores (> 1.000 anos). Nesse sentido, a paleohidrologia permite aumentar a escala temporal dos dados hidrológicos por meio da determinação da frequência e magnitude de eventos extremos ocorridos antes do período histórico.

De acordo com Thorndycraft et al. (2005), as enchentes são os eventos mais destrutivos dentre os desastres naturais na Europa. A cheia ocorrida naquele continente em 2002, por exemplo, devastou muitas partes da Europa Central, com perdas estimadas em € 55 bilhões. Ainda de acordo com os mesmos autores, os métodos estatísticos de definição de cheias que usam somente dados de estações fluviométricas, em geral, subestimam o risco de enchentes, acarretando uma considerável perda de propriedades e vidas. Os dados fluviométricos são, geralmente, limitados a poucas dezenas de anos e as grandes cheias são sub-representadas nesse conjunto de informações. Contrariamente, as informações não sistemáticas podem

abranjer uma escala temporal superior a 1.000 anos, o que permite uma melhor descrição da frequência de eventos extremos.

Ouarda et al. (1998) classificaram a informação não contemporânea, ou não sistemática, relativa às cheias em três categorias: (1) evidências geológicas de cheias pré-históricas (paleohidrologia), (2) evidências botânicas de cheias pré-históricas, especialmente aquelas encontradas nos troncos das árvores (dendrohidrologia), e (3) observações registradas em jornais, arquivos e outros documentos. No entanto, a classificação proposta por Baker (1987), em função da ordem cronológica dos eventos, é mais adequada aos propósitos desta tese. De acordo com essa classificação, o período completo de informações sobre cheias é dividido em três momentos distintos: pré-histórico, histórico e contemporâneo.

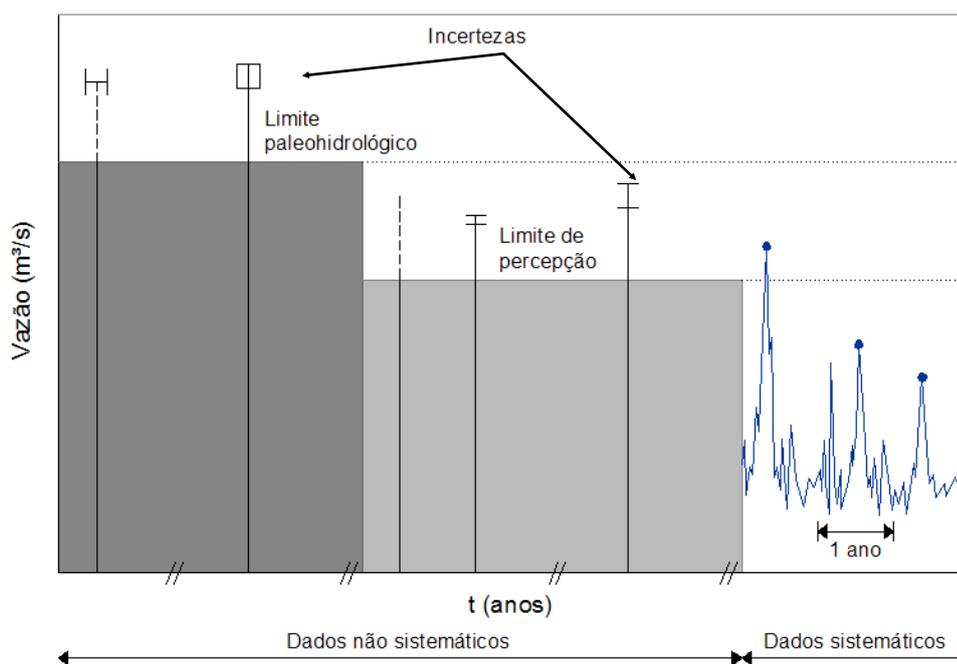
O período pré-histórico fornece informações sobre eventos antigos de cheias, os quais não foram diretamente observados por seres humanos. Tais informações são obtidas de maneira indireta, por meio de evidências físicas da ocorrência das denominadas “paleocheias”. As técnicas empregadas para a determinação dessas vazões, conforme será visto nos próximos itens, combinam sedimentologia, geomorfologia e geobotânica, e se dividem em dois grupos principais: estudos de depósitos de sedimentos e estudos botânicos.

O período histórico se refere aos eventos de cheias diretamente observados por seres humanos que ocorreram e foram documentados antes do período de observações fluviométricas regulares. A coleta de informações sobre cheias históricas requer um trabalho metuculoso, que envolve a consulta a diversos volumes manuscritos, jornais e outros periódicos antigos, diários de viagens, órgãos administrativos locais, arquivos históricos e, até mesmo, relatos pessoais (Lima, 2005).

De acordo com Lima (2005), as informações históricas incluem desde descrições dos danos causados pela cheia, até a data de ocorrência do evento, ou a última vez que uma cheia daquela magnitude ocorreu, ou, ainda, referências ao nível atingido durante a passagem do pico da cheia. Essas informações, associadas à modelagem hidráulica, permitem avaliar a magnitude e a frequência de cheias ocorridas durante o período histórico (em geral, alguns séculos), que foram documentadas especialmente por terem excedido certos limiares de referência ou de percepção.

O período contemporâneo se refere às observações fluviométricas regulares, provenientes das estações de medição. Essas observações são geralmente expressas em termos de elevações do nível d'água, as quais são convertidas em vazões por meio das respectivas curvas-chave.

A figura 3.1 mostra os diferentes tipos de informações relativas às cheias, de acordo com a classificação cronológica dos eventos descrita anteriormente.



Período	Pré-histórico	Histórico	Contemporâneo
Técnica de coleta de dados	Paleohidrologia (paleocheias)	Observações ocasionais documentadas ou relatadas (cheias históricas)	Estações de medição (cotas e vazões)
Tipos de dados	Amostras censuradas, com eventos de intensidade conhecida ou não		Amostras não censuradas
Incertezas	Relativas à data e à intensidade	Relativas à intensidade	

Figura 3.1 – Classificação cronológica das informações relativas às cheias (Fonte: Lima, 2005)

3.2 Indicadores paleohidrológicos e métodos de determinação

A paleohidrologia tem sido empregada em várias partes do mundo para melhorar a estimativa do risco de cheias, aumentando a quantidade de dados sobre eventos extremos. No Brasil, não se tem conhecimento da realização de estudos paleohidrológicos, a despeito de eles terem se mostrado de extrema importância na análise de frequência de cheias (Stedinger e Cohn, 1986).

Benito et al. (2004) recomendam os seguintes passos para a elaboração de estudos paleohidrológicos:

- 1) interpretação e análise de fotos aéreas e mapas topográficos em várias escalas na região de interesse;
- 2) visita em campo e pesquisa para a identificação e seleção de indicadores de cheias (marcas e depósitos de sedimentos);
- 3) descrição estratigráfica do solo com ênfase na identificação de vestígios de cheias;
- 4) coleta de amostras para datação dos vestígios de cheias;
- 5) investigação topográfica para a identificação de locais atingidos por cheias;
- 6) modelagem hidráulica e estimação das cheias;
- 7) comparação dos resultados com dados históricos disponíveis;
- 8) análise de frequência de cheias; e
- 9) outras aplicações, tais como, definição de áreas inundáveis e estudos de mudança climática, entre outras.

Os canais em leitos rochosos são os que possuem as melhores configurações para a reconstrução das descargas paleohidrológicas devido à estabilidade de suas seções transversais e à mudança nítida das condições do fluxo entre o canal principal e o secundário, onde as velocidades de escoamento são significativamente reduzidas. Nas áreas marginais dos cursos d'água, durante eventos de cheia, a velocidade reduzida da água favorece a deposição de argila, silte e areia, constituindo um registro ou marca da cheia que possibilita a sua respectiva reconstrução. A figura 3.2 mostra uma seção fluvial esquemática com as principais fontes de indicadores paleohidrológicos.

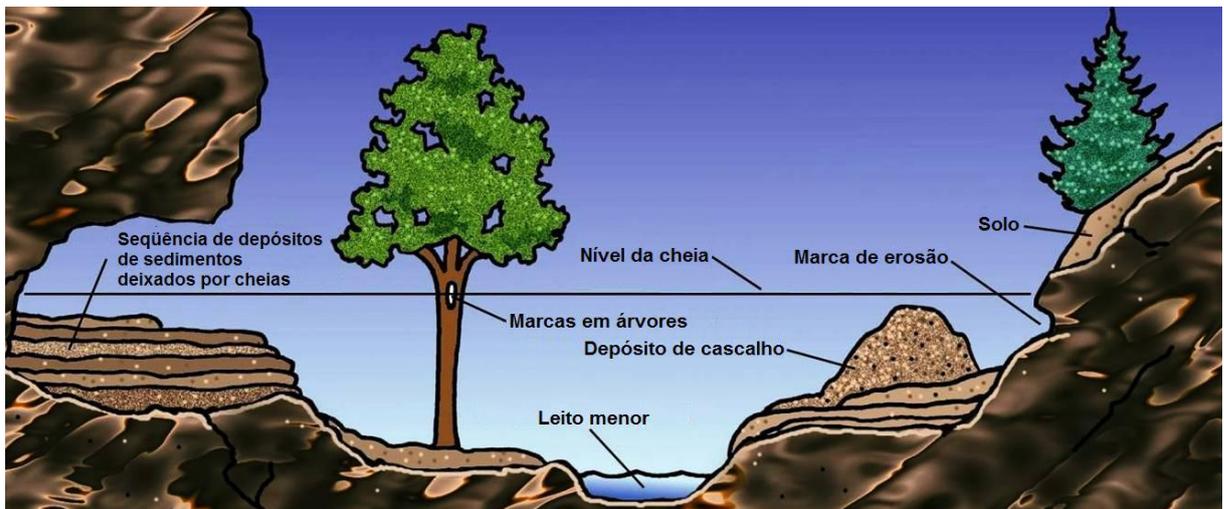


Figura 3.2 – Seção fluvial com os indicadores de paleohidrologia (Adap.: Jarrett e England, 2002)

A seqüência de depósitos de sedimentos deixados por cheias ou, simplesmente, unidades sedimentares, quando preservadas, permitem estabelecer o número de vezes que uma determinada cota da seção transversal foi atingida por um evento fluvial extremo. A figura 3.3 mostra uma seqüência de depósitos de sedimentos obtida na bacia do rio Llobregat, na Espanha, que será objeto de estudo no capítulo 7.

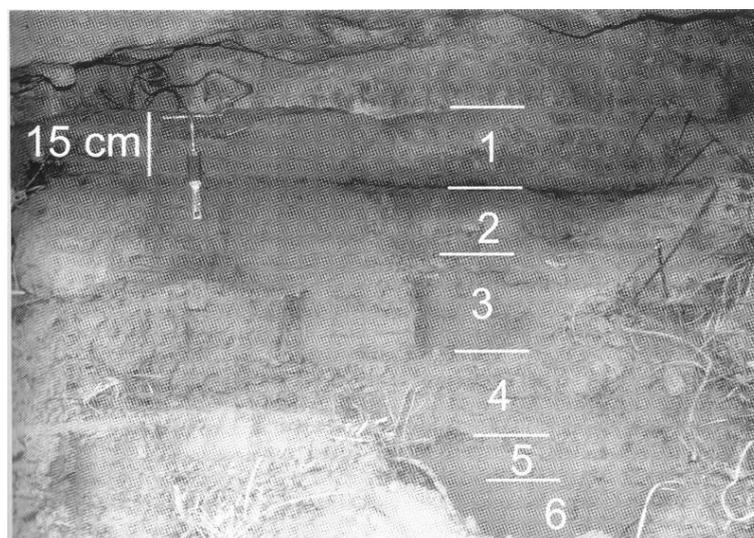


Figura 3.3 – Unidades de sedimentos deixadas por cheias (Fonte: Benito et al., 2004)

Benito et al. (2004) citam pelo menos 4 critérios para a identificação das unidades de sedimentos:

- Identificação de camadas distintas de argila no topo das unidades sedimentares, que representa o nível mínimo atingido pela cheia;
- Presença de atividade biológica (plantas e animais), indicando que aquela unidade já esteve com sua superfície exposta;

- Marcas de erosão em encostas, mostrando que uma determinada unidade sedimentar foi exposta por uma cheia maior; e
- Mudanças nas características físicas do sedimento, tais como cor e tamanho das partículas do solo. Isso indica que as unidades sedimentares foram formadas por sedimentos de diferentes fontes, trazidos para o local sob diferentes condições de fluxo da água.

Outra forma de identificação de unidades sedimentares é a presença de artefatos históricos, deixados por comunidades instaladas na bacia em passado remoto. Esse método, embora pouco utilizado, foi aplicado à bacia do rio American, que será objeto de estudo no capítulo 7.

Uma vez identificadas as unidades sedimentares provenientes de cheias extremas, as mesmas devem ser datadas. Assim, é possível estabelecer a frequência com que a cheia ocorreu na seção. De acordo com Benito et al. (2004), o método mais utilizado para o estabelecimento da cronologia das cheias é a datação por isótopos radioativos de carbono.

Na datação por radiocarbonos, é medido o decaimento de isótopos de ^{14}C presentes em amostras orgânicas das unidades sedimentares. O ^{14}C é produzido na atmosfera por meio de uma reação nuclear entre nêutrons térmicos, produzidos por raios cósmicos, com um núcleo estável de nitrogênio ($^{14}_7\text{N} + n \rightarrow ^{14}_6\text{C} + p$). O ^{14}C é, então, oxidado em CO_2 e introduzido no ciclo do carbono.

As plantas assimilam o CO_2 atmosférico, via fotossíntese, formando compostos orgânicos e os animais, ao se alimentarem dos vegetais, incorporam o carbono. Esse processo faz com que a quantidade de isótopos de carbono nos seres vivos esteja em equilíbrio com a da atmosfera. Quando a planta (ou outro ser vivo) morre, o equilíbrio é interrompido e os isótopos de ^{14}C começam a decair para ^{14}N , com uma meia vida de 5.730 anos (Pessenda et al., 2005). Dessa maneira, sabendo-se da quantidade inicial de ^{14}C presente na amostra (na morte da planta) e supondo que o ciclo do carbono é fechado, ou seja, não há entradas e saídas de carbono na Terra, pode-se avaliar a quantidade de carbono após t anos pela lei do decaimento radiativo dada por:

$$A = A_0 e^{-\lambda t}, \quad (3.1)$$

onde A e A_0 são, respectivamente, as quantidades de carbono após t anos e na morte da planta, e λ é uma constante de decaimento dependente das características físicas do elemento (carbono) em análise.

Na verdade, no estabelecimento da cronologia de cheias a equação 3.1 é aplicada de forma inversa. A partir de uma amostra de material orgânico retirado da unidade sedimentar estabelece-se a quantidade atual (A) de carbono. Pessenda et al. (2005) citam entre os métodos mais utilizados para a detecção da atividade do ^{14}C o acelerador de partículas acoplado a um espectômetro de massa (AMS – *Accelerator Mass Spectrometry*). Supondo-se que a quantidade inicial de carbono é a mesma da atmosfera atual, a data de morte da planta é obtida diretamente de 3.1.

O passo seguinte na identificação das paleocheias é definir a vazão capaz de produzir cada unidade sedimentar. Com efeito, em estudos paleohidrológicos são identificados vários pontos, ao longo do rio, com as mesmas características geomorfológicas, formando um perfil de linha d'água. A partir de uma modelagem hidráulica do trecho fluvial são identificadas as vazões capazes de produzir tal perfil. Webb e Jarret (2002) afirmam que o método mais utilizado para estimar vazões a partir de indicadores paleohidrológicos é o *step-backwater*. Esse método utiliza a equação da conservação de massa, de energia e a equação de Manning, juntamente com as características hidráulicas (rugosidade) e geométricas (forma, declividade etc.) do trecho fluvial, num processo iterativo no qual são definidos perfis de linha d'água para várias condições de vazão. O perfil que mais se aproxima daquele obtido pelos indicadores paleohidrológicos é adotado para futuros estudos. Detalhes desse método podem ser encontrados em Chow (1959).

Modelos hidráulicos bidimensionais, tal qual o TRIM2D descrito por Denlinger et al. (2002), também são aplicados na definição de paleocheias. No entanto, esse tipo de modelagem exige um maior detalhamento das características do trecho fluvial à época da ocorrência da cheia, que nem sempre é possível para a maioria das aplicações práticas.

O maior problema na modelagem hidráulica de linhas d'água históricas é estabelecer as características geométricas do trecho fluvial à época da ocorrência da cheia. Fatores como a urbanização e mudanças do uso do solo impossibilitam a modelagem hidráulica sob as mesmas condições da ocorrência da paleocheia, fazendo com que as incertezas sobre o passado hidrológico da bacia aumentem significativamente. Sob esse aspecto, as melhores

condições para a aplicação de estudos paleohidrológicos são encontradas em bacia rurais, com poucas alterações geomorfológicas ao longo do tempo.

Por fim, a modelagem hidráulica fornece uma série de dados não sistemáticos de vazão que podem ser classificados em 4 tipos:

- Dados EX (exatos): quando é possível determinar, a menos de incertezas de modelagem, o valor exato da cheia que ocorreu no passado. Esse tipo de dado aparece com bastante raridade em registros paleohidrológicos. São mais comuns em registros históricos de cheias ocorridas em meio urbano.
- Dados UB (*upper bounded*): quando o valor da cheia não é conhecido. Sabe-se somente que a cheia não atingiu um determinado nível.
- Dados LB (*lower bounded*): quando o valor da cheia não é conhecido. Sabe-se somente que a cheia foi maior que um determinado nível.
- Dados DB (*double bounded*): quando ocorre uma composição dos casos LB e UB.

A figura a seguir exemplifica cada tipo de dado.

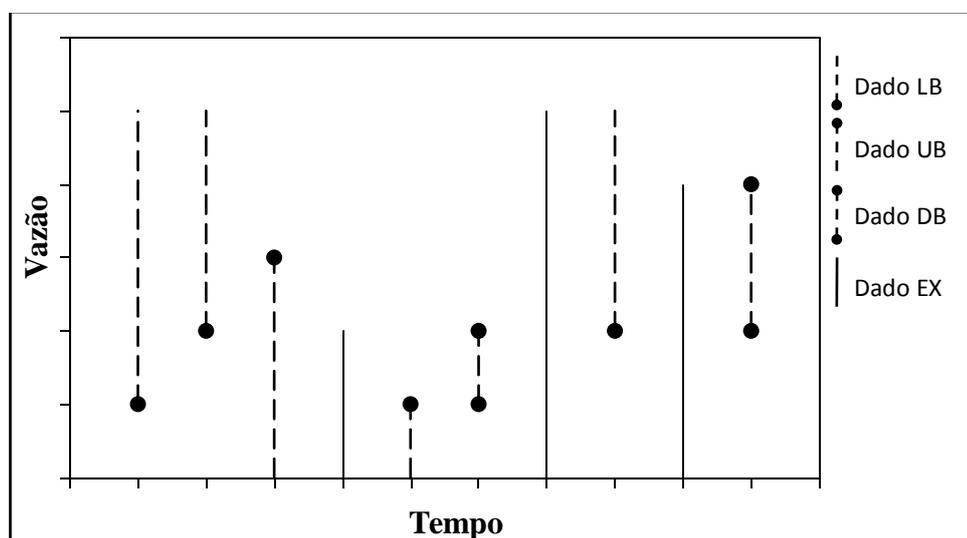


Figura 3.4 – Tipos de dados paleohidrológicos

3.3 O uso de dados paleohidrológicos e a existência de um limite superior para as vazões

A viabilidade de se usar dados paleohidrológicos, abrangendo centenas ou milhares de anos, para melhorar as estimativas na análise de cheias extremas tem sido objeto de debates contrastando a variabilidade natural do clima em grandes escalas temporais e as condições de

estacionariedade, independência e distribucionais requeridas para as variáveis aleatórias empregadas na análise de frequência de cheias (NRC, 1999; Baker, 2003). Mesmo no caso de amostras sistemáticas recentes (últimos 50 anos), essas condições nem sempre são atendidas. Associado a esse fato está a ausência de teste estatísticos robustos para a avaliação de não estacionariedades e para a verificação da condição IID (independente e identicamente distribuída) de variáveis hidrológicas, sobretudo quando incluídos dados não sistemáticos.

Reconhecendo a complexidade inerente a essa questão por um lado, mas verificando a freqüente situação da completa ausência de informação sistemática sobre cheias extremas por outro, compartilha-se, aqui, o ponto de vista de que “... *parece mais pragmático não desconsiderar ou ignorar os dados paleohidrológicos, mas, ao contrário, pelo menos avaliar o que esses dados têm a dizer sobre o fenômeno de cheias extremas*”, tal como expressado por Baker (2003). Nesse mesmo sentido, Hirschboeck (2003) argumenta que os registros de paleocheias fornecem um dos melhores indicadores de eventos extremos históricos. Além disso, as marcas de cheias preservam ou registram somente os eventos mais extremos ocorridos na bacia, os quais são, precisamente, as informações faltantes nas pequenas amostras de dados sistematicamente observados.

Além de aumentarem significativamente o período de dados a ser analisado, as paleocheias exercem um papel central na caracterização e definição de um limite natural para as vazões, que é o foco desta tese. Com efeito, a própria existência de um limite superior para as magnitudes de cheias e a capacidade de determiná-lo em uma dada região ou bacia são fontes de uma longa controvérsia em hidrologia de enchentes (Horton, 1936; Yevjevich, 1968; Klemeš, 1987; Yevjevich and Harmancioglu, 1987).

Essas controvérsias têm dividido a análise de cheias em duas linhas: a estocástica, na qual uma distribuição de probabilidades é ajustada aos dados, assumindo implicitamente a hipótese de que qualquer vazão futura, por maior que seja, tem probabilidade de ocorrência diferente de zero, ou seja, empregando distribuições ilimitadas superiormente, e a determinística, a qual é mais bem representada pela análise hidrológica com base na PMP/PMF.

A completa predominância de distribuições de probabilidades ilimitadas superiormente na análise de frequência convencional se deve principalmente às dificuldades de se estimar o limite superior com base em poucos dados observados. A esse 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} . No entanto, quando as enchentes extremas são o objeto de estudo, afirmar que as vazões em um rio podem ou não crescer até o infinito tem um profundo impacto na análise de frequência.

Nesse contexto, Enzel et al. (1993) avaliaram um total de 32.120 dados de vazão máxima anual em várias estações fluviométricas da bacia do rio Colorado, juntamente com dados de 25 estudos paleohidrológicos, e encontraram evidências que nesse local a hipótese da existência de um limite superior para as vazões pode ser confirmada. O trabalho desenvolvido por Enzel et al. (1993), um dos únicos e talvez o mais extenso no sentido de confirmação da existência de um limite, procurou confirmar a hipótese de que a aparente estabilização da curva envoltória de vazões nos EUA indica que essas variáveis são limitadas superiormente.

A curva envoltória de vazões, que será detalhada no item 4.3.2, reflete o conhecimento ou a experiência atual no que se refere a eventos extremos. Enzel et al. (1993) construíram a curva envoltória para sub-bacias do rio Colorado com áreas de drenagem variando entre 1 e 10^6 km² e dados de vazão com período de registro de até 70 anos. Na seqüência, os autores avaliaram dados de 25 estudos paleohidrológicos, sendo que a data de ocorrência das paleocheias estende até um período de 4.000 anos atrás. A comparação entre a curva envoltória para vazões recentes e as paleocheias indicou que a curva envoltória não foi alterada no período analisado. Ou seja, o grande aumento no comprimento dos registros de vazão, de aproximadamente 70 anos para mais de 4.000 anos, não elevou a curva envoltória. Assim, de acordo com os autores, a síntese regional dos dados modernos é suficiente para definir o limite superior das magnitudes de vazão.

O fato de a curva envoltória para a bacia do rio Colorado não se alterar com a inclusão de dados de até 4.000 anos atrás levou Enzel et al. (1993) a concluir que, naquela bacia, existe um limite físico para as vazões que tem persistido através dos últimos milênios. Em um trabalho recente, Jacoby et al. (2008) realizaram um estudo semelhante para a bacia do rio Nahal Arava, em Israel, e chegaram às mesmas conclusões. Embora esses estudos não erradiquem a controvérsia quanto à existência de um limite superior, eles constituem um valioso argumento para a hipótese adotada nesta tese.

Por fim, vale ressaltar uma afirmação do estudo clássico de Horton em hidrologia: *“a magnitude das cheias sempre aumenta com o aumento do período de retorno, mas aumentam em direção a um limite finito e não na direção do infinito”*.

3.4 Conclusão

Neste capítulo foram mostradas as principais características dos denominados dados não sistemáticos de vazões, os quais são objetos de estudo da paleohidrologia. Conforme visto, a paleohidrologia aborda a reconstrução de cheias não registradas pelos métodos convencionais de monitoramento de dados hidrológicos. A reconstrução de cheias permite avaliar inundações ocorridas até um período de 10.000 anos atrás.

As informações analisadas em estudos paleohidrológicos originam-se de evidências botânicas e geológicas da ocorrência de grandes cheias na bacia, além de registros jornalísticos e relatos pessoais. Essas informações possibilitam o estabelecimento de indicadores paleohidrológicos, tal como explicitado no item 3.2, que, após serem datados por técnicas diversas, constituem os elementos necessários para a modelagem hidráulica do curso d'água estudado. Como produto, a modelagem hidráulica fornece as paleovazões, que podem ser do tipo UB, LB, DB ou EX.

Conforme explicitado no item 3.3, as paleovazões são informações importantes que permitem, além de aumentar significativamente o período de dados disponível, avaliar a plausibilidade de um limite superior para as vazões de uma determinada bacia. Nesse sentido, Enzel et al. (1993) e Jacoby et al. (2008) mostraram que é plausível admitir a existência de um limite superior para as bacias do rio Colorado, nos EUA, e do rio Nahal Arava, em Israel.

A paleohidrologia é uma ciência complexa, que envolve o conhecimento aprofundado em geologia e geomorfologia, bem como dos processos de formação dos cursos d'água. Obviamente, não foram apresentados, no capítulo, todos os elementos que envolvem estudos paleohidrológicos. Pelo contrário, buscou-se uma descrição geral das principais características de tais estudos. O leitor interessado em detalhes sobre esse tipo de estudo devem remeter-se às referências bibliográficas citadas no capítulo, sobretudo aquelas produzidas por V. R. Baker.

4 HIDROLOGIA DE ENCHENTES

4.1 Introdução

A hidrologia de enchentes começou a entrar em evidência nas publicações científicas no início do século XX, quando o número e o tamanho dos reservatórios construídos nos EUA começaram a aumentar significativamente. Com o aumento dos reservatórios, verificou-se a necessidade da construção de estruturas extravasoras mais seguras ou, em outra frente, da melhor descrição dos riscos associados a estruturas existentes. Esse quadro impulsionou fortemente o desenvolvimento de métodos mais refinados de análise hidrológica. Myers (1967) dividiu a evolução da hidrologia de enchentes em quatro períodos distintos:

Período inicial: trata-se de um período anterior a 1900, onde a estimativa de cheia era feita pelo julgamento pessoal do engenheiro projetista de uma barragem. Tal julgamento tinha como base o conhecimento do engenheiro sobre a fluviometria local e os dados históricos de níveis d'água atingidos no local do projeto. Os dados históricos dos níveis d'água eram, na maioria das vezes, obtidos por marcas em construções próximas ao local ou através de informações fornecidas pelas pessoas que moravam na bacia. Myers (1967) salienta que os projetos feitos nessa época eram bem menos cautelosos devido, por um lado, à inexistência de dados monitorados e, por outro, pelo fato de a maioria dos reservatórios serem construídos em locais inabitados, sem risco à população a jusante.

Período da vazão regional: esse período teve início a partir da constatação de que dados monitorados em bacias distintas ao local do projeto constituíam uma importante informação para se estabelecer as possibilidades de ocorrência de cheias na bacia de interesse. O trabalho de Fuller, de 1914, que será discutido posteriormente neste capítulo, marca o início do uso de métodos padronizados em hidrologia de enchentes. Conceitos como risco de falha, regionalização de vazões, curvas envoltórias e limite superior para vazões foram introduzidos na prática da hidrologia nesse período. O problema desses métodos, à época em que foram propostos, é que as fórmulas e parâmetros eram estimados com base em uma pequena quantidade de dados hidrológicos, o que tornava incerta a generalização da metodologia proposta.

Período da análise de frequência: esse período teve início na década de 1930, com os trabalhos de Hazen, Gumbel e Horton. Embora amplamente utilizados atualmente, Myers (1967) aponta que a dificuldade de se prever vazões em amostras com grandes *outliers* fez

com que esses métodos fossem preteridos em relação àqueles de transposição de dados hidrometeorológicos.

Período da transposição de tormentas extremas: esse período é mais bem representado pelo aparecimento do conceito da PMP. De fato, o conceito de PMP surgiu da verificação de que a transposição de vazão era afetada por fatores muito mais complexos que a transposição de chuvas. A transposição de vazão é influenciada por fatores tais como a geomorfologia local e a topografia das bacias, enquanto a transposição de chuvas é menos afetada por fatores locais, refletindo principalmente as características climatológicas regionais. Isso faz com que a transposição de chuvas seja fisicamente mais realista e acurada. Esse período se estende com a consolidação dos métodos de cálculo de PMP e o uso do seu análogo para vazões, a PMF, como critério de projeto para grandes estruturas hidráulicas.

O trabalho de Myers (1967) aborda o desenvolvimento dos métodos de cálculo de cheia máxima até o fim da década de 1960. É interessante notar que, nesse trabalho, é feita uma previsão para o futuro da hidrologia de enchentes extremas, onde o autor ressalta os avanços já então estabelecidos pelo método hidrometeorológico da PMP e sua transformação em PMF, em relação aos estudos convencionais de frequência. O autor argumenta que a PMP, quando calculada de maneira precisa, com dados suficientes e métodos apropriados, fornece estimativas que se aproximam do máximo que a natureza pode produzir em termos de chuva. O autor argumenta, ainda, que o futuro da hidrologia de enchentes estaria no aperfeiçoamento do cálculo da PMP, na obtenção de dados mais precisos e no desenvolvimento de modelos chuva-vazão mais realistas. Passadas quatro décadas, verifica-se que o método hidrometeorológico de determinação de enchentes extremas é de uso corrente em diversos países do mundo, com destaque para seu emprego dominante como critério de cálculo de cheias de projeto de vertedores das grandes barragens de usinas hidrelétricas brasileiras (Eletrobrás, 1987). No entanto, ao contrário do previsto por Myers (1967), os métodos estatísticos também tiveram grande desenvolvimento a partir da década de 1970, por meio do aperfeiçoamento das técnicas de regionalização hidrológica e de estimação estatística.

Na seqüência, é feita uma descrição sucinta dos principais métodos utilizados em hidrologia. Não se pretende abordar cada método em detalhes, e sim estabelecer um quadro geral sobre suas principais características.

4.2 Métodos estocásticos

4.2.1 Análise de frequência convencional

A análise de frequência convencional ou clássica busca, principalmente, encontrar uma relação única entre a magnitude de um evento extremo e seu correspondente tempo de retorno, bem como um intervalo que descreve a incerteza relativa à estimativa. Essa relação é identificada a partir de informações obtidas de eventos observados em uma determinada seção fluvial.

A relação entre o evento extremo e o tempo de retorno pode ser obtida localmente através de modelos baseados nas séries dos máximos anuais, onde somente a vazão máxima observada em cada ano hidrológico é considerada, ou através de modelos baseados nas séries de duração parcial, onde os maiores valores acima de um determinado limiar são considerados. Em ambos os casos, a análise de frequência busca determinar a distribuição de probabilidades que caracteriza a amostra, bem como as propriedades dessa distribuição e os parâmetros que a descrevem.

Dentre os métodos de estimação dos parâmetros das distribuições, destacam-se três: o método dos momentos (MOM), o método de máxima verossimilhança (MLE) e o método dos momentos-L (MML).

a) Método dos momentos

O método dos momentos é o meio mais simples de estimação dos parâmetros. Seja uma densidade de probabilidades $f_X(x)$ determinada por k parâmetros, com momentos populacionais μ_i . O método dos momentos consiste em se igualar os k primeiros momentos populacionais aos seus respectivos estimadores amostrais. Formalmente, para uma distribuição $f_X(x/\theta_1, \theta_2, \dots, \theta_k)$, com momentos populacionais $\mu_i(\theta_1, \theta_2, \dots, \theta_k)$, tem-se:

$$\mu_i(\theta_1, \theta_2, \dots, \theta_k) = m_i ; \quad i = 1, 2, \dots, k, \quad (4.1)$$

onde m_i denota o momento amostral de ordem i .

A solução do sistema, definido por (4.1), fornece as estimativas dos parâmetros da distribuição.

b) Método da máxima verossimilhança

Considere uma amostra aleatória simples $\{X_1, X_2, X_3, \dots, X_n\}$, retirada de uma população com densidade de probabilidades $f_X(x/\theta_1, \theta_2, \dots, \theta_k)$. Se os elementos $X_1, X_2, X_3, \dots, X_n$ forem considerados independentes, a distribuição de probabilidades conjunta desses elementos pode ser escrita da seguinte forma:

$$L(\theta_1, \theta_2, \dots, \theta_k) = \prod_{i=1}^N f_X(x_i/\theta_1, \theta_2, \dots, \theta_k). \quad (4.2)$$

O método da máxima verossimilhança consiste em se encontrar o conjunto de parâmetros Θ que maximiza a função de verossimilhança dada em (4.2), ou seja, consiste na busca da solução do seguinte sistema de equações:

$$\frac{\partial L(\theta_1, \theta_2, \dots, \theta_k)}{\partial \theta_j} = 0; \quad j = 1, \dots, k. \quad (4.3)$$

c) Método dos momentos-L

O método dos momentos-L de estimação de parâmetros de distribuições de probabilidades é semelhante ao método dos momentos convencionais. No entanto, para a estimação dos parâmetros, são utilizadas as estatísticas-L e seus estimadores amostrais. Detalhes desse método de estimação são dados em Hosking e Wallis (1997).

De uma forma geral, e para grandes amostras, o MLE é mais eficiente que os demais, pois produz estimadores de menor variância. No entanto, para amostras pequenas os estimadores do MML são geralmente mais acurados. Além disso, o cálculo dos parâmetros pelo MML exige um esforço computacional significativamente menor que o MLE. Em geral, o método MOM, embora de construção mais simples, fornece os estimadores menos eficientes dentre os três métodos apresentados.

4.2.2 Análise de frequência regional

A regionalização de vazões é a inclusão, na análise de frequência, de dados de locais diferentes daquele onde se deseja obter estimativas de quantis. Conforme apontado por Hosking e Wallis (1997), uma vez que a análise regional incorpora mais dados que a análise local, ela tem o potencial de fornecer estimativas mais acuradas para os quantis de vazão.

Essa técnica de análise tem sido usada em hidrologia há muito tempo, sendo o método *index-flood*, apresentado por Dalrymple (1960), um dos exemplos mais antigos da sua aplicação em

estimativa de quantis de vazão. Vários métodos de regionalização têm sido indicados por agências nacionais ligadas a recursos hídricos para uso corrente em hidrologia de enchentes. Hosking e Wallis (1997) citam, por exemplo, o Boletim 17 (USWRC, 1982), nos EUA, e o método do NERC (1975), no Reino Unido. No caso do Boletim 17, admite-se que o logaritmo das vazões máximas anuais se distribui de acordo com a distribuição Pearson III, com o coeficiente de assimetria sendo obtido por meio de mapas regionais dessa medida. O método recomendado pelo NERC (1975) tem uma característica regional mais forte, já que divide o Reino Unido em 7 regiões, nas quais presume-se que as vazões máximas anuais de cada posto, após serem adimensionalizadas pela vazão média de cada local, se distribuem de acordo com uma mesma distribuição de probabilidade. Ou seja, admite-se uma distribuição de probabilidades regional e adimensional para cada região.

De acordo com Stedinger (2000), várias pesquisas têm demonstrado as vantagens dos métodos regionais com base no *index-flood* (Lettenmaier et al., 1987; Stedinger e Lu, 1995; Hosking e Wallis, 1997; Madsen e Rosbjerg, 1997). A idéia por trás do método *index-flood* é usar dados de bacias hidrologicamente similares para estimar uma distribuição adimensional para as cheias ou para a precipitação. Ou seja, “substituir o tempo pelo espaço” para compensar o fato de cada local possuir registros relativamente curtos. O conceito do método reside na hipótese de que diferentes locais na mesma região possuem a mesma distribuição para as cheias, a menos de um parâmetro de escala, ou *index-flood*, que reflete a magnitude média dos eventos máximos de precipitação ou de escoamento em cada bacia.

CPRM (2001) afirma que a análise de frequência regional não se restringe apenas à necessidade de transferência espacial de variáveis hidrológicas, mas também à obtenção de estimativas mais confiáveis de parâmetros e quantis, identificação de regiões com carência de postos de observação e verificação da consistência das séries hidrológicas.

De uma forma geral, conforme apontado por Naghettini e Pinto (2007), os métodos regionais de análise podem ser divididos em três classes: métodos de regionalização dos parâmetros da distribuição de probabilidades, métodos de regionalização do evento com um determinado risco e métodos de regionalização da curva adimensional de frequências.

No que se refere aos métodos utilizados atualmente, verifica-se uma profusão daqueles baseados nos conceitos do *index-flood*. Nesse sentido, Hosking e Wallis (1997) desenvolveram um conjunto metodológico para análise regional que utiliza as estatísticas-L de modo unificado e conciso em todas as etapas de sua aplicação. A regionalização por

momentos-L permite uma análise menos subjetiva do comportamento regional das variáveis hidrológicas e, por esse fato, tem sido amplamente utilizada em projetos de engenharia.

4.2.3 Análise de frequência com informações não sistemáticas

Os métodos convencionais de análise de frequência e os métodos de análise regional, da forma como foram expostos anteriormente, não permitem incorporar informações não sistemáticas ao conjunto de dados hidrológicos para análise de cheias. No entanto, conforme visto no capítulo 3, os dados não sistemáticos, além de aumentarem significativamente o período e a quantidades de dados de vazão, fornecem informações importantes a respeito das cheias máximas experimentadas em uma bacia. Assim, sua inclusão na análise de frequência é de fundamental interesse, uma vez que permite uma melhor descrição da variabilidade dos dados extremos de vazão.

A partir da década de 80 do século passado, os métodos de análise de frequência que incorporam informação não sistemática começaram a ser desenvolvidos. Dentre esses métodos destacam-se: o método dos momentos ponderados historicamente (MPH), apresentado em USWRC (1982); o método do algoritmo dos momentos esperados (EMA), apresentado por Cohn et al. (1997), e o método da máxima verossimilhança com informação não sistemática (MLE), apresentado por Stedinger e Cohn (1986). A seguir é feita uma descrição sucinta de cada método.

4.2.3.1 Método dos momentos ponderados historicamente (MPH)

O método dos momentos ponderados historicamente, também denominado método dos momentos ajustados, foi apresentado pelo USWRC (1982), na edição revisada do Boletim 17, com o emprego da distribuição Log-Pearson III. Assim como o método dos momentos, o método MPH consiste em igualar os momentos populacionais aos correspondentes momentos amostrais, calculados por meio da atribuição de “pesos” distintos aos elementos constituintes da amostra de vazões máximas anuais.

Nesse método, não são permitidos eventos censurados no período das cheias sistemáticas, nem cheias históricas de intensidade conhecida, inferiores a um limiar de referência y_U . A amostra é dividida em eventos de magnitude superior e inferior a um único limiar y_U , definido como a menor cheia histórica de intensidade conhecida. Dessa forma, eventos maiores que y_U , observados no período sistemático, são tratados como cheias históricas e o período de observações sistemáticas fica reduzido aos eventos menores que y_U .

O ajuste previsto no cálculo dos momentos amostrais visa preencher a porção do período histórico cujas magnitudes dos eventos são desconhecidas e, por definição, inferiores a y_U , com um número apropriado de cópias da porção inferior a y_U pertencente ao período sistemático. Esse preenchimento é alcançado aplicando-se um fator de ponderação (W) às cheias sistemáticas inferiores ao limiar de referência y_U , definido por:

$$W = \frac{N - N^{\bullet>}}{N_S^{\bullet<}} \geq 1 \quad (4.4)$$

Para o cálculo dos momentos amostrais pelo método MPH, utiliza-m-se:

- um fator de ponderação W para os $N_S^{\bullet<}$ eventos máximos anuais $x_i^<$ do período sistemático para representar artificialmente os $N_H^<$ anos de informação truncada do período histórico; e
- um fator de ponderação igual a 1 para os $N^{\bullet>} = N_H^{\bullet>} + N_S^{\bullet>}$ eventos maiores que o limiar de referência y_U .

Se p denota a probabilidade de não excedência de y_U , tal que $P[X < y_U] = P[Y < y_U] = p$, então, dado que $N^{\bullet>}/N$ representa a probabilidade empírica de excedência de y_U , p pode ser estimado por $\hat{p} = 1 - (N^{\bullet>}/N)$. Em seguida, as médias das cheias de intensidade conhecida são calculadas do seguinte modo:

- superiores ao limiar y_U (período histórico e sistemático):

$$m^{\bullet>} = \frac{1}{N^{\bullet>}} \left(\sum_{i=1}^{N_S^{\bullet>}} x_i^> + \sum_{j=1}^{N_H^{\bullet>}} y_j \right) \quad (4.5)$$

- inferiores ao limiar y_U (período sistemático):

$$m^{\bullet<} = \frac{1}{N_S^{\bullet<}} \left(\sum_{i=1}^{N_S^{\bullet<}} x_i^< \right) \quad (4.6)$$

A média empírica da amostra global, denominada média ponderada historicamente e denotada por \tilde{m} , é obtida ponderando-se as médias calculadas anteriormente pelas respectivas probabilidades de ocorrência:

$$\tilde{m} = \hat{p} \cdot m^{\bullet <} + (1 - \hat{p}) \cdot m^{\bullet >} \quad (4.7)$$

$$\tilde{m} = \frac{1}{N} \left(W \sum_{i=1}^{N_s^{\bullet <}} x_i^{\bullet <} + \sum_{i=1}^{N_s^{\bullet >}} x_i^{\bullet >} + \sum_{j=1}^{N_H} y_j \right) \quad (4.8)$$

A variância e o coeficiente de assimetria ponderados historicamente são obtidos de modo análogo e são expressos respectivamente pelas seguintes equações:

$$\tilde{s}^2 = \frac{1}{(N-1)} \left[W \sum_{i=1}^{N_s^{\bullet <}} (x_i^{\bullet <} - \tilde{m})^2 + \sum_{i=1}^{N_s^{\bullet >}} (x_i^{\bullet >} - \tilde{m})^2 + \sum_{j=1}^{N_H} (y_j - \tilde{m})^2 \right] \quad (4.9)$$

$$\tilde{g} = \frac{N}{\tilde{s}^3 (N-1)(N-2)} \left[W \sum_{i=1}^{N_s^{\bullet <}} (x_i^{\bullet <} - \tilde{m})^3 + \sum_{i=1}^{N_s^{\bullet >}} (x_i^{\bullet >} - \tilde{m})^3 + \sum_{j=1}^{N_H} (y_j - \tilde{m})^3 \right] \quad (4.10)$$

4.2.3.2 Método do algoritmo dos momentos esperados (EMA)

O método do algoritmo dos momentos esperados, apresentado por Cohn et al. (1997), é um procedimento de estimação de parâmetros que, tal como o método dos momentos, tem como princípio a igualdade entre os momentos populacionais e os correspondentes momentos amostrais. Ao contrário do método MPH, no método do algoritmo dos momentos esperados as cheias históricas e sistemáticas são tratadas da mesma maneira, e todos os tipos de informações censuradas podem ser utilizados. Para os eventos de magnitude desconhecida, os momentos são estimados por meio dos valores esperados da variável aleatória, descrita por uma função densidade de probabilidade truncada.

Nesse método, supõe-se que as cheias sistemáticas e históricas são descritas pela mesma função densidade de probabilidade $f_X(x)$. Assim, os momentos esperados para os vários casos de informações censuradas são dados por:

Período histórico

- cheia inferior a um limiar y_{Uj} :

$$\mu'_{H,r,j}^{\bullet <} = E[Y^r / Y < y_{Uj}] = \frac{1}{F_X(y_{Uj} / \Theta)} \int_0^{y_{Uj}} y^r f_X(y / \Theta) dy \quad (4.11)$$

- cheia superior a um limiar y_{Lj} :

$$\mu'_{H,r,j} > = E[Y^r / Y > y_{Lj}] = \frac{1}{1 - F_X(y_{Lj} / \Theta)} \int_{y_{Lj}}^{+\infty} y^r f_X(y / \Theta) dy \quad (4.12)$$

- cheia compreendida em um intervalo (y_{Lj}, y_{Uj}) :

$$\mu'_{H,r,j} < = E[Y^r / y_{Lj} < Y < y_{Uj}] = \frac{1}{F_X(y_{Uj} / \Theta) - F_X(y_{Lj} / \Theta)} \int_{y_{Lj}}^{y_{Uj}} y^r f_X(y / \Theta) dy \quad (4.13)$$

Período sistemático

- cheia inferior a um limiar x_{Ui} :

$$\mu'_{S,r,i} < = E[X^r / X < x_{Ui}] = \frac{1}{F_X(x_{Ui} / \Theta)} \int_0^{x_{Ui}} x^r f_X(x / \Theta) dx \quad (4.14)$$

- cheia superior a um limiar x_{Li} :

$$\mu'_{S,r,i} > = E[X^r / X > x_{Li}] = \frac{1}{1 - F_X(x_{Li} / \Theta)} \int_{x_{Li}}^{+\infty} x^r f_X(x / \Theta) dx \quad (4.15)$$

- cheia compreendida em um intervalo (x_{Li}, x_{Ui}) :

$$\mu'_{S,r,i} < = E[X^r / x_{Li} < X < x_{Ui}] = \frac{1}{F_X(x_{Ui} / \Theta) - F_X(x_{Li} / \Theta)} \int_{x_{Li}}^{x_{Ui}} x^r f_X(x / \Theta) dx \quad (4.16)$$

Portanto, para a amostra completa, os momentos em relação à origem de ordem r , denotados por \tilde{m}'_r , são calculados somando-se os dados de intensidade conhecida, conforme é feito para o cálculo dos momentos amostrais com base em uma amostra não censurada, e os momentos esperados, dados por (4.11) a (4.16). Assim:

$$\tilde{m}'_r = \frac{1}{N} \left(\begin{array}{l} \sum_{i=1}^{N_S^*} x_i^r + \sum_{i=1}^{N_S^<} \mu'_{S,r,i} < + \sum_{i=1}^{N_S^>} \mu'_{S,r,i} > + \sum_{i=1}^{N_S^<} \mu'_{S,r,i} < + \\ + \sum_{j=1}^{N_H^*} y_j^r + \sum_{j=1}^{N_H^<} \mu'_{H,r,j} < + \sum_{j=1}^{N_H^>} \mu'_{H,r,j} > + \sum_{j=1}^{N_H^<} \mu'_{H,r,j} < \end{array} \right) \quad (4.17)$$

A expressão para o cálculo de \tilde{m}'_r é função dos parâmetros desconhecidos Θ . Desta forma, o método do algoritmo dos momentos esperados parte de um processo iterativo para a determinação dos parâmetros, tal como mostrado no fluxograma da figura 4.1.

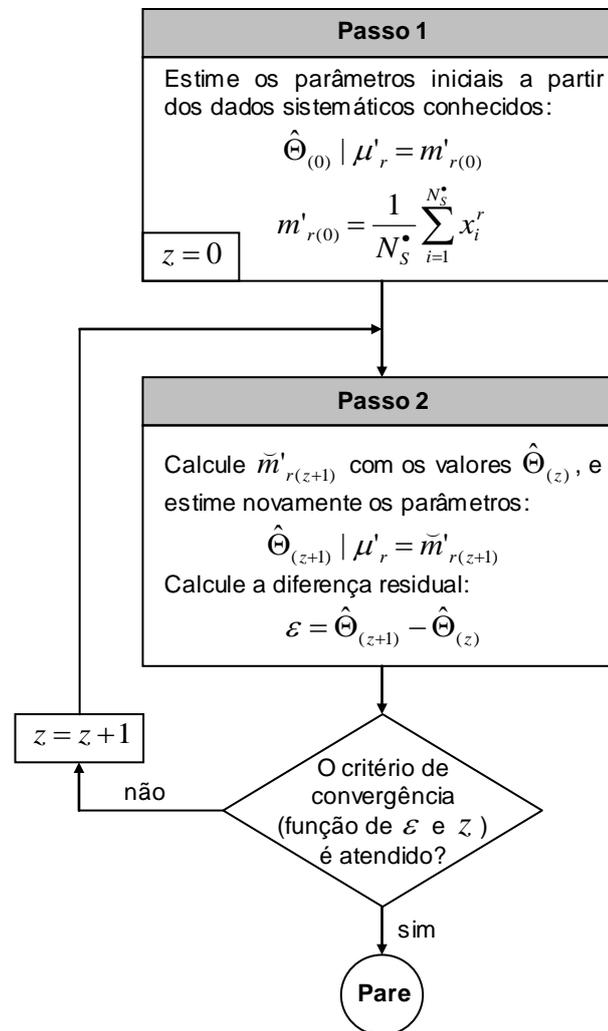


Figura 4.1 – Processo iterativo do método do algoritmo dos momentos esperados (Fonte: Lima, 2005)

4.2.3.3 Método da máxima verossimilhança com dados não sistemáticos (MLE)

Este método é semelhante ao método da máxima verossimilhança para dados sistemáticos discutido no item 4.2.1. Trata-se de buscar um conjunto de parâmetros que maximize a função de verossimilhança. A diferença reside na forma de construção da função de verossimilhança. Naulet (2002) desenvolveu uma função de verossimilhança que permite agregar todo tipo de informação na análise. Tal função tem a seguinte forma:

$$L = \left(L_H^\bullet \cdot L_H^< \cdot L_H^> \cdot L_H^\diamond \cdot L_S^\bullet \cdot L_S^< \cdot L_S^> \cdot L_S^\diamond \right), \quad (4.18)$$

onde o subscrito “ H ” indica os dados históricos, o subscrito “ S ” indica os dados sistemáticos, o sobrescrito “ c ” indica as cheias conhecidas, o sobrescrito “ u ” indica as cheias desconhecidas inferiores a um determinado limiar, o sobrescrito “ $^>$ ” indica as cheias desconhecidas superiores a um determinado limiar e o sobrescrito “ $^<$ ” indica as cheias desconhecidas pertencentes a um intervalo.

No capítulo 6, mostrar-se-á com detalhes a construção da função de verossimilhança com dados não sistemáticos.

4.3 Métodos determinísticos

4.3.1 Modelos de transformação chuva-vazão

A idéia central desta abordagem é analisar séries de vazões obtidas pela simulação de longas séries de precipitação em modelos chuva-vazão. Os modelos chuva-vazão são de especial interesse na análise hidrológica uma vez que, na maioria dos casos, os dados de chuva são mais abundantes e espacialmente melhor distribuídos que os dados de vazão. Além disso, os dados de chuva são menos sujeitos a erros amostrais e a interferências antrópicas, o que permite a simulação de cenários próximos às condições naturais de uma dada bacia.

A figura 4.2 mostra o esquema geral dos modelos de transformação chuva-vazão. Trata-se da representação matemática da fase terrestre do ciclo hidrológico. Os modelos chuva-vazão passaram a ser empregados em larga escala a partir da década de 1960 com a formulação do modelo Stanford (*Stanford Watershed Model* – SWM) por Crawford e Linsley (1966). A partir do modelo Stanford uma grande variedade de modelos foram propostos. Todos, no entanto, são representações matemáticas de diferentes partes do ciclo hidrológico apresentadas na figura 4.2. Uma descrição do desenvolvimento histórico do uso de modelos de transformação chuva-vazão é apresentada em Singh e Woolhiser (2002).

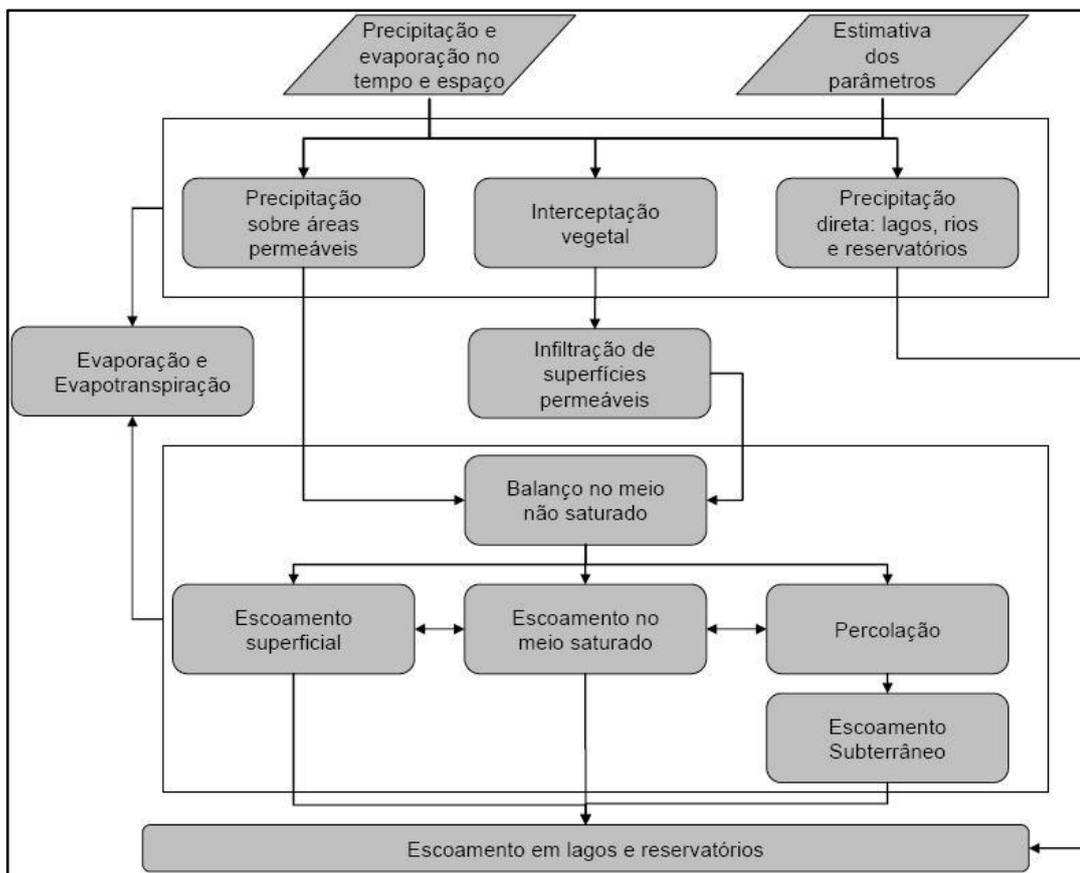


Figura 4.2 – Parte terrestre do ciclo hidrológico (adap. Tucci, 1998)

A resposta desses modelos de transformação é uma série de dados de vazão com a mesma extensão da série de precipitação. Nesse sentido, Lima (2004) desenvolveu um método que utiliza um modelo de geração estocástica de chuva e um modelo de transformação chuva-vazão de forma integrada, que permite obter séries extensas de vazão para uma determinada bacia. Tal modelo permite estabelecer uma série sintética de precipitações com as mesmas características estatísticas (variância, média, sazonalidade, dispersão, entre outras) dos valores observados. Na seqüência, a série sintética de chuva é transformada em uma série de vazões por meio de um modelo e essa, por sua vez, é avaliada sob a ótica dos métodos de análise de freqüência apresentados anteriormente.

Essa abordagem oferece vantagens no que diz respeito à possibilidade de análise em locais não monitorados. No entanto, é difícil avaliar o quanto da variabilidade dos dados de vazão é devido às incertezas do modelo chuva-vazão, o quanto é devido às incertezas dos dados de entrada (chuva e evaporação) e o quanto é devido às incertezas do modelo de geração de chuvas. Assim, esses métodos são pouco utilizados para análise de freqüência de vazões máximas anuais. Suas principais aplicações são no campo de planejamento de recursos hídricos, na análise de consistência de dados hidrológicos e no preenchimento de falhas em séries históricas.

4.3.2 Curvas envoltórias de vazão

As curvas envoltórias refletem o conjunto da experiência de observações dos recordes de vazões (maior vazão registrada) ou de estimativas de PMF de uma determinada região. Trata-se, de acordo com England (2005), de uma relação simples entre a máxima vazão de pico (recorde) e a área de drenagem.

Trabalhos recentes, tais como Vogel et al. (2001), Castellarin et al. (2007), Castellarin et al. (2005) e England (2005), têm buscado desenvolver métodos que possibilitem a análise dessas curvas envoltórias. As curvas envoltórias permitem avaliar o limite superior das vazões a partir dos recordes observados ou das estimativas de PMF na região da qual a bacia faz parte. Se por um lado as curvas envoltórias não permitem a definição exata do limite, uma vez que estão associadas a uma probabilidade de não excedência diferente de zero (Castellarin et al., 2005), por outro elas permitem avaliar um possível valor mínimo para o limite.

Além da aplicação natural do conceito da curva envoltória à estimação do limite superior para as vazões, England (2005) descreve pelo menos três outras aplicações: (1) em estudos de cheias máximas em bacias não-monitoradas; (2) em estudos onde se deseja comparar estimativas probabilísticas de vazões de projeto; e (3) como uma forma de avaliar a pertinência dos valores encontrados para a PMF.

A curva envoltória pode ser construída a partir das vazões recordes ou das estimativas de PMF de uma região ou, alternativamente, a partir dos valores de PMF unitária (descarga por unidade de área). A figura 4.3 mostra um exemplo dos dois casos para as estimativas de PMF nos Estados Unidos. Nessa figura, a curva envoltória é representada pela linha contínua. England (2005) aponta que usar descargas unitárias facilita, em alguns casos, a interpretação e o desenvolvimento matemático da curva envoltória.

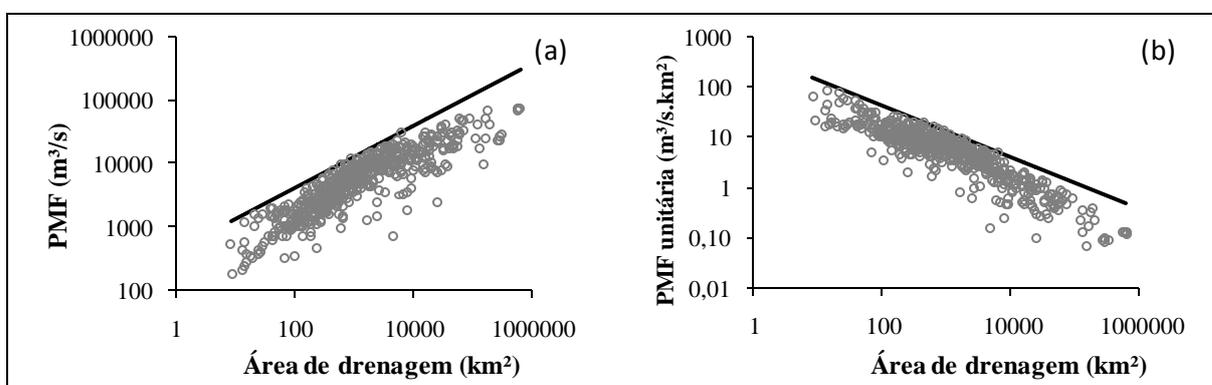


Figura 4.3 – Exemplo de curva envoltória (linha contínua) para estimativas de PMF nos Estados Unidos; o gráfico (a) mostra as estimativas de PMF sem transformação e o gráfico (b) mostra as PMF's unitárias.

As diversas formas de determinação das curvas envoltórias podem ser divididas em dois grupos principais: a) relações simples e diretas com as características físicas da bacia, notadamente a área de drenagem; e b) relações probabilísticas.

No primeiro grupo destaca-se a fórmula proposta por Myers (Jarvis e outros, 1936), dada por:

$$Q = CA^n, \quad (4.19)$$

onde Q é a vazão de pico (recorde ou PMF), A é a área de drenagem da bacia, C é um coeficiente que depende das características da bacia e n é um expoente inferior a um. Uma modificação da fórmula de Myers, de uso bastante freqüente, é dada por:

$$Q = 100b\sqrt{A}, \quad (4.20)$$

onde b é uma constante denominada de taxa de Myers, com valores variando entre 1 e 300. Com relação ao uso dessas fórmulas, Linsley et al. (1958) apontam que é bastante difícil escolher um valor adequado para a constante b e que fórmulas desse tipo não devem ser usadas em projetos de dimensionamento de estruturas de engenharia importantes.

England (2005) sugere que a fórmula de Myers deva ser usada como escolha inicial no estudo de curvas envoltórias. Além disso, recomenda-se que a mesma seja aplicada a bacias com áreas de drenagem inferiores a uma ordem de magnitude em relação à área de drenagem da bacia de interesse. Para outros casos, England (2005) aponta que o uso da fórmula de Crippen pode ser uma opção. Crippen (1982) propôs uma fórmula mais flexível para a curva envoltória que envolve a estimação, por vezes algo arbitrária, de cinco parâmetros. Sua equação é dada por:

$$Q = K_1 A^{K_2} (A^{C_1} + C_2)^{K_3}, \quad (4.21)$$

onde C e K são constantes empíricas. Crippen (1982) recomenda valores de 0,5 e de 5, respectivamente, para C_1 e C_2 , para curvas envoltórias de vazões recordes dos Estados Unidos. As demais constantes são determinadas a partir dos dados de vazões recordes.

Além dessas equações, muitas outras têm sido propostas (England, 2005), várias das quais envolvendo outros fatores além da área de drenagem, tais como a precipitação média, a altitude, entre outras. No entanto, é necessário aprofundar os estudos de curva envoltória no sentido de se desenvolver métodos que possibilitem escolher entre essas diversas equações.

No que se refere à estimação probabilística da curva envoltória, England (2005) lista os seguintes métodos:

- 1) Atribuição arbitrária de uma probabilidade: trata-se de definir uma probabilidade de excedência para a curva envoltória com base na probabilidade de excedência dos dados que a compõem. Com base nesse método, England et al. (2001) atribuíram um período de retorno entre 100 e 500 anos para as curvas envoltórias nos Estados Unidos.
- 2) Método de Fuller: Fuller (1914) analisou os dados de vazão de centenas de rios americanos e propôs um método empírico que relaciona a vazão média anual, a área de drenagem e o período de retorno. As equações a seguir resumem o método de Fuller:

$$\bar{Q} = CA^{0,8}, \quad (4.22)$$

$$Q = \bar{Q}(1 + 0,8\log_{10} T), \quad (4.23)$$

$$Q_{max} = Q(1 + 2A^{-0,3}), \quad (4.24)$$

onde \bar{Q} é a vazão média anual, Q é a maior vazão média diária provável para o período de T anos, Q_{max} é a vazão de pico diária e C é um coeficiente admitido como constante para uma bacia de interesse.

A principal crítica ao método de Fuller refere-se à quantidade e à qualidade dos dados utilizados na elaboração das fórmulas empíricas. Com efeito, o conjunto de estações utilizadas por Fuller possuem, em geral, poucos registros de vazão (menos de 100 anos), o que torna o método questionável quando de sua aplicação em projetos que envolvam grandes períodos de retorno ($T > 1.000$ anos). Além disso, não foram utilizadas estações de rios localizados em regiões áridas e semi-áridas, o que limita a aplicação do método para esses locais.

- 3) Método de Jarrett e Tomlinson: a idéia central do método proposto por Jarret e Tomlinson (2000) é atribuir uma probabilidade de excedência à curva envoltória por meio da análise de frequência dos dados sistemáticos e não sistemáticos de cada estação da região considerada. Os passos a seguir resumem o método.

- a) construção da curva de frequência, para cada bacia, considerando os dados sistemáticos e não sistemáticos;

- b) construção da curva envoltória para as vazões de pico, incluindo as paleocheias, das bacias da região. Pode-se utilizar o método de Myers, por exemplo;
- c) plotagem dos quantis de vazão obtidos em (a) para alguns períodos de retorno (100, 5.000 e 10.000 anos, por exemplo) e para bacias com áreas de drenagem selecionadas (100, 500, 1.000 km², por exemplo). Na seqüência, os quantis devem ser conectados de forma a verificar o comportamento do mesmo com a variação da área de drenagem e com o período de retorno; e
- d) comparação das curvas obtidas em (c) com a curva envoltória obtida em (b) de forma a verificar com qual curva de quantis a curva envoltória mais se assemelha.

England (2005) mostra que, por esse método, a curva envoltória nos Estados Unidos, para bacias com áreas de drenagem superiores a 100 km², tem um período de retorno por volta de 10.000 anos.

4) Método de Castellarin, Vogel e Matalas: Castellarin et al. (2005) propuseram um método para estimar a probabilidade de excedência da curva envoltória usando duas hipóteses básicas: (1) a região para a qual a curva envoltória é construída pode ser considerada homogênea, no sentido do método do *index-flood*; e (2) a curva envoltória é dada pela equação 4.19 usando uma transformação logarítmica. O método é construído com base na teoria dos recordes, desenvolvida por Chandler (1952), que aborda a análise do comportamento do maior valor contido em uma amostra. O método desenvolvido por Castellarin et al. (2005) é bastante complexo e não será descrito em detalhes aqui. No entanto, comparando com os demais métodos de análise de curva envoltória, esse parece ser o mais adequado do ponto de vista estatístico e conceitual. Uma aplicação desse método comparando os resultados para curvas envoltórias de vazões máximas e curvas envoltórias para PMF é dada em Vogel et al. (2007).

4.3.3 Precipitação máxima provável (PMP) e enchente máxima provável (PMF)

A PMP é definida, de acordo com WMO (1986), como o limite máximo teórico, fisicamente possível, da precipitação, para uma determinada duração, em uma determinada área geográfica, em uma determinada época do ano. A PMF, por sua vez, decorre da transformação da PMP em vazão por meio de modelos de transformação chuva-vazão. Embora não haja um consenso entre os pesquisadores quanto à existência de um limite para a precipitação, os conceitos de PMP e PMF têm sido correntemente empregados na prática da

engenharia hidrológica como elementos essenciais para o projeto de vertedores de grandes barragens e para controle de inundações de outras estruturas, tais como as centrais nucleares.

Berod et al. (1992) afirmam que o problema da PMP não está relacionado à existência de um limite superior para a precipitação, como implícito na sua definição, e sim no cálculo desse limite. Segundo os mesmos autores, a complexidade dos processos físicos atmosféricos e as características aleatórias dos campos de precipitação tornam impossível determinar o verdadeiro valor do limite superior. Assim, os métodos de estimação (e não de cálculo) da PMP consistem em técnicas de aproximação do valor do limite superior para a precipitação. Além disso, a sua estimação depende da disponibilidade de um conjunto de observações históricas, fazendo com que a PMP seja suscetível às incertezas impostas pelas amostras disponíveis. Por esse ponto de vista, a PMP deve ser vista como uma variável aleatória, cujo comportamento pode ser descrito por uma distribuição de probabilidades.

Os métodos de estimação da PMP podem ser classificados em dois grupos: (i) os métodos hidrometeorológicos, que agrupam aqueles baseados na maximização de tormentas; extremas observadas e aqueles que simulam condições extremas através de modelos de tormentas e (ii) os métodos estatísticos, desenvolvidos por Hershfield (1961) e posteriormente modificados por Koutsoyiannis (1999).

De acordo com Bertoni e Tucci (1993), os métodos hidrometeorológicos consideram que o total precipitado tende a crescer à medida que aumenta o teor de umidade do fluxo de ar que alimenta as tempestades. Admite-se que a coincidência entre a máxima precipitação e a máxima umidade ainda não ocorreu no passado, devido às flutuações dos demais fatores que influenciam o fenômeno, mas nada impede que tal coincidência venha a ocorrer no futuro. Os métodos hidrometeorológicos mais empregados na prática são descritos por Bertoni e Tucci (1993) da seguinte forma:

- 1) Maximização de tormentas severas: o método consiste em se selecionar as maiores tormentas na região de interesse e multiplicá-las por um fator de maximização f_m dado por:

$$f_m = \frac{w_m}{w_r} , \quad (4.25)$$

onde w_m é a altura de água precipitável nas condições hidrometeorológicas mais críticas para a época do ano em que se deseja estimar a PMP e w_r é a altura de água precipitável nas condições em que foi observada a precipitação que se deseja maximizar.

A massa de água precipitável por unidade de área (w) ou altura de água precipitável é calculada integrando-se a umidade específica ao longo da coluna de ar atmosférico da seguinte forma:

$$w = \frac{1}{g} \int_{P_0}^{P_1} q(P, t) dP , \quad (4.26)$$

onde g é a aceleração da gravidade, q é a umidade específica em função da distribuição da pressão e da temperatura no interior da coluna de ar e P_0 e P_1 são, respectivamente, as pressões nos níveis inferior e superior da coluna. De acordo com Bertoni e Tucci (1993), a máxima quantidade de água precipitável é obtida para uma temperatura igual à temperatura do ponto de orvalho, a uma pressão atmosférica de 1.000 *mb*.

- 2) Transposição de tormentas severas: esse método é aplicável quando o histórico de precipitação da bacia de interesse não é suficiente para caracterizar as condições mais severas possíveis para a chuva. A transposição é válida se existem reais condições de que as tormentas de outras bacias possam ocorrer na bacia de interesse. Nesse sentido, as bacias devem estar expostas à incursão das mesmas massas de ar e aos mesmos tipos de tormentas. A transposição é feita multiplicando-se a precipitação da bacia de origem por um fator que compensa as diferenças hidroclimatológicas entre as bacias. Uma opção para o cálculo desse fator é usar a relação entre água precipitável da tormenta de origem e a água precipitável da bacia de interesse, obtida à mesma época e sob as mesmas condições hidroclimatológicas.
- 3) Maximização das seqüências de tormentas severas: aplica-se às grandes bacias, nas quais a área de drenagem supera significativamente a extensão das tormentas. Para tanto, é preciso definir o sincronismo mais adverso entre as máximas precipitações acumuladas e o desenvolvimento e propagação das enchentes. Na definição desse sincronismo, analisa-se os registros históricos relativos às grandes enchentes, diagnostica-se as tormentas geradoras das máximas precipitações acumuladas e determina-se as maximizações a se realizar, abrangendo realocação e transposição de tormentas (Bertoni e Tucci, 1993).

Entre os métodos estatísticos, o mais utilizado é o procedimento de Hershfield (1961), que é o método estatístico sugerido pela Organização Meteorológica Mundial (WMO, 1986) para estimar a PMP. A principal vantagem do método é a sua facilidade de aplicação e o fato de serem levadas em consideração as características hidrometeorológicas locais por meio de parâmetros estatísticos. O procedimento é baseado na seguinte equação:

$$h_m = \bar{h}_n + k_m s_n, \quad (4.27)$$

onde h_m é a chuva máxima observada no local de interesse, \bar{h}_n e s_n são, respectivamente, a média e o desvio padrão da série de máximos anuais de precipitação do local e k_m é um fator freqüência.

Para avaliar o fator de freqüência, Hershfield (1961) analisou um total de 95.000 dados de máximos anuais de chuva pertencentes a 2.645 estações, das quais cerca de 90% localizadas nos Estados Unidos, e constatou que o valor máximo observado para o fator k_m é 15. Hershfield (1961) concluiu que a PMP pode ser estimada, para todos os casos, adotando $k_m = 15$ em (4.27). Posteriormente, Hershfield (1965) avaliou que o fator k_m varia de acordo com a duração da precipitação e com a média \bar{h}_n . De acordo com o autor, $k_m = 15$ é um valor muito alto para locais com fortes chuvas e muito baixo para locais áridos, além de ser alto para precipitações com duração inferior a 24 horas. Para contornar esse problema, Hershfield (1965) construiu uma curva onde k_m varia com a precipitação média anual e com a duração. Essa curva (mostrada na figura 4.4) e a equação 4.27 constituem a base do método estatístico de estimativa da PMP padronizado pela Organização Meteorológica Mundial.

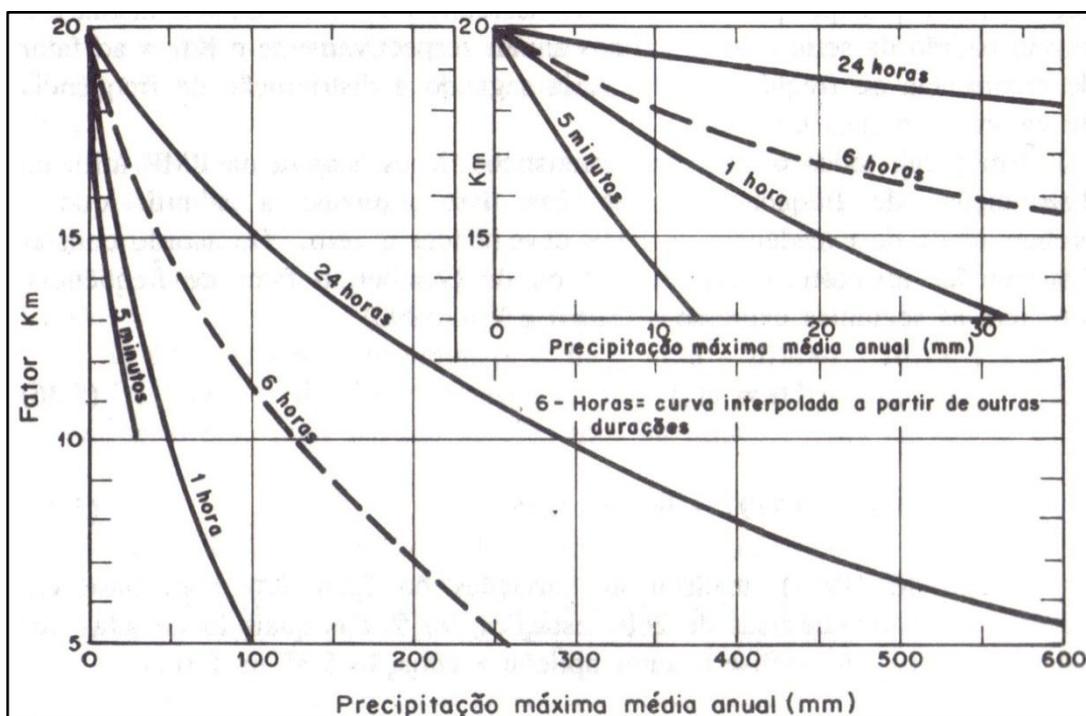


Figura 4.4 – Estimativa do fator k_m para o método estatístico de determinação da PMP (Fonte: Bertoni e Tucci, 1993)

Koutsoyiannis (1999) reavaliou o método de Hershfield e propôs uma alternativa para estimar a PMP. O autor demonstrou que os resultados obtidos por Hershfield não estão de acordo com

a hipótese de que existe um limite físico para a PMP, ou seja, um limite superior para o valor de k_m , e que o tratamento puramente estatístico da precipitação é uma alternativa mais viável. Além disso, usando o mesmo conjunto de dados utilizado por Hershfield, Koutsoyiannis (1999) mostrou que a estimativa de PMP de Hershfield pode ser obtida usando a distribuição GEV (Generalizada de Valores Extremos), com parâmetro de forma dado por uma função linear da precipitação média anual máxima e com um período de retorno de cerca de 60.000 anos. O método proposto por Koutsoyiannis, além de mais consistente sob o ponto de vista estatístico, tem a vantagem adicional de substituir completamente o uso dos gráficos empíricos propostos por Hershfield.

Por fim, a cheia máxima provável (PMF) é obtida pela transformação da PMP em vazão por modelos chuva-vazão, tais quais aqueles discutidos no item 4.3.1. Vale salientar que os parâmetros desses modelos devem ser calibrados para condições extremas de produção de vazão na bacia, refletindo, assim, a coincidência entre a maior chuva possível e as condições mais favoráveis de umidade do solo para a produção de vazão.

4.4 Métodos mistos

4.4.1 Método PVP

O método PVP (Pico-Volume-Precipitação) surgiu da tentativa de atender às recomendações do Comitê de Estudos dos Métodos de Estimação de Probabilidade de Cheias Extremas, formado pelo Conselho Nacional de Pesquisas norte-americano (NRC, 1988), no que se refere à proposição de novos métodos para análise de enchentes raras. NRC (1988) identificou três princípios gerais a serem perseguidos na proposição de novos métodos:

- a) “substituição do tempo pelo espaço”, uma indicação de uso preferencial das técnicas de estimação espacialmente regionalizada, em contraposição à estimação pontual;
- b) “introdução de maior estrutura aos modelos”, uma alusão, por exemplo, à possibilidade de se equacionar a transformação chuva vazão em condições extremas de transporte e armazenamento de umidade na bacia; e
- c) “ênfase para os extremos em detrimento, ou até mesmo exclusão, das características centrais”, de forma a evitar que as observações amostrais mais frequentes possam vir a distorcer a estimação das características essenciais da cauda superior das distribuições de probabilidade.

Naghetini et al. (1996) desenvolveram um método misto que atende às recomendações do NRC e que faz uso de três métodos: método pico-volume, método GRADEX e séries de duração parcial.

No contexto desse método, as vazões de pico excedentes sobre um valor limiar arbitrário u , denotadas por Y_i , e os volumes de cheia X_i , associados a essas excedências, são individualizados e modelados como um processo estocástico pontual marcado usando-se a representação de um processo composto de Poisson de intensidade $\lambda(t)$. A essência do método proposto por Naghetini et al. (1996) consiste em se estimar separadamente a função densidade marginal de probabilidade $g_u(x)$ dos volumes de cheia, para uma duração equivalente ao tempo de base da bacia, e a função densidade $f_{Y|X}(y/x)$ das vazões de pico condicionadas aos volumes.

Na seqüência, a função de distribuição de probabilidades das vazões de pico pode ser estimada por meio da integração do produto entre a densidade marginal de probabilidade dos volumes de cheia e a densidade das vazões de pico condicionadas aos volumes. Formalmente,

$$H_u(y_p) = \int_0^{y_p} \int_0^{\infty} f_{Y|X}(y|x) g_u(x) dx dy, \quad (4.28)$$

e

$$F_M(y_p) = \exp \left\{ - [1 - H_u(y_p)] \int_0^1 \lambda(t) dt \right\}, \quad (3.29)$$

onde $H_u(y_p)$ representa a função acumulada das excedências y_p sobre o valor limiar u e $F_M(y_p)$ a função acumulada anual das vazões de pico. A solução do sistema formado por (4.28) e (4.29) fornece a probabilidade anual correspondente a um quantil y_p , ou inversamente y_p em função do tempo de retorno.

Os princípios (a), (b) e (c) identificados pelo NRC são incorporados ao método proposto por Naghetini et al. (1996), principalmente na estimativa de $g_u(x)$, da seguinte forma: inicialmente, utiliza-se um modelo regional [princípio (a)], baseado em estatísticas amostrais de ordem superior [princípio (c)] de um conjunto de estações pluviométricas, para se testar a hipótese de exponencialidade da cauda superior da distribuição de probabilidades da altura de

precipitação, cuja duração deve ser tomada como igual ao tempo de base de uma bacia dentro da área em estudo. Trata-se de um teste baseado na razão das funções logaritmos de verossimilhança sob a hipótese $H_0 \{ \xi = 0 \}$ e sob a alternativa $H_1 \{ \xi \neq 0 \}$, onde ξ representa a estimativa regional do parâmetro de forma de uma distribuição generalizada de Pareto, da qual a exponencial é um caso particular ($\xi = 0$). Caso não se possa rejeitar a hipótese $H_0 \{ \xi = 0 \}$, atende-se então à premissa básica do método GRADEX, desenvolvida por Guillot e Duband (1967), para se transferir a informação hidrometeorológica para as curvas de frequência de volumes de cheia. De acordo com o método GRADEX, a distribuição de probabilidades dos volumes de cheia $g_u(x)$ pode ser deduzida da distribuição das precipitações de mesma duração efetuando-se a translação das distribuições acumuladas de uma distância fixa ao longo do eixo da variável. A distância de translação é função do *gradex*, ou parâmetro de escala, da distribuição de precipitações, a qual deve ter necessariamente um comportamento assintótico exponencial. Supõe-se que em condições de saturação e, portanto, para elevados tempos de retorno, qualquer incremento da altura de chuva irá provocar igual incremento do volume de cheia, desde que tomados sob a mesma duração; observa-se uma clara aplicação do princípio (b) do NRC.

A despeito de seus atributos potenciais, o método de Naghettini et al. (1996) apresenta algumas dificuldades para sua implementação em aplicações genéricas. A primeira refere-se à recomendação de extrapolar a relação pico-volume para além dos dados observados por meio de simulação chuva-vazão de um conjunto de tormentas transpostas para a bacia de interesse. Nesse particular, além da complexidade inerente à transposição de tormentas de uma região para outra, constata-se que, em diversos países e regiões, é incomum a existência de um extenso catálogo de eventos extremos de precipitação tal como aquele empregado por Naghettini et al. (1996) na aplicação efetuada para a bacia do rio Blue, localizada no estado americano de Oklahoma. A segunda dificuldade diz respeito ao modelo regional de estatísticas superiores usado para inferir o parâmetro de escala de cauda superior exponencial das precipitações de dada duração. De fato, a prescrição de usar somente os 10 ou 20% maiores valores amostrais, além de restringir consideravelmente os dados disponíveis para a inferência, introduz subjetividade na estimação do parâmetro de escala da cauda superior exponencial.

Visando generalizar a sua aplicação, Fernandes e Naghettini (2008) desenvolveram o método denominado de PVP, que preserva a estrutura geral da seqüência metodológica proposta por Naghettini et al. (1996), principalmente no que diz respeito à transferência de informação

hidrometeorológica por meio das hipóteses do método GRADEX. No método PVP são introduzidas duas modificações importantes: a primeira refere-se ao emprego do modelo regional TCEV (*Two-Component Extreme Value*) para a estimação do parâmetro de escala das precipitações de duração igual ao tempo de base e a segunda refere-se à modelação, em escala regional, da relação entre vazões máximas e volumes de cheia, de duração igual ao tempo de base, ou seja, da relação pico-volume.

O método PVP tem a vantagem, em relação à análise estatística convencional, de incorporar, de forma lógica, os três principais fatores que afetam as distribuições de probabilidade dos picos de vazão, a saber: a hidrometeorologia local, a transformação chuva-vazão e a hidráulica fluvial. Entretanto, cabe ressaltar que a inclusão de novas informações traz também incertezas, particularmente aquelas relacionadas às entradas do modelo de transformação chuva-vazão, seus parâmetros e sua estrutura.

O uso de séries de chuva para extrapolar a distribuição de probabilidades dos volumes escoados incorpora um número maior de dados à análise. Uma vez que os dados de chuva são mais abundantes e mais facilmente medidos, a estimativa da distribuição dos volumes pode ser feita de modo relativamente mais confiável. Alterações antropogênicas afetam, em grau menor, a precipitação, relativamente aos efeitos sobre a vazão escoada. Como consequência, os dados de precipitação são menos sujeitos a não estacionariedade e mais facilmente regionalizáveis que os dados de vazão.

A transformação chuva-vazão em eventos críticos é mais simples de se modelar do que em eventos comuns. Para valores altos de períodos de retorno, o clima exerce uma influência dominante, fazendo com que os demais fatores intervenientes tenham um papel relativamente secundário. Como consequência, a hipótese de proporcionalidade direta entre os incrementos de chuva e volumes escoados pode ser aceita. Assim, a principal hipótese do método GRADEX torna-se plausível.

Finalmente, a hidráulica fluvial, que é a forma que os excessos de chuva são propagados pela rede de canais da bacia, é representada pela relação pico-volume. A correta modelação da relação entre os picos de vazão e os volumes escoados pode incorporar, ainda que indiretamente, os principais fatores geomorfológicos que influem na hidráulica fluvial.

Uma vantagem adicional do método PVP é que a análise probabilística envolvendo chuva e vazão evita o absurdo de se ter um volume escoado superior ao precipitado, para uma mesma duração e um mesmo período de retorno.

4.4.2 Análise de frequência com o uso conjunto de distribuições limitadas superiormente, informações não sistemáticas e PMF

Botero (2006) propôs um modelo onde dois conceitos, virtualmente antagônicos, são utilizados em conjunto na estimação de cheias extremas. Por um lado a análise de frequência de dados sistemáticos e não sistemáticos, as quais são vistas como eventos estocásticos, e por outro, as estimativas de PMF, as quais são obtidas deterministicamente por meio de variáveis climatológicas e modelos de transformação chuva vazão.

A idéia central do trabalho de Botero (2006), também utilizada nesta tese, é que as vazões máximas anuais são variáveis limitadas, com um limite fisicamente finito, e devem ser analisadas sob esse aspecto. Assim, as distribuições de probabilidades limitadas superiormente são uma escolha natural para o modelo a ser utilizado na descrição dessas variáveis.

O principal obstáculo na modelagem de variáveis limitadas é a definição do limite superior. Essa dificuldade se deve, em parte, à pouca disponibilidade ou até ausência de dados, sobretudo aqueles relacionados a eventos extremos, nas bacias em estudo. Botero (2006) buscou contornar esse problema utilizando a PMF como uma variável auxiliar na definição do limite superior.

As distribuições utilizadas por Botero (2006) também serão empregadas nesta tese e, por essa razão, serão objeto de detalhamento no capítulo 6. Como será visto, são três as distribuições limitadas superiormente utilizadas para modelar as vazões de enchentes: a distribuição Log-Normal de 4 parâmetros (LN4), a distribuição de valores extremos transformada (TDF) e a distribuição de valores extremos do tipo IV (EV4). Botero (2006) lista três possibilidades para a estimação dos parâmetros dessas distribuições:

1) Método de máxima verossimilhança

Nesse caso, trata-se de encontrar, por algum método numérico, os parâmetros que maximizam a função de verossimilhança dada por (4.18). Botero (2006) verificou que o valor do limite superior (α) que maximiza a função de verossimilhança depende do tipo de dado incluído na análise e da distribuição utilizada. As seguintes conclusões foram estabelecidas:

a) Somente dados sistemáticos (EX)

Nesse caso, Botero (2006) verificou que não há como determinar o valor de α que maximiza a verossimilhança. Em alguns casos, dependendo dos demais parâmetros das distribuições analisadas, o valor de α pode ser o máximo observado ou o “máximo possível”, tomado como a PMF. É importante ressaltar que no método proposto por Botero (2006), diferentemente do método aqui proposto, que será visto no capítulo 6, está implícito que a PMF é exata e que seu valor é o limite superior para as vazões.

b) Inclusão de dados UB

Nesse caso, para todas as distribuições, o valor de α que maximiza a função de verossimilhança é dado pelo máximo amostral, o qual inclui os valores UB.

c) Inclusão de dados LB

Com a presença de dados EX e LB a função de verossimilhança, para todas as distribuições, tem um máximo quando α é igual ao máximo possível, ou seja, igual à PMF.

d) Inclusão de dados DB

Com a presença de dados EX e DB, a função de verossimilhança, para todas as distribuições, tem um máximo quando α é igual ao máximo observado.

Quando há uma combinação de diferentes tipos de dados não há como saber o valor de α que maximiza a função de verossimilhança. No entanto, Botero (2006), afirma que na presença de dados EX, UB e DB, os modelos EV4 e TDF têm um máximo quando α é igual ao máximo observado. Já o modelo LN4 é bastante dependente dos dados EX, sendo que o máximo da função de verossimilhança não está, necessariamente, no máximo amostral.

2) Atribuição de um valor extremo ao parâmetro α

Neste caso, pode-se atribuir ao limite superior o valor da PMF, que é, por suposição, o maior valor possível para as vazões em um determinado local. No entanto, o método de estimação da PMF está longe de ser um consenso entre os pesquisadores e, assim, a estimativa de α por meio desse valor certamente não irá refletir um valor inequívoco para o limite superior. A

despeito dessa consideração, Botero (2006) verificou que, na falta de informações paleohidrológicas, uma estimativa plausível para o limite superior pode ser a PMF.

Definido o valor de α , pode-se encontrar os demais parâmetros por meio do método da máxima verossimilhança.

3) Equação genérica

Kijko (2004) desenvolveu um método para a estimação da maior magnitude possível de um terremoto, o equivalente à PMF para as vazões de um rio, baseado na distribuição do recorde (maior valor) de uma variável. Tal método, denominado de equação genérica, é descrito a seguir.

Seja uma variável X pertencente ao intervalo (x_{\min}, x_{\max}) cuja distribuição acumulada de probabilidade é dada por $F_X(x)$. Dada uma amostra com n valores, a distribuição de probabilidade do valor correspondente ao máximo observado X_n é dada por:

$$F_{X_n}(x) = \begin{cases} 0 & \text{para } x < x_{\min} \\ [F_X(x)]^n & \text{para } x_{\min} \leq x \leq x_{\max} \\ 1 & \text{para } x > x_{\max} \end{cases} \quad (4.30)$$

O valor esperado de X_n é dado por:

$$E[X_n] = \int_{x_{\min}}^{x_{\max}} x \cdot d[F_X(x)]^n \quad (4.31)$$

Integrando (4.31) por partes, tem-se:

$$E[X_n] = x_{\max} - \int_{x_{\min}}^{x_{\max}} [F_X(x)]^n dx \quad (4.32)$$

Partindo do pressuposto que o melhor estimador para $E[X_n]$ é o máximo valor observado, conforme sugerido por Kijko (2004), a equação 4.32 torna-se:

$$x_{\max} = x_{\max, obs} + \int_{x_{\min}}^{x_{\max}} [F_X(x)]^n dx \quad (4.33)$$

Seja $\bar{\Theta}$ o conjunto de parâmetros das distribuições limitadas, excluindo o limite superior α . A estimação dos parâmetros, pelo método da equação genérica, é feita, iterativamente, utilizando (4.33), conforme as seguintes etapas:

- a) estima-se $\bar{\Theta}$ pelo método de máxima verossimilhança, fazendo $\alpha = \textit{máximo observado}$. Assim, a primeira estimativa para o conjunto de parâmetros é dada por $\{\bar{\Theta}_i, \alpha_i = \textit{máx.obs.}\}$;
- b) com os parâmetros $\{\bar{\Theta}_i, \alpha_i\}$, estima-se o valor de α_{i+1} resolvendo a equação genérica (4.33);
- c) com o valor de α_{i+1} estima-se os novos parâmetros $\bar{\Theta}_{i+1}$ pelo método de máxima verossimilhança; e
- d) repete-se os passos b e c até que algum critério de convergência seja atingido. Por exemplo, até que se tenha $|\alpha_{i+1} - \alpha_i| \leq \delta$, onde δ é um número pequeno ($<10^{-4}$).

O método da equação genérica não pode ser aplicado quando a amostra inclui dados LB, já que, nesse caso, não é possível estabelecer a magnitude do máximo amostral. Outro problema com esse método se refere à convergência do algoritmo iterativo mostrado anteriormente. Com efeito, não há qualquer garantia de convergência e, em alguns casos, o limite superior pode tender para valores fisicamente impossíveis ou, até mesmo, para o infinito.

Botero (2006) aplicou os três métodos de estimação de parâmetros a seis bacias espanholas. De uma forma geral, a distribuição EV4 com o limite superior fixado na PMF e no máximo amostral foi a que apresentou os melhores resultados. Por outro lado, a distribuição TDF, independente do método de estimação dos parâmetros, foi a que apresentou os piores resultados. A distribuição LN4 apresentou resultados variados, dependentes fortemente do tipo de dado presente na amostra.

O método desenvolvido por Botero (2006) pode ser visto com um ponto inicial do método proposto nesta tese. No capítulo 6 será vista uma nova abordagem para a análise de eventos extremos de vazão que preserva os conceitos de base utilizados por Botero (2006), porém dentro do contexto da análise de frequência bayesiana, a qual permite uma interpretação mais abrangente das incertezas presentes na estimação de quantis de cheias extremas.

4.5 Conclusão

Neste capítulo, foram descritas as principais técnicas para a análise de eventos hidrológicos extremos. Buscou-se, sempre que possível, estabelecer a perspectiva histórica do desenvolvimento de tais técnicas. Esse campo de estudo da hidrologia está em constante evolução, contando com uma série de técnicas não abordadas aqui. Observa-se, entretanto, que, os métodos aqui descritos são os mais utilizados na prática e, por isso, foram detalhados ao longo do capítulo.

Dentre os métodos descritos destacam-se: a análise de frequência regional, a análise de frequência com informações não sistemáticas, a PMP/PMF meteorológica e o PVP. Esses métodos buscam a melhoria das estimativas de vazões extremas pelo aumento do conjunto de dados e informações disponíveis para a análise. Por um lado, tem-se a substituição do tempo pelo espaço nos métodos regionais, pelo outro, tem-se o aumento do período analisado nos métodos que incluem informações não sistemáticas. O método hidrometeorológico da PMP/PMF se destaca pelo fato da PMP ser obtida por meio da maximização de observações regionais de variáveis hidrometeorológicas e se constituir em um limite teórico para a produção local de precipitações.

Não se buscou, aqui, esgotar a análise dos métodos no que se refere às suas qualidades e deficiências. De modo geral, um exaustivo estudo comparativo dos métodos de análise de eventos hidrológicos extremos, com critérios absolutamente objetivos, permanece dificultado pela escassez ou pela própria inexistência de dados inequívocos sobre esses eventos e pela sempre presente necessidade de extrapolação.

Por fim, foi descrita a técnica proposta por Botero (2006), que contém alguns princípios seguidos nas etapas metodológicas desta tese. Conforme visto, a principal deficiência do método proposto por Botero (2006) é a definição do limite superior para as vazões. Com efeito, nesse caso, o limite superior é fixado em um valor extremo ou estimado pela equação genérica, que frequentemente fornece estimativas não realistas para o limite superior. Com isso, as incertezas relativas ao limite superior não são adequadamente avaliadas e o receio, por parte do analista, quanto à especificação de uma vazão com probabilidade de excedência igual a zero não é dirimido. Ciente dessas limitações, são propostas novas formas de análise do limite superior no capítulo 6.

5 ANÁLISE DE FREQUÊNCIA BAYESIANA

5.1 *Conceitos Gerais*

De acordo com Brooks (2003), os métodos bayesianos tiveram seu início em 1763, quando o teorema de Bayes - Rev. Thomas Bayes (1701-1761) - foi apresentado à Royal Statistical Society, em Londres. Os conceitos bayesianos despertaram o interesse de pesquisadores renomados, como Laplace, Gauss e Pearson, e dominaram o pensamento estatístico no século XIX. No início do século XX, os métodos bayesianos perderam importância devido, sobretudo, à oposição de pesquisadores como Neyman e Fisher, que tinham objeções filosóficas quanto à subjetividade da abordagem bayesiana. Mesmo assim, como aponta Brooks (2003), Jeffreys, Savage, Lindley, de Finetti, entre outros, continuaram a advogar em favor dos métodos bayesianos, desenvolvendo-os e indicando-os como alternativa para sanar as deficiências da abordagem frequentista.

A partir do final da década de 80 do século passado, a abordagem bayesiana voltou a ter importância no cenário das pesquisas envolvendo métodos estatísticos. Esse ressurgimento foi devido, principalmente, ao rápido desenvolvimento computacional ocorrido nessa década e ao crescente desejo de se descrever e modelar processos cada vez mais complexos, para os quais a teoria frequentista não oferecia meios para tal (Brooks, 2003). Com a abordagem bayesiana recebendo mais atenção em pesquisas em Estatística, mais ferramentas computacionais foram desenvolvidas e meios mais flexíveis de inferência foram apresentados. Assim, atualmente, a abordagem bayesiana se constitui em uma estrutura de análise que vai ao encontro da crescente complexidade das pesquisas científicas do século XXI.

Tanto a Escola Clássica quanto a bayesiana entendem que as incertezas sobre os objetos aleatórios devem ser mensuradas via probabilidade. No entanto, no contexto clássico, a medida de probabilidade capta a variabilidade inerente ao processo e, no contexto bayesiano, tal medida captura o desconhecimento do indivíduo sobre o objeto em estudo, ou seja, a probabilidade é dita subjetiva. Como consequência, surge uma das diferenças entre a abordagem bayesiana e a frequentista, que é o modo como cada uma vê os parâmetros dos modelos.

Os frequentistas vêem o parâmetro θ como um valor fixo (não variável) e tentam estimar esse valor, por exemplo, maximizando a função de verossimilhança. Os bayesianos, por outro lado, também vêem os parâmetros como um valor fixo e, por serem desconhecidos, são aleatórios e, assim, estão associados a uma distribuição de probabilidades a qual resume o conhecimento

que se tem sobre essas quantidades. Os bayesianos também acreditam que existe um valor verdadeiro para o parâmetro e utilizam sua correspondente distribuição *a posteriori* para obter estimativas pontuais de um certo parâmetro θ , por exemplo. Assim, à medida que o conhecimento sobre o parâmetro cresce, espera-se que a incerteza sobre ele diminua. No limite, pelo menos em teoria, o total conhecimento sobre o parâmetro implicaria em uma distribuição degenerada, ou seja, o parâmetro assumiria um único valor com probabilidade 1.

Com isso, do ponto de vista bayesiano, uma quantidade aleatória é uma quantidade desconhecida que pode variar e assumir diferentes valores (uma variável aleatória, por exemplo) ou simplesmente ser uma quantidade fixa sobre a qual há alguma ou nenhuma informação disponível (um parâmetro, por exemplo). As incertezas sobre essas quantidades 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. Em termos formais, seja θ um parâmetro de interesse assumindo valores no espaço paramétrico Θ . Denote por \mathcal{H} a informação prévia e o conhecimento sobre θ . Baseado em \mathcal{H} , a incerteza sobre θ é resumida pela distribuição *a priori* $\pi(\theta/\mathcal{H})$ que descreve o estado de conhecimento sobre a quantidade aleatória antes de qualquer dado ter sido observado. Se Θ é um conjunto finito, essa inferência representa a chance de ocorrência de cada valor de θ . É importante salientar que a distribuição *a priori* não modela a variabilidade do parâmetro, que é uma quantidade fixa, e sim o grau de conhecimento do analista sobre o verdadeiro valor do parâmetro.

Em geral, \mathcal{H} não contém toda informação relevante sobre o parâmetro, e nesse 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 neste caso não há razão para se fazer inferência, o que não ocorre em situações reais. Se a informação contida em \mathcal{H} não é suficiente, um experimento deve ser realizado para obter informações adicionais sobre o parâmetro. Suponha que uma variável aleatória X , a qual é relacionada a θ , possa ser amostrada ou observada. Antes de amostrar de X e supondo que o atual valor de θ seja conhecido, a incerteza sobre a quantidade X é resumida pela função de verossimilhança $f(X/\theta, \mathcal{H})$. A função de verossimilhança fornece a probabilidade de ocorrência de cada amostra particular x de X , no

caso de θ ser o valor verdadeiro do parâmetro. Após realizar o experimento, o conhecimento *a priori* sobre θ deve ser atualizado usando a nova informação x . A ferramenta usual para atualizar a distribuição *a priori* é o teorema de Bayes. De acordo com esse teorema, a distribuição *a posteriori*, que agrega o conhecimento atualizado sobre θ , é dada por:

$$\pi(\theta/x, \mathcal{H}) = \frac{f(x/\theta, \mathcal{H})\pi(\theta/\mathcal{H})}{f(x/\mathcal{H})}, \quad (5.1)$$

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

$$f(x/\mathcal{H}) = \int_{\Theta} f(X/\theta, \mathcal{H})\pi(\theta/\mathcal{H})d\theta. \quad (5.2)$$

A distribuição *a posteriori* calculada em (5.1) descreve a incerteza sobre θ depois de se observar os dados, ou seja, $\pi(\theta/X, \mathcal{H})$ é a inferência *a posteriori* sobre θ , a partir da qual é possível estabelecer qualquer característica de θ .

No que se refere 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 e apontam que a abordagem bayesiana oferece a solução mais coerente para tal. Nesse mesmo 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 faz sentido argumentar se essa abordagem oferece estimativas mais precisas que a freqüentista. Em outras palavras, a análise de eventos extremos em engenharia é um problema de predição e não de estimação paramétrica.

Ainda de acordo com McRobbie (2004), antecipar ou avaliar a probabilidade de ocorrência de eventos futuros, os quais estão 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 freqüentista, 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. Nesse contexto, há uma subjetividade intrínseca sobre o 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. Assim, 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 existentes sobre o evento (Vick, 2002).

5.2 Estimação bayesiana e intervalos de credibilidade

De acordo com Bernardo e Smith (1994), a estimação bayesiana é um problema de decisão. Assim, para a estimação de um determinado parâmetro θ , a inferência bayesiana requer a especificação de uma função de perda $\mathcal{L}(\delta, \theta)$, a qual representa o erro, ou penalidade, associado à escolha de δ como estimador de θ . Nessa abordagem busca-se um estimador que minimiza o denominado risco de Bayes, o qual é definido da seguinte forma:

$$RB = \iint \mathcal{L}(\delta, \theta) f(x/\theta) \pi(\theta) dx d\theta, \quad (5.3)$$

onde a perda é integrada em x e em θ . Uma inversão na ordem de integração (veja Robert e Casella (2004), para detalhes dessa transformação) permite avaliar o estimador de θ em termos da perda esperada *a posteriori*. Ou seja, o estimador δ_B de θ é tal que:

$$\delta_B = \min_{\delta} E[\mathcal{L}(\delta, \theta) / x] = \min_{\delta} \int \mathcal{L}(\delta, \theta) \pi(\theta / x) d\theta. \quad (5.4)$$

A escolha da função de perda é feita de forma subjetiva e reflete o modo como o decisor acha justo ser penalizado por suas decisões. As principais funções de perda utilizadas na estimação de parâmetros são, de acordo com Bernardo e Smith (1994), as seguintes:

- Quadrática: nesse caso tem-se que $\mathcal{L}(\delta, \theta) = (\delta - \theta)^2$ e o estimador de Bayes para θ é a média *a posteriori* $E[\pi(\theta / x)]$, admitindo que a média existe;
- Valor absoluto: nesse caso tem-se que $\mathcal{L}(\delta, \theta) = |\delta - \theta|$ e o estimador de Bayes para θ é a mediana de $\pi(\theta / x)$, admitindo que a mediana existe; e
- Zero-Um: nesse caso tem-se que $\mathcal{L}(\delta, \theta) = \mathbf{1}_{(\delta \neq \theta)}$, onde $\mathbf{1}_{(a)}$ é a função indicadora, e o estimador de Bayes para θ é a moda de $\pi(\theta / x)$, admitindo que a moda existe.

Robert e Casella (2004) citam duas dificuldades relacionadas com o cálculo δ : a primeira é que a distribuição *a posteriori* de θ , $\pi(\theta/x)$, em geral, não tem uma forma analítica fechada; e a segunda é que, em muitos casos, a integração em (5.4) não pode ser feita analiticamente. No item 5.3, serão vistos alguns métodos que buscam contornar essas dificuldades.

Uma importante questão referente à estimação de parâmetros que mostra as vantagens da análise bayesiana sobre a freqüentista é a forma como cada uma avalia as incertezas em relação à escolha do estimador. Com efeito, na análise freqüentista, esse problema é abordado através do princípio da repetição da amostra, sendo que o desempenho do estimador, para uma única amostra, é avaliado a partir do comportamento esperado de um conjunto hipotético de amostras coletadas sob condições idênticas, supondo ser possível realizar tal experimento. Também baseia-se no princípio da repetição da amostra a construção do intervalo de confiança (IC) freqüentista. A confiança de um IC para um parâmetro θ é interpretada como o percentual de intervalos que, se construídos com base em dados coletados em condições idênticas, conteria o verdadeiro valor de θ . Na Escola Clássica, os parâmetros dos modelos são vistos como quantidades fixas, para as quais não se pode atribuir probabilidade. Desta forma, não faz sentido interpretar um IC como sendo “a probabilidade do verdadeiro valor do parâmetro pertencer ao intervalo de confiança”. A figura 5.1 exemplifica a interpretação gráfica do intervalo de confiança freqüentista para a média aritmética de uma amostra aleatória simples (AAS) extraída de uma população Normal de média μ e desvio padrão σ .

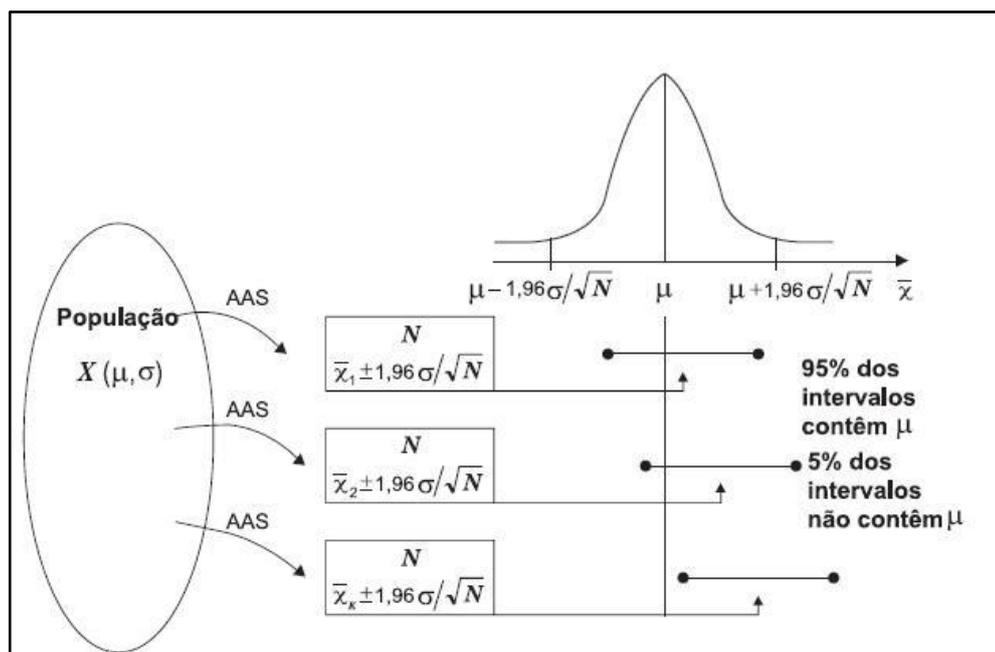


Figura 5.1 – Interpretação gráfica do intervalo de confiança freqüentista (Fonte: Naghettini e Pinto, 2007)

Por outro lado, a abordagem bayesiana fornece uma estrutura mais natural para avaliar as incertezas na estimativa dos parâmetros. Com efeito, a variância da distribuição *a posteriori* dos parâmetros fornece uma medida direta da incerteza associada a essa quantidade. Um indicador mais apropriado dessa incerteza é o intervalo de credibilidade, o equivalente bayesiano para o intervalo de confiança. O intervalo de credibilidade de $100(1-\alpha)\%$ para um parâmetro θ é construído com base na distribuição *a posteriori* de θ e, portanto, leva em conta a única amostra que foi de fato observada. Como o parâmetro é um objeto aleatório, o intervalo de credibilidade é aquele no qual θ está com probabilidade $(1-\alpha)$. Não há aqui nenhuma suposição sobre a possibilidade de replicação da amostra original. Além disso, a interpretação do intervalo de credibilidade é bem mais natural que a interpretação para seu análogo freqüentista.

O intervalo de credibilidade pode ser construído para qualquer quantidade aleatória e não somente para os parâmetros do modelo. Seja ω uma quantidade aleatória e $p(\omega)$ a distribuição de probabilidade (*a priori*, *a posteriori* ou preditiva) dessa quantidade. O intervalo de credibilidade para ω , com $(1-\alpha)$ de probabilidade, é o intervalo (L, U) tal que:

$$\int_L^U p(\omega)d\omega = 1 - \alpha, \quad (5.5)$$

onde L e U são, respectivamente, o limite inferior e superior do intervalo. Claramente, para um quantil α fixo não há um valor único para os limites L e U , mesmo no caso de $p(\omega)$ ser unimodal. Assim, adota-se o intervalo de mais alta densidade ou, brevemente, intervalo HPD do acrônimo em inglês *Highest Probability Density*. O intervalo HPD é definido em Bernardo e Smith (1994) da seguinte forma:

um intervalo $I \subseteq \Omega$, onde Ω é o domínio de ω , é o intervalo de mais alta densidade (HPD) a um nível $100(1-\alpha)\%$ para ω , com relação a $p(\omega)$ se:

- i) $P(\omega \in I) = 1 - \alpha$; e
- ii) $p(\omega_1) \geq p(\omega_2)$ para todo $\omega_1 \in I$ e $\omega_2 \notin I$, exceto possivelmente para algum subconjunto de Ω com probabilidade zero.

O intervalo HPD é o mais curto dos intervalos com massa $(1-\alpha)$, quanto mais curto for o intervalo HPD, mais certeza se tem sobre ω .

5.3 Métodos de cálculo

A principal dificuldade na aplicação da teoria bayesiana é o cálculo da constante de normalização ou distribuição preditiva *a priori* dada em (5.2). Mais precisamente, para se fazer qualquer inferência sobre o modelo (momentos, quantis, intervalos de credibilidade etc.), é necessário o cálculo do valor esperado de uma função h sobre a distribuição *a posteriori* dos parâmetros. Formalmente, tem-se:

$$E[h(\theta)/x] = \frac{\int_{\Theta} h(\theta)f(x/\theta)\pi(\theta)d\theta}{\int_{\Theta} f(x/\theta)\pi(\theta)d\theta} = \int_{\Theta} h(\theta)\pi(\theta/x)d\theta. \quad (5.6)$$

A distribuição h varia de acordo com a inferência que se deseja fazer. No caso da estimação pontual, h pode representar uma das funções de perda discutidas anteriormente. No caso de predição de valores futuros da variável x , h pode representar a distribuição de x_{n+1} dado θ . No caso de vazões máximas anuais, por exemplo, h pode representar o quantil com um determinado período de retorno T . Neste caso, sejam X as vazões máximas anuais cuja função densidade de probabilidade é representada por $k(x/\theta)$ e a função de probabilidade acumulada é representada por $K(x/\theta)$. O quantil *a posteriori* de X , representado por x_T , para um período de retorno de T anos dado por:

$$T = \frac{1}{1 - P(X < x_T)} = \frac{1}{1 - K(x_T/\theta)}, \quad (5.7)$$

é expresso por:

$$x_T = E[K^{-1}(p/\theta)/x] = \int_{\Theta} K^{-1}(p/\theta)\pi(\theta/x)d\theta, \quad (5.8)$$

onde $K^{-1}(p/\theta)$ é a função inversa de $K(x/\theta)$ e $p = P(X < x_T) = 1 - \frac{1}{T}$.

O cálculo analítico de integrais do tipo mostrado em (5.6) é impossível na maioria das aplicações práticas, especialmente nos casos multidimensionais. O cálculo dessas integrais multidimensionais pode ser evitado utilizando algoritmos de amostragem tais como aqueles

que empregam a integração de Monte Carlo via cadeias de Markov – MCMC do acrônimo 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 $\pi(\theta / X)$ em (5.1), através de uma cadeia de Markov construída de forma 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 θ , $\pi(\theta / X)$, a estimativa do valor esperado em (5.6) é obtida por integração de Monte Carlo da seguinte forma:

Seja $\theta_1, \theta_2, \dots, \theta_m$ uma amostra da distribuição *a posteriori* de θ . Uma aproximação do valor esperado mostrado em (5.6) pela integração de Monte Carlo é dada por:

$$E[h(\theta) / x] \approx \frac{1}{m} \sum_{i=1}^m h(\theta_i). \quad (5.9)$$

Assim, a média populacional de h é estimada pela média da amostra gerada da distribuição *a posteriori*. Quando a amostra $\{\theta_i\}$ é independente, a lei dos grandes números garante que a aproximação pode ser feita de modo tão acurado quanto se queira, aumentando-se o tamanho m da amostra (Gilks et al., 1996).

Em geral, conforme aponta Gilks et al. (1996), obter amostras independentes de $\pi(\theta / X)$ não é fácil, uma vez que $\pi(\theta / X)$ pode ter formas bastante complexas. No entanto, $\{\theta_i\}$ não precisa ser necessariamente independente. Basta que $\{\theta_i\}$ seja gerada por um processo que amostra sobre todo o suporte de $\pi(\theta / X)$ em proporções corretas, o que pode ser feito por uma cadeia de Markov que tem $\pi(\theta / X)$ como sua distribuição estacionária.

Uma cadeia de Markov é um processo estocástico $\{\theta_t, t \in T, \theta_t \in S\}$, onde $T = \{1, 2, \dots\}$ e S representa o conjunto de possíveis estados para θ , que respeita a seguinte condição:

$$P(\theta_{t+1} \in A / \theta_t, \theta_{t-1}, \dots, \theta_0) = P(\theta_{t+1} \in A / \theta_t), \quad A \subseteq S. \quad (5.9)$$

Em outras palavras, uma cadeia de Markov é um processo estocástico onde o próximo estado depende somente do estado atual e de nenhum outro estado anterior. As cadeias de Markov, para uso em algoritmos de amostragem, devem ser:

- irredutíveis: o que significa que, a despeito do seu estado inicial, a cadeia é capaz de alcançar qualquer outro estado em um número finito de iterações com uma probabilidade maior que zero;
- aperiódicas: o que significa que a cadeia não fica oscilando entre um conjunto de estados em movimentos regulares; e
- recorrente: o que significa que para todos os estado i , se o processo se inicia em i , ele retornará ao estado i , com probabilidade 1, em um número finito de iterações.

Uma cadeia de Markov com as características acima é dita ergódica. A idéia básica de todos os algoritmos de amostragem é obter uma amostra de $\pi(\theta / X)$ construindo uma cadeia de Markov ergódica com as seguintes propriedades:

- a cadeia deve ter o mesmo conjunto de estados de θ ;
- a cadeia deve ser de fácil simulação; e
- a distribuição de equilíbrio deve ser $\pi(\theta / X)$.

O algoritmo de Metropolis tem a propriedade de criar cadeias com as características apresentadas acima. O algoritmo de Metropolis (Metropolis et al., 1953) foi desenvolvido nos laboratórios de Los Alamos com o objetivo de resolver problemas referentes ao estado de energia de materiais nucleares. A principal motivação do método era fazer uso da “grande” capacidade de cálculo do primeiro computador programável, MANIAC (*Mathematical Analyzer, Numerical Integrator and Computer*), desenvolvido durante a 2^o Guerra Mundial. Embora o método tenha ganhado dimensão a partir do trabalho de Nicholas Metropolis e seus colaboradores, em 1953, seu desenvolvimento teve a colaboração de vários pesquisadores que trabalharam no Projeto Manhattan, notadamente Stanislaw Ulam, John Von Neumann, Enrico Fermi, Richard Feynman, entre outros. Aliás, conforme o próprio Metropolis admitiu (Metropolis, 1987), a idéia básica do método foi desenvolvida, porém não publicada, cerca de 15 anos antes por Enrico Fermi. Detalhes sobre os momentos históricos do desenvolvimento do algoritmo de Metropolis e do método de Monte Carlo podem ser encontrados em Hitchcock (2003), Anderson (1986) e Metropolis (1987).

O algoritmo de Metropolis foi generalizado por Hastings (1970), de onde surgiu a versão amplamente utilizada nos dias atuais. O algoritmo é construído a partir de uma distribuição de referência $g(\theta^* / \theta_t, x)$, da qual é fácil obter amostras de θ da seguinte forma:

Algoritmo Metropolis-Hastings para obter uma amostra de $\pi(\theta / X)$

Inicialize θ_0 ; $t \leftarrow 0$

Repita{

 Amostre $\theta^* \sim g(\theta^* / \theta_t, x)$

 Amostre $u \sim \text{Uniforme}(0, 1)$

 Calcule $\alpha_{MH}(\theta^* / \theta_t, x) = \text{MIN}\left\{1, \frac{\pi(\theta^* / x)}{\pi(\theta_t / x)} \cdot \frac{g(\theta_t / \theta^*, x)}{g(\theta^* / \theta_t, x)}\right\}$

 Se $u \leq \alpha_{MH}(\theta^* / \theta_t, x)$ faça

$\theta_{t+1} \leftarrow \theta^*$

 Caso contrário faça

$\theta_{t+1} \leftarrow \theta_t$

$t \leftarrow (t + 1)$

}

Uma importante característica do algoritmo é que as distribuições são avaliadas apenas em termos de taxas, $\pi(\theta^* / x) / \pi(\theta_t / x)$, dispensando o cálculo da constante de normalização em (5.1). A generalização proposta por Hastings (1970) refere-se basicamente às propriedades da distribuição de referência $g(\cdot / \cdot)$. Com efeito, no trabalho original de Metropolis et al. (1953), somente distribuições simétricas eram permitidas para g . Ou seja, g deve ser tal que $g(\theta_i / \theta_j) = g(\theta_j / \theta_i)$. Neste caso, a taxa de aceitação do algoritmo se reduz a:

$$\alpha_{MH}(\theta^* / \theta_t, x) = \text{MIN}\left\{1, \frac{\pi(\theta^* / x)}{\pi(\theta_t / x)}\right\}. \quad (5.10)$$

Robert e Casella (2004) mostram que o algoritmo acima, após um grande número de iterações m , alcança seu equilíbrio, tendo como distribuição estacionária $\pi(\theta / X)$. Depois de alcançado o equilíbrio, todas as realizações do algoritmo serão uma amostra da distribuição *a posteriori* de θ e os valores esperados em (5.6) podem ser estimados, via integração de Monte Carlo, com a precisão que se desejar, bastando aumentar o número de realizações do algoritmo.

A escolha da distribuição de referência $g(\cdot / \cdot)$ é a chave para a eficiência do algoritmo. Conforme apontado por Gilks et al. (1996), qualquer distribuição de referência permite obter

amostras da distribuição *a posteriori* de θ . Em outras palavras, para qualquer forma de $g(\cdot/\cdot)$, a cadeia de Markov alcança seu estado de equilíbrio em π . No entanto, a taxa de convergência depende fortemente da relação entre $g(\cdot/\cdot)$ e da distribuição alvo $\pi(\theta/X)$. Assim, quanto mais semelhante for g de π , mais rapidamente a cadeia alcançará seu equilíbrio. Além disso, mesmo quando a cadeia alcança o equilíbrio, suas realizações podem se mover lentamente sobre o suporte de π , sendo necessário um número muito grande de realizações para se ter uma correta amostra da distribuição alvo. Do ponto de vista computacional, g deve ser escolhida de forma que seja facilmente avaliada em qualquer ponto e que seja fácil obter amostras aleatórias em todo o seu suporte. Além disso, a distribuição de referência deve ter caudas mais pesadas que $\pi(\theta/X)$ para se ter maior garantia de que a amostra candidata, gerada a partir de $g(\cdot/\cdot)$, percorrerá todo o espaço paramétrico de $\pi(\theta/X)$. Detalhes sobre como construir a distribuição de referência podem ser encontrados em Gilks et al. (1996).

A partir do algoritmo Metropolis-Hastings surgiram vários outros esquemas de simulação estocástica, entre eles o amostrador de Gibbs. Basicamente, esses outros algoritmos se diferenciam pela forma de construção da distribuição de referência. O amostrador Gibbs, por exemplo, utiliza as distribuições condicionais completas *a posteriori* como distribuição de referência. Esses outros algoritmos não serão abordados neste texto. Uma ampla discussão sobre esses algoritmos e o Metropolis-Hasting é feita em Gilks et al. (1996), Liu (2001) e Robert e Casella (2004), referências às quais sugere-se ao leitor remeter-se, em busca de maior aprofundamento teórico e aplicações mais detalhadas.

5.4 Conclusão

Neste capítulo, foram descritas as principais características da análise de frequência bayesiana. Embora os conceitos primordiais da análise bayesiana datem do século XIX, sua aplicação começou a crescer somente na segunda metade do século XX, com o desenvolvimento de ferramentas computacionais e com a necessidade de modelar processos estocásticos cada vez mais complexos.

A principal diferença entre a análise clássica e a bayesiana está na forma de avaliar as incertezas presentes em objetos aleatórios. Enquanto na Escola Clássica, a medida de probabilidade capta a variabilidade inerente ao processo, no contexto bayesiano, tal medida captura o desconhecimento do indivíduo sobre o objeto em estudo. Daí decorre uma série de diferenças entre a análise clássica e a bayesiana, muitas das quais foram apresentadas nos itens 5.1 e 5.2. entre a análise clássica e a bayesiana.

Conforme visto, uma das maiores dificuldades para a aplicação da análise bayesiana é o cálculo da constante de normalização ou distribuição preditiva *a priori* dada em (5.2). Para contornar essa dificuldade, foi mostrado que algoritmos de amostragem, tais como o de Metropolis, permitem obter amostras das distribuições *a posteriori* sem, no entanto, calcular a constante de normalização. Conforme menção anterior, as demonstrações da validade do algoritmo Metropolis e outras questões de maior rigor matemático podem ser obtidas, por exemplo, em Robert e Casella (2004).

A principal justificativa para a adoção da análise bayesiana frente à clássica advém do fato de que o objetivo principal da inferência estatística é a predição de valores futuros da variável em análise, sendo a análise bayesiana, conforme os argumentos do item 5.1, uma abordagem mais racional e prudente.

Por fim, verifica-se que, a despeito de suas qualidades, a análise bayesiana ainda é pouco aplicada em problemas relacionados à engenharia de recursos hídricos.

6 ABORDAGEM BAYESIANA PARA ESTIMAÇÃO DE QUANTIS DE ENCHENTES EXTREMAS COM O USO DE DISTRIBUIÇÕES LIMITADAS SUPERIORMENTE E INFORMAÇÕES NÃO SISTEMÁTICAS

6.1 Introdução

Nos capítulos 1 e 3 foram dados argumentos em favor da premissa de que as vazões máximas anuais são variáveis aleatórias limitadas superiormente e que devem ser modeladas por funções compatíveis. O principal obstáculo na modelagem de variáveis limitadas é a definição do limite superior. Essa dificuldade se deve, em parte, à pouca disponibilidade ou até ausência de dados, sobretudo aqueles relacionados a eventos extremos nas bacias em estudo, e ao receio por parte do hidrólogo de se prescrever uma vazão cuja probabilidade de excedência seja nula.

Este capítulo contém o desenvolvimento de um método que permite incorporar um limite superior à modelagem das vazões máximas anuais de uma forma lógica e consistente, visando uma melhor estimativa dos quantis de alto período de retorno. Nesse sentido, a abordagem bayesiana, discutida no capítulo 5, oferece uma estrutura de análise mais coerente, uma vez que o limite superior pode ser visto como uma quantidade cujas incertezas podem ser quantificadas por um especialista na forma de uma distribuição de probabilidades. Nesse capítulo propõe-se construir a distribuição de probabilidades do limite superior a partir das estimativas de PMF. A vantagem desse tipo de abordagem é isentar o especialista da difícil, se não impossível, tarefa de se estimar uma vazão cuja probabilidade de excedência seja nula, a partir unicamente de uma amostra reduzida de cheias máximas anuais.

A aplicação da análise bayesiana, que pode ser resumida pela equação 5.1, pressupõe a escolha de um modelo distributivo para modelar as vazões máximas anuais e a definição das distribuições *a priori* dos parâmetros do modelo estatístico. Assim, na seqüência, serão descritas as distribuições de probabilidades limitadas superiormente e a função de verossimilhança que agrega os dados sistemáticos e não sistemáticos, compondo o modelo estatístico. Depois serão apresentados os métodos para a construção da distribuição *a priori* para o limite superior, que, como será visto, é o único, dentre os parâmetros das distribuições limitadas, para o qual foi encontrada uma relação evidente com as características físicas da bacia, permitindo que se façam conjecturas sobre seu comportamento.

As distribuições limitadas superiormente são aquelas que incorporam um limite superior α à variável. Em outras palavras, a distribuição acumulada de probabilidades F é um para valores da variável maiores ou iguais a α .

Botero (2006) analisou o comportamento das vazões máximas anuais com base em três distribuições de probabilidades limitadas superiormente: a distribuição EV4, proposta por Kanda (1981) para analisar o comportamento probabilístico de abalos sísmicos e de ventos extremos no Japão; a distribuição de valores extremos transformada (TDF), proposta por Eliasson (1994), que a utilizou na análise de frequência de precipitações extremas na Islândia e nos Estados Unidos; e a distribuição Log-Normal de 4 parâmetros (LN4), proposta por Takara e Loebis (1996), com base na variável transformada de Slade (Slade, 1936 *apud* Botero, 2006), empregada na análise de frequência de precipitações extremas no Japão e Indonésia. Além dessas, a distribuição Generalizada de Valores Extremos (GEV) e a distribuição de Pareto Generalizada (DGP) também apresentam limites superiores para determinados valores de seus respectivos parâmetros de forma.

Essas distribuições foram selecionadas pelo fato de já terem sido aplicadas em hidrologia e terem demonstrado capacidade de descrever fenômenos naturais. Na seqüência, é descrita cada uma dessas distribuições de probabilidades e suas principais características. Nas aplicações feitas no capítulo 7 somente as distribuições LN4 e a EV4 serão utilizadas. De fato, conforme será visto, a GEV e a DGP, em geral, não são limitadas superiormente para as características amostrais presentes nas séries de vazões máximas anuais. Além disso, no trabalho de Botero (2006), demonstrou-se que a TDF não permite uma boa caracterização de variáveis limitadas. Assim, com o intuito de melhor focalizar os resultados e as vantagens do método proposto, o elenco das distribuições limitadas superiormente restringiu-se aos modelos LN4 e EV4.

6.2 Distribuições limitadas superiormente

6.2.1 Distribuições GEV e DGP

Uma variável aleatória X tem distribuição GEV com parâmetro de posição $\mu \in \mathfrak{R}$, parâmetro de escala $\sigma \in \mathfrak{R}_+$ e parâmetro de forma $\xi \in \mathfrak{R}$, a qual será denotada por $X \sim \text{GEV}(\mu, \sigma, \xi)$, se sua função de distribuição acumulada de probabilidades é dada por:

$$F_X(x/\bar{\Theta}) = \begin{cases} \exp\left\{-\left[1-\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{1/\xi}\right\} & \text{se } \xi \neq 0, \\ \exp\left\{-\exp\left\{-\left(\frac{x-\mu}{\sigma}\right)\right\}\right\} & \text{se } \xi = 0. \end{cases} \quad (6.1)$$

Para essa distribuição, o sinal de ξ determina o domínio da distribuição conforme mostrado a seguir:

$$\begin{cases} x > \mu + \sigma/\xi & \text{se } \xi < 0, \\ x < \mu + \sigma/\xi & \text{se } \xi > 0, \\ x \in \mathfrak{R} & \text{se } \xi = 0. \end{cases}$$

É importante frisar que a GEV é limitada superiormente somente se $\xi > 0$. Para $\xi = 0$, a GEV é conhecida como distribuição de Gumbel e tem coeficiente de assimetria constante igual a 1,1396. Para $\xi > -1/3$ seu coeficiente de assimetria é dado por:

$$\gamma = \langle \text{sinal de } \xi \rangle \frac{-\Gamma(1+3\xi) + 3\Gamma(1+\xi)\Gamma(1+2\xi) - 2\Gamma^3(1+\xi)}{[\Gamma(1+2\xi) - \Gamma^2(1+\xi)]^{3/2}}, \quad \text{se } \xi > -\frac{1}{3}. \quad (6.2)$$

Nesse caso, o coeficiente de assimetria depende somente do parâmetro de forma. De (6.2) verifica-se que, para $\xi > 0$, o coeficiente de assimetria é menor que 1,1396. Ou seja, a GEV será limitada superiormente somente para valores do coeficiente de assimetria inferiores a 1,1396.

Uma variável aleatória X tem distribuição DGP com parâmetro de posição $\mu \in \mathfrak{R}$, parâmetro de escala $\sigma \in \mathfrak{R}_+$ e parâmetro de forma $\xi \in \mathfrak{R}$, a qual será denotada por $X \sim \text{DGP}(\mu, \sigma, \xi)$, se sua função de distribuição acumulada de probabilidades é dada por:

$$F_X(x/\bar{\Theta}) = \begin{cases} 1 - \left[1 - \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{1/\xi} & \text{se } \xi \neq 0, \\ 1 - \exp\left\{-\left(\frac{x-\mu}{\sigma}\right)\right\} & \text{se } \xi = 0. \end{cases} \quad (6.3)$$

Analogamente ao caso da GEV, o domínio da distribuição DGP depende do parâmetro de forma ξ da seguinte maneira:

$$\begin{cases} x \geq \mu & \text{se } \xi \leq 0, \\ \mu < x \leq \mu + \sigma/\xi & \text{se } \xi > 0. \end{cases}$$

Destaca-se que a DGP é limitada superiormente somente se $\xi > 0$. Como para a GEV, o coeficiente de assimetria da DGP depende unicamente do parâmetro de forma. Se $\xi > -1/3$, o coeficiente de assimetria da DGP é dado por:

$$\gamma = \frac{2(1-\xi)(1+2\xi)^{1/2}}{1+3\xi}. \quad (6.4)$$

De (6.4) segue-se que o coeficiente de assimetria é menor que 2,0 para $\xi > 0$. Ou seja, a DGP será limitada superiormente somente para valores do coeficiente de assimetria inferiores a 2,0.

De acordo com Naghettini e Pinto (2007), as séries hidrológicas referentes a eventos máximos, em geral, possuem coeficientes de assimetria positivos. Ainda de acordo com esses autores, no caso de séries de vazões máximas anuais, há uma grande concentração de valores próximos à cheia média anual que correspondem aos níveis d'água contidos pelo leito menor da seção fluvial. Entretanto, a rara combinação de condições hidrometeorológicas excepcionais e de elevado teor de umidade do solo pode determinar a ocorrência de uma grande enchente, com vazão máxima muitas vezes superior ao valor modal. Bastam apenas algumas ocorrências de tais grandes enchentes para determinar valores muito positivos para o coeficiente de assimetria. Na prática, as séries referentes às vazões máximas anuais, com muita frequência, possuem coeficientes de assimetria superiores a 1,1396 e, com menor frequência, superiores a 2,0. Dessa forma, as distribuições GEV e DGP não apresentam limites superiores para essas séries hidrológicas.

Devido à falta de generalidade apontada, essas distribuições de probabilidades não serão consideradas neste trabalho para modelar o comportamento de vazões máximas anuais.

6.2.2 Distribuição EV4

A distribuição de valores extremos do tipo IV (EV4), denominada assim por Kanda (1981), não é propriamente uma distribuição de valores extremos, uma vez que não é derivada da teoria dos valores extremos. Essa denominação se deve ao fato de sua forma paramétrica ter sido proposta como uma modificação da distribuição de valores extremos do tipo II, ou EV3.

A função de probabilidades acumulada de uma variável distribuída de acordo com o modelo EV4, a qual será denotada por $X \sim \text{EV4}(\sigma, \xi, \alpha, \varepsilon)$, com parâmetro de escala $\sigma \in \mathfrak{R}_+^*$, de forma $\xi \in \mathfrak{R}_+$, limite superior $\alpha \in \mathfrak{R}_+$ e limite inferior $\varepsilon \in \mathfrak{R}_+$, é dada pela seguinte equação:

$$F_X(x/\bar{\Theta}) = \exp\left\{-\left[\frac{\alpha-x}{\sigma(x-\varepsilon)}\right]^\xi\right\}, \quad \varepsilon < x \leq \alpha. \quad (6.5)$$

A função densidade de probabilidade e os quantis de ordem p da EV4 são dados respectivamente, por:

$$f_X(x/\bar{\Theta}) = \frac{\xi(\alpha-x)^{\xi-1}(\alpha-\varepsilon)}{\sigma^\xi(x-\varepsilon)^{\xi+1}} \exp\left\{-\left[\frac{\alpha-x}{\sigma(x-\varepsilon)}\right]^\xi\right\}, \quad \varepsilon < x \leq \alpha \quad (6.6)$$

e

$$F_X^{-1}(p/\bar{\Theta}) = \frac{\alpha + \sigma\varepsilon[-\ln(p)]^{1/\xi}}{\sigma[-\ln(p)]^{1/\xi} + 1}, \quad 0 < p < 1. \quad (6.7)$$

Embora a EV4 não seja uma distribuição de valores extremos, pode-se estabelecer um paralelo com essa última quando X tende ao limite inferior ou ao limite superior. Com efeito, quando X tende ao limite inferior ε , a EV4 tende para uma EV2 (distribuição de Fréchet), com parâmetro de escala igual a $(\alpha - \varepsilon)/\sigma$. Por outro lado, quando X tende ao limite superior α , a EV4 tende a uma EV3 (distribuição de Weibull), com parâmetro de escala igual a $\sigma(\alpha - \varepsilon)$.

As vazões de um curso d'água podem assumir somente valores não negativos. Desta forma, torna-se razoável considerar $\varepsilon = 0$ para aplicações da EV4 com dados fluviométricos. A rigor, ε deve ser superior a zero, uma vez que, na prática, a menos que o rio fique seco por mais de um ano, sempre haverá um valor mínimo de vazão maior do que zero. Por outro lado, a prescrição de um limite inferior para as vazões máximas anuais de um rio é controversa e não há um consenso quanto a um método a ser empregado para tal. Takara e Tosa (1999) verificaram que os melhores ajustes são obtidos quando $\varepsilon = 0$ e que, à medida que o limite inferior se aproxima do mínimo amostral, a EV4 perde a capacidade de descrever o comportamento das variáveis máximas anuais. Por essa razão e por maior conveniência do ponto de vista matemático, nas aplicações que se seguem será fixado o limite inferior em zero.

A figura 6.1 mostra a influência de cada parâmetro na forma da distribuição. Nota-se que a EV4 tende a ter caudas superiores relativamente mais pesadas para valores pequenos de σ e ξ e para valores grandes de α .

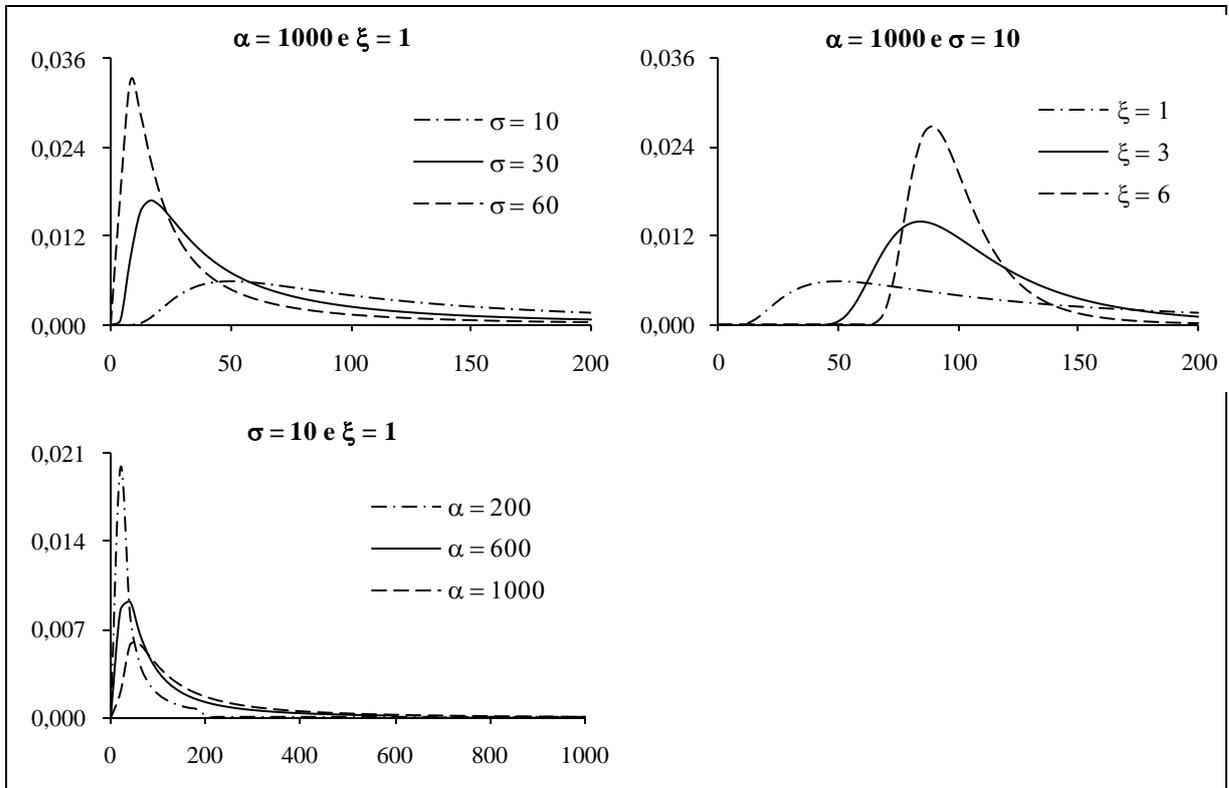


Figura 6.1 – Efeito de cada parâmetro na forma da distribuição EV4

Botero (2006) verificou que não é possível obter uma equação de momentos para a EV4 escrita de acordo com (6.5). No entanto, definindo $\varepsilon = 0$ e fazendo uma mudança de variáveis da forma $Y = 1/X$, é possível obter uma forma analítica para a equação de momentos de Y , conforme é mostrado na dedução a seguir.

Proposição 6.1: Seja $X \sim \text{EV4}(\sigma, \xi, \alpha, 0)$. Nesse caso, os momentos de ordem r de $Y = 1/X$ são:

$$\mu_r = \frac{1}{\alpha^r} \sum_{i=0}^r \binom{r}{i} \cdot \sigma^{r-i} \cdot G(r-i), \quad (6.8)$$

onde $G(j) = \Gamma\left(\frac{j}{\xi} + 1\right)$ e $\Gamma(\cdot)$ denota a função gama.

Prova: Fazendo $\varepsilon = 0$ em (6.5) e (6.6), tem-se que a FDA e FDP de X são, respectivamente:

$$F_X(x/\bar{\Theta}) = \exp\left\{-\left(\frac{\alpha-x}{\sigma x}\right)^\xi\right\}, \quad 0 < x \leq \alpha, \quad (6.9)$$

e

$$f_X(x/\bar{\Theta}) = \frac{\xi\alpha}{\sigma} \cdot \left(\frac{\alpha-x}{\sigma x}\right)^{\xi-1} \cdot \frac{1}{x^2} \cdot \exp\left\{-\left(\frac{\alpha-x}{\sigma x}\right)^\xi\right\}, \quad 0 < x \leq \alpha. \quad (6.10)$$

Rearranjando os termos em (6.9), tem-se que a FDA de X pode ser reescrita da seguinte forma:

$$F_X(x/\bar{\Theta}) = \left[\exp\left\{\left(\frac{\alpha}{x}-1\right)^\xi\right\} \right]^\kappa, \quad (6.11)$$

onde $\kappa = -\left(\frac{1}{\sigma}\right)^\xi$.

Fazendo $Y = \frac{1}{X}$, a FDA de Y é dada por:

$$F_Y(y/\bar{\Theta}) = 1 - F_X(1/y), \text{ ou} \quad (6.12)$$

$$F_Y(y/\bar{\Theta}) = 1 - \left[\exp\left\{\left(\alpha \cdot y - 1\right)^\xi\right\} \right]^\kappa, \quad \frac{1}{\alpha} \leq y < +\infty. \quad (6.13)$$

A inversa de (6.13) é dada por:

$$F_Y^{-1}(p/\bar{\Theta}) = \frac{1}{\alpha} \left\{ \left[\frac{\ln(1-p)}{\kappa} \right]^{1/\xi} + 1 \right\}, \quad 0 < p < 1. \quad (6.14)$$

O valor esperado de y é, por definição:

$$E[y] = \int_{-\infty}^{+\infty} y \cdot f_Y(y/\bar{\Theta}) dy = \int_{-\infty}^{+\infty} y \cdot dF_Y(y/\bar{\Theta}). \quad (6.15)$$

Fazendo $p = F_Y(y/\bar{\Theta})$, segue-se que $y = F_Y^{-1}(p/\bar{\Theta})$. Conseqüentemente, de (6.15) segue-se que:

$$E[y] = \int_0^1 F_Y^{-1}(p/\bar{\Theta}) dp. \quad (6.16)$$

Substituindo (6.14) em (6.16), tem-se:

$$\mu_1 = \int_0^1 F_Y^{-1}(p/\bar{\Theta}) dp = \frac{1}{\alpha} \int_0^1 \left\{ \left[\frac{\ln(1-p)}{\kappa} \right]^{1/\xi} + 1 \right\} dp = \frac{1}{\alpha} \left[1 + \sigma \cdot \Gamma\left(\frac{1}{\xi} + 1\right) \right]. \quad (6.17)$$

O momento de ordem 2 pode ser obtido a partir de um desenvolvimento semelhante ao realizado para o momento de ordem 1. Neste caso, tem-se que:

$$\mu_2 = \int_0^1 \left[F_Y^{-1}(p/\bar{\Theta}) \right]^2 dp = \frac{\sigma^2}{\alpha^2} \Gamma\left(\frac{2}{\xi} + 1\right) + 2 \frac{\sigma}{\alpha^2} \Gamma\left(\frac{1}{\xi} + 1\right) + \frac{1}{\alpha^2}. \quad (6.18)$$

A proposição 6.1 segue-se, então, por indução.

A proposição 6.1 fornece um limite inferior para os momentos de uma variável aleatória $X \sim \text{EV4}(\sigma, \xi, \alpha, 0)$, uma vez que, pela desigualdade de Jensen, segue-se que $E(X^r) \geq [E(1/X^r)]^{-1} = [E(Y^r)]^{-1}$.

6.2.3 Distribuição TDF

A distribuição TDF (*Transformed Distribution Function*) foi proposta inicialmente por Elíasson (1994) para a análise de frequência de precipitações máximas. Elíasson (1997) utilizou esta distribuição na análise de frequência de precipitações máximas de 24 horas de duração na Islândia e Estados Unidos, obtendo resultados promissores.

A TDF pode ser construída da seguinte forma: seja Y uma variável ilimitada distribuída de acordo com o modelo Gumbel. A função de probabilidades acumulada de Y , a qual será denotada por $Y \sim \text{GUM}(\mu, \sigma)$, com parâmetro de posição $\mu \in \mathfrak{R}$ e parâmetro de escala $\sigma \in \mathfrak{R}_+^*$, é dada pela seguinte equação:

$$F_Y(y/\bar{\Theta}) = \exp\left\{-\exp\left\{-\left(\frac{y-\mu}{\sigma}\right)\right\}\right\}, \quad y \in \mathfrak{R}. \quad (6.19)$$

Elíasson (1997) propõe a seguinte transformação de Y :

$$Y = X - \frac{\sigma^2\beta}{\alpha - X}, \quad (6.20)$$

onde $Y \sim \text{GUM}(\mu, \sigma)$, $\alpha \in \mathfrak{R}_+$ é o limite superior de X e $\beta \in \mathfrak{R}_-$ é um parâmetro.

A TDF é obtida a partir de (6.19) e (6.20). Ou seja, se Y é distribuída de acordo com o modelo Gumbel, então X é distribuída de acordo com o modelo TDF. A função de probabilidades acumulada de uma variável distribuída de acordo com o modelo TDF, a qual será denotada por $X \sim \text{TDF}(\mu, \sigma, \beta, \alpha)$, é dada pela seguinte equação:

$$F_X(x/\bar{\Theta}) = \exp\left\{-\exp\left\{-\frac{x}{\sigma} + \frac{\sigma\beta}{\alpha - x} + \frac{\mu}{\sigma}\right\}\right\}, \quad 0 < x < \alpha. \quad (6.21)$$

A função densidade de probabilidade da TDF é dada por:

$$f_X(x/\bar{\Theta}) = \exp\left\{-\exp\left\{-\frac{x}{\sigma} + \frac{\sigma\beta}{\alpha - x} + \frac{\mu}{\sigma}\right\}\right\} \cdot \exp\left\{-\frac{x}{\sigma} + \frac{\sigma\beta}{\alpha - x} + \frac{\mu}{\sigma}\right\} \cdot \left[\frac{1}{\sigma} - \frac{\sigma\beta}{(\alpha - x)^2}\right]. \quad (6.22)$$

Os quantis de ordem $p \in (0, 1)$ podem ser obtidos resolvendo (6.21) para x . Nesse caso, obtém-se a seguinte equação quadrática:

$$x^2 + C_1x + C_0 = 0, \quad (6.23)$$

onde

$$C_0 = -\sigma\alpha \left\{ \ln[-\ln(p)] - \frac{\mu}{\sigma} + \beta \right\} \text{ e}$$

$$C_1 = \sigma \{ \ln[-\ln(p)] - \alpha \}.$$

Resolvendo (5.23) e tomando somente o valor positivo de x , tem-se:

$$F_X^{-1}(p/\bar{\Theta}) = -\frac{C_0}{2} + \sqrt{\frac{C_0^2}{4} - C_1}. \quad (6.24)$$

A figura 6.2 mostra a influência de cada parâmetro na forma da distribuição.

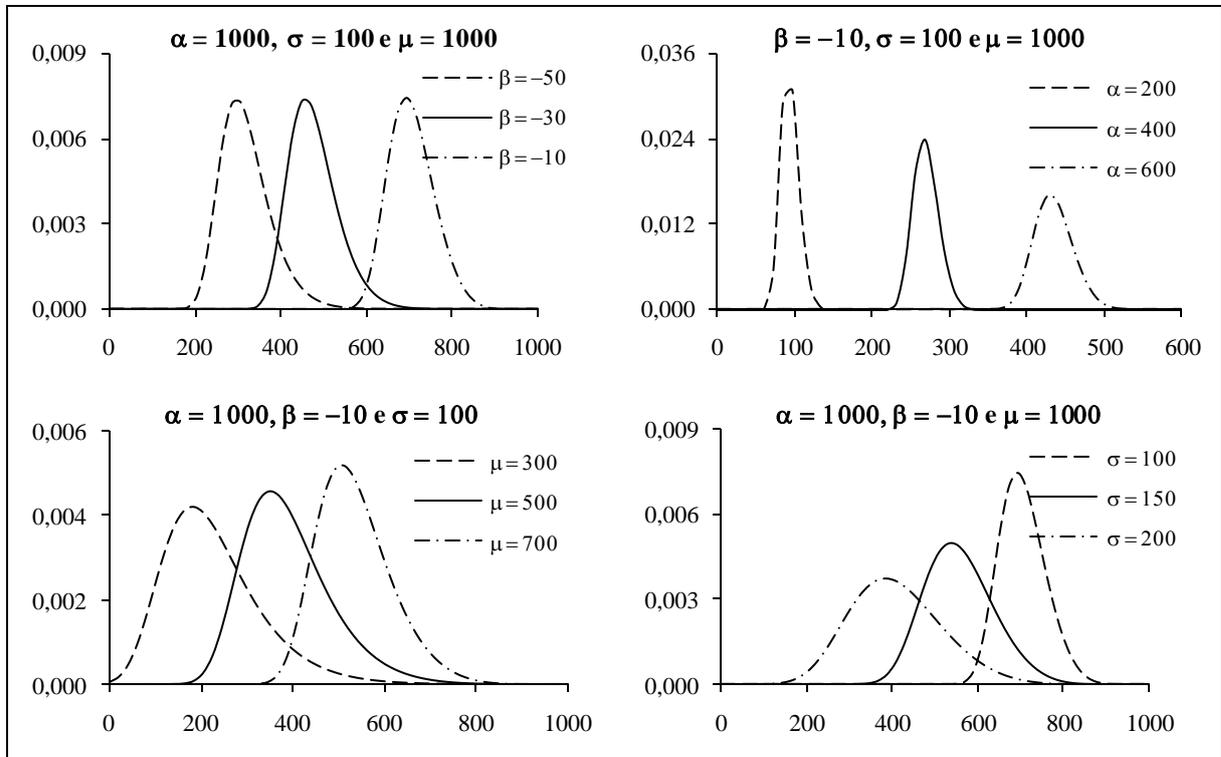


Figura 6.2 – Efeito de cada parâmetro na forma da distribuição TDF

Não é possível obter uma forma analítica para os momentos da TDF, nem mesmo um limite inferior, tal como na EV4. Assim, os momentos devem ser encontrados, numericamente, a partir de suas respectivas equações de definição.

6.2.4 Distribuição Log-Normal de 4 parâmetros (LN4)

Essa distribuição foi proposta por Slade (1936) a partir de uma transformação de uma variável normalmente distribuída. A hipótese básica adotada por Slade é a de que uma distribuição pode ser caracterizada pelo seu desvio padrão e pelas flutuações máxima e mínima possíveis em torno da média. Quando essas flutuações tendem para o infinito (máximo e mínimo infinitos), a distribuição tende para a distribuição Normal.

Takara e Loebis (1996) introduziram essa distribuição em hidrologia para a análise de dados extremos de precipitação da Indonésia e Japão. Os autores obtiveram bons resultados quando comparados com aqueles obtidos pelo uso da distribuição Log-Normal de 3 parâmetros, a qual não é limitada superiormente. Posteriormente, Takara e Tosa (1999) utilizaram a LN4 para dados de vazão obtendo resultados semelhantes àqueles de precipitação.

A distribuição LN4 é obtida partindo-se da seguinte transformação:

$$Y = \ln\left(\frac{X - \varepsilon}{\alpha - X}\right), \quad (6.25)$$

onde $\varepsilon \in \mathfrak{R}_+$ é o limite inferior de X , $\alpha \in \mathfrak{R}_+$ é o limite superior de X e $Y \sim \text{NOR}(\mu_y, \sigma_y^2)$.

Por conveniência, os parâmetros μ_y e σ_y serão denotados, doravante, somente por μ e σ . A função densidade de probabilidade de uma variável distribuída de acordo com o modelo LN4, a qual será denotada por $X \sim \text{LN4}(\mu, \sigma, \alpha, \varepsilon)$, com parâmetro de escala $\sigma \in \mathfrak{R}_+^*$, de posição $\mu \in \mathfrak{R}$, limite superior $\alpha \in \mathfrak{R}_+$ e limite inferior $\varepsilon \in \mathfrak{R}_+$, é dada pela seguinte equação:

$$f_X(x/\vec{\Theta}) = \frac{\alpha}{x(\alpha - x)\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}\left[\ln\left(\frac{x - \varepsilon}{\alpha - x}\right) - \mu\right]^2\right\}, \quad \varepsilon < x < \alpha. \quad (6.26)$$

A distribuição acumulada de probabilidade é dada por:

$$F_X(x/\vec{\Theta}) = \Phi\left[\frac{1}{\sigma}\ln\left(\frac{x - \varepsilon}{\alpha - x}\right) - \frac{\mu}{\sigma}\right], \quad \varepsilon < x < \alpha, \quad (6.27)$$

onde Φ representa a distribuição Normal padrão acumulada. Pelas mesmas razões apontadas para o caso da EV4, o limite inferior ε é tomado igual a zero também para a LN4. Não é possível obter uma forma analítica para os momentos da LN4. Assim, os mesmos devem ser obtidos por meio da solução numérica de suas respectivas equações de definição.

A figura 6.3, mostra a influência de cada parâmetro na forma da distribuição LN4. Nota-se pelo gráfico que os parâmetros σ e α controlam o peso da cauda superior, sendo que, quanto maiores os valores desses parâmetros mais pesada é a cauda. Já o parâmetro μ controla a assimetria da distribuição. Para valores negativos de μ , a assimetria da distribuição é positiva, enquanto que para valores positivos de μ , a assimetria da distribuição é negativa; para μ igual a zero a distribuição é simétrica.

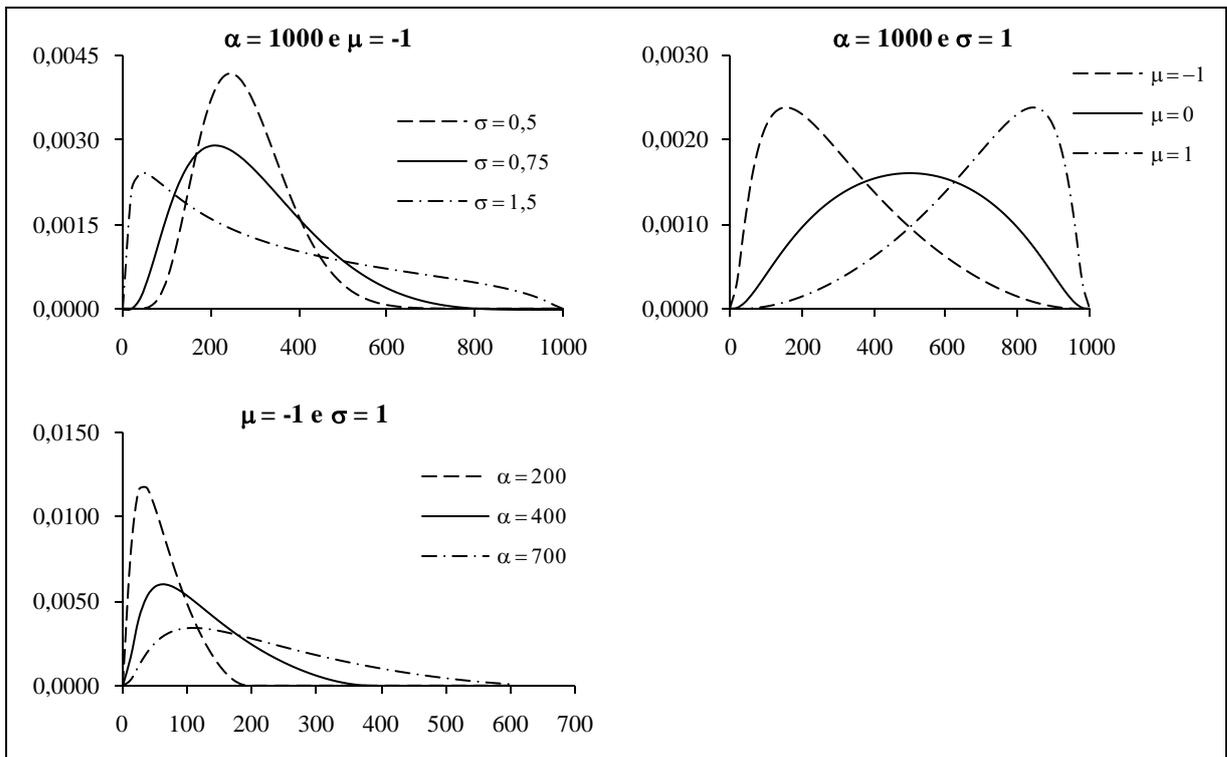


Figura 6.3 – Efeito de cada parâmetro na distribuição LN4

6.3 Função de verossimilhança para dados sistemáticos e não sistemáticos

A função de verossimilhança que agrega todos os dados e informações hidrológicas mostrados no capítulo 3 pode ser construída de acordo com o desenvolvimento feito por Nault (2002). Na seqüência, cada tipo de informação é discutido separadamente, considerando-se que a intensidade das cheias tem seu comportamento descrito por uma FDA $F_X(\cdot/\bar{\Theta})$, com densidade $f_X(\cdot/\bar{\Theta})$.

6.3.1 Cheias de intensidade conhecida, superior a um limiar fixo

Num período histórico de N_H anos, a amostra é constituída por k cheias de intensidade conhecida (y_1, y_2, \dots, y_k) , as quais foram registradas exatamente por terem excedido um limiar de referência y_H , e $(N_H - k)$ cheias de intensidade desconhecida, porém inferiores a y_H , ou seja, uma amostra censurada, ou truncada do tipo I.

Sejam os eventos $A: [Y \geq y_H]$, $A^c: [0 < Y < y_H]$ e $B: [y \leq Y < y + dy]$. Dada a condição que as cheias históricas e sistemáticas sejam descritas pela mesma função densidade de probabilidade f_X , as probabilidades de ocorrência dos eventos A e A^c são expressas, respectivamente, por:

$$P[Y \geq y_H] = \int_{y_H}^{+\infty} f_X(y/\bar{\Theta}) dy = 1 - F_X(y_H/\bar{\Theta}) \quad (6.28)$$

e

$$P[0 < Y < y_H] = \int_0^{y_H} f_X(y/\bar{\Theta}) dy = F_X(y_H/\bar{\Theta}). \quad (6.29)$$

A probabilidade de ocorrência do evento B , dado que A ocorreu, é expressa por:

$$P[y \leq Y < y + dy, Y \geq y_H] = \frac{P[y \leq Y < y + dy, Y \geq y_H]}{P[Y \geq y_H]}, \quad (6.30)$$

onde $P[y \leq Y < y + dy, Y \geq y_H] = P[y \leq Y < y + dy]$, uma vez que B está contido em A ($B \subset A$). Assim,

$$P[y \leq Y < y + dy, Y \geq y_H] = \frac{f_X(y/\bar{\Theta}) dy}{1 - F_X(y_H/\bar{\Theta})}. \quad (6.31)$$

Considerando que os elementos da amostra são independentes, a probabilidade P_C de se observar, nos N_H anos, exatamente k cheias de intensidade conhecida y_m , superiores a y_H , e $(N_H - k)$ cheias de intensidade desconhecida, inferiores a y_H , é expressa por:

$$P_C = \binom{N_H}{k} \cdot [1 - F_X(y_H/\bar{\Theta})]^k \cdot [F_X(y_H/\bar{\Theta})]^{N_H - k} \cdot \prod_{m=1}^k \frac{f_X(y_m/\bar{\Theta}) dy_m}{1 - F_X(y_H/\bar{\Theta})} \quad (6.32)$$

e

$$P_C = \binom{N_H}{k} \cdot [F_X(y_H/\bar{\Theta})]^{N_H - k} \cdot \prod_{m=1}^k f_X(y_m/\bar{\Theta}) dy_m. \quad (6.33)$$

Portanto, a função de verossimilhança L_C é dada pela seguinte expressão:

$$L_C = \binom{N_H}{k} \cdot [F_X(y_H/\Theta)]^{N_H - k} \cdot \prod_{m=1}^k f_X(y_m/\Theta). \quad (6.34)$$

6.3.2 Cheias de intensidade desconhecida, superior a um limiar fixo

Num período histórico de N_H anos, a amostra é constituída por k cheias superiores a um limiar de referência y_H e $(N_H - k)$ cheias inferiores a esse limiar. Todos os eventos são de intensidade desconhecida, ou seja, uma amostra binomial censurada, ou truncada do tipo I.

Considerando as propriedades dos eventos A e A^c , descritos no item anterior, e a independência entre os elementos da amostra, a probabilidade P_{BC} de se observar, nos N_H anos, exatamente k cheias superiores e $(N_H - k)$ cheias inferiores a y_H é dada pela seguinte distribuição binomial:

$$P_{BC} = \binom{N_H}{k} \cdot [1 - F_X(y_H / \bar{\Theta})]^k \cdot [F_X(y_H / \bar{\Theta})]^{N_H - k}. \quad (6.35)$$

Portanto, a função de verossimilhança L_{BC} é dada por:

$$L_{BC} = \binom{N_H}{k} \cdot [1 - F_X(y_H / \bar{\Theta})]^k \cdot [F_X(y_H / \bar{\Theta})]^{N_H - k}. \quad (6.36)$$

6.3.3 Cheias de intensidade compreendida em um intervalo, superior a um limiar fixo

Num período histórico de N_H anos, a amostra é constituída por k cheias de intensidade compreendida em um intervalo (y_{Lm}, y_{Um}) , as quais foram registradas exatamente por terem excedido um limiar de referência y_H , e $(N_H - k)$ cheias de intensidade desconhecida, porém inferiores a y_H , ou seja, uma amostra censurada em um intervalo, ou truncada do tipo I.

Sejam os eventos A e A^c , cujas propriedades foram descritas anteriormente, e o evento $C: [y_L \leq Y < y_U]$. A probabilidade de ocorrência do evento C , dado que A ocorreu, é expressa por:

$$P[y_L \leq Y < y_U, Y \geq y_H] = \frac{P[y_L \leq Y < y_U, Y \geq y_H]}{P[Y \geq y_H]}, \quad (6.37)$$

onde $y_L \geq y_H$ e $P[y_L \leq Y < y_U, Y \geq y_H] = P[y_L \leq Y < y_U]$, uma vez que C está contido em A ($C \subset A$). Assim:

$$P[y_L \leq Y < y_U, Y \geq y_H] = \frac{F_X(y_U / \vec{\Theta}) - F_X(y_L / \vec{\Theta})}{1 - F_X(y_H / \vec{\Theta})}. \quad (6.38)$$

Considerando que os elementos da amostra são independentes, a probabilidade P_{CI} de se observar, nos N_H anos, exatamente k cheias de intensidade compreendida em um intervalo (y_{Lm}, y_{Um}) , cujo limite inferior excede y_H , e $(N_H - k)$ cheias de intensidade desconhecida, inferiores a y_H , é expressa por:

$$P_{CI} = \binom{N_H}{k} \cdot [1 - F_X(y_H / \vec{\Theta})]^k \cdot [F_X(y_H / \vec{\Theta})]^{N_H - k} \cdot \prod_{m=1}^k \frac{F_X(y_{Um} / \vec{\Theta}) - F_X(y_{Lm} / \vec{\Theta})}{1 - F_X(y_H / \vec{\Theta})} \quad (6.39)$$

e

$$P_{CI} = \binom{N_H}{k} \cdot [F_X(y_H / \vec{\Theta})]^{N_H - k} \cdot \prod_{m=1}^k [F_X(y_{Um} / \vec{\Theta}) - F_X(y_{Lm} / \vec{\Theta})]. \quad (6.40)$$

Portanto, a função de verossimilhança L_{CI} é dada por:

$$L_{CI} = \binom{N_H}{k} \cdot [F_X(y_H / \vec{\Theta})]^{N_H - k} \cdot \prod_{m=1}^k [F_X(y_{Um} / \vec{\Theta}) - F_X(y_{Lm} / \vec{\Theta})]. \quad (6.41)$$

6.3.4 Generalização

Os casos particulares apresentados nos três itens anteriores podem ser generalizados, de forma a levar em conta a existência de diferentes limiares y_{Hj} e diferentes intervalos $(y_{Lm,j}, y_{Um,j})$, associados a t períodos históricos de N_{Hj} anos, $j = 1, 2, \dots, t$, tal que k_j eventos tenham igualado ou excedido y_{Hj} e $\sum_{j=1}^t N_{Hj} = N_H$. Nessas condições, as funções de verossimilhança são dadas por:

Cheias de intensidade conhecida, superior a um limiar fixo

$$L_C = \prod_{j=1}^t \left\{ \binom{N_{Hj}}{k_j} \cdot [F_X(y_{Hj} / \vec{\Theta})]^{N_{Hj} - k_j} \cdot \prod_{m=1}^{k_j} f_X(y_{m,j} / \vec{\Theta}) \right\} \quad (6.42)$$

Se forem considerados vários períodos sucessivos de um ano, $N_{Hj} = 1$, com $k_j = 1$ ou 0, caso o evento do ano j seja, respectivamente, superior ou inferior ao limiar y_{Hj} , o período histórico de N_H anos será composto por:

- N_H^* anos de cheias de intensidade conhecida y_j , superior ao limiar; e
- $N_H^<$ anos de cheias de intensidade desconhecida, inferior ao limiar, o qual, nesse caso, é denotado por y_{Uj} .

A expressão da função de verossimilhança pode, então, ser reescrita como:

$$L_C = \prod_{j=1}^{N_H^<} F_X(y_{Uj} / \bar{\Theta}) \cdot \prod_{j=1}^{N_H^*} f_X(y_j / \bar{\Theta}) \quad (6.43)$$

Cheias de intensidade desconhecida, superior a um limiar fixo

$$L_{BC} = \prod_{j=1}^t \left\{ \binom{N_{Hj}}{k_j} \cdot [1 - F_X(y_{Hj} / \bar{\Theta})]^{k_j} \cdot [F_X(y_{Hj} / \bar{\Theta})]^{N_{Hj}-k_j} \right\} \quad (6.44)$$

Se forem considerados vários períodos sucessivos de um ano ($N_{Hj} = 1$), com $k_j = 1$ ou 0, caso o evento do ano j seja, respectivamente, superior ou inferior ao limiar y_{Hj} , o período histórico de N_H anos será composto por:

- $N_H^>$ anos de cheias de intensidade desconhecida, superior ao limiar, o qual, nesse caso, é denotado por y_{Lj} ; e
- $N_H^<$ anos de cheias de intensidade desconhecida, inferior ao limiar, o qual, nesse caso, é denotado por y_{Uj} .

A expressão da função de verossimilhança pode, então, ser reescrita como:

$$L_{BC} = \prod_{j=1}^{N_H^>} [1 - F_X(y_{Lj} / \bar{\Theta})] \cdot \prod_{j=1}^{N_H^<} F_X(y_{Uj} / \bar{\Theta}) \quad (6.45)$$

Cheias de intensidade compreendida em um intervalo, superior a um limiar fixo

$$L_{CI} = \prod_{j=1}^t \left\{ \binom{N_{Hj}}{k_j} \cdot [F_X(y_{Hj} / \vec{\Theta})]^{N_{Hj}-k_j} \cdot \prod_{m=1}^{k_j} [F_X(y_{Um,j} / \vec{\Theta}) - F_X(y_{Lm,j} / \vec{\Theta})] \right\} \quad (6.46)$$

Se forem considerados vários períodos sucessivos de um ano ($N_{Hj} = 1$), com $k_j = 1$ ou 0, caso o evento do ano j seja, respectivamente, superior ou inferior ao limiar y_{Hj} , o período histórico de N_H anos será composto por:

- $N_H^>$ anos de cheias de intensidade compreendida no intervalo (y_{Lj}, y_{Uj}) , superior ao limiar; e
- $N_H^<$ anos de cheias de intensidade desconhecida, inferior ao limiar, o qual, nesse caso, é denotado por y_{Uj} .

A expressão da função de verossimilhança pode, então, ser reescrita como:

$$L_{CI} = \prod_{j=1}^{N_H^<} F_X(y_{Uj} / \vec{\Theta}) \cdot \prod_{j=1}^{N_H^>} [F_X(y_{Uj} / \vec{\Theta}) - F_X(y_{Lj} / \vec{\Theta})] \quad (6.47)$$

Para os dados do período sistemático, as funções de verossimilhança podem ser obtidas utilizando um raciocínio análogo ao desenvolvido para as cheias históricas. As expressões (6.43), (6.45) e (6.47) podem ser decompostas em expressões elementares, que correspondem às funções de verossimilhança dos diferentes tipos de informações, ou seja:

- Informação não censurada

$$L_H^* = \prod_{j=1}^{N_H^*} f_X(y_j / \vec{\Theta}) \quad \text{Período histórico} \quad (6.48)$$

$$L_S^* = \prod_{i=1}^{N_S^*} f_X(x_i / \vec{\Theta}) \quad \text{Período sistemático} \quad (6.49)$$

- Informação censurada

Inferior a um limiar

$$L_H^< = \prod_{j=1}^{N_H^<} F_X(y_{Uj} / \vec{\Theta}) \quad \text{Período histórico} \quad (6.50)$$

$$L_S^< = \prod_{i=1}^{N_S^<} F_X(x_{Ui} / \vec{\Theta}) \quad \text{Período sistemático} \quad (6.51)$$

Superior a um limiar

$$L_H^> = \prod_{j=1}^{N_H^>} [1 - F_X(y_{Lj} / \vec{\Theta})] \quad \text{Período histórico} \quad (6.52)$$

$$L_S^> = \prod_{i=1}^{N_S^>} [1 - F_X(x_{Li} / \vec{\Theta})] \quad \text{Período sistemático} \quad (6.53)$$

Compreendida em um intervalo

$$L_H^{\diamond} = \prod_{j=1}^{N_H^{\diamond}} [F_X(y_{Uj} / \vec{\Theta}) - F_X(y_{Lj} / \vec{\Theta})] \quad \text{Período histórico} \quad (6.54)$$

$$L_S^{\diamond} = \prod_{i=1}^{N_S^{\diamond}} [F_X(x_{Ui} / \vec{\Theta}) - F_X(x_{Li} / \vec{\Theta})] \quad \text{Período sistemático} \quad (6.55)$$

A função de verossimilhança da amostra completa, denotada por L_{TOTAL} , é dada pelo produto das funções acima, ou seja:

$$L_{TOTAL} = L_H^{\bullet} \cdot L_H^< \cdot L_H^> \cdot L_H^{\diamond} \cdot L_S^{\bullet} \cdot L_S^< \cdot L_S^> \cdot L_S^{\diamond} \quad (6.56)$$

Vale salientar que, em situações reais, a disponibilidade de informações é bastante limitada, reduzindo, assim, o número de componentes da equação 6.56.

O modelo estatístico se completa com a especificação das distribuições *a priori* dos parâmetros que compõem a função de verossimilhança. Essa especificação será vista no item 6.5.

6.4 Probabilidades de excedência empíricas

6.4.1 Probabilidades de excedência empíricas na ausência de dados não sistemáticos

As equações que permitem o cálculo das probabilidades empíricas de excedência, ou simplesmente posições de plotagem, especificam, com base em uma amostra de tamanho N , a frequência com que determinado evento é igualado ou excedido. Trata-se de atribuir uma probabilidade de excedência às observações sem, no entanto, se comprometer com a distribuição da qual elas se originaram.

As principais equações para o cálculo da probabilidade empírica, na ausência de informação não sistemática são apresentadas por Cunnane (1978). De uma forma geral, tais expressões têm a seguinte forma:

$$\hat{p}_i = \frac{i - a}{N_s^* + 1 - 2a} \quad i = 1, 2, \dots, N_s^* \quad (6.57)$$

onde N_s^* representa o tamanho de uma amostra, ordenada de maneira decrescente, e a é uma constante, cujos diferentes valores correspondem aos casos particulares mostrados na tabela 6.1.

Tabela 6.1 – Valores de a para as equações de probabilidade empírica (Adap. de Naghettini e Pinto, 2007)

a	Posição de plotagem	Atributo
0	Weibull	probabilidades de excedência sem viés
3/8	Blom	população Normal
0,4	Cunnane	quantis sem viés
0,44	Gringorten	populações Exponencial, Gumbel ou GEV
0,5	Hazen	população Gama

6.4.2 Probabilidades de excedência empíricas com dados não sistemáticos

Hirsch e Stedinger (1987) fizeram uma revisão dos diferentes métodos para o cálculo da probabilidade empírica quando a amostra engloba as informações censuradas e não censuradas. Os autores propuseram, ainda, um novo método, posteriormente modificado por

Naulet (2002), para ser aplicado a amostras com dados não sistemáticos. A descrição resumida do método é apresentada a seguir.

Considere uma amostra constituída por dados sistemáticos e não sistemáticos representados pela variável aleatória Z . O cálculo das probabilidades empíricas de excedência pode ser feito de acordo com as seguintes etapas:

- 1) Classificação, em ordem decrescente, da amostra das N^* cheias ($z_1 \geq z_2 \geq \dots \geq z_{N^*}$), constituída
 - pelas N_S^* cheias sistemáticas x_i e N_H^* cheias históricas y_j de intensidade conhecida e
 - pelos valores médios das cheias de intensidade compreendida nos intervalos (x_{L_i}, x_{U_i}) e (y_{L_j}, y_{U_j}) ;
- 2) Classificação, em ordem crescente, do conjunto dos t limiares de referência distintos $(y_{H_1} = 0 < y_{H_2} < \dots < y_{H_t} < y_{H_{t+1}} = +\infty)$, de forma que o limiar y_{H_j} se aplique aos $N_j^<$ anos;
- 3) Cálculo da probabilidade de excedência do limiar j , $p_{e_j} = P[Z \geq y_{H_j}]$, de acordo com a expressão

$$\hat{p}_{e_j} = \hat{p}_{e_{j+1}} + \hat{p}_{c_j} \cdot (1 - \hat{p}_{e_{j+1}}), \quad j = t, t-1, \dots, 1, \quad (6.58)$$

com

$$p_{e_{t+1}} = 0, \text{ pois } y_{H_{t+1}} = +\infty;$$

$$p_{e_1} = 1, \text{ pois } y_{H_1} = 0; \text{ e}$$

$$\hat{p}_{c_j} = \frac{A_j}{A_j + B_j + C_j}, \quad j = t, t-1, \dots, 1, \quad (6.59)$$

onde

$$A_j \text{ é o número de cheias conhecidas no intervalo } (y_{H_j}, y_{H_{j+1}});$$

B_j é o número de cheias conhecidas inferiores a y_{H_j} ; e

$C_j = \sum_{k=1}^j N_k^<$ é o número de cheias desconhecidas inferiores a y_{H_j} .

4) Finalmente, cálculo das probabilidades empíricas de excedência das A_j cheias compreendidas entre os limiares $(y_{H_j}, y_{H_{j+1}})$, utilizando a seguinte expressão:

$$\hat{p}_i = p_{e_{j+1}} + (p_{e_j} - p_{e_{j+1}}) \cdot \left(\frac{i - a}{A_j + 1 - 2a} \right), \quad j = t, t-1, \dots, 1 \text{ e } i = 1, 2, \dots, A_j. \quad (6.60)$$

6.5 Construção da distribuição a priori para o limite superior

A construção da distribuição *a priori* para os parâmetros das distribuições abordadas neste trabalho é a etapa mais importante da inferência bayesiana. Por outro lado, essa é a etapa que envolve maior subjetividade, uma vez que depende do conhecimento do especialista que está realizando a modelagem e de informações extras, tais como as características de eventos extremos em locais diferentes daqueles em que se está realizando o estudo.

Dentre os parâmetros das distribuições discutidas no item 6.2, aquele para o qual se tem informação disponível possibilitando a construção de uma distribuição *a priori* informativa é o limite superior α . Esse parâmetro apresenta uma relação interpretável com as características físicas da bacia, o que facilita a construção de uma distribuição *a priori* que realmente traduza a incerteza que se tem sobre o mesmo.

A distribuição *a priori*, por definição, é a representação matemática do conhecimento do especialista a respeito da quantidade de interesse antes de se observar realizações dessa quantidade. No caso da vazão máxima anual, 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, podem fornecer indícios de cheias extremas (paleocheias) nunca superadas em um longo intervalo de tempo, possibilitando a 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 6.56) e seu uso na construção da

distribuição *a priori* para o limite superior acarretaria em uma duplicidade no uso de informações no modelo estatístico.

As características hidráulicas do trecho fluvial, sobretudo a forma da sua seção fluvial, podem dar informações sobre a cheia máxima fisicamente possível naquele trecho. Essas informações vêm 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 uma tarefa 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 há limite superior. Obviamente, essa 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. Essa definição, em última análise, leva-nos a um problema tão ou mais complexo quanto o problema inicial.

Com isso, resta a definição da distribuição *a priori* por meio da análise de eventos extremos, observados ou estimados, em bacias similares à bacia de interesse. Essa opção é abordada no restante deste item.

6.5.1 Distribuição *a priori* para o limite superior para os Estados Unidos

Conforme será visto no capítulo 7, uma aplicação do método desenvolvido nesta tese é feita na bacia do rio American, no estado americano da Califórnia. Essa bacia possui uma grande disponibilidade de dados. No entanto, são poucas as informações sobre um provável limite superior para as vazões máximas anuais.

Essa bacia conta com uma estimativa de PMF, um conjunto de dados paleohidrológicos e um conjunto de dados de vazão diária observados nos últimos 100 anos. Os dados paleohidrológicos e as vazões observadas são partes integrantes da função de verossimilhança e, assim, não podem ser utilizadas na construção da distribuição *a priori* para o limite superior. O dado restante, a estimativa da PMF, não é suficiente para caracterizar toda a incerteza relacionada ao limite superior. Dessa forma, faz-se necessário recorrer à análise de dados de outras bacias.

A PMF, discutida no item 4.3.3, pode ser considerada como um estimador natural para o limite superior. Entretanto, o limite dado pela PMF não é inequívoco e depende de um conjunto limitado de dados amostrais, o qual, em geral, é insuficiente para caracterizar a complexa variabilidade espaço-temporal das variáveis que lhe dão origem. Assim, a alternativa adotada por Botero (2006) de fixar o limite superior pela PMF e tratá-lo como uma constante, embora válida, não permite a completa análise de seu comportamento. Com efeito, neste trabalho o limite superior é visto como uma quantidade fixa, porém desconhecida, cuja incerteza sobre o seu verdadeiro valor é descrita por uma distribuição de probabilidade e, assim, é passível de inferência e análise por meio de diversas ferramentas estatísticas. Essa vantagem adicional acarreta, no entanto, uma maior complexidade ao modelo.

A situação ideal, no que se refere à construção de uma distribuição *a priori* para o limite superior com base na estimativa da PMF, seria alcançada caso se dispusesse de uma grande amostra de estimativas de PMF para uma mesma bacia, calculadas por métodos idênticos, em diferentes épocas, usando toda informação hidrometeorológica e hidrológica disponível à época de seu cálculo. Nessa situação, seria possível avaliar objetivamente a variabilidade da PMF e elaborar modelos estatísticos capazes de descrever tal variabilidade ou, sob o ponto de vista bayesiano, a variabilidade da PMF constituir-se-ia na base de conhecimento do especialista sobre o limite superior, possibilitando-lhe descrever as incertezas sobre essa variável de maneira mais precisa. Obviamente essa situação ideal não existe, o que leva a adotar abordagens baseadas em dados regionais. Nesse sentido, são propostos dois procedimentos para a construção da distribuição *a priori* para o limite superior: um baseado na distribuição das PMF de várias bacias americanas e outro baseado na curva envoltória de PMF construída para os EUA.

6.5.1.1 Procedimento A – Distribuição das PMF a nível regional

A Comissão de Regulação Nuclear dos Estados Unidos (USNRC, 1977) possui uma compilação de 561 estimativas de PMF em várias bacias americanas. A figura 6.4 mostra a distribuição espacial dessas estimativas. Nessa figura, mostra-se a localização de 494 PMF uma vez que o catálogo do USNRC não define as coordenadas das demais estimativas. As estimativas que não possuem coordenadas localizam-se, em sua maioria, nos estados da região noroeste dos EUA.

Como pode ser visto pelo mapa da figura 6.4, as estimativas de PMF cobrem uma grande diversidade de condições climáticas e hidrológicas. Além disso, essas estimativas foram feitas

para bacias com diferentes áreas de drenagem, o que possibilita a definição da distribuição *a priori* para o limite superior de maneira bastante geral.

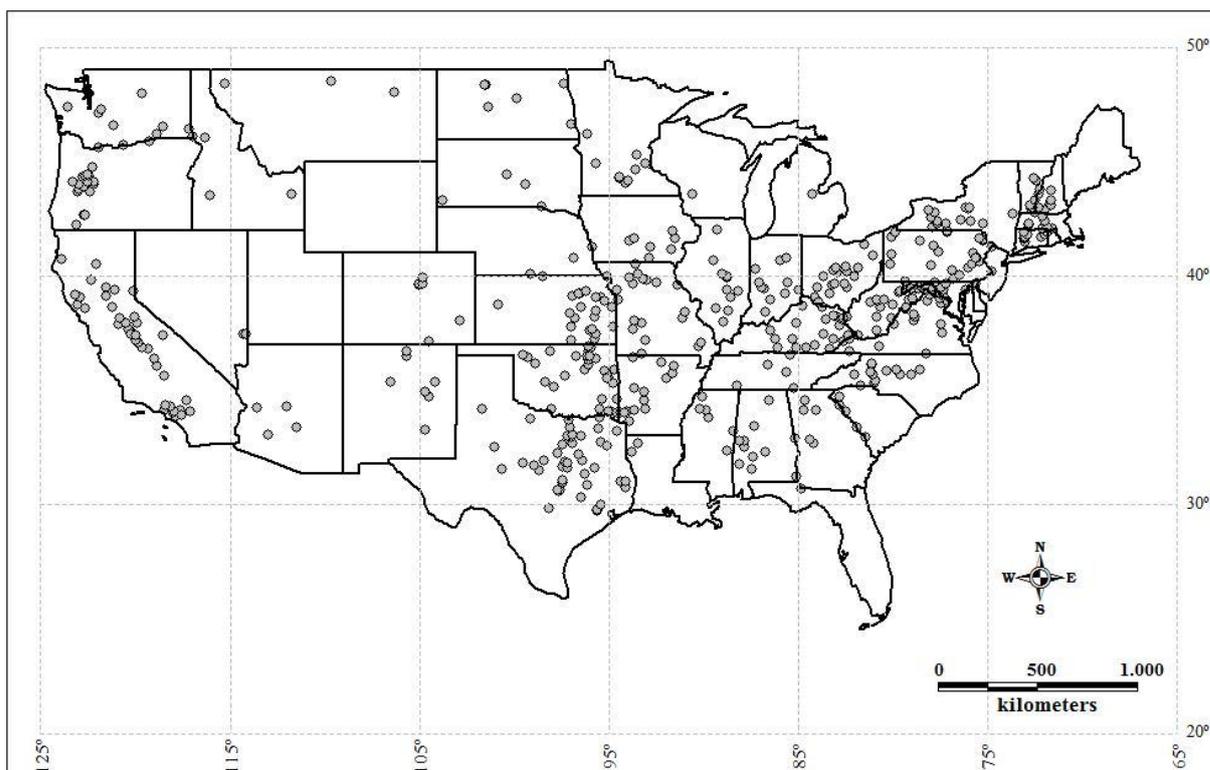


Figura 6.4 – Distribuição espacial das estimativas de PMF nos EUA

Para evitar avaliar as estimativas de PMF calculadas em áreas de drenagem muito pequenas com outras estimadas em áreas de drenagem muito grandes sob o mesmo contexto, dividiu-se a amostra em dez grupos de áreas aproximadamente iguais. Esse procedimento busca, mesmo que de maneira indireta, dividir a amostra em grupos mais homogêneos, sob o ponto de vista de escala e das características hidrometeorológicas utilizadas na estimativa de cada PMF. No entanto, conforme visto no item 4.3.3, cada PMF é estimada a partir de informações particulares de cada local e não tem uma relação evidente com outras estimativas de PMF, exceto nos casos onde a PMF foi estimada por transposição de tormentas de bacias vizinhas. A tabela 6.2 mostra as principais informações de cada grupo.

Tabela 6.2 – Características principais de cada grupo de PMF's analisado

Grupo	N	AD (km ²)	PMF (m ³ /s)			DP	CV
			Mín.	Méd.	Máx.		
1	60	(0, 100)	184	1.056	2.312	552	0,52
2	46	(100, 200)	351	1.828	4.221	824	0,45
3	51	(200, 350)	482	2.753	5.893	1.234	0,45
4	61	(350, 600)	992	3.990	8.272	1.521	0,38
5	55	(600, 900)	1.377	5.626	10.057	2.063	0,37
6	54	(900, 1500)	3.116	7.138	12.748	2.368	0,33
7	48	(1500, 2500)	1.331	9.224	16.119	3.570	0,39
8	55	(2500, 4500)	2.450	11.958	19.462	3.949	0,33
9	57	(4500, 15000)	722	13.286	31.417	6.439	0,48
10	74	(15000, 625000)	2.524	27.561	77.056	17.406	0,63

N - Número de PMF no grupo, AD - Intervalo de área de drenagem das bacias do grupo, DP - Desvio padrão e CV - Coeficiente de variação.

A figura 6.5 mostra os histogramas das estimativas de PMF contidas em cada grupo. As linhas contínuas representam a distribuição Gama ajustada às estimativas de PMF. De uma forma geral, os histogramas mostram que as estimativas de PMF são distribuídas assimetricamente e que distribuições Gama fornecem uma boa descrição da incerteza que se tem sobre ela. Para verificar a aderência da distribuição Gama às estimativas de PMF, foi realizado o teste de Kolmogorov-Smirnov a um nível de significância de 5%, com a não rejeição da hipótese nula para todos os grupos.

Essa simples análise do comportamento dos histogramas de cada grupo mostra que a distribuição Gama é uma candidata razoável para modelar as estimativas de PMF para bacias com áreas de drenagens similares. Certamente, a distribuição *a priori* para o limite superior não será idêntica à distribuição das PMF uma vez que, conforme mencionado neste trabalho, a PMF não é o limite superior. No entanto, considerando um grupo de bacias cujas áreas de drenagem limites estão bastante próximas, o histograma resultante poderia ainda ser assimétrico para direita e refletir a frequência de estimativas de PMF de bacias com áreas de drenagem aproximadamente iguais, porém geograficamente localizadas em regiões distintas. Esse fato pode ser visualmente verificado a partir dos dados da tabela 6.2 e da figura 6.5: o tamanho de cada grupo (faixa de áreas de drenagem das bacias que o compõem) diminui do grupo 10 para o grupo 1, da mesma forma, os dados são mais assimétricos para direita nesse sentido e se ajustam melhor à distribuição Gama. Supondo, agora, que essas bacias estejam localizadas em regiões muito similares sob o ponto de vista climático e hidrológico, pode-se imaginar que o histograma correspondente teria a mesma forma assimétrica, possivelmente com uma menor variância. Em outros termos, no caso limite em que há uma amostra de PMF

para a mesma bacia, a distribuição correspondente se assemelharia àquelas representadas pelas distribuições Gama da figura 6.5, porém com menores amplitudes no eixo da variável.

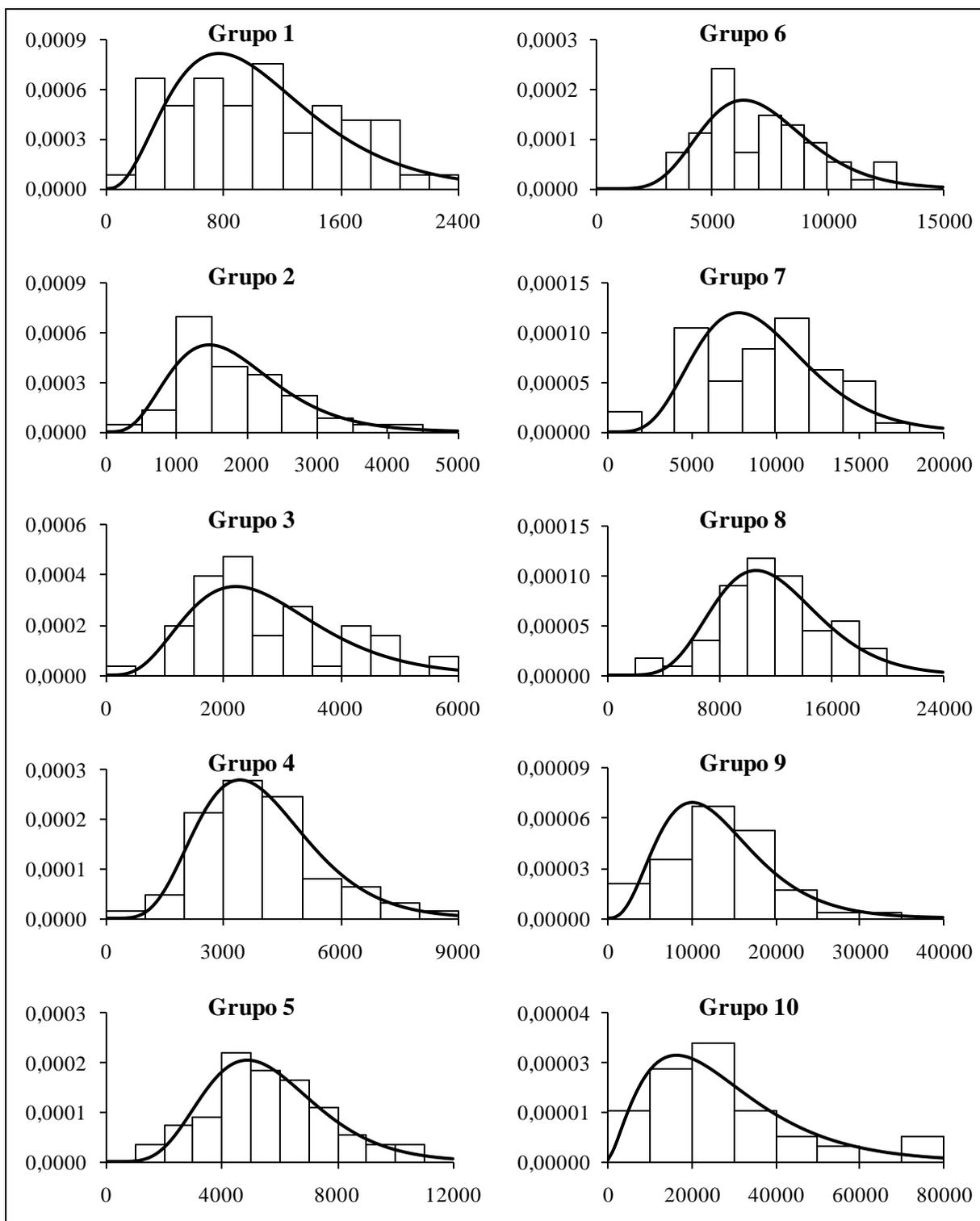


Figura 6.5 – Histogramas para as PMF's de cada grupo e ajuste da distribuição Gama

Com base nesses argumentos e pelo fato de se assumir nesta tese que a PMF é um estimador para o limite superior, é plausível admitir que uma distribuição não limitada e assimétrica a direita, como a Gama, é uma candidata razoável para modelar o limite superior α .

A função densidade de probabilidade de uma variável distribuída de acordo com o modelo Gama, a qual será denotada por $X \sim \text{GAM}(\rho, \beta)$, com parâmetro de escala $\beta \in \mathfrak{R}_+^*$ e de forma $\rho \in \mathfrak{R}_+$, é dada pela seguinte equação:

$$f_x(x/\bar{\Theta}) = \frac{\beta^\rho}{\Gamma(\rho)} x^{\rho-1} \exp\{-\beta x\}, \quad x > 0. \quad (6.61)$$

Para a estimativa dos dois parâmetros da distribuição Gama são necessários os momentos amostrais de ordem 1 e 2, os quais não estão disponíveis para uma determinada bacia, ou seja, não há uma amostra de PMF para uma bacia a partir da qual se pode inferir a média e a variância. Como alternativa, pode-se descrever a distribuição Gama através de duas de suas características: o coeficiente de variação (CV) e a probabilidade de excedência da PMF local. De fato, pelo método dos momentos tem-se que:

$$\rho = \frac{1}{CV^2} \quad (6.62)$$

Embora não seja possível estimar o coeficiente de variação da PMF para uma determinada bacia, pode-se inferir seu valor a partir de sua estimativa regional, tal como aquela apresentada na tabela 6.2.

O parâmetro β pode ser estimado admitindo-se uma probabilidade de excedência p para a estimativa local da PMF, ou seja, o parâmetro β deve ser tal que $P(\alpha \leq \text{PMF}/\rho, \beta) = p$.

A distribuição *a priori* do limite superior deve refletir o conhecimento sobre a vazão máxima que é possível ocorrer em uma dada seção transversal de um certo rio. É necessário assumir, também, que a estimativa da PMF foi feita a partir das melhores informações hidrometeorológicas e hidrológicas disponíveis para a bacia, usando ferramentas de modelagem apropriadas. No entanto, não importando o quão acurado é o valor da PMF, não é possível estabelecer a probabilidade de o limite superior α ser superior à sua estimativa. Dadas essas constatações, é razoável admitir uma probabilidade de 50% de a PMF ser menor que o limite superior, ou seja, $P(\alpha \leq \text{PMF}/\rho, \beta) = 0,5$. Essa declaração probabilística de certo modo reflete a completa incerteza a respeito do evento “o limite superior é menor que a estimativa local para a PMF”, independentemente da melhor ou pior qualidade da estimativa existente da PMF, variável em função da extensão das informações disponíveis e/ou das diferentes características dos métodos empregados.

Sobre o coeficiente de variação, a tabela 6.2 mostra que seus valores variam de 0,33 a 0,63, com uma média de 0,43, sendo que não há nenhuma relação evidente com o tamanho das bacias. Uma vez que esses valores levam em consideração a dispersão horizontal, entre diferentes áreas de drenagem na mesma categoria, e vertical, entre estimativas de PMF de bacias localizadas em áreas geográficas completamente distintas, eles estão certamente superestimados dentro de cada grupo. Apesar desses fatos, de forma a avaliar a variabilidade do limite superior α , pode-se adotar os seguintes valores para CV: 0,3, 0,5 e 0,7. O coeficiente de variação pode ser visto, nesse contexto, como uma medida do conhecimento acerca da estimativa da PMF. Valores menores do coeficiente de variação implicam em uma menor dispersão da distribuição em torno da PMF, ou seja, há um maior conhecimento sobre o valor provável do limite superior.

No capítulo 7 serão vistos mais detalhes sobre as distribuições *a priori* para o limite superior obtidas pelo procedimento apresentado neste item.

6.5.1.2 Procedimento B – Transposição de estimativas de PMF

O procedimento B é baseado na transposição de estimativas de PMF de outras bacias para a bacia de interesse. Analogamente ao procedimento A, são utilizadas as 561 estimativas de PMF compiladas pelo USNRC. Diferentemente do método convencional de se fazer transposição de dados hidrológicos por uma relação linear entre áreas de drenagem, emprega-se aqui a curva envoltória para as PMF's dos EUA construída pelo USNRC (1977). A idéia central desse procedimento é buscar qual seria a estimativa da PMF para a bacia de interesse, caso essa experimentasse as mesmas condições hidrometeorológicas e hidrológicas de outras bacias.

Conforme visto no item 4.3.2, a curva envoltória, seja ela construída a partir das cheias recordes ou a partir de estimativas de PMF, representa o conhecimento disponível a respeito da capacidade máxima de produção de vazão por área de drenagem de uma determinada região. Assim, o uso da curva envoltória, em detrimento à transposição linear pelas áreas de drenagem, tem como objetivo evitar que as PMF's transpostas sejam superiores à experiência regional.

A figura 6.6 mostra as 561 estimativas de PMF plotadas contra suas respectivas área de drenagem, juntamente com a curva envoltória construída pelo USNRC (1977). São mostradas, também, as principais características do método de transposição. Formalmente, se A_i e PMF_i denotam, respectivamente a área de drenagem e a estimativa de PMF para a bacia i e A_0 é a

área de drenagem da bacia de interesse, então a estimativa de PMF transposta, denotada por PMF_0 , é dada por:

$$PMF_0 = PMF_i \left(\frac{A_0}{A_i} \right)^{0,4891} \quad (6.63)$$

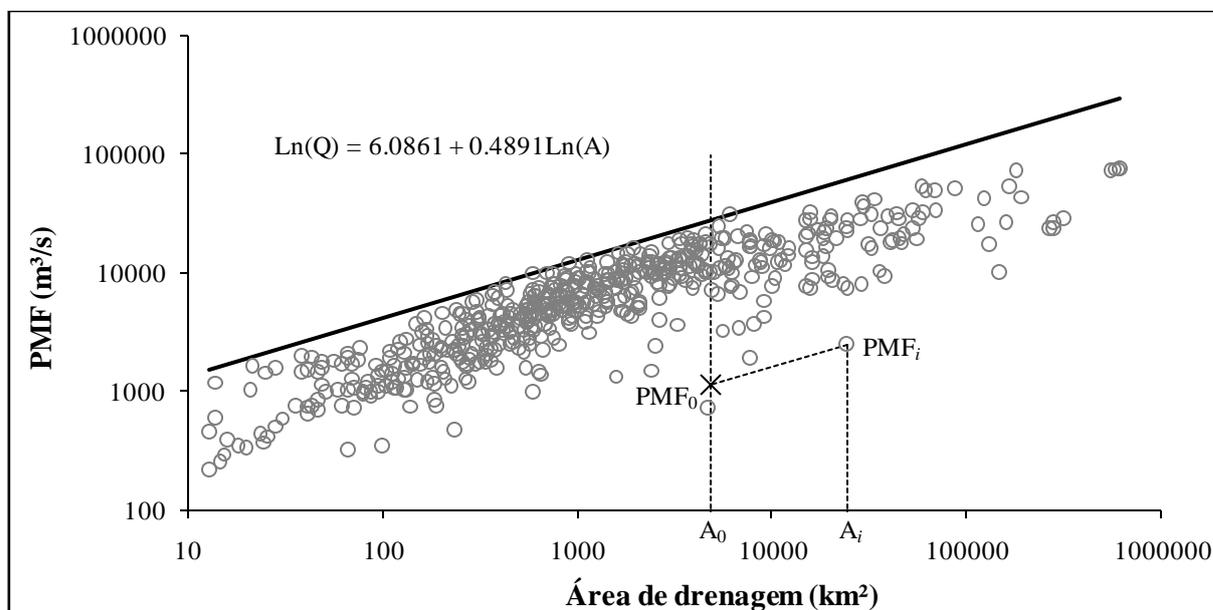


Figura 6.6 – Estimativas de PMF versus área de drenagem para 561 bacias americanas

Se a equação 6.63 for aplicada a todas as 561 bacias, no final, ter-se-á uma amostra de 561 estimativas de PMF para uma mesma área de drenagem. Isso equivale a dizer que uma área de drenagem em particular está sujeita a todas as características climáticas e hidrológicas incorporadas pelo conjunto de bacias analisadas.

Na seqüência, a partir desse conjunto de PMF transpostas, são feitas conjecturas a respeito da forma da distribuição *a priori* para o limite superior das vazões máximas anuais. No capítulo 7 serão vistos mais detalhes sobre as distribuições *a priori* para o limite superior obtidas pelo procedimento apresentado neste item.

6.5.2 Distribuição *a priori* para o limite superior para a Espanha

Conforme será visto no capítulo 7, uma segunda aplicação do método desenvolvido nesta tese é feita à bacia do rio Llobregat, em Pont Du Vilomara, na Espanha. Essa bacia foi objeto do estudo de Botero (2006) e possui uma boa disponibilidade de dados. No entanto, são poucas as informações sobre um provável limite superior para as vazões máximas anuais.

Diferentemente do rio American, essa bacia não possui uma estimativa de PMF. Para suprir essa deficiência, seguindo as recomendações de Botero (2006), o limite superior foi estimado

através da curva envoltória de vazões recordes. Além dessa estimativa, a bacia conta com um conjunto de dados paleohidrológicos e um conjunto de dados de vazão diária observada entre os anos de 1946 e 1988. Os dados paleohidrológicos e as vazões observadas são partes integrantes da função de verossimilhança e, assim, não podem ser utilizados na construção da distribuição *a priori* para o limite superior. O dado restante, a estimativa de cheia recorde, não é suficiente para caracterizar toda a incerteza relacionada ao limite superior. Dessa forma, faz-se necessário recorrer à análise de dados de outras bacias. Da mesma forma que no rio American, a distribuição *a priori* para o limite superior foi construída de acordo com os procedimentos A e B, discutidos nos itens 6.5.1.1 e 6.5.1.2.

6.5.2.1 Procedimento A – Distribuição das cheias recordes a nível regional

O *Centro de Estudios y Experimentación de Obras Públicas (CEDEX)*, que é o órgão responsável pelo monitoramento de vazões na Espanha, possui uma compilação de 486 registros de vazões recordes (maior valor observado) em várias bacias espanholas. Desses registros, 242 são cheias naturais, enquanto o restante é proveniente de bacias com regularização de vazões. O gráfico da figura 6.7 mostra os registros de vazões naturais em escala logarítmica.

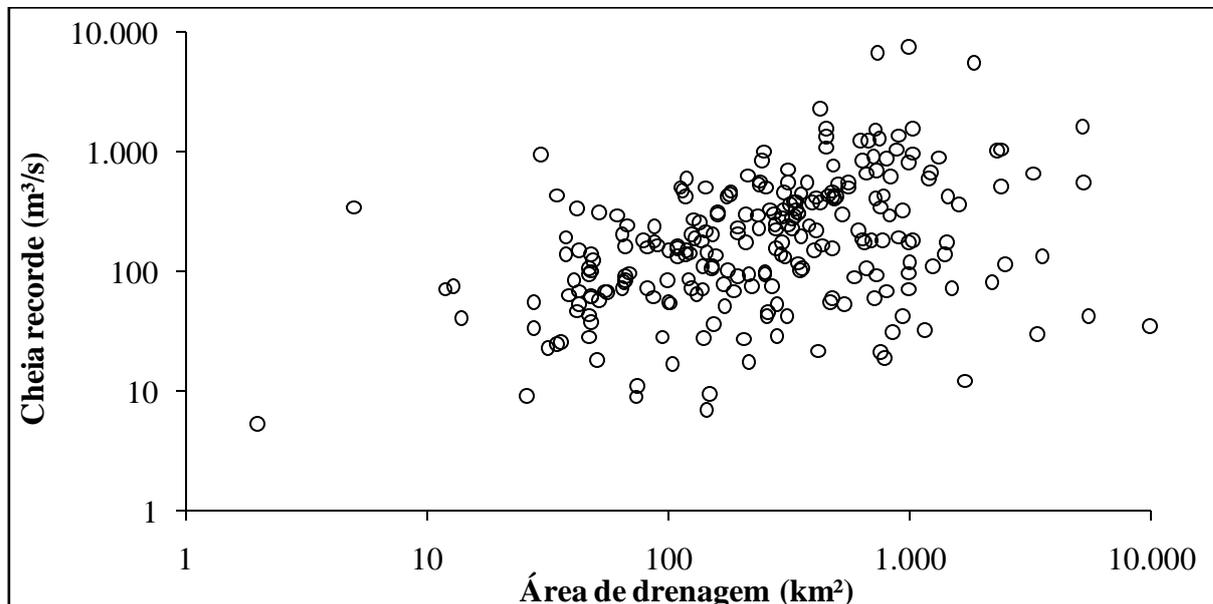


Figura 6.7 – Cheias recordes observadas em bacias da Espanha, em função de suas respectivas áreas de drenagem

Para evitar avaliar as cheias recordes observadas em áreas de drenagem muito pequenas com outras estimadas em áreas de drenagem muito grandes, sob o mesmo contexto, dividiu-se a amostra em quatro grupos. Esse procedimento busca, mesmo que de maneira indireta, dividir a amostra em grupos mais homogêneos, sob os pontos de vista de escala e das características

hidrológicas formadoras das cheias recordes. A tabela 6.3 mostra as principais informações de cada grupo.

Tabela 6.3 – Características principais de cada grupo de cheias recordes

Grupo	N	AD (km ²)	Cheia recorde (m ³ /s)			DP	CV
			Mín.	Méd.	Máx.		
1	84	(0, 150)	5,20	144,96	932,60	153,68	1,06
2	88	(150, 500)	17,40	322,55	2.267,30	348,47	1,08
3	42	(500, 1.000)	18,80	626,98	7.576,00	1.177,90	1,89
4	28	(1.000, 10.000)	11,90	632,36	5.600,00	1.076,12	1,70

N - Número de cheias recordes no grupo, AD - Intervalo de área de drenagem das bacias do grupo, DP - Desvio padrão e CV - Coeficiente de variação

A figura 6.8 mostra os histogramas das cheias recordes observadas em cada grupo. As linhas contínuas representam a distribuição Gama ajustada em cada grupo. De uma forma geral, os histogramas mostram que as cheias recordes, da mesma forma que as estimativas de PMF dos EUA, são distribuídas assimetricamente e que distribuições Gama fornecem uma boa descrição da incerteza sobre o comportamento de tais cheias recordes. Para verificar a aderência da distribuição Gama às cheias recordes observadas, foi realizado o teste de Kolmogorov-Smirnov a um nível de significância de 5%, sendo que os todos os grupos passaram no teste. Essa simples análise do comportamento dos histogramas de cada grupo mostra que a distribuição Gama é uma candidata razoável para modelar as cheias recordes.

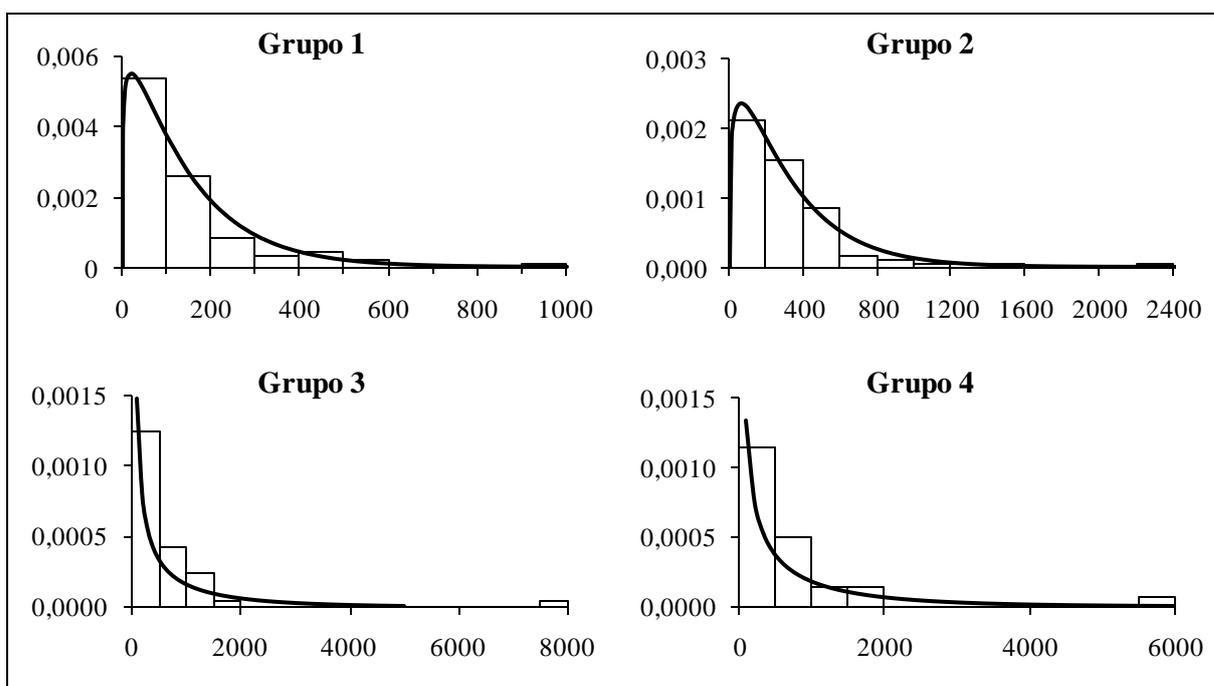


Figura 6.8 – Histogramas para as cheias recordes de cada grupo e ajuste da distribuição Gama

Com base nesses argumentos e pelo fato de se assumir nesta tese que a cheia recorde também é um estimador para o limite superior, é plausível admitir que uma distribuição não limitada e assimétrica a direita, como a Gama, é uma candidata razoável para modelar o limite superior α . Neste ponto, vale salientar que, embora a cheia recorde seja um estimador para o limite superior, é esperado que esse seja um estimador pior que a PMF. De fato, as cheias recordes dependem fortemente do tamanho das amostras de vazões observadas. Uma vez que, em geral, essas amostras são pequenas, não se espera que as mesmas contenham valores próximos ao limite superior para a bacia. Apesar desse fato, no caso da Espanha, não há outra alternativa para a construção da distribuição *a priori* para o limite superior.

Para a aplicação do procedimento A, conforme visto no item 6.5.1.1, é necessária uma estimativa para o coeficiente de variação e uma estimativa da probabilidade de o limite superior ser maior que seu estimador. Com base nos dados da tabela 6.3, adotou-se $CV = 1,0$ e $CV = 2,0$. Além disso, adotou-se uma probabilidade $p = 50\%$ de o limite superior α ser maior que a cheia recorde pontual.

No capítulo 7 serão vistos mais detalhes sobre as distribuições *a priori* para o limite superior obtidas pelo procedimento apresentado neste item.

6.5.2.2 Procedimento B – Transposição de cheias recordes

O procedimento B baseia-se na transposição de estimativas de cheias recordes de outras bacias para aquela de interesse. Para a aplicação do método, foram utilizadas as 242 cheias recordes naturais observadas na Espanha, compiladas pelo CEDEX. A transposição das cheias foi feita por meio da curva envoltória definida por Rodier e Roche (1984), com base no trabalho de Francou e Rodier (1967), para as 41 maiores cheias registradas no mundo.

A figura 6.9 mostra as 242 cheias observadas na Espanha juntamente com a curva envoltória construída pelo Rodier e Roche (1984). De acordo com a curva envoltória, a cheia transposta é dada por:

$$Q_0 = Q_i \left(\frac{A_0}{A_i} \right)^{0,4}, \quad (6.64)$$

onde A_i e Q_i denotam, respectivamente, a área de drenagem e a cheia recorde observada na bacia i e A_0 e Q_0 denotam, respectivamente, a área de drenagem e a cheia recorde na bacia de interesse.

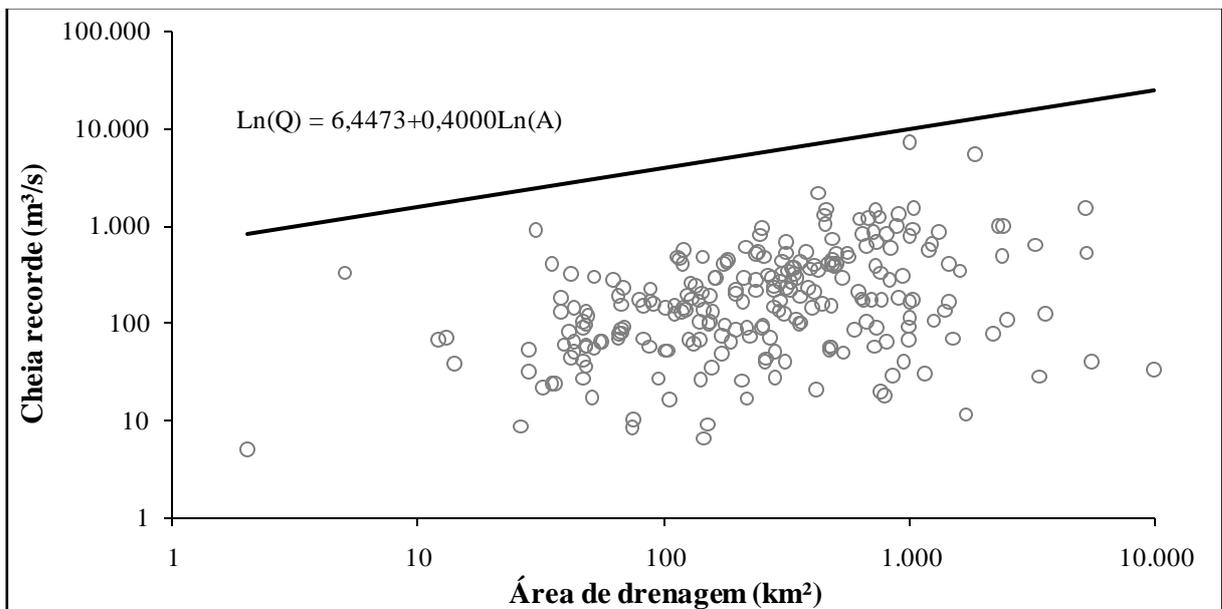


Figura 6.9 – Cheias recordes observadas na Espanha e curva envoltória para as 41 cheias recordes mundiais

Na seqüência, a partir desse conjunto de cheias transpostas, são feitas conjecturas a respeito da forma da distribuição *a priori* para o limite superior das vazões máximas anuais. No capítulo 7 serão vistos mais detalhes sobre as distribuições *a priori* para o limite superior obtidas pelo procedimento apresentado neste item.

6.5.3 Distribuição *a priori* para o limite superior para o Brasil

Conforme será visto no capítulo 7, uma última aplicação do método desenvolvido nesta tese é feita na bacia do rio Pará, em Ponte do Vilela, Minas Gerais, Brasil. Essa bacia foi objeto de outro estudo por Fernandes e Naghettini (2008), os quais aplicaram uma série de métodos de análise de frequência de cheias máximas e se diferencia das aplicações anteriores por não haver nenhum estudo paleohidrológico disponível.

6.5.3.1 Procedimento A – Distribuição das PMF a nível regional

O Comitê Brasileiro de Barragens (CBDB, 2002) publicou, em 2002, uma compilação de dados sobre 61 aproveitamentos energéticos do Brasil. Desses, apenas 34 contam com estimativas de PMF, constituindo a base para a especificação do limite superior para Ponte do Vilela. A tabela 6.4 mostra os dados de PMF utilizados.

Tabela 6.4 – Características principais das estimativas de PMF no Brasil (Fonte: CBDB, 2002)

Local	Rio - Bacia	Área de Drenagem km²	PMF diária m³/s
Cajuru	Paranaíba	2.440	1.210
Queimado	São Francisco	3.773	2.100
Paraibuna/Paraitinga	Paraibuna	4.160	5.400
Euclides da Cunha	Pardo	4.366	3.470
APM - Manso	Manso/Cuiabá	9.365	6.271
Funil (RJ)	Paraíba do Sul	13.410	4.900
Funil (MG)	Grande	15.153	7.750
Nova Ponte	Paranaíba	15.300	8.630
Irapé	Jequitinhonha	16.200	7.950
Miranda	Paranaíba	17.800	8.800
Capim Branco I	Paranaíba	18.300	8.940
Capim Branco II	Paranaíba	19.100	9.218
Corumbá I	Corumbá	27.800	8.200
Emborcação	Paranaíba	28.900	9.280
Machadinho	Pelotas/Uruguai	35.800	39.750
Salto Santiago	Iguaçu/Paraná	43.300	24.600
Itá	Uruguai	44.500	52.800
Três Marias	São Francisco	50.600	9.980
Furnas	Grande	52.000	18.802
Pedra do Cavalo	Paraguaçu	53.650	15.160
Salto Caxias	Iguaçu/Paraná	57.000	49.600
Serra da Mesa	Tocantins	57.062	22.780
Cana Brava	Tocantins	57.780	17.802
Aimorés	Doce	62.167	16.813
Salto da Divisa	Jequitinhonha	66.700	17.755
Mascarenhas de Moraes	Grande	74.300	16.107
Porto Colombia	Grande	78.400	16.000
Itumbiara	Paranaíba	95.000	22.100
São Simão	Paranaíba	171.000	27.200
Ilha Solteira	Paraná	375.460	55.230
Jupia	Paraná	470.000	60.790
Porto Primavera	Paraná	575.000	62.040
Tucuruí	Tocantins	758.000	114.300
Itaipú	Paraná	820.000	72.020

Uma vez que a amostra é composta de apenas 34 valores, não foi feita uma divisão dos dados em grupos. Assim, a análise se estende a bacias com áreas de drenagem bastante distintas. A figura 6.10 mostra o histograma das PMF com o ajuste da distribuição Gama. Para verificar a aderência da distribuição Gama às estimativas de PMF, foi realizado o teste de Kolmogorov-Smirnov a um nível de significância de 5%, sendo que o grupo das 34 PMF passou no teste.

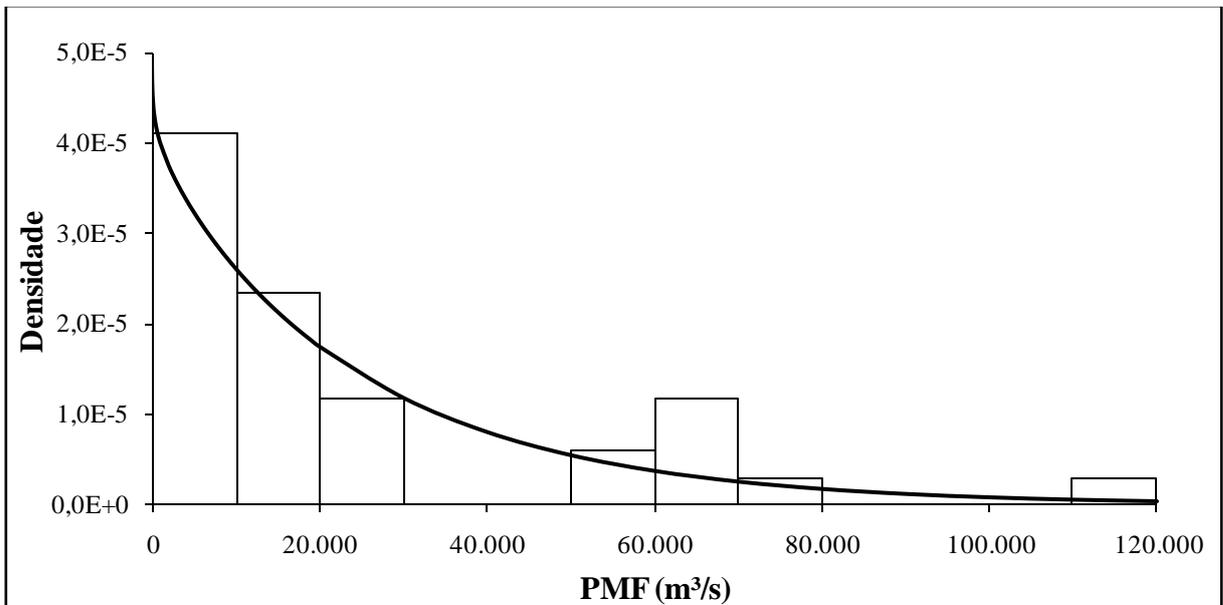


Figura 6.10 – Histograma para as PMF's do Brasil e ajuste da distribuição Gama

O coeficiente de variação calculado para a amostra de PMF é de 1,01. No entanto, conforme mostrado na tabela 6.4, a amostra reúne bacias com áreas de drenagem bastante diferentes e PMF's variando entre 1.210 m³/s e 114.300 m³/s. Esse fato, com certeza, infla demasiadamente o valor do coeficiente de variação. A figura 6.11 mostra a variação do CV com a área de drenagem. Para bacias com até 200.000 km², o CV está limitado a 0,6 e a partir desse ponto há um salto para 1,0. Uma vez que a bacia do rio Pará, em Ponte do Vilela, tem uma área de 1.620 km², adotou-se um CV igual a 0,6 para a especificação da distribuição *a priori* de acordo com o procedimento A. O CV adotado é da mesma ordem de grandeza dos valores encontrados para os EUA.

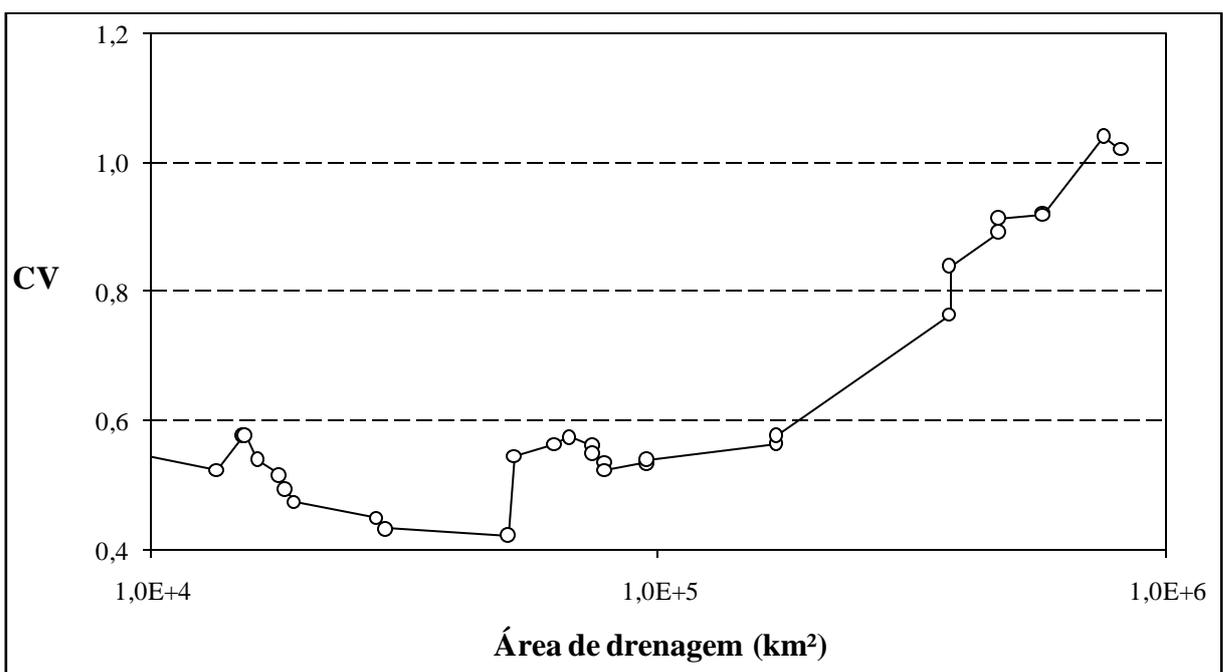


Figura 6.11 – Variação do CV das PMF's do Brasil com a área de drenagem

6.5.3.2 Procedimento B – Transposição de estimativas de PMF

Para a aplicação do procedimento B, construiu-se inicialmente a curva envoltória para o Brasil. A curva foi obtida pela aplicação da fórmula de Myers (equação 4.19) aos logaritmos da PMF e da área de drenagem, de acordo com o método discutido no item 4.3.2. O procedimento adotado foi o de construir uma curva envoltória para todos os dados e outra somente para aqueles de bacias próximas à região sudeste. A figura 6.12 mostra os resultados encontrados. A inclinação de cada curva foi estimada pelo coeficiente angular da reta obtida pela regressão linear dos dados de cada conjunto.

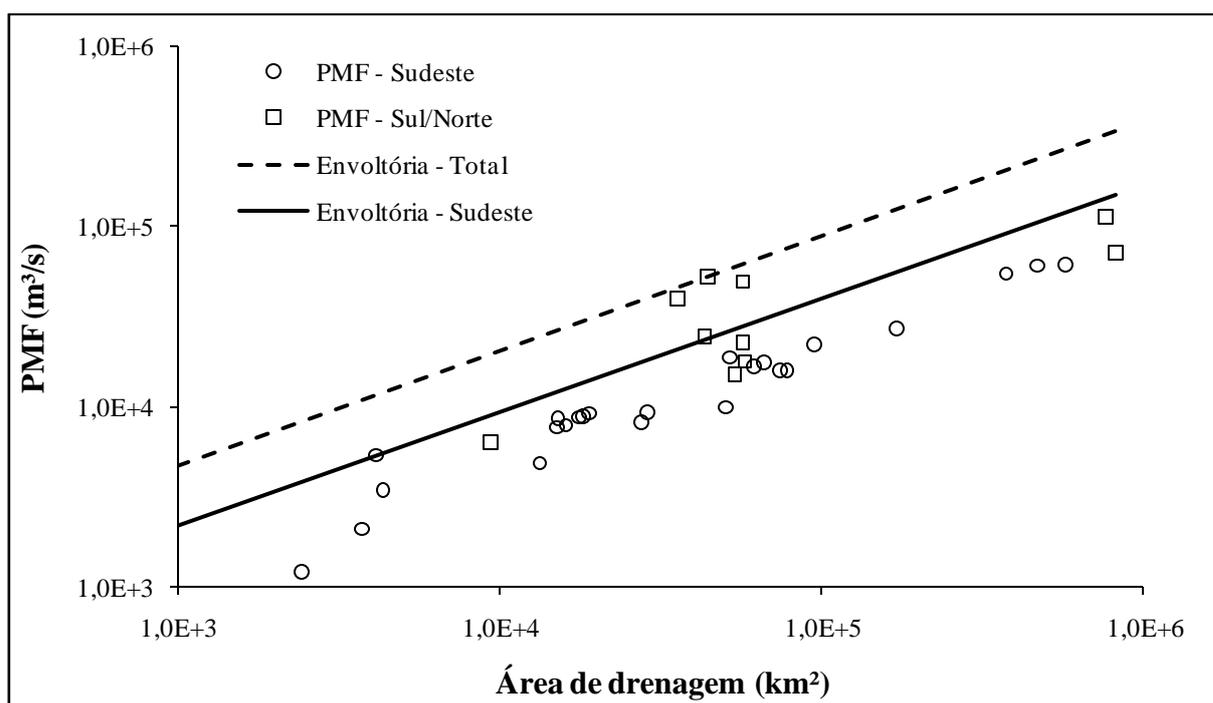


Figura 6.12 – Curvas envoltórias de PMF para o Brasil e para a região sudeste

As curvas mostradas na figura 6.12 têm as seguintes formas:

$$Q_{BRASIL} = 57,2300A^{0,6379} \quad (6.65)$$

e

$$Q_{SUDESTE} = 28,9222A^{0,6275} \quad (6.66)$$

A curva envoltória para o Brasil fornece valores aproximadamente 125% maiores que a curva envoltória para o sudeste. Dada essa grande diferença e o fato de que a aplicação do método será no sudeste, adotou-se a curva envoltória do sudeste como referência para a aplicação do procedimento B na avaliação da distribuição *a priori* para Ponte do Vilela.

No capítulo 7 serão vistos mais detalhes sobre a construção da distribuição *a priori* pelos procedimentos A e B para a bacia do rio Pará.

6.6 Conclusão

Neste capítulo foram descritas as partes constituintes do método proposto. Inicialmente foram descritas as distribuições de probabilidades limitadas superiormente, sendo que a LN4 e a EV4 foram adotadas para as aplicações propostas. Na seqüência, demonstrou-se a construção da função de verossimilhança que agrega a informação referente aos dados não sistemáticos e sistemáticos de vazão. Por fim, foram descritos os procedimentos propostos para a construção das distribuições *a priori* para o limite superior. Além disso, foram apresentadas as características das distribuições *a priori* para o limite superior para os EUA, Espanha e Brasil.

As etapas a serem seguidas para a aplicação do método proposto nesta tese, que tem como arcabouço lógico a análise bayesiana, podem ser assim resumidas:

1. Coleta e avaliação dos dados referentes às cheias máximas anuais: nesta etapa são avaliados os estudos paleohidrológicos disponíveis e as estações fluviométricas localizadas na bacia de interesse;
2. Construção das distribuições *a priori* para o limite superior: esta etapa se refere à aplicação dos procedimentos A e B descritos no item 6.5. Em suma tem-se:
 - Coleta de estimativas de PMF (ou vazões recordes) na região onde está inserida a bacia em estudo;
 - Divisão da amostra de PMFs, quando constituída por estimativas feitas em bacias com áreas de drenagem muito diferentes, em grupos com áreas de drenagem aproximadamente iguais;
 - Avaliação dos histogramas das estimativas em cada grupo de maneira a identificar sua forma;
 - Definição do coeficiente de variação para o limite superior e da probabilidade de a PMF local (ou vazão recorde) ser maior que o limite superior;
 - Transposição da amostra de PMFs (ou de vazões recordes), por meio da curva envoltória, para o local de interesse;

- Ajuste de uma distribuição de probabilidades aos dados transpostos de PMFs (ou de vazões recordes);
3. Construção das distribuições *a priori* para os demais parâmetros: quando for possível encontrar uma relação evidente entre os demais parâmetros e as características físicas da bacia pode-se construir distribuições *a priori* informativas para estes. Quando não for possível, devem ser adotadas distribuições *a priori* não informativas, baseadas nos valores possíveis para o parâmetro de interesse;
 4. Obtenção das estatísticas *a posteriori* das quantidades de interesse (parâmetros e quantis): as estatísticas *a posteriori* são obtidas pela aplicação do Teorema de Bayes. Neste caso, de forma a evitar o cálculo da constante de normalização, podem ser utilizados algoritmos MCMC tal como o de Metropolis. Na aplicação desses algoritmos devem ser consideradas as seguintes questões:
 - Tamanho da amostra – deve ser obtida uma amostra suficientemente grande para que as estatísticas *a posteriori* (média, CV, etc.) dos parâmetros e quantis não sejam influenciadas pelo seu tamanho. Não há um critério para a definição do tamanho da amostra e este deve ser obtido por meio da análise de amostras de vários tamanhos;
 - *Lag* – uma vez que os dados gerados via MCMC são bastante correlacionados, devem ser descartadas algumas realizações do algoritmo em intervalos pré-estabelecidos (*lag*). A escolha do tamanho do *lag* é feita por tentativa e erro, avaliando a autocorrelação entre as realizações do algoritmo;
 - *Burn in* – o termo *burn in* refere-se ao número de realizações necessárias para que a cadeia de Markov “esqueça” seu estado inicial. Gilks et al. (1996) avaliam que valores entre 1 e 2% do total de realizações são suficientes para que a cadeia de Markov atenda a esse pré-requisito;
 5. Análise dos resultados para cada parâmetro: esta etapa busca a identificação de possíveis inconsistências nos modelos adotados e estimativas pontuais para os parâmetros das distribuições de probabilidades. Deve ser avaliada a média e o coeficiente de variação *a posteriori* dos parâmetros, bem como o intervalo de credibilidade da estimativa. Uma atenção especial deve ser dada à análise dos resultados *a posteriori* para o limite superior já que este exerce um papel central em todas as etapas metodológicas listadas;

6. Construção da curva de quantis: a curva de quantis ou a distribuição preditiva *a posteriori* é estimada por integração de Monte Carlo (vide item 5.3) a partir dos valores da distribuição *a posteriori* dos parâmetros. Os quantis estimados devem ser plotados juntamente com os quantis observados de forma a avaliar a aderência do modelo em análise. Os dados observados não sistemáticos são plotados de acordo com o método descrito no item 6.4.2. Quando a disponibilidade se restringir aos dados sistemáticos, esses devem de plotados de acordo com o método exposto no item 6.4.1. A qualidade do modelo pode ser avaliada pela sua capacidade preditiva e descritiva, assim como pelo intervalo de credibilidade para cada quantil;
7. Avaliação global do método: por fim, os resultados obtidos devem ser avaliados sob à luz de outros métodos para a estimação de cheias extremas. Obviamente, esta avaliação não deve ser feita apenas pela aderência de cada modelo aos dados observados, mas também levando em consideração a estrutura de cada método.

7 APLICAÇÕES DO MÉTODO PROPOSTO: RESULTADOS E DISCUSSÃO

7.1 Introdução

Neste capítulo, faz-se uma descrição das aplicações do método proposto nesta tese. Serão descritas as aplicações feitas às 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 permitirá uma comparação com o método proposto. 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 permite 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.

7.2 Aplicação para a bacia do rio American em Folsom

7.2.1 A bacia do rio American

O método proposto nesta tese foi aplicado à bacia do rio American, na seção do eixo da barragem de Folsom, localizada no estado americano da Califórnia. Essa bacia foi escolhida pelo grande número de informações sistemáticas e não sistemáticas disponíveis, além dos dados de PMF em nível local e regional. Outro fato que favorece a escolha dessa bacia é o grande número de estudos ali realizados, o que permite a comparação do método aqui proposto em relação aos resultados de outros métodos.

O reservatório de Folsom, cuja barragem é mostrada na figura 7.1, localiza-se imediatamente a montante da cidade de Sacramento. A figura 7.2 mostra a localização da bacia em estudo. De acordo com NRC (1999), a cidade de Sacramento foi fundada no final do século XVIII, sobre áreas ribeirinhas dos rios American e Sacramento, sendo que seu grande desenvolvimento ocorreu logo após a descoberta de ouro na bacia, em 1848. Desde então, a cidade tem sofrido com frequentes inundações, o que tem levado as autoridades locais a

realizarem estudos para avaliar o risco de cheias, bem como a promoção de políticas públicas no sentido de ordenar o uso e a ocupação da bacia.



Figura 7.1 – Vista de jusante da barragem de Folsom
(Fonte: <http://www.panoramio.com/photo/2596279>, acessado em 12/08/2008)

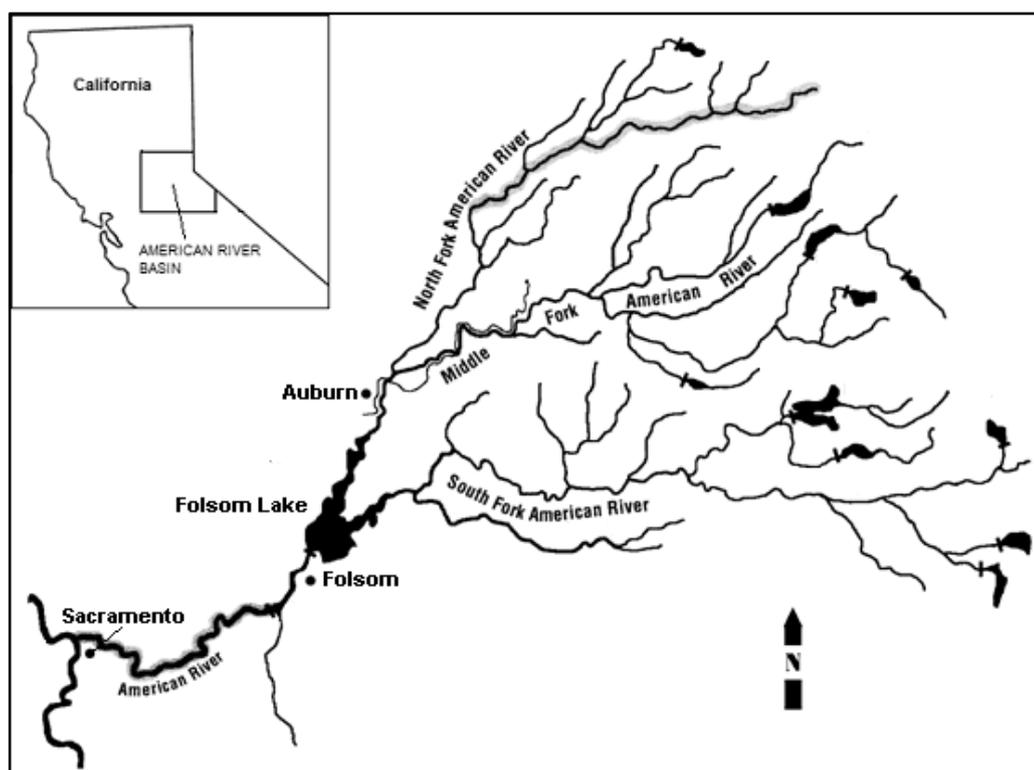


Figura 7.2 – Localização da bacia do rio American (Fonte: USBR, 2002)

Atualmente, mais de 400.000 pessoas e cerca de \$40 bilhões em propriedades estão vulneráveis à inundação, incluindo a maior parte do comércio da cidade e áreas governamentais, uma vez que se trata da capital do estado da Califórnia. Medidas estruturais, tais como diques de contenção nos rios Sacramento e American, têm sido implantadas desde a fundação da cidade. Adicionalmente, a construção do reservatório de Folsom, em 1955, e a grande quantidade de pequenos reservatórios a montante de Sacramento têm contribuído para a atenuação dos efeitos das cheias na bacia. No entanto, cheias recentes, como a de 1997 e a de 1986, mostraram que as medidas estruturais, por si só, são incapazes de garantir a segurança da população e das construções às margens dos rios da bacia. Assim, novos estudos e novas medidas estão sendo tomadas. O NRC (1999) cita, por exemplo, a opção da construção de uma bacia de detenção seca com custo estimado em mais de \$1 bilhão (em 1999) em Auburn, no braço norte do rio American, a montante do reservatório de Folsom.

A preocupação constante dos habitantes de Sacramento e vizinhanças com os danos causados por cheias levou a um grande esforço no que se refere à coleta de dados e informações sobre cheias e à elaboração de estudos visando um melhor gerenciamento do risco hidrológico. Como mencionado anteriormente, a grande disponibilidade de dados e estudos relativos a eventos hidrológicos foram fatores preponderantes na escolha dessa bacia para a aplicação proposta nesta tese.

7.2.2 Dados hidrológicos sistemáticos e não sistemáticos

7.2.2.1 Dados sistemáticos

Na bacia do rio American, o USGS (*United States Geological Service*) possui registros de vazão em 62 pontos monitorados por estações fluviométricas. De acordo com USBR (2002), atualmente 20 estações continuam em operação. A estação fluviométrica de interesse para a aplicação do método é a American River, em Fair Oaks, cujo código dado pelo USGS é 11446500. Essa estação possui dados desde 1905 até os dias atuais e está localizada imediatamente a jusante do reservatório de Folsom.

A área de drenagem a montante de Fair Oaks é de 4.890 km², aproximadamente 1,5% maior que a área de drenagem na seção da barragem de Folsom (4.820 km²). Assim, as vazões não regularizadas nos dois pontos podem ser consideradas equivalentes (USBR, 2002). Na presente aplicação, não será feita distinção entre as vazões em Folsom e em Fair Oaks.

As vazões no rio American são bastante influenciadas pela regularização promovida pelos reservatórios da bacia. USBR (2002) aponta que, a montante de Folsom, existem 58

reservatórios, dos quais 16 exercem um relativo impacto nas vazões afluentes a Folsom. Cinco reservatórios controlam 90% do armazenamento a montante de Folsom e 14% da área de drenagem. Há pouca informação sobre o controle e operação desses reservatórios, uma vez que em sua maioria são de propriedade privada. No entanto, observações do USACE durante grande eventos chuvosos mostraram que os reservatórios a montante de Folsom reduzem o pico e o volume de grandes cheias entre 14 e 18% na entrada do lago de Folsom.

As vazões em Fair Oaks foram reavaliadas por USACE (1998) de forma a estabelecer sua condição natural, sem o efeito de regularização. Para tal, as vazões médias diárias em Fair Oaks foram combinadas com a variação diária do armazenamento do lago de Folsom e com a variação diária do armazenamento dos demais reservatórios.

Seguindo os critérios adotados por USBR (2002), os anos hidrológicos de 1910, 1912, 1913, 1918, 1929, 1977 e de 1987 a 1996 não foram incluídos na análise devido a dados faltosos, impossibilidade de se obter estimativas não regularizadas de cheias ou por essas terem sido originadas por degelo. Para esses anos, USBR (2002) estimou um nível de censura na marca de 4.248 m³/s. As principais características estatísticas para essa série são:

Mínimo:	184 m ³ /s
Máximo:	8.438 m ³ /s
Média:	1.943 m ³ /s
Desvio padrão:	1.853 m ³ /s
Coefficiente de variação:	0,95
Coefficiente de assimetria:	1,76

A figura 7.3 mostra os dados sistemáticos considerados na análise. De forma a manter essa aplicação na mesma base de tempo do estudo do USBR (2002) e assim possibilitar uma comparação entre os resultados, foram utilizados os dados até o ano hidrológico de 2000.

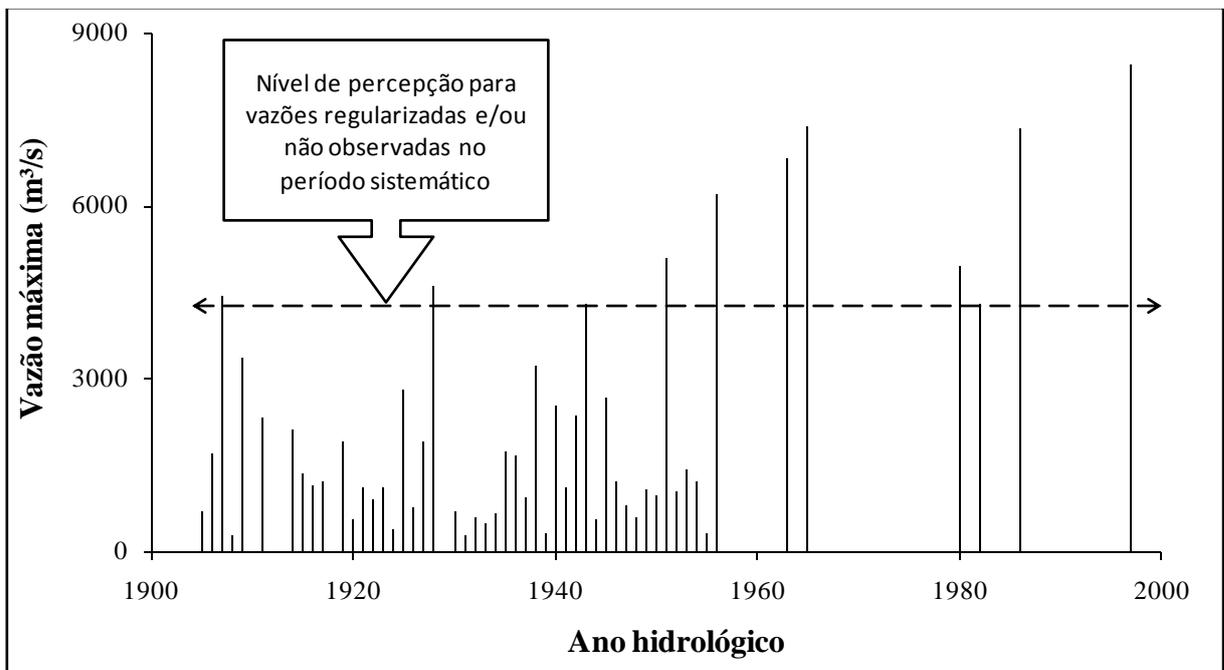


Figura 7.3 – Dados sistemáticos do rio American em Fair Oaks

7.2.2.2 Dados não sistemáticos

USBR (2002) realizou uma extensa análise de indicadores de cheias históricas na bacia do rio American. Foram analisadas as evidências arqueológicas, obtidas por meio de escavação em várias partes da bacia, de comunidades que se estabeleceram na região nos últimos 2500 anos e perfis estratigráficos de depósitos sedimentares ao longo do curso d'água. Três locais foram escolhidos para se estabelecer a cronologia das cheias no rio American:

- 1) Fair Oaks: este local está centrado na estação fluviométrica 11446500. Nesse ponto, o estudo teve como objetivo estabelecer a magnitude da cheia de 1862 e permitir a calibração do modelo hidráulico utilizado nos demais locais;
- 2) Rossmoor: este local possui dois sítios arqueológicos e está localizado em uma ampla área plana na margem esquerda do rio American. Uma característica favorável desse local é a relativa preservação das áreas marginais ao rio em relação ao desenvolvimento urbano experimentado pela região no último século. Além disso, as atividades agrária e minerária de antigas comunidades fornecem importantes informações para a reconstrução das cheias paleohidrológicas; e
- 3) Escola da Cordova: este local foi escolhido pelo USBR devido à presença de material do período pré-histórico em boas condições de preservação em quatro sítios arqueológicos.

A figura 7.4 mostra os locais considerados por USBR (2002) para estabelecer a cronologia das cheias do rio American.



Figura 7.4 – Locais disponíveis para análise de paleocheias no rio American próximo a Fair Oaks (adapt. USBR, 2002)

Os estudos arqueológicos realizados nos três locais acima descritos permitiram estabelecer vários níveis d'água alcançados pelas cheias passadas nessa bacia. A combinação dos dados históricos, arqueológicos, estratigráficos e geomorfológicos possibilitou estabelecer o número de cheias em cada nível, bem como estimar suas respectivas datas aproximadas de ocorrência.

Na seqüência, USBR (2002) estimou a vazão necessária para atingir cada um dos níveis paleohidrológicos por meio do modelo hidráulico bidimensional TRIMR2D. Uma aplicação desse modelo no estabelecimento de limites paleohidrológicos pode ser encontrada em Denlinger et al. (2002). A essência da modelagem realizada pelo USBR está na determinação das vazões mínima e máxima capazes de produzir uma força de arraste tal que sedimentos e objetos arqueológicos possam ser deslocados e transportados com o pico da cheia. No caso do rio American, os perfis estratigráficos e os estudos arqueológicos de Rossroom e de Cordova permitiram avaliar o número e a cronologia das cheias e a modelagem hidráulica possibilitou a determinação do pico de vazão capaz de produzir tal perfil estratigráfico.

Ainda no que se refere à modelagem hidráulica, vale salientar que todos os limites paleohidrológicos foram estabelecidos a partir de mapas altimétricos confeccionados pelo USACE no início do século XX. De acordo com USBR (2002), embora os mapas atuais sejam mais precisos, eles agregam as mudanças geomorfológicas da bacia ocorridas no último século. Entre essas mudanças, destacam-se: as mudanças das áreas ribeirinhas pela agricultura e pela mineração, modificação do curso d'água após a construção do reservatório de Folsom e alterações da bacia devido ao desenvolvimento urbano de Sacramento e da construção de

diques ao longo do rio. Assim, os mapas antigos, embora menos precisos, permitem avaliar as vazões históricas de acordo com as condições naturais da bacia.

No local onde está instalada a estação fluviométrica de Fair Oaks há, ainda, o registro de uma de uma cheia datada de 1862 cuja vazão de pico, avaliada por USBR (2002), é de 7.504 m³/s. Informações históricas mostram que essa é a maior cheia observada no rio American, anterior à cheia de 1997, desde 1848.

Os resultados encontrados por USBR (2002) mostraram que, na bacia do rio American, a jusante de Folsom, há evidências de pelo menos três paleocheias entre 11.327 e 15.574 m³/s ocorridas entre 150 e 700 anos atrás (ano de referência 2000) e pelo menos uma paleocheia entre 16.990 e 24.069 m³/s ocorrida entre 700 e 2000 anos atrás (ano de referência 2000). A figura 7.5 resume os resultados encontrados por USBR (2002). A tabela 7.1 mostra cada uma das paleocheias consideradas.

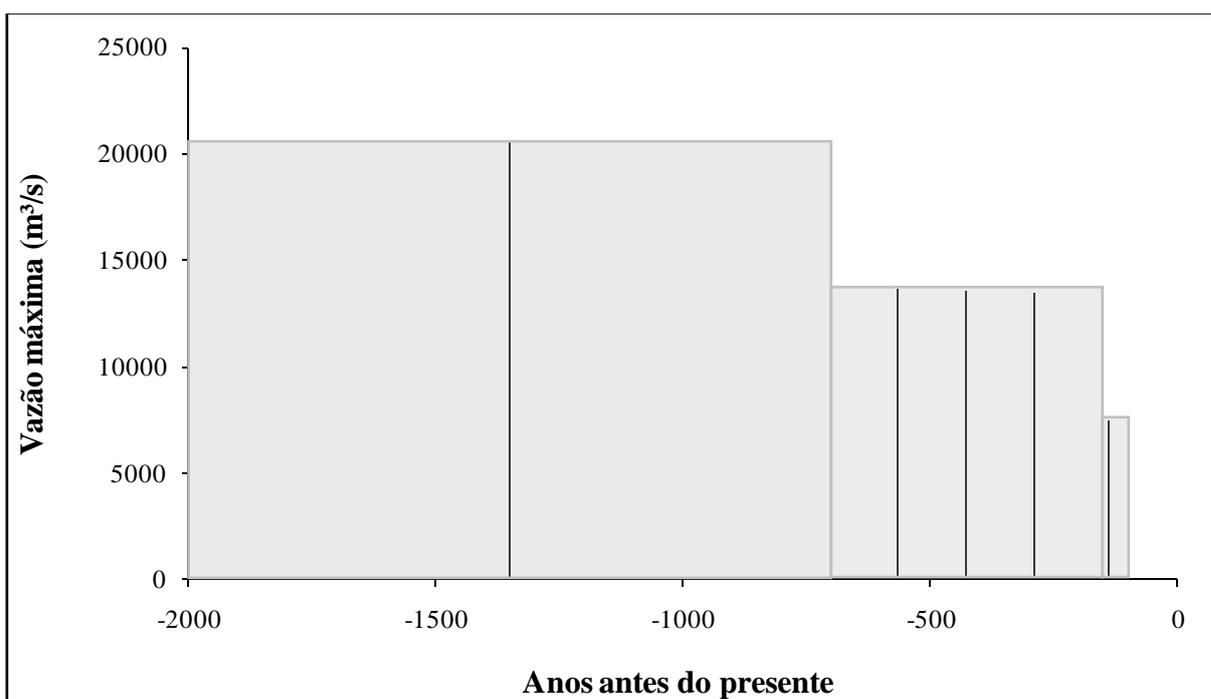


Figura 7.5 – Dados não sistemáticos do rio American em Fair Oaks (Ano de referência: 2000)

Tabela 7.1 – Paleocheias individuais em Fair Oaks

Mínimo (m ³ /s)	Máximo (m ³ /s)	Média (m ³ /s)	Idade* (anos AP)
7.419	8.495	7.504	136
11.327	15.574	13.451	289
11.327	15.574	13.535	426
11.327	15.574	13.592	563
16.990	24.069	20.530	1350

* a idade se refere ao número de anos antes do presente, sendo o ano de 2000 a referência

7.2.3 Distribuição *a priori* para o limite superior em Fair Oaks

A estimativa da PMF para o rio American, em Fair Oaks, dada pelo USACE (2001) é de 25.655 m³/s. Seguindo o procedimento A descrito no item 6.5.1.1, com $p = 0,5$ e $CV = 0,3$, $0,5$ e $0,7$, têm-se, respectivamente, as seguintes distribuições *a priori* Gama para o limite superior:

$$I) \quad \alpha \sim \text{GAM}(11,11, 4,17 \times 10^{-4})$$

$$II) \quad \alpha \sim \text{GAM}(4,00, 1,43 \times 10^{-4})$$

$$III) \quad \alpha \sim \text{GAM}(2,04, 6,64 \times 10^{-5})$$

Seguindo, agora, o procedimento B descrito no item 6.5.1.2, as 561 estimativas de PMF compiladas por USNRC (1977) foram transpostas para a bacia do rio American, em Folsom. A transposição dessas tormentas foi feita a partir da equação 6.63, com $A_0 = 4.820 \text{ km}^2$.

A figura 7.6 mostra o histograma das PMF's transpostas para Folsom. A forma desse histograma é semelhante à forma dos histogramas mostrados na figura 6.5 para grupos de PMF com áreas de drenagem semelhantes. Ou seja, as PMF's são distribuídas assimetricamente para a direita, o que permite definir a distribuição Gama como modelo para os valores transpostos. Os parâmetros da distribuição Gama foram estimados pelo método dos momentos convencionais. A distribuição *a priori* para o limite superior, de acordo com o procedimento B, é a seguinte:

$$IV) \quad \alpha \sim \text{GAM}(5,20, 4,33 \times 10^{-4})$$

A figura 7.6 mostra a distribuição Gama ajustada às estimativas transpostas de PMF.

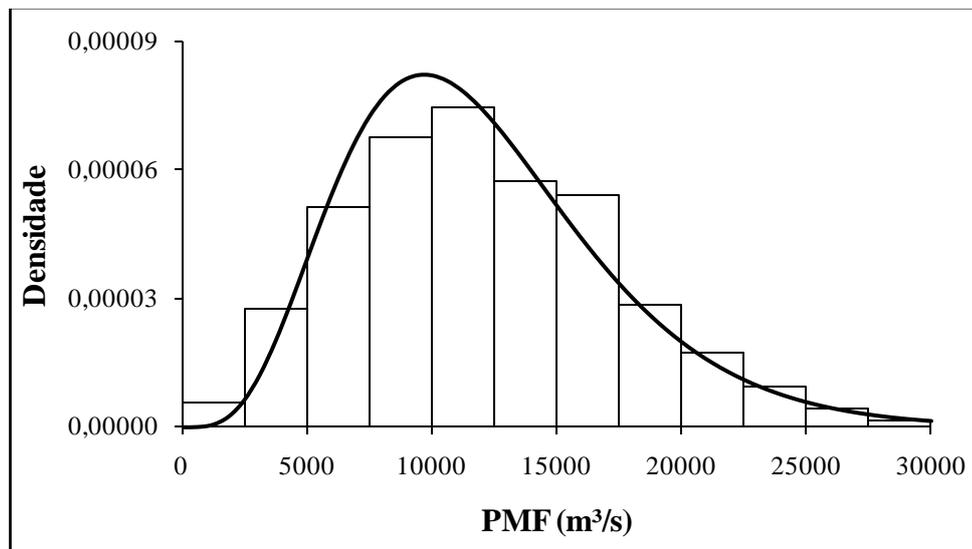


Figura 7.6 – Histograma das estimativas de PMF transpostas para a bacia do rio American e ajuste da distribuição Gama

Adicionalmente às distribuições *a priori* definidas a partir dos procedimentos A e B, foi considerada, também, uma distribuição não informativa para o limite superior. Esta é incluída com o objetivo de fornecer uma análise quando o decisor não tem informação *a priori* ou não compartilha da opinião de que a PMF tem uma associação direto com o limite superior. A distribuição não informativa adotada neste trabalho é uma Gama com grande variância. Ou seja, a distribuição *a priori* não informativa, ou mais rigorosamente pouco informativa, é uma Gama relativamente plana com os seguintes parâmetros:

$$V) \quad \alpha \sim \text{GAM}(1,00, 1,00 \times 10^{-8})$$

A tabela 7.2 resume as principais características das distribuições usadas como especificação *a priori* para o limite superior. A figura 7.7 mostra suas respectivas formas.

Tabela 7.2 – Parâmetros e características das distribuições *a priori* para o limite superior na bacia do rio American, em Folsom

Distribuição <i>a priori</i>	ρ_α	β_α	Média	Mediana	CV_α	DP	Proc.
I	11,111	$4,17 \times 10^{-4}$	26.633	25.655	0,3	7.990	A
II	4,0000	$1,43 \times 10^{-4}$	27.952	25.655	0,5	13.976	A
III	2,0408	$6,64 \times 10^{-5}$	30.717	25.655	0,7	21.502	A
IV	5,1992	$4,33 \times 10^{-4}$	12.018	11.257	0,4	5.271	B
V	1,0000	$1,00 \times 10^{-8}$	$1,00 \times 10^8$	$\sim 6,9 \times 10^7$	1,0	$1,00 \times 10^8$	-

CV: coeficiente de variação; DP: desvio padrão; Proc.: procedimento de cálculo

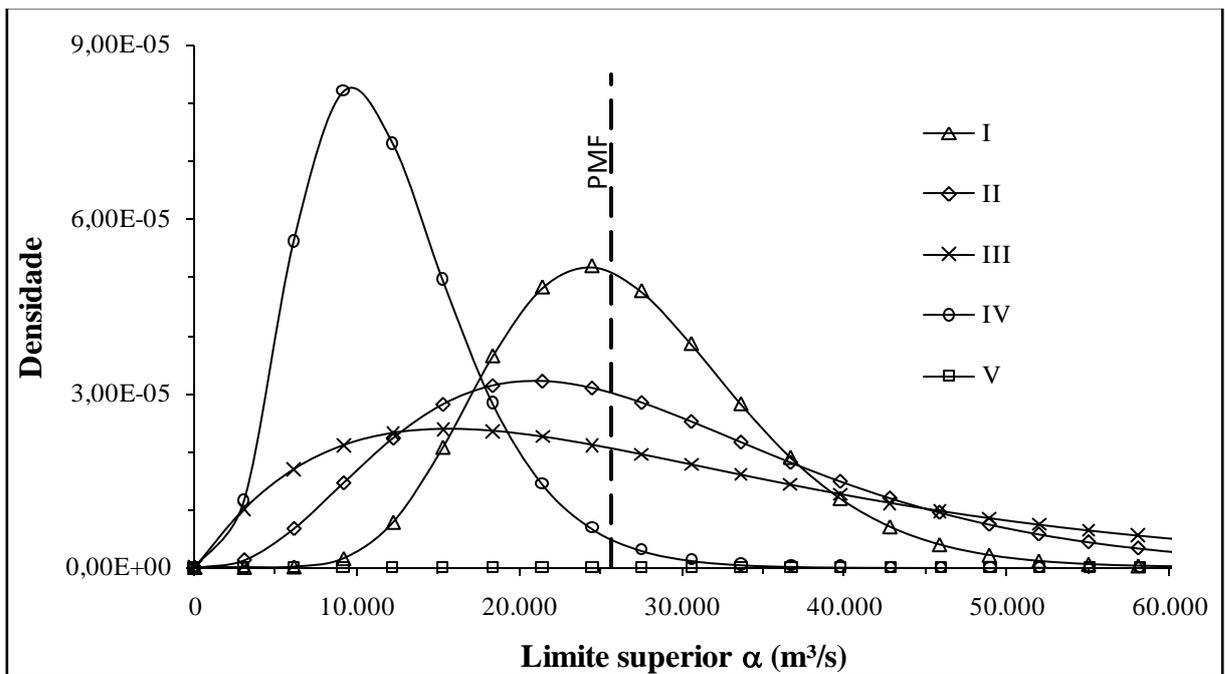


Figura 7.7 – Distribuições *a priori* de I a V para o limite superior na bacia do rio American, em Folsom

7.2.4 Estatísticas *a posteriori*

7.2.4.1 Modelo LN4

Conforme visto no capítulo 5, a distribuição *a posteriori* dos parâmetros é proporcional ao produto da função de verossimilhança pela distribuição *a priori*, ou seja:

$$p(\Theta/x) \propto p(x/\Theta)p(\alpha)p(\mu)p(\sigma), \quad (7.1)$$

onde $p(x/\Theta)$ é a função de verossimilhança dada em (6.56), $p(\alpha)$ é a distribuição *a priori* para o limite superior dada pelas distribuições de I a V do item 7.2.3, $p(\mu)$ e $p(\sigma)$ são, respectivamente, a distribuição *a priori* para os parâmetros μ e σ .

A interpretação dos parâmetros μ e σ – respectivamente, média e desvio padrão da variável transformada LN4 (vide item 6.2.4) – em termos hidrológicos ou hidrometeorológicos é bastante complicada. Na verdade, ao se avaliar a distribuição LN4 é difícil encontrar alguma relação evidente entre esses parâmetros e as características físicas das bacias de interesse. Assim, construir uma distribuição *a priori* informativa para esses parâmetros é uma tarefa, no mínimo, complexa. No entanto, uma vez que o parâmetro μ pode assumir qualquer valor real e o parâmetro σ é sempre positivo, admitiu-se uma distribuição Normal, não informativa, para o primeiro e uma distribuição Gama, não informativa, para o segundo, ou seja:

$$\mu \sim N(1,00, 1,00 \times 10^{-6}) \text{ e}$$

$$\sigma \sim \text{Gama}(1,00, 1,00 \times 10^{-8}).$$

A solução da equação 7.1 envolve o cálculo de integrais multidimensionais complexas, por vezes impossíveis de serem obtidas analiticamente. Esse é o caso do modelo proposto nesta tese. A alternativa à integração analítica são os algoritmos capazes de produzir amostras da distribuição *a posteriori*, tais como os métodos MCMC, discutido no capítulo 5. Para o modelo LN4 foi adotado o algoritmo Metropolis-Hastings, cuja implementação de domínio público é dada no software WinBUGS (Lunn et al., 2000).

Foram construídos 5 modelos a partir das distribuições *a priori* para o limite superior α utilizando todos os dados disponíveis em Folsom (sistemáticos e não sistemáticos) e 5 modelos utilizando somente os dados sistemáticos. A construção dos modelos com somente dados sistemáticos tem por objetivo avaliar a contribuição da informação não sistemática na análise.

Para cada modelo foi gerada, no WinBUGS, uma amostra de tamanho 600.000, sendo que, a cada 10 realizações do algoritmo, uma era selecionada para análise ($lag = 10$). Ou seja, selecionou-se para análise uma amostra de tamanho 60.000. A escolha do tamanho do lag foi feita por tentativa e erro, avaliando a autocorrelação entre as realizações do algoritmo. Verificou-se que um lag igual a 10 é suficiente para produzir amostras não correlacionadas.

Da amostra de 60.000 valores, os 10.000 primeiros foram descartados como *burn in*. O termo *burn in* é o número de realizações necessárias para que a cadeia de Markov “esqueça” seu estado inicial. Gilks et al. (1996) avaliam que valores entre 1 e 2% do total de realizações é suficiente para que a cadeia de Markov atenda a esse pré-requisito. Desta forma, a amostra final é composta de 50.000 realizações da distribuição conjunta *a posteriori* dos parâmetros da LN4.

Estatísticas a posteriori para o parâmetro μ

A tabela 7.3 mostra os resultados para os parâmetros μ . A figura 7.8 mostra as distribuições *a posteriori* do parâmetro μ para os modelos completos e com dados sistemáticos.

Tabela 7.3 – Estatísticas *a posteriori* para o parâmetro μ da LN4

Conjunto de dados	Modelo	Média	CV	DP	95% HPD
Sistemáticos e não sistemáticos	I	-3,07	-0,073	0,225	(-3,51, -2,65)
	II	-3,22	-0,095	0,305	(-3,82, -2,67)
	III	-3,35	-0,115	0,387	(-4,12, -2,68)
	IV	-2,92	-0,061	0,178	(-3,28, -2,59)
	V	-10,42	-0,156	1,629	(-12,97, -7,35)
Sistemáticos	I*	-2,68	-0,140	0,375	(-3,37, -1,92)
	II*	-2,67	-0,199	0,530	(-3,64, -1,60)
	III*	-2,75	-0,236	0,649	(-3,92, -1,49)
	IV*	-2,01	-0,184	0,371	(-2,74, -1,31)
	V*	-10,52	-0,121	1,271	(-12,68, -7,95)

95% HPD - Intervalo de Credibilidade; os modelos marcados com '*' são aqueles formulados somente com informação sistemática

Uma vez que não foi possível estabelecer uma distribuição *a priori* para o parâmetro μ , torna-se difícil analisar os resultados *a posteriori*. A figura 7.8 mostra que esse parâmetro é sensível à especificação da distribuição *a priori* para o limite superior. Com efeito, com o aumento do coeficiente de variação da distribuição *a priori* para o limite superior, há um aumento no coeficiente de variação da distribuição *a posteriori* de μ . Além disso, há um deslocamento da distribuição para esquerda e um aumento do intervalo de credibilidade. Isso mostra, mesmo que indiretamente, que um maior conhecimento acerca do limite superior (menor coeficiente de variação) melhora as estimativas do parâmetro μ .

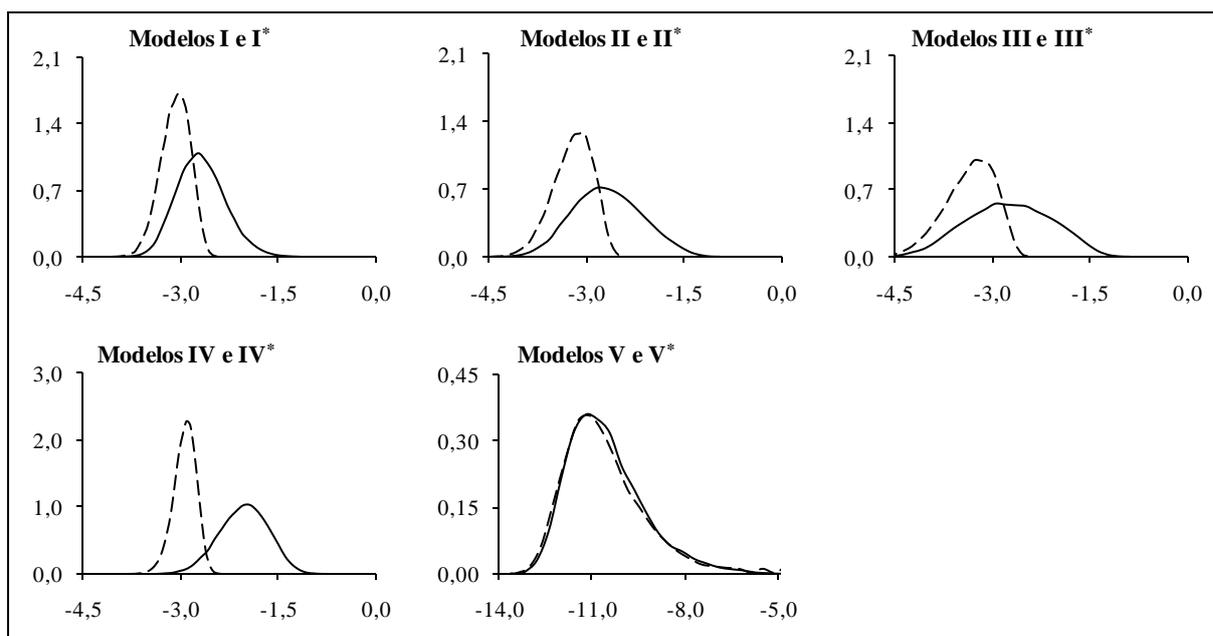


Figura 7.8 – Distribuições *a posteriori* do parâmetro μ para os modelos completos (linha tracejada) e para os modelos com somente dados sistemáticos (linha contínua)

Conclusões semelhantes podem ser obtidas para os modelos que consideram apenas dados sistemáticos. No entanto, nesse caso, há um aumento significativo do coeficiente de variação e do intervalo de credibilidade em relação aos modelos completos, mostrando que o uso de informação não sistemática é uma alternativa válida para melhorar as estimativas desse parâmetro, levando a uma diminuição da incerteza sobre ele.

Para os modelos onde foi especificada uma distribuição não informativa para o limite superior (modelos V e V^{*}), não houve diferenças entre as distribuições *a posteriori* para o parâmetro μ , o que mostra que, neste caso, não há ganho em se utilizar dados não sistemáticos. No entanto, a forma bastante discrepante dessa distribuição em relação aos demais modelos e o grande intervalo de credibilidade leva a concluir em favor dos modelos com distribuição informativa para o limite superior.

Estatísticas a posteriori para o parâmetro σ

A tabela 7.4 mostra os resultados para os parâmetros σ . A figura 7.9 mostra as distribuições *a posteriori* do parâmetro σ para os modelos completos e com dados sistemáticos.

Tabela 7.4 – Estatísticas *a posteriori* para o parâmetro σ da LN4

Conjunto de dados	Modelo	Média	CV	DP	95% HPD
Sistemáticos e não sistemáticos	I	0,99	0,0760	0,0753	(0,85, 1,14)
	II	0,97	0,0789	0,0765	(0,82, 1,12)
	III	0,95	0,0820	0,0783	(0,81, 1,11)
	IV	1,02	0,0739	0,0753	(0,88, 1,17)
	V	0,84	0,0619	0,0520	(0,74, 0,94)
Sistemáticos	I [*]	1,03	0,1105	0,1137	(0,82, 1,26)
	II [*]	1,04	0,1231	0,1281	(0,81, 1,30)
	III [*]	1,04	0,1301	0,1353	(0,80, 1,31)
	IV [*]	1,15	0,1279	0,1465	(0,88, 1,44)
	V [*]	0,93	0,1024	0,0951	(0,76, 1,13)

95% HPD - Intervalo de Credibilidade; os modelos marcados com '*' são aqueles formulados somente com informação sistemática

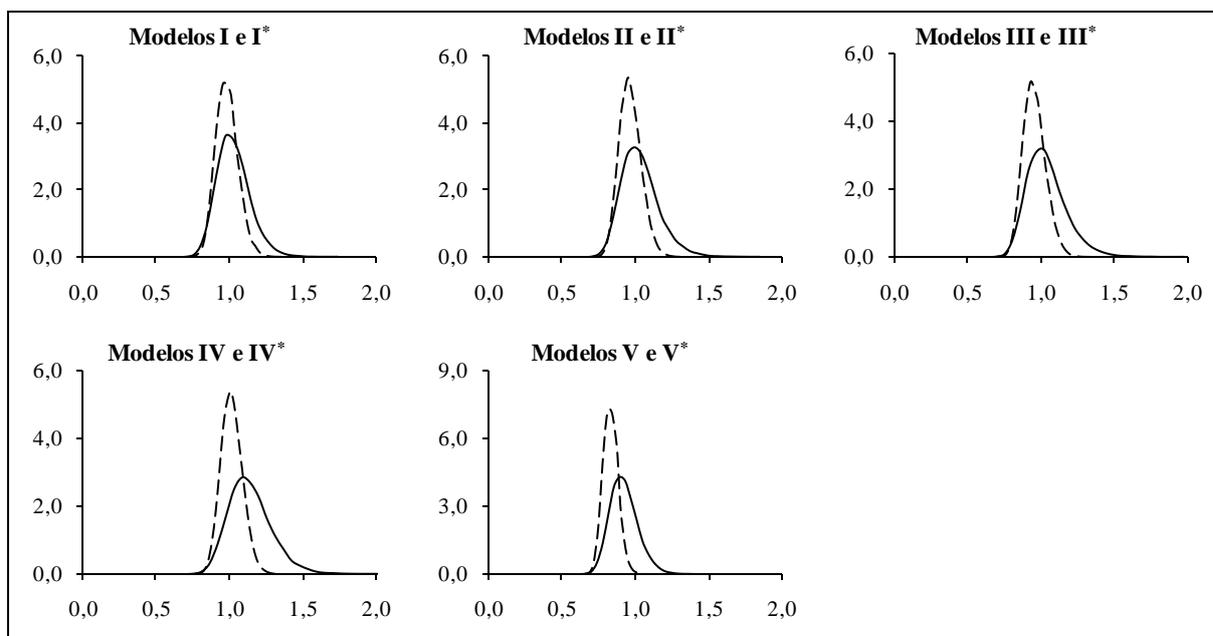


Figura 7.9 – Distribuições *a posteriori* do parâmetro σ para os modelos completos (linha tracejada) e para os modelos com somente dados sistemáticos (linha contínua)

Assim como para o parâmetro μ , o fato de não ter sido possível estabelecer uma distribuição *a priori* para o parâmetro σ dificulta a análise dos resultados *a posteriori*. Para esse parâmetro as distribuições *a posteriori* apresentaram um comportamento bastante semelhante (figura 7.9). Isso mostra que esse parâmetro é pouco sensível à especificação da distribuição *a priori* para o limite superior. No entanto, apesar de pouco pronunciado, há um aumento do coeficiente de variação com o aumento do coeficiente de variação da distribuição *a priori* do limite superior.

Os modelos V e V* apresentaram um comportamento não esperado no que se refere ao parâmetro σ . Esses modelos apresentaram um coeficiente de variação bastante inferior aos demais modelos, implicando, a princípio, em um ganho na estimativa de σ . No entanto, a melhor estimativa de σ nos modelos V e V* implica em uma pior estimativa de μ e, como será visto a diante, de α .

Estatísticas a posteriori para o parâmetro α

A tabela 7.5 e figura 7.10 mostram os resultados para o limite superior α .

Tabela 7.5 – Estatísticas *a posteriori* para o limite superior α da LN4

Conjunto de dados	Modelo	Média	CV	DP	95% HPD
Sistemáticos e não sistemáticos	I	32.538	0,19	6.057	(24.070, 44.140)
	II	38.109	0,29	11.064	(24.080, 59.810)
	III	44.735	0,40	17.909	(24.070, 80.290)
	IV	28.104	0,12	3.499	(24.070, 34.980)
	V	$9,45 \times 10^7$	1,03	$9,73 \times 10^7$	$(5,91 \times 10^4, 2,88 \times 10^8)$
Sistemáticos	I*	25.947	0,30	7.872	(11.710, 41.370)
	II*	27.296	0,47	12.712	(8.785, 51.770)
	III*	31.478	0,62	19.439	(8.664, 70.190)
	IV*	14.628	0,29	4.265	(8.609, 22.960)
	V*	$9,95 \times 10^7$	0,99	$9,82 \times 10^7$	$(5,65 \times 10^4, 2,97 \times 10^8)$

95% HPD - Intervalo de Credibilidade; os modelos marcados com '*' são aqueles formulados somente com informação sistemática

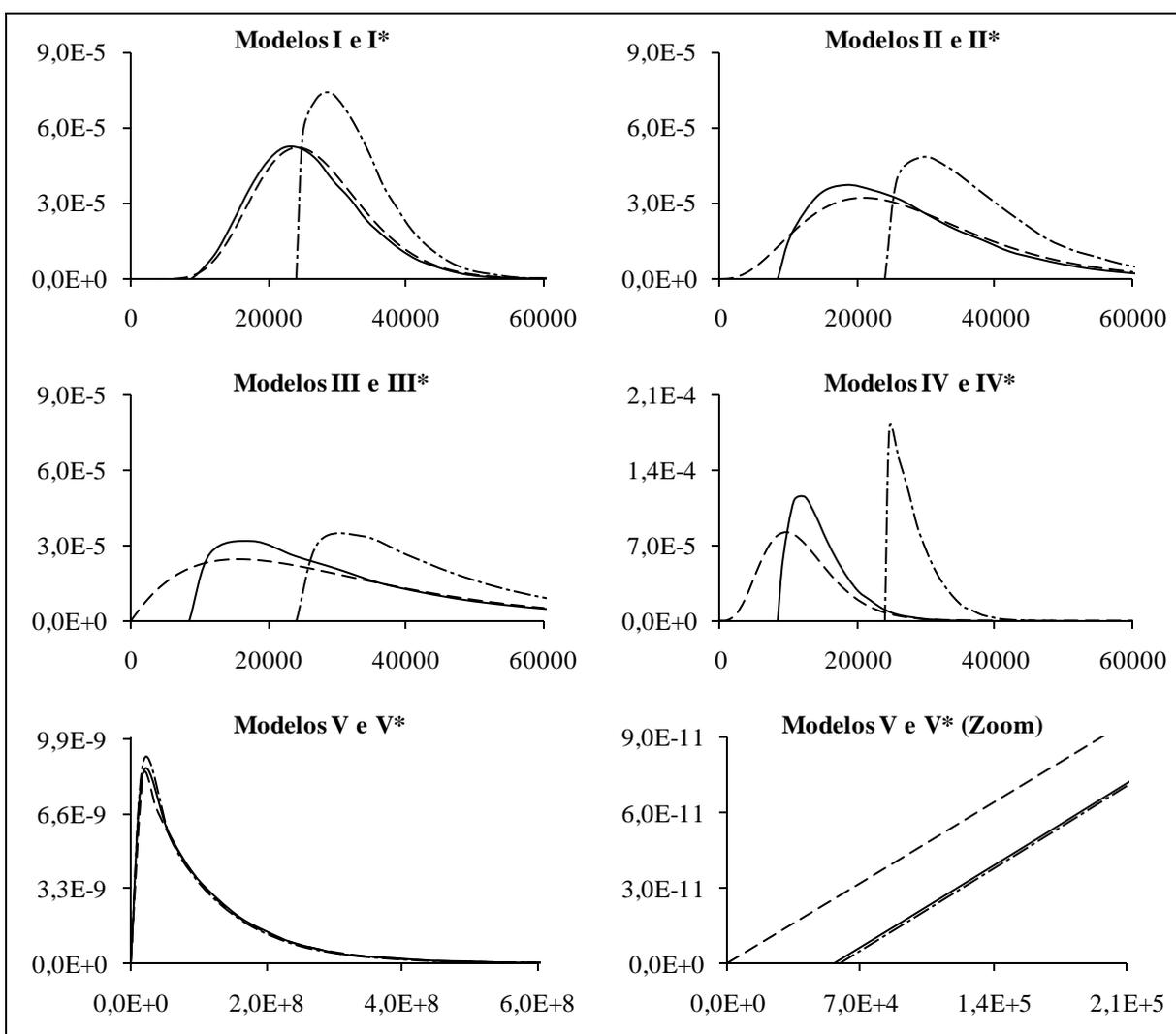


Figura 7.10 – Distribuição *a priori* (linha tracejada), distribuição *a posteriori* para os modelos com todos os dados (linha pontilhada-tracejada) e distribuição *a posteriori* para os modelos com dados sistemáticos (linha contínua)

A tabela 7.5 mostra uma sensível redução no coeficiente de variação quando os dados não sistemáticos são incluídos na análise. Isso pode ser verificado, em particular, nos modelos I e I*. Como os dados sistemáticos são muito menores que a estimativa da PMF, não é esperado que esses contribuam com novas informações sobre o limite superior. De fato, a figura 7.10 mostra que a distribuição *a posteriori* para o parâmetro α no modelo I* é bastante semelhante à distribuição *a priori*. A tabela 7.5 mostra que não há uma redução do coeficiente de variação. Isso sugere que o conhecimento sobre o limite superior não aumenta se somente os dados sistemáticos são incluídos na análise.

Por outro lado, os dados não sistemáticos exercem uma grande influência da estimativa *a posteriori* do limite superior. A tabela 7.5 mostra que no modelo I há uma redução de mais de 35% no coeficiente de variação, indicando um ganho significativo na estimação desse parâmetro. Conclusões semelhantes podem ser feitas para os modelos II, III e IV, embora nesses casos haja uma redução maior no CV para os modelos com somente dados sistemáticos. Ou seja, a distribuição *a posteriori* não é tão semelhante à distribuição *a priori* como no modelo I*.

Outra maneira de avaliar a quantidade de novas informações incluídas na análise pelos dados não sistemáticos é o intervalo de credibilidade. Com efeito, os intervalos de credibilidade para os modelos completos são muito menores que aqueles dos modelos com somente dados sistemáticos. O modelo I, por exemplo, tem um intervalo de credibilidade de cerca de 30% mais curto que aquele dado pelo modelo I*. Ou seja, há menos incerteza sobre α quando os modelos completos são utilizados.

A análise dos modelos V e V* mostra que, independentemente do tipo de dado hidrológico incluído na análise, o uso de uma distribuição não informativa para o limite superior é uma prática não recomendada. De fato, a distribuição não informativa gera uma distribuição *a posteriori* que dá altas probabilidades de ocorrência para valores muito pouco plausíveis do limite superior. A figura 7.10 mostra que as distribuições *a posteriori* para os modelos V e V* possuem uma significativa cauda para valores acima de 100.000.000 m³/s, o que é fisicamente impossível para a bacia em questão ou para qualquer outro curso d'água terrestre.

Outro fato a ser observado nos modelos V e V* é a semelhança entre as distribuições *a posteriori*. Independentemente do tipo de dado incluído na análise, a estimativa *a posteriori* para o limite superior é a mesma, sugerindo que os dados hidrológicos por si só não são capazes de capturar as características essenciais da variabilidade do limite superior. Em outras

palavras, quando não há nenhuma informação *a priori* sobre o limite superior, qualquer valor acima do maior pico de cheia observado na bacia é um candidato plausível a limite superior.

Curva de quantis

A curva de quantis ou a distribuição preditiva *a posteriori* foi estimada por integração de Monte Carlo (vide item 5.3) a partir dos valores da distribuição *a posteriori* dos parâmetros da LN4. A figura 7.11 mostra os resultados para cada um dos modelos estudados. No anexo 1 são mostrados os resultados numéricos para cada período de retorno e para cada modelo. Os dados observados, sistemáticos e não sistemáticos, foram plotados de acordo com a probabilidade de excedência calculada a partir do método descrito no item 6.4.

De uma forma geral, todos os modelos formulados com base em dados sistemáticos e não sistemáticos apresentaram um ajuste satisfatório para todas as faixas de período de retorno. Entre esses, o modelo III aparentemente é o que apresentou o melhor ajuste. A figura 7.12 mostra uma comparação entre os quantis estimados e os observados. Observa-se que os modelos de I a IV apresentam resultados sistematicamente melhores que os modelos de I* a IV*.

As curvas de quantis da figura 7.11 e os gráficos da figura 7.12 permitem avaliar a adequação das distribuições *a priori* para o limite superior utilizadas nesta tese. Com efeito, as curvas de quantis para os modelos I* a IV* permitem prever, razoavelmente bem, algumas cheias não sistemáticas, a despeito do fato de esses modelos terem sido formulados somente em termos de dados sistemáticos. O modelo III*, por exemplo, prediz com bastante precisão a cheia de 1.000 anos de período de retorno mesmo que cheias dessa magnitude não tenham sido incluídas na análise.

A capacidade preditiva desses modelos é contrastada com os resultados do modelo V*, para o qual nenhuma informação *a priori* sobre o limite superior foi fornecida, além de nenhuma informação não sistemática. De fato, mesmo para vazões com período de retorno inferiores a 100 anos, o ajuste do modelo V* não é satisfatório. Isso demonstra que especificar uma distribuição informativa para o limite superior é vantajoso não somente para prever vazões extremas, mas também para vazões de menor magnitude.

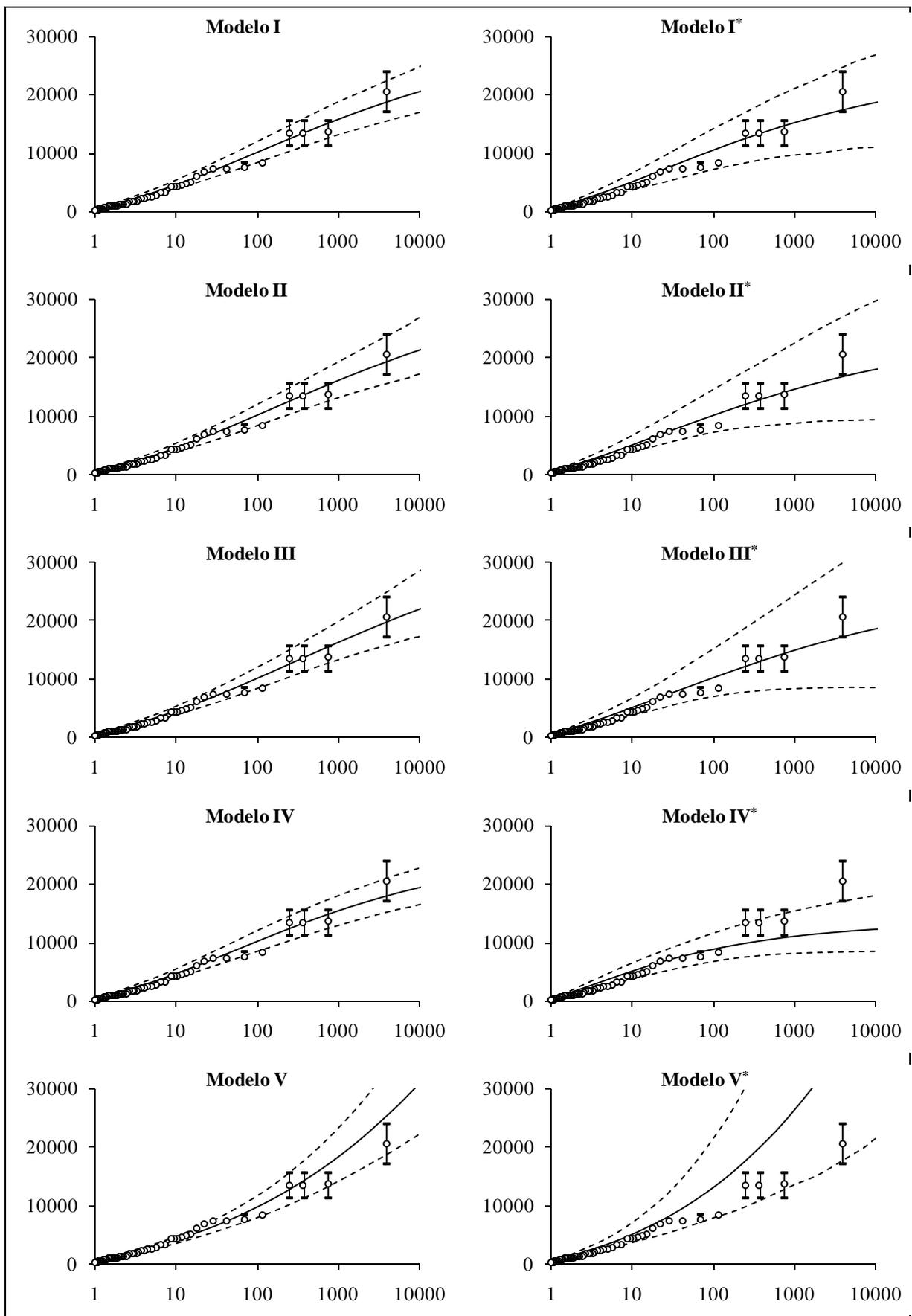


Figura 7.11 – Distribuição preditiva *a posteriori* (média) ou curva de quantis (linha contínua), intervalo de credibilidade de 95% (linha tracejada). Os círculos representam os dados sistemáticos e as barras com círculos representam os dados não sistemáticos. No eixo horizontal está o período de retorno em anos e no eixo vertical está a vazão em m³/s

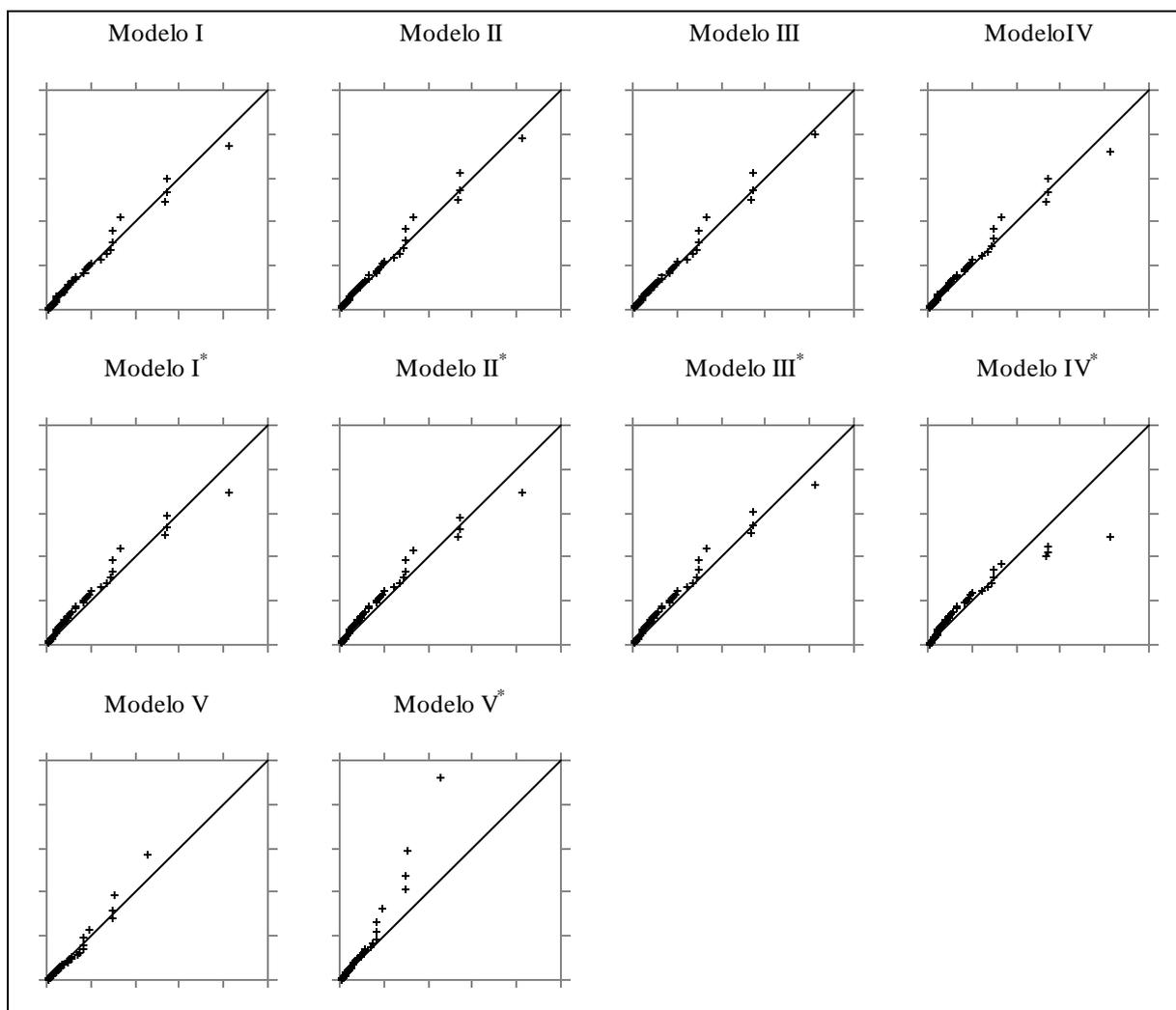


Figura 7.12 – Comparação entre os quantis estimados (eixo vertical) e observados (eixo horizontal) para o modelo LN4

A média *a posteriori* para o limite superior da LN4, nos modelos de I a IV, varia entre 28.104 m³/s e 44.735 m³/s, superior à estimativa da PMF de 25.655 m³/s. No entanto, uma vez que os limites paleohidrológicos são 16.990 e 24.069 m³/s, é esperado que o limite superior da LN4 seja maior que a estimativa da PMF. As diferenças entre as estimativas refletem as diferentes formas da distribuição *a priori* para esse parâmetro e como cada distribuição agrega o conhecimento acerca do limite superior. Do ponto de vista da curva de quantis somente, é difícil optar por uma dessas distribuições *a priori*, embora o modelo III tenha apresentado uma capacidade descritiva e preditiva um pouco superior que os demais modelos.

7.2.4.2 Modelo EV4

Para o modelo EV4, a distribuição *a posteriori* dos parâmetros torna-se:

$$p(\Theta/x) \propto p(x/\Theta)p(\alpha)p(\sigma)p(\xi), \quad (7.2)$$

onde, $p(x/\Theta)$ é a função de verossimilhança dada em (6.56), $p(\alpha)$ é a distribuição *a priori* para o limite superior dada pelas distribuições de I a V do item 7.2.3, $p(\sigma)$ e $p(\xi)$ são, respectivamente, a distribuição *a priori* para os parâmetros σ e ξ .

Da mesma forma que para o modelo LN4, a interpretação dos parâmetros σ e ξ em termos hidrológicos ou hidrometeorológicos é bastante complexa, inviabilizando a obtenção de alguma relação evidente entre esses parâmetros e as características físicas das bacias de interesse. Assim, uma vez que esses parâmetros são sempre positivos, admitiu-se uma distribuição Gama, não informativa, como especificação *a priori* para ambos os parâmetros.

Assim como no caso da LN4, foram construídos 5 modelos a partir das distribuições *a priori* para α . Para cada modelo foi gerada, no WinBUGS, uma amostra de tamanho 1.200.000 sendo que, a cada 20 realizações do algoritmo, uma era selecionada para análise ($lag = 20$). Ou seja, selecionou-se para análise uma amostra de tamanho 60.000. A escolha do tamanho do lag foi feita por tentativa e erro, avaliando a autocorrelação entre as realizações do algoritmo. Verificou-se que um lag igual a 20 é suficiente para produzir amostras não correlacionadas. As 10.000 primeiras foram descartadas como *burn in*, restando uma amostra final composta de 50.000 realizações da distribuição conjunta *a posteriori* dos parâmetros da EV4.

Estatísticas a posteriori para o parâmetro ξ

A tabela 7.6 mostra os resultados para os parâmetros ξ . A figura 7.13 mostra as distribuições *a posteriori* desse parâmetro para os modelos completos e com dados sistemáticos.

Esses resultados – mostrados na tabela 7.6 e na figura 7.13 – são semelhantes àqueles obtidos para os parâmetros de escala e posição da LN4. Com efeito, verifica-se que o parâmetro ξ é pouco sensível à especificação da distribuição *a priori* para o limite superior. Além disso, a distribuição *a posteriori* de ξ é praticamente a mesma em todos os casos.

Para os modelos V e V*, contrariando o resultado esperado, o coeficiente de variação para a distribuição *a posteriori* de ξ apresentou seu menor valor. Associado a esse fato, verifica-se que o intervalo de credibilidade para ξ sob esses modelos possui o mesmo comprimento que os demais. Assim, analisando somente as estimativas para ξ , os resultados mostram que não há vantagens em se utilizar uma distribuição *a priori* informativa para o limite superior. Por outro lado, como será visto na seqüência, os resultados para os demais parâmetros contrariam

esta conclusão, sobretudo para o limite superior, onde os modelos V e V* não fornecem boas estimativas.

Tabela 7.6 – Estatísticas *a posteriori* para o parâmetro ξ da EV4

Conjunto de dados	Modelo	Média	CV	DP	95% HPD
Sistemáticos e não sistemáticos	I	1,46	0,061	0,090	(1,29, 1,64)
	II	1,46	0,062	0,091	(1,29, 1,65)
	III	1,46	0,063	0,093	(1,29, 1,65)
	IV	1,45	0,060	0,086	(1,29, 1,62)
	V	1,73	0,051	0,089	(1,57, 1,91)
Sistemáticos	I*	1,14	0,113	0,128	(0,89, 1,39)
	II*	1,10	0,128	0,141	(0,84, 1,38)
	III*	1,09	0,133	0,145	(0,81, 1,37)
	IV*	1,01	0,124	0,126	(0,78, 1,27)
	V*	1,23	0,105	0,128	(0,98, 1,48)

95% HPD - Intervalo de Credibilidade; os modelos marcados com '*' são aqueles formulados somente com informação sistemática

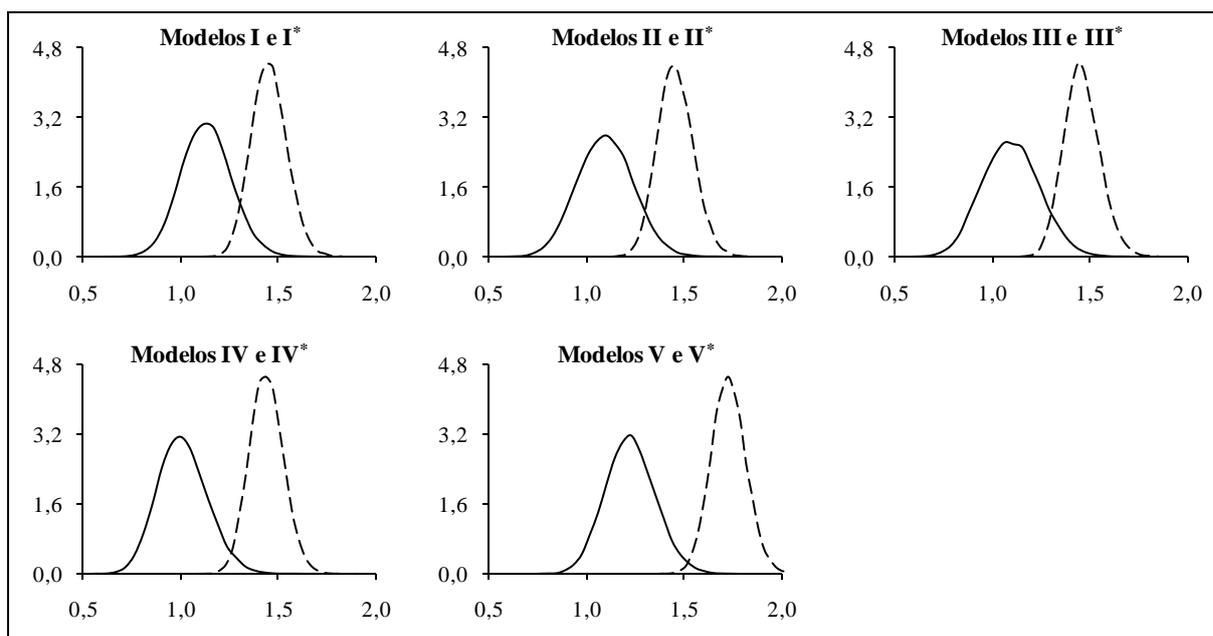


Figura 7.13 – Distribuições *a posteriori* do parâmetro ξ para os modelos completos (linha tracejada) e para os modelos com somente dados sistemáticos (linha contínua)

Estatísticas a posteriori para o parâmetro σ

A tabela 7.7 mostra os resultados para os parâmetros σ . A figura 7.14 mostra as distribuições *a posteriori* do parâmetro σ para os modelos completos e com dados sistemáticos.

Tabela 7.7 – Estatísticas *a posteriori* para o parâmetro σ da EV4

Conjunto de dados	Modelo	Média	CV	DP	95% HPD
Sistemáticos e não sistemáticos	I	31,11	0,1554	4,8351	(23,27, 40,83)
	II	31,69	0,1863	5,9025	(22,98, 42,98)
	III	32,09	0,2186	7,0143	(22,92, 44,58)
	IV	29,71	0,1201	3,5692	(23,43, 36,93)
	V	$2,60 \times 10^5$	0,7136	$1,85 \times 10^5$	$(2,5 \times 10^3, 6,20 \times 10^5)$
Sistemáticos	I*	24,55	0,3936	9,6625	(7,06, 42,82)
	II*	22,81	0,6794	15,4965	(5,08, 53,54)
	III*	25,85	0,9007	23,2807	(4,87, 74,46)
	IV*	10,19	0,3789	3,8612	(4,99, 18,03)
	V*	$2,08 \times 10^5$	0,7229	$1,50 \times 10^5$	$(4,7 \times 10^3, 5,0 \times 10^5)$

95% HPD - Intervalo de Credibilidade; os modelos marcados com '*' são aqueles formulados somente com informação sistemática.

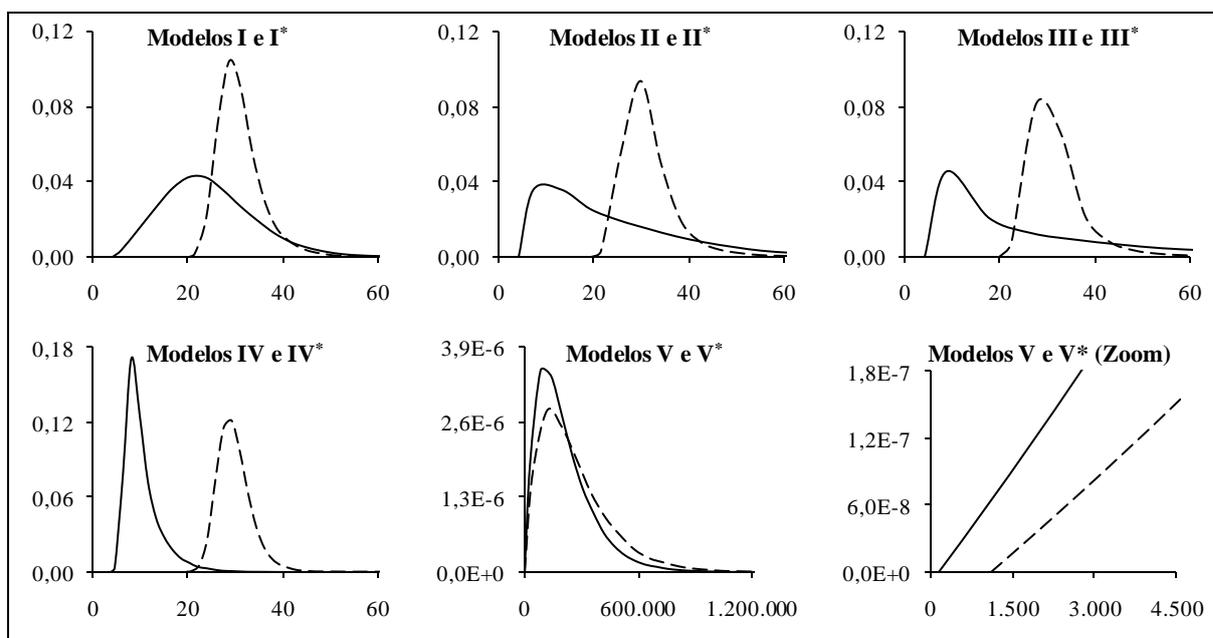


Figura 7.14 – Distribuições *a posteriori* do parâmetro σ para os modelos completos (linha tracejada) e para os modelos com somente dados sistemáticos (linha contínua)

Esses resultados – tabela 7.7 e figura 7.14 – mostram que nos modelos completos, com exceção do modelo V, as distribuições *a posteriori* de σ são bastante similares, indicando uma relativa independência desse parâmetro com a especificação *a priori* para α . Os modelos com somente dados sistemáticos, no entanto, apresentaram sensíveis diferenças na distribuição *a posteriori* de σ , destacando-se o modelo IV*. Para esses modelos, a especificação *a priori* de α influencia significativamente a distribuição *a posteriori* de σ , levando à conclusão que há uma dependência entre esses parâmetros.

Assim, verifica-se que o parâmetro σ está ligado às características da cauda superior da EV4. Quando os dados não sistemáticos são incluídos na análise, esses exercem maior influência na cauda superior da distribuição, condicionando as estimativas de σ . Quando o modelo contempla somente os dados sistemáticos, a distribuição *a priori* de α domina a cauda superior da EV4, fazendo com que σ seja condicionado pela especificação *a priori* de α .

No que se refere aos modelos V e V*, os resultados mostram que o uso de uma distribuição não informativa para α implica em distribuições *a posteriori* para σ com características bastante diferentes daquelas dos demais modelos. Além disso, o intervalo de credibilidade bastante largo para esse parâmetro é um indicador desfavorável ao uso de distribuições não informativas para α .

Estadísticas a posteriori para o parâmetro α

A tabela 7.8 e figura 7.15 mostram os resultados para o limite superior α .

Tabela 7.8 – Estadísticas *a posteriori* para o limite superior α da EV4

Conjunto de dados	Modelo	Média	CV	DP	95% HPD
Sistemáticos e não sistemáticos	I	26.677	0,11	2.885	(24.070, 32.370)
	II	27.098	0,14	3.814	(24.070, 34.170)
	III	27.427	0,18	4.819	(24.070, 35.620)
	IV	25.608	0,06	1.633	(24.070, 28.790)
	V	2,0x10 ⁸	0,71	1,4x10 ⁸	(2,0x10 ⁶ , 4,8x10 ⁸)
Sistemáticos	I*	24.846	0,35	8.586	(8.456, 40.100)
	II*	23.242	0,62	14.352	(8.438, 51.450)
	III*	26.151	0,83	21.812	(8.438, 71.340)
	IV*	11.288	0,29	3.326	(8.438, 18.120)
	V*	2,0x10 ⁸	0,71	1,4x10 ⁸	(4,3x10 ⁶ , 4,7x10 ⁸)

95% HPD - Intervalo de Credibilidade; os modelos marcados com '*' são aqueles formulados somente com informação sistemática

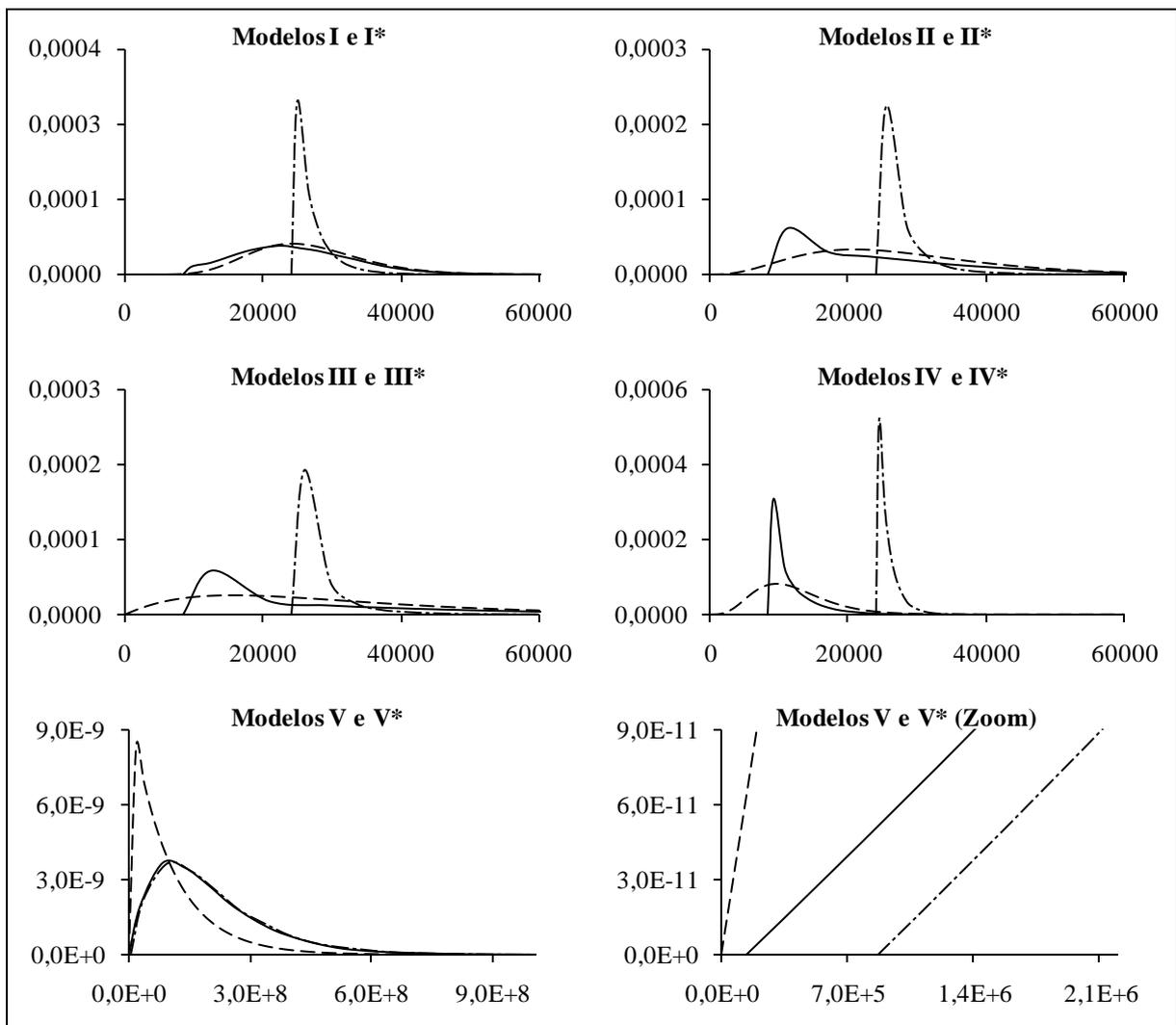


Figura 7.15 – Distribuição *a priori* (linha tracejada), distribuição *a posteriori* para os modelos com todos os dados (linha pontilhada-tracejada) e distribuição *a posteriori* para os modelos com dados sistemáticos (linha contínua)

As conclusões sobre a distribuição para o limite superior para o modelo EV4 são as mesmas obtidas para o modelo LN4, a saber:

- Redução significativa do coeficiente de variação para os modelos completos (>60% para a EV4);
- Aumento do intervalo de credibilidade com o aumento do coeficiente de variação *a priori*;
- Estimativas para o limite superior incompatíveis com as características físicas da bacia quando a especificação *a priori* de α é não informativa; e
- Aumento do intervalo de credibilidade quando somente os dados sistemáticos são usados na análise.

Comparando os resultados entre os modelos LN4 e EV4 verifica-se que o intervalo de credibilidade para o último é cerca de um terço do intervalo do primeiro, quando se utiliza toda a base de dados disponível. Quando somente os dados sistemáticos são utilizados, o comprimento do intervalo de credibilidade é o mesmo para ambos os modelos. Esse fato reforça a idéia de que, pelo menos para esta aplicação, os dados sistemáticos não fornecem informações sobre o limite superior, sendo o modelo dominado pela distribuição *a priori* de α .

Outro fato decorrente da comparação entre os modelos EV4 e LN4 é que, a despeito de a estimativa pontual (média) do limite superior para o modelo LN4 ser sistematicamente superior à estimativa para o modelo EV4, o modelo LN4 aproxima-se mais suavemente do limite superior que o modelo EV4. Com isso, como será visto a seguir, há uma distorção da curva de quantis da EV4 para períodos de retorno pequenos.

Curva de quantis

A curva de quantis ou a distribuição preditiva *a posteriori* foi estimada por integração de Monte Carlo (vide item 5.3) a partir dos valores da distribuição *a posteriori* dos parâmetros da EV4. A figura 7.16 mostra os resultados para cada um dos modelos estudados. No anexo 2 são mostrados os resultados numéricos para cada período de retorno e para cada modelo. Os dados observados, sistemáticos e não sistemáticos, foram plotados de acordo com a probabilidade de excedência calculada a partir do método descrito no item 6.4.

De uma forma geral, os modelos não apresentaram um ajuste satisfatório. Mesmo no caso dos modelos completos, há uma perda no ajuste a partir do período de retorno de 100 anos. A figura 7.17 mostra uma comparação entre os quantis estimados e os observados. Nota-se que os modelos de I a IV apresentam resultados sistematicamente melhores que os modelos de I* a IV*. Para os modelos V e V*, por questões de escala e para facilitar a visualização, foram plotados somente os dados sistemáticos.

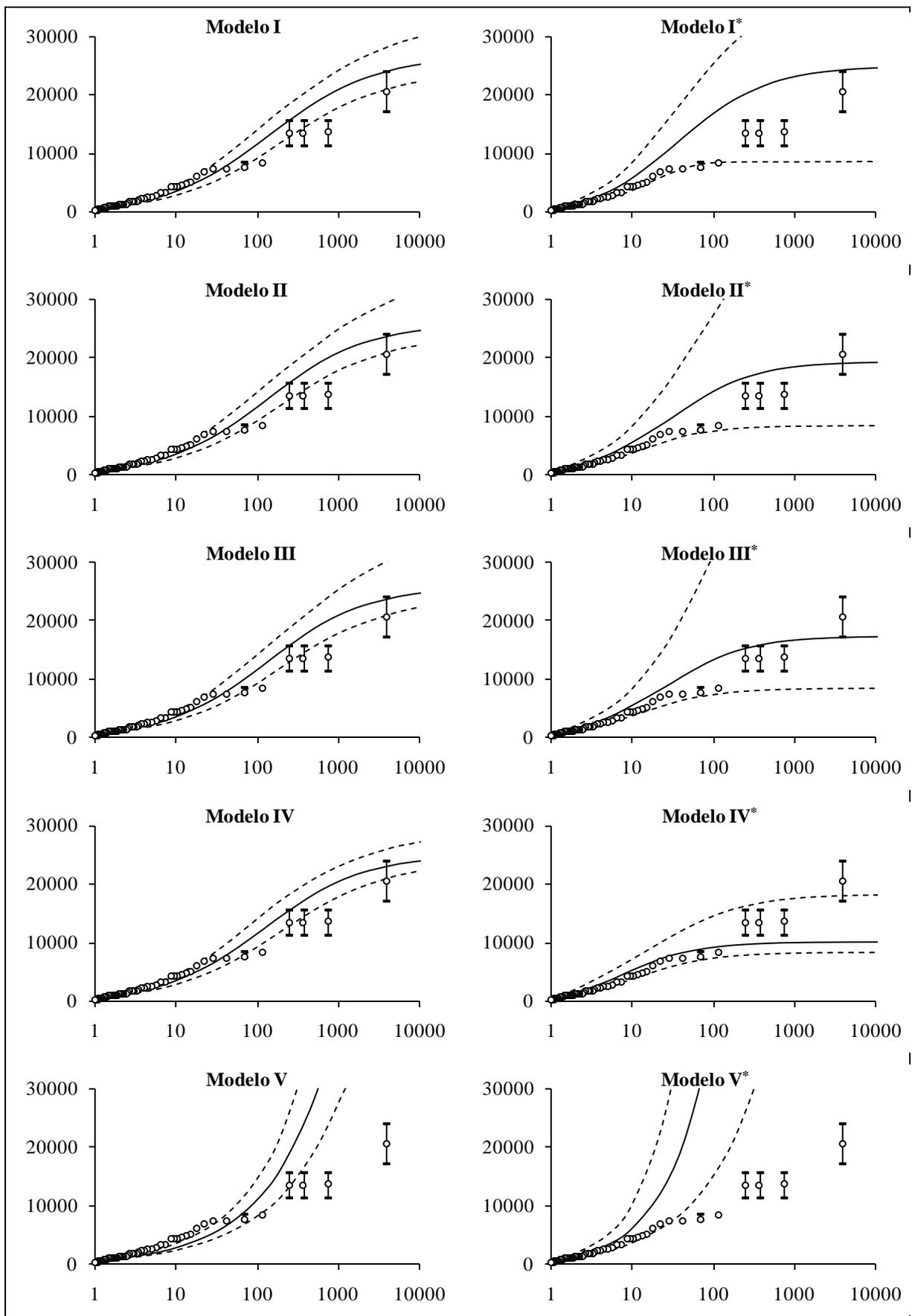


Figura 7.16 – Distribuição preditiva *a posteriori* (média) ou curva de quantis (linha contínua), intervalo de credibilidade de 95% (linha tracejada). Os círculos representam os dados sistemáticos e as barras com círculos representam os dados não sistemáticos. No eixo horizontal está o período de retorno em anos e no eixo vertical está a vazão em m³/s

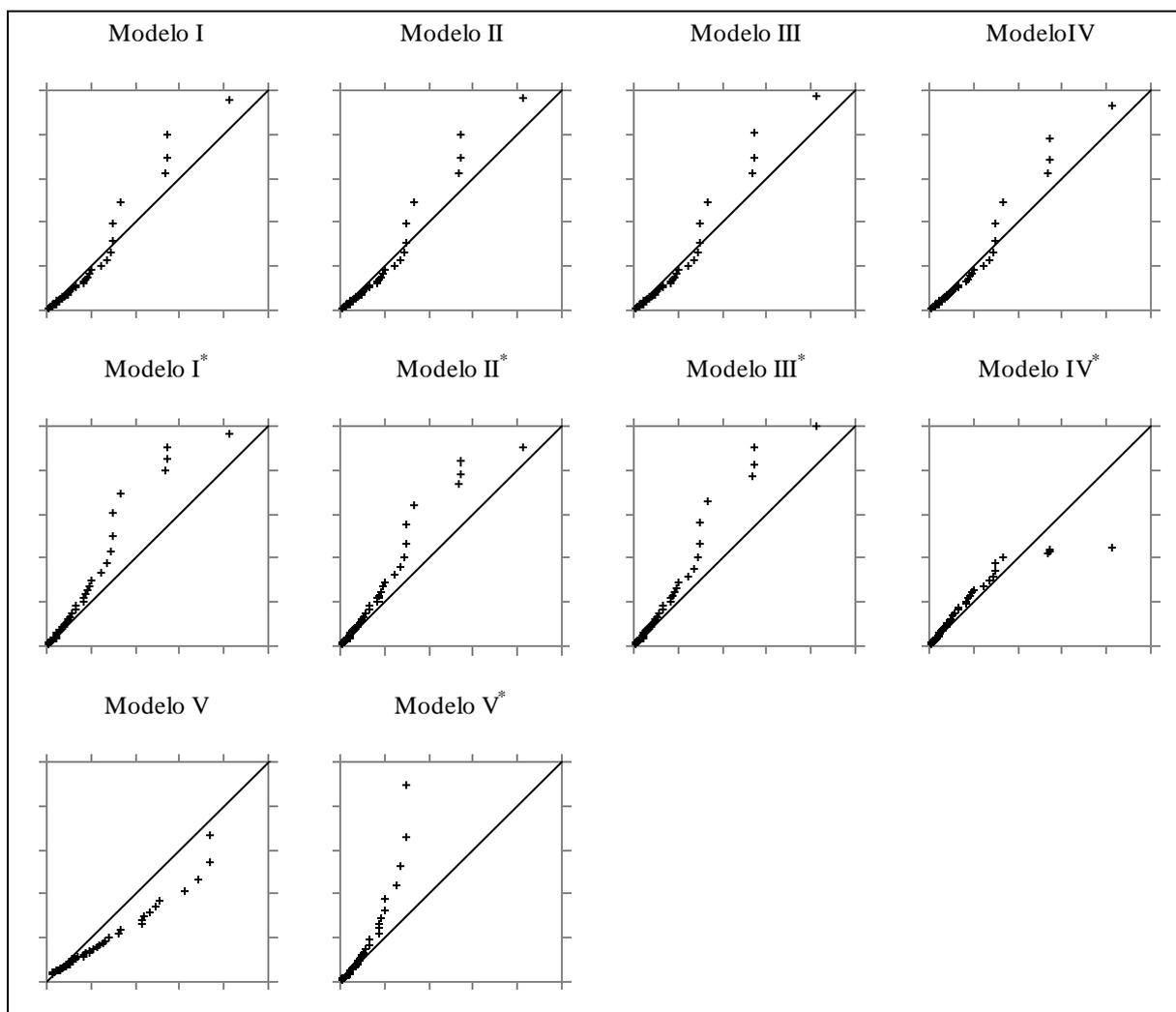


Figura 7.17 – Comparação entre os quantis estimados (eixo vertical) e observados (eixo horizontal) para o modelo EV4

As curvas de quantis da figura 7.16 e os gráficos da figura 7.17 permitem avaliar a adequação do modelo EV4. Como pode ser visto, tanto do ponto de vista da capacidade descritiva (I a IV) quanto da capacidade preditiva (I* a IV*), o modelo EV4 não é uma boa opção para a análise de vazões máximas anuais.

Dentre as distribuições *a priori* utilizadas, aquela que proporcionou o melhor ajuste, tanto para os modelos completos quanto para os modelos com somente dados sistemáticos, foi a II. No entanto, mesmo nesse caso, verifica-se que o modelo superestima sistematicamente as vazões com período de retorno superior a 100 anos. Para o modelo II*, verifica-se que não há qualquer garantia em relação à previsão de vazões com período de retorno superior a 50 anos.

A baixa capacidade descritiva e preditiva da EV4 pode ser relacionada à maneira como esse modelo se aproxima do limite superior. Com efeito, conforme mencionado anteriormente, a EV4 aproxima-se mais rapidamente do limite superior que a LN4. Assim, para períodos de retorno relativamente pequenos (< 1.000 anos), a EV4 fornece estimativas de quantis bastante

próximas do limite superior. Essa característica particular da EV4 implica em uma perda de ajuste, uma vez que sua forma em “S” não permite modelar as partes baixa, média e alta da curva de quantis simultaneamente.

A média *a posteriori* para o limite superior da EV4, nos modelos de I a IV, varia entre 25.608 m³/s e 27.427 m³/s, bastante semelhante à estimativa da PMF. Para os modelos I* a III*, a média *a posteriori* de α apresentou valores na mesma faixa, indicando que o modelo EV4 é pouco sensível às distribuições *a priori* para o limite superior escolhidas neste trabalho. Na verdade, a estimativa *a posteriori* de α parece ser mais influenciada pelo valor do máximo amostral de que à distribuição *a priori* de α .

7.2.5 Análise de sensibilidade da probabilidade de excedência da PMF

Com o objetivo de avaliar o efeito da locação da PMF, foram construídas mais duas distribuições *a priori* para α . Em ambos os casos, é feita a hipótese de que a estimativa da PMF descreve, com grande precisão, o limite superior. Ou seja, a probabilidade de excedência p da PMF assume valores extremos, como, por exemplo, 5% e 95%. No primeiro caso sabe-se que há uma boa chance de o limite superior ser maior que a estimativa PMF e no segundo há uma boa chance de o limite superior ser menor que a estimativa da PMF. Assim, adotando CV igual 0,50 e p igual a 5% e 95%, as distribuições *a priori* VI e VII são dadas, respectivamente, da seguinte forma:

$$\text{VI) } \alpha \sim \text{GAM}(4,00, 5,33 \times 10^{-5}) \rightarrow P(\alpha \leq \text{PMF}) = 5\%$$

$$\text{VII) } \alpha \sim \text{GAM}(4,00, 3,02 \times 10^{-4}) \rightarrow P(\alpha \leq \text{PMF}) = 95\%$$

A tabela 7.9 mostra as principais características dessas distribuições *a priori* e a figura 7.18 mostra suas respectivas formas juntamente com a distribuição II, que tem o mesmo coeficiente de variação.

Tabela 7.9 – Parâmetros e características das distribuições *a priori* VI e VII para o limite superior na bacia do rio American, em Folsom

Distribuição <i>a priori</i>	ρ_α	β_α	Média	Mediana	CV _{α}	DP	Proc.
VI	4,0000	5,33x10 ⁻⁵	75.103	68.894	0,5	37.552	A
VII	4,0000	3,02x10 ⁻⁴	13.235	12.151	0,5	6.617	A

CV: coeficiente de variação; DP: desvio padrão; Proc.: procedimento de cálculo

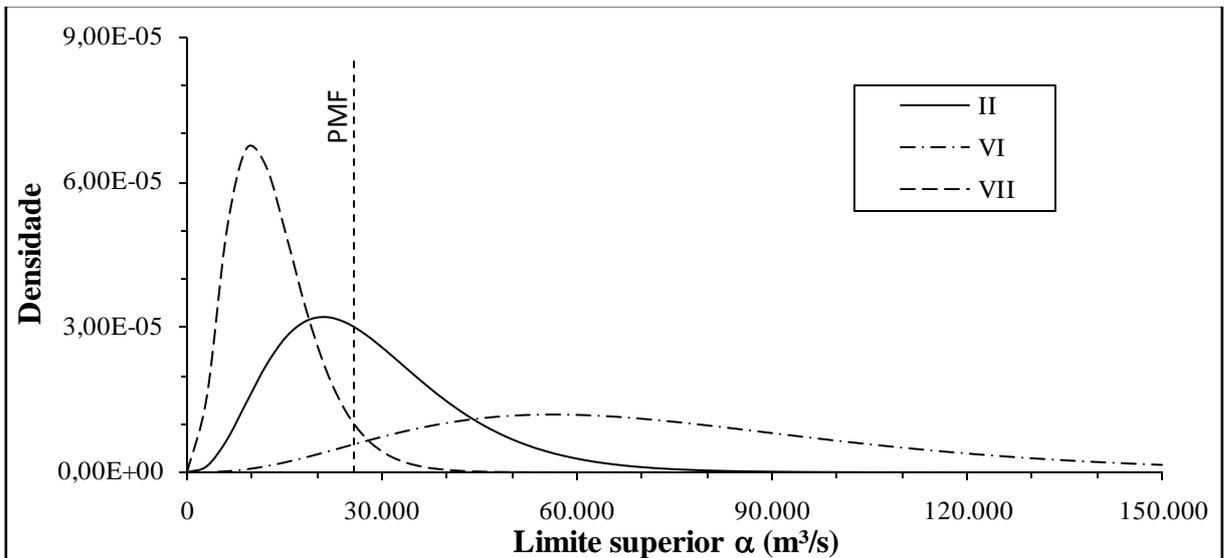


Figura 7.18 – Distribuições *a priori* II, VI e VII para o limite superior na bacia do rio American, em Folsom

A figura 7.19 mostra a distribuição *a posteriori* de α para os modelos VI e VII com as distribuições LN4 e EV4.

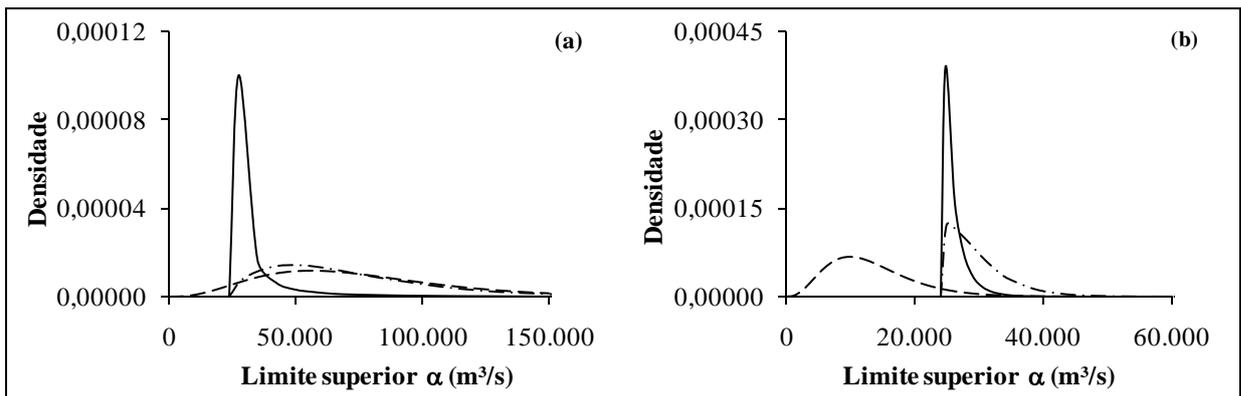


Figura 7.19 – Distribuições *a priori* (linha tracejada) e *a posteriori* de α dos modelos VI (gráfico a) e VII (gráfico b) para as distribuições LN4 (linha tracejada-pontilhada) e EV4 (linha contínua)

No modelo VI a probabilidade de o limite superior ser maior que a PMF é grande. Uma vez que o máximo amostral é menor que a estimativa da PMF, a informação *a posteriori* sobre o limite superior é dominada pela distribuição *a priori* de α , sendo a amostra pouco informativa em relação a esse parâmetro. Para o modelo LN4 (figura 7.19a), a distribuição *a posteriori* reflete somente o conhecimento *a priori*, mostrando que os dados não acrescentam à análise novas informações sobre o limite superior.

Para a EV4, as análises anteriores mostraram que esse modelo tende a produzir estimativas para o limite superior bastante próximas ao máximo amostral. Assim, a distribuição *a priori* exerce uma influência relativamente menor nesse caso. A figura 7.19a mostra esse efeito através da concentração da distribuição entre 24.069 m³/s e 50.000 m³/s. No entanto, vale

salientar que mesmo que pouco pronunciado, o efeito da distribuição *a priori* ainda pode ser notado. Para isso, basta verificar que as distribuições *a posteriori* do limite superior para a EV4 mostradas na figura 7.15 se concentram entre 24.069 m³/s e 40.000 m³/s.

No modelo VII, a probabilidade de o limite superior ser maior que a PMF é pequena. Uma vez que o máximo amostral é menor que a estimativa da PMF, porém próxima a essa, a informação *a posteriori* sobre o limite superior é dominada pela função de verossimilhança, ou seja, pelos dados amostrais. Esse fato é verificado na figura 7.19b, onde as distribuições *a posteriori* de α para os modelos LN4 e EV4 exibem claramente um comportamento de concentração próximo ao máximo amostral.

Por fim, foram construídas as curvas de quantis para ambos os modelos. A figura 7.20 mostra essas curvas juntamente com o intervalo de credibilidade para cada quantil. Os resultados em relação a essas foram semelhantes àqueles obtidos para os modelos de I a IV anteriormente discutidos. Assim, não serão feitas análises específicas para essas curvas.

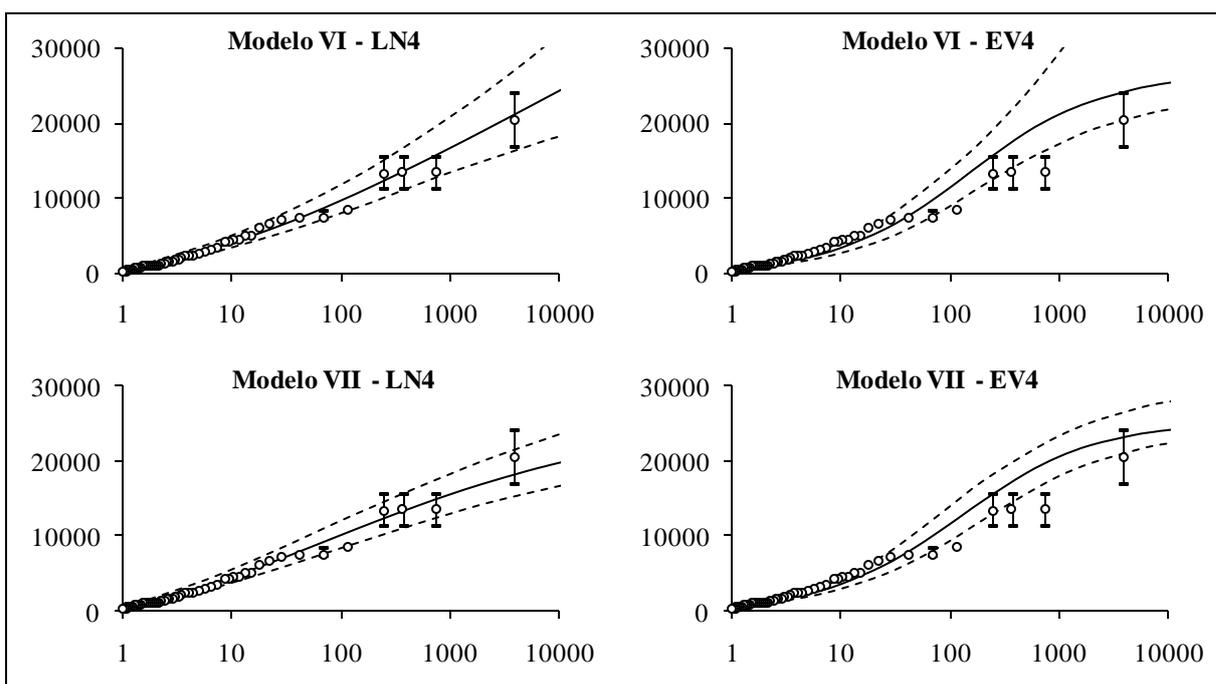


Figura 7.20 – Distribuição preditiva *a posteriori* (linha contínua), intervalo de credibilidade de 95% (linha tracejada) para os modelos VI e VII. No eixo horizontal está o período de retorno em anos e no eixo vertical está a vazão em m³/s

7.2.6 Comparação com modelos ilimitados superiormente

Os resultados obtidos para os modelos propostos neste trabalho foram contrastados com os resultados obtidos em outros estudos feitos para as cheias do rio American, em Folsom. Um estudo de referência para essa bacia foi feito pelo *United States Bureau of Reclamation* (USBR, 2002). Nesse estudo o USBR ajustou a distribuição Log-Pearson III (LPIII), com

assimetria positiva e, portanto, ilimitada superiormente, aos dados sistemáticos e não sistemáticos de Folsom usando um procedimento bayesiano de estimação de parâmetros, que busca uma estimativa pontual para os mesmos por meio da maximização do produto da função de verossimilhança pela distribuição *a priori* dos parâmetros. Ou seja, a estimativa pontual do parâmetro é dada pela moda de sua distribuição *a posteriori*. O procedimento, descrito em detalhes em O'Connell et al. (2002), emprega uma abordagem bayesiana para incluir diferentes tipo de incertezas na análise. Na aplicação em Folsom, USBR (2002) incluiu as seguintes fontes de incertezas: erros na medição de vazões, erros na estimação das paleocheias e na data de sua ocorrência e incertezas relacionadas aos parâmetros da distribuição LPIII.

Foram conduzidas duas comparações: a) entre os resultados obtidos pelo USBR (2002) e os resultados do modelo LN4-V, onde nenhuma informação *a priori* para o limite superior foi fornecida; e b) entre os resultados obtidos pelo USBR (2002) e os resultados do modelo LN4-III, o qual foi o melhor modelo pelo método proposto. Não foram feitas comparações com o modelo EV4 pelo fato desse ter dado resultados sistematicamente piores que aqueles obtidos com a LN4. A figura 7.21 mostra os resultados da primeira comparação e a figura 7.22 mostra os resultados da segunda.

Os resultados para os dois modelos não podem ser comparados diretamente uma vez que USBR (2002) incluiu incertezas que não foram levadas em consideração nos modelos propostos aqui. No entanto, alguns comentários gerais podem ser feitos. Em primeiro lugar, pelo menos para o caso estudado, o modelo limitado com o uso de uma distribuição *a priori* informativa para α (figura 7.22) se ajusta melhor aos dados que o modelo ilimitado, particularmente para pequenos períodos de retorno em que o modelo ilimitado claramente subestima os quantis de vazão. Outro ponto a ser notado se refere ao intervalo de credibilidade para esse modelo limitado, o qual é significativamente mais curto que o intervalo de credibilidade para o modelo ilimitado. Não obstante o fato de incertezas adicionais estarem incluídas no intervalo de credibilidade do modelo ilimitado, ele é semelhante ao intervalo de credibilidade determinado para o modelo V (figura 7.21), com uma distribuição *a priori* plana para o limite superior.

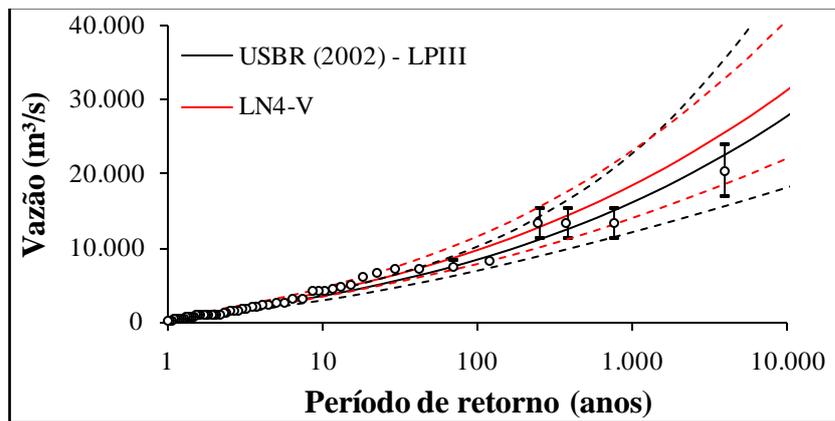


Figura 7.21 – Curva de quantis para o modelo limitado superiormente LN4 (Modelo V) e para o modelo ilimitado LP III; as linhas tracejadas correspondem aos intervalos de credibilidade de 95% de probabilidade

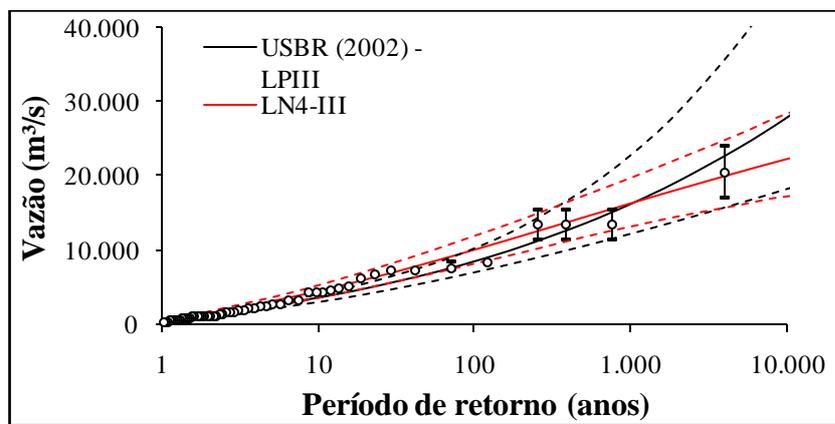


Figura 7.22 – Curva de quantis para o modelo limitado superiormente LN4 (Modelo III) e para o modelo ilimitado LP III; as linhas tracejadas correspondem aos intervalos de credibilidade de 95% de probabilidade

Quando não há informação *a priori* para o limite superior, como mostrado na figura 7.21, os resultados para os modelos limitados e ilimitados são equivalentes. No entanto, como desejável, o modelo ilimitado LP III se ajusta melhor aos dados extremos do que o modelo limitado. Este último mostra um melhor ajuste para vazões com período de retorno pequeno. Por outro lado, o modelo limitado apresentou um intervalo de credibilidade mais curto que o modelo ilimitado.

Conforme mencionado anteriormente, os resultados obtidos por USBR (2002) não podem ser comparados diretamente com os resultados obtidos uma vez que o conjunto de informações utilizado na análise é diferente. De forma a contornar esse problema, foi feita a análise dos dados de Folsom a partir da distribuição GEV, sendo a função de verossimilhança construída da mesma forma que para o modelo LN4. Essa distribuição foi escolhida devido à sua flexibilidade quanto à variável em análise, podendo ser limitada ou não, e pela sua ampla utilização em hidrologia de extremos (Martins e Stedinger, 2000). O modelo para a GEV tem a seguinte propriedade:

$$p(\Theta/x) \propto p(x/\Theta)p(\mu)p(\sigma)p(\xi), \quad (7.3)$$

onde $p(x/\Theta)$ é a função de verossimilhança para $\xi \neq 0$, sendo dada por:

$$\begin{aligned} L_{TOTAL}^{GEV} = & \left(\frac{1}{\sigma}\right)^{N_S^*} \prod_{i=1}^{N_S^*} \left[1 - \frac{\xi}{\sigma}(x_i - \mu)\right]^{\frac{1}{\xi}-1} \cdot \exp\left\{-\left[1 - \frac{\xi}{\sigma}(x_i - \mu)\right]^{\frac{1}{\xi}}\right\} \\ & \prod_{i=1}^{N_S^<} \exp\left\{-\left[1 - \frac{\xi}{\sigma}(x_{Ui} - \mu)\right]^{\frac{1}{\xi}}\right\} \\ & \prod_{i=1}^{N_H^<} \exp\left\{-\left[1 - \frac{\xi}{\sigma}(y_{Ui} - \mu)\right]^{\frac{1}{\xi}}\right\} \\ & \prod_{i=1}^{N_H^>} \left\{ \exp\left\{-\left[1 - \frac{\xi}{\sigma}(y_{URi} - \mu)\right]^{\frac{1}{\xi}}\right\} - \exp\left\{-\left[1 - \frac{\xi}{\sigma}(y_{LRi} - \mu)\right]^{\frac{1}{\xi}}\right\} \right\}, \end{aligned} \quad (7.4)$$

e $p(\mu)$, $p(\sigma)$ e $p(\xi)$ são, respectivamente, a distribuição *a priori* para os parâmetros μ , σ e ξ .

Martins e Stedinger (2000) estabeleceram uma distribuição *a priori* para ξ , denominada de geofísica, que limita os valores desse parâmetro a uma faixa fisicamente razoável. A distribuição *a priori* para ξ é tal que $\xi \in \underline{v} \cup -0,5$, onde $v \sim \text{Beta}(p, q)$. Assim, a distribuição *a priori* de ξ toma a seguinte forma:

$$p(\xi) = \frac{(0,5 + \xi)^{p-1} (0,5 - \xi)^{q-1}}{B(p, q)}, \quad (7.5)$$

onde $B(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$, $p = 6$ e $q = 9$. Essa distribuição limita os valores de ξ a $-0,5 \leq \xi \leq +0,5$.

Para os demais parâmetros não foram construídas distribuições *a priori* informativas. Assim, uma distribuição plana foi tomada como especificação *a priori* para os parâmetros com base em seus domínios. Para μ e σ , tem-se:

$$\mu \sim \text{NOR}(1,00, 1,00 \times 10^{-6}) \text{ e}$$

$$\sigma \sim \text{GAM}(1,00, 1,00 \times 10^{-8}).$$

Foi gerada, no WinBUGS, uma amostra de tamanho 600.000 sendo que, a cada 10 realizações do algoritmo, uma era selecionada para análise ($lag = 10$). Ou seja, selecionou-se para análise uma amostra de tamanho 60.000. As 10.000 primeiras foram descartadas como *burn in*, restando uma amostra final composta de 50.000 realizações da distribuição conjunta *a posteriori* dos parâmetros da GEV. A tabela a seguir mostra as estimativas *a posteriori* para os parâmetros da GEV.

Tabela 7.10 – Estimativas *a posteriori* para os parâmetros da GEV

Parâmetro	Média	CV	DP	95% HPD
μ	1.069	0,12	128	(824, 1.326)
ξ	-0,31	-0,21	0,064	(-0,43, -0,18)
σ	826	0,15	128	(583, 1.077)

A figura 7.23 mostra os resultados comparativos entre a GEV e o modelo LN4-V. Em ambos os modelos não há qualquer informação *a priori* sobre os parâmetros, o que permite compará-los sob a mesma base de conhecimento.

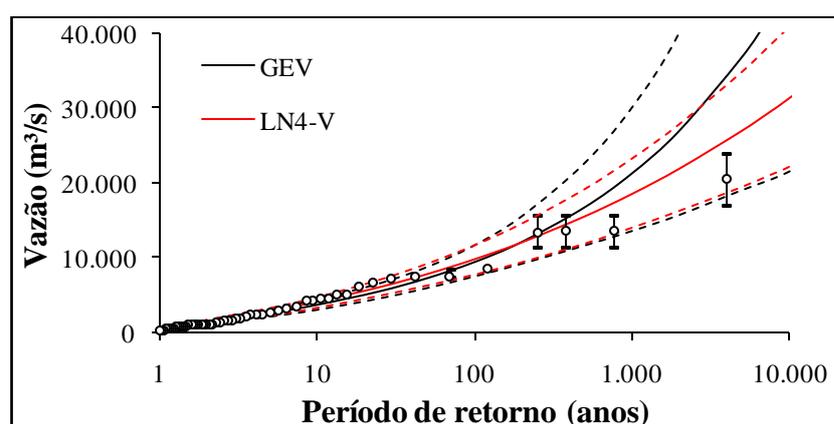


Figura 7.23 – Curva de quantis para o modelo limitado superiormente LN4 (Modelo V) e para o modelo ilimitado GEV; as linhas tracejadas correspondem aos intervalos de credibilidade de 95% de probabilidade

A figura 7.23 mostra que os modelos são similares até um período de retorno em torno de 300 anos. A partir desse ponto a GEV perde sua capacidade preditiva superestimando os quantis com alto período de retorno. A LN4-V, por outro lado, tem uma melhor capacidade preditiva, a despeito do fato deste também superestimar os quantis com alto período de retorno. Outro fato digno de menção é o intervalo de credibilidade para a LN4-V ser mais curto que o da GEV. Isso resulta em uma maior confiança nos resultados obtidos pelo modelo limitado.

A comparação entre os modelos LN4-V e a GEV através da curva de quantis somente leva à conclusão de que o modelo limitado é indicado mesmo no caso de não haver qualquer informação *a priori* acerca do limite superior. No entanto, apontar um modelo como melhor

que outro é uma tarefa um tanto complexa que requer, além da avaliação da capacidade descritiva e preditiva, uma análise física e/ou hidrológica do fenômeno que se está modelando. No caso do modelo LN4-V, por exemplo, a estimativa *a posteriori* para o limite superior, mostrada na tabela 7.5, está completamente fora da realidade hidrológica do rio American, fazendo com que esse modelo seja descartado, independentemente do ajuste obtido.

Por outro lado, ao avaliar a GEV verifica-se que houve um ajuste aceitável, principalmente para vazões com período de retorno pequeno. No entanto, a premissa adotada nesta tese é de que as vazões máximas anuais são variáveis aleatórias limitadas superiormente e, assim, não faz sentido, do ponto de vista conceitual, modelá-las com uma GEV com parâmetro de forma ξ menor que zero, como foi o caso encontrado aqui.

Essa discussão levanta uma questão que transcende o escopo deste trabalho: o que é mais indicado: adotar um modelo conceitualmente inadequado que, por outro lado, possui uma melhor capacidade preditiva ou um modelo conceitualmente adequado, porém com uma pior capacidade preditiva?

Obviamente, essa questão não surge do estudo de caso do rio American, onde o modelo proposto, além de conceitualmente mais adequado (variáveis limitadas \rightarrow modelo limitado), foi o que mostrou maior capacidade preditiva. Para verificar esse fato, basta comparar o modelo LN4-III com a GEV, como mostrado na figura 7.24.

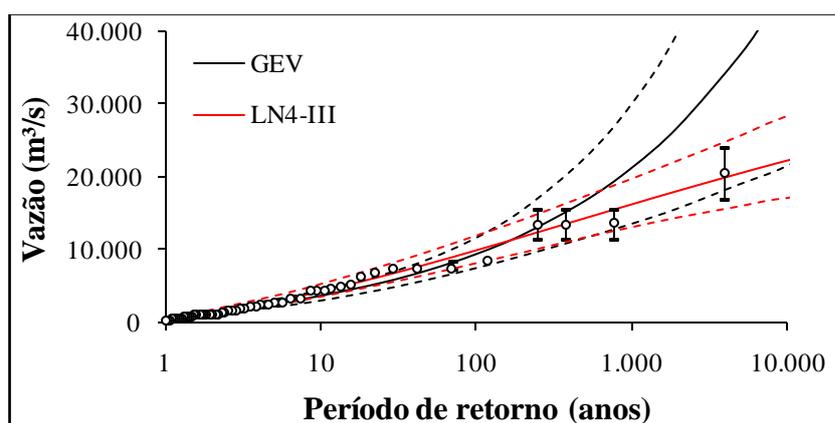


Figura 7.24 – Curva de quantis para o modelo limitado superiormente LN4 (Modelo III) e para o modelo ilimitado GEV; as linhas tracejadas correspondem aos intervalos de credibilidade de 95% de probabilidade

A resposta à questão levantada anteriormente passa pela verificação da capacidade de generalização do resultado mostrado na figura 7.24. Na conclusão desta tese é feito um esboço de um experimento capaz de demonstrar essa capacidade de generalização.

7.3 Aplicação para a bacia do rio Llobregat, em Pont Du Vilomara

7.3.1 A bacia do rio Llobregat

Outra aplicação do método proposto nesta tese foi feita na bacia do rio Llobregat, em Pont Du Vilomara, localizada na região da Catalunha, nordeste da Espanha. Essa bacia foi escolhida pelo fato de ter sido objeto de estudo de Botero (2006), que empregou as distribuições limitadas superiormente consideradas nesta tese sob a ótica da análise de frequência clássica. Com isso, será possível avaliar as principais características do método bayesiano frente aos métodos frequentistas. A figura 7.25 mostra a bacia do rio Llobregat, enquanto a figura 7.26 ilustra alguns trechos do rio.

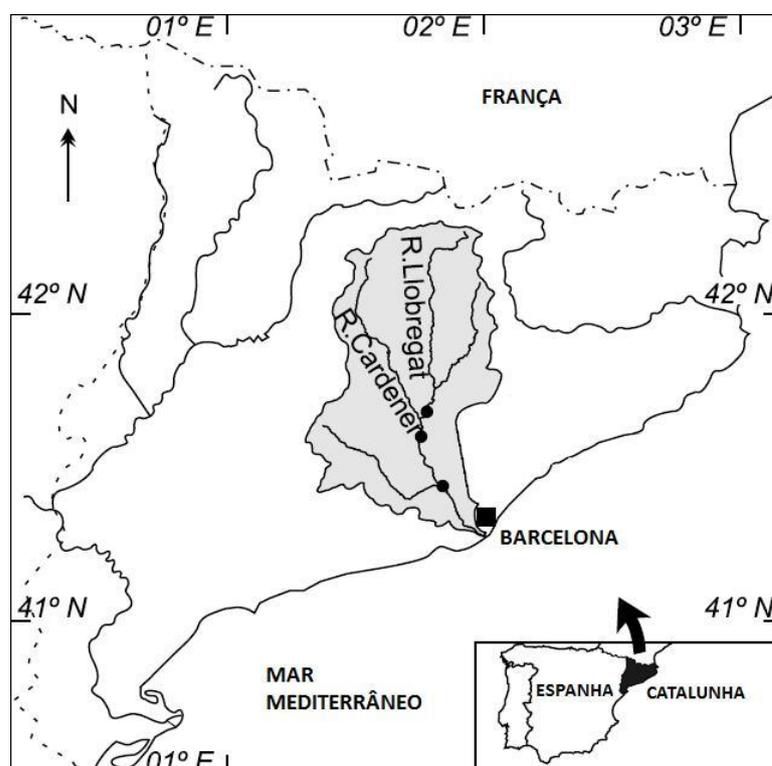


Figura 7.25 – Localização da bacia do rio Llobregat (adapt. Thorndycraft et al., 2005)



Figura 7.26 – Trechos da bacia do rio Llobregat (Fonte: ACA, 2001)

7.3.2 Dados hidrológicos sistemáticos e não sistemáticos

7.3.2.1 Dados sistemáticos

A bacia do rio Llobregat é monitorada pelo CEDEX, que possui uma extensa rede de monitoramento na Espanha. A estação fluviométrica de interesse para a aplicação do método é a Vilomara, cujo código dado pelo CEDEX é 10031. Essa estação possui dados de 1916 até 1989, e a área de drenagem a montante da mesma é de 1.885 km².

Botero (2006) fez uma reconstituição da série de vazões máximas diárias em Vilomara de forma a obter a série de máximos instantâneos em regime natural. De maneira geral, Botero (2006) obteve uma relação entre os máximos diários e instantâneos observados, o que permitiu estimar os máximos instantâneos quando não monitorados. Na seqüência, foi feito o balanço hídrico dos reservatórios da bacia e, assim, foi possível obter uma estimativa não regularizada de máximos instantâneos. As principais características para essa série são:

Mínimo:	25 m ³ /s
Máximo:	1.148 m ³ /s
Média:	221 m ³ /s
Desvio padrão:	194 m ³ /s
Coefficiente de variação:	0,89
Coefficiente de assimetria:	2,70

A figura 7.27 mostra os dados sistemáticos considerados na análise.

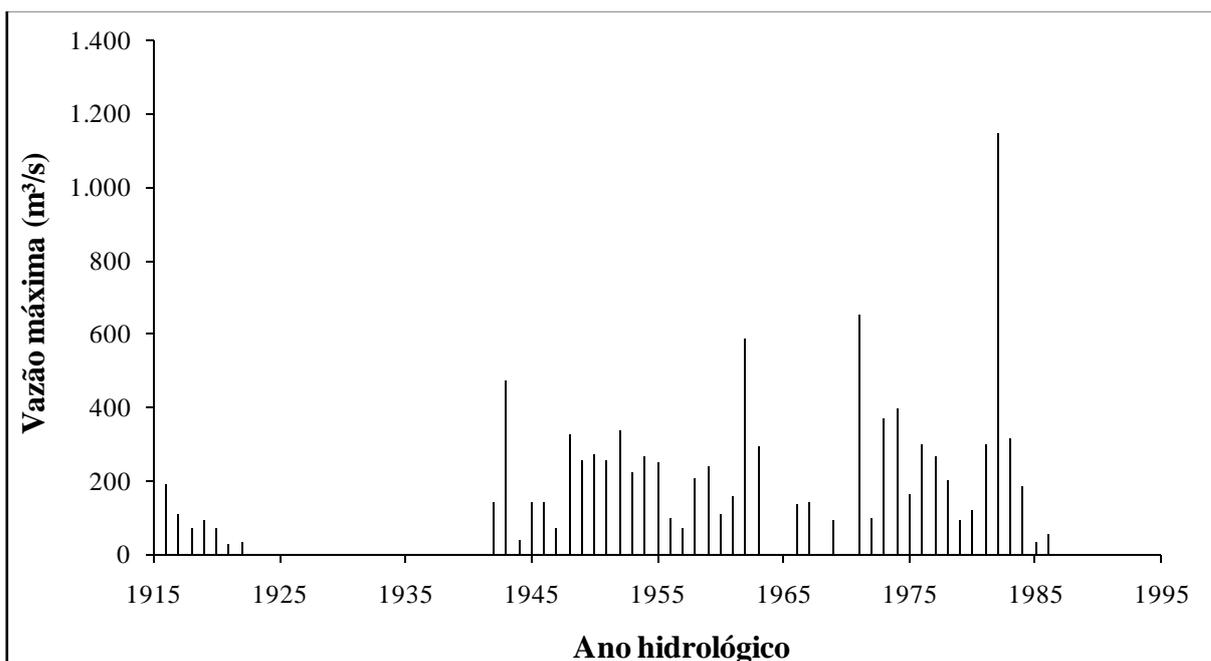


Figura 7.27 – Dados sistemáticos do rio Llobregat, em Pont Du Vilomara

7.3.2.2 Dados não sistemáticos

Os dados não sistemáticos para Vilomara correspondem aos dados históricos compilados pelo projeto europeu SPHERE (*Systematic, Paleoflood and Historical Data for the Improvement of Flood Risk Estimation* - <http://www.ccma.csic.es/dpts/suelos/hidro/sphere/home.html>). Os dados não sistemáticos para Vilomara foram obtidos por Thorndycraft et al. (2005) a partir da datação, por meio de ^{14}C , dos perfis estratigráficos de Pont Du Vilomara, e posterior modelagem hidráulica através do *software* HEC-RAS, desenvolvido pelo *Hydrologic Engineering Center do U. S. Army Corps of Engineers* (<http://www.hev.usace.army.mil>). 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 tabela 7.11 mostra as cheias não sistemáticas consideradas na análise. A figura 7.28 mostra a faixa de dados históricos em Pont Du Vilomara.

Tabela 7.11 – Cheias históricas em Pont Du Vilomara.

Ano	Vazão (m³/s)	Tipo
669 bC	(4850; ...)	LB
1600	(4850; ...)	LB
1650	(4950; ...)	LB
1700	(5100; ...)	LB
734 bC - 1600	(... ; 4850)	UB
1600-1650	(... ; 4950)	UB
1650-1700	(... ; 5100)	UB
1700-1915	(... ; 2300)	UB
1970	(2600; 4850)	DB
1987	(602; 2399)	DB
1965	(396; 2300)	DB
1968	(257; 2300)	DB
1989	(233; 2300)	DB
1988	(51; 2300)	DB

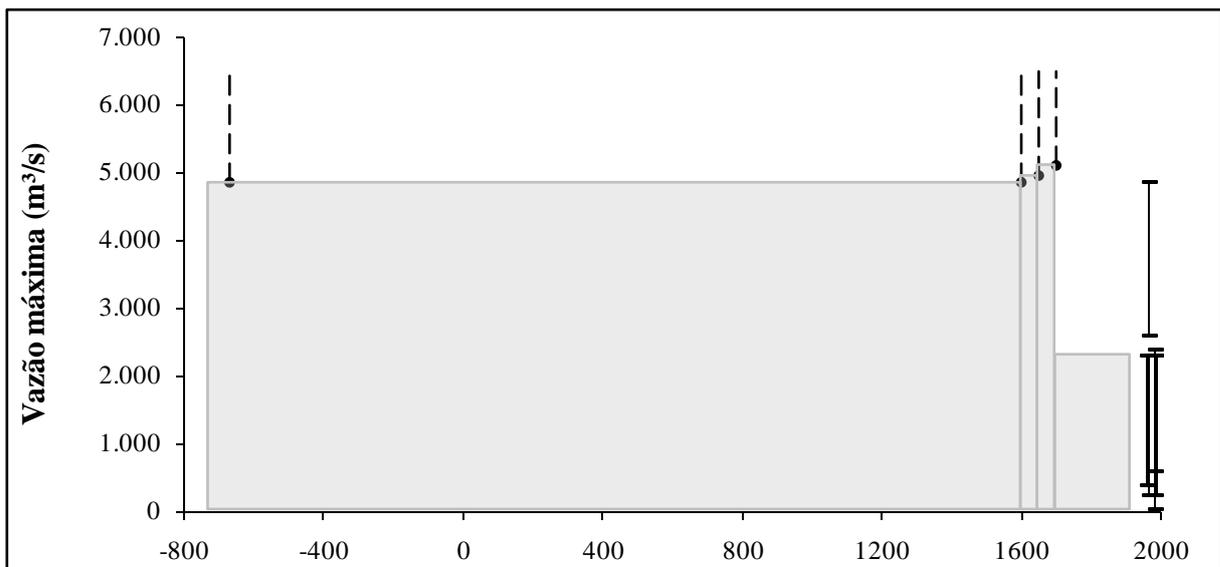


Figura 7.28 – Dados não sistemáticos do rio Llobregat, em Pont Du Vilomara

7.3.3 Distribuição *a priori* para o limite superior em Pont Du Vilomara

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 m³/s. Seguindo o procedimento A do item 6.5.2.1, com $p = 0,5$ e $CV = 1,0$ e $2,0$, têm-se, respectivamente, as seguintes distribuições *a priori* para o limite superior:

$$I) \quad \alpha \sim \text{GAM}(1,00, 5,38 \times 10^{-5})$$

$$II) \quad \alpha \sim \text{GAM}(0,25, 3,27 \times 10^{-6})$$

Seguindo, agora, o procedimento B do item 6.5.2.2, as 242 cheias recordes compiladas pelo CEDEX foram transpostas para a bacia do rio Llobregat, em Pont Du Vilomara. A transposição dessas tormentas foi feita a partir da equação 6.64, com $A_0 = 1.885 \text{ km}^2$.

A figura 7.29 mostra o histograma das cheias recordes transpostas para Pont Du Vilomara. A forma desse histograma é semelhante à forma dos histogramas mostrados na figura 6.8, para grupos de cheias recordes com áreas de drenagem semelhantes, ou seja, as cheias recordes são distribuídas assimetricamente para a direita, o que permite recomendar a distribuição Gama como modelo para as estimativas transpostas. Os parâmetros da distribuição Gama foram estimados pelo método dos momentos convencionais. A distribuição *a priori* para o limite superior, de acordo com o procedimento B, é a seguinte:

$$\text{III) } \alpha \sim \text{GAM}(0,51, 7,60 \times 10^{-4})$$

A figura 7.29 mostra a distribuição Gama ajustada às estimativas transpostas de cheias recordes.

Adicionalmente às distribuições *a priori* definidas a partir dos procedimentos A e B, foi considerada também uma distribuição não informativa para o limite superior, tal como para a aplicação do rio American, em Folsom.

A tabela 7.12 resume as principais características das distribuições usadas como especificação *a priori* para o limite superior. A figura 7.30 mostra suas respectivas formas.

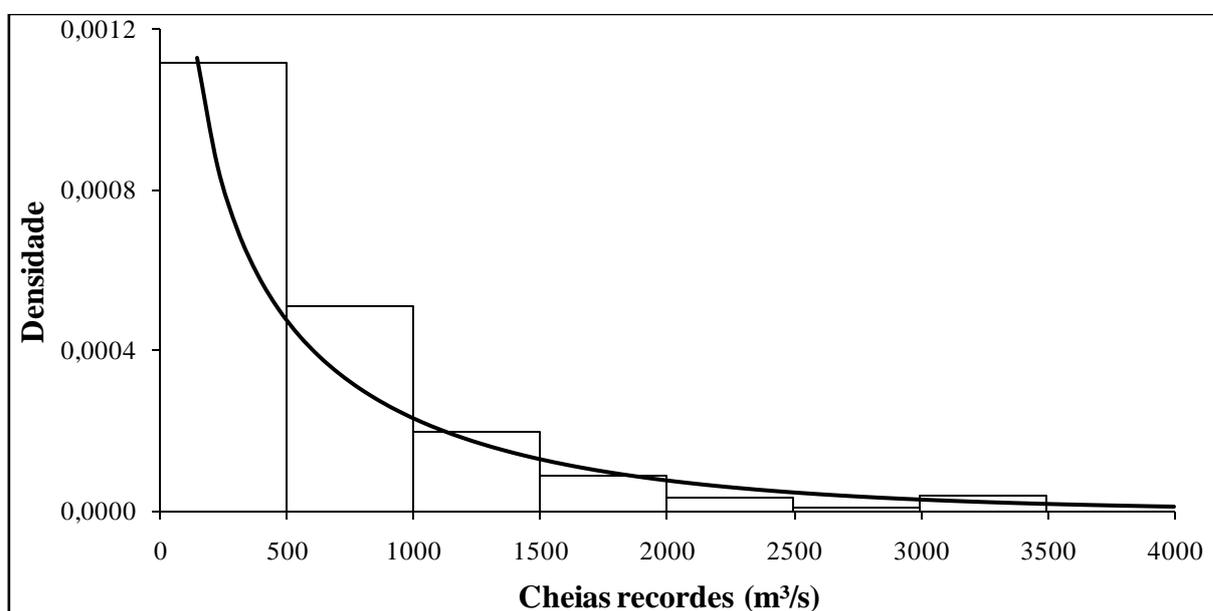


Figura 7.29 – Histograma das estimativas de cheias recordes transpostas para a bacia do rio Llobregat e ajuste da distribuição Gama

Tabela 7.12 – Parâmetros e características das distribuições *a priori* para o limite superior na bacia do rio Llobregat, em Pont Du Vilomara

Distribuição <i>a priori</i>	ρ_α	β_α	Média	Mediana	CV_α	DP	Proc.
I	1,00	$5,38 \times 10^{-5}$	18.592	12.887	1,0	18.592	A
II	0,25	$3,27 \times 10^{-6}$	76.371	12.887	2,0	152.742	A
III	0,51	$7,60 \times 10^{-4}$	668	308	1,4	938	B
IV	1,00	$1,00 \times 10^{-8}$	$1,00 \times 10^8$	$\sim 6,9 \times 10^7$	1,0	$1,00 \times 10^{-8}$	-

CV: coeficiente de variação; DP: desvio padrão; Proc.: procedimento de cálculo

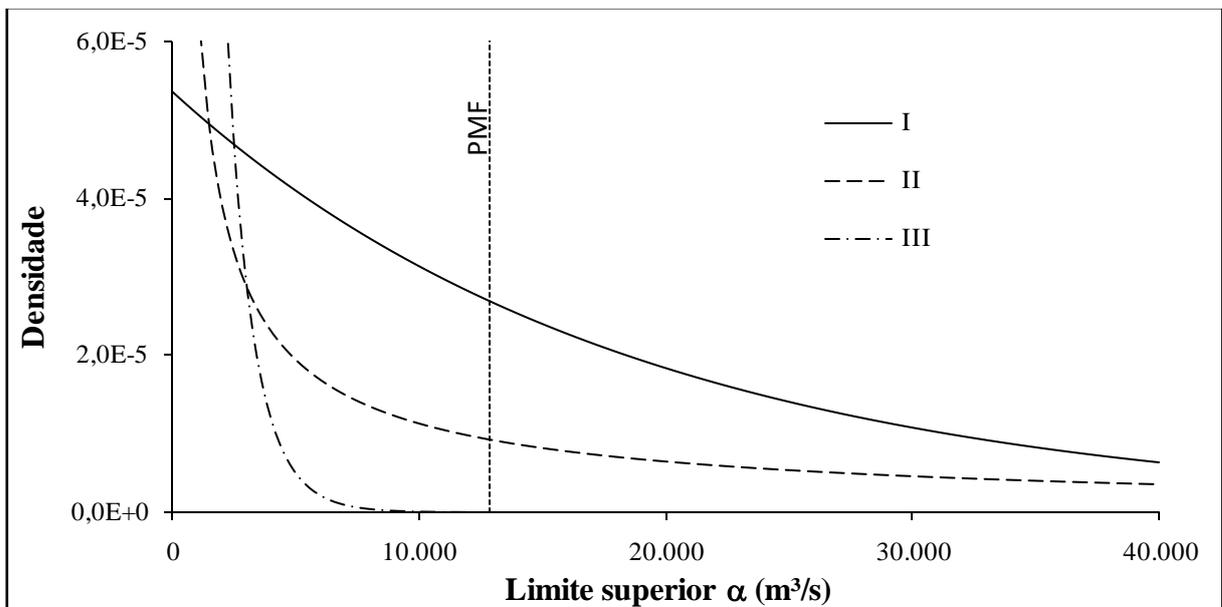


Figura 7.30 – Distribuições *a priori* de I a III para o limite superior na bacia do rio Llobregat

7.3.4 Estatísticas *a posteriori*

As principais conclusões sobre as estatísticas *a posteriori* para Pont Du Vilomara são bastante semelhantes às feitas para Folsom. Assim, os resultados para a Espanha serão apresentados de forma simplificada.

7.3.4.1 Modelo LN4

Da mesma forma que para os modelos do rio American, foram empregadas distribuições não informativas como especificação *a priori* para os parâmetros μ e σ do modelo LN4.

A partir das distribuições *a priori* para α , definidas anteriormente, foram construídos 4 modelos utilizando todos os dados disponíveis (sistemáticos e não sistemáticos) e 4 modelos utilizando somente os dados sistemáticos. A construção dos modelos com somente dados sistemáticos tem por objetivo avaliar a contribuição da informação não sistemática na análise.

Para cada modelo foi gerada, no WinBUGS, uma amostra de tamanho 1.200.000 sendo que, a cada 20 realizações do algoritmo, uma era selecionada para análise ($lag = 20$), ou seja, selecionou-se para análise uma amostra de tamanho 60.000. Dessa amostra de 60.000 valores, os 10.000 primeiros foram descartados como *burn in*, compondo uma amostra final de 50.000 realizações da distribuição conjunta *a posteriori* dos parâmetros da LN4.

A tabela 7.13 mostra os resultados para o parâmetro μ . A figura 7.31 mostra as distribuições *a posteriori* do parâmetro μ para os modelos completos e somente com os dados sistemáticos.

Tabela 7.13 – Estatísticas *a posteriori* para o parâmetro μ da LN4

Conjunto de dados	Modelo	Média	CV	DP	95% HPD
Sistemáticos e não sistemáticos	I	-5,12	-0,110	0,561	(-6,18, -4,01)
	II	-6,39	-0,162	1,032	(-8,32, -4,42)
	III	-3,86	-0,067	0,259	(-4,36, -3,36)
	IV	-12,61	-0,097	1,221	(-14,74, -10,13)
Sistemáticos	I*	-4,31	-0,232	1,000	(-6,10, -2,38)
	II*	-5,26	-0,315	1,657	(-8,04, -2,17)
	III*	-2,56	-0,172	0,439	(-3,43, -1,77)
	IV*	-12,74	-0,105	1,335	(-14,92, -10,13)

95% HPD - Intervalo de Credibilidade; os modelos marcados com '*' são aqueles formulados somente com informação sistemática

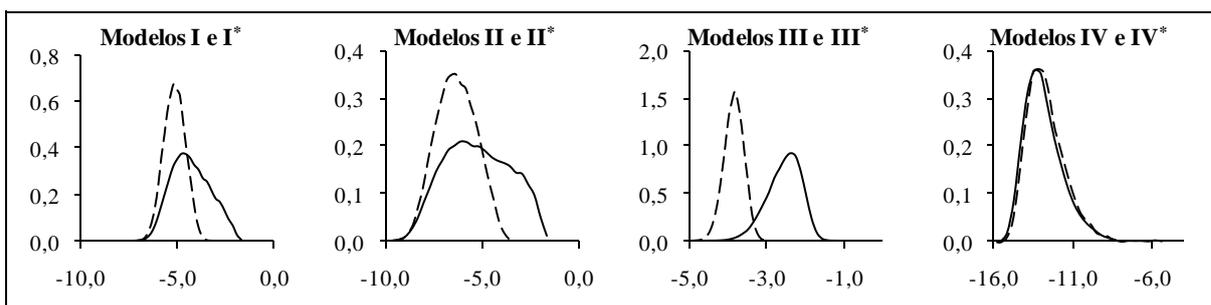


Figura 7.31 – Distribuições *a posteriori* do parâmetro μ para os modelos completos (linha tracejada) e para os modelos com somente dados sistemáticos (linha contínua)

Dois fatos relevantes a serem destacados para o parâmetro μ são: (i) há uma redução significativa do intervalo de credibilidade com a inclusão de informação não sistemática e (ii) há uma redução do CV *a posteriori* de μ com a redução do CV *a priori* de α , ou seja, há uma relação direta entre esses dois parâmetros.

Como era esperado, para o modelo IV (distribuição não informativa para α), não há um ganho aparente em se utilizar os dados não sistemáticos. No entanto, a forma bastante discrepante dessa distribuição em relação aos demais modelos e o grande intervalo de credibilidade leva a concluir em favor dos modelos com distribuição informativas para o limite superior.

Estatísticas a posteriori para o parâmetro σ

A tabela 7.14 mostra os resultados para o parâmetro σ . A figura 7.32 mostra as distribuições *a posteriori* do parâmetro σ para os modelos completos e somente com os dados sistemáticos.

Tabela 7.14 – Estatísticas *a posteriori* para o parâmetro σ da LN4

Conjunto de dados	Modelo	Média	CV	DP	95% HPD
Sistemáticos e não sistemáticos	I	1,08	0,0646	0,0694	(0,94, 1,22)
	II	1,05	0,0624	0,0654	(0,93, 1,18)
	III	1,19	0,0717	0,0854	(1,03, 1,36)
	IV	1,03	0,0587	0,0606	(0,92, 1,16)
Sistemáticos	I*	0,86	0,1123	0,0963	(0,68, 1,05)
	II*	0,85	0,1128	0,0959	(0,68, 1,04)
	III*	0,94	0,1192	0,1119	(0,74, 1,16)
	IV*	0,83	0,1066	0,0887	(0,67, 1,01)

95% HPD - Intervalo de Credibilidade; os modelos marcados com '*' são aqueles formulados somente com informação sistemática.

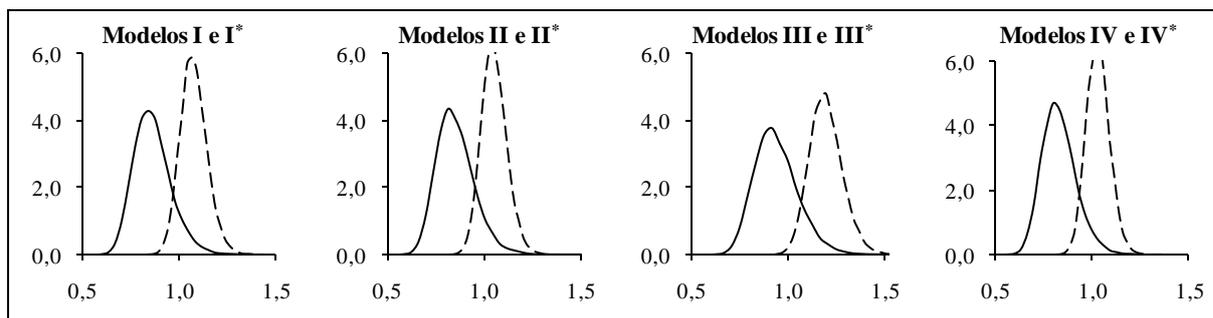


Figura 7.32 – Distribuições *a posteriori* do parâmetro σ para os modelos completos (linha tracejada) e para os modelos com somente dados sistemáticos (linha contínua).

Para o parâmetro σ , os resultados *a posteriori* se mostraram bastantes semelhantes para todos os modelos. Esse fato é um indicativo de que esse parâmetro é pouco sensível à especificação *a priori* para α . Os resultados da tabela 7.14 mostram que a inclusão de dados não sistemáticos proporciona uma sensível diminuição do coeficiente de variação e do intervalo de credibilidade. Mais uma vez, a comparação entre os modelos completos e somente dados com sistemáticos permite concluir em favor dos primeiros.

A tabela 7.15 e figura 7.33 mostram os resultados para o limite superior α .

Tabela 7.15 – Estatísticas a posteriori para o limite superior α da LN4

Conjunto de dados	Modelo	Média	CV	DP	95% HPD
Sistemáticos e não sistemáticos	I	38.116	0,57	21.689	(8.704, 81.390)
	II	191.151	1,13	215.395	(7.806, 610.600)
	III	9.991	0,21	2.064	(6.574, 14.120)
	IV	$1,00 \times 10^8$	0,99	$9,93 \times 10^7$	$(2,81 \times 10^5, 2,97 \times 10^8)$
Sistemáticos	I*	19.080	0,95	18.130	(1.219, 55.420)
	II*	99.257	1,65	163.559	(1.241, 419.600)
	III*	2.546	0,44	1.125	(1.212, 4.760)
	IV*	$1,00 \times 10^8$	1,00	$1,00 \times 10^8$	$(3,39 \times 10^4, 3,02 \times 10^8)$

95% HPD - Intervalo de Credibilidade; os modelos marcados com '*' são aqueles formulados somente com informação sistemática

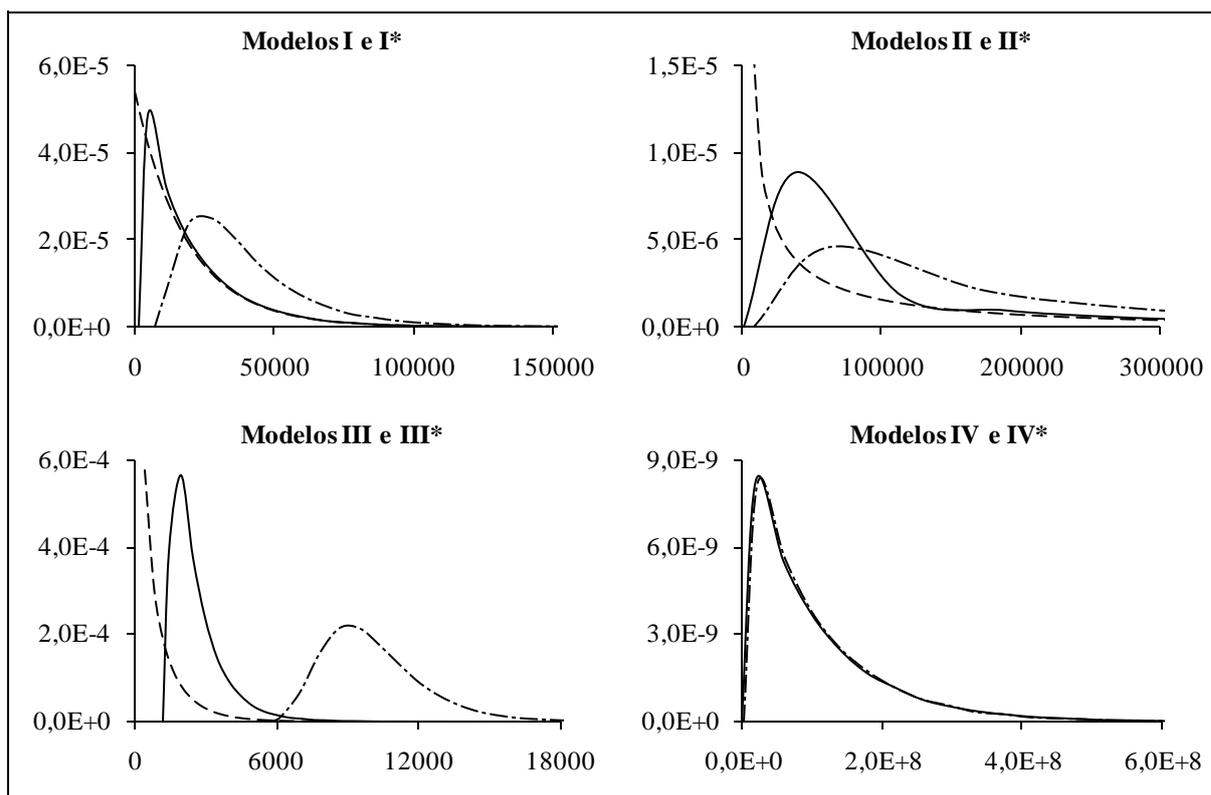


Figura 7.33 – Distribuições a priori (linha tracejada), a posteriori para os modelos com todos os dados (linha tracejada-pontilhada) e a posteriori para os modelos com dados sistemáticos (linha contínua)

A tabela 7.15 mostra uma sensível redução no coeficiente de variação quando os dados não sistemáticos são incluídos na análise. Isso pode ser verificado, em particular, para os modelos I e I*, com uma redução de quase 50% para o CV do primeiro e uma redução insignificante para o segundo. Assim como em Folsom, uma vez que os dados sistemáticos são muito

menores que a estimativa para o limite superior, não é esperado que estes contribuam com novas informações sobre o limite superior. De fato, a figura 7.33 mostra que a distribuição *a posteriori* para o parâmetro α no modelo I* é bastante semelhante à distribuição *a priori*. A tabela 7.15 mostra que não há uma redução do coeficiente de variação. Isso sugere que o conhecimento sobre o limite superior não aumenta se somente os dados sistemáticos são incluídos na análise.

Um resultado não esperado se refere ao intervalo de credibilidade, o qual possui valores significativamente menores para os modelos com somente dados sistemáticos. Avaliando os modelos somente sob esse aspecto, poderia se concluir a favor dos modelos com somente dados sistemáticos. No entanto, a sensível redução do coeficiente de variação da distribuição do limite superior para os modelos completos e a curva de quantis, vista a seguir, levam à conclusão contrária.

A análise dos modelos IV e IV* mostra que, independentemente do tipo de dado hidrológico incluído na análise, o uso de uma distribuição não informativa para o limite superior é uma prática não recomendada. De fato, a análise dos resultados encontrados para Pont Du Vilomara é bastante semelhantes àquela elaborada para Folsom.

Curva de quantis

A curva de quantis, ou a distribuição preditiva *a posteriori*, foi estimada por integração de Monte Carlo (vide item 5.3) a partir dos valores da distribuição *a posteriori* dos parâmetros da LN4. A figura 7.34 mostra os resultados para cada um dos modelos estudados. No anexo 3, são mostrados os resultados numéricos para cada período de retorno e para cada modelo. Os dados observados, sistemáticos e não sistemáticos, foram plotados de acordo com a probabilidade de excedência calculada a partir do método descrito no item 6.4. No caso de Pont Du Vilomara, foi feita uma simplificação, uma vez que existem dados do tipo LB. Para esse tipo de dado, não é possível estabelecer a probabilidade de excedência empírica e, assim, os mesmos não foram incluídos na curva de quantis.

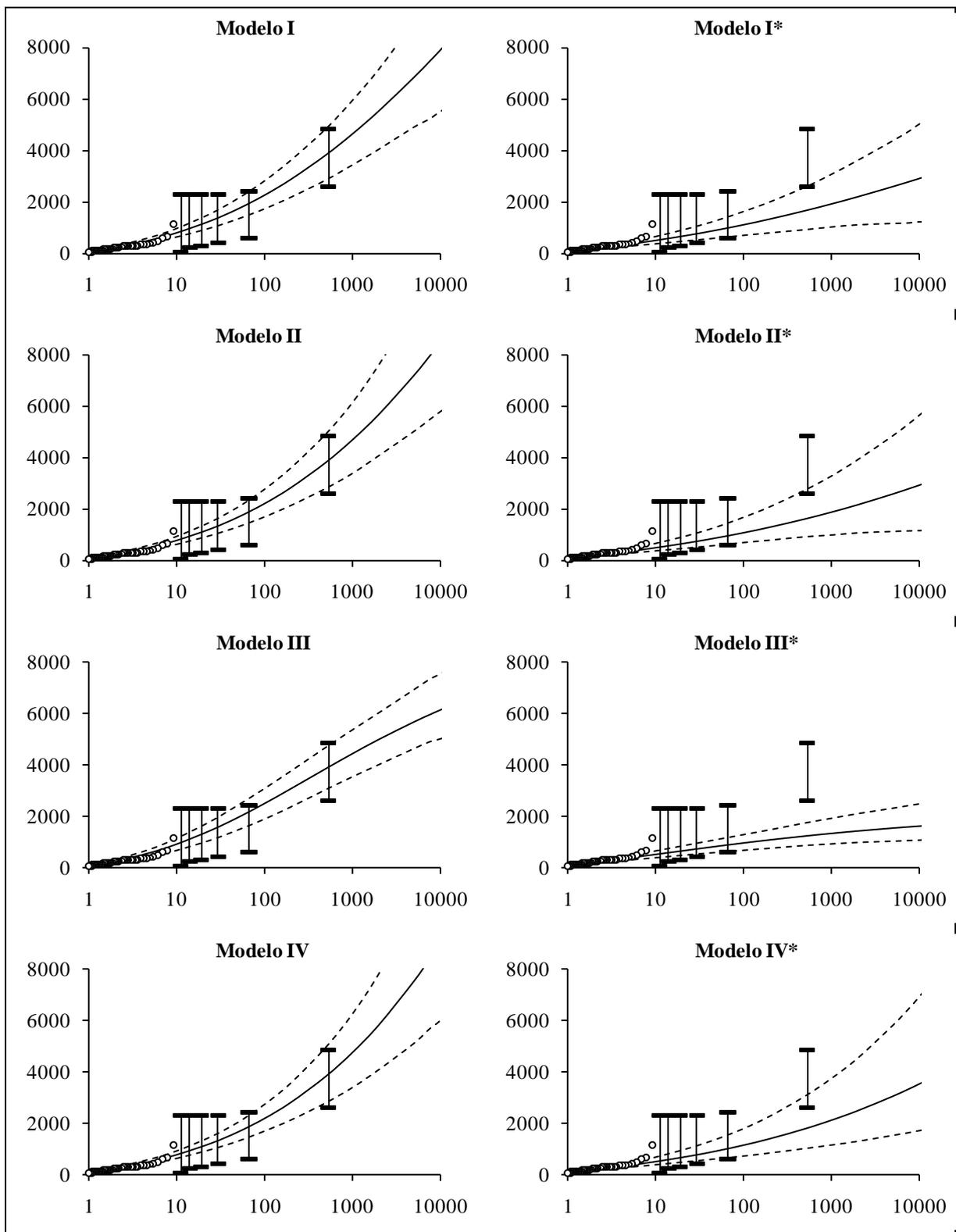


Figura 7.34 – Distribuição preditiva *a posteriori* (média) ou curva de quantis (linha contínua), intervalo de credibilidade de 95% (linha tracejada). Os círculos representam os dados sistemáticos e as barras com círculos representam os dados não sistemáticos. No eixo horizontal está o período de retorno em anos e no eixo vertical está a vazão em m³/s

De uma forma geral, todos os modelos completos apresentaram um ajuste satisfatório para todas as faixas de período de retorno. Entre esses, os modelo I e II aparentemente foram aqueles que apresentaram os melhores ajustes. A figura 7.35 mostra uma comparação entre os

quantis estimados e os observados. Observa-se que os modelos de I a IV apresentam resultados sistematicamente melhores que os modelos de I* a IV*.

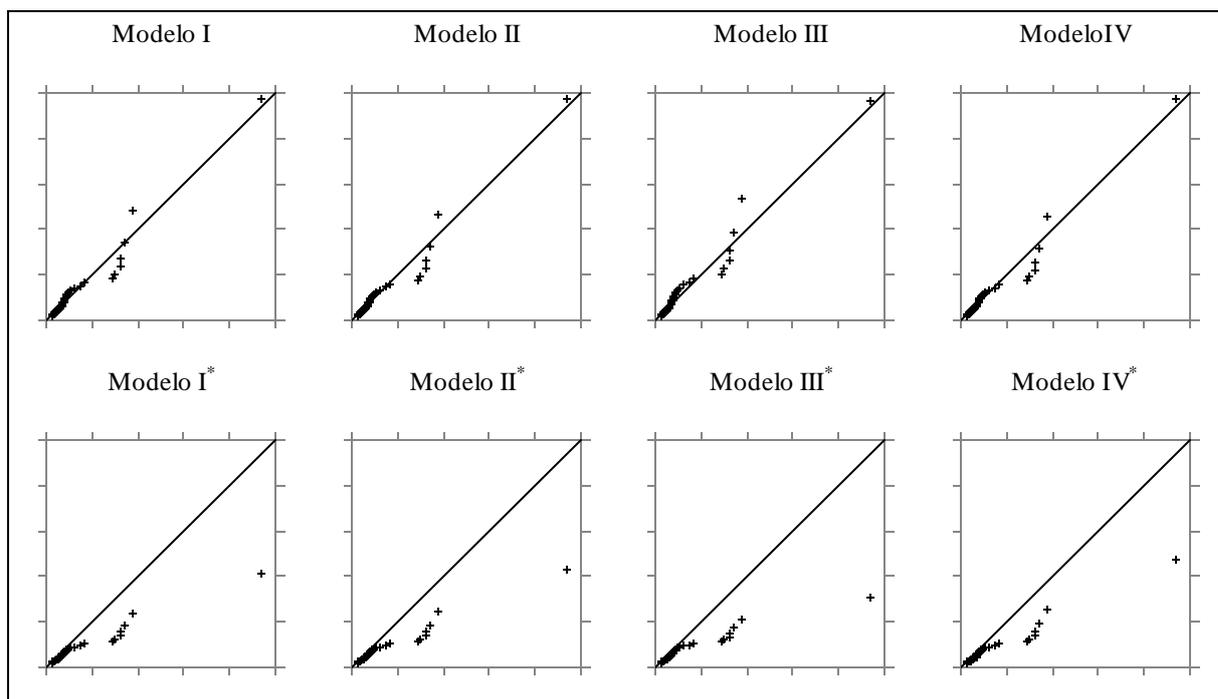


Figura 7.35 – Comparação entre os quantis estimados (eixo vertical) e observados (eixo horizontal) para o modelo LN4

Os resultados mostrados nas figuras 7.34 e 7.35 permitem verificar que os modelos completos têm uma boa capacidade descritiva. No entanto, a capacidade preditiva, possível de ser avaliada pelos modelos de I* a IV*, não é tão satisfatória quanto aquela encontrada em Folsom. Nesse ponto vale salientar algumas diferenças marcantes em relação à aplicação feita em Folsom:

- a) Em primeiro lugar, as distribuições *a priori* para o limite superior em Folsom são bem mais informativas que as distribuições em Pont Du Vilomara. Com efeito, o coeficiente de variação em Folsom abrange uma faixa de 0,3 a 0,7, enquanto em Pont Du Vilomara abrange uma faixa de 1,0 a 2,0;
- b) A estimativa pontual para o limite superior em Folsom é dada pela PMF local, que engloba as características hidroclimatológicas de forma bem mais verossímil que a estimativa em Pont Du Vilomara, dada pela curva envoltória de vazões recordes; e
- c) Os dados não sistemáticos em Folsom fornecem mais informação sobre o limite superior que os dados em Pont Du Vilomara. Em Folsom, o maior dado não sistemático está próximo à estimativa da PMF, enquanto que a estimativa para o limite superior em Pont Du Vilomara é mais que o dobro do maior dado não sistemático.

7.3.4.2 Modelo EV4

Assim como para a LN4, a partir das distribuições *a priori* para α definidas anteriormente foram construídos 4 modelos utilizando todos os dados disponíveis (sistemáticos e não sistemáticos) e 4 modelos utilizando somente os dados sistemáticos. Para cada modelo foi gerada, no WinBUGS, uma amostra de tamanho 1.200.000, sendo que, a cada 20 realizações do algoritmo, uma era selecionada para análise (*lag* = 20), ou seja, selecionou-se para análise uma amostra de tamanho 60.000. Dessa amostra de 60.000 valores, os 10.000 primeiros foram descartados como *burn in*, compondo uma amostra final de 50.000 realizações da distribuição conjunta *a posteriori* dos parâmetros da EV4.

Estatísticas a posteriori para o parâmetro ξ

A tabela 7.16 mostra os resultados para o parâmetro ξ . A figura 7.36 mostra as distribuições *a posteriori* do parâmetro ξ para os modelos completos e somente com os dados sistemáticos.

Tabela 7.16 – Estatísticas *a posteriori* para o parâmetro ξ da EV4

Conjunto de dados	Modelo	Média	CV	DP	95% HPD
Sistemáticos e não sistemáticos	I	1,34	0,082	0,110	(1,12, 1,55)
	II	1,43	0,072	0,103	(1,23, 1,61)
	III	1,23	0,070	0,086	(1,06, 1,40)
	IV	1,48	0,049	0,073	(1,34, 1,62)
Sistemáticos	I*	1,21	0,105	0,127	(0,96, 1,46)
	II*	1,22	0,103	0,125	(0,97, 1,47)
	III*	1,11	0,113	0,125	(0,87, 1,35)
	IV*	1,22	0,104	0,127	(0,97, 1,47)

95% HPD - Intervalo de Credibilidade; os modelos marcados com '*' são aqueles formulados somente com informação sistemática

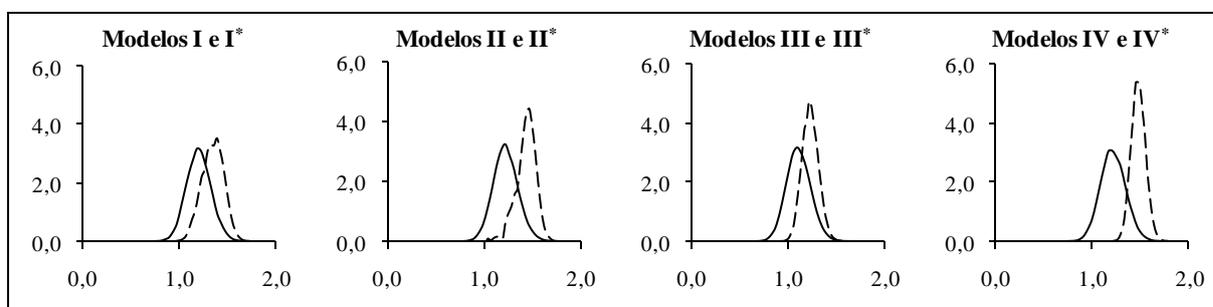


Figura 7.36 – Distribuições *a posteriori* do parâmetro ξ para os modelos completos (linha tracejada) e para os modelos com somente dados sistemáticos (linha contínua)

A tabela 7.16 e a figura 7.36 mostram que o parâmetro ξ é pouco sensível à especificação *a priori* de α . No que se refere ao tipo de dado incluído na análise, verifica-se uma redução do

intervalo de credibilidade e do coeficiente de variação com a adição dos dados não sistemáticos.

Estatísticas a posteriori para o parâmetro σ

A tabela 7.17 mostra os resultados para o parâmetro σ . A figura 7.37 mostra as distribuições *a posteriori* do parâmetro σ para os modelos completos e somente com os dados sistemáticos.

Tabela 7.17 – Estatísticas *a posteriori* para o parâmetro σ da EV4

Conjunto de dados	Modelo	Média	CV	DP	95% HPD
Sistemáticos e não sistemáticos	I	173,6	1,05	181,8	(37,1, 554,9)
	II	2.735,7	1,25	3.415,0	(36,4, 9.688,0)
	III	59,5	0,22	12,8	(39,2, 85,2)
	IV	$2,1 \times 10^6$	0,71	$1,5 \times 10^6$	$(5,3 \times 10^4, 4,9 \times 10^6)$
Sistemáticos	I*	310,7	0,83	257,0	(7,3, 810,6)
	II*	3.524,9	0,94	3.330,7	(12,8, 10.070,0)
	III*	16,6	0,59	9,9	(6,7, 36,6)
	IV*	$1,9 \times 10^6$	0,72	$1,3 \times 10^6$	$(3,1 \times 10^4, 4,5 \times 10^6)$

95% HPD - Intervalo de Credibilidade; os modelos marcados com '*' são aqueles formulados somente com informação sistemática

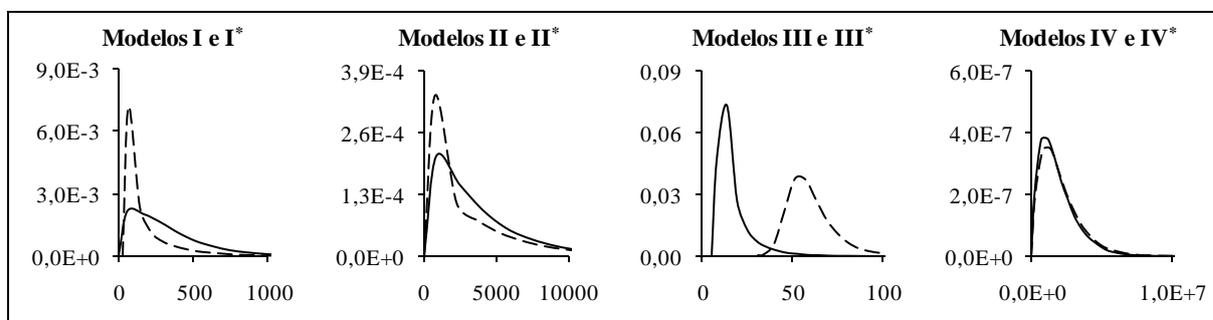


Figura 7.37 – Distribuições *a posteriori* do parâmetro σ para os modelos completos (linha tracejada) e para os modelos com somente dados sistemáticos (linha contínua)

A tabela 7.17 e a figura 7.37 mostram que o parâmetro σ é bastante sensível à especificação *a priori* de α . No que se refere ao tipo de dado incluído na análise, verifica-se uma redução do intervalo de credibilidade e do coeficiente de variação com a adição dos dados não sistemáticos. Destaca-se, entre os resultados encontrados, o modelo III, que apresentou o menor CV. Como será visto a seguir, os resultados para o parâmetro σ são semelhantes àqueles obtidos para o parâmetro α .

Estatísticas a posteriori para o parâmetro α

A tabela 7.18 mostra os resultados para o parâmetro α . A figura 7.38 mostra as distribuições *a posteriori* do parâmetro α para os modelos completos e somente com os dados sistemáticos.

Tabela 7.18 – Estatísticas *a posteriori* para o parâmetro α da EV4

Conjunto de dados	Modelo	Média	CV	DP	95% HPD
Sistemáticos e não sistemáticos	I	17.363	1,00	17.360	(5.169, 53.580)
	II	262.912	1,24	325.535	(5.147, 924.700)
	III	6.489	0,16	1.013	(5.160, 8.450)
	IV	$1,97 \times 10^8$	0,70	$1,38 \times 10^8$	$(4,6 \times 10^6, 4,7 \times 10^8)$
Sistemáticos	I*	32.732	0,81	26.413	(1.149, 84.170)
	II*	369.997	0,93	344.106	(1.395, 1.046.000)
	III*	1.903	0,52	997	(1.148, 3.888)
	IV*	$1,98 \times 10^8$	0,70	$1,38 \times 10^8$	$(4,2 \times 10^6, 4,7 \times 10^8)$

95% HPD - Intervalo de Credibilidade; os modelos marcados com '*' são aqueles formulados somente com informação sistemática.

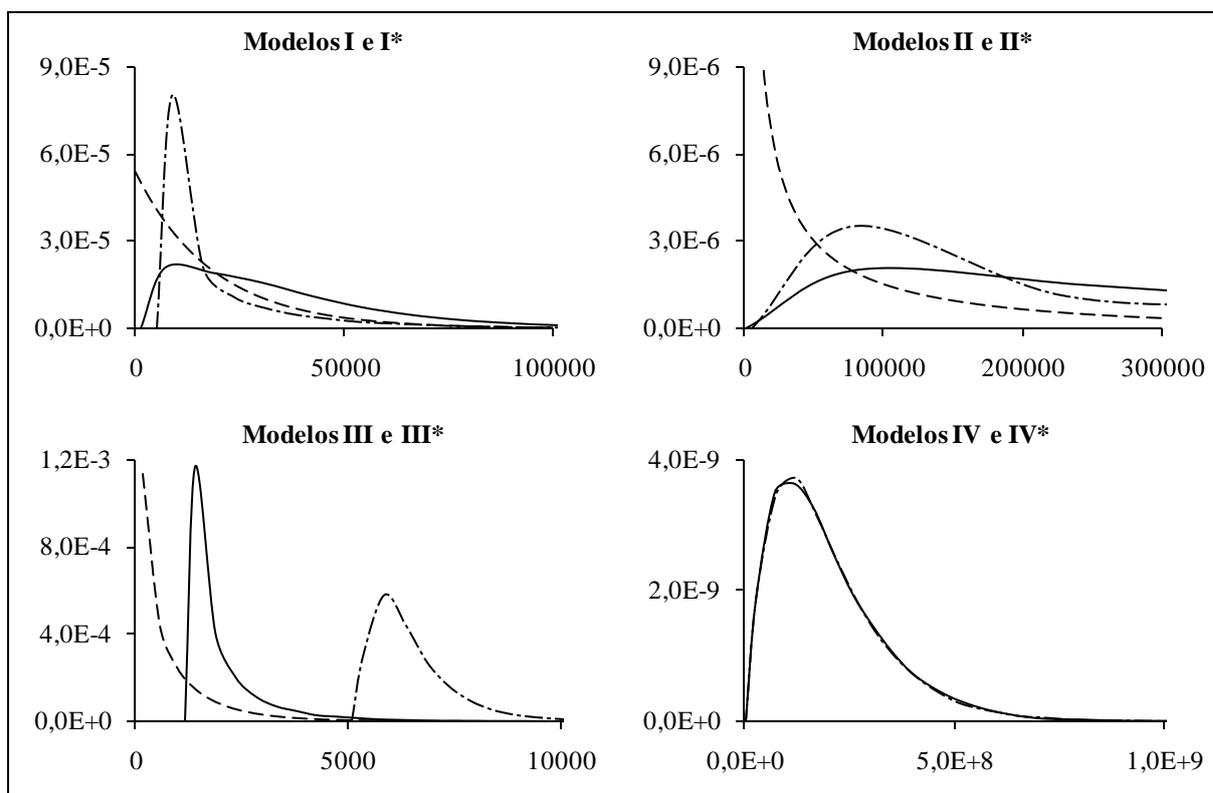


Figura 7.38 – Distribuições *a priori* (linha tracejada) e *a posteriori* do parâmetro α para os modelos completos (linha tracejada-pontilhada) e para os modelos com somente dados sistemáticos (linha contínua)

As estatísticas *a posteriori* para o limite superior se mostraram bastantes discrepantes quando comparadas entre os quatro modelos. A inclusão de informação não sistemática, ao contrário

do esperado, não reduziu o coeficiente de variação. No modelo I, por exemplo, não houve qualquer redução. O único modelo que se comportou como esperado foi o III, onde a inclusão da informação não sistemática proporcionou uma redução de cerca de 90% no CV.

O modelo II, apesar da redução do coeficiente de variação, resultou em estatísticas *a posteriori* para α pouco plausíveis para a realidade física da bacia. Desta forma, analisando somente os resultados para o limite superior, esse modelo deve ser descartado. O mesmo ocorreu para o modelo IV; no entanto, isso já era esperado.

Outro fato a ser notado em relação ao limite superior se refere aos resultados para os modelos com somente dados sistemáticos. Nesses casos, verificou-se uma sensível redução do coeficiente de variação, embora os intervalos de credibilidade, com exceção do modelo III*, tenham comprimentos maiores que os análogos com todos os dados.

Curva de quantis

A curva de quantis ou a distribuição preditiva *a posteriori* foi estimada por integração de Monte Carlo (vide item 5.3), a partir dos valores da distribuição *a posteriori* dos parâmetros da EV4. A figura 7.39 mostra os resultados para cada um dos modelos estudados. No anexo 4, são mostrados os resultados numéricos para cada período de retorno e para cada modelo. Os dados observados, sistemáticos e não sistemáticos, foram plotados de acordo com a probabilidade de excedência calculada a partir do método descrito no item 6.4. No caso de Pont Du Vilomara, foi feita uma simplificação, uma vez que existem dados do tipo LB. Para esse tipo de dado não é possível estabelecer a probabilidade de excedência empírica e, assim, os mesmos não foram incluídos na curva de quantis.

De uma forma geral, os modelos completos apresentaram um ajuste bem melhor que os modelos com somente dados sistemáticos. Além disso, verifica-se que o enorme intervalo de credibilidade para quantis acima de 100 anos de período de retorno nos modelos I* a IV* é um fator determinante para a não utilização dos mesmos.

Dentre todos os modelos analisados, aquele que apresentou os melhores resultados foi o III. Para esse modelo, tanto o ajuste quanto o intervalo de credibilidade são bastante razoáveis. Para os demais modelos completos, o intervalo de credibilidade também é um fator que compromete a inferência de quantis com altos períodos de retorno.

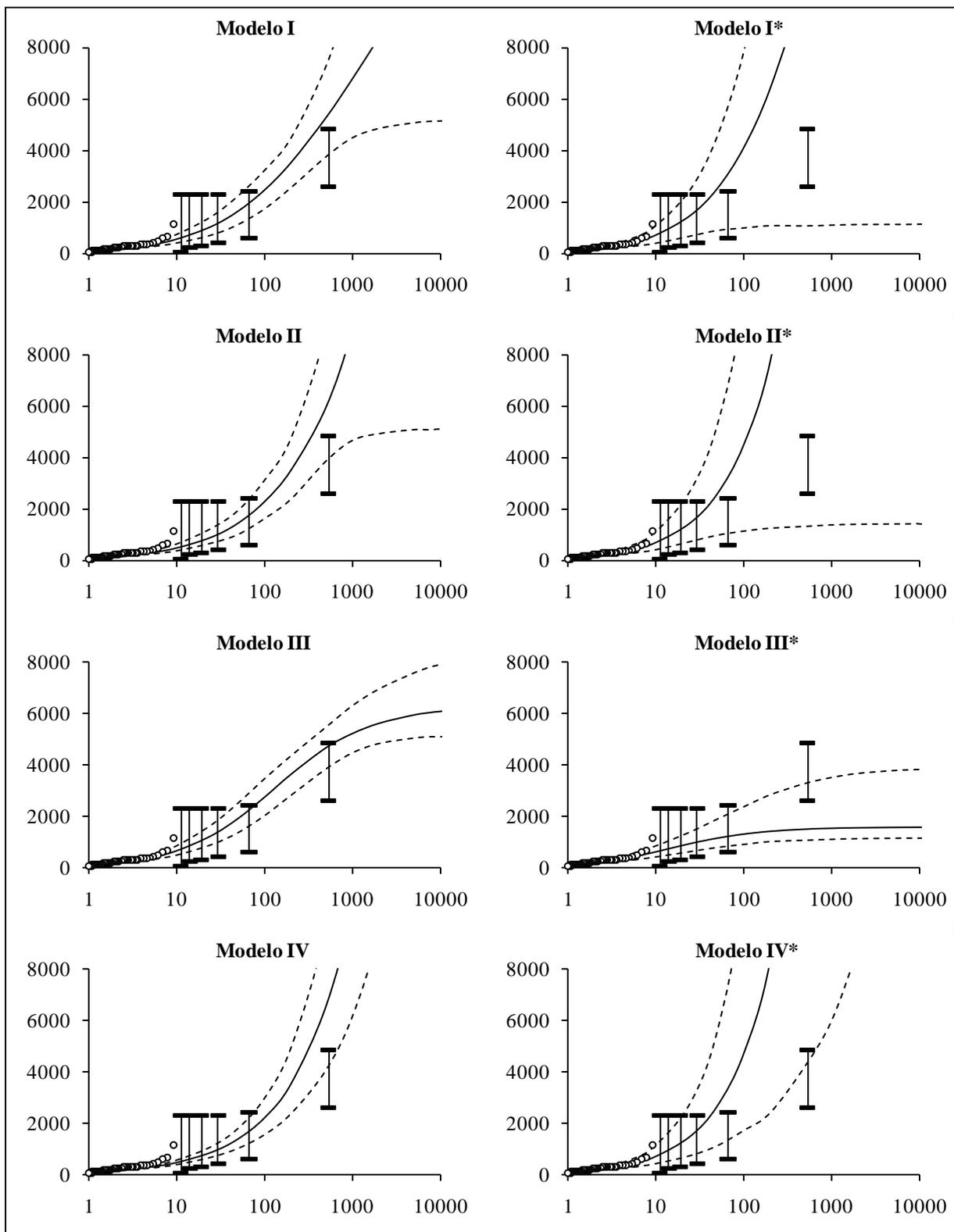


Figura 7.39 – Distribuição preditiva *a posteriori* (média) ou curva de quantis (linha contínua), intervalo de credibilidade de 95% (linha tracejada). Os círculos representam os dados sistemáticos e as barras com círculos representam os dados não sistemáticos. No eixo horizontal está o período de retorno em anos e no eixo vertical está a vazão em m³/s

A figura 7.40 mostra uma comparação entre os quantis estimados e os observados. É possível constatar que os modelos de I a IV apresentam resultados sistematicamente melhores que os modelos de I* a IV*.

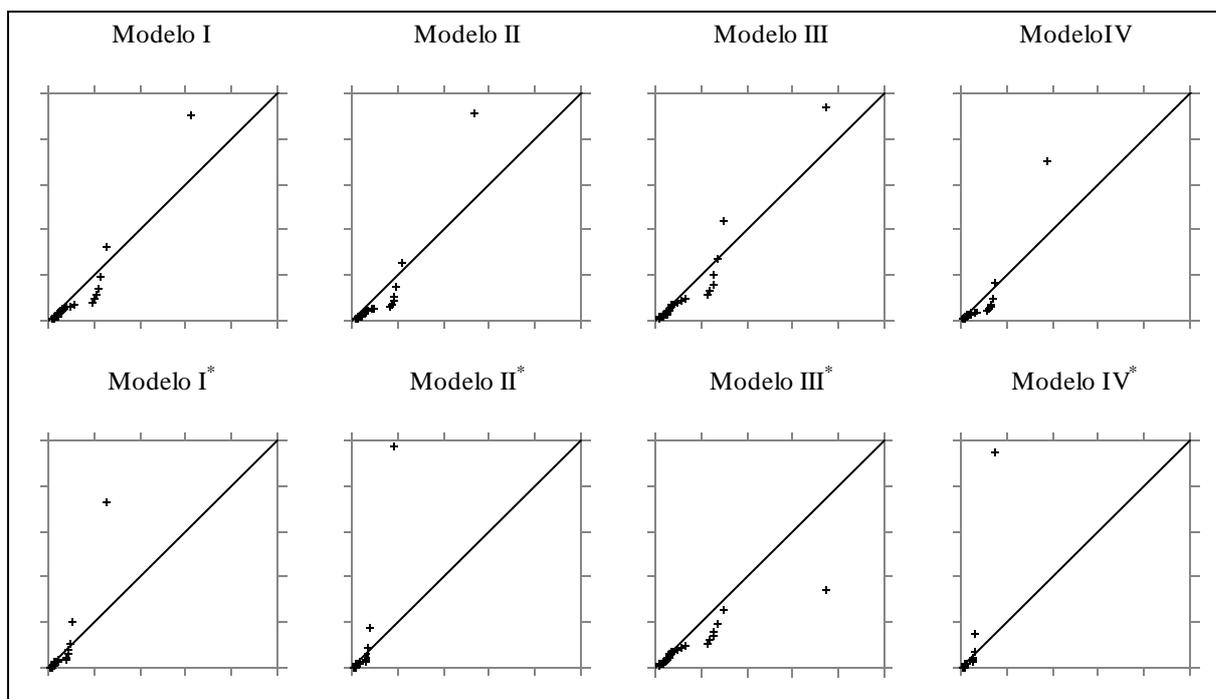


Figura 7.35 – Comparação entre os quantis estimados (eixo vertical) e observados (eixo horizontal) para o modelo EV4

Os resultados mostrados nas figuras 7.34 e 7.35 permitem verificar que os modelos completos têm uma boa capacidade descritiva. No entanto, a capacidade preditiva, passível de ser avaliada pelos modelos de I* a IV*, da mesma forma que para a distribuição LN4, não é tão satisfatória quanto àquela encontrada em Folsom.

7.3.5 Comparação com os resultados da análise de frequência clássica

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 as distribuições LN4 e EV4 aos dados de Pont Du Vilomara a partir dos métodos discutidos em 4.4.2. De acordo com Botero (2006), os métodos que produziram os melhores resultados foram os da máxima verossimilhança (ML) e o da atribuição de um valor extremo ao parâmetro α (PMF). A tabela 7.19 mostra uma comparação dos resultados obtidos para o limite superior de acordo com cada método utilizado. A estimativa pelo método bayesiano com o uso da LN4 é a média *a posteriori* do modelo I (obtida da tabela 7.15), enquanto que com o uso da EV4 é a média *a posteriori* do modelo III (obtida da tabela 7.18).

Tabela 7.19 – Estimativas para o limite superior em Pont Du Vilomara por diferentes métodos

Método	LN4	EV4
ML	34.627	4.850
PMF	12.887	12.887
bayesiano	38.116	6.489

A tabela 7.19 mostra que os resultados obtidos pelo método desenvolvido nesta tese são comparáveis àqueles obtidos pelo método da máxima verossimilhança. A capacidade descritiva de cada método pode ser avaliada pela curva de quantis. A figura 7.36 mostra a comparação para a LN4 e a figura 7.37 mostra os resultados para a EV4.

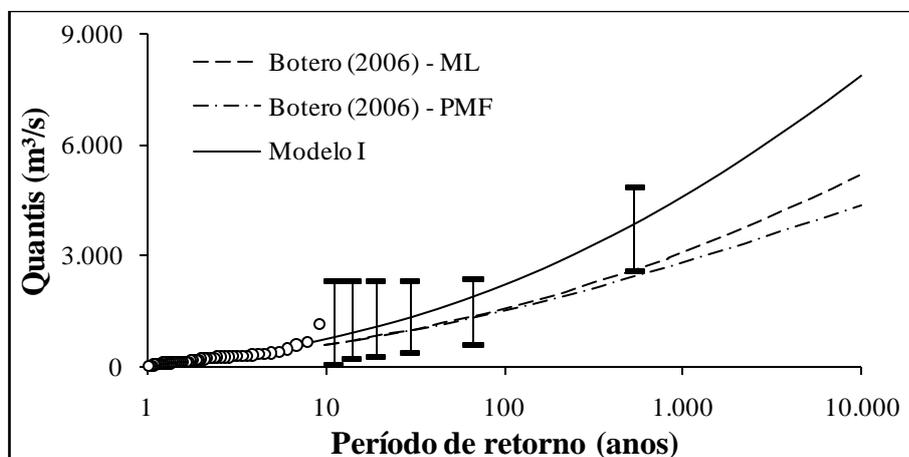


Figura 7.36 – Comparação entre os métodos bayesiano e clássico para a LN4

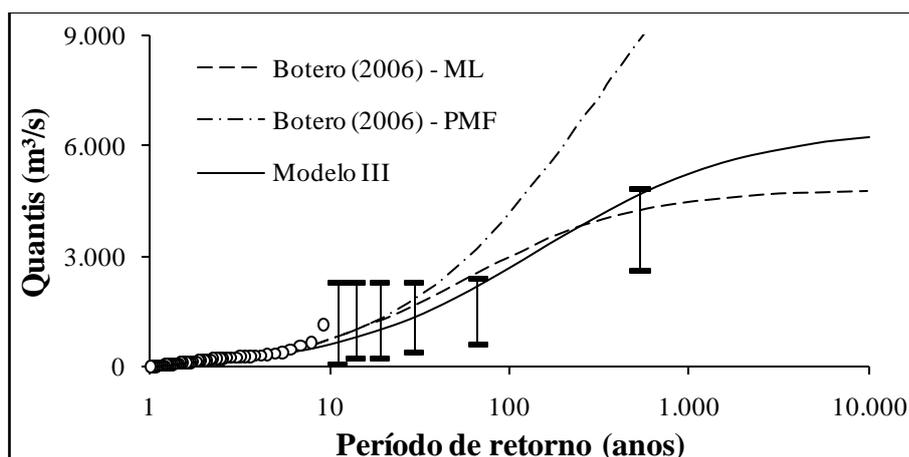


Figura 7.37 – Comparação entre os métodos bayesiano e clássico para a EV4

Para a LN4, o método bayesiano, representado pelo modelo I na figura 7.36, apresentou um ajuste bem superior aos métodos clássicos. Esses últimos não conseguiram se ajustar ao maior dado não sistemático, fornecendo uma estimativa menor que o limite inferior do último dado DB.

Para a EV4, o método da máxima verossimilhança apresentou resultados comparáveis ao método bayesiano, sobretudo para quantis com períodos de retorno inferiores a 1.000 anos. Ressalta-se que a fixação de um valor extremo para o limite superior no caso da EV4 tem um agravante na aplicação feita para essa bacia espanhola. Conforme visto anteriormente, a EV4 se aproxima rapidamente do limite superior. Como em Pont Du Vilomara a estimativa para o

limite superior é bem maior que os dados observados, verifica-se a perda de ajuste para os quantis com período de retorno superior a 50 anos, tal como ilustrada na figura 7.37.

A comparação dos métodos clássicos com o bayesiano apenas pela curva de quantis, como feita aqui, não mostra as vantagens de um sobre o outro. Deve ser considerado, nesse 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.

7.4 Aplicação para a bacia do rio Pará, em Ponte do Vilela

7.4.1 A bacia do rio Pará

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 possibilitará 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, será possível comparar os resultados do método proposto com os resultados obtidos por outros métodos.

O rio Pará se localiza entre as longitudes 44° e 45° e latitudes 20° e 21° na bacia do alto rio São Francisco, no estado de Minas Gerais. A nascente é na Serra da Cebola, a uma altitude de 1.160 m, no município de Resende Costa, no estado de Minas Gerais. Seus principais afluentes são os rios Itapecerica, Lambari e Picão, pela margem esquerda, e o rio São João, pela margem direita. O rio Pará percorre uma extensão de cerca de 310 km até desaguar no rio São Francisco. A figura 7.38 mostra a localização da bacia, enquanto a figura 7.39 ilustra alguns trechos do rio.

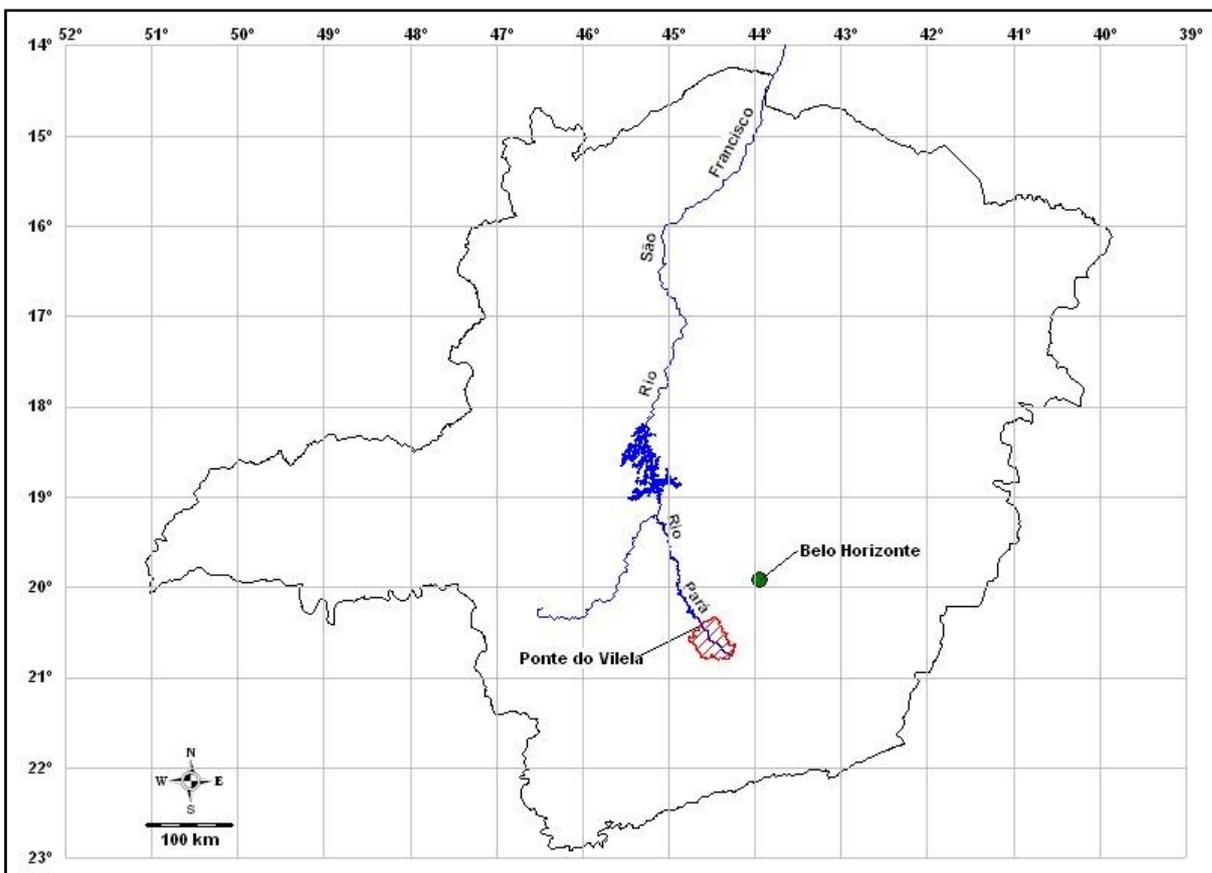


Figura 7.38 – Localização da bacia do rio Pará



Figura 7.39 – Trechos do rio Pará (Fonte: www.cbhpara.org.br)

7.4.2 Dados hidrológicos sistemáticos

A bacia do rio Pará no ponto de interesse dessa aplicação é monitorada pela Companhia Energética de Minas Gerais – CEMIG. A estação fluviométrica utilizada foi a de Ponte do Vilela, a qual possui uma área de drenagem de 1.620 km², abrangendo quase a totalidade da área a montante da represa de Cajuru. O rio Pará, até esse ponto, tem um comprimento de 73 km e uma declividade média de 0,30%. Essa estação possui dados diários de vazão de julho de 1938 a dezembro de 2000, dispondo, portanto, de 63 anos de observações. As principais características para a série de vazão são:

Mínimo:	80 m ³ /s
Máximo:	620 m ³ /s
Média:	196 m ³ /s
Desvio padrão:	108 m ³ /s
Coefficiente de variação:	0,55
Coefficiente de assimetria:	1,78

A figura 7.40 mostra os dados sistemáticos considerados na análise.

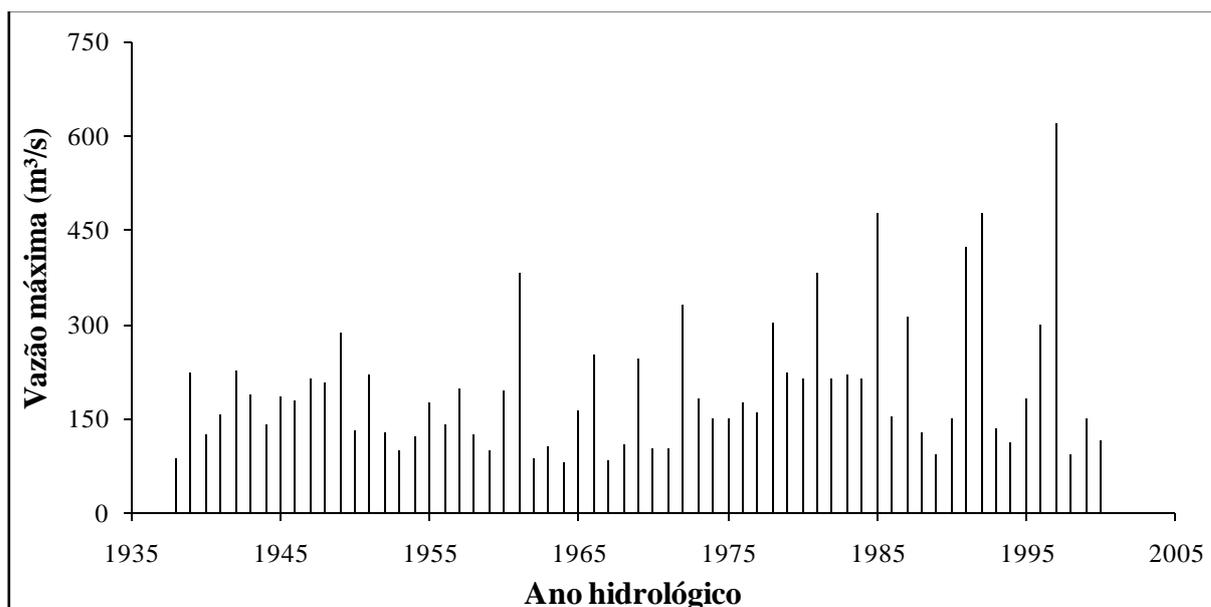


Figura 7.40 – Dados sistemáticos do rio Pará em Ponte do Vilela

7.4.3 Distribuição *a priori* para o limite superior em Ponte do Vilela

Uma vez que não há estimativa de PMF para Ponte do Vilela, a mesma foi estimada pela curva envoltória para o sudeste brasileiro, dada pela equação 6.66. Com isso, tem-se uma PMF de 2.988 m³/s. Seguindo o procedimento A, com $p = 0,5$ e $CV = 0,6$, tem-se a seguinte distribuição *a priori* para o limite superior:

$$D) \quad \alpha \sim \text{GAM}(2,78, 8,54 \times 10^{-4})$$

Seguindo o procedimento B, as 34 estimativas de PMF compiladas pelo CBDB (2002) foram transpostas para a bacia do rio Pará, em Ponte do Vilela. A transposição dessas tormentas foi feita pela curva envoltória do sudeste.

A figura 7.41 mostra o histograma das PMF transpostas para Ponte do Vilela, com o ajuste da distribuição Gama. A forma desse histograma tem características distributivas bastante semelhantes à distribuição Normal, não ficando evidente a assimetria positiva como nos casos

anteriores (rio American e rio Llobregat). No entanto, uma vez que o limite superior é sempre positivo, definiu-se a distribuição Gama como especificação para o limite superior. Os parâmetros da distribuição Gama foram estimados pelo método dos momentos convencionais. A distribuição resultante tem a seguinte forma:

$$\text{II) } \alpha \sim \text{GAM}(19,9064, 1,18 \times 10^{-2})$$

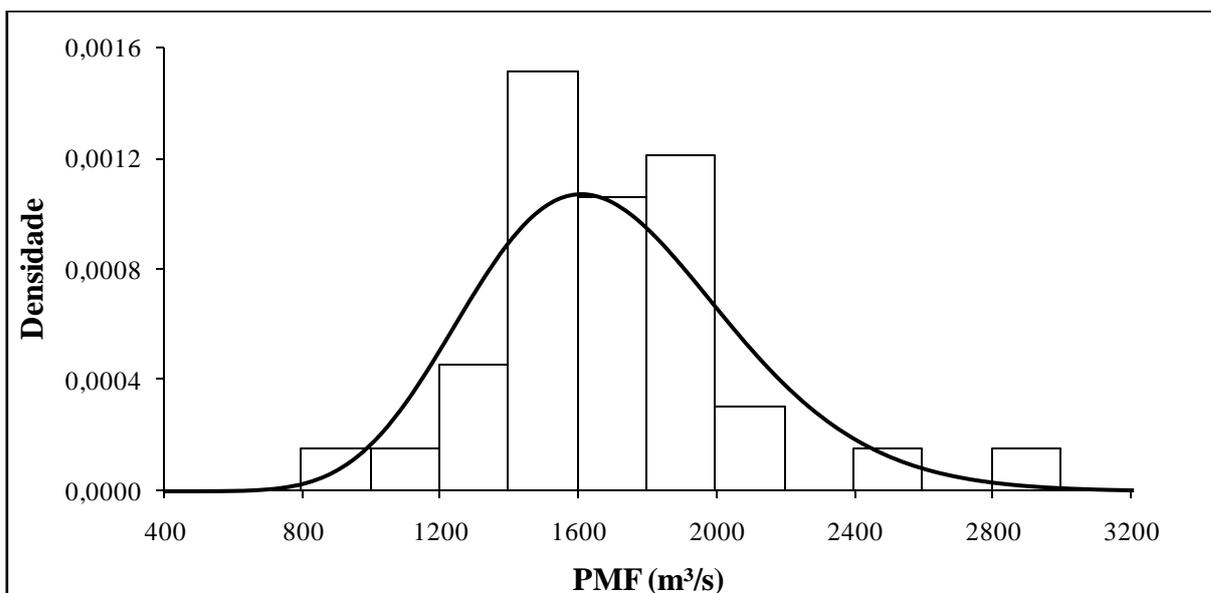


Figura 7.41 – Histograma das PMF's transpostas para a bacia do rio Pará e ajuste da distribuição Gama

Adicionalmente às distribuições *a priori* definidas a partir dos procedimentos A e B, foi considerada, também, uma distribuição não informativa para o limite superior, tal como para as demais aplicações.

A tabela 7.20 resume as principais características das distribuições usadas como especificação *a priori* para o limite superior. A figura 7.42 mostra suas respectivas formas.

Tabela 7.20 – Parâmetros e características das distribuições *a priori* para o limite superior na bacia do rio Pará, em Ponte do Vilela

Distribuição <i>a priori</i>	ρ_α	β_α	Média	Mediana	CV_α	DP	Proc.
I	2,78	$8,54 \times 10^{-4}$	3.255	2.988	0,6	1.953	A
II	19,91	$1,18 \times 10^{-2}$	1.693	1.664	0,2	379	B
III	1,00	$1,00 \times 10^{-8}$	$1,00 \times 10^8$	$\sim 6,9 \times 10^7$	1,0	$1,00 \times 10^{-8}$	-

CV: coeficiente de variação; DP: desvio padrão; Proc.: procedimento de cálculo

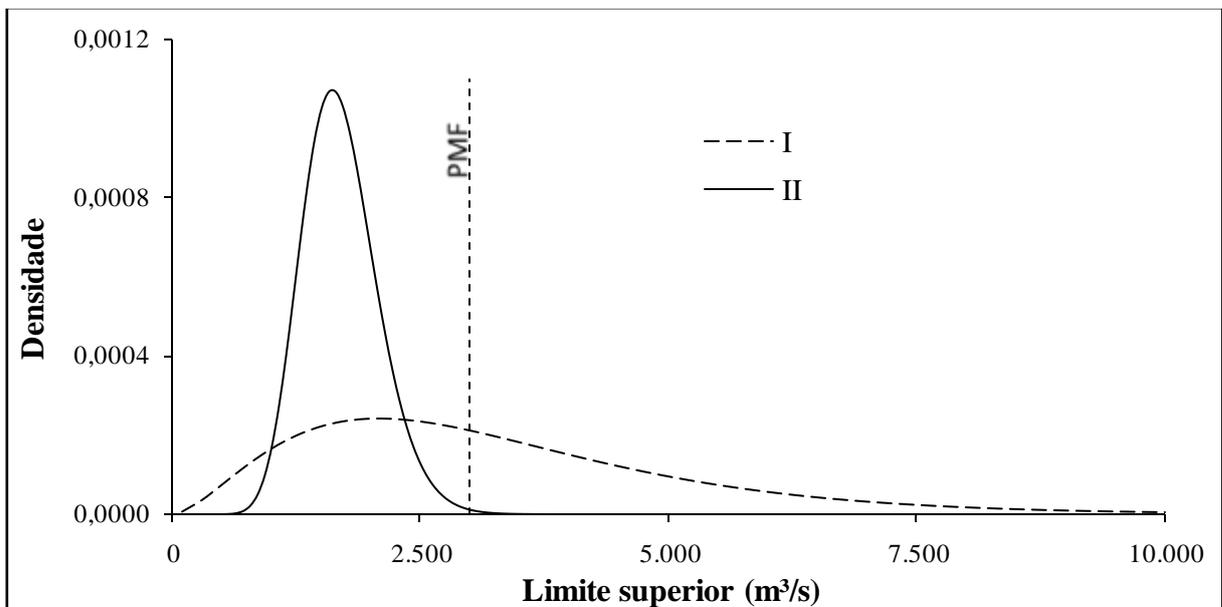


Figura 7.42 – Distribuições *a priori* de I e II para o limite superior na bacia do rio Pará

7.4.4 Estatísticas *a posteriori*

As conclusões sobre as estatísticas *a posteriori* para Ponte do Vilela são bastante semelhantes às feitas para os demais casos. Assim, os resultados para essa bacia serão apresentados de forma simplificada.

7.4.4.1 Modelo LN4

A partir das distribuições *a priori* para α definidas anteriormente, foram construídos três modelos utilizando todos os dados disponíveis, os quais, nesse caso, são somente sistemáticos. Para cada modelo foi gerada, no WinBUGS, uma amostra de tamanho 1.200.000, sendo que, a cada 20 realizações do algoritmo, uma era selecionada para análise ($lag = 20$), ou seja, selecionou-se para análise uma amostra de tamanho 60.000. Dessa amostra de 60.000 valores, os 10.000 primeiros foram descartados como *burn in*, compondo uma amostra final de 50.000 realizações da distribuição conjunta *a posteriori* dos parâmetros da LN4.

Estatísticas a posteriori para os parâmetros da LN4

A tabela 7.21 mostra os resultados para os parâmetros da LN4. A figura 7.43 mostra as distribuições *a posteriori* para cada um dos parâmetros.

Tabela 7.21 – Estatísticas *a posteriori* para os parâmetros da LN4

Parâmetro	Modelo	Média	CV	DP	95% HPD
α	I	3.922	0,48	1.902	(1.065, 7.715)
	II	1.806	0,21	372	(1.120, 2.544)
	III	$1,00 \times 10^8$	1,00	$1,00 \times 10^8$	$(1,3 \times 10^5, 3,0 \times 10^8)$
μ	I	-2,94	-0,18	0,52	(-3,89, -1,89)
	II	-2,20	-0,11	0,24	(-2,66, -1,71)
	III	-12,70	-0,10	1,26	(-14,88, -10,15)
σ	I	0,52	0,10	0,05	(0,43, 0,63)
	II	0,56	0,10	0,06	(0,46, 0,67)
	III	0,49	0,09	0,04	(0,40, 0,57)

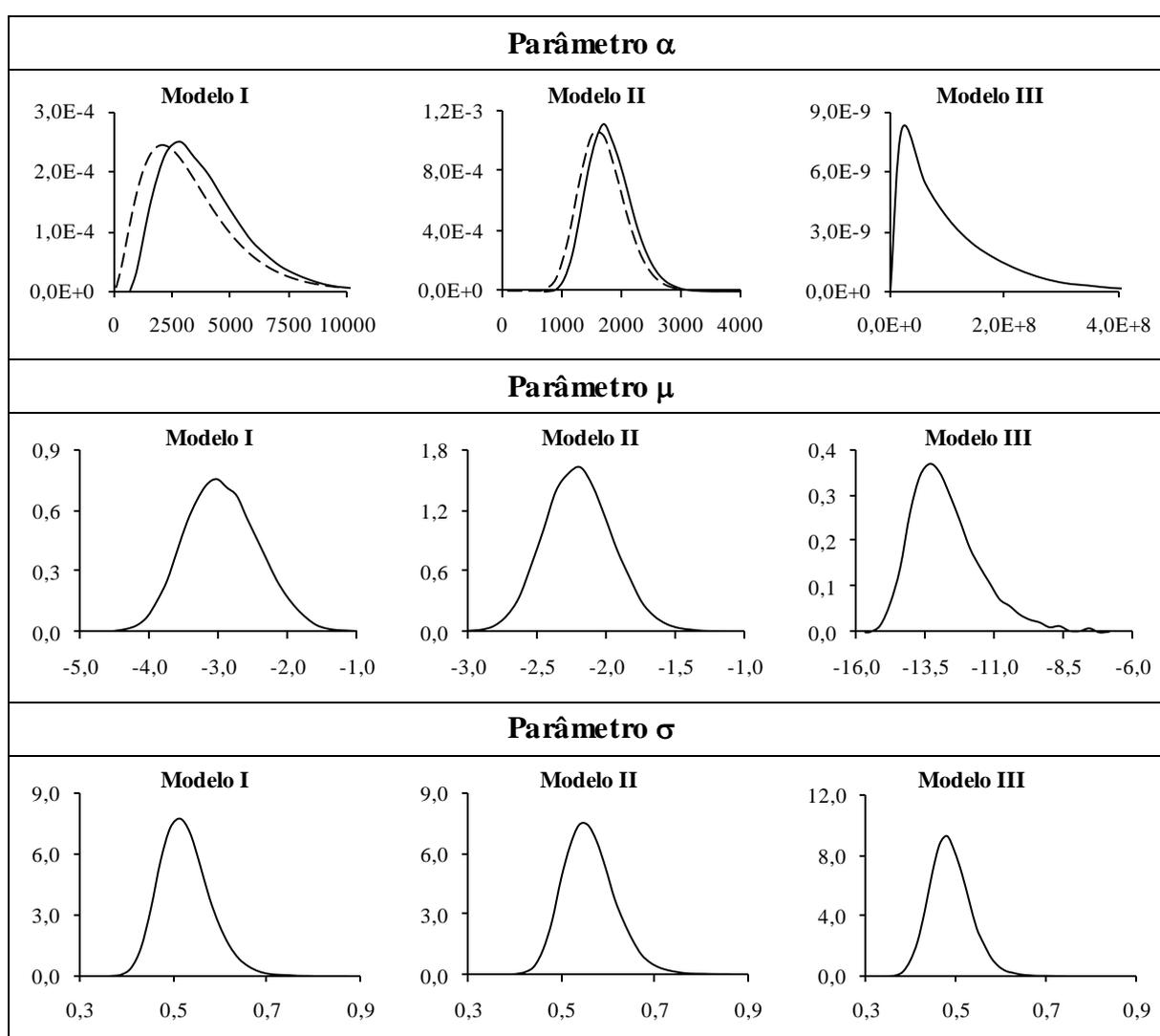


Figura 7.43 – Distribuições *a posteriori* para os parâmetros da LN4; a linha tracejada representa a distribuição *a priori*

Os parâmetros μ e σ têm distribuições *a posteriori* bastante semelhantes para cada modelo. Isso mostra, de forma parcial, que a especificação *a priori* para α não influencia esses parâmetros.

Os resultados *a posteriori* para o limite superior foram bastante semelhantes àqueles obtidos para os modelos com somente dados sistemáticos na aplicação em Folsom, ou seja, os dados observados não fornecem informações importantes para o limite superior. Isso é refletido na figura 7.43 pela proximidade entre as distribuições *a priori* e *a posteriori* para α .

Curva de quantis

A curva de quantis, ou a distribuição preditiva *a posteriori*, foi estimada por integração de Monte Carlo (vide item 5.3) a partir dos valores da distribuição *a posteriori* dos parâmetros da LN4. A figura 7.44 mostra os resultados para cada um dos modelos estudados. No anexo 5 são mostrados os resultados numéricos para cada período de retorno e para cada modelo. Os dados observados foram plotados de acordo com a probabilidade de excedência calculada a partir da equação de Gringorten descrita no item 6.4.

De uma forma geral, os modelos apresentaram um ajuste satisfatório aos dados observados. Dentre os modelos analisados, aquele que apresentou os melhores resultados foi o III. No entanto, o limite superior para esse modelo é pouco plausível para a realidade física da bacia, sendo, portanto, descartado. O fato de o modelo III ter fornecido os melhores resultados pode ser um indicativo de que os dados de Ponte do Vilela, por si só, não apresentam características de variáveis limitadas. Na prática, fixar um limite superior em $1,00 \times 10^8$, para dados com as magnitudes de Ponte do Vilela, pode ser visto como um expediente numérico para eliminar o efeito desse na análise.

Dentre os modelos I e II, o segundo foi o que aparentemente mostrou o melhor ajuste. Com efeito, a aderência dos dois modelos aos dados observados é bastante semelhante, sendo que o segundo apresentou o menor intervalo de credibilidade. Por outro lado, o modelo III, apesar da boa aderência, foi o modelo que apresentou o maior intervalo de credibilidade.

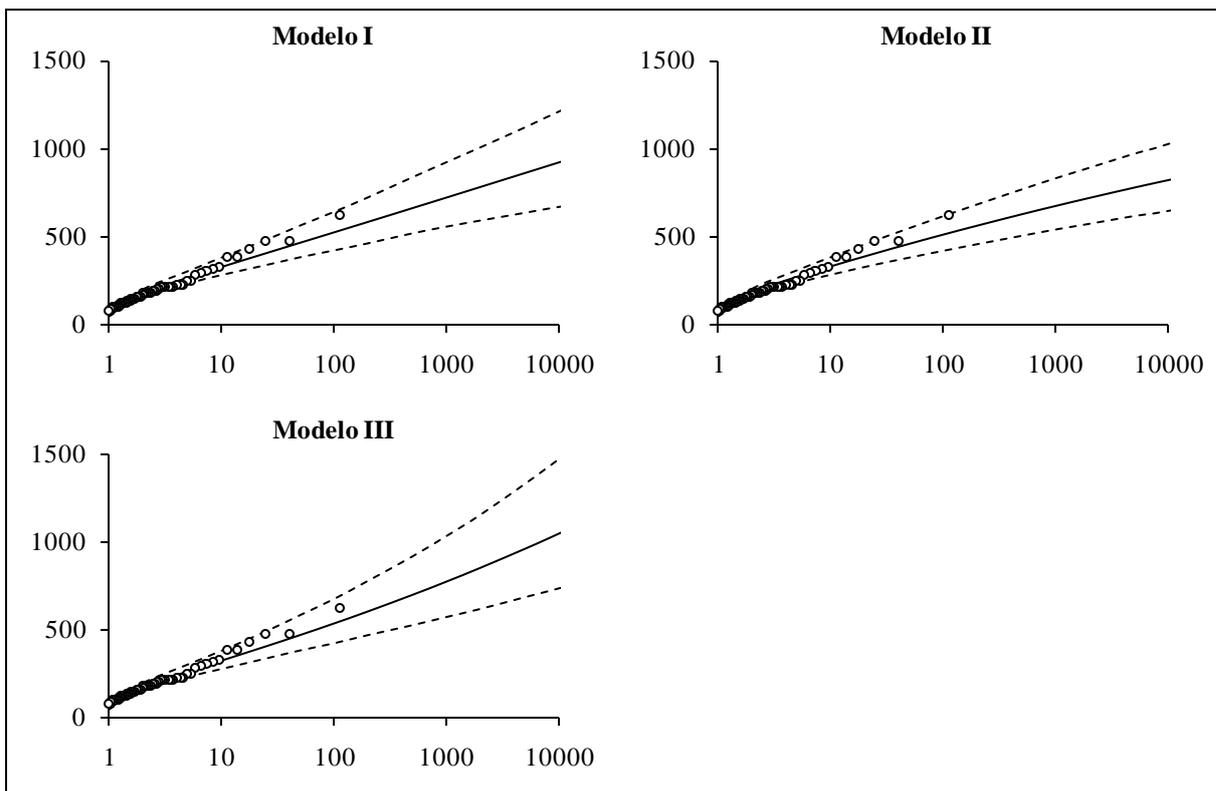


Figura 7.44 – Distribuição preditiva *a posteriori* (média) ou curva de quantis (linha contínua), intervalo de credibilidade de 95% (linha tracejada). Os círculos representam os dados sistemáticos e as barras com círculos representam os dados não sistemáticos. No eixo horizontal está o período de retorno em anos e no eixo vertical está a vazão em m³/s

A figura 7.45 mostra uma comparação entre os quantis estimados e os observados. Observa-se que todos os modelos apresentaram um ajuste equivalente.

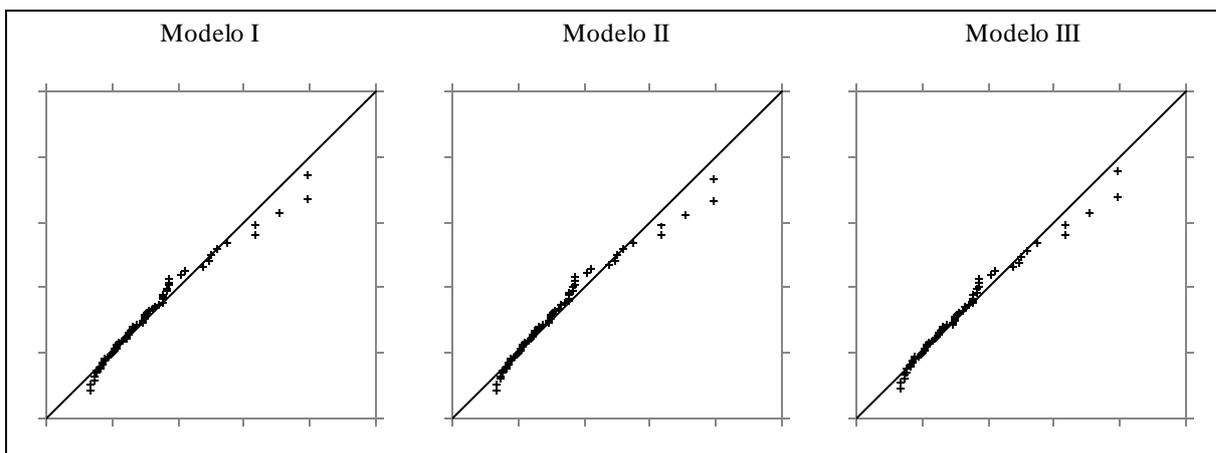


Figura 7.45 – Comparação entre os quantis estimados (eixo vertical) e observados (eixo horizontal) para o modelo LN4

7.4.4.2 Modelo EV4

Para cada modelo foi gerada, no WinBUGS, uma amostra de tamanho 1.200.000, com *lag* = 20 e *burn in* igual a 10.000. Assim, obteve-se uma amostra final de 50.000 realizações da distribuição conjunta *a posteriori* dos parâmetros da EV4.

Estadísticas a posteriori para os parâmetros da EV4

A tabela 7.22 mostra os resultados para os parâmetros da EV4. A figura 7.46 mostra as distribuições *a posteriori* para cada um dos parâmetros.

Tabela 7.22 – Estadísticas *a posteriori* para os parâmetros da EV4

Parâmetro	Modelo	Média	CV	DP	95% HPD
α	I	4.193	0,53	2.238	(676, 8.476)
	II	1.748	0,22	387	(1.023, 2.514)
	III	$2,0 \times 10^8$	0,70	$1,4 \times 10^8$	$(5,8 \times 10^6, 4,8 \times 10^8)$
ξ	I	2,38	0,11	0,25	(1,88, 2,88)
	II	2,25	0,10	0,23	(1,82, 2,73)
	III	2,52	0,10	0,25	(2,04, 3,01)
σ	I	29,26	0,56	16,32	(4,23, 61,04)
	II	11,55	0,25	2,89	(6,11, 17,23)
	III	$1,4 \times 10^6$	0,71	$1,0 \times 10^6$	$(3,3 \times 10^4, 3,4 \times 10^6)$

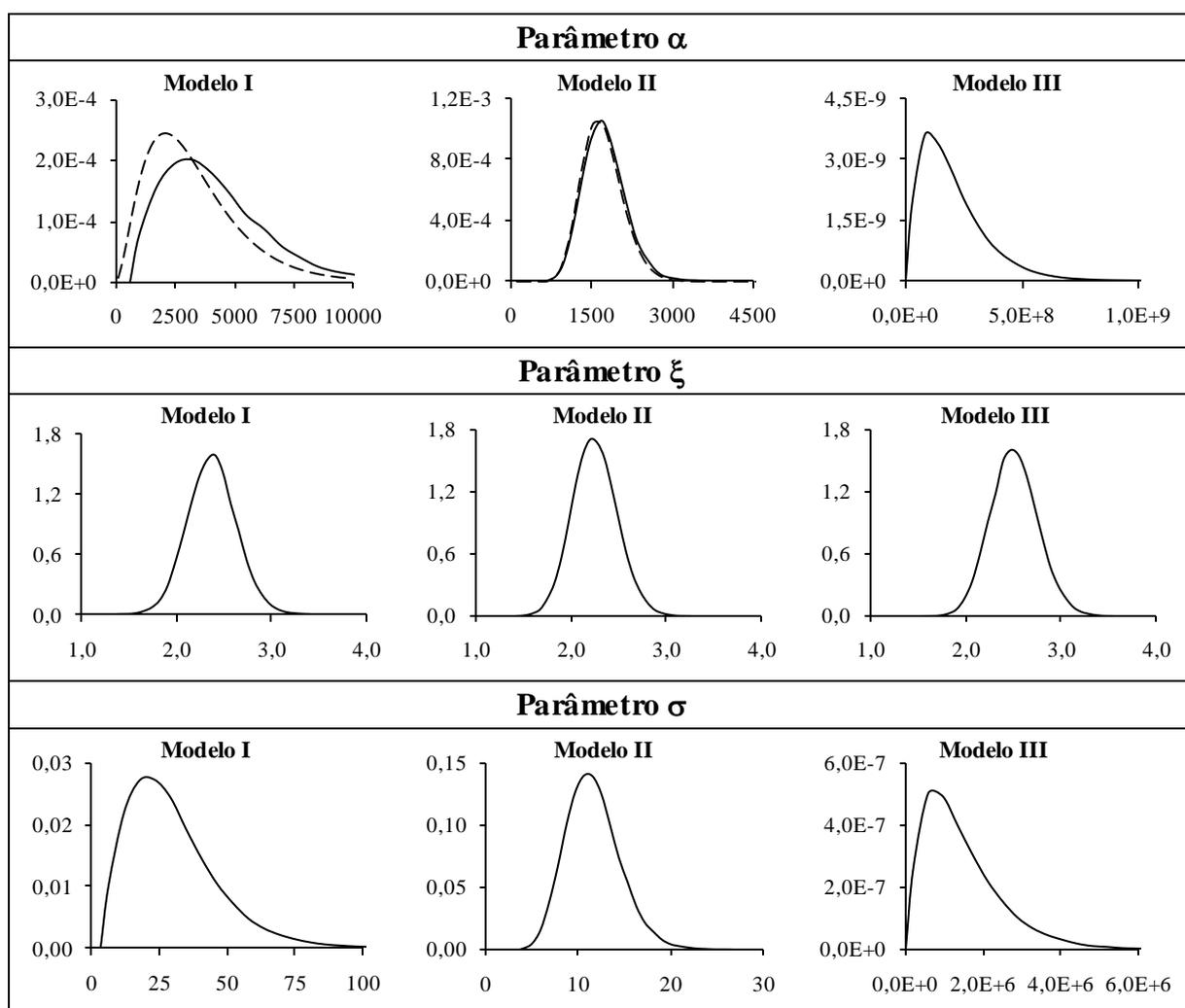


Figura 7.46 – Distribuições *a posteriori* para os parâmetros da EV4; a linha tracejada representa a distribuição *a priori*

As mesmas conclusões para a LN4 são válidas aqui. A exceção ficou por conta da distribuição *a posteriori* para o limite superior no modelo I, a qual não foi tão semelhante à distribuição *a priori*, como no caso da LN4.

Curva de quantis

A curva de quantis ou a distribuição preditiva *a posteriori* foi estimada por integração de Monte Carlo (vide item 5.3) a partir dos valores da distribuição *a posteriori* dos parâmetros da EV4. A figura 7.47 mostra os resultados para cada um dos modelos estudados. No anexo 6 são mostrados os resultados numéricos para cada período de retorno e para cada modelo. Os dados observados foram plotados de acordo com a probabilidade de excedência calculada a partir da equação de Gringorten descrita no item 6.4.

De uma forma geral, os modelos apresentaram um ajuste satisfatório aos dados observados. Dentre os modelos analisados, aquele que apresentou os melhores resultados foi o II. No caso de Ponte do Vilela, a EV4 deu resultados melhores que a LN4. Além disso, para a EV4, o modelo não informativo, como esperado, resultou em um ajuste pior que os demais modelos, diferentemente da LN4.

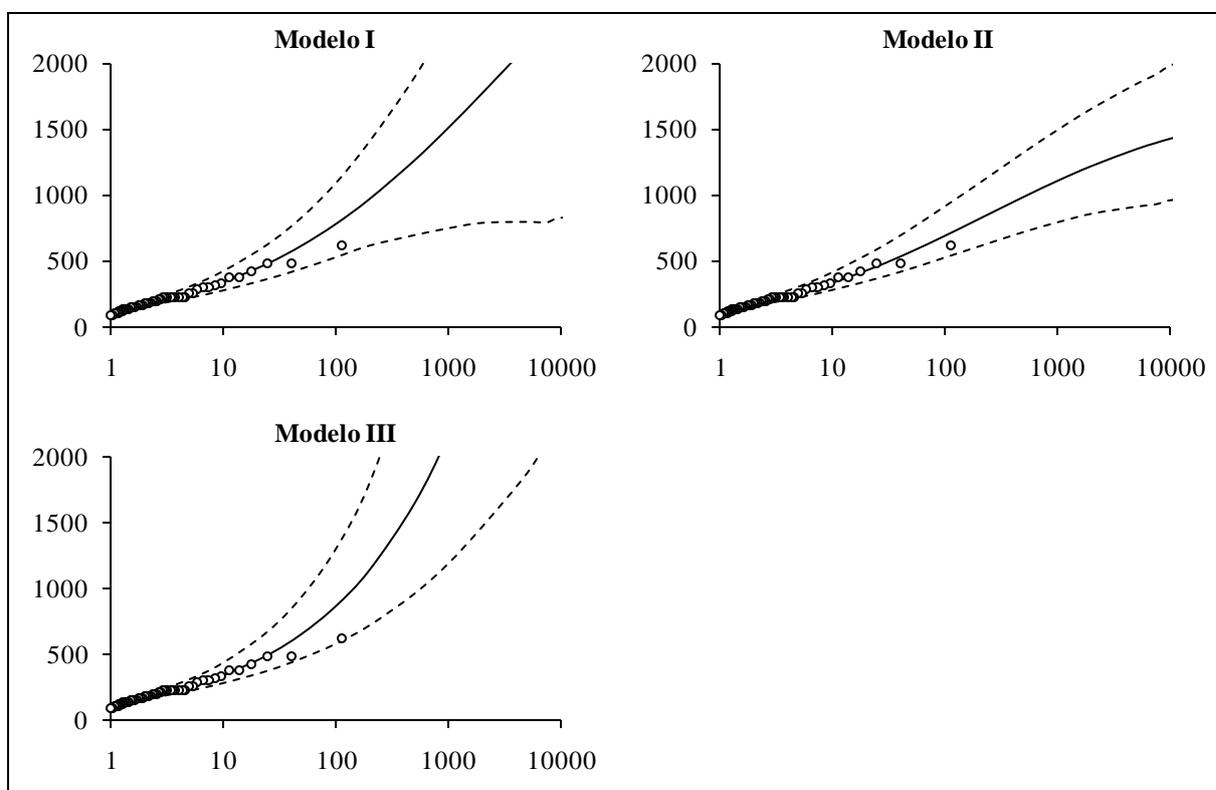


Figura 7.47 – Distribuição preditiva *a posteriori* (média) ou curva de quantis (linha contínua), intervalo de credibilidade de 95% (linha tracejada). Os círculos representam os dados sistemáticos e as barras com círculos representam os dados não sistemáticos. No eixo horizontal está o período de retorno em anos e no eixo vertical está a vazão em m³/s

A figura 7.48 mostra uma comparação entre os quantis estimados e os observados. Observa-se que o modelo II apresentou uma maior aderência aos dados.

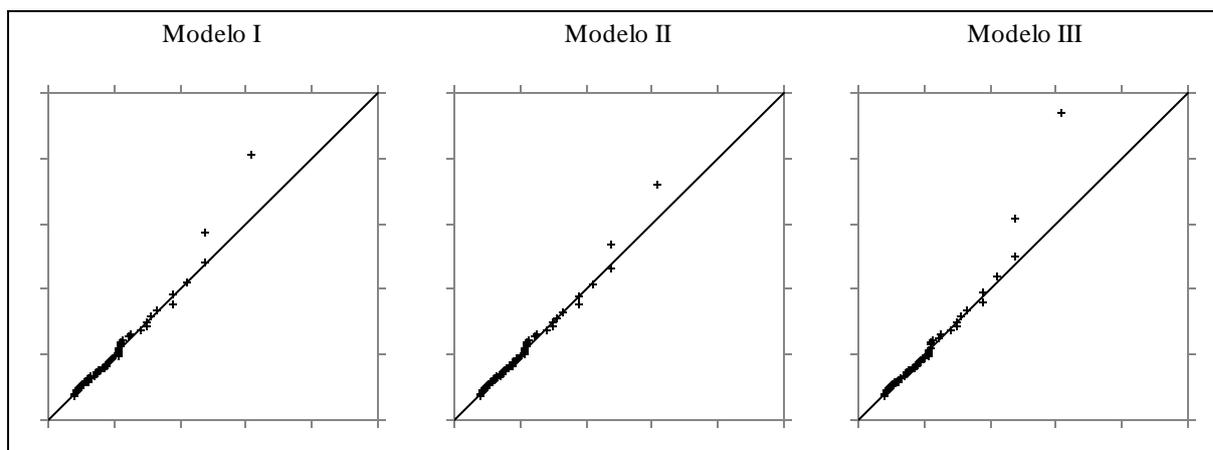


Figura 7.48 – Comparação entre os quantis estimados (eixo vertical) e observados (eixo horizontal) para o modelo EV4

7.4.5 Comparação com outros métodos de análise

Os resultados para Ponte do Vilela foram comparados a cinco diferentes métodos de estimativas 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, tal como exposto no item 4.3.1. 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): nesse 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;
- Série de duração parcial (SDP): Ajuste do modelo clássico de Poisson-Pareto à série de duração parcial de vazões máximas anuais de Ponte do Vilela com taxa média de excedência igual a 1,5; e

- Pico-Volume-Precipitação (PVP): aplicação do método proposto por Fernandes e Naghettini (2008), tal como descrito no item 4.4.1.

A figura 7.49 mostra a comparação entre as curvas de quantis obtidas pelos métodos listados acima e com os modelos LN4-I e EV4-II.

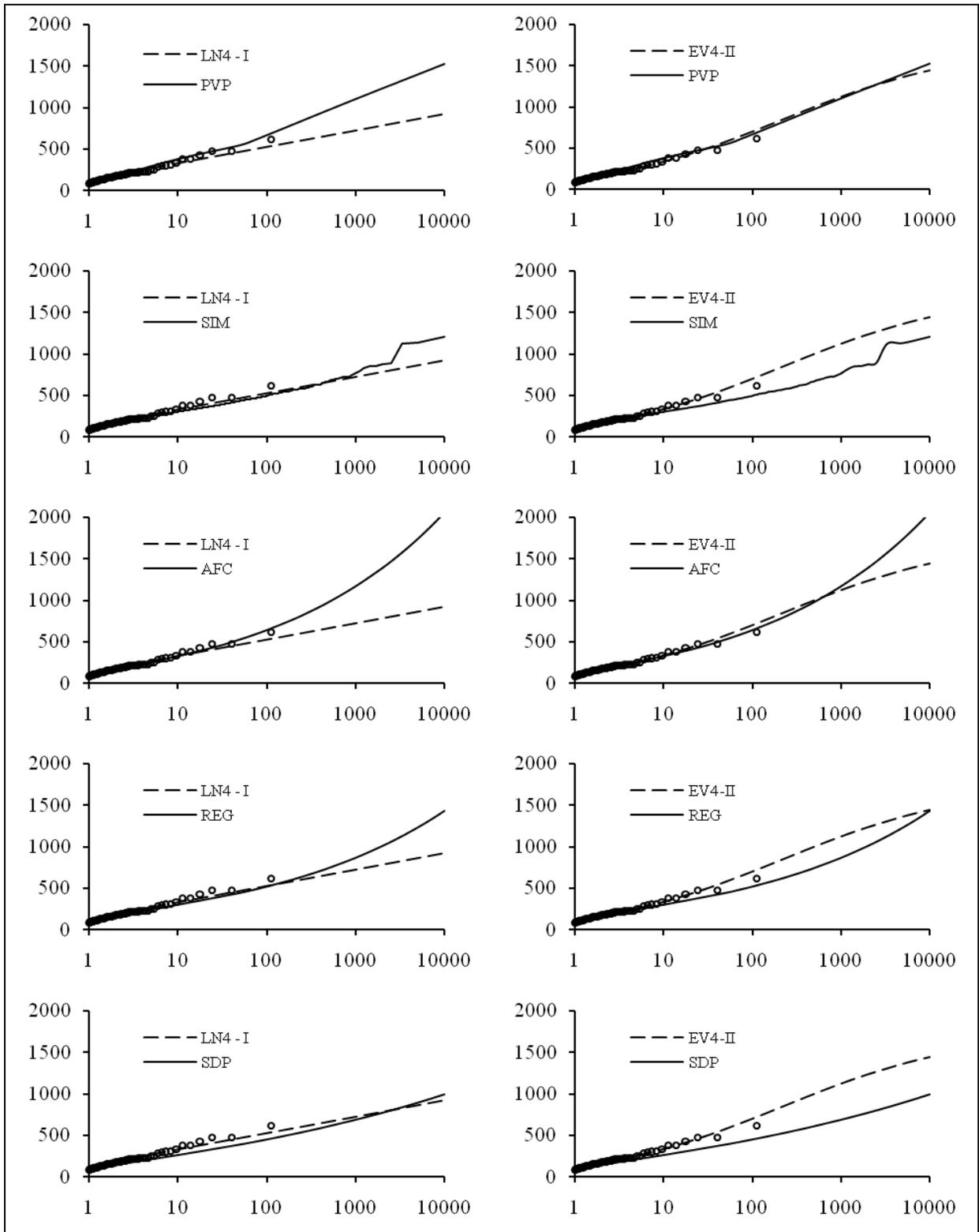


Figura 7.49 – Comparação entre diversos métodos para os dados de Ponte do Vilela

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 extrapolação, as diferenças entre os métodos começam a aparecer. As questões envolvendo a engenharia de recursos hídricos residem precisamente nessa faixa, onde é necessário fazer predições 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.

Nesse sentido, destacam-se os modelos REG e o PVP que agregam dados de diferentes fontes em suas respectivas estruturas. No que se refere à predição, o modelo que mais distoou dos demais foi a análise de frequência convencional (AFC), com os maiores valores para quantis com altos períodos de retorno. Em geral, a EV4 foi a que mais se aproximou dos demais resultados, sendo que a maior diferença foi observada no modelo SDP.

Vale salientar a grande similaridade entre os resultados válidos para os modelos EV4-II e PVP, tendo em vista a completa diferença entre as estruturas de análise dos mesmos. Conforme apontado por Fernandes e Naghettini (2008), o método PVP tem a vantagem, em relação à análise estatística convencional, de incorporar, de forma lógica, os três principais fatores que afetam as distribuições de probabilidade dos picos de vazão, a saber: a hidrometeorologia local, a transformação chuva-vazão e a hidráulica fluvial. Esse fato, associado ao fato de o modelo EV4-II ter sido o melhor dentre os modelos LN4 e EV4, pode ser visto, mesmo que de forma parcial, como uma confirmação da qualidade do método proposto nesta tese.

8 CONCLUSÕES E RECOMENDAÇÕES

Este trabalho teve como objetivo a concepção e a construção de um modelo probabilístico com a capacidade de descrever variáveis aleatórias limitadas superiormente. Mais do que isso, buscou-se um método para a estimação do limite superior que vai além da prática comum de se fixar um valor extremo para a variável a partir de análises determinísticas das condições físicas impostas ao problema. No caso específico desta tese, a variável de interesse é a vazão máxima anual, sendo que seu limite superior pode ser inicialmente estimado pelo valor corrente da PMF.

Embora haja uma controvérsia sobre a existência ou não de um limite superior para a precipitação em uma determinada região, existem fortes evidências de que tais variáveis sejam limitadas superiormente. A esse respeito, é bastante conhecida a controvérsia quanto à existência da PMP, originalmente formulada como um limite superior de produção de precipitação por uma coluna de ar atmosférico. Se de fato existe um limite inequívoco para a PMP, sua determinação fica comprometida pela insuficiente quantificação da variabilidade espaço-temporal das variáveis que lhe dão origem, o que permite dizer que a PMP é uma variável aleatória, cuja variabilidade pode ser descrita por uma distribuição de probabilidades. Sendo as precipitações limitadas superiormente, é natural que as vazões máximas anuais também o sejam. Com efeito, sob condições extremas de umidade no solo, próximas à saturação, as vazões em trânsito em um curso d'água dependem quase exclusivamente do volume de água precipitado. Essa é a idéia implícita no cálculo da PMF, o equivalente da PMP para as vazões máximas anuais, que também é vista como uma variável aleatória.

O fato de as vazões máximas anuais serem consideradas como variáveis aleatórias limitadas superiormente motivou o uso de distribuições limitadas para descrever seu comportamento. Dentre as distribuições limitadas, foram utilizados nesta pesquisa os modelos EV4 e LN4, seguindo a proposta de Botero (2006). De fato, o trabalho de Botero (2006) foi a base inicial para o desenvolvimento metodológico proposto nesta tese.

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 as variáveis do modelo (vazões observadas, paleocheias e parâmetros) foi feita seguindo o paradigma bayesiano. Essa abordagem tornou mais flexível e intuitiva a inclusão da PMF na

análise, uma vez que seu valor é visto como uma informação *a priori* sobre o limite superior e não como o próprio limite superior. A idéia defendida nesta tese é que a distribuição do limite superior das vazões máximas anuais pode ser estimada a partir da variabilidade regional da PMF e do grau de conhecimento do analista a respeito da chance da estimativa local da PMF ser superada. Essa idéia é corroborada pelos resultados obtidos nas aplicações feitas no rio American, no rio Llobregat e no rio Pará, conforme descrito no capítulo 7.

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 capítulo 3, 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 do intervalo de credibilidade com a inclusão de informações não sistemáticas, como, por exemplo, ilustrado na figura 7.11, é um indicativo de que esse tipo de dado contribui não só para a melhoria das estimativas dos parâmetros, mas também para uma melhor predição de valores futuros de vazão, sobretudo aqueles com baixa probabilidade de excedência.

A inclusão da PMF na análise foi feita através da especificação de uma distribuição *a priori* para o limite superior. Dois procedimentos foram propostos: um baseado na distribuição espacial da PMF (Procedimento A) e outro baseado na transposição da PMF de outras bacias para o local de interesse por meio da curva envoltória (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 o quão precisa é a estimativa atual da PMF e o quão 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* dada no item 6.5. 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. No entanto, nesse caso os resultados foram piores que aqueles obtidos com as PMF's estimadas em distintas bacias.

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, a utilização do método proposto nesta tese, quando não for possível estabelecer uma distribuição informativa para o limite superior, é desencorajada.

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

- Pelo menos para as aplicações realizadas nesta tese, as distribuições de probabilidade limitadas superiormente, sobretudo a LN4, além de serem mais plausíveis para a modelagem de vazões máximas anuais, se ajustam melhor aos dados sistemáticos e não sistemáticos quando comparadas ao modelo não limitado LPIII e GEV;
- 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 sobre a distribuição *a posteriori* dos parâmetros é muito mais informativa que a estimação pontual da análise freqüentista. Esse fato pode ser observado na aplicação em Pont Du Vilomara, onde foram confrontados os métodos clássico e bayesiano sob as mesmas distribuições de probabilidades. Além disso, a possibilidade da inclusão de informações subjetivas, 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 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. Uma atenção especial deve ser dada aos projetos de grandes

estruturas hidráulicas, para os quais o risco hidrológico é avaliado para vazões com alto período de retorno.

Embora as aplicações feitas 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. Nesse sentido, Hosking e Wallis (1997) afirmam que a avaliação de métodos estatísticos deve ser feita por meio de experimentos de Monte Carlo. A principal vantagem desse tipo de abordagem é que se conhece a “verdade” sobre o problema em questão, ou seja, parte-se de uma situação cujo resultado é conhecido e aplica-se o método proposto, avaliando sua capacidade de predizer a resposta.

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. A principal dificuldade no uso dessa técnica para testar o método proposto parece ser a relação entre a distribuição *a priori* para o limite superior e os dados gerados. Com efeito, em um contexto bayesiano, a distribuição *a priori* não tem relação direta com os dados observados. Em outras palavras, poderia ocorrer de a população gerada no experimento possuir valores muito maiores ou muito menores que a média da distribuição *a priori* para o limite superior. Assim, o experimento não representaria a realidade física para o local estudado. Uma possível solução para esse problema seria a fixação da distribuição *a priori* e a fixação ou, pequena variação, da média populacional em valores abaixo do limite superior. No entanto, isso limitaria a avaliação da capacidade de generalização dos resultados obtidos.

Uma última questão que surge na realização de um experimento de Monte Carlo se refere ao tempo de processamento. Para os modelos onde foi necessário estabelecer um *lag* de 10, o tempo de processamento foi de aproximadamente 2 horas para gerar uma amostra final de 50.000 realizações da distribuição conjunta *a posteriori* dos parâmetros. Para um *lag* de 20, o tempo ultrapassou 3 horas. Assim, em um experimento de Monte Carlo, onde são necessárias milhares de simulações, não seria possível o uso de um software tal como o WinBUGS. Nesse caso, deveria ser implementada uma rotina de cálculo mais rápida e eficiente ou simplificar o experimento.

Esses problemas, bem como o experimento de Monte Carlo como um todo, não foi abordado nesta tese, mas seguem como recomendação de estudo em futuras pesquisas. Uma das principais respostas de um experimento de Monte Carlo seria a avaliação do desempenho do método proposto sob a condição de dados ilimitados (gerados por uma GEV, por exemplo).

Outro resultado interessante seria a avaliação do desempenho das distribuições ilimitadas sob a condição de dados limitados (gerados por uma LN4, por exemplo).

Por fim, vale salientar que a pesquisa não foi realizada com a intenção de se esgotar o assunto. Pelo contrário, as etapas metodológicas seguidas nesta tese estão sujeitas a diferentes abordagens, podendo ser exploradas sob diversos contextos. A esse 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 item 6.3.4), 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.

Com essas conclusões e recomendações, espera-se que esta tese tenha contribuído para o avanço científico nas aplicações da teoria de probabilidades e estatística na análise hidrológica.

9 REFERÊNCIAS BIBLIOGRÁFICAS

- ACA – Agència Catalana de l'Aigua. Delimitació de zones inundables per a la redacció de l'inuncat: conques internes de Catalunya – Volum III, Generalitat de Catalunya, Departament de Medi Ambient, Espanha, 2001.
- ANDERSON, H. L. Metropolis, Monte Carlo, and The MANIAC. *Los Alamos Science*, U.S GOVERNMENT PRINTING OFFICE, p. 96-108, 1986.
- BAKER, V. R. Paleoflood hydrology and extraordinary flood events. *Journal of Hydrology*, v. 96, p. 79-99, 1987.
- 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.; 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.
- BENITO, G.; THORNDYCRAFT, V. R. Use of systematic, paleoflood and historical data for the improvement of flood risk estimation: An introduction. In: BENITO, G.; THORNDYCRAFT, V. R (Ed.) *Systematic, paleoflood and historical data for the improvement of flood risk estimation: Methodological guidelines*. CSIC – Centro de Ciencias Medioambientales, Madrid, p. 5-14, 2004.
- BENITO, G.; THORNDYCRAFT, V. R.; ENZEL, Y.; SHEFFER, N. A.; RICO, M.; SOPENA, A.; SÁNCHEZ-MOYA, Y. Paleoflood data collection and analysis. In: BENITO, G.; THORNDYCRAFT, V. R (Ed.) *Systematic, paleoflood and historical data for the improvement of flood risk estimation: Methodological guidelines*. CSIC – Centro de Ciencias Medioambientales, Madrid, p. 15-28, 2004.
- BERNARDO, J.; SMITH, A. *Bayesian theory*. New York: John Wiley and Sons, 586 p., 1994.
- BEROD, D.; DEVRED, D.; LAGLAINE, V.; CHAIX, O.; ALTINAKAR, M.; DELLEY, P. Calcul des crues extrêmes par des methodes deterministes du type pluie maximale probable (PMP)/crue maximale probable (PMF): application au cas de la Suisse. Lausanne, 1992.
- BERTONI, J. C.; TUCCI, C. E. M. Precipitação In: TUCCI, C. E. M. (Org.) *Hidrologia: ciência e aplicação*. Porto Alegre: UFRGS/ABRH/EDUSP, p. 177-241, 1993.
- 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.
- BOUGHTON, W. C. A frequency distribution for annual floods. *Water Resources Research*, v. 16, p. 347-354, 1980.
- BROOKS, S. P. Bayesian computation: a statistical revolution. *Phil. Trans. R. Soc. Lond.*, v. 361, p. 2681-2697, 2003.
- CASTELLARIN, A.; VOGEL, R. M.; MATALAS, N. C. Multivariate probabilistic regional envelopes of extreme floods. *Journal of Hydrology*, v. 336, p. 376-390, 2007.

- CASTELLARIN, A.; VOGEL, R. M.; MATALAS, N. C. Probabilistic behavior of a regional envelope curve. *Water Resources Research*, v. 41, doi: 10.1029/2004WR003042, 2005.
- CBDB – Comitê Brasileiro de Barragens. Large Brazilian spillways: an overview of Brazilian practice and experience in designing and building spillways for large dams. CBDB/ICOLD, Rio de Janeiro, Brasil, 205 p., 2002.
- CHANDLER, K. N. The distribution and frequency of record values. *J. R. Stat. Soc.*, B 14, p. 220-228, 1952.
- CHOW, V. T. *Open channel hydraulics*, New York: McGraw Hill, 680 p., 1959.
- COHN, T. A.; LANE, W. L.; BAIER, W. G. An algorithm for computing moments-based flood quantile estimates when historical flood information is available. *Water Resources Research*, v. 33, n. 9, p. 2089-2096, 1997.
- COLES, S. G.; POWELL, E. A. Bayesian methods in extreme value modelling: a review and new developments. *Intern. Statist. Review*, v. 64, n. 1, p. 119-136, 1996.
- 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.
- CRAWFORD, N. H.; LINSLEY, R. K. Digital simulation in hydrology: Stanford Watershed Model IV. *Tech. Rep. No. 39*, Stanford Univ., Palo Alto, California, 1966.
- CRIPPEN, J. R. Envelope curves for extreme flood events. *Journal of Hydraulic Engineering*, v. 108, p. 1208-1212, 1982.
- CUNNANE, C. Unbiased plotting positions, a review. *Journal of Hydrology*, v. 37, p. 205-222, 1978.
- DALRYMPLE, T. Flood frequency analysis. U. S. Geological Survey, *Water supply paper 1543-A*. U. S. Government Printing Office, Washington, DC, 1960.
- DENLINGER, R. P., O'CONNELL, D. R. H.; HOUSE, P. K. Robust determination of stage and discharge: an example from an extreme flood on the Verde river, Arizona . 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. 127-146, 2002.
- ELETROBRÁS. Guia para Cálculo de Cheia de Projeto de Vertedores, Ministério das Minas e Energia, 1987.
- ELIASSON, J. A statistical model for extreme precipitation. *Water Resources Research*, v. 33, n. 3, p. 449-455, 1997.
- ELIASSON, J. Statistical estimates of PMP values. *Nordic Hydrology*, v. 25, p. 301-312, 1994.
- ENGLAND, J. F. *Envelope curve probabilities for dam safety*, U.S. Department of the Interior, Bureau of Reclamation, 21 p., 2005.
- ENGLAND, J. F.; KLINGER, R. E.; CAMRUD, M.; KLAWON, J. E. *Guidelines for preparing preliminary flood frequency analysis reports for comprehensive facility reviews. Version 1.0*. Bureau of Reclamation, Denver, Colorado, 16 p., 2001.
- ENZEL, Y.; ELY, L. L.; HOUSE, P. K.; BAKER, 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, n. 7, p. 2287-2297, 1993.

- 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, n. 3, p. 171-179, doi: 10.1061/(ASCE)1084-0699(2008)13:3(171), 2008.
- FRANCOU, J.; RODIER, J. A. Essai de classification des crues maximales observées dans le monde. In: Cah. O.R.S.T.O.M. sér. Hydrol, v. IV, n. 3, 1967.
- FULLER, W.E. Flood flows. *Trans. Am. Soc. Civ. Eng.*, v. 77, p. 564-617, p. 618-694, 1914.
- GILKS, W. R.; RICHARDSON, S.; SPIEGELHALTER, D. J. *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall/CRC, UK, 286 p., 1996.
- GUILLOT, P.; DUBAND, D. La méthode du gradex pour le calcul de la probabilité des crues à partir des pluies. In: FLOODS AND THEIR COMPUTATION, 1967, Leningrad, *Proceedings of the Leningrad Symposium*, IASH Publication no. 84, p. 560-569, 1967.
- HASTINGS, W. K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, v. 57, n. 1, p. 97-109, doi: 10.1093/biomet/57.1.97, 1970.
- HERSHFIELD, D. M. Estimating the probable maximum precipitation, *J. Hydraul. Div. Am. Soc. Civ. Eng.*, v. 87, p. 99-106, 1961.
- HERSHFIELD, D. M. Method for estimating probable maximum precipitation, *J. Am. Waterworks Assoc.*, v. 57, p. 965-972, 1965.
- HIRSCH, R. M.; STEDINGER, J. R. Plotting positions for historical floods and their precision. *Water Resources Research*, v. 23, n. 4, p. 715-727, 1987.
- 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.
- HITCHCOCK, D. B. A history of Metropolis-Hastings Algorithm. *The American Statistician*, v. 57, n. 4, p. 254-257, doi: 10.1198/0003130032413, 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.
- 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.
- JARRETT, R. D.; ENGLAND, J. F. Reliability of paleostage indicators for paleoflood studies. 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. 91-109., 2002.
- JARRETT, R. D.; TOMLINSON, E. M. Regional interdisciplinary paleoflood approach to assess extreme flood potential, *Water Resources Research*, v. 36, n. 10, p. 2957-2984, 2000.
- JARVIS, C.S. AND OTHERS *Floods in the United States, magnitude and frequency*. U.S. Geological Survey Water-Supply Paper 771, 497 p., 1936.

- 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.
- KIJKO, A. Estimation of the maximum earthquake magnitude, m_{\max} . *Pure Applied Geophysics*. v. 161, p. 1655–1681, doi: 10.1007/s00024-004-2531-4, 2004.
- 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.
- KOUTSOYIANNIS, D. A Probabilistic view of Hershfield's method for estimating probable maximum precipitation. *Water Resources Research*, v. 35, n. 4, p. 1313–1322, 1999.
- LAURSEN, E. M. Comment on “Paleohydrology of southwestern Texas” by KOCHER, R. C.; BAKER, V. R.; PATTON, P. C. *Water Resources Research*, v. 19, n. 5, p. 1339-1339, doi: 10.1029/WR019i005p01339, 1983.
- LETTENMAIER, D. P.; WALLIS, J. R.; WOOD, E. F. Effect of regional heterogeneity on flood frequency estimation. *Water Resources Research*, v. 23, n. 2, p. 313-324, 1987.
- 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.
- LIMA, F. A. Análise bayesiana de frequência de vazões máximas anuais com informações históricas: Aplicação à bacia do rio São Francisco em São Francisco. Dissertação (Mestrado em Saneamento, Meio Ambiente e Recursos Hídricos) – Escola de Engenharia, Universidade Federal de Minas Gerais, Belo Horizonte, 2005.
- LINSLEY, R. K.; KOHLER, M. A.; PAULHUS, J. L. H. *Hydrology for engineers*. New York: McGraw-Hill, 340 p., 1958.
- LIU, J. S. *Monte Carlo Strategies in Scientific Computing*, New York: Springer, 343 p., 2001.
- 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.
- MADSEN, H.; ROSBJERG, D. The partial duration series method in regional index flood modeling. *Water Resources Research*, v. 33, n. 4, p. 737-746, 1997.
- MARTINS, E. S.; STEDINGER, J. R. Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research*, v. 36, n. 3, p. 737-744, 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, n. 10, p. 2559-2567, 2001.
- 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.
- METROPOLIS, N. The beginning of the Monte Carlo Method. *Los Alamos Science*, U.S GOVERNMENT PRINTING OFFICE 1986-676-104/40022, Special Issue, 1987.

- METROPOLIS, N.; ROSENBLUTH, A.; ROSENBLUTH, M.; TELLER, A.; TELLER, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, v. 21, n. 6, p. 1087-1092, 1953.
- MYERS, V. A. Meteorological estimation os extreme precipitation for spillway design floods. Weather Bureau Technical Memorandum WBTM HYDRO-5, Office of Hydrology, Washington, D.C., USA, 1967.
- NAGHETTINI, M. C.; PINTO, E. J. A. *Hidrologia estatística*, CPRM – Serviço Geológico do Brasil, Belo Horizonte, 561 p., 2007.
- NAGHETTINI, M. C.; POTTER, K. W.; ILLANGASEKARE, T. Estimating the upper-tail of flood-peak frequency distributions using hydrometeorological information. *Water Resources Research*, v. 32, n. 6, p. 1729-1740, 1996.
- 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.
- 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.
- NERC – National Environmental Research Council. *Flood studies report – Vol. I: Hydrological studies*. Washington, DC, 550 p., 1975.
- NRC – National Research Council. *Estimating probabilities of extreme floods*, Washington: National Academy Press, 1988.
- NRC – National Research Council. *Improving American river flood frequency analyses*, Committee on American River Flood Frequencies, Washington: National Academy Press, 132 p., 1999.
- 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, n. 5, p. 16.1-16.13, 2002.
- OUARDA, T. B. M. J.; RASMUSSEN, P. F.; BOBÉE, B.; BERNIER, J. Utilisation de l'information historique en analyse hydrologique fréquentielle. *Revue des Sciences de l'Eau* (spécial), p. 41-49, 1998.
- PESSEDA, L. C. R.; GOUVEIA, S. E. M.; FREITAS, H. A.; RIBEIRO, A. S.; ARAVENA, R.; BENDASSOLLI, J. A.; LEDRU, M.; SIEFEDDINE, A. F.; SCHEEL-YBERT, R. Isótopos do carbono e suas aplicações em estudos paleoambientais. In: SOUZA, C. R. G.; SUGUIO, K.; OLIVEIRA, A. M. S.; OLIVEIRA, P. E. (Ed.) *Quaternário do Brasil*, Associação Brasileira de Estudos do Quaternário, Editora Holos, Ribeirão Preto, p. 75-93, 2005.
- PINTO, E. J. A.; NAGHETTINI, M. C. Definition of homogeneous regions and frequency analysis of annual maximum daily precipitation over the upper São Francisco river basin, in southeastern Brazil. In: INTERNATIONAL WATER RESOURCES ENGINEERING CONFERENCE, *ASCE-American Society of Civil Engineer*, Seattle, 1999.
- ROBERT, C. P.; CASELLA, G. *Monte Carlo statistical methods*. New York: Springer, 2° ed., 645 p., 2004.

- RODIER, J. A.; ROCHE, M. World catalogue of maximum observed floods, International Association of Hydrological Sciences, IAHS, Publ. n. 143, 1984.
- SAINT-LAURENT, D. Palaeoflood hydrology: an emerging science. *Progress in physical geography*, v. 28, n. 4, p. 531-543, 2004.
- SINGH, V. P.; WOOLHISER, D. A. Mathematical modeling of watershed hydrology. *Journal of Hydrologic Engineering*. v. 7, n. 4, doi: 10.1061/(ASCE)1084-0699(2002)7:4(270), 2002.
- SLADE, J. J., An asymmetric probability function. *Transactions: American Society of Civil Engineers*, n. 62, p. 35-104, 1936.
- STEDINGER, J. R. Flood frequency analysis and statistical estimation of flood risk. In: WOHL, E. E. (Ed.) *Inland flood hazards: human, riparian, and aquatic communities*, Cambridge University Press, Cambridge, UK, p. 334-358, 2000.
- STEDINGER, J. R.; COHN, T. A. Flood frequency analysis with historical and paleoflood information. *Water Resources Research*, v. 22, n. 5, p. 785-794, 1986.
- STEDINGER, J. R.; LU, L. A. Appraisal of regional and index flood quantile estimators. *Stochastic Hydrology and Hydraulics*, v. 9, n. 1, p. 49-75, 1995.
- TAKARA, K.; LOEBIS, 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.
- 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.
- TUCCI, C. E. M. *Modelos hidrológicos*. Porto Alegre: Editora da UFRGS, ABRH, 1998.
- USACE – United States Army Corps of Engineers. *American river, California, rain flood flow frequency analysis: Civil Design Branch*, Office Report, USACE Sacramento District, Sacramento, USA, 1998.
- 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.
- 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.
- USWRC – United States Water Resources Council. *Guidelines for determining flood flow frequency*, Hydrology Committee, Bulletin 17B (revised), U.S. Government Printing Office, Washington DC, 1982.

- 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.
- VOGEL, R. M.; ZAFIRAKOU-KOULOURIS, A.; MATALAS, N. C. Frequency of record-breaking floods in the United States. *Water Resources Research*, v. 37, n. 6, p. 1723-1731, 2001.
- WEBB, R. H.; JARRETT, R. D. One-dimensional estimation techniques for discharges of paleofloods and historical floods. 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. 111-125, 2002.
- WMO – World Meteorological Organization. *Manual for estimation of probable maximum precipitation*, Operational Hydrologic Report No. 1, WMO No. 332, 2^o ed., Geneva, 270 p., 1986.
- YEVJEVICH, V. Misconceptions in hydrology and their consequences. *Water Resources Research*, v. 4, n. 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.

ANEXO 1 – QUANTIS DE VAZÃO MÁXIMA ANUAL PARA O MODELO LN4 EM FOLSOM

MODELO I - (CV = 0,30 - p = 0,50)				MODELO I* - (CV = 0,30 - p = 0,50)			
T	Média	Mediana	95% HPD	T	Média	Mediana	95% HPD
2	1.429	1.427	(1.106, 1.757)	2	1.598	1.584	(1.187, 2.032)
10	4.543	4.529	(3.741, 5.382)	10	5.032	4.942	(3.612, 6.633)
50	8.363	8.339	(6.833, 9.879)	50	8.838	8.634	(6.120, 11.933)
100	10.114	10.090	(8.272, 11.941)	100	10.439	10.186	(7.203, 14.222)
500	14.141	14.096	(11.653, 16.720)	500	13.831	13.485	(9.155, 19.183)
1.000	15.783	15.723	(13.062, 18.706)	1.000	15.115	14.747	(9.706, 21.115)
5.000	19.236	19.094	(15.812, 22.913)	5.000	17.654	17.283	(10.837, 25.415)
10.000	20.547	20.351	(16.899, 24.698)	10.000	18.568	18.188	(11.075, 26.885)
50.000	23.170	22.857	(18.806, 28.239)	50.000	20.323	19.926	(11.577, 29.927)
100.000	24.127	23.764	(19.444, 29.563)	100.000	20.941	20.531	(11.708, 31.031)

MODELO II - (CV = 0,50 - p = 0,50)				MODELO II* - (CV = 0,50 - p = 0,50)			
T	Média	Mediana	95% HPD	T	Média	Mediana	95% HPD
2	1.425	1.422	(1.109, 1.748)	2	1.605	1.590	(1.192, 2.051)
10	4.478	4.461	(3.680, 5.318)	10	5.044	4.958	(3.638, 6.618)
50	8.284	8.256	(6.761, 9.790)	50	8.799	8.546	(6.123, 12.038)
100	10.068	10.037	(8.263, 11.943)	100	10.378	10.042	(7.024, 14.456)
500	14.294	14.229	(11.671, 16.987)	500	13.762	13.282	(8.253, 19.963)
1.000	16.076	15.973	(13.083, 19.209)	1.000	15.064	14.522	(8.565, 22.345)
5.000	19.961	19.712	(15.910, 24.370)	5.000	17.696	17.040	(9.097, 27.562)
10.000	21.491	21.174	(17.163, 26.787)	10.000	18.665	17.944	(9.190, 29.559)
50.000	24.666	24.170	(18.903, 31.476)	50.000	20.567	19.692	(9.054, 33.419)
100.000	25.868	25.288	(19.569, 33.479)	100.000	21.252	20.286	(9.014, 34.903)

MODELO III - (CV = 0,70 - p = 0,50)				MODELO III* - (CV = 0,70 - p = 0,50)			
T	Média	Mediana	95% HPD	T	Média	Mediana	95% HPD
2	1.419	1.416	(1.105, 1.740)	2	1.606	1.592	(1.197, 2.051)
10	4.426	4.408	(3.625, 5.257)	10	5.054	4.970	(3.616, 6.641)
50	8.223	8.190	(6.722, 9.758)	50	8.867	8.571	(6.161, 12.418)
100	10.033	9.997	(8.211, 11.926)	100	10.503	10.102	(6.956, 15.000)
500	14.420	14.342	(11.737, 17.331)	500	14.106	13.484	(8.154, 21.379)
1.000	16.320	16.191	(13.131, 19.731)	1.000	15.537	14.820	(8.350, 24.183)
5.000	20.584	20.261	(16.015, 25.664)	5.000	18.527	17.528	(8.597, 30.640)
10.000	22.317	21.906	(17.147, 28.377)	10.000	19.667	18.507	(8.553, 33.168)
50.000	26.026	25.361	(19.122, 34.513)	50.000	21.978	20.424	(8.627, 38.937)
100.000	27.475	26.699	(19.768, 37.054)	100.000	22.839	21.082	(8.700, 41.249)

MODELO IV - (PMF adimensional)				MODELO IV* - (PMF adimensional)			
T	Média	Mediana	95% HPD	T	Média	Mediana	95% HPD
2	1.435	1.432	(1.107, 1.769)	2	1.664	1.647	(1.218, 2.141)
10	4.636	4.622	(3.796, 5.468)	10	5.071	5.016	(3.801, 6.463)
50	8.476	8.447	(6.985, 10.001)	50	7.970	7.822	(6.014, 10.109)
100	10.184	10.156	(8.428, 12.015)	100	8.967	8.751	(6.798, 11.535)
500	13.967	13.934	(11.666, 16.339)	500	10.749	10.423	(7.957, 14.341)
1.000	15.447	15.408	(12.912, 17.960)	1.000	11.325	10.962	(8.175, 15.303)
5.000	18.435	18.352	(15.583, 21.423)	5.000	12.335	11.883	(8.457, 17.222)
10.000	19.523	19.400	(16.533, 22.712)	10.000	12.660	12.168	(8.524, 17.923)
50.000	21.621	21.373	(18.401, 25.338)	50.000	13.233	12.652	(8.590, 19.184)
100.000	22.360	22.050	(19.059, 26.308)	100.000	13.419	12.797	(8.615, 19.639)

MODELO V - (Não informativa)				MODELO V* - (Não informativa)			
T	Média	Mediana	95% HPD	T	Média	Mediana	95% HPD
2	1.382	1.383	(1.106, 1.669)	2	1.539	1.526	(1.163, 1.946)
10	4.042	4.037	(3.380, 4.693)	10	5.101	4.963	(3.514, 7.054)
50	7.727	7.695	(6.347, 9.177)	50	10.579	10.102	(6.305, 15.942)
100	9.714	9.668	(7.789, 11.643)	100	13.704	12.988	(7.757, 21.419)
500	15.436	15.277	(11.894, 19.128)	500	23.190	21.573	(11.510, 38.629)
1.000	18.444	18.235	(14.011, 23.219)	1.000	28.408	26.232	(13.463, 48.601)
5.000	26.902	26.554	(19.427, 34.807)	5.000	43.745	39.663	(18.504, 78.949)
10.000	31.259	30.816	(22.104, 40.919)	10.000	51.965	46.768	(21.400, 96.256)
50.000	43.298	42.563	(28.824, 58.113)	50.000	75.584	66.819	(27.069, 145.300)
100.000	49.407	48.456	(30.890, 66.521)	100.000	88.002	77.218	(29.992, 172.234)

MODELO VI - (CV = 0,50 - p = 0,05)				MODELO VI* - (CV = 0,50 - p = 0,05)			
T	Média	Mediana	95% HPD	T	Média	Mediana	95% HPD
2	1.433	1.429	(1.096, 1.759)	2	1.655	1.638	(1.214, 2.137)
10	4.604	4.589	(3.786, 5.456)	10	5.072	5.010	(3.755, 6.490)
50	8.440	8.410	(6.957, 9.985)	50	8.108	7.934	(6.046, 10.535)
100	10.167	10.136	(8.405, 12.013)	100	9.190	8.936	(6.841, 12.143)
500	14.043	14.000	(11.668, 16.448)	500	11.186	10.800	(7.977, 15.380)
1.000	15.583	15.525	(12.984, 18.251)	1.000	11.852	11.418	(8.211, 16.578)
5.000	18.741	18.606	(15.691, 22.076)	5.000	13.048	12.496	(8.418, 18.853)
10.000	19.910	19.721	(16.699, 23.586)	10.000	13.442	12.839	(8.513, 19.747)
50.000	22.196	21.860	(18.479, 26.528)	50.000	14.150	13.418	(8.666, 21.444)
100.000	23.012	22.609	(19.246, 27.794)	100.000	14.384	13.608	(8.654, 21.985)

MODELO VII - (CV = 0,50 - p = 0,95)				MODELO VII* - (CV = 0,50 - p = 0,95)			
T	Média	Mediana	95% HPD	T	Média	Mediana	95% HPD
2	1.407	1.403	(1.098, 1.703)	2	1.563	1.549	(1.181, 1.980)
10	4.281	4.265	(3.536, 5.071)	10	5.059	4.948	(3.520, 6.780)
50	8.040	8.005	(6.571, 9.576)	50	9.718	9.397	(6.165, 13.850)
100	9.916	9.874	(8.067, 11.844)	100	12.034	11.592	(7.386, 17.529)
500	14.754	14.657	(11.833, 17.957)	500	17.914	17.176	(9.685, 27.128)
1.000	16.995	16.851	(13.450, 20.896)	1.000	20.582	19.698	(10.939, 32.137)
5.000	22.407	22.144	(16.834, 28.327)	5.000	26.859	25.720	(12.637, 43.582)
10.000	24.776	24.467	(18.206, 31.906)	10.000	29.528	28.286	(12.838, 48.261)
50.000	30.229	29.816	(20.291, 40.113)	50.000	35.479	34.042	(12.840, 58.608)
100.000	32.518	32.053	(21.210, 44.019)	100.000	37.897	36.390	(13.546, 63.814)

ANEXO 2 – QUANTIS DE VAZÃO MÁXIMA ANUAL PARA O MODELO EV4 EM FOLSOM

MODELO I - (CV = 0,30 - p = 0,50)				MODELO I* - (CV = 0,30 - p = 0,50)			
T	Média	Mediana	95% HPD	T	Média	Mediana	95% HPD
2	1.069	1.067	(889, 1.262)	2	1.353	1.336	(1.005, 1.735)
10	3.537	3.519	(2.815, 4.281)	10	5.735	5.558	(3.544, 8.237)
50	8.581	8.547	(6.703, 10.516)	50	13.448	13.019	(7.527, 20.031)
100	11.536	11.511	(9.120, 13.926)	100	16.731	16.302	(8.250, 25.106)
500	18.470	18.444	(15.455, 21.397)	500	21.878	21.418	(8.405, 33.608)
1.000	20.841	20.707	(17.821, 24.182)	1.000	23.038	22.502	(8.394, 35.892)
5.000	24.336	23.779	(21.369, 28.731)	5.000	24.313	23.635	(8.430, 38.726)
10.000	25.149	24.468	(22.276, 29.937)	10.000	24.535	23.820	(8.435, 39.292)
50.000	26.131	25.296	(23.396, 31.467)	50.000	24.758	23.990	(8.438, 39.871)
100.000	26.330	25.465	(23.640, 31.798)	100.000	24.794	24.018	(8.439, 39.953)

MODELO II - (CV = 0,50 - p = 0,50)				MODELO II* - (CV = 0,50 - p = 0,50)			
T	Média	Mediana	95% HPD	T	Média	Mediana	95% HPD
2	1.067	1.065	(879, 1.256)	2	1.376	1.357	(1.017, 1.781)
10	3.524	3.505	(2.793, 4.274)	10	5.631	5.455	(3.622, 8.074)
50	8.558	8.523	(6.723, 10.550)	50	12.419	11.545	(6.810, 20.739)
100	11.521	11.493	(9.129, 13.955)	100	15.309	14.100	(7.422, 27.058)
500	18.539	18.485	(15.458, 21.673)	500	20.116	17.637	(8.215, 40.406)
1.000	20.970	20.767	(17.798, 24.799)	1.000	21.288	18.314	(8.317, 44.147)
5.000	24.601	23.850	(21.295, 29.966)	5.000	22.641	18.966	(8.404, 49.134)
10.000	25.459	24.547	(22.128, 31.271)	10.000	22.888	19.072	(8.418, 50.101)
50.000	26.506	25.392	(23.350, 33.127)	50.000	23.140	19.169	(8.433, 51.077)
100.000	26.719	25.563	(23.617, 33.523)	100.000	23.183	19.188	(8.436, 51.209)

MODELO III - (CV = 0,70 - p = 0,50)				MODELO III* - (CV = 0,70 - p = 0,50)			
T	Média	Mediana	95% HPD	T	Média	Mediana	95% HPD
2	1.067	1.064	(885, 1.259)	2	1.380	1.359	(1.018, 1.787)
10	3.521	3.504	(2.798, 4.276)	10	5.611	5.417	(3.614, 8.091)
50	8.557	8.523	(6.724, 10.558)	50	12.477	11.012	(6.600, 22.477)
100	11.527	11.503	(9.112, 13.949)	100	15.630	13.224	(7.292, 30.975)
500	18.599	18.516	(15.438, 21.905)	500	21.502	16.100	(8.147, 50.838)
1.000	21.069	20.792	(17.700, 25.136)	1.000	23.119	16.627	(8.265, 57.432)
5.000	24.797	23.878	(21.221, 30.733)	5.000	25.148	17.126	(8.394, 66.658)
10.000	25.690	24.585	(22.174, 32.295)	10.000	25.547	17.209	(8.415, 68.627)
50.000	26.791	25.428	(23.307, 34.318)	50.000	25.970	17.283	(8.435, 70.565)
100.000	27.018	25.599	(23.561, 34.749)	100.000	26.044	17.294	(8.437, 70.884)

MODELO IV - (PMF adimensional)				MODELO IV* - (PMF adimensional)			
T	Média	Mediana	95% HPD	T	Média	Mediana	95% HPD
2	1.074	1.073	(887, 1.262)	2	1.429	1.409	(1.057, 1.855)
10	3.573	3.553	(2.856, 4.318)	10	5.314	5.253	(3.734, 6.988)
50	8.640	8.604	(6.779, 10.562)	50	8.960	8.471	(6.676, 12.505)
100	11.562	11.538	(9.242, 13.990)	100	9.882	9.162	(7.414, 14.395)
500	18.232	18.248	(15.547, 20.865)	500	10.901	9.869	(8.166, 16.928)
1.000	20.438	20.423	(17.840, 23.005)	1.000	11.071	9.972	(8.292, 17.425)
5.000	23.595	23.330	(21.419, 26.397)	5.000	11.233	10.061	(8.404, 17.932)
10.000	24.309	23.944	(22.300, 27.221)	10.000	11.257	10.074	(8.421, 18.015)
50.000	25.154	24.689	(23.423, 28.273)	50.000	11.280	10.087	(8.435, 18.097)
100.000	25.321	24.838	(23.627, 28.433)	100.000	11.284	10.088	(8.437, 18.108)

MODELO V - (Não informativa)			
T	Média	Mediana	95% HPD
2	954	953	(808, 1.097)
10	2.843	2.829	(2.335, 3.365)
50	7.420	7.355	(5.680, 9.338)
100	11.140	11.012	(8.159, 14.407)
500	28.531	28.000	(19.073, 39.627)
1.000	42.776	41.803	(27.398, 61.310)
5.000	109.636	105.997	(60.506, 165.827)
10.000	164.495	158.167	(86.100, 256.263)
50.000	422.049	399.914	(196.314, 702.864)
100.000	633.031	595.527	(281.243, 1.085.164)

MODELO V* - (Não informativa)			
T	Média	Mediana	95% HPD
2	1.314	1.297	(975, 1.662)
10	6.381	6.063	(3.532, 9.930)
50	26.147	23.417	(9.690, 48.942)
100	47.819	41.444	(14.950, 96.719)
500	196.225	155.004	(38.365, 461.893)
1.000	362.325	273.079	(58.533, 908.977)
5.000	1.506.456	1.011.624	(152.774, 4.245.497)
10.000	2.751.382	1.768.704	(183.805, 8.051.275)
50.000	10.186.632	6.297.035	(473.018, 31.684.061)
100.000	16.824.396	10.566.103	(774.199, 52.655.238)

MODELO VI - (CV = 0,50 - p = 0,05)			
T	Média	Mediana	95% HPD
2	1.073	1.071	(883, 1.259)
10	3.561	3.540	(2.863, 4.328)
50	8.618	8.580	(6.809, 10.592)
100	11.547	11.524	(9.227, 13.973)
500	18.292	18.300	(15.491, 20.967)
1.000	20.545	20.501	(17.875, 23.380)
5.000	23.798	23.443	(21.336, 26.986)
10.000	24.540	24.078	(22.253, 27.956)
50.000	25.424	24.847	(23.401, 29.133)
100.000	25.600	24.996	(23.634, 29.377)

MODELO VI* - (CV = 0,50 - p = 0,05)			
T	Média	Mediana	95% HPD
2	1.422	1.402	(1.056, 1.853)
10	5.338	5.263	(3.700, 7.107)
50	9.247	8.607	(6.672, 13.567)
100	10.307	9.362	(7.409, 15.983)
500	11.545	10.141	(8.183, 19.456)
1.000	11.764	10.254	(8.250, 20.137)
5.000	11.977	10.359	(8.411, 20.928)
10.000	12.011	10.373	(8.421, 21.052)
50.000	12.042	10.386	(8.429, 21.180)
100.000	12.047	10.388	(8.437, 21.208)

MODELO VII - (CV = 0,50 - p = 0,95)			
T	Média	Mediana	95% HPD
2	1.057	1.054	(867, 1.243)
10	3.454	3.440	(2.712, 4.238)
50	8.442	8.418	(6.534, 10.439)
100	11.474	11.459	(8.968, 13.930)
500	19.163	18.826	(15.156, 23.954)
1.000	22.112	21.232	(17.318, 29.354)
5.000	27.124	24.649	(20.952, 41.715)
10.000	28.514	25.467	(21.990, 45.918)
50.000	30.454	26.457	(23.149, 52.367)
100.000	30.910	26.655	(23.536, 54.012)

MODELO VII* - (CV = 0,50 - p = 0,95)			
T	Média	Mediana	95% HPD
2	1.327	1.309	(995, 1.701)
10	6.122	5.860	(3.429, 9.196)
50	19.426	18.241	(7.783, 32.724)
100	28.627	26.932	(8.048, 48.856)
500	52.970	50.388	(8.379, 93.362)
1.000	61.924	59.101	(8.361, 110.657)
5.000	75.262	71.601	(8.421, 140.512)
10.000	78.265	73.987	(8.436, 148.177)
50.000	81.665	76.496	(8.440, 157.557)
100.000	82.289	76.935	(8.440, 159.414)

ANEXO 3 – QUANTIS DE VAZÃO MÁXIMA ANUAL PARA O MODELO LN4 EM PONT DU VILOMARA

MODELO I - (CV = 1,0 - p = 0,50)				MODELO I* - (CV = 1,0 - p = 0,50)			
T	Média	Mediana	95% HPD	T	Média	Mediana	95% HPD
2	198	197	(147, 251)	2	166	165	(127, 206)
10	768	763	(601, 943)	10	479	469	(340, 638)
50	1.702	1.692	(1.309, 2.116)	50	886	853	(570, 1.267)
100	2.236	2.223	(1.702, 2.798)	100	1.091	1.044	(676, 1.610)
500	3.806	3.779	(2.833, 4.832)	500	1.640	1.549	(901, 2.554)
1.000	4.624	4.585	(3.421, 5.935)	1.000	1.905	1.791	(1.008, 3.055)
5.000	6.832	6.736	(4.902, 9.003)	5.000	2.584	2.409	(1.149, 4.363)
10.000	7.900	7.770	(5.514, 10.505)	10.000	2.901	2.696	(1.211, 5.011)
50.000	10.588	10.380	(6.891, 14.552)	50.000	3.688	3.407	(1.241, 6.611)
100.000	11.811	11.563	(7.515, 16.611)	100.000	4.045	3.730	(1.246, 7.367)

MODELO II - (CV = 2,0 - p = 0,50)				MODELO II* - (CV = 2,0 - p = 0,50)			
T	Média	Mediana	95% HPD	T	Média	Mediana	95% HPD
2	196	196	(146, 247)	2	166	164	(127, 207)
10	745	742	(585, 906)	10	479	469	(343, 644)
50	1.654	1.645	(1.271, 2.046)	50	894	859	(565, 1.290)
100	2.185	2.172	(1.658, 2.736)	100	1.109	1.056	(675, 1.659)
500	3.812	3.775	(2.772, 4.888)	500	1.700	1.597	(907, 2.717)
1.000	4.702	4.651	(3.354, 6.103)	1.000	1.996	1.865	(975, 3.264)
5.000	7.273	7.155	(4.995, 9.793)	5.000	2.787	2.579	(1.119, 4.835)
10.000	8.614	8.453	(5.769, 11.779)	10.000	3.174	2.926	(1.153, 5.634)
50.000	12.312	12.029	(7.738, 17.642)	50.000	4.187	3.818	(1.247, 7.904)
100.000	14.167	13.830	(8.347, 20.498)	100.000	4.673	4.245	(1.259, 9.016)

MODELO III - (PMF adimensional)				MODELO III* - (PMF adimensional)			
T	Média	Mediana	95% HPD	T	Média	Mediana	95% HPD
2	205	203	(149, 266)	2	170	169	(130, 214)
10	872	867	(669, 1.095)	10	479	471	(349, 621)
50	1.920	1.909	(1.462, 2.387)	50	802	786	(567, 1.065)
100	2.463	2.453	(1.878, 3.030)	100	937	914	(652, 1.256)
500	3.822	3.817	(3.020, 4.633)	500	1.225	1.182	(847, 1.710)
1.000	4.409	4.398	(3.510, 5.310)	1.000	1.337	1.284	(912, 1.892)
5.000	5.678	5.637	(4.599, 6.871)	5.000	1.567	1.490	(1.022, 2.285)
10.000	6.164	6.105	(4.976, 7.488)	10.000	1.653	1.567	(1.065, 2.454)
50.000	7.131	7.040	(5.627, 8.806)	50.000	1.826	1.716	(1.122, 2.795)
100.000	7.477	7.372	(5.840, 9.341)	100.000	1.889	1.768	(1.150, 2.945)

MODELO IV - (Não informativa)				MODELO IV* - (Não informativa)			
T	Média	Mediana	95% HPD	T	Média	Mediana	95% HPD
2	195	194	(146, 245)	2	165	163	(126, 204)
10	731	727	(581, 885)	10	481	470	(340, 650)
50	1.623	1.616	(1.255, 2.001)	50	925	887	(574, 1.355)
100	2.153	2.140	(1.646, 2.698)	100	1.166	1.111	(682, 1.756)
500	3.817	3.780	(2.753, 4.927)	500	1.868	1.750	(978, 3.013)
1.000	4.757	4.703	(3.344, 6.219)	1.000	2.240	2.084	(1.116, 3.700)
5.000	7.595	7.478	(5.065, 10.271)	5.000	3.299	3.021	(1.496, 5.763)
10.000	9.150	8.989	(5.962, 12.538)	10.000	3.850	3.501	(1.682, 6.875)
50.000	13.715	13.420	(8.671, 19.579)	50.000	5.390	4.828	(2.117, 10.032)
100.000	16.162	15.779	(10.037, 23.392)	100.000	6.179	5.497	(2.343, 11.707)

ANEXO 4 – QUANTIS DE VAZÃO MÁXIMA ANUAL PARA O MODELO EV4 EM PONT DU VILOMARA

MODELO I - (CV = 1,0 - p = 0,50)				MODELO I* - (CV = 1,0 - p = 0,50)			
T	Média	Mediana	95% HPD	T	Média	Mediana	95% HPD
2	136	134	(107, 167)	2	145	143	(107, 186)
10	537	523	(377, 728)	10	689	655	(377, 1.055)
50	1.627	1.596	(1.094, 2.226)	50	2.490	2.281	(890, 4.518)
100	2.451	2.432	(1.702, 3.224)	100	4.081	3.699	(993, 7.792)
500	5.272	5.065	(3.725, 7.279)	500	10.539	9.464	(1.076, 21.974)
1.000	6.790	6.300	(4.462, 10.434)	1.000	14.309	12.873	(1.108, 30.602)
5.000	10.534	8.593	(5.045, 21.313)	5.000	22.974	20.483	(1.136, 52.223)
10.000	12.041	9.180	(5.114, 27.239)	10.000	25.892	22.747	(1.145, 60.547)
50.000	14.785	9.890	(5.156, 40.119)	50.000	30.137	25.410	(1.148, 74.280)
100.000	15.580	10.028	(5.162, 44.153)	100.000	31.096	25.910	(1.150, 77.895)

MODELO II - (CV = 2,0 - p = 0,50)				MODELO II* - (CV = 2,0 - p = 0,50)			
T	Média	Mediana	95% HPD	T	Média	Mediana	95% HPD
2	128	127	(101, 153)	2	145	143	(107, 186)
10	477	464	(349, 626)	10	704	670	(383, 1.095)
50	1.460	1.430	(987, 1.930)	50	2.832	2.553	(947, 5.214)
100	2.294	2.268	(1.610, 3.106)	100	5.070	4.455	(1.107, 9.941)
500	6.139	5.939	(3.857, 8.902)	500	18.485	15.334	(1.291, 41.484)
1.000	9.175	9.013	(4.667, 14.202)	1.000	30.858	25.148	(1.349, 72.058)
5.000	22.425	22.935	(5.087, 40.540)	5.000	85.108	68.712	(1.386, 207.790)
10.000	32.250	33.160	(5.113, 62.297)	10.000	120.115	97.229	(1.390, 296.816)
50.000	68.885	68.282	(5.141, 155.326)	50.000	213.884	174.224	(1.240, 541.293)
100.000	90.894	85.533	(5.144, 218.957)	100.000	251.668	204.013	(1.241, 646.252)

MODELO III - (PMF adimensional)				MODELO III* - (PMF adimensional)			
T	Média	Mediana	95% HPD	T	Média	Mediana	95% HPD
2	146	145	(116, 181)	2	150	148	(112, 194)
10	633	624	(453, 825)	10	592	578	(386, 823)
50	1.893	1.877	(1.331, 2.481)	50	1.200	1.110	(767, 1.870)
100	2.708	2.704	(1.995, 3.428)	100	1.418	1.268	(890, 2.353)
500	4.634	4.623	(3.825, 5.432)	500	1.731	1.460	(1.057, 3.268)
1.000	5.251	5.177	(4.456, 6.266)	1.000	1.798	1.494	(1.095, 3.502)
5.000	6.071	5.897	(5.035, 7.538)	5.000	1.872	1.527	(1.138, 3.773)
10.000	6.237	6.036	(5.087, 7.859)	10.000	1.885	1.533	(1.141, 3.817)
50.000	6.414	6.177	(5.142, 8.252)	50.000	1.897	1.538	(1.145, 3.868)
100.000	6.445	6.200	(5.148, 8.326)	100.000	1.900	1.538	(1.146, 3.877)

MODELO IV - (Não informativa)				MODELO IV* - (Não informativa)			
T	Média	Mediana	95% HPD	T	Média	Mediana	95% HPD
2	124	124	(102, 146)	2	145	143	(106, 186)
10	446	444	(354, 544)	10	709	673	(387, 1.113)
50	1.376	1.360	(987, 1.776)	50	2.929	2.621	(1.080, 5.514)
100	2.216	2.184	(1.522, 2.951)	100	5.376	4.653	(1.672, 10.945)
500	6.684	6.522	(4.033, 9.568)	500	22.285	17.547	(4.177, 52.751)
1.000	10.753	10.436	(6.127, 15.885)	1.000	41.422	31.048	(5.940, 103.795)
5.000	32.480	31.080	(16.000, 51.509)	5.000	177.984	116.631	(13.151, 499.420)
10.000	52.323	49.684	(24.308, 85.632)	10.000	335.478	206.010	(19.724, 981.762)
50.000	158.531	147.710	(64.122, 279.282)	50.000	1.443.705	768.073	(53.810, 4.614.465)
100.000	255.617	236.142	(96.158, 462.887)	100.000	2.647.197	1.349.647	(79.511, 8.752.568)

ANEXO 5 – QUANTIS DE VAZÃO MÁXIMA ANUAL PARA O MODELO LN4 EM PONTE DO VILELA

MODELO I - (CV = 0,6 - p = 0,50)			
T	Média	Mediana	95% HPD
2	176	175	(155, 198)
10	326	324	(277, 380)
50	465	459	(378, 562)
100	524	517	(418, 640)
500	663	652	(515, 838)
1.000	724	711	(553, 923)
5.000	864	847	(632, 1.122)
10.000	925	906	(666, 1.212)
50.000	1.066	1.043	(742, 1.431)
100.000	1.126	1.101	(779, 1.531)

MODELO II - (PMF adimensional)			
T	Média	Mediana	95% HPD
2	177	177	(155, 199)
10	327	325	(279, 379)
50	458	453	(378, 545)
100	511	506	(418, 613)
500	628	621	(504, 765)
1.000	677	669	(541, 828)
5.000	783	773	(618, 968)
10.000	826	815	(648, 1.023)
50.000	920	908	(714, 1.147)
100.000	958	946	(739, 1.198)

MODELO III - (Não informativa)			
T	Média	Mediana	95% HPD
2	174	174	(154, 196)
10	325	323	(274, 380)
50	475	468	(379, 582)
100	543	534	(420, 673)
500	711	697	(523, 916)
1.000	790	772	(569, 1.031)
5.000	986	959	(682, 1.328)
10.000	1.077	1.046	(732, 1.469)
50.000	1.304	1.261	(853, 1.828)
100.000	1.410	1.360	(904, 1.994)

ANEXO 6 – QUANTIS DE VAZÃO MÁXIMA ANUAL PARA O MODELO EV4 EM PONTE DO VILELA

MODELO I - (CV = 0,6 - p = 0,50)			
T	Média	Mediana	95% HPD
2	161	161	(143, 181)
10	339	335	(269, 417)
50	618	604	(437, 826)
100	781	761	(524, 1.083)
500	1.264	1.225	(695, 1.885)
1.000	1.512	1.465	(748, 2.324)
5.000	2.134	2.068	(797, 3.482)
10.000	2.403	2.326	(829, 4.036)
50.000	2.973	2.859	(815, 5.209)
100.000	3.182	3.043	(816, 5.700)

MODELO II - (PMF - adimensional)			
T	Média	Mediana	95% HPD
2	162	162	(143, 182)
10	335	332	(270, 407)
50	576	567	(433, 737)
100	699	688	(517, 907)
500	998	982	(713, 1.314)
1.000	1.120	1.104	(787, 1.485)
5.000	1.360	1.344	(906, 1.833)
10.000	1.440	1.423	(955, 1.972)
50.000	1.575	1.558	(1.001, 2.194)
100.000	1.615	1.598	(1.007, 2.259)

MODELO III - (Não informativa)			
T	Média	Mediana	95% HPD
2	161	160	(142, 180)
10	344	338	(270, 428)
50	673	653	(458, 930)
100	895	863	(568, 1.290)
500	1.738	1.641	(948, 2.766)
1.000	2.314	2.162	(1.177, 3.838)
5.000	4.516	4.103	(1.863, 8.135)
10.000	6.032	5.406	(2.274, 11.259)
50.000	11.855	10.256	(3.659, 23.999)
100.000	15.883	13.511	(4.520, 33.280)