

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

**ESTIMAÇÃO INDIRETA DE QUANTIS DE
ENCHENTES EXTREMAS A PARTIR DE
MODELOS CHUVA-VAZÃO COM EMPREGO
CONJUNTO DE UM GERADOR ESTOCÁSTICO
DE PRECIPITAÇÃO DIÁRIA, ANÁLISE
BAYESIANA E DISTRIBUIÇÕES LIMITADAS
SUPERIORMENTE**

Veber Afonso Figueiredo Costa

Belo Horizonte

2015

**ESTIMAÇÃO INDIRETA DE QUANTIS DE
ENCHENTES EXTREMAS A PARTIR DE MODELOS
CHUVA-VAZÃO COM EMPREGO CONJUNTO DE UM
GERADOR ESTOCÁSTICO DE PRECIPITAÇÃO
DIÁRIA, ANÁLISE BAYESIANA E DISTRIBUIÇÕES
LIMITADAS SUPERIORMENTE**

Veber Afonso Figueiredo Costa

Veber Afonso Figueiredo Costa

**ESTIMAÇÃO INDIRETA DE QUANTIS DE
ENCHENTES EXTREMAS A PARTIR DE MODELOS
CHUVA-VAZÃO COM EMPREGO CONJUNTO DE UM
GERADOR ESTOCÁSTICO DE PRECIPITAÇÃO
DIÁRIA, ANÁLISE BAYESIANA E 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: Wilson dos Santos Fernandes

Belo Horizonte

Escola de Engenharia da UFMG

2015

Página com as assinaturas dos membros da banca examinadora, fornecida pelo Colegiado do Programa

AGRADECIMENTOS

Agradeço à minha esposa Camila e minha filha Ana Beatriz, pelo incentivo, apoio e compreensão durante a elaboração desta tese.

Aos meus pais, Antonio e Marília, e meus irmãos, Vinícius e Mariana, por estarem sempre presentes.

Ao meu orientador, professor Wilson Fernandes, pela constante dedicação e apoio no decorrer deste trabalho.

Ao professor Mauro Naghettini, co-orientador não oficial desta pesquisa, pelas ideias e contribuições essenciais ao desenvolvimento do método.

Ao colega Francisco Eustáquio Silva, pela ajuda providencial com o código de calibração.

Aos amigos do EHR.

À CAPES, pela concessão da bolsa de estudos.

Por fim, ao professor Upmanu Lall, pelas valiosas sugestões de melhoria no âmbito geral da pesquisa.

RESUMO

A estimação de quantis de enchentes extremas é absolutamente necessária para o projeto de grandes estruturas hidráulicas. No entanto, sua descrição probabilística por meio de análise de frequência convencional é envolta em grandes dificuldades, uma vez que as amostras disponíveis são geralmente reduzidas, tornando necessária a extrapolação até quantis que ultrapassam um “limite crível”. Uma alternativa para contornar esse problema é a estimação indireta dos quantis de cheias por meio de modelos de transformação chuva-vazão, abordagem essa que permite uma caracterização mais adequada da variabilidade de respostas hidrológicas de uma bacia mediante a simulação sob condições distintas, ainda que equiprováveis, de *input*.

Na presente pesquisa, propõe-se uma estrutura metodológica Bayesiana com esse propósito, a qual compreende um novo modelo de geração estocástica de precipitação diária, cuja principal particularidade reside no uso de modelos distributivos limitados superiormente na descrição de eventos extremos, e um método de calibração baseado na função de verossimilhança generalizada, a qual sintetiza de maneira mais precisa o comportamento dos parâmetros do modelo hidrológico. Essa estrutura confere fundamentação física aos processos modelados, o que proporciona estimativas mais confiáveis de quantis de enchentes extremas e das incertezas associadas aos mesmos.

O método proposto foi aplicado à bacia do rio Pará em Minas Gerais e à bacia do rio American, no estado americano da Califórnia. Os resultados mostraram a eficiência do modelo de geração na simulação das características amostrais da precipitação diária e na extrapolação a quantis com reduzida probabilidade de superação. Além disso, o modelo chuva-vazão se mostrou apto a reproduzir até mesmo aquelas cheias oriundas de monitoramento não sistemático, o que constitui um indicativo da adequação da estrutura metodológica proposta.

ABSTRACT

The estimation of extreme flood quantiles is absolutely necessary for the design of large hydraulic structures. However, their probabilistic description by means of conventional flood frequency analysis is a complex task, since the size of the available samples is usually small, which means that it is necessary the extrapolation to quantiles beyond a credible limit. An alternative to circumvent this problem is to proceed an indirect estimation of the flood quantiles with the use of rainfall-runoff models, since such an approach provides a more suitable characterization of the hydrological response of a catchment through simulation under distinct yet equally probable input conditions.

In this research, it is proposed a Bayesian framework for this purpose. This framework comprises a new model for daily rainfall stochastic generation, which employs upper-bounded distribution functions to describe extreme events, and a calibration method based on the generalized likelihood function, which synthetizes in a more accurate manner the behavior of the parameters of the hydrological model. This structure of analysis has a physically based nature, which leads to more reliable estimates of extreme flood quantiles and their uncertainties.

Applications of the proposed method were conducted for the Pará river basin, in the Brazilian state of Minas Gerais, and for the American river basin, in California. Results showed the efficiency of the stochastic generator in simulating the empirical characteristics of daily rainfall, as well as in the extrapolation of frequency curves to quantiles with very small exceedance probabilities. Furthermore, the rainfall-runoff model proved to be capable of reproducing even very large non-systematic floods, thus showing the suitability of the methodological framework proposed herein.

SUMÁRIO

1	INTRODUÇÃO	1
2	OBJETIVOS	9
2.1	OBJETIVO GERAL	9
2.2	OBJETIVOS ESPECÍFICOS.....	9
3	REVISÃO DE LITERATURA	10
3.1	GERAÇÃO DE SÉRIES SINTÉTICAS DE PRECIPITAÇÃO.....	10
3.1.1	<i>Introdução</i>	10
3.1.2	<i>Modelos para geração estocástica de precipitação diária</i>	12
3.2	DISTRIBUIÇÕES LIMITADAS SUPERIORMENTE	19
3.2.1	<i>O uso de distribuições superiormente limitadas em hidrologia: fundamentação física e estimação paramétrica</i>	19
3.2.2	<i>Modelos distributivos dotados de limite superior explícito</i>	22
3.3	INFERÊNCIA BAYESIANA	28
3.3.1	<i>Introdução</i>	28
3.3.2	<i>Estimação Bayesiana</i>	31
3.3.3	<i>Intervalos de credibilidade</i>	32
3.3.4	<i>Métodos de cálculo</i>	33
3.4	PRECIPITAÇÃO MÁXIMA PROVÁVEL: CONCEITO E MÉTODOS DE ESTIMAÇÃO	37
3.4.1	<i>Introdução</i>	37
3.4.2	<i>Métodos hidrometeorológicos de estimação de PMP</i>	38
3.4.3	<i>Métodos estatísticos de estimação de PMP</i>	41
3.5	MODELOS DE SIMULAÇÃO HIDROLÓGICA	43
3.5.1	<i>Introdução</i>	43
3.5.2	<i>Calibração de parâmetros modelos hidrológicos sob o paradigma Bayesiano</i>	45
3.5.3	<i>O modelo Rio Grande de transformação chuva-vazão</i>	48
4	METODOLOGIA	57
4.1	GERAÇÃO ESTOCÁSTICA DE PRECIPITAÇÃO DIÁRIA.....	57
4.2	CONSTRUÇÃO DAS DISTRIBUIÇÕES A PRIORI DOS PARÂMETROS DOS MODELOS LN4 E EV4	66
4.2.1	<i>Construção do limite superior para a bacia do rio Pará</i>	66
4.2.2	<i>Construção do limite superior para a bacia do rio American</i>	72
4.3	CALIBRAÇÃO DOS PARÂMETROS DO MODELO RIO GRANDE DE TRANSFORMAÇÃO CHUVA-VAZÃO	75
4.4	CONSTRUÇÃO DAS CURVAS DE QUANTIS DE ENCHENTES.....	80
5	APLICAÇÃO DO MÉTODO À BACIA DO RIO PARÁ	82
5.1	CARACTERIZAÇÃO DA BACIA DO RIO PARÁ, EM PONTE DO VILELA	82

5.2	ESTAÇÕES PLUVIOMÉTRICAS, FLUVIOMÉTRICAS E EVAPORIMÉTRICAS UTILIZADAS NO ESTUDO	84
5.3	DISTRIBUIÇÃO A PRIORI PARA O LIMITE SUPERIOR	85
5.4	ESTATÍSTICAS A POSTERIORI	85
5.4.1	<i>Modelo LN4</i>	85
5.4.2	<i>Modelo EV4</i>	91
5.5	GERAÇÃO ESTOCÁSTICA DE PRECIPITAÇÃO DIÁRIA	96
5.5.1	<i>Definição do limiar entre chuvas convencionais e chuvas extremas</i>	97
5.5.2	<i>Definição do tamanho das janelas de reamostragem</i>	103
5.5.3	<i>Avaliação de desempenho do modelo misto de geração estocástica de precipitação diária</i> 104	
5.5.4	<i>Comparação com outros modelos de geração estocástica de precipitação diária</i>	109
5.6	CALIBRAÇÃO DOS PARÂMETROS DO MODELO RIO GRANDE	111
5.7	CONSTRUÇÃO DAS CURVAS DE QUANTIS DE ENCHENTES	121
6	APLICAÇÃO DO MÉTODO À BACIA DO RIO AMERICAN	127
6.1	CARACTERIZAÇÃO DA BACIA DO RIO AMERICAN, EM FOLSOM	127
6.2	ESTAÇÕES PLUVIOMÉTRICAS, FLUVIOMÉTRICAS E EVAPORIMÉTRICAS UTILIZADAS NO ESTUDO	128
6.3	DISTRIBUIÇÃO A PRIORI PARA O LIMITE SUPERIOR	130
6.4	ESTATÍSTICAS A POSTERIORI	131
6.4.1	<i>Modelo LN4</i>	131
6.4.2	<i>Modelo EV4</i>	135
6.5	GERAÇÃO ESTOCÁSTICA DE PRECIPITAÇÃO DIÁRIA	140
6.5.1	<i>Definição do limiar entre chuvas convencionais e chuvas extremas</i>	140
6.5.2	<i>Definição do tamanho das janelas de reamostragem</i>	145
6.5.3	<i>Avaliação de desempenho do modelo misto de geração estocástica de precipitação diária</i> 146	
6.5.4	<i>Comparação com outros modelos de geração estocástica de precipitação diária</i>	150
6.6	CALIBRAÇÃO DOS PARÂMETROS DO MODELO RIO GRANDE	151
6.7	CONSTRUÇÃO DAS CURVAS DE QUANTIS DE ENCHENTES	160
7	CONCLUSÕES E RECOMENDAÇÕES	166
8	REFERÊNCIAS BIBLIOGRÁFICAS	171

LISTA DE FIGURAS

Figura 3.1 – Efeito de cada parâmetro na distribuição LN4.....	24
Figura 3.2 – Efeito de cada parâmetro na distribuição LN4.....	26
Figura 3.3 – Efeito de cada parâmetro na distribuição TDF.....	28
Figura 3.4 – Estimativa do fator k_m para o método estatístico de estimação da PMP.....	42
Figura 3.5 – Fluxograma do módulo de produção do modelo Rio Grande. Fonte: Pinheiro (2009)	52
Figura 4.1 – Histograma de frequências das estimativas de PMP de 1 dia em Minas Gerais.....	70
Figura 4.2 – Histograma de frequências das estimativas de PMP de 1 dia no estado da Califórnia	74
Figura 5.1 – Localização da bacia do rio Pará, na estação de Ponte do Vilela (40130002).....	83
Figura 5.2 – Perfil longitudinal do rio Pará.....	83
Figura 5.3 – Variação da estatística de Brooks-Gelman-Rubin ao longo da simulação. O eixo x representa o número de iterações em milhares e o eixo y os valores da referida estatística ...	87
Figura 5.4 – Variação dos valores dos parâmetros do modelo LN4 ao longo da simulação. O eixo x representa o número de iterações em milhares	87
Figura 5.5 – Distribuições <i>a posteriori</i> para o parâmetro μ	88
Figura 5.6 - Distribuições <i>a posteriori</i> para o parâmetro σ	89
Figura 5.7 – Distribuição <i>a priori</i> e distribuição <i>a posteriori</i> do parâmetro α . (a) Modelo I; (b) Modelo II.	90
Figura 5.8 – Distribuições <i>a posteriori</i> para o parâmetro σ . (a) Modelo I; (b) Modelo II.	93
Figura 5.9 – Distribuições <i>a posteriori</i> para o parâmetro ξ	94
Figura 5.10 – Distribuição <i>a priori</i> e distribuição <i>a posteriori</i> do parâmetro α . (a) Modelo I; (b) Modelo II.	96
Figura 5.11 – Precipitação diária máxima anual gerada para diferentes limiares entre chuvas convencionais e extremas para o modelo LN4	98
Figura 5.12 – Comparação entre a precipitação média mensal das séries simuladas com o modelo LN4 e da série observada para diferentes limiares entre chuvas convencionais e extremas.....	99
Figura 5.12 (cont.) – Comparação entre a precipitação média mensal das séries simuladas com o modelo LN4 e da série observada para diferentes limiares entre chuvas convencionais e extremas.....	100
Figura 5.13 – Precipitação diária máxima anual gerada para diferentes limiares entre chuvas convencionais e extremas para o modelo EV4	101
Figura 5.14 – Comparação entre a precipitação média mensal das séries simuladas com o modelo EV4 e da série observada para diferentes limiares entre chuvas convencionais e extremas.....	102

Figura 5.14 (cont.) – Comparação entre a precipitação média mensal das séries simuladas com o modelo EV4 e da série observada para diferentes limiares entre chuvas convencionais e extremas.....	103
Figura 5.15 – Comparação entre estatísticas das séries geradas e observada para diferentes tamanhos de janela de reamostragem para o modelo LN4.....	104
Figura 5.16 – Comparação das estatísticas das séries geradas com o modelo LN4 com a série observada para cada mês do ano. (a) chuva média diária; (b) altura máxima de chuva diária; (c) coeficiente de variação; (d) coeficiente de assimetria; (e) número médio de dias chuvosos. O eixo X representa os meses do ano.....	105
Figura 5.17 – Comparação entre as médias (a) e variâncias (b) mensais das séries simuladas com o modelo LN4 e a série observada	106
Figura 5.18 – Curva de quantis de precipitação diária obtida com o modelo LN4	107
Figura 5.19 – Comparação das estatísticas das séries geradas a partir do modelo EV4 com a série observada para cada mês do ano. (a) chuva média diária; (b) altura máxima de chuva diária; (c) coeficiente de variação; (d) coeficiente de assimetria; (e) número médio de dias chuvosos. O eixo X representa os meses do ano.	108
Figura 5.20 – Comparação entre as médias (a) e variâncias (b) mensais das séries simuladas com o modelo EV4 e a série observada	108
Figura 5.21 – Curva de quantis de precipitação diária obtida com o modelo EV4	109
Figura 5.22 – Curvas de quantis para diferentes modelos de geração estocástica de precipitação diária.....	111
Figura 5.23 – Comparação das distribuições <i>a priori</i> e <i>a posteriori</i> dos parâmetros do modelo Rio Grande.....	115
Figura 5.23 (cont.) – Comparação das distribuições <i>a priori</i> e <i>a posteriori</i> dos parâmetros do modelo Rio Grande	116
Figura 5.23 (cont.) – Comparação das distribuições <i>a priori</i> e <i>a posteriori</i> dos parâmetros do modelo Rio Grande	117
Figura 5.24 – Comparação entre os hidrogramas simulado e observado para o período de calibração	118
Figura 5.25 – Vazões simuladas versus vazões observadas	118
Figura 5.26 – Resíduos de calibração com a função de verossimilhança generalizada versus vazões observadas	119
Figura 5.27 – Autocorrelação dos resíduos de calibração	120
Figura 5.28 – Distribuições empírica (círculos) e teórica (linha contínua) dos resíduos de calibração padronizados	120
Figura 5.29 – Comparação entre os hidrogramas simulado e observado para o período de validação	121
Figura 5.30 – Curva de quantis de enchentes na bacia do rio Pará, em Ponte do Vilela. As linhas tracejadas correspondem ao intervalo de credibilidade de 95%.....	123
Figura 5.31 – Comparação das curvas de quantis de enchentes obtidas com diferentes métodos	126
Figura 6.1 – Localização da bacia do rio American, na estação de Fair Oaks (11446500).....	128

Figura 6.2 – Polígonos de Thiessen para a determinação da chuva média espacializada na bacia do rio American, em Folsom	129
Figura 6.3 – Distribuições <i>a posteriori</i> para o parâmetro μ	132
Figura 6.4 – Distribuições <i>a posteriori</i> para o parâmetro σ	133
Figura 6.5 – Distribuição <i>a priori</i> e distribuição <i>a posteriori</i> do parâmetro α . (a) Modelo I; (b) Modelo II.	135
Figura 6.6 – Distribuições <i>a posteriori</i> para o parâmetro σ	137
Figura 6.7 – Distribuições <i>a posteriori</i> para o parâmetro ξ	138
Figura 6.8 – Distribuição <i>a priori</i> e distribuição <i>a posteriori</i> do parâmetro α . (a) Modelo I; (b) Modelo II.	139
Figura 6.9 – Precipitação diária máxima anual gerada para diferentes limiares entre chuvas convencionais e extremas para o modelo LN4	141
Figura 6.10 – Comparação entre a precipitação média mensal das séries simuladas com o modelo LN4 e da série observada para diferentes limiares entre chuvas convencionais e extremas.....	142
Figura 6.10 (cont.) – Comparação entre a precipitação média mensal das séries simuladas com o modelo LN4 e da série observada para diferentes limiares entre chuvas convencionais e extremas.....	143
Figura 6.11 – Precipitação diária máxima anual gerada para diferentes limiares entre chuvas convencionais e extremas para o modelo EV4	143
Figura 6.12 – Comparação entre a precipitação média mensal das séries simuladas com o modelo EV4 e da série observada para diferentes limiares entre chuvas convencionais e extremas.....	144
Figura 6.12 (cont.) – Comparação entre a precipitação média mensal das séries simuladas com o modelo EV4 e da série observada para diferentes limiares entre chuvas convencionais e extremas.....	145
Figura 6.13 – Comparação entre estatísticas das séries geradas e observada para diferentes tamanhos de janela de reamostragem para o modelo LN4.....	145
Figura 6.14 – Comparação das estatísticas das séries geradas com o modelo LN4 com a série observada para cada mês do ano. (a) chuva média diária; (b) altura máxima de chuva diária; (c) coeficiente de variação; (d) coeficiente de assimetria; (e) número médio de dias chuvosos. O eixo X representa os meses do ano.....	147
Figura 6.15 – Comparação entre as médias (a) e variâncias (b) mensais das séries simuladas com o modelo EV4 e a série observada	147
Figura 6.16 – Curva de quantis de precipitação diária obtida com o modelo LN4	148
Figura 6.17 – Comparação das estatísticas das séries geradas com o modelo EV4 com a série observada para cada mês do ano. (a) chuva média diária; (b) altura máxima de chuva diária; (c) coeficiente de variação; (d) coeficiente de assimetria; (e) número médio de dias chuvosos. O eixo X representa os meses do ano.....	149
Figura 6.18 – Comparação entre as médias (a) e variâncias (b) mensais das séries simuladas com o modelo EV4 e a série observada	149
Figura 6.19 – Curva de quantis de precipitação diária obtida com o modelo EV4	150

Figura 6.20 – Curvas de quantis para diferentes modelos de geração estocástica de precipitação diária.....	151
Figura 6.21 – Comparação das distribuições a priori e a posteriori dos parâmetros do modelo Rio Grande.....	154
Figura 6.21 (cont.) – Comparação das distribuições a priori e a posteriori dos parâmetros do modelo Rio Grande	155
Figura 6.21 (cont.) – Comparação das distribuições a priori e a posteriori dos parâmetros do modelo Rio Grande	156
Figura 6.22 – Comparação entre os hidrogramas simulado e observado para o período de calibração	157
Figura 6.23 – Vazões simuladas versus vazões observadas	157
Figura 6.24 – Resíduos de calibração com a função de verossimilhança generalizada <i>versus</i> vazões observadas	158
Figura 6.25 – Autocorrelação dos resíduos de calibração	159
Figura 6.26 – Distribuições empírica (círculos) e teórica (linha contínua) dos resíduos de calibração padronizados	159
Figura 6.27 – Comparação entre os hidrogramas simulado e observado para o período de validação	160
Figura 6.28 – Curva de quantis de enchentes na bacia do rio American, em Folsom. As linhas tracejadas correspondem ao intervalo de credibilidade de 95%.....	162
Figura 6.29 – Comparação das curvas de quantis de enchentes obtidas com diferentes métodos	164
Figura 6.30 – Comparação dos intervalos de credibilidade dos métodos de estimação de cheias	165

LISTA DE TABELAS

Tabela 4.1 – Estimativas de PMP estatística de 1 dia em Minas Gerais	68
Tabela 4.2 – Estimativas de PMP meteorológica em bacias da Califórnia, reduzidas para a área da bacia do rio American, em Folsom.....	74
Tabela 4.3 – Informações necessárias à utilização do método DREAM.....	78
Tabela 5.1 – Estações pluviométricas utilizadas para determinação da chuva média espacializada na bacia do rio Pará, em Ponte do Vilela. Fonte: Lima (2004).....	84
Tabela 5.2 – Estações evaporimétricas utilizadas no estudo. Fonte: Lima (2004)	84
Tabela 5.3 – Parâmetros e características das distribuições <i>a priori</i> do limite superior	85
Tabela 5.4 – Estatísticas <i>a posteriori</i> para o parâmetro μ	87
Tabela 5.5 – Estatísticas <i>a posteriori</i> do parâmetro σ	89
Tabela 5.6 – Estatísticas <i>a posteriori</i> para o parâmetro α	90
Tabela 5.7 – Estatísticas <i>a posteriori</i> para o parâmetro σ	92
Tabela 5.8 – Estatísticas <i>a posteriori</i> para o parâmetro ξ	94
Tabela 5.9 – Estatísticas <i>a posteriori</i> para o parâmetro α	95
Tabela 5.10 – Intervalos de variação dos parâmetros do modelo Rio Grande.....	112
Tabela 5.11 – Intervalos de variação dos parâmetros da função de verossimilhança generalizada.....	113
Tabela 5.12 – Parâmetros para aplicação do algoritmo DREAM	113
Tabela 5.13 – Índices de desempenho do período de calibração.....	117
Tabela 5.14 – Índices de desempenho para o período de validação.....	121
Tabela 5.15 – Estimativa de quantis de enchentes para diversos tempos de retorno.....	122
Tabela 6.1 – Estações pluviométricas utilizadas para determinação da chuva média espacializada na bacia do rio American, em Folsom.....	129
Tabela 6.2 – Parâmetros e características das distribuições <i>a priori</i> do limite superior	131
Tabela 6.3 – Estatísticas <i>a posteriori</i> para o parâmetro μ	132
Tabela 6.4 – Estatísticas <i>a posteriori</i> para o parâmetro σ	133
Tabela 6.5 – Estatísticas <i>a posteriori</i> para o parâmetro α	133
Tabela 6.6 – Estatísticas <i>a posteriori</i> para o parâmetro σ	136
Tabela 6.7 – Estatísticas <i>a posteriori</i> para o parâmetro ξ	138
Tabela 6.8 – Estatísticas <i>a posteriori</i> para o parâmetro α	138
Tabela 6.9 – Intervalos de variação dos parâmetros do modelo Rio Grande.....	152
Tabela 6.10 – Intervalos de variação dos parâmetros da função de verossimilhança	152
Tabela 6.11 – Parâmetros para aplicação do algoritmo DREAM	152
Tabela 6.12 – Índices de desempenho do período de calibração.....	156
Tabela 6.13 – Índices de desempenho para o período de validação.....	160

Tabela 6.14 – Estimativa de quantis de enchentes para diversos tempos de retorno..... 161

LISTA DE SÍMBOLOS

A – Área de drenagem

CPRM – Companhia de Pesquisa de Recursos Minerais – Serviço Geológico do Brasil

CV – Coeficiente de variação

DGP – Distribuição generalizada de Pareto

DP – Desvio padrão

EMP – Enchente Máxima Provável

EV2 – Distribuição de Fréchet ou de valores extremos do tipo II

EV3 – Distribuição de Weibull ou de valores extremos do tipo III

EV4 – Distribuição de valores extremo do tipo IV

f, f_x – Função densidade de probabilidade

F, F_x – Função acumulada de probabilidade

GAM – Distribuição Gama

GEV – Distribuição Generalizada de Valores Extremos

HPD – Highest Probability Density ou intervalo de credibilidade

L – Função de verossimilhança

LN4 – Distribuição Log-normal de 4 parâmetros

LP3 – Distribuição Log-Pearson do tipo III

MCMC – Markov Chain Monte Carlo

PMP – Precipitação Máxima Provável

PVP – Método Pico-Volume-Precipitação

Q – Vazão máxima anual

R_A – Coeficiente de armazenamento

T – Tempo de retorno

TC – Tempo de concentração da bacia

USACE – United States Army Corps of Engineers

USBR – United States Bureau of Reclamation

USGS – United States Geological Service

WMO – Organização Meteorológica Mundial

α – Limite superior das precipitações máximas anuais

ε – Limite inferior das precipitações máximas anuais

μ – Parâmetro de posição

σ, β – Parâmetro de escala

ξ, ρ – Parâmetro de forma

b – Expoente da curva de distribuição espacial da retenção capilar

c – coeficiente de evapotranspiração profunda

ci – Coeficiente de recessão do fluxo sub-superficial diário

cg – Coeficiente de recessão do fluxo subterrâneo diário

ex – expoente da curva de distribuição espacial de capacidade de água livre

imp – Fração da área impermeável da bacia

k – Coeficiente de tanque

kss – Coeficiente do fluxo diário do escoamento sub-superficial

kg – Coeficiente do fluxo diário do escoamento subterrâneo

sm – capacidade média espacial de água livre

wdm – capacidade de retenção capilar espacial média na zona profunda do solo

wlm – capacidade de retenção capilar espacial média na zona inferior do solo

wum – Capacidade de retenção capilar média espacial na zona superior do solo

1 INTRODUÇÃO

Um dos maiores desafios da Engenharia de Recursos Hídricos nos dias atuais é a modelagem de eventos extremos de precipitação e vazão. Tais eventos são de fundamental importância no estabelecimento de diretrizes de projeto e avaliação de risco de grandes estruturas hidráulicas para aproveitamento de recursos hídricos e mitigação de cheias. No entanto, sua descrição probabilística encontra grandes dificuldades, uma vez que as amostras disponíveis são geralmente reduzidas, o que torna necessária a extrapolação até quantis que ultrapassam um “limite crível” para tal. Diante disso, diversas abordagens têm sido avaliadas com objetivo de prover estimativas mais confiáveis de quantis e das incertezas a eles associadas.

Dentre tais abordagens, vem ganhando destaque a modelagem de eventos raros e extremos com distribuições de probabilidade limitadas superiormente (ELÍASSON, 1994, 1997; TAKARA & TOSA, 1999; BOTERO, 2006; FERNANDES *et al.*, 2010; GUSE *et al.*, 2010). O uso de modelos com limite superior para precipitações e vazões máximas anuais é baseado na hipótese de que tanto o suprimento de umidade à atmosfera quanto a resposta das bacias hidrográficas em condições extremas de formação de enchentes são fisicamente limitados. A existência de tal limite, contudo, é uma questão bastante controversa na Engenharia Hidrológica. Ainda mais controversa é a capacidade de estimá-lo com as ferramentas estatísticas de uso corrente na análise de frequência, tendo-se em vista as grandes incertezas impostas pelas amostras disponíveis e pelos métodos de estimação de parâmetros. Entretanto, pesquisas recentes, como as de Botero (2006), Fernandes *et al.* (2010) e Guse *et al.* (2010), mostraram que a incorporação de um limite superior à modelagem de máximos anuais traz melhorias significativas às estimativas de cheias com menores probabilidades de superação.

Trabalhos contemplando limites superiores se fazem presentes na literatura científica desde a década de 1920, quando Jarvis (1925) introduziu o conceito de curvas envoltórias de vazões máximas. Essas curvas correspondem a limites empíricos de cheias observadas, em um conjunto de estações fluviométricas dispersas em uma dada região. Na década seguinte, Horton (1936) propôs um modelo superiormente limitado argumentando que essa é uma forma de modelagem mais racional, uma vez que: “... *nenhum corpo d'água terrestre pode produzir uma cheia infinita*”. O modelo de Horton incorpora implicitamente em sua formulação as hipóteses de limitação física da

produção de chuva e da resposta hidrológica da bacia. Outros modelos, tais como a envoltória de Creager (Creager & Justin, 1950) e a fórmula de Fuller (Fuller, 1914), foram bastante empregados naquela época para o projeto de grandes estruturas hidráulicas.

Esses modelos, contudo, são essencialmente empíricos e se mostraram inadequados a uma vasta gama de estudos de enchentes extremas (FERNANDES, 2009). Dessa forma, os hidrólogos se voltaram para os modelos de natureza determinística, comprometidos com a tentativa de descrição da realidade física do fenômeno em estudo. Essa transição foi grandemente facilitada pelo desenvolvimento das ciências meteorológicas e de técnicas de modelagem nas décadas seguintes, e culminou com a publicação do “Manual de Estimação da Precipitação Máxima Provável (PMP)”, da WMO, no ano de 1973. Em linhas gerais, a PMP pode ser definida como a altura máxima teórica de precipitação, para uma dada duração, cuja ocorrência em uma determinada área, em uma certa região geográfica e época do ano, é meteorologicamente possível (WMO, 1986). As hipóteses básicas implícitas na estimação da PMP são: (1) existe um limite físico para a produção de precipitação; e (2) esse limite pode ser estimado a partir de considerações determinísticas, com base no conhecimento atual dos processos hidrometeorológicos sob condições extremas. Métodos de estimação da PMP contemplam a maximização meteorológica local de eventos históricos, com ou sem transposição de tormentas, ou a estimação estatística, tal como o método proposto por Hershfield (1961, 1965) e posteriormente modificado por Koutsoyiannis (1999).

A “Enchente Máxima Provável (EMP)” (WMO, 1973), por sua vez, representa o limite superior de uma enchente potencial em uma dada seção fluvial, resultante de uma tempestade hipotética de duração e altura críticas, a PMP, antecedida por severíssimas, porém, fisicamente possíveis condições hidrológicas e hidrometeorológicas. A transformação de PMP em EMP é realizada usualmente através de modelos chuva-vazão calibrados para a bacia hidrográfica em estudo.

Os conceitos de PMP e EMP têm sido globalmente empregados no projeto de estruturas hidráulicas, como vertedores de grandes barragens e centrais nucleares, sendo considerados, de um ponto de vista prático, como valores máximos de precipitação e vazão, respectivamente (FERNANDES *et al.*, 2010). Entretanto, ainda que estejam associadas, por definição, a limites superiores, as estimativas de PMP e EMP não são

inequívocas e dependem intrinsecamente das amostras disponíveis e das técnicas empregadas em sua estimação, o que faz com que as mesmas apresentem incertezas passíveis de serem interpretadas de modo probabilístico. Nesse contexto, a PMP e a EMP devem ser entendidas como variáveis aleatórias, cujas probabilidades de excedência, ainda que reduzidas, não são nulas, e, assim, tais variáveis não devem ser tratadas como limites superiores para o propósito de análise de frequência.

Mais recentemente, pesquisas relacionadas a limites superiores ganharam impulso, com destaque para três vertentes: (1) modelagem probabilística de curvas envoltórias de vazão (Ohnishi *et al.*, 2004; Pegram & Parak, 2004; Bayazit & Önöz, 2004; Vogel *et al.*, 2007) e precipitação (Castellarin *et al.*, 2009); (2) modelagem probabilística da teoria dos recordes para cheias (Koutsoyiannis & Baloutsos, 2000; Vogel *et al.*, 2001; Douglas & Vogel, 2006); e (3) emprego de distribuições de probabilidade dotadas de limite superior explícito (Elíasson, 1994, 1997; Takara & Loebis, 1996; Takara & Tosa, 1999; Botero, 2006; Fernandes *et al.*, 2010, Guse *et al.*, 2010). O presente trabalho se inclui na terceira categoria, com destaque para o emprego da distribuição log-normal de 4 parâmetros, ou LN4 (Slade, 1936), e da distribuição de valores extremos do tipo 4, ou EV4 (Kanda, 1981). Essas distribuições são particularmente interessantes em função da possibilidade de se associar o limite superior a características hidrometeorológicas de uma bacia.

Apesar dos avanços na utilização de modelos superiormente limitados em análises hidrológicas, a aceitação desses modelos e de suas premissas não é unânime na comunidade científica e tem promovido intensos debates entre pesquisadores desde o início do século XX. Yevjevich (1968), por exemplo, argumenta que não há prova definitiva de limites superiores físicos para precipitações e que essa consideração pode levar a equívocos na modelagem de eventos extremos. Papalexiou & Koutsoyiannis (2006) também sustentam que não há evidências de limites superiores para o suprimento de umidade do ar e, conseqüentemente, para alturas de precipitação, e questionam o uso da PMP no projeto de estruturas hidráulicas. Hosking & Wallis (1997), por sua vez, consideram errôneo o uso de distribuições limitadas sob um ponto de vista matemático, indicando que, em função da posição do limite superior no domínio das variáveis hidrológicas, as estimativas de quantis para tempos de retorno de

interesse para a maioria das estruturas, em torno de 100 anos, podem ser severamente comprometidas.

Por outro lado, há autores que defendem o uso dos modelos limitados na análise de frequência, argumentando que não há uma razão científica para que as cheias em um dado curso d'água possam crescer indefinidamente. Nesse sentido, pesquisadores como Boughton (1980), Laursen (1983) e Takara & Tosa (1999) sustentam que os limites superiores físicos das enchentes e precipitações devem ser descritos por valores finitos nos modelos probabilísticos empregados na análise de frequência.

Diante das controvérsias acerca do limite superior de precipitação e vazão, diversos autores têm recorrido a evidências físicas e geomorfológicas que demonstrem sua existência e forneçam fundamentação à adoção de distribuições com essa característica na análise de frequência. Enzel *et al.* (1993), por exemplo, empregaram as chamadas paleocheias em conjunto com séries de observações sistemáticas na construção de curvas envoltórias de vazão para a bacia do rio Colorado, e mostraram que tais curvas tendem a se estabilizar à medida que o tempo de retorno cresce, indicando um possível limite físico tanto para a precipitação quanto para a resposta hidrológica da bacia. Jacoby *et al.* (2008), empregando vestígios de cheias extremas obtidos em árvores e depósitos aluviais, chegaram a conclusões similares quanto à existência de um limite natural para as cheias, ao construir curvas envoltórias para bacias israelenses. Esses dois trabalhos fornecem argumentos para o uso de modelos probabilísticos limitados superiormente na modelagem de precipitações e cheias máximas na presente pesquisa.

A dificuldade que se impõe à utilização de modelos distributivos limitados é a estimação do limite superior. Métodos de estimação pontual, tais como a maximização da função de verossimilhança, são fortemente influenciados pelo tamanho das amostras disponíveis, o que pode agregar imensas incertezas à análise caso as mesmas sejam reduzidas. Em lugar da estimação pontual, autores como Elíasson (1994, 1997), Takara & Tosa (1999) e Botero (2006) sugerem a fixação de estimativas de PMP ou EMP como limites superiores nos modelos probabilísticos. Entretanto, esse procedimento não é inteiramente adequado uma vez que essas variáveis, ainda que interpretadas como estimadores convenientes para os limites superiores de precipitação e vazão, possuem uma probabilidade não-nula de superação, e essa consideração deve ser devidamente introduzida na análise.

Fernandes *et al.* (2010) sugeriram a adoção do paradigma Bayesiano para abordar o problema de estimação do limite superior de variáveis hidrológicas. A análise Bayesiana permite a inclusão da PMP e da EMP em estruturas lógicas de análise, nas quais os referidos estimadores são acomodados em modelos probabilísticos adequadamente construídos, de forma a refletir as probabilidades de excedência associadas aos mesmos.

No método proposto por Fernandes *et al.* (2010), estimativas de EMP para um conjunto de bacias hidrográficas foram empregadas na descrição da incerteza acerca do limite superior das enchentes em uma seção do curso d'água. Entretanto, ainda que o método tenha apresentado resultados promissores na análise de frequência de cheias raras e extremas, a construção da distribuição *a priori* para o referido parâmetro constitui um objeto de questionamento, uma vez que a mesma faz uso de estimativas de EMP obtidas em bacias com áreas de drenagem muito variadas, nas quais os processos hidrológicos de formação de cheias podem apresentar natureza física bastante distinta. Em outras palavras, ao se combinar no modelo probabilístico estimativas de uma variável fortemente afetada pelo fator de escala, tal como a EMP, pode-se perder o realismo físico do fenômeno a ser modelado na construção da distribuição *a priori*.

Uma alternativa para se contornar esse problema é empregar a precipitação como variável de interesse na análise, uma vez que extensas áreas, abrangendo um conjunto significativo de bacias hidrográficas, são caracterizadas por regimes pluviométricos similares, o que faz com que essa variável hidrológica seja praticamente independente do fator de escala. As premissas fundamentais do método proposto por Fernandes *et al.* (2010) podem ser devidamente ajustadas a essa nova perspectiva: as precipitações máximas anuais podem ser modeladas por distribuições limitadas superiormente e a estimação de parâmetros pode ser realizada sob o paradigma Bayesiano, empregando-se estimativas de PMP na construção da distribuição *a priori* para o limite superior das alturas de chuva. Na sequência, estimativas de quantis de cheias podem então ser obtidas a partir dos quantis de precipitação com auxílio de um modelo hidrológico.

Em um contexto mais amplo, os modelos distributivos limitados superiormente podem ser empregados para a geração estocástica de precipitação diária. Nesse tipo de metodologia, características estatísticas, tais como média, desvio padrão e correlação temporal, são extraídas de uma amostra e utilizadas para a geração de séries sintéticas com comportamento similar ao da série original. Empregadas em conjunto com um

modelo de simulação hidrológica contínua devidamente calibrado para a bacia em estudo, as séries de precipitação estocasticamente geradas podem ser utilizadas para a obtenção de séries de vazão e análise de frequência de cheias. Essa abordagem se mostra interessante uma vez que as enchentes máximas anuais usualmente não são formadas por um único evento de precipitação, mas por uma combinação particular de eventos, com distribuições temporais e espaciais tais que levam a bacia a condições de saturação críticas antes da ocorrência de uma tormenta severa.

Geradores estocásticos de chuva têm sido empregados desde a década de 1950, e podem ser baseados em duas abordagens distintas: a abordagem paramétrica, que faz uso de distribuições teóricas de probabilidade para ajustar os dados observados, e a abordagem não paramétrica, que emprega modelos de reamostragem para a simulação. Tais abordagens têm se mostrado adequadas à reprodução de características médias das amostras em uma ampla gama de aplicações (WILKS & WILBY, 1999). Contudo, ambas apresentam limitações na modelagem dos eventos extremos (SHARIF & BURN, 2006; LI *et al.*, 2012). A abordagem paramétrica convencional, via de regra, emprega distribuições com cauda superior exponencial, o que leva à subestimação sistemática das precipitações de maior tempo de retorno (LI *et al.*, 2012). Modelos paramétricos híbridos, por outro lado, fazem uso de distribuições com cauda superior pesada para modelar as alturas de chuva mais elevadas, e dessa forma valores fisicamente implausíveis podem ser gerados com relativa frequência (LI *et al.*, 2013; CHEN & BRISSETTE, 2014). Por fim, a abordagem não paramétrica apresenta capacidade nula ou bastante reduzida de extrapolação dos valores observados (WILKS & WILBY, 1999), e esse é um grande problema na caracterização da cauda superior da distribuição das precipitações (SHARMA & LALL, 1999).

Na presente pesquisa, é proposto um novo modelo de gerador estocástico de precipitação diária, que combina as principais vantagens das abordagens paramétrica e não paramétrica e busca contornar as limitações das mesmas. O gerador se baseia em uma estrutura que trata separadamente a ocorrência de chuva e a modelagem dos volumes precipitados. A ocorrência de precipitação é modelada por um processo estocástico de múltiplos estados discretos, que contemplam a ausência de precipitação, a ocorrência de chuvas convencionais e a ocorrência de chuvas extremas. Havendo chuva convencional, as alturas precipitadas são modeladas por um procedimento de

reamostragem, que torna o modelo mais parcimonioso e permite maior flexibilidade de ajuste. Se, por outro lado, ocorrer chuva extrema, as alturas precipitadas são modeladas por distribuições limitadas superiormente, as quais permitem a extrapolação a valores bastante superiores ao máximo observado e, assim, restringem a geração a alturas de precipitação fisicamente plausíveis.

A conversão das séries de precipitação obtidas com o gerador proposto em séries de vazão requer o emprego de um modelo hidrológico contínuo. E nesse ponto reside uma das limitações do método aqui proposto: se por um lado o uso da precipitação como variável primária de interesse possibilita um maior grau de realismo físico, reduzindo as incertezas inerentes ao processo modelado, por outro lado, o uso de um modelo chuva-vazão constitui uma nova fonte de incertezas, decorrentes das simplificações adotadas na estrutura do modelo para a representação do fenômeno físico de formação dos escoamentos e da escolha do conjunto de parâmetros no processo de calibração. De fato, uma vez que os modelos hidrológicos são meras abstrações da realidade, aproximações inapropriadas ou não consideradas para os mecanismos físicos envolvidos na transformação chuva-vazão podem resultar em erros significativos (SILVA *et al.*, 2014). Dessa forma, a correta identificação dos parâmetros e a avaliação das incertezas oriundas da modelagem chuva-vazão se tornam absolutamente necessárias para a obtenção de estimativas confiáveis de quantis de enchentes, em especial as mais raras e extremas.

Também nesse contexto o paradigma Bayesiano oferece uma alternativa coerente para análise. Com efeito, os parâmetros do modelo hidrológico podem ser interpretados como variáveis aleatórias, cujo conhecimento prévio é resumido por suas respectivas distribuições *a priori*. A partir de uma amostra de realizações do modelo e de um critério objetivo para avaliação, tal como uma função de verossimilhança adequada, obtém-se, em lugar do tradicional conjunto de valores ótimos dos parâmetros, os quais raramente constituem seus valores corretos, uma amostra da distribuição conjunta *a posteriori* dos mesmos, a qual caracteriza completamente suas incertezas após a inferência. Dessa forma, é possível proceder a uma avaliação mais acurada da incerteza global da simulação e na estimação dos quantis de enchentes.

A presente tese introduz duas inovações metodológicas. A primeira diz respeito ao novo modelo para geração estocástica de precipitação diária, construído a partir da

combinação de abordagens paramétricas e não paramétricas, e que emprega distribuições limitadas superiormente para a geração de alturas de precipitação raras e extremas, em um contexto Bayesiano de estimação paramétrica. No gerador proposto, conceitos de natureza aparentemente antagônica, como a PMP determinística e um processo estocástico para geração de chuva, são combinados em modelo probabilístico com maior fundamentação física, o que permite a obtenção de estimativas mais confiáveis de quantis de precipitação e das incertezas associadas. A segunda inovação reside no método para a quantificação das incertezas de todo o processo de simulação hidrológica, da ocorrência da precipitação à concentração do escoamento na seção exutória da bacia, também em uma estrutura de análise Bayesiana. Partindo-se de séries de precipitação obtidas com o gerador estocástico e de conjuntos de parâmetros aleatoriamente gerados em simulações de sua distribuição conjunta *a posteriori*, é possível resumir de maneira mais precisa as incertezas nos quantis de cheias. A metodologia proposta constitui um importante passo na busca de um instrumento confiável para avaliação das probabilidades associadas às cheias de projeto de estruturas hidráulicas e de mitigação de enchentes.

Além deste capítulo introdutório, esta tese se organiza da seguinte forma: no capítulo 2 são apresentados os objetivos geral e específicos da pesquisa. No capítulo 3 é realizada a revisão de literatura, englobando os principais métodos de geração estocástica de chuva diária, as principais características da abordagem Bayesiana, as distribuições limitadas superiormente, a conceituação e os métodos de cálculo da Precipitação Máxima Provável e o modelo Rio Grande de transformação chuva-vazão. O capítulo 4 descreve a metodologia proposta, com ênfase no gerador estocástico de chuva e na construção das distribuições *a priori* do limite superior das precipitações. Nos capítulos 5 e 6 são realizadas aplicações completas da metodologia às bacias do rio Pará, em Minas Gerais, e do rio American, no estado americano da Califórnia. No capítulo 7 são apresentadas as principais conclusões e recomendações para futuros desenvolvimento da pesquisa. Por fim, no capítulo 8 são apresentadas as referências bibliográficas utilizadas no trabalho.

2 OBJETIVOS

2.1 Objetivo Geral

Desenvolver um método para estimação indireta de quantis de vazões a partir de modelos de transformação chuva-vazão com emprego conjunto de um gerador estocástico de precipitação, abordagem Bayesiana de inferência estatística e distribuições limitadas superiormente.

2.2 Objetivos específicos

- Avaliar diversas abordagens para a geração estocástica de chuva diária;
- Desenvolver um método para construção de distribuições *a priori* informativas para o limite superior de precipitação diária, com base na variabilidade da PMP;
- Comparar os resultados obtidos para as distribuições limitadas superiormente com aqueles obtidos para distribuições ilimitadas na geração estocástica de precipitação;
- Avaliar os resultados obtidos com a metodologia proposta em relação a outros métodos de estimação de cheias;
- Quantificar as incertezas relacionadas à simulação hidrológica e na estimação dos quantis de cheias; e
- Avaliar as vantagens da abordagem Bayesiana frente à abordagem frequentista.

3 REVISÃO DE LITERATURA

3.1 Geração de séries sintéticas de precipitação

3.1.1 Introdução

Séries suficientemente longas de registros de precipitação diária são necessárias para estudos de gestão de recursos hídricos, estabelecimento de diretrizes de projeto e avaliação de risco de grandes estruturas hidráulicas, tais como vertedores e estruturas de contenção de cheias, além da caracterização e predição de secas. Entretanto, as amostras disponíveis são, em geral, de tamanho reduzido, o que torna difícil a modelagem de processos hidrológicos associados a essa variável e introduz grandes incertezas à estimação de quantis de chuva, em especial aqueles com menor probabilidade de superação.

Uma alternativa para se contornar a disponibilidade limitada de dados é extrair da amostra características como média, desvio padrão, coeficiente de assimetria e autocorrelação, e utilizá-las para a geração de séries sintéticas mais longas, que guardem semelhança estatística com a série original, e, ao mesmo tempo, permitam a observação de um maior número de eventos raros, possibilitando uma caracterização mais precisa da cauda superior da distribuição das precipitações máximas anuais. Em um contexto mais amplo, uma vez que as séries geradas são estatisticamente similares à amostra, as mesmas podem ser interpretadas como realizações do fenômeno tão prováveis quanto a série original, e assim serem empregadas, em conjunto com um modelo de simulação hidrológica, para avaliar a variabilidade de respostas de uma dada bacia hidrográfica (WILKS & WILBY, 1999).

A ideia de se gerar séries de precipitação estocasticamente não é recente. Já na década de 1950, pesquisadores como Williams (1952) e Longley (1953) propuseram modelos com esse propósito. Contudo, somente se introduziu um enfoque estatístico formal à abordagem do problema na década seguinte, com a publicação do trabalho de Gabriel & Newman (1962). O modelo proposto por esses autores foi o ponto de partida para o desenvolvimento de uma enorme gama de técnicas de simulação, incluindo modelos de geração multivariada (Richardson, 1981, 1985; Nonhebel, 1994; Semenov & Barrow, 1997; Semenov *et al.*, 1998; Buishand & Brandsma, 2001; Furrer & Katz, 2007; Kim *et al.*, 2012), nos quais outras variáveis climáticas são simuladas na análise, condicionadas

à ocorrência de precipitação, e os chamados modelos “*multisite*”, que preservam a correlação espacial tanto na ocorrência de chuva quanto nos volumes precipitados no processo de geração (Wilks, 1998; Buishand & Brandsma, 2001; Srikanthan, 2005; Mehrotra *et al.*, 2006; Thompson *et al.*, 2007).

Sob uma perspectiva de modelagem estatística, as séries de precipitação diária exibem duas características que tornam sua descrição especialmente complexa. A primeira delas diz respeito à presença de valores nulos ao longo do período de observação dos dados, o que resulta na ocorrência de uma descontinuidade entre estes e os valores não nulos na distribuição de probabilidades modela as alturas de chuva, seja ela teórica ou empírica (WILKS & WILBY, 1999). A segunda característica é relacionada à persistência de eventos chuvosos e secos, que impõe relações de dependência temporal aos estados de precipitação em dias sucessivos. Geradores estocásticos devem ser capazes de acomodar essas duas características em sua estrutura.

Para contornar problemas de descontinuidade, a grande maioria dos geradores trata separadamente a ocorrência de chuva e a estimação dos volumes precipitados. Nesse contexto, a distribuição de probabilidades que modela as alturas de chuva, seja ela teórica ou empírica, somente se aplica àqueles valores não nulos dessa variável. O problema de persistência, por sua vez, pode ser abordado por meio da utilização de processos estocásticos de estados discretos, condicionados a estados precedentes, tais como cadeias de Markov, ou a partir de modelos de alternância de intervalos de dias secos e chuvosos, os quais procuram reproduzir as frequências amostrais de ocorrência dos diversos comprimentos dessas grandezas (WOOLHISER & ROLDÁN, 1982a).

No que diz respeito à simulação das alturas de chuva, duas abordagens merecem destaque: a abordagem paramétrica, que faz uso de modelos probabilísticos teóricos, e a abordagem não paramétrica, que emprega métodos de reamostragem para estimar a distribuição empírica dos dados. Modelos paramétricos de uso corrente incluem as distribuições exponencial, gama de dois parâmetros, normal assimétrica e normal truncada. Já na abordagem não paramétrica grande parte dos modelos emprega variações da técnica de “*bootstrap*” em conjunto com estimadores de núcleo, ou “*kernel*” (SRIKANTHAN & McMAHON, 2000).

Pesquisas recentes envolvendo geradores estocásticos de precipitação têm conferido amplo destaque à modelagem de eventos extremos (SHARIF & BURN, 2006; LI *et al.*, 2012, 2013; GOYAL *et al.*, 2013). Com efeito, as técnicas de simulação estabelecidas na literatura apresentam limitações para reproduzir acuradamente aqueles quantis com menor probabilidade de superação. Modelos com cauda superior exponencial, em geral, subestimam os extremos amostrais (LI *et al.*, 2012). Modelos híbridos com cauda superior do tipo potência, por outro lado, podem gerar alturas de chuva extremamente elevadas, inconsistentes com os processos físicos de formação de tormentas severas na bacia hidrográfica (LI *et al.*, 2013; CHEN & BRISSETE, 2014). Por fim, modelos de reamostragem possuem capacidade reduzida de extrapolação (Markovich, 2007), e, assim, são pouco efetivos na caracterização da cauda superior das precipitações máximas, especialmente na simulação de séries longas.

No tópico seguinte é apresentada uma compilação das principais técnicas de geração estocástica de chuva, contemplando as abordagens paramétrica e não paramétrica para estimação das alturas precipitadas. Modelos de geração multivariada e “*multisite*” não são mencionados, uma vez que as especificidades dos mesmos fogem ao escopo da presente pesquisa.

3.1.2 Modelos para geração estocástica de precipitação diária

3.1.2.1 Modelos bipartidos

Modelos bipartidos constituem os modelos mais simples e difundidos para a geração estocástica de chuva (SRIKANTAHAN & McMAHON, 2000). Em termos gerais, tais modelos compreendem duas fases em sua estrutura: na primeira delas é simulada a ocorrência de precipitação, enquanto na segunda modelam-se as alturas de chuva com auxílio de uma distribuição de probabilidades teórica. A sazonalidade verificada nas séries de precipitação diária é usualmente acomodada na estrutura do modelo por meio da estimação dos parâmetros das duas fases descritas em intervalos de tempo mensais ou de menor duração.

Modelos para ocorrência de chuva

De acordo com Srikanthan & McMahon (2000), os modelos mais difundidos para a ocorrência de precipitação são aqueles baseados em cadeias de Markov e aqueles baseados em processos de renovação alternada. Cadeias de Markov são processos

estocásticos nos quais as probabilidades de transição entre os diversos estados discretos que compõem o espaço de estados, no passo de tempo atual, são condicionadas aos estados de um certo número de passos de tempo precedentes. O uso de cadeias de Markov na modelagem de ocorrência de chuva foi inicialmente sugerido por Gabriel & Newman (1962).

De acordo com Srikanthan & McMahon (2000), a imensa maioria dos modelos de cadeia de Markov discutidos na literatura (Katz 1974, 1977; Richardson, 1981; Stern & Coe, 1984; Wilks, 1998) é de primeira ordem e esses modelos, em geral, são suficientes para reproduzir as mudanças entre estados de precipitação. Entretanto, Wilks & Wilby (1999) argumentam que, para condições climáticas peculiares, caracterizadas por períodos excepcionalmente longos de seca ou precipitações extremas, cadeias de Markov de primeira ordem são capazes de simular adequadamente o número médio de dias chuvosos, mas se mostram ineficazes na reprodução de todo o espectro de comprimentos dos intervalos de dias secos e chuvosos observados na amostra.

Essa deficiência pode ser contornada com a adoção de cadeias de Markov de ordem superior, tal como sugerido por Chin (1977), Mimikou (1983) e Wilks (1999). Esses trabalhos mostraram que especificidades climáticas das áreas de estudo podem ter forte influência na definição da ordem das cadeias de Markov, implicando na necessidade de condicionar o estado atual de ocorrência de precipitação a um número maior de estados precedentes para se capturar adequadamente as características de variabilidade de intervalos de dias secos e chuvosos. Contudo, segundo Wilks & Wilby (1999), essa abordagem apresenta graves problemas relacionados à modelagem, devido ao elevado número de parâmetros a serem estimados à medida que a ordem da cadeia cresce. Além disso, Hutchinson (1987) e Gregory *et al.* (1993) advogam que a estimação paramétrica para cadeias de maior ordem constitui um expediente pouco confiável, em especial para amostras de tamanho reduzido.

Uma alternativa ao uso de cadeias de Markov são os modelos de renovação alternada. Nesses modelos, a ocorrência de chuva não é modelada com base nos estados de dias sucessivos, mas a partir da construção e alternância de intervalos de dias secos e chuvosos. A ideia básica do método, atribuída a Green (1964), é a de se ajustar distribuições de probabilidade aos diversos comprimentos de intervalos secos e chuvosos observados na amostra, e então construir sequências de eventos que

reproduzam as frequências de ocorrência de cada comprimento, ou, em outras palavras, que caracterizem de maneira adequada os diversos níveis de persistência de dias secos e chuvosos. Em função das características distintas dos tipos de evento, diferentes distribuições podem ser ajustadas aos comprimentos de intervalos de dias secos e úmidos.

Modelos de renovação alternada incluem a distribuição binomial negativa truncada (Buishand, 1978; Woolhiser & Roldán, 1982a), distribuição geométrica truncada (Woolhiser & Roldán, 1982a), a distribuição gama (Small & Morgan, 1986), a distribuição geométrica (Foufoula-Georgiou & Lettenmainer, 1987), a distribuição geométrica mista (Rackso & Szeidl, 1991) e a distribuição binomial negativa (Wilby *et al.*, 1998; Wilks, 1999). De acordo com Wilks & Wilby (1999), o uso de modelos mais complexos, como os exemplificados acima, pode resultar na simulação mais eficiente de períodos de seca com maior duração, em comparação aos modelos de Markov. Contudo, sustentam os autores, os primeiros são sujeitos a maiores incertezas, oriundas da escolha do modelo distributivo e do método de estimação paramétrica, o que pode acarretar um desempenho especialmente ruim na representação de intervalos de maior comprimento para amostras de tamanho reduzido.

Modelos para estimação das alturas precipitadas

A segunda etapa da simulação dos modelos bipartidos compreende o estabelecimento de um modelo probabilístico contínuo para simulação das alturas precipitadas nos dias chuvosos. De acordo com Wilks & Wilby (1999), a característica mais evidente das amostras de alturas de precipitação diária é que as mesmas são fortemente assimétricas à direita, característica essa que deve ser verificada na distribuição de probabilidades selecionada para a modelagem.

Modelos distributivos usuais no referido contexto incluem a distribuição exponencial (Todorovic & Woolhiser, 1975; Richardson, 1981; Woolhiser & Roldán, 1982b), a distribuição exponencial mista de três parâmetros (Woolhiser & Roldán, 1982b; Chapman, 1994; Detzel, 2009), a distribuição gama de dois parâmetros (Richardson, 1981; Woolhiser & Roldán, 1982b; Stern & Coe, 1984; Wilks, 1992; Chapman, 1994; Detzel, 2009) e a distribuição normal assimétrica (Chapman, 1994).

Apesar da simplicidade de aplicação, alguns problemas têm sido associados aos modelos bipartidos. Em primeiro lugar, essa classe de modelos, ainda que modele acuradamente as estatísticas diárias das séries de precipitação, se mostra inapta a reproduzir as variâncias amostrais para intervalos de tempo de maior duração, tais como mensais ou anuais (SRIKANTHAN & McMAHON, 2000). Esse fato decorre das reduzidas probabilidades de geração de longos intervalos secos ou úmidos associadas aos modelos de ocorrência de chuva usualmente empregados (WILKS & WILBY, 1999).

Além disso, para uma dada sequência de dias chuvosos, as alturas precipitadas geradas em cada dia são independentes e modeladas por uma mesma distribuição de probabilidades (WILKS & WILBY, 1999; FURRER & KATZ, 2008). Essa estrutura se mostra conveniente em razão do reduzido número de parâmetros a serem estimados. No entanto, a mesma é inapta a reproduzir a correlação existente entre as alturas de chuva em um dado intervalo, e, dessa forma, a persistência usualmente verificada na ocorrência de eventos extremos não é capturada por esse tipo de modelo, o que pode comprometer as estimativas de cheias quando da simulação hidrológica.

Por fim, estudos como o de Furrer & Katz (2008), Hundecha *et al.* (2009) e Li *et al.* (2012) mostram que modelos com cauda exponencial, tais como os descritos anteriormente, são aptos a representar o núcleo da distribuição das precipitações, mas subestimam consistentemente aqueles eventos mais extremos. De fato, Papalexiou & Koutsouyannis (2012) e Papalexiou *et al.* (2013) demonstraram que as alturas de precipitação diária mais elevadas são melhor descritas por modelos probabilísticos dotados de cauda superior pesada, tais como a distribuição generalizada de Pareto com parâmetro de forma negativo. Esses últimos, porém, não se prestam à modelagem das chuvas com maior probabilidade de superação. Diante disso, o uso de modelos híbridos vem sendo apontado como a alternativa mais coerente para abordar o problema (Furrer & Katz, 2008; Hundecha *et al.*, 2009; Li *et al.*, 2012, Papalexiou *et al.*, 2013, Li *et al.*, 2013), como será visto a seguir.

3.1.2.2 Modelos híbridos

Nesse caso, uma distribuição assimétrica à direita com cauda superior exponencial é utilizada para modelar chuvas moderadas e uma distribuição com cauda suficientemente pesada é empregada na modelagem dos eventos cujos valores excedem um determinado

limiar. Essa abordagem assume de maneira explícita que chuvas moderadas e extremas compreendem processos físicos distintos e, assim, as alturas de precipitação associadas a cada um desses processos são amostradas de populações diferentes (FURRER & KATZ, 2008). Tal estrutura acarreta uma melhora significativa nas estimativas de quantis mais raros em comparação aos modelos bipartidos convencionais (HUNDECHA *et al.*, 2009; LI *et al.*, 2012).

Ainda que constitua uma alternativa promissora para a simulação de precipitações extremas, o uso de modelos híbridos incorre em alguns problemas. O primeiro e mais crítico deles é a definição do limiar entre chuvas moderadas e extremas (PAPALEXIOU *et al.*, 2013). De fato, sob uma perspectiva estritamente física, é extremamente difícil estabelecer um valor que diferencie essas classes de alturas de precipitação. Em um contexto matemático, por outro lado, a imposição de um limiar arbitrário pode ocasionar uma descontinuidade entre os modelos distributivos nesse ponto específico, o que torna inviável a estimação de parâmetros por métodos de uso corrente, tais como o de máxima verossimilhança (FURRER & KATZ, 2008). Nesse sentido, um número restrito de modelos, tais como o proposto por Li *et al.* (2012), oferecem soluções analíticas para a determinação do limiar, e autores como Furrer & Katz (2008) e Chen & Brissette (2014) sugerem a calibração do gerador para esse propósito. Já Hundecha *et al.* (2009) sugerem que, em lugar da fixação de limiares, a estrutura do modelo híbrido seja baseada em uma mistura de distribuições, na qual a função de ponderação confere maior peso ao modelo com cauda superior pesada à medida que os valores das alturas de precipitação aumentam.

O segundo problema é relacionado à extrapolação aos quantis com probabilidade de superação muito reduzida. Com efeito, o emprego de distribuições com caudas superiores muito pesadas pode acarretar a geração de valores fisicamente implausíveis com relativa frequência (LI *et al.*, 2013; CHEN & BRISSETTE, 2014), o que compromete de maneira severa as estimativas de cheias extremas na simulação hidrológica.

Por fim, tal como os modelos bipartidos, os modelos híbridos são inaptos a reproduzir tanto a correlação temporal dos eventos extremos quanto as variâncias amostrais em escalas mensal e anual (CHEN & BRISSETTE, 2014), ainda que sejam bastante efetivos na simulação de estatísticas diárias.

3.1.2.3 Modelos de matriz de probabilidade de transição

Modelos de matriz de probabilidade de transição (MPT) constituem uma extensão natural dos modelos bipartidos, tendo-se em vista que, nesse tipo de modelo, a dicotomia chuva-ausência de chuva é substituída por um conjunto com maior número de estados discretos, cada qual associado a uma classe de alturas de precipitação. Tal expediente permite que a correlação entre os volumes precipitados em dias sucessivos seja preservada ao longo da simulação (GREGORY *et al.*, 1993; WILKS & WILBY, 1999). Ademais, a estrutura do modelo possibilita a reprodução mais acurada das alturas de precipitação intermediárias, tendo-se em vista que cada estado é descrito por uma distribuição com vetor paramétrico específico (CHAPMAN, 2001).

O modelo MPT foi introduzido por Haan *et al.* (1976), e desde então tem sido empregado com modificações pontuais por autores como Selvalingam & Miura (1978), Srikanthan & McMahon (1985), Boughton & Hill (1997), Boughton (1999), Lima (2004) e Gontijo (2009). As principais distinções entre os estudos citados residiram no número de estados discretos que compunham as matrizes de probabilidade de transição e o modelo distributivo empregado no último estado das mesmas, os quais compreenderam distribuições com cauda superior exponencial (Haan *et al.*, 1976), distribuições com cauda superior pesada (Lima, 2004), e uma distribuição superiormente limitada (Boughton & Hill, 1997; Boughton, 1999).

Wilks & Wilby (1999) argumentam que, em função do elevado número de parâmetros a serem estimados, as vantagens dos modelos MPT frente aos modelos bipartidos e híbridos somente se manifestam mediante a correta especificação dos limites dos intervalos de alturas de chuva. Contudo, tal especificação constitui um dos grandes entraves à aplicação do modelo. Com efeito, nenhum critério objetivo com esse propósito foi encontrado na literatura, e, nas aplicações discutidas, a mesma foi realizada de maneira empírica. Técnicas como análise de *clusters* podem ser usadas nesse contexto, mas a distribuição empírica das alturas de chuva pode não ser reproduzida adequadamente (LIMA, 2004).

Além dessa limitação, modelos MPT, tal como aqueles discutidos anteriormente, são inaptos a reproduzir as variâncias observadas em escala mensal e anual (BOUGHTON, 1999; SRIKANTHAN & McMAHON, 2000), em função das pequenas probabilidades de geração de longos intervalos secos ou úmidos inerentes à sua estrutura.

Já com relação aos extremos, modelos MPT podem apresentar as mesmas deficiências associadas a modelos bipartidos e híbridos, em função do tipo de distribuição adotado na modelagem do último estado da matriz de probabilidades de transição (CHEN & BRISSETTE, 2014). Nesse sentido, uma distribuição com cauda superior exponencial pode acarretar a subestimação dos eventos extremos, ao passo que distribuições com cauda potencial podem levar à geração de valores fisicamente implausíveis.

3.1.2.4 Modelos não paramétricos

Modelos não paramétricos simulam as alturas de chuva por meio de técnicas de reamostragem, tais como o “*bootstrap*” e suas variações. Esse tipo de modelo é especialmente interessante por excluir da análise as incertezas acerca da seleção do modelo distributivo do qual se supõe que a amostra tenha sido retirada e da estimação dos parâmetros do mesmo (SHARIF & BURN, 2006). Em adição, esse tipo de modelo é apto a representar as mais variadas formas assumidas pela distribuição empírica dos dados, acomodando características amostrais peculiares, tais como a bimodalidade (APIPATTANAVIS *et al.*, 2007).

Dentre as técnicas não paramétricas de estimação de densidades, aquela mais difundida no contexto de geração estocástica de precipitação diária é o “*bootstrap*” do k-ésimo vizinho mais próximo (Lall *et al.*, 1996; Rajagopalan *et al.*, 1996; Lall & Sharma, 1996; Buishand & Brandsma, 2001; Yates *et al.*, 2003; Apipattanavis *et al.*, 2007; Basinger *et al.*, 2010). Modelos baseados na referida técnica são especialmente hábeis na reprodução de estatísticas mensais e anuais, além daquelas em escala diária, as quais são adequadamente simuladas também por modelos paramétricos (CHEN & BRISSETTE, 2014). Mais do que isso, esse tipo de modelo não paramétrico geralmente se mostra apto a preservar a correlação entre as alturas de chuva em dias sucessivos, o que é de grande interesse no contexto de simulação hidrológica (WILKS & WILBY, 1999).

Por outro lado, essa classe de modelos de simulação, em geral, não traz à análise informações distintas daquelas oriundas da amostra, o que restringe severamente sua capacidade de extrapolação. Técnicas alternativas nesse contexto, tais como a adição de perturbações às alturas de chuva (Sharif & Burn, 2006; Goyal *et al.*, 2013), têm sido empregadas com intuito de se contornar essa deficiência. Entretanto, ainda que este procedimento pareça conveniente, Furrer & Katz (2008) argumentam que não foram encontradas na literatura avaliações mais conclusivas sobre sua eficiência na

representação da cauda superior das precipitações. Além disso, Markovich (2007) sustenta que, ainda que alguma extrapolação seja possível, as probabilidades de geração de eventos muito raros são extremamente reduzidas, tendo-se em vista que os estimadores de núcleo usualmente empregados com o referido propósito são dotados de cauda superior exponencial.

3.2 Distribuições limitadas superiormente

3.2.1 O uso de distribuições superiormente limitadas em hidrologia: fundamentação física e estimação paramétrica

A descrição probabilística de eventos extremos de precipitação e vazão por meio de modelos distributivos superiormente limitados constitui um dos temas mais controversos das ciências hidrológicas. Tal controvérsia se manifesta principalmente em duas vertentes. A primeira delas diz respeito à existência de limites superiores para as alturas de chuva e magnitudes de enchentes, determinados com base em considerações físicas acerca dos mecanismos de formação de tormentas e da resposta hidrológica da bacia. A segunda vertente compreende a adequação das ferramentas estatísticas disponíveis para a estimação do limite superior, assumindo-se que o mesmo exista.

A existência de um limite superior para precipitações e vazões tem sido advogada desde o estabelecimento dos primeiros modelos empíricos para a estimação de cheias, tais como as curvas envoltórias de Jarvis (1925), a envoltória de Creager e o modelo de análise de frequência de Horton (1936). A esse último autor, inclusive, são atribuídas declarações especialmente favoráveis a essa hipótese, segundo as quais “... *nenhum corpo d’água terrestre pode produzir uma cheia infinita*”, e “... *a magnitude das cheias sempre aumenta com o período de retorno, mas em direção a um limite finito e não na direção do infinito*”.

Inserem-se nesse contexto também modelos de natureza determinística, tais como a PMP e a EMP (WMO, 1986), os quais se valem de forte fundamentação física, incorporada em sua estrutura por meio das melhores informações meteorológicas e hidrológicas disponíveis, para argumentar favoravelmente à existência de limites superiores de precipitação e vazão, respectivamente.

Por outro lado, um amplo grupo de autores se opõe com veemência ao uso de modelos limitados (Yevjevich, 1968; Klemes, 1987; Yevjevich & Harmancioglu, 1987, Hosking

& Wallis, 1997; Papalexiou & Koutsouyiannis, 2006), argumentando que tais modelos não são dotados de fundamentação física suficiente para se determinar com exatidão o valor do limite superior de precipitações e vazões e sustentando que, de um ponto de vista matemático, a imposição desse limite pode comprometer severamente as estimativas de quantis associados a tempos de retorno de interesse para o projeto da grande maioria das estruturas hidráulicas. De fato, a qualidade de ajuste de modelo limitado é intrinsecamente relacionada à posição do limite superior no domínio da variável hidrológica em estudo (Gosh & Resnick, 2010), a qual, em última análise, determina a maneira como a curva de quantis se comporta à medida que tais variáveis se aproximam do valor máximo.

Em face dessas considerações, fica evidente que, com base em fundamentos estritamente matemáticos, é extremamente difícil aceitar de maneira definitiva a existência de um limite superior para variáveis hidrológicas. Contudo, hidrólogos e engenheiros frequentemente se deparam com a necessidade de prescrever valores máximos práticos para tais variáveis em projetos de grandes estruturas para aproveitamentos hídricos e contenção de cheias. Dessa maneira, alguns autores têm recorrido a argumentos físicos que forneçam subsídios consistentes à adoção de modelos limitados.

Sob tal perspectiva, Enzel *et al.* (1993) construíram a curva envoltória de vazões para o rio Colorado, considerando-se duas situações distintas. A primeira delas contemplou somente o uso de dados sistemáticos, os quais abrangiam um período de 70 anos. Na segunda, foram incluídas 25 paleocheias à análise, cuja ocorrência remonta a até 4.000 anos. Os autores verificaram que a inclusão das paleocheias não alterou a curva envoltória, o que sugere que a mesma tem se mantido estável ao longo desse período. Esse fato, sustentam os autores, é uma evidência de que existe um limite físico finito para as enchentes na bacia.

Ainda nesse contexto, Jacoby *et al.* (2008), por meio de vestígios de cheias extremas obtidos em árvores e depósitos aluviais no rio Nahal Arava, em Israel, chegaram a conclusões similares à de Enzel *et al.* (1993): a inclusão de dados não sistemáticos não eleva a curva envoltória de vazões, o que corrobora a existência de um limite superior finito para as enchentes. Ainda que aplicados ao estudo de cheias, esses dois trabalhos constituem a fundamentação física para o uso de modelos probabilísticos limitados para

precipitações máximas na presente pesquisa, tendo-se em vista que as enchentes máximas somente serão limitadas se o suprimento de água precipitada na bacia também o for.

Assumindo-se a existência de um limite superior para variáveis hidrológicas, impõe-se à análise o problema de estimação dos parâmetros dos modelos distributivos. Métodos estatísticos convencionais, tais como o de máxima verossimilhança, são fortemente dependentes do tamanho da amostra (Naghetini & Pinto, 2007), o que pode agregar imensas incertezas à inferência caso esse tamanho seja reduzido, tornando a estimativa desse parâmetro uma mera abstração matemática. Além disso, é reconhecido na literatura (Slade, 1936; Botero, 2006) que as estimativas do limite superior explícito obtidas com o método de máxima verossimilhança não diferem muito do valor máximo amostral. Tendo-se em vista que, via de regra, os períodos de monitoramento pluviométrico e fluviométrico são pouco extensos, não é razoável supor que as condições mais severas de formação de tormentas ou enchentes tenham sido registradas, o que inviabiliza o emprego desse método para a estimação paramétrica.

Como alternativa, autores como Takara & Tosa (1999) e Botero (2006) sugerem o uso de estimativas de PMP e EMP como um valor fixo para o limite superior de precipitações e vazões. Esse expediente, contudo, é questionado por grande número de estudiosos, tais como Dingman (1994), Koutsoyiannis (1999) e Papalexou & Koutsoyiannis (2006), uma vez que estimativas dessas variáveis dependem das amostras disponíveis e das ferramentas de estimação, o que lhes confere uma natureza “quase-determinística” (FERNANDES *et al.*, 2010). Em outras palavras, em função das incertezas na estimação dessas variáveis, uma probabilidade de excedência não-nula é associada às mesmas, o que contradiz seu conceito de limite superior.

Outro problema com esse raciocínio é que, com frequência, as estimativas de PMP e EMP são superadas em períodos de observação relativamente curtos, mesmo em bacias dotadas de extensas redes de monitoramento. Tal situação é exemplificada no relatório hidrometeorológico HMR59 (USACE, 1999), no qual diferenças da ordem de 22% são verificadas nas estimativas de PMP diária, considerando-se um período de 38 anos.

O paradigma Bayesiano oferece uma abordagem coerente para esse problema. Diferentemente da estatística convencional, esse paradigma oferece uma estrutura lógica

para resumir as incertezas relacionadas ao limite superior, por meio da especificação de uma distribuição *a priori* informativa, a qual agrega o conhecimento prévio acerca desse parâmetro com base na caracterização dos fenômenos físicos de formação de tormentas ou cheias em condições extremas. Nesse contexto, estimadores para o limite superior, como a PMP e a EMP, são incorporados ao modelo não como valores fixos, mas em uma descrição probabilística completa, que acomoda as incertezas oriundas de sua estimação e, assim, a probabilidade de superação inerente a tais estimadores, a qual certamente refletirá na inferência do limite superior, pode ser incluída na análise.

O próximo tópico da presente seção tratará de alguns modelos distributivos dotados de limite superior explícito. A abordagem Bayesiana de inferência, por sua vez, será abordada em detalhes na seção 3.3.

3.2.2 Modelos distributivos dotados de limite superior explícito

Dentre as distribuições de probabilidade frequentemente utilizadas por hidrólogos para modelar variáveis relacionadas à precipitação e vazão, algumas podem apresentar limite superior. Elas incluem distribuições de 3, 4 ou mesmo 5 parâmetros.

A distribuição generalizada de valores extremos de 3 parâmetros (GEV), por exemplo, apresentará limite superior se seu parâmetro de forma for positivo, o que irá ocorrer quando seu coeficiente de assimetria for negativo ou mesmo se for positivo e inferior a 1,1396. Também a distribuição LogPearson do tipo III será limitada superiormente se seu coeficiente de assimetria for negativo. Essa característica particular das referidas distribuições restringe severamente o uso das mesmas como modelos limitados gerais, uma vez que as variáveis hidrológicas relacionadas a máximos são habitualmente assimétricas à direita, com coeficientes de assimetria geralmente superiores a 1,1396.

Outros modelos possivelmente limitados são a distribuição Kappa de 4 parâmetros (Hosking & Wallis, 1997), e a distribuição Wakeby de 5 parâmetros (Houghton, 1978). Essas duas distribuições, no entanto, podem acomodar formas distributivas distintas, as quais incluem modelos ilimitados. Dessa forma, a existência de um limite superior finito fica condicionada a características amostrais, traduzidas em uma combinação particular de valores de parâmetros, o que dificulta, se não inviabiliza, o estabelecimento de uma relação física clara entre o referido limite e a bacia hidrográfica.

Por outro lado, existem distribuições que incorporam o limite superior de maneira explícita em sua formulação. Para essas distribuições é possível associar o parâmetro citado aos processos de formação de precipitações e enchentes extremas, o que constitui uma propriedade de grande interesse em um contexto de modelagem estatística, uma vez que confere fundamentação física aos modelos distributivos. Dentre tais distribuições merecem destaque a distribuição de valores extremos do tipo IV (EV4), a distribuição log-normal de 4 parâmetros (LN4) e a distribuição *Transformed Distribution function* (TDF), as quais são apresentadas em detalhes nos tópicos seguintes.

3.2.2.1 Distribuição EV4

A distribuição de valores extremos do tipo IV (EV4) foi introduzida por Kanda (1981) e empregada na modelagem de terremotos e velocidade de ventos no Japão. Essa não é uma distribuição de valores extremos propriamente dita, uma vez que não deriva da teoria do valores extremos. Sua denominação é devida ao fato de sua forma paramétrica ter sido proposta como uma modificação da distribuição de valores extremos do tipo III, ou EV3.

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

$$F_X(x) = \exp\left\{-\left[\frac{\alpha - x}{\sigma(x - \varepsilon)}\right]^\xi\right\}, \quad \varepsilon < x \leq \alpha \quad (3.1)$$

A função densidade de probabilidade da EV4 é dada por:

$$f_X(x) = \frac{\xi(\alpha - x)^{\xi-1}(\alpha - \varepsilon)}{\sigma^\xi(x - \varepsilon)^{\xi+1}} \exp\left\{-\left[\frac{\alpha - x}{\sigma(x - \varepsilon)}\right]^\xi\right\}, \quad \varepsilon < x \leq \alpha \quad (3.2)$$

Os quantis de ordem $p \in (0,1)$ da EV4 são dados pela seguinte equação:

$$F_X^{-1}(x) = \frac{\alpha + \sigma\varepsilon[-\ln(p)]^{1/\xi}}{\sigma[-\ln(p)]^{1/\xi} + 1} \quad (3.3)$$

Embora a EV4 não seja uma distribuição de valores extremos, pode-se estabelecer um paralelo com esta 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 $\sigma(\alpha - \varepsilon)$.

A Figura 3.1 mostra a influência de cada um dos parâmetros na forma da distribuição. É possível perceber que a EV4 tende a ter caudas superiores mais pesadas para valores pequenos de σ e ξ , e para valores grandes de α .

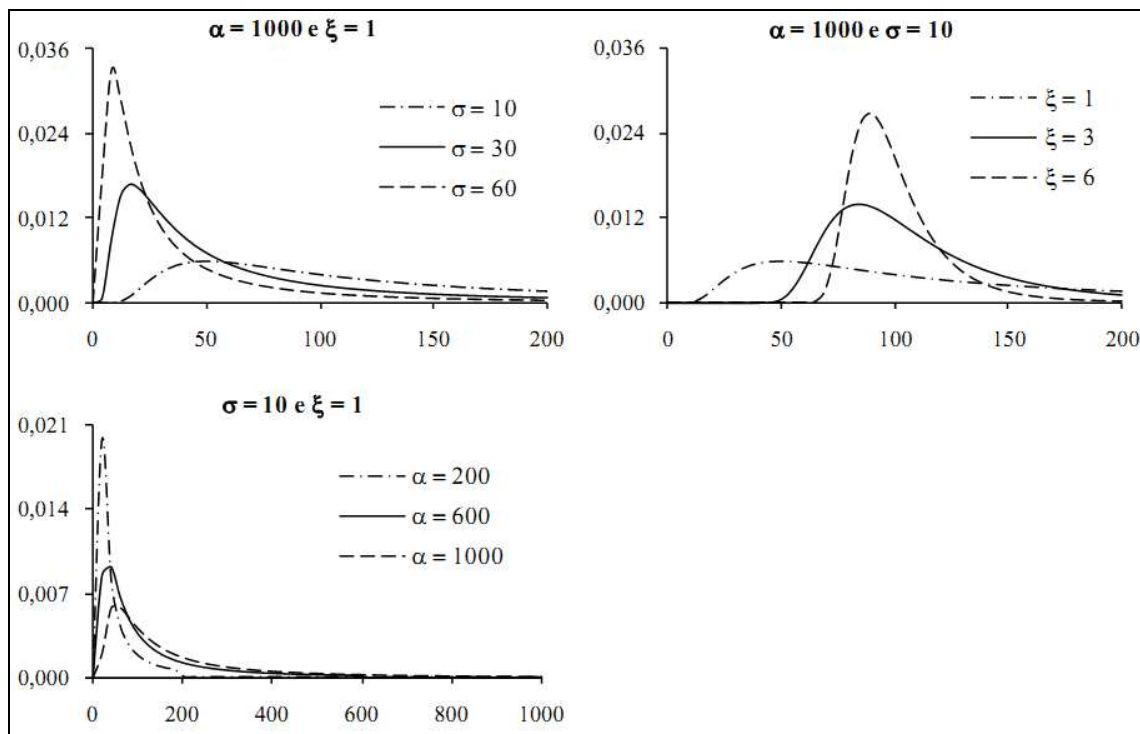


Figura 3.1 – Efeito de cada parâmetro na distribuição LN4

Fonte: Fernandes (2009)

3.2.2.2 Distribuição Log-normal de 4 parâmetros (LN4)

A distribuição LN4 foi proposta por Slade (1936), a partir de uma transformação de uma variável normalmente distribuída. A hipótese básica adotada pelo autor 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.

A primeira utilização da LN4 em hidrologia é atribuída a Takara & Loebis (1996), em um trabalho que contemplou a análise de eventos extremos de precipitação na Indonésia e no Japão. Os autores obtiveram melhores resultados quando comparados àqueles oriundos do uso da distribuição log-normal de 3 parâmetros, a qual não é limitada superiormente. Posteriormente, Takara & Tosa (1999) utilizaram a LN4 na modelagem de eventos extremos de vazão, obtendo conclusões semelhantes às dos estudos de precipitação.

A distribuição LN4 é obtida a partir da seguinte transformação:

$$Y = \ln\left(\frac{X - \varepsilon}{\alpha - X}\right), \quad (3.4)$$

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

$$f_X(x) = \frac{\alpha}{x(\alpha - x)\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2} \left[\ln\left(\frac{x - \varepsilon}{\alpha - x}\right) - \mu\right]^2\right\}, \quad \varepsilon < x < \alpha \quad (3.5)$$

A função acumulada de probabilidades é dada por:

$$F_X(x) = \Phi\left[\frac{1}{\sigma} \ln\left(\frac{x - \varepsilon}{\alpha - x}\right) - \frac{\mu}{\sigma}\right], \quad \varepsilon < x < \alpha \quad (3.6)$$

na qual Φ denota a distribuição normal padrão.

A Figura 3.2 mostra a influência de cada um dos parâmetros na forma da distribuição LN4. É possível perceber que os parâmetros σ e α controlam o peso da cauda superior, sendo que a mesma fica mais pesada à medida que os valores desses parâmetros aumentam. O parâmetro μ controla a assimetria da distribuição. Para

valores negativos de μ , a assimetria é positiva, enquanto que para valores positivos de μ , a assimetria é negativa. No caso de μ ser igual a zero, a distribuição é simétrica.

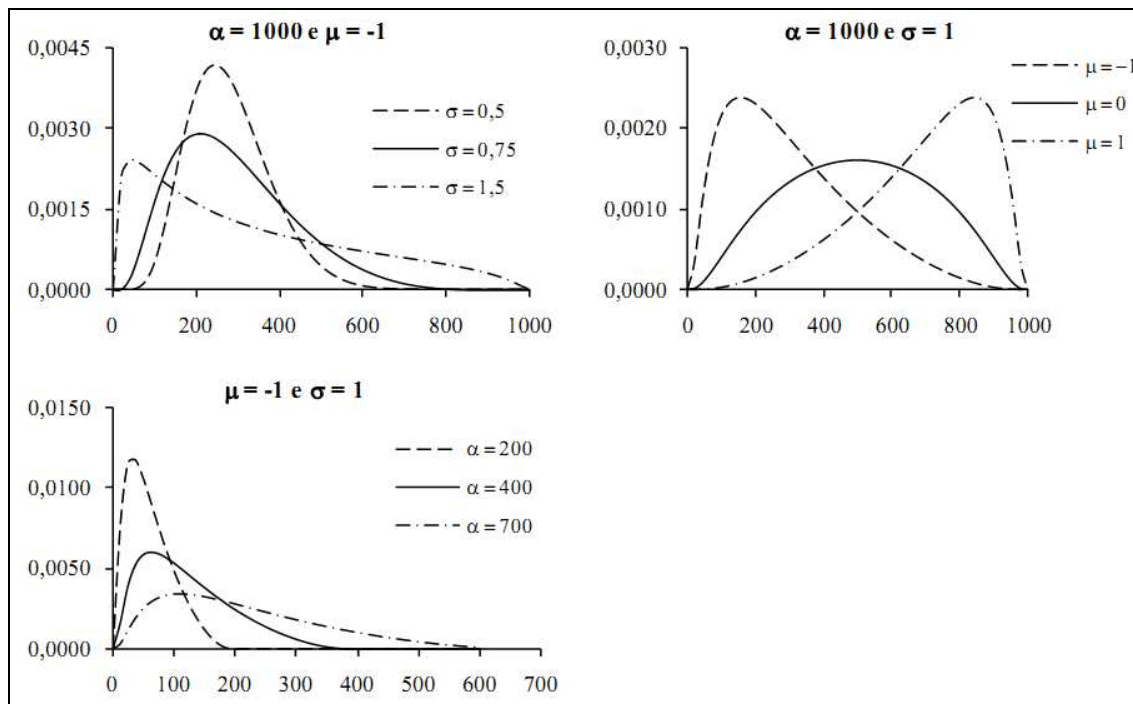


Figura 3.2 – Efeito de cada parâmetro na distribuição LN4

Fonte: Fernandes (2009)

De acordo com Fernandes (2009), não é possível obter uma forma analítica para os momentos da distribuição LN4. Dessa forma, os mesmos devem ser calculados numericamente, a partir das respectivas equações de definição.

3.2.2.3 Distribuição Transformed Distribution Function (TDF)

A distribuição TDF foi introduzida por Elíasson (1994) para a análise de frequência de precipitações máximas. Em estudo posterior, Elíasson (1997) utilizou a referida distribuição na análise de frequência de precipitações máximas de 24 horas de duração nos Estados Unidos e na Islândia.

A TDF pode ser construída da seguinte maneira: seja Y uma variável superiormente ilimitada distribuída segundo o modelo Gumbel. A função acumulada de probabilidades de Y , 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:

$$F_Y(y) = \exp\left\{-\exp\left\{-\left(\frac{y-\mu}{\varepsilon}\right)\right\}\right\}, \quad y \in \mathfrak{R} \quad (3.7)$$

Elíasson (1997) propõe a seguinte transformação para Y :

$$Y = X - \frac{\sigma^2 \beta}{\alpha - X} \quad (3.8)$$

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 3.7 e 3.8: se Y é distribuída de acordo com o modelo Gumbel, então X é distribuída segundo a TDF.

A função densidade de probabilidades de uma variável distribuída de acordo com o modelo TDF, denotada por $X \sim \text{TDF}(\mu, \sigma, \beta, \alpha)$, é dada pela seguinte equação:

$$F_X(x) = \exp\left\{-\exp\left\{-\frac{x}{\sigma} + \frac{\sigma\beta}{\alpha-x} + \frac{\mu}{\sigma}\right\}\right\}, \quad 0 < x < \alpha \quad (3.9)$$

A função densidade de probabilidade de TDF é dada por:

$$f_X(x) = \exp\left\{-\exp\left\{-\frac{x}{\sigma} + \frac{\sigma\beta}{\alpha-x} + \frac{\mu}{\sigma}\right\}\right\} \exp\left\{-\frac{x}{\sigma} + \frac{\sigma\beta}{\alpha-x} + \frac{\mu}{\sigma}\right\} \left[\frac{1}{\sigma} - \frac{\sigma\beta}{(\alpha-x)^2}\right] \quad (3.10)$$

Os quantis de ordem $p \in (0,1)$ podem ser derivados a partir da solução de 3.9 com relação a x . Ao fazê-lo, obtém-se a seguinte equação quadrática:

$$x^2 + C_1 x + C_0 = 0 \quad (3.11)$$

na qual

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

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

Resolvendo-se 3.11 e tomando-se somente o valor positivo de x , tem-se:

$$F_X^{-1}(p) = -\frac{C_0}{2} + \sqrt{\frac{C_0^2}{4} - C_1} \quad (3.12)$$

A Figura 3.3 mostra a influência de cada um dos parâmetros na forma da distribuição TDF.

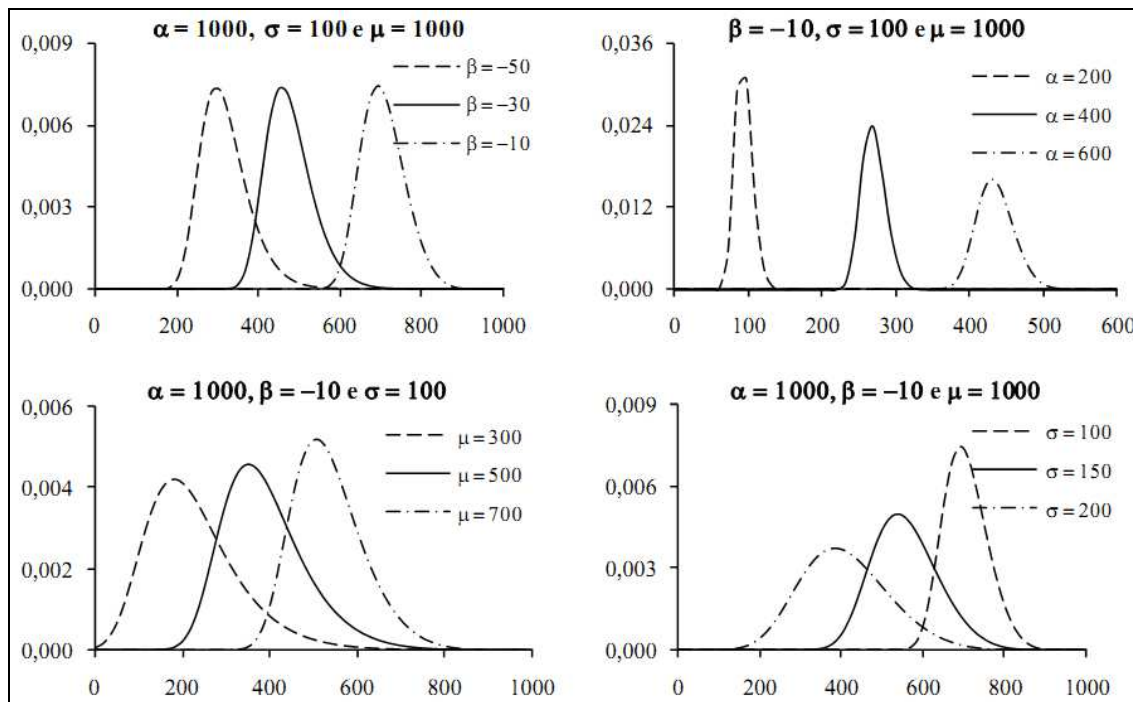


Figura 3.3 – Efeito de cada parâmetro na distribuição TDF

Fonte: Fernandes (2009)

Não é possível obter uma forma analítica para os momentos da distribuição TDF. Dessa forma, os mesmos devem ser encontrados numericamente, a partir das respectivas equações de definição. Além disso, a TDF não apresenta um limite inferior explícito.

3.3 Inferência Bayesiana

3.3.1 Introdução

O emprego dos métodos Bayesianos teve início em 1763, quando o trabalho do Reverendo Thomas Bayes (1702-1761) foi apresentado à Royal Statistical Society. Os conceitos Bayesianos despertaram o interesse de pesquisadores influentes como LaPlace, Gauss e Pearson, dentre outros, e dominaram o pensamento estatístico durante todo o século XIX (BROOKS, 2003).

A partir do início do século XX, no entanto, a abordagem Bayesiana perdeu importância em razão, principalmente, da oposição de pesquisadores como Newman e Fisher, que tinham severas restrições quanto à subjetividade inerente à referida abordagem. Ainda assim, segundo Brooks (2003), nomes como Jeffreys, Savage, Lindley e de Finetti advogavam a favor da corrente Bayesiana, apontando-a como alternativa para contornar deficiências dos métodos frequentistas.

A partir do final da década de 1980, a abordagem Bayesiana ganhou novo impulso, em função do rápido desenvolvimento computacional verificado nessa década e também do desejo de se modelarem processos cada vez mais complexos, os quais não encontravam soluções coerentes na abordagem frequentista (Brooks, 2003). Com a abordagem Bayesiana recebendo cada vez mais atenção nas pesquisas em Estatística, novas ferramentas computacionais foram desenvolvidas e meios mais flexíveis de inferência foram apresentados. Atualmente, a abordagem Bayesiana constitui uma estrutura de análise que vem de encontro à crescente complexidade das pesquisas científicas (FERNANDES, 2009).

Ambas as escolas, frequentista e Bayesiana, entendem que as incertezas sobre objetos aleatórios devem ser avaliadas por meio de probabilidades. No contexto frequentista, entretanto, a medida de probabilidade captura a variabilidade inerente ao processo, ao passo que, no contexto Bayesiano, tal medida reflete o desconhecimento do indivíduo com relação ao objeto em estudo. Reside aí uma das principais distinções entre as escolas clássica e Bayesiana: a maneira como cada uma interpreta os parâmetros de um modelo. Para os frequentistas, o parâmetro θ é uma quantidade fixa (não variável) e seu valor pode ser estimado, por exemplo, pela maximização da função de verossimilhança. Essa abordagem fornece estimativas pontuais do valor do parâmetro. Os Bayesianos, por outro lado, entendem o parâmetro como uma quantidade fixa, porém desconhecida, e que, por ser aleatória, deve ser descrita por uma distribuição de probabilidades, que resume o conhecimento a seu respeito. À medida que aumenta o nível de informação em relação ao parâmetro, espera-se que sua incerteza diminua e, no limite, ao menos em teoria, o total conhecimento do parâmetro implicaria em uma distribuição degenerada, na qual o parâmetro assume um valor único com probabilidade igual a 1.

Outra diferença entre as escolas diz respeito às fontes de informação. Para a corrente frequentista, a única informação a ser considerada é aquela trazida pelos dados. A escola Bayesiana, por outro lado, incorpora outras fontes de informação para resolver problemas de inferência. Formalmente, seja θ um parâmetro com domínio no espaço paramétrico Θ e H a informação ou conhecimento prévio do especialista a respeito de θ . Com base em H , a incerteza sobre θ é resumida pela distribuição a priori $\pi(\theta|H)$,

que descreve o estado de conhecimento sobre a quantidade aleatória antes da observação dos dados.

De maneira geral, H não contém toda informação relevante sobre o parâmetro, e, assim, a distribuição *a priori* não constitui uma boa inferência para θ . Se a informação contida em H não for suficiente, deve-se realizar um experimento para se obter informações adicionais sobre o parâmetro. Seja X uma variável aleatória relacionada a θ passível de ser amostrada. Supondo que o atual valor do parâmetro seja conhecido, a incerteza sobre a quantidade X é resumida pela função de verossimilhança $f(X|\theta, H)$, que fornece a probabilidade de ocorrência de cada amostra particular x de X . Após a realização do experimento, o conhecimento sobre o parâmetro pode ser atualizado com auxílio do teorema de Bayes. De acordo com esse teorema, a distribuição *a posteriori*, que agrega o conhecimento atualizado sobre θ , é dada por:

$$\pi(\theta/x, H) = \frac{f(x/\theta, H)\pi(\theta|H)}{f(x|H)} \quad (3.13)$$

onde a função $f(x|H)$ representa a distribuição preditiva *a priori* e é calculada por:

$$f(x|H) = \int_{\Theta} f(x/\theta, H)\pi(\theta|H)d\theta \quad (3.14)$$

A distribuição *a posteriori* descreve a incerteza sobre o parâmetro após a observação dos dados, ou seja, $\pi(\theta/x, H)$ constitui a inferência *a posteriori* sobre θ , a partir da qual é possível estabelecer qualquer característica de θ .

No caso de precipitações e cheias extremas, engenheiros e hidrólogos necessitam tomar decisões que envolvem eventos extraordinários com base em informações abstratas e incompletas. Nesse contexto, há uma subjetividade intrínseca acerca do evento, que, de maneira distinta da escola clássica, pode ser coerentemente analisada sob a ótica Bayesiana (FERNANDES, 2009). De fato, sob tal paradigma, a subjetividade é avaliada a partir do conhecimento do analista a respeito das características probabilísticas do evento em estudo e, dessa forma, a correta descrição da subjetividade inerente a um evento extremo depende da habilidade do especialista em selecionar, criticar, interpretar e julgar o conjunto de informações existentes sobre o evento (VICK, 2002).

3.3.2 Estimação Bayesiana

Segundo Bernardo & Smith (1994), a estimação Bayesiana é um problema de decisão. Dessa forma, para a estimação pontual de um parâmetro θ a inferência Bayesiana requer a especificação de uma função de perda $L(\delta, \theta)$, que representa a penalidade associada à escolha de δ como estimador de θ . Neste contexto, procura-se adotar um estimador que minimiza o chamado risco de Bayes, definido da seguinte maneira:

$$RB = \iint L(\delta, \theta) f(x|\theta) \pi(\theta) dx d\theta \quad (3.15)$$

onde a perda é integrada em x e θ . Uma inversão na ordem de integração permite avaliar o estimador de θ em termos da perda esperada *a posteriori* (ROBERT & CASELLA, 2004). Assim, o estimador δ_B de θ é tal que:

$$\delta_B = \min_{\delta} E[L(\delta, \theta)|x] = \min_{\delta} \int L(\delta, \theta) \pi(\theta|x) d\theta \quad (3.16)$$

A escolha da função de perda é subjetiva, refletindo o modo como o especialista acha justo ser penalizado por suas decisões. As principais funções de perda empregadas na estimação de parâmetros são:

- Quadrática: a função de perda é dada por $L(\delta, \theta) = (\delta - \theta)^2$ e o estimador de Bayes para θ é a média da distribuição *a posteriori*, admitindo-se que a mesma exista;
- Valor absoluto: a função de perda é dada por $L(\delta, \theta) = |\delta - \theta|$ e o estimador de Bayes para θ é a mediana da distribuição *a posteriori*, admitindo-se que a mesma exista;
- Zero-um: a função de perda é dada por $L(\delta, \theta) = \mathbf{1}_{(\delta \neq \theta)}$, onde $\mathbf{1}_{(a)}$ denota a função indicadora, e o estimador de Bayes para θ é a moda da distribuição *a posteriori*, admitindo-se que a mesma exista.

De acordo com Robert & Casella (2004), dois problemas se apresentam para o cálculo de δ : (1) a distribuição *a posteriori* de θ , $\pi(\theta|x)$, em geral, não apresenta forma analítica fechada; e (2) na maioria dos casos, a integração da equação (3.16) não pode

ser feita analiticamente. Métodos para contornar esses problemas serão apresentados na seção 3.3.4.

3.3.3 Intervalos de credibilidade

Uma importante distinção referente à estimação de parâmetros das abordagens Bayesiana e frequentista é a forma como cada uma avalia as incertezas em relação à escolha do estimador. Na análise frequentista, o problema é tratado por meio do princípio de 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-se ser possível a realização de tal experimento. É também baseada no princípio de repetição da amostra a construção do intervalo de confiança (IC) frequentista. A confiança de um IC é interpretada como o percentual de intervalos que, caso construído de maneira idêntica, conteria o verdadeiro valor do parâmetro. Cabe ressaltar que na escola clássica os parâmetros são vistos como quantidades fixas, para as quais não se atribui probabilidades. Assim, não se pode interpretar a confiança de um IC como sendo a probabilidade do verdadeiro valor de θ pertencer ao referido intervalo. A abordagem Bayesiana, por outro lado, fornece uma estrutura mais natural para se avaliar as incertezas na estimativa dos parâmetros. Com efeito, a variância da distribuição *a posteriori* dos mesmos fornece uma medida direta da incerteza associada a essas quantidades.

Um indicador mais apropriado da incerteza na estimativa dos parâmetros é o intervalo de credibilidade, o equivalente Bayesiano para o intervalo de confiança. O intervalo de credibilidade para um parâmetro é construído com base na distribuição *a posteriori* de θ , e, portanto, leva em consideração apenas a amostra de fato observada. Uma vez que o parâmetro é um objeto aleatório, o intervalo de credibilidade é aquele em que θ está com probabilidade $(1-\alpha)$. Não ocorre aqui nenhuma suposição de repetição da amostra original. Dessa forma, a interpretação do intervalo de credibilidade é bem mais natural que a de seu equivalente frequentista.

O intervalo de credibilidade pode ser construído para qualquer quantidade aleatória, e não somente para parâmetros de um modelo. Seja ω uma quantidade aleatória e $p(\omega)$ a distribuição de probabilidade dessa quantidade. Essa distribuição pode ser *a priori*, *a*

posteriori ou preditiva. O intervalo de credibilidade para ω , com $(1-\alpha)$ de probabilidade, é o intervalo (L,U) tal que:

$$\int_L^U p(\omega)d\omega = 1 - \alpha \quad (3.17)$$

onde L e U são os limites inferior e superior do intervalo, respectivamente. É possível perceber que, para um quantil fixo α , não há um valor único para os limites L e U , ainda que $p(\omega)$ seja unimodal. Com isso em vista, é prática corrente a adoção do intervalo de mais alta densidade, ou intervalo HPD (do inglês, *Highest Probability Density*). Um intervalo $I \subseteq \Omega$, sendo Ω o domínio de ω , é caracterizado como HPD, a um nível $100(1-\alpha)\%$, com relação a $p(\omega)$ se (BERNARDO & SMITH, 2004):

1. $P(\omega \in I) = 1 - \alpha$; e
2. $p(\omega_1) \geq p(\omega_2)$ 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 intervalo com massa $(1-\alpha)$, e quanto mais curto for esse intervalo, menor a incerteza acerca de ω .

3.3.4 Métodos de cálculo

A principal dificuldade na aplicação da teoria Bayesiana é o cálculo da distribuição preditiva *a priori* dada na equação (3.14). Mais precisamente, para se fazer qualquer tipo de inferência sobre o modelo (momentos, quantis, intervalos de credibilidade), é necessário o cálculo do valor esperado de uma função h sobre a distribuição *a posteriori* dos parâmetros. Formalmente:

$$E[h(\theta)|x] = \frac{\int_{\Theta} h(\theta)f(x|\theta)\pi(\theta)d\theta}{\int_{\Theta} f(x|\theta)\pi(\theta)d\theta} = \int_{\Theta} h(\theta)\pi(\theta|x)d\theta \quad (3.18)$$

O cálculo analítico de integrais do tipo mostrado na equação (3.18) é impossível na quase totalidade das aplicações práticas, em especial nos casos multidimensionais. O cálculo dessas integrais pode ser evitado utilizando-se algoritmos de amostragem, tais

como os que empregam a integração de Monte Carlo via cadeias de Markov (MCMC, do inglês *Markov Chain Monte Carlo*). Segundo Gilks *et al.* (1996), essa classe de algoritmos permite se obter uma amostra de uma distribuição de probabilidades, como, por exemplo, a distribuição *a posteriori* expressa na equação (3.13), através de uma cadeia de Markov construída de forma que, após um grande número de realizações, sua distribuição de equilíbrio seja $\pi(\theta|x)$. De posse da amostra da distribuição *a posteriori* de θ a estimativa do valor esperado na equação (3.8) é obtida por integração de Monte Carlo da seguinte maneira: seja $\theta_1, \theta_2, \dots, \theta_m$ uma amostra da distribuição *a posteriori* de θ ; uma aproximação do valor esperado mostrado na equação (3.18) pela integração de Monte Carlo é dada por:

$$E(h(\theta)|x) = \frac{1}{m} \sum_{i=1}^m h(\theta_i) \quad (3.19)$$

Dessa forma, a média populacional de h é estimada pela média da amostra gerada da distribuição *a posteriori*. De acordo com Gilks *et al.* (1996), quando a amostra $\{\theta_i\}$ é independente, a lei dos grandes números garante que a aproximação possa ser feita de modo tão acurado quando se queira, a partir do aumento do tamanho da amostra.

De maneira geral, é difícil obter amostras independentes de $\pi(\theta|x)$, uma vez que a mesma pode ter formas bastante complexas (GILKS *et al.*, 1996). Entretanto, $\{\theta_i\}$ não precisa ser necessariamente independente. É suficiente que $\{\theta_i\}$ seja gerada por um processo que amostrasse 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 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:

$$P(\theta_{t+1} \in A | \theta_t, \theta_{t-1}, \dots, \theta_0) = P(\theta_{t+1} \in A | \theta_t), \quad A \subseteq S \quad (3.20)$$

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. Para serem empregadas em algoritmos de amostragem, as cadeias de Markov devem ser:

- Irredutíveis, que significa que, independentemente de seu estado inicial, a cadeia é capaz de atingir qualquer outro estado em um número finito de iterações com probabilidade maior que zero;
- Aperiódicas, que significa que a cadeia não oscila entre um conjunto de estados em movimentos regulares; e
- Recorrente, que significa que para todos os estados, se o processo se inicia em i , ele retornará a i em um número finito de iterações.

Uma cadeia de Markov com as características acima é denominada ergódica. A ideia básica de todos os algoritmos de amostragem é obter uma amostra da distribuição *a posteriori* construindo uma cadeia de Markov ergódica com as seguintes propriedades:

- A cadeia deve ter o mesmo número 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 (Metropolis *et al.*, 1953) tem a propriedade de criar cadeias com as características apresentadas acima. O referido algoritmo foi desenvolvido nos laboratórios de Los Alamos com objetivo de resolver problemas relacionados ao estado de energia de materiais nucleares. Ainda que o método tenha ganhado dimensão a partir do trabalho de Nicholas Metropolis e colaboradores, em 1953, seu desenvolvimento teve a colaboração de vários pesquisadores que trabalharam no projeto Manhattan, como Stanislaw Ulam, John Von Neumann, Enrico Fermi e outros. Segundo o próprio Metropolis, a ideia básica do método foi desenvolvida, mas não publicada, por Enrico Fermi, cerca de 15 anos antes. Detalhes a respeito do algoritmo podem ser encontrados nos trabalhos de Anderson (1986), Metropolis (1987) e Hitchcock (2003).

O algoritmo de Metropolis foi generalizado por Hastings (1970), dando origem à versão amplamente empregada atualmente. O algoritmo é construído a partir de uma distribuição de referência $g(\theta^*|\theta, x)$, da qual é fácil obter amostras de θ , da seguinte forma:

Inicialize θ_0 ; $t \leftarrow 0$

Repita {

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

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

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

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 por meio de taxas, $\pi(\theta^* | x) / \pi(\theta_t | x)$, o que dispensa o cálculo da constante de normalização dada pela equação (3.14). A generalização proposta por Hastings (1970) se refere basicamente às propriedades da distribuição de referência $g(\bullet | \bullet)$, uma vez que no trabalho original de Metropolis *et al.* (1953) somente distribuições simétricas eram permitidas para $g(g(\theta_i | \theta_j) = g(\theta_j | \theta_i))$, e, neste caso, a taxa de aceitação do algoritmo se reduz a:

$$\alpha_{MH}(\theta^* | \theta_t, x) = \text{MIN} \left\{ 1, \frac{\pi(\theta^* | x) g(\theta_t | \theta^*, x)}{\pi(\theta_t | x) g(\theta^* | \theta_t, x)} \right\} \quad (3.21)$$

Robert & Casella (2004) mostraram que o algoritmo anterior alcança o equilíbrio, após um grande número de iterações, tendo $\pi(\theta | x)$ como distribuição estacionária. Depois de alcançado o equilíbrio, todas as realizações do algoritmo constituirão amostras da distribuição *a posteriori* de θ e os valores esperados na equação (3.18) podem ser estimados por meio de integração de Monte Carlo, com a precisão definida pelo número de realizações do algoritmo.

A escolha da distribuição de referência $g(\bullet|\bullet)$ é a principal questão para a eficiência do algoritmo. Segundo Gilks *et al.* (1996), qualquer distribuição de referência permite obter amostras da distribuição *a posteriori* de θ . Entretanto, a taxa de convergência depende fortemente da relação entre $g(\bullet|\bullet)$ e a distribuição alvo $\pi(\theta|x)$. Dessa forma, quanto maior a semelhança entre g e π , mais rapidamente a cadeia alcança seu equilíbrio. Além disso, mesmo quando a cadeia alcança o equilíbrio, as realizações podem se mover lentamente sobre o suporte de π , sendo necessário um grande número de realizações para uma amostragem adequada da distribuição alvo.

Do ponto de vista computacional, g deve ser escolhida de maneira 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 garantir que a amostra candidata, obtida de $g(\bullet|\bullet)$, percorra todo o espaço paramétrico de $\pi(\theta|x)$. Detalhes adicionais relativos à construção de distribuições de referência podem ser encontrados em Gilks *et al.* (1996).

A partir do algoritmo de Metropolis-Hastings surgiram inúmeros esquemas de simulação, dentre eles o amostrador de Gibbs. Tais esquemas se diferenciam pela forma de construção da distribuição de referência. O amostrador de Gibbs, por exemplo, utiliza as distribuições condicionais completas *a posteriori* com esse intuito. Esses outros algoritmos não serão abordados neste trabalho. Para detalhes adicionais a respeito dos mesmos, podem ser consultados os trabalhos de Gilks *et al.* (1996), Liu (2001) e Robert & Casella (2004).

3.4 Precipitação máxima provável: conceito e métodos de estimação

3.4.1 Introdução

A PMP é definida, segundo a WMO (1986), como a altura máxima teórica de precipitação, para uma dada duração, meteorologicamente possível para uma dada região geográfica, em uma determinada época do ano. Por sua vez, a EMP representa o limite superior de uma enchente potencial, em uma seção fluvial, resultante de uma tempestade hipotética de duração e altura críticas, a PMP, antecedida por condições hidrológicas e hidrometeorológicas severíssimas, porém fisicamente possíveis. A

transformação de PMP em EMP é realizada usualmente através de modelos chuva-vazão devidamente calibrados para a bacia hidrográfica em estudo.

Os conceitos de PMP e EMP têm sido amplamente empregados na prática da Engenharia Hidrológica, como, por exemplo, no projeto de estruturas hidráulicas, como vertedores de grandes barragens, estruturas para controle de inundações e centrais nucleares, como valores máximos de ordem prática para precipitação e vazão. Contudo, ainda que, por definição, a PMP esteja associada a um limite superior, sua determinação não é inequívoca e depende da disponibilidade de um conjunto adequado de observações históricas, fazendo com que suas estimativas sejam suscetíveis às incertezas impostas pelas amostras disponíveis (FERNANDES *et al.*, 2010). Nesse sentido, a PMP deve ser tratada como uma variável aleatória, e seu comportamento deve ser modelado por uma distribuição de probabilidades.

Os métodos de estimação de PMP podem ser classificados em 2 grupos: (1) métodos hidrometeorológicos, que envolvem a maximização de tormentas extremas observadas, e (2) métodos estatísticos baseados em fatores de frequência, tais como o proposto por Hershfield (1961, 1965), e posteriormente modificado por Koutsoyiannis (1999).

3.4.2 Métodos hidrometeorológicos de estimação de PMP

Os métodos hidrometeorológicos consideram que o total precipitado cresce à medida que aumenta o teor de umidade do fluxo de ar que alimenta as tempestades (BERTONI & TUCCI, 1993). Admite-se que a coincidência entre a máxima precipitação e a máxima umidade não ocorreu no passado, em função de flutuações dos demais fatores intervenientes no fenômeno. Entretanto, essa coincidência pode ocorrer em algum tempo futuro.

Bertoni & Tucci (1993) classificam os métodos hidrometeorológicos da seguinte maneira: (1) maximização de tormentas severas; (2) transposição de tormentas severas; e (3) maximização de sequências de tormentas severas. Esses métodos são tratados em detalhes a seguir.

3.4.2.1 Maximização de tormentas severas

Este método consiste em se identificarem as maiores tormentas em uma dada região e multiplicá-las por um fator de maximização dado por:

$$f_m = \frac{W_m}{W} \quad (3.22)$$

onde W_m é a altura máxima de água precipitável para a época do ano e W é a altura de água precipitável nas condições em que a precipitação foi observada.

A altura de água precipitável, por unidade de área, pode ser determinada com auxílio da seguinte equação:

$$W = \frac{1}{g} \int_{P_0}^{P_1} q(P, T) dP \quad (3.23)$$

na qual g é a aceleração da gravidade, q é a umidade específica em função da pressão P e da temperatura T , e P_0 e P_1 são as pressões nos níveis inferior e superior da coluna de ar.

O processo de maximização de umidade requer a identificação de duas adiabáticas saturadas. A primeira delas é representativa da temperatura e umidade específica da tormenta a ser maximizada. Já a segunda corresponde à adiabática mais quente esperada na região em uma dada época do ano. As referidas adiabáticas são identificadas pelas temperaturas de bulbo úmido de um psicrômetro, que corresponde à temperatura do ponto de orvalho à pressão de 1.000 mb (WMO, 1986).

De acordo com a WMO (1986), o ponto de orvalho representativo de uma dada tormenta é o maior ponto de orvalho persistente por 12 horas. Já o ponto de orvalho representativo da adiabática mais quente é estimado por meio do máximo ponto de orvalho persistente por 12 horas, obtido em séries de tamanho superior a 50 anos. Para amostras menores, a WMO (1986) recomenda a análise de frequência dos valores máximos observados e a adoção de um quantil com tempo retorno entre 50 e 100 anos.

Para a maximização em regiões montanhosas, devem ser também considerados os efeitos do vento, uma vez que as alturas de precipitação orográfica aumentam com a velocidade do vento normal às escarpas. O fator de ajuste do vento, f_v , é dado por:

$$f_v = \frac{V_{\max}}{V_m} \quad (3.24)$$

na qual V_{\max} representa a velocidade média máxima e V_m representa a velocidade média observada durante a tormenta a ser maximizada. As velocidades são correspondentes às durações das tormentas e tomadas na direção mais crítica. A velocidade média máxima é obtida por meio de análise de frequência e corresponde a um quantil com tempo de retorno entre 50 e 100 anos.

3.4.2.2 Transposição de tormentas severas

Este método se aplica quando o número de observações de tormentas severas não é suficiente na bacia de interesse. A transposição somente é válida se existem reais condições de que a tormenta possa ocorrer nas duas regiões. Nesse caso, as regiões são ditas meteorologicamente homogêneas, estando expostas à incursão das mesmas massas de ar e aos mesmos tipos de tormentas (BERTONI & TUCCI, 1993).

O ajuste de transposição é realizado pela multiplicação da altura pluviométrica da tormenta observada por um ou mais fatores, que tem por objetivo compensar as diferenças de condições nas duas áreas de estudo. De acordo com Bertoni & Tucci (1993), a frequência de ocorrência das tormentas nas duas áreas não precisa ser necessariamente a mesma. No entanto, as diferenças topográficas entre as regiões podem dificultar ou mesmo inviabilizar a transposição. Dessa forma, é necessária uma análise prévia dos fatores topográficos, morfológicos, geométricos e de orientação das bacias.

3.4.2.3 Maximização de sequências de tormentas severas

Este método é aplicável a grandes bacias, nas quais a área de drenagem supera significativamente a extensão das tormentas. Nesse caso, é necessário definir o sincronismo mais adverso entre as máximas precipitações acumuladas e o desenvolvimento e propagação das enchentes.

Na definição do referido sincronismo, analisam-se os registros históricos relacionados às grandes enchentes, diagnosticam-se as tormentas geradoras das máximas precipitações acumuladas e determinam-se as maximizações a serem efetuadas, abrangendo-se a relocação e transposição de tormentas.

3.4.3 Métodos estatísticos de estimação de PMP

Os métodos estatísticos, que podem ser empregados sempre que a disponibilidade de dados pluviométricos for suficiente, são particularmente úteis para realizar estimativas de PMP na ausência de dados meteorológicos (WMO, 1986). Dentre os métodos estatísticos, o mais difundido é aquele proposto por Hershfield (1961), que se tornou o procedimento padrão sugerido pela WMO (1986). As principais vantagens do método são o fato de levar em consideração as características do local de interesse, por meio dos parâmetros estatísticos, e a facilidade de aplicação (KOUTSOYIANNIS, 1999). O procedimento é baseado na seguinte equação:

$$X_m = \bar{X}_n + k_m s_n \quad (3.25)$$

na qual:

X_m é a chuva máxima observada na bacia de interesse;

\bar{X}_n é a média da série de precipitações máximas anuais;

s_n é o desvio padrão da série de precipitações máximas anuais;

k_m é o fator de frequência.

Para a estimação do fator de frequência k_m , Hershfield (1961) empregou 95.000 dados de precipitações máximas anuais de 24 horas, oriundos de 2.645 estações, sendo cerca de 90% delas localizadas em território norte-americano. O autor constatou que o valor máximo de k_m era igual a 15 e sugeriu que o mesmo fosse adotado em todos os casos de estimação de PMP. Análises posteriores mostraram que o fator de frequência não é independente da magnitude da precipitação, variando inversamente com a mesma: o valor 15 é muito alto para locais com fortes chuvas e muito baixo para regiões áridas (WMO, 1986). Além disso, o referido valor é muito alto para precipitações com duração inferior a 24 horas (HERSHFIELD, 1965).

Para contornar esse problema, Hershfield (1965) construiu um ábaco de variação de k_m com a precipitação média anual e com a duração. Esse ábaco, mostrado na figura 3.4, é recomendado pelo manual da WMO para aplicação do método estatístico de estimação de PMP.

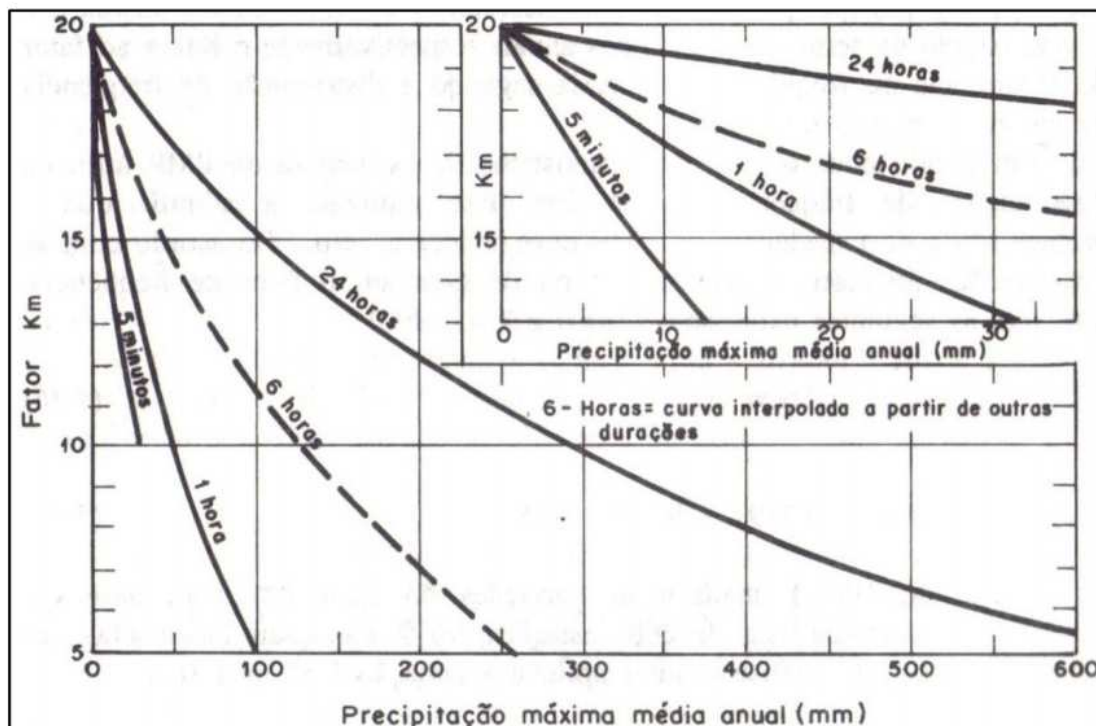


Figura 3.4 – Estimativa do fator k_m para o método estatístico de estimação da PMP (Fonte: Bertoni & Tucci, 1993)

Koutsoyiannis (1999), utilizando a mesma base de dados de Hershfield (1961), realizou uma série de análises relacionadas à determinação de k_m . O estudo demonstrou que não existem evidências acerca da existência de um limite superior para k_m e conseqüentemente para a PMP. O autor demonstrou ainda que a estimativa de PMP de Hershfield é obtida com o uso da distribuição GEV (Generalizada de Valores Extremos), com o parâmetro de forma dado por uma função linear da precipitação média anual máxima, e para período de retorno igual a 60.000 anos.

Segundo a WMO (1986), alguns cuidados devem ser tomados ao se empregar o método estatístico na estimação da PMP. Em primeiro lugar, deve-se observar a presença de “outliers” na amostra, uma vez que esses podem influenciar consideravelmente os cálculos da média e do desvio padrão das precipitações máximas anuais. Outro problema é que a média e o desvio padrão das séries de precipitação anual tendem a crescer com o tamanho da amostra, uma vez que a distribuição de chuvas extremas é, de maneira geral, assimétrica à direita. Assim, a probabilidade de se observar maiores valores para as alturas precipitadas torna-se mais elevada à medida que novos registros são disponibilizados (WMO, 1896).

Além disso, as medições de chuva são realizadas, em geral, para intervalos de tempos fixos, que usualmente não contêm as reais quantidades máximas de precipitação para uma dada duração. Segundo o manual da WMO (1986), a multiplicação das alturas máximas anuais, definidas em intervalos fixos, por um fator igual a 1,13 gera resultados bastante próximos dos máximos reais para a duração avaliada.

Por fim, deve-se levar em conta que as estimativas de PMP obtidas com o método estatístico são pontuais e, assim, devem ser ajustadas para diferentes áreas de drenagem em concordância com as curvas altura-área-duração estabelecidas para a região de estudo.

3.5 Modelos de simulação hidrológica

3.5.1 Introdução

Os processos de formação de escoamento em uma bacia hidrográfica são oriundos da interação de um complexo conjunto de fatores causais, os quais refletem a variabilidade espaço-temporal do aporte de precipitação, das condições de infiltração e percolação no solo, do armazenamento em depressões do terreno e da evapotranspiração no balanço hídrico da bacia. Em função dessa complexidade, a descrição dos referidos processos a partir da solução das equações diferenciais que governam cada uma dessas etapas constitui uma impossibilidade. Dessa forma, é necessário recorrer a uma conceitualização simplificada da síntese hidrológica na bacia, que seja apta a representar adequadamente os processos físicos de interesse, mediante a solução de um número reduzido de equações. Tal conceitualização é base dos modelos de simulação hidrológica, também denotados modelos de transformação chuva-vazão.

Modelos chuva-vazão têm sido amplamente utilizados em análises hidrológicas, tais como procedimentos de regionalização, extensão de séries temporais de vazões, avaliação da resposta hidrológica de bacias a alterações climáticas e de uso do solo, determinação de cheias de projeto, avaliação de estratégias de gestão de recursos hídricos, dentre outros (WAGENER *et al.*, 2004).

De maneira geral, modelos de simulação hidrológica são classificados de acordo com a escala temporal, escala espacial, natureza dos processos envolvidos e estrutura. Com relação à escala temporal, os modelos podem ser contínuos ou de eventos isolados. No primeiro caso, utilizam-se séries contínuas de vazões, precipitações e/ou

evapotranspiração, independentemente do intervalo de tempo adotado, para realizar as simulações desejadas. No segundo caso, os modelos simulam a resposta de uma bacia a um evento pré-determinado, tal como uma chuva de projeto.

Com relação à escala espacial, os modelos podem ser concentrados ou distribuídos. No primeiro caso, a bacia hidrográfica é tratada como uma única unidade espacial, e as variações dos fenômenos modelados no espaço são desconsideradas nas análises. Já para um modelo distribuído, a bacia é dividida em várias unidades espaciais, e os fenômenos são modelados como funções do espaço e do tempo.

Quanto à natureza dos processos, pode-se classificar os modelos como estocásticos, determinísticos ou híbridos. Modelos estocásticos são aqueles em que a teoria de probabilidades é introduzida na formulação do problema, e para os quais diferentes *outputs* podem ser originados a partir do mesmo conjunto de *inputs*. Modelos determinísticos são regidos por leis físicas específicas, que não envolvem a teoria de probabilidades. Modelos híbridos combinam processos estocásticos e determinísticos.

Finalmente, quanto à estrutura, os modelos podem ser empíricos, conceituais ou paramétricos, ou fisicamente fundamentados. Modelos empíricos são aqueles em que os valores simulados são relacionados às séries observadas por meio de funções que não possuem relação física com o fenômeno descrito (TUCCI, 1998). A estrutura do modelo e os valores dos parâmetros são deduzidos das séries temporais disponíveis, não exigindo nenhum conhecimento *a priori* do comportamento da bacia e dos processos de formação do escoamento. Em geral, modelos empíricos são concentrados e não são adequados, ao menos em teoria, para procedimentos de transferência espacial de informações relacionadas às séries de vazões (WAGENER *et al.*, 2004).

Modelos conceituais são aqueles cuja estrutura é definida segundo um entendimento prévio de como o processo hidrológico ocorre. A representação dos processos é frequentemente realizada com auxílio de reservatórios fictícios, os quais podem ser lineares ou não lineares. Tais reservatórios são alimentados por processos como precipitação e infiltração, e exauridos por evapotranspiração e escoamentos superficial, sub-superficial e subterrâneo. Os parâmetros do modelo descrevem as propriedades de armazenamento dos reservatórios.

A maioria dos modelos conceituais é do tipo concentrado, embora não seja incomum a utilização de modelos distribuídos, nos quais é realizada a discretização espacial da bacia (WAGENER *et al.*, 2004). Na presente tese, será conferido destaque à utilização de modelos do primeiro tipo, uma vez que, segundo argumentam Wagener *et al.* (2004), os mesmos são suficientes para se descrever adequadamente a síntese hidrológica de uma bacia.

3.5.2 Calibração de parâmetros modelos hidrológicos sob o paradigma Bayesiano

Modelos conceituais têm por objetivo reproduzir as principais características da dinâmica de transformação chuva-vazão em uma bacia a partir de uma estruturação simplificada do sistema hidrológico. Nesse contexto, as complexas interações que governam o processo de conversão da precipitação em escoamento, em escalas temporal e espacial, são sintetizadas por um número reduzido de componentes que buscam representar, em termos de balanço hídrico, a resposta da bacia aos aportes de chuva. A transferência de água entre os diversos componentes do modelo de simulação hidrológica se dá mediante relações funcionais estabelecidas pelos parâmetros do mesmo.

Em essência, os parâmetros constituem os principais fatores que controlam a síntese hidrológica de um modelo chuva-vazão (MORADKHANI & SOROOSHIAN, 2008). Sob essa perspectiva, a correta representação do fenômeno físico modelado é intrinsecamente relacionada à determinação dos parâmetros. Esses, contudo, em geral não possuem um significado físico interpretável e, assim, não são passíveis de medição direta. Diante disso, os mesmos devem ser inferidos a partir dos registros observados, tendo-se em vista a obtenção do melhor ajuste possível entre as vazões simulada e observada. Esse procedimento é denominado calibração de parâmetros.

A calibração de parâmetros tem sido tradicionalmente tratada como um problema de otimização (KAVETSKI *et al.*, 2003). De fato, o que usualmente se busca é um conjunto de parâmetros tal que os resíduos da simulação sejam minimizados, mediante a especificação de uma função-objetivo. Essa abordagem preconiza que os resíduos oriundos da simulação hidrológica sejam de natureza essencialmente aleatória e que possam ser acomodados na incerteza acerca dos *outputs* do modelo. Em última análise, tal simplificação conduz à hipótese de que os erros de calibração são independentes e

identicamente distribuídos segundo um modelo Gaussiano $r \sim N(0, \sigma^2_e)$, o que corresponde a uma regressão via método dos mínimos quadrados (SILVA *et al.*, 2014).

Ainda que extremamente difundida em aplicações práticas, a busca pelo conjunto “ótimo” de parâmetros segundo esse modelo de regressão tem sido objeto de severas críticas na comunidade científica (Beven & Binley, 1992; Kavetski *et al.*, 2002; Kavetski *et al.*, 2006a, b; Vrugt *et al.*, 2013), tendo-se em vista que as hipóteses básicas sobre as quais se constrói a regressão são frequentemente violadas após a calibração. Com efeito, os resíduos da simulação são, em geral, não normais, correlacionados e heterocedásticos. Além disso, grande parte dos algoritmos empregados com propósitos de otimização faz uso de um indicador estatístico único, a função objetivo, para resumir a informação extraída dos dados (Vrugt *et al.*, 2013), e, dessa forma, desempenhos muito similares na calibração podem ser obtidos para conjuntos paramétricos distintos, situação essa nomeada “equifinalidade” por Beven & Binley (1992). Por fim, métodos tradicionais de otimização podem ser fortemente afetados pela geometria da função objetivo, a qual pode apresentar formas especialmente complexas, descontinuidades e um conjunto expressivo de máximos locais em decorrência da multidimensionalidade e da não linearidade do modelo hidrológico (KAVETSKI *et al.*, 2006c). Os efeitos combinados dessas limitações podem levar a estimativas bastante inconsistentes para os parâmetros do modelo hidrológico.

Métodos Bayesianos têm despontado como uma ferramenta promissora para abordar o problema de calibração (KAVETSKI *et al.*, 2003). Sob tal paradigma, os parâmetros do modelo hidrológico são tratados como variáveis aleatórias, cuja incerteza pode ser resumida por meio de uma distribuição *a priori*, e a função de verossimilhança é interpretada como um critério de avaliação da similaridade entre a realização do fenômeno, representada pelas vazões observadas, e os resultados oriundos da simulação (SILVA *et al.*, 2014), à qual podem ser agregadas, implícita ou explicitamente, as diversas fontes de incertezas da modelagem chuva-vazão (*inputs*, *outputs*, parâmetros e estruturais). O resultado da calibração é uma distribuição conjunta *a posteriori* que expressa intervalos de variação plausíveis para os parâmetros, tanto do modelo hidrológico quanto dos modelos de resíduos, mediante a informação trazida pelos dados.

A principal questão que se impõe à calibração sob o paradigma Bayesiano é o tratamento da função de verossimilhança no contexto da inferência. Duas abordagens se distinguem em razão do rigor estatístico empregado em sua construção (Vrugt *et al.*, 2008): a abordagem informal, a qual sintetiza a influência de todas as fontes de erro por meio de medidas de verossimilhança, e pode ser exemplificada pelo método GLUE, do acrônimo em inglês para “*Generalized Likelihood Uncertainty Estimation*” (BEVEN & BINLEY, 1992); e a abordagem formal, a qual se vale de distribuições teóricas na especificação do modelo de resíduos, e pode ser exemplificada pelo método BATEA, do acrônimo em inglês para “*Bayesian Total Error Analysis*” (KAVETSKI *et al.*, 2003, 2006a,b; RENARD *et al.*, 2011).

Modelos baseados na abordagem informal podem apresentar resultados bastante efetivos em procedimentos de calibração (VRUGT *et al.*, 2008). Contudo, diversos autores (Montanari, 2005; Blasone *et al.*, 2008; Stedinger *et al.*, 2008; Vrugt *et al.*, 2008) têm se mostrado contrários ao uso da referida abordagem, argumentando que as hipóteses estabelecidas *a priori* não podem ser testadas e validadas *a posteriori*. Por outro lado, ainda que a abordagem formal seja adotada, há o problema da escolha de uma função de verossimilhança adequada, a qual seja apta a reproduzir o comportamento dos resíduos após a calibração. Nesse contexto, o modelo Gaussiano, apesar de amplamente empregado (KAVETISKI *et al.*, 2006a,b; Renard *et al.*, 2011; Silva *et al.*, 2014), via de regra não corrobora as hipóteses de normalidade, independência e homocedasticidade assumidas *a priori*.

Diante desse fato, pesquisadores têm buscado funções verossimilhança capazes de acomodar a complexa estrutura dos erros oriundos da modelagem hidrológica. Um modelo bastante promissor sob essa perspectiva é aquele proposto por Schoups & Vrugt (2010), denotado por função de verossimilhança generalizada por Silva *et al.* (2014). O referido modelo incorpora em sua estrutura um polinômio autoregressivo, que descreve a autocorrelação dos resíduos, um modelo probabilístico dotado de parâmetros relacionados à assimetria e à curtose, que controlam a não normalidade dos erros, e uma relação de desvio padrão expressa em função da magnitude das vazões, que assim acomoda sua heterocedasticidade. Essa função de verossimilhança foi empregada por Silva *et al.* (2014) e se mostrou apta a validar *a posteriori* as hipóteses assumidas para os resíduos de calibração.

3.5.3 O modelo Rio Grande de transformação chuva-vazão

O modelo Rio Grande de transformação chuva-vazão pode ser classificado como conceitual, determinístico, contínuo e semi-distribuído. O modelo é composto pelos módulos de produção, concentração e propagação em cursos d'água. Esses módulos são descritos detalhadamente nos próximos tópicos. As informações que se seguem foram obtidas do relatório desenvolvido por Naghettini *et al.* (2006) para o sistema de previsão de vazões de curto prazo da Companhia Energética de Minas Gerais (CEMIG).

3.5.3.1 Módulo de produção

O módulo de produção do modelo Rio Grande realiza o balanço de umidade do solo ao longo do tempo. Esse módulo é inteiramente baseado na conceitualização do modelo Xinanjiang, originalmente descrito por Zhao *et al.* (1980 *apud* Naghettini *et al.*, 2006). De acordo com os autores, a estrutura conceitual adotada no modelo Xinanjiang é adequada a regiões úmidas e semi-úmidas, possuindo forte analogia física com características da bacia.

As principais fases do módulo de produção do modelo Rio Grande são tratadas em detalhes a seguir.

Formação do escoamento

A formação de escoamento no modelo Rio Grande somente se dá nas situações em que o volume máximo de armazenamento de água capilar é superado. Em outras palavras, não ocorre escoamento antes que a umidade do solo em sua zona aerada alcance a chamada capacidade de campo. A partir desse ponto, todo o volume precipitado, exceto a evapotranspiração, é convertido em escoamento.

A capacidade de campo pode ser definida como o conteúdo de umidade presente em uma amostra de solo depois de completada a drenagem do excesso de água gravitacional. Esse volume de água fica retido no solo pela ação de forças capilares. Na estrutura do modelo Xinanjiang, uma parcela da capacidade de campo pode ser eliminada por meio da ação da evapotranspiração. Essa parcela é denominada capacidade de tensão capilar da água do solo.

O balanço de umidade no solo pode ser descrito resumidamente da seguinte forma:

- Antes de atingir a capacidade de campo:

$$P - ET = WC_2 - WC_1 \quad (3.26)$$

- Depois de atingir a capacidade de campo:

$$P - ET - R = WM - WC_1 \quad (3.27)$$

sendo:

P - precipitação;

E - evapotranspiração;

R - escoamento;

WM - capacidade de campo; e

WC_1, WC_2 - conteúdo de umidade do solo no início e no fim do intervalo, respectivamente.

Volume de escoamento

As equações 3.26 e 3.27 representam o balanço de umidade em um dado ponto da bacia. Ao se avaliar toda a área de drenagem, no entanto, é necessário levar em consideração que a distribuição da capacidade de retenção capilar não é uniforme, o que faz com que a produção de escoamento varie de maneira significativa nos diversos pontos da bacia. Tendo tal fato em vista, foi empregada no modelo Rio Grande uma curva paramétrica para distribuição espacial da capacidade de retenção capilar, que indica a fração da área da bacia que alcança a capacidade de campo em função de um evento chuvoso e contribui assim para a formação de escoamento.

Os parâmetros que controlam a curva de distribuição espacial da capacidade de retenção capilar são:

wm - capacidade de retenção capilar média espacial. Esse parâmetro constitui uma medida da aridez da bacia;

b - expoente da curva de distribuição espacial de retenção capilar, que representa uma medida da não-uniformidade da bacia.

imp - fração de área impermeável da bacia.

Escoamentos superficial, sub-superficial e subterrâneo

Nos pontos da bacia em que a capacidade de retenção capilar é atingida, a introdução de qualquer suprimento adicional de água, aqui denominada água “livre”, dá origem ao escoamento. O escoamento pode ser dividido em três tipos distintos: superficial, sub-superficial e subterrâneo. A ocorrência de escoamento superficial se dá quando a umidade do solo atinge o máximo armazenamento de água gravitacional ou livre.

Em áreas permeáveis, o cálculo do escoamento superficial é realizado com auxílio de uma curva paramétrica de distribuição espacial de água livre. Essa curva determina a fração da área da bacia que atinge o máximo armazenamento de água livre em função de um evento chuvoso, originando assim escoamento superficial. No restante da área, onde o máximo armazenamento não é atingido, o suprimento adicional de água livre é convertido em escoamento sub-superficial e subterrâneo, cujas contribuições são determinadas pelos coeficientes k_{ss} e k_g , respectivamente. Em áreas impermeáveis, todo escoamento gerado é superficial.

Os parâmetros introduzidos nessa fase são:

- sm : capacidade média espacial de água livre, que controla a magnitude do escoamento superficial. Seu valor depende da profundidade da camada superficial de maior condutividade hidráulica;
- ex : expoente da curva de distribuição espacial de capacidade de água livre;
- k_{ss} : coeficiente do fluxo diário do escoamento sub-superficial; e
- k_g : coeficiente do fluxo diário do escoamento subterrâneo.

Afluxos à rede de drenagem

Essa fase representa o movimento da água oriunda das colunas de solo em direção ao sistema de drenagem. O componente superficial do escoamento flui diretamente aos elementos do sistema de drenagem de forma muito mais rápida que as componentes sub-superficial e subterrânea. Para levar esse fato em consideração, no modelo Rio

Grande as parcelas sub-superficial e subterrânea do escoamento são submetidas à propagação através de reservatórios lineares. O deflúvio total CIN é dado pela soma dos componentes superficial, sub-superficial e subterrâneo no intervalo de tempo considerado.

Os parâmetros introduzidos nessa fase são:

- ci : coeficiente de recessão do fluxo sub-superficial diário. Seu valor pode ser obtido por meio da análise de hidrogramas observados. O valor CI , correspondente ao intervalo de tempo de cálculo, pode ser obtido por exponenciação do parâmetro ci à razão entre os intervalos sub-diário e diário;
- cg : coeficiente de recessão do fluxo subterrâneo diário, cujo valor pode igualmente ser obtido por meio da análise de hidrogramas observados. O valor CG , correspondente ao intervalo de tempo de cálculo, pode ser obtido por exponenciação do parâmetro cg à razão entre os intervalos sub-diário e diário.

Evapotranspiração

À medida que a coluna de solo é submetida à ação da evapotranspiração, inicia-se o processo de depleção da umidade armazenada sob forma de água capilar. A análise da evapotranspiração é realizada considerando-se a zona aerada do solo dividida em zona superior, inferior e profunda, com capacidades de retenção espaciais médias dadas, respectivamente, por WUM , WLM e WDM . Tanto o armazenamento quanto a depleção da umidade da região aerada do solo ocorrem inicialmente na zona superior, em seguida na zona inferior e finalmente na zona profunda.

A evapotranspiração potencial (EM) pode ser estimada tanto de maneira direta, por meio de tanques evaporimétricos, quanto de maneira indireta, por meio de dados meteorológicos. Na zona superior, a evapotranspiração real é igual à potencial. Na zona inferior, a evapotranspiração real é igual ao produto da evapotranspiração potencial pela razão entre o conteúdo atual e a capacidade de água capilar. Por fim, na zona profunda, a evapotranspiração real corresponde a uma fração da evapotranspiração potencial proporcional à área da bacia com vegetação de raízes profundas. A evapotranspiração efetiva total, denotada por E , é dada pela soma dos termos correspondentes às zonas superior, inferior e profunda.

Os parâmetros introduzidos nessa fase são:

- k : coeficiente de ajuste da evapotranspiração potencial, estimada por tanques evaporimétricos ou a partir de dados meteorológicos;
- wum : capacidade de retenção capilar espacial média na zona superior;
- wlm : capacidade de retenção capilar espacial média na zona inferior;
- wdm : capacidade de tensão espacial média na zona profunda;
- c : coeficiente de evapotranspiração profunda. Esse parâmetro permanece inativo durante os períodos de cheia, mas é de grande importância em períodos de seca.

A Figura 3.5 ilustra de maneira esquemática o módulo de produção do modelo Rio Grande.

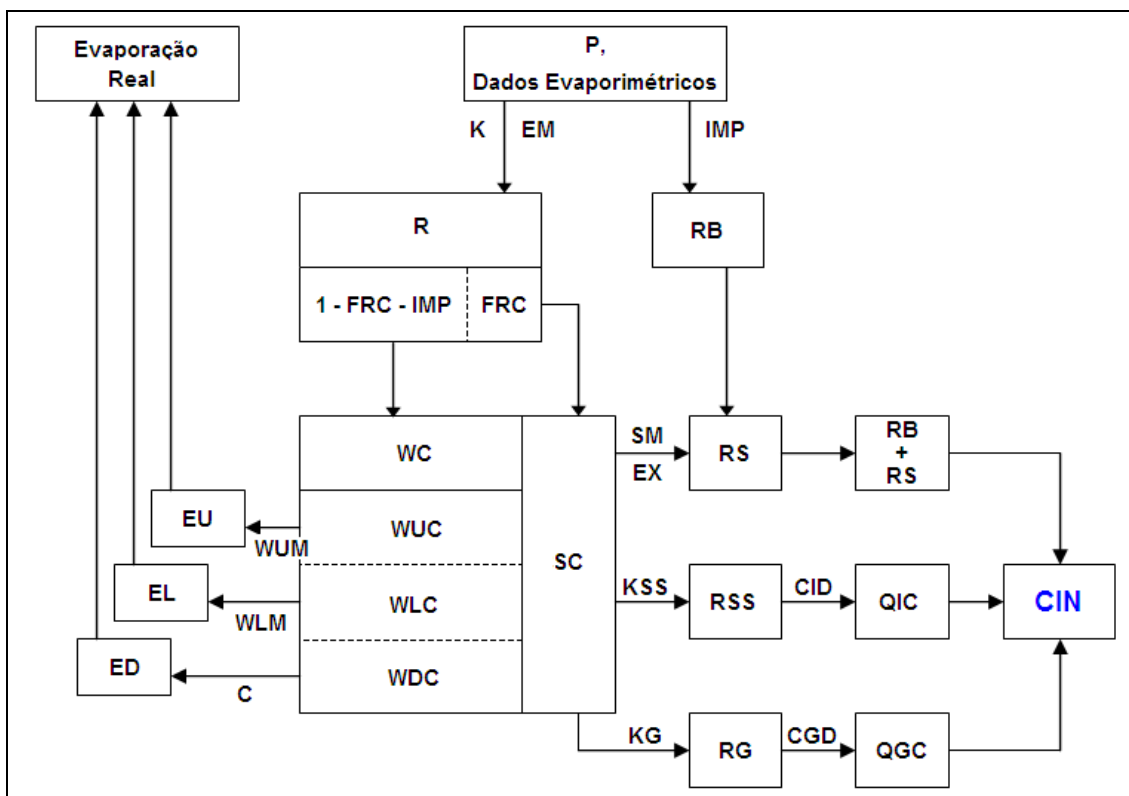


Figura 3.5 – Fluxograma do módulo de produção do modelo Rio Grande. Fonte: Pinheiro (2009)

3.5.3.2 Módulo de concentração

O módulo de produção do modelo Rio Grande gera um conjunto de pulsos discretos de lâminas de escoamento total CIN sobre uma área unitária da sub-bacia, em N intervalos de tempo Δt . A unidade usual de CIN é $mm/\Delta t \times km^2$.

Ainda que se assuma que as lâminas de escoamento geradas pelo módulo de produção sejam uniformes na sub-bacia, deve-se levar em consideração que as contribuições de diferentes áreas se concentram na seção exutória em diferentes intervalos de tempo. A introdução dessa consideração na estrutura do modelo Rio Grande se dá pela utilização de uma função de transferência baseada no hidrograma unitário sintético de Clark, de acordo com a formulação original do *Hydrologic Engineering Center* HEC (1981, *apud* NAGHETTINI *et al.*, 2006).

A construção do hidrograma unitário sintético de Clark requer a determinação dos seguintes parâmetros:

TC : tempo de concentração da bacia, expresso como um número inteiro de horas, e utilizado para computar os efeitos de translação; e

R_A : coeficiente de armazenamento, expresso como um número não necessariamente inteiro de horas, e utilizado para computar os efeitos de armazenamento e translação;

Histograma tempo-área, que representa a distribuição temporal das áreas que contribuem para a formação das vazões na seção exutória.

O tempo de concentração da bacia pode ser determinado por análise de hidrogramas observados ou com base em fórmulas empíricas. Uma vez determinado o tempo de concentração, pode-se estimar o coeficiente de armazenamento por meio da formulação proposta por HEC (1981):

$$TC + R_A = 0,625\sqrt{A} \quad (3.28)$$

na qual:

A - área, em km^2 .

O histograma tempo – área pode ser construído a partir de mapas topográficos e curvas hipsométricas da bacia. A estimação direta do mesmo, no entanto, é trabalhosa e

envolve muitas incertezas e simplificações. Para contornar esse inconveniente, o modelo Rio Grande utiliza o histograma sintético sugerido pelo HEC (1981, *apud* Naghettini *et al.*, 2006), que se descreve a seguir.

Seja T uma fração do tempo de concentração da bacia, conforme mostrado na equação que se segue:

$$T = \frac{t}{TC}, \quad (0 \leq T \leq 1) \quad (3.29)$$

e AI a área acumulada, obtida pela seguinte equação:

$$AI = \frac{A}{AT} \quad (3.30)$$

O histograma tempo – área sintético é dado por:

$$AI = aT^n, \quad \text{se } (0 \leq T \leq 0,5) \quad (3.31)$$

$$AI = 1 - a(1 - T)^n, \quad \text{se } (0,5 \leq T \leq 1)$$

Na equação 3.31, o expoente n é uma função da forma da bacia, podendo assumir os seguintes valores:

- $n = 1$, para bacias de forma retangular;
- $n = 2$, para bacias de forma losangular; e
- $n = 1,5$, para bacias de forma elipsoidal.

O parâmetro a , por sua vez, pode ser estimado a partir do valor de n , assumindo estes valores:

- $a = 1$, para bacias de forma retangular;
- $a = 2$, para bacias em forma losangular; e
- $a = 1,414$, para bacias de forma elipsoidal.

Após a síntese do histograma tempo – área, é necessário determinar as ordenadas incrementais $(AI_i - AI_{i-1})$, para cada t_i , e convertê-las para a unidade $\frac{m^3}{s \times mm \times km^2}$, dando origem ao hidrograma unitário de translação, cujas ordenadas são denotadas por y_i .

Para se incorporar os efeitos de armazenamento, o hidrograma de translação obtido é propagado em um reservatório linear de constante R_A horas, de acordo com a equação a seguir:

$$Q_i = CA \times y_i + CB \times Q_{i-1} \quad (3.32)$$

na qual:

$$CA = \frac{\Delta t}{R + 0,5 \times \Delta t} \quad (3.33)$$

e

$$CB = 1 - CA \quad (3.34)$$

Na equação 3.33, Δt corresponde ao intervalo de tempo dos pulsos de chuva efetiva ou lâminas de escoamento total em mm/km^2 .

Por fim, as ordenadas do hidrograma unitário de Clark, denotadas por x_i , são determinadas por meio da seguinte equação:

$$x_i = \frac{Q_{i-1} + Q_i}{2} \quad (3.35)$$

É importante observar que, na construção do hidrograma de Clark, são impostas as seguintes restrições pelo HEC (1981, *apud* NAGHETTINI *et al.*, 2006):

$$TC = 1,03\Delta t \quad (3.36)$$

$$R = 0,52\Delta t \quad (3.37)$$

3.5.3.3 Módulo de propagação em cursos d'água

A propagação de uma onda de cheia por um trecho de rio é descrita pelos fenômenos de translação e armazenamento. O primeiro fenômeno é caracterizado pelo movimento de

água paralelamente à superfície e resulta na concentração do escoamento em um ponto, como a seção exutória da bacia. Já o segundo fenômeno caracteriza-se pelo movimento da água em direção perpendicular ao escoamento, resultando em um processo de armazenamento. A modelagem matemática do processo de propagação é realizada com auxílio de modelos de escoamento em canais, que descrevem o fenômeno da translação, modelos de reservatório, que descrevem o fenômeno da difusão, ou uma combinação de ambos. Uma vez que o módulo de propagação não é empregado nas simulações realizadas nesta tese, o mesmo não será abordado em maiores detalhes.

4 METODOLOGIA

O método proposto na presente tese é constituído de três etapas principais: (1) desenvolvimento de um novo modelo de geração estocástica de precipitação diária, apto a representar de maneira adequada chuvas moderadas e extremas; (2) calibração do modelo Rio Grande de transformação chuva—vazão, empregando a denominada função de verossimilhança generalizada (SCHOUPS & VRUGT, 2010), a qual considera a correlação temporal, não normalidade e heterocedasticidade dos resíduos na simulação; e (3) conversão das séries de precipitação em vazão, empregando-se a distribuição conjunta *a posteriori* dos parâmetros do modelo Rio Grande obtidas na calibração para a geração de um grande número de séries de vazão, com intuito de se caracterizar de maneira mais acurada as incertezas nas estimativas de quantis de cheias. Essas três etapas são tratadas em detalhes nos tópicos a seguir.

4.1 Geração estocástica de precipitação diária

A principal motivação para a construção de um novo gerador estocástico é estabelecer um modelo que consiga representar de maneira acurada tanto chuvas moderadas quanto chuvas extremas na simulação de séries longas. Essa característica é essencial para a futura aplicação do gerador proposto na estimação indireta de quantis de enchentes extremas. No entanto, a maioria dos modelos estabelecidos na literatura apresentam limitações para simular todo o espectro de alturas de precipitação (SHARIF & BURN, 2006; LI *et al.*, 2012).

A grande limitação dos geradores tradicionais é que os mesmos tentam modelar os diferentes processos físicos de formação de chuvas moderadas e chuvas extremas sob o mesmo paradigma, por meio de uma única distribuição de probabilidades, seja ela teórica ou empírica. Contudo, acomodar processos de naturezas tão distintas como os que produzem chuvas moderadas e aqueles relacionados a chuvas extremas em uma única abordagem matemática não é uma tarefa trivial, uma vez que a maior parte dos modelos de uso corrente não possui flexibilidade suficiente para tal. Diante disso, fica evidente que os volumes de precipitação diária são mais acuradamente simulados por modelos híbridos de distribuições. Nesse contexto, uma distribuição é empregada na simulação de alturas de precipitação baixas à moderadas, via de regra associadas a eventos frontais, e uma segunda distribuição é utilizada para simular as chuvas extremas, oriundas de processos convectivos, mistos ou ainda aqueles resultantes da

intensificação de convergências de massas de ar de diferentes características hidrometeorológicas.

Modelos híbridos estritamente paramétricos representam uma opção conveniente para a construção de geradores estocásticos, uma vez que tais modelos, de maneira geral, são facilmente ajustados ao conjunto de dados e podem fornecer uma descrição bastante acurada de eventos extremos, por meio do uso de uma distribuição com cauda superior adequada. Contudo, esses modelos requerem a especificação das distribuições teóricas das quais supõe-se que as amostras foram retiradas, o que introduz à análise uma ampla gama de incertezas, associadas à escolha do modelo probabilístico e à estimação de seus parâmetros. Outro problema é a pouca capacidade de reprodução das variâncias mensal e anual desse tipo modelo, característica essa inerente à estrutura de geração independente das alturas de chuva em dias consecutivos, tal como apontado por Boughton (1999) e Srikanthan & McMahon (2000). Além disso, modelos distributivos usuais podem não ser aptos a modelar características particulares de certas amostras, tais como a bimodalidade (APIPATTANAVIS *et al.*, 2007). Por fim, a imposição de um limiar entre as duas distribuições do modelo híbrido pode fazer com que a densidade resultante seja descontínua nesse ponto de transição, o que inviabiliza a estimação de parâmetros pelo método de máxima verossimilhança e pode levar a instabilidades numéricas ao longo das simulações (FERRER & KATZ, 2008).

Modelos não paramétricos, por outro lado, são menos restritivos no que se refere à forma das distribuições empíricas e se mostram mais aptos a reproduzir estatísticas de longo termo, ao mesmo tempo em que excluem da análise a especificação de um modelo probabilístico teórico. No entanto, sua capacidade de extrapolação além do valor máximo observado é reduzida e, mesmo quando essa se faz possível, a probabilidade de geração de alturas de precipitação de grande magnitude é pequena, em função do comportamento exponencial das caudas dos estimadores de núcleo usualmente empregados na estimação de densidade não paramétrica (WILKS & WILBI, 1999; LI *et al.*, 2012). Assim, uma vez que poucos eventos extremos são gerados ao longo da simulação, a cauda superior das precipitações diárias não é adequadamente descrita.

Diante do exposto, a construção de um modelo que combine em sua estrutura as abordagens paramétrica e não paramétrica parece ser uma alternativa coerente para se

abordar o problema de geração de precipitação diária. Nesse caso, chuvas de ocorrência mais frequente podem ser simuladas por meio de técnicas não paramétricas de reamostragem, ao passo que as chuvas com maior tempo de retorno e que serão objeto de extrapolação, podem ser modeladas por distribuições teóricas de probabilidade. Essa estrutura permite que se aproveite a flexibilidade de ajuste e a capacidade de reprodução de estatísticas de longo termo dos modelos não paramétricos e também possibilita a geração de eventos de grande magnitude, com valores significativamente superiores ao da máxima altura de precipitação observada. Esse novo modelo de geração será doravante denotado por modelo misto.

O modelo misto é construído a partir de uma matriz de probabilidade de transição de 3 estados, a saber, ausência de chuva, ocorrência de chuva moderada e ocorrência de chuva extrema, denotados respectivamente por (s), (u) e (e). A adoção de um modelo multi-estados tem por objetivos permitir a modelagem das chuvas de naturezas físicas distintas por meio de diferentes paradigmas e representar mais acuradamente a persistência de eventos de precipitação extrema, usualmente verificada em meses da estação chuvosa e que pode responder por uma parcela significativa das variâncias mensais observadas. Essa característica é expressa, sob uma perspectiva formal, pela probabilidade de permanência no estado de chuvas extremas na matriz de probabilidade de transição, em geral elevada nos meses mais úmidos, e pelo uso de um conjunto distinto de parâmetros para a distribuição das chuvas extremas, com relação às demais classes de chuva. Modelos de dois estados, por outro lado, geram alturas de chuva em dias consecutivos de maneira independente, a partir de um conjunto único de parâmetros, e assim, em função da natureza aleatória do processo de geração, raramente conseguem reproduzir algum tipo de estrutura de correlação temporal nos volumes precipitados.

O uso de um número maior de estados possibilitaria a reprodução ainda mais adequada da correlação serial entre as alturas de chuva (Gregory *et al.*, 1992; Wilks & Wilbi, 1999). Contudo, a ausência de um critério objetivo para a definição do número adequado de estados e dos limites entre os mesmos torna a solução desse problema uma tarefa extremamente complexa, inerentemente sujeita a procedimentos *ad hoc*, que assim impediriam uma generalização do método. Dessa forma, fez-se a opção de um

modelo que, ainda que de maneira subjetiva, somente faz distinção entre duas classes de precipitação, indiretamente associadas aos processos físicos de formação de chuva.

Para a estimação das probabilidades de transição, foi empregado um modelo de cadeia de Markov de primeira ordem, no qual o estado de ocorrência de chuva no dia atual depende apenas do estado do dia precedente. Essa escolha se fundamentou em duas questões práticas: (1) a dificuldade de estimação confiável dos parâmetros para modelos de ordem superior, cujo número cresce exponencialmente com o aumento da ordem do modelo de Markov, e se manifesta de maneira especialmente forte em amostras de tamanhos reduzidos (WILKS & WILBY, 1999); e (2) a conclusão de diversos estudos (Gregory *et al.*, 1992; Wilks & Wilbi, 1999; Apipattanavis *et al.*, 2007) que o desempenho de modelos de segunda ordem, baseado em índices como o critério de informação de Akaike (AIC) ou o critério de informação Bayesiano (BIC), não é, em geral, superior ao daqueles de primeira ordem, para uma ampla gama de bacias hidrográficas, com diferentes características climáticas e meteorológicas.

As probabilidades de transição entre os três estados são estimadas para cada dia do ano, com base na frequência histórica de transição e em um limiar estabelecido para diferenciar chuvas moderadas e extremas. A base diária para construção das matrizes de probabilidade de transição foi empregada com intuito de se permitir que essas probabilidades variem de maneira mais suave ao longo do ano, minimizando os efeitos das discontinuidades entre as mesmas na mudança entre intervalos discretos de grande duração, como quinzenas ou meses, durante a simulação.

A seleção de estados na matriz de probabilidade de transição, construção das sequências de dias secos e chuvosos, é realizada com auxílio de números aleatórios (NA) sorteados no intervalo (0,1). Para o primeiro dia da simulação, é necessário estabelecer se ocorre precipitação. Um simples modelo binário discreto, com probabilidade de sucesso especificada, pode ser empregado para esse fim. A probabilidade de sucesso pode ser obtida, por exemplo, pela frequência histórica de dias secos. Nesse caso, se o número aleatório sorteado for inferior à probabilidade de sucesso, o primeiro dia é seco. Caso contrário, o primeiro dia é úmido, e é considerada, por simplicidade, uma chuva convencional.

Para determinar a ocorrência de chuva no segundo dia, identifica-se o estado do primeiro dia e escolhe-se a linha correspondente na matriz de probabilidade de transição. Em seguida, sorteia-se um NA e verifica-se sua posição na referida linha. Considerando-se como exemplo o primeiro dia seco, o segundo dia será seco se $NA < p_{dd}$. Se $p_{dd} < NA < p_{dd} + p_{du}$, o segundo dia terá chuva convencional. Por fim, se $NA > p_{dd} + p_{du}$, o segundo dia terá uma chuva extrema. Para o terceiro dia em diante, o processo se repete.

Com relação à simulação das alturas de precipitação, abordagens distintas são empregadas em função do processo físico de formação de chuva. No caso de chuvas com alturas baixas a moderadas, as quais representam a grande maioria dos eventos chuvosos ao longo do ano, é adotada a abordagem não paramétrica. Essa escolha é baseada em três considerações. A primeira delas diz respeito às dificuldades na seleção de modelos distributivos para ajuste dos dados e ao tratamento das incertezas associadas à utilização de um dado modelo. A segunda consideração se relaciona ao esforço computacional necessário para se estimar os parâmetros das distribuições teóricas para cada dia do ano, intervalo esse que constitui a discretização temporal adotada na construção das matrizes de probabilidade de transição. A última consideração é relacionada à maior capacidade de modelos não paramétricos de reproduzir variâncias mensais e anuais. Em outras palavras, modelos não paramétricos são capazes de simular anos excepcionalmente secos ou chuvosos sem a necessidade de se recorrer a fatores de correção empíricos para o ajuste de variâncias.

As alturas de chuva baixas a moderadas são calculadas por meio de um procedimento de “*bootstrap*” convencional. O processo de reamostragem ocorre no interior de janelas móveis centradas no dia de interesse ao longo do período histórico. O gerador estocástico pode ser ajustado para selecionar entre todos os valores não nulos dentro da janela ou somente entre pares com combinação idêntica de ocorrência de chuva. No primeiro caso, a altura de chuva estimada em um dado dia não depende de sua posição no intervalo de dias chuvosos, o que faz com que o comportamento do gerador se aproxime daquele observado em modelos bipartidos paramétricos. No segundo caso, o gerador indiretamente condiciona a altura de chuva estimada em um dia à sua posição em um intervalo chuvoso, conferindo ao modelo um comportamento similar ao de modelos que empregam o “*bootstrap*” do k-ésimo vizinho mais próximo, considerando-

se que o número k de possíveis vizinhos é igual ao tamanho da amostra de pares idênticos de combinação. Nenhuma dessas abordagens, entretanto, é capaz de reproduzir integralmente a correlação entre as alturas de chuva no interior de um dado intervalo.

Métodos mais complexos, tais como o método de “*bootstrap*” do k -ésimo vizinho mais próximo, tem se mostrado mais eficientes que o “*bootstrap*” convencional em estruturas integralmente não paramétricas, que contemplam a modelagem da ocorrência de chuva e as alturas precipitadas e podem até mesmo empregar outras variáveis climáticas na simulação. Contudo, uma vez que no modelo proposto a ocorrência de chuva é modelada sob uma perspectiva paramétrica e em base diária, em lugar de intervalos de dias secos e chuvosos, e tendo-se em vista que a única variável de interesse é altura precipitada, torna-se difícil estabelecer um critério para se identificar os vizinhos mais próximos na simulação de volumes de precipitação. Diante disso, considerou-se que o “*bootstrap*” convencional é suficiente para a presente aplicação.

O procedimento de “*bootstrap*” é intrinsecamente relacionado ao tamanho das janelas móveis de reamostragem. Definir esse tamanho, contudo, é uma questão complexa, uma vez que uma janela de poucos dias pode não representar adequadamente a variabilidade das alturas de chuva para o dia em questão, enquanto uma janela de grande tamanho demandará esforço computacional significativamente maior. Dessa forma, deve-se avaliar o tamanho da janela por meio de algum índice objetivo. Propõe-se aqui comparar a média e a variância dos valores gerados para cada mês do ano com as referidas estatísticas dos registros observados. Tais índices serão aferidos para janelas de reamostragem de tamanho 8, 14 e 30 dias.

Uma deficiência comumente associada a procedimentos de “*bootstrap*” convencional é a repetição sistemática dos valores da amostra. Com intuito de se aumentar a variabilidade nas simulações, preservando-se as características estatísticas amostrais, introduziu-se no módulo não paramétrico do gerador uma função que permite associar pequenas perturbações aos valores selecionados em uma dada janela de reamostragem. As perturbações são obtidas da inversa de uma distribuição normal padrão com probabilidade de não excedência dada por um número aleatório.

No caso de chuvas extremas, a modelagem é realizada com auxílio de distribuições de probabilidades teóricas, tendo-se como principal objetivo a possibilidade de extrapolação para valores significativamente maiores que o máximo observado. Na presente tese, são empregados modelos probabilísticos com limite superior explícito, a saber, a distribuição Log-normal de 4 parâmetros (LN4), e a distribuição de valores extremos do tipo IV (EV4), tratados em detalhes nos itens 3.2.2.1 e 3.2.2.2. Tais distribuições são caracterizadas pela assimetria à direita e se adequam aos critérios estabelecidos por Koutsouyiannis (2004a,b), Papalexiou & Koutsouyiannis (2012) e Papalexiou *et al.* (2013) para a modelagem de precipitação diária.

Tal como visto em detalhes no capítulo introdutório, o uso de modelos distributivos limitados superiormente é envolto em imensas controvérsias. Argumentos favoráveis e contrários ao emprego desse tipo de distribuição foram apresentados e discutidos e, ainda que conclusões definitivas acerca do tema não possam ser obtidas a partir de critérios estritamente matemáticos, evidências físicas e geológicas (Enzel *et al.*, 1993; Jacoby *et al.*, 2008) fornecem subsídios para a adoção de modelos limitados. Além disso, autores como Li *et al.* (2013) e Chen & Brissette (2014) têm enfatizado que o uso de distribuições ilimitadas com caudas superiores demasiadamente pesadas leva à superestimação tanto da frequência quanto da magnitude dos eventos chuvosos. Em outras palavras, alturas de chuva fisicamente improváveis podem ser geradas inúmeras vezes ao longo da simulação, comprometendo a representação das precipitações máximas anuais e, em última análise, a estimação indireta dos quantis de enchentes. Sob tal perspectiva, o uso de um modelo limitado impede a geração de valores considerados improváveis.

O principal desafio associado ao uso de modelos limitados é a estimação do limite superior. Contudo, o paradigma Bayesiano, adotado no presente trabalho, oferece uma alternativa coerente para se realizar a estimação do limite superior. Essa abordagem permite que variáveis associadas ao limite superior, tais como a PMP, sejam incorporadas ao modelo não como valores fixos, mas associadas a uma descrição probabilística capaz de acomodar as incertezas relacionadas à sua estimação. Sob essa perspectiva, é possível construir uma estrutura de análise na qual a estimativa local de PMP é empregada como estimador do limite superior e a variabilidade das estimativas de PMP em uma dada região fornece indícios acerca da variabilidade do referido

parâmetro. Dessa forma, a estimativa local de PMP é associada a uma probabilidade de excedência que certamente se refletirá na estimativa do limite superior. O procedimento para construção da distribuição *a priori* para o limite superior é apresentado em detalhes no próximo item.

Uma vez que não usualmente não se dispõe de estimativas de PMP em base mensal, um único conjunto de parâmetros é considerado para cada um dos modelos distributivos empregados. Essa limitação, entretanto, pode ser contornada por meio da adequada fixação do limiar entre chuvas moderadas e extremas. Nesse contexto, deve-se estabelecer um valor tal que a probabilidade de chuvas de ordem de magnitude da estimativa de PMP seja bastante reduzida, ainda que não nula, durante a estação seca. Em outras palavras, o valor do limiar deve ser definido de maneira a se restringir a geração de valores de precipitação extremamente elevados em meses secos, e, ao mesmo tempo, possibilite a caracterização adequada dos eventos extremos nos meses chuvosos.

Apesar de fundamental ao correto funcionamento de geradores estocásticos, a definição de limiares em modelos híbridos é uma questão problemática sob um ponto de vista matemático, tendo-se em vista a continuidade das duas densidades no ponto de transição (FERRER & KATZ, 2008; GHOSH & RESNICK, 2010). Para alguns modelos particulares, é possível estimar limiares analiticamente, empregando-se o método de máxima verossimilhança (LI *et al.*, 2012). Na grande maioria dos casos, contudo, tal expediente é impossível. Uma alternativa para solucionar esse problema, apontada por Ferrer & Katz (2008), é calibrar o modelo, ou seja, realizar simulações para diferentes limiares e comparar com os dados por meio de índices definidos ou funções-objetivo. Tal procedimento é, por natureza, subjetivo, uma vez que diferentes limiares podem levar a resultados similares nas simulações e a continuidade das densidades não é garantida para qualquer um dos valores testados. No entanto, em função da estrutura mista na qual se baseia o gerador proposto, o problema de continuidade é automaticamente contornado, uma vez que não há necessidade de compatibilização de estimativas de parâmetros nos dois módulos de estimação de alturas de chuva. Assim, a calibração pode ser empregada como critério “objetivo” na definição do limiar. Na presente pesquisa foram selecionados como funções-objetivo a precipitação máxima anual, a qual sintetiza a influência do limiar na geração de valores extremos com

variados tempos de retorno, e estatísticas de longo termo, tais como a chuva média mensal e anual, que permitem aferir o efeito das diferentes frequências de ocorrência de eventos extremos na quantificação das variâncias para intervalos de tempo de maior duração.

A etapa final da geração estocástica de precipitação diária é a avaliação de desempenho dos modelos de geração. Em função da natureza estocástica das séries geradas, não é possível a comparação direta de uma série sintética qualquer com a amostra, por meio de testes estatísticos de aderência usuais. Dessa forma, estabeleceu-se na literatura uma série de índices para avaliação de desempenho de geradores estocásticos baseados em estatísticas, que avaliam a semelhança estatística entre as séries geradas e a amostra. Os índices empregados neste trabalho, similares aos propostos por Srikanathan & Pegram (2009), são: (1) estatísticas diárias, tais como média, desvio padrão, coeficiente de assimetria, alturas máximas de precipitação e número médio de dias chuvosos, para cada mês do ano; (2) estatísticas de longo termo, como média e variância das precipitações médias mensais e anuais; e (3) comportamento das precipitações máximas anuais, por meio da curva de quantis.

Para validação dos modelos, serão geradas 1.000 séries, de mesmo tamanho das amostras observadas em cada uma das bacias hidrográficas em estudo. Para as estatísticas diárias, serão tomados os valores médios das simulações para a comparação com os dados. Constitui uma exceção a esse procedimento a altura máxima de precipitação diária, cujos valores observados serão confrontados com os máximos simulados em todo o conjunto de 1.000 simulações, considerando-se o ano hidrológico. As estatísticas mensais e anuais e as curvas de quantis também serão avaliadas em função de valores médios do conjunto de simulações.

Por fim, após a validação dos modelos, serão geradas 1.000 séries de 10.000 anos de precipitação diária, as quais serão empregadas na estimativa indireta dos quantis de enchentes e avaliação das incertezas relacionadas a esse quantis após a simulação hidrológica.

4.2 Construção das distribuições a priori dos parâmetros dos modelos LN4 e EV4

A construção da distribuição *a priori* para os parâmetros dos modelos probabilísticos abordados nesta pesquisa constitui a etapa mais importante da inferência Bayesiana. Por outro lado, é a etapa que envolve maior subjetividade, uma vez que depende do conhecimento do especialista que está realizando a modelagem e também de informações extras, tais como aquelas relacionadas às características de eventos extremos em locais distintos daqueles em que se está realizando o estudo.

Dentre os parâmetros das distribuições LN4 e EV4, aquele para o qual se tem informação disponível, possibilitando a construção de uma distribuição *a priori* informativa, é o limite superior α . Tal parâmetro apresenta uma relação interpretável com as características hidrometeorológicas da bacia, evidenciada por um estimador “quase-determinístico”, a PMP, cuja variabilidade fornece indícios para se resumir a incerteza associada ao limite superior. Nesse contexto, é possível se atribuir maior densidade de probabilidade a uma dada região do domínio do parâmetro, a qual se caracteriza pela maior concentração de estimativas de PMP. No caso dos demais parâmetros dos modelos LN4 e EV4, é extremamente difícil realizar qualquer associação física com características climáticas da bacia. Diante disso, não é possível se atribuir massa a uma região específica do domínio dos parâmetros, o que leva à adoção das chamadas distribuições *a priori* não informativas para os mesmos.

4.2.1 Construção do limite superior para a bacia do rio Pará

Conforme será visto no capítulo 5, uma aplicação do método é realizada para a bacia do rio Pará, na estação de Ponte do Vilela. Essa bacia dispõe de uma quantidade significativa de dados pluviométricos e também estudos de geração estocástica de chuva e estimação indireta de cheias. No entanto, são poucas as informações acerca de um provável limite superior para as precipitações máximas anuais.

A bacia em estudo conta com uma estimativa de PMP e um conjunto de dados de precipitação diária observados entre 1942 e 2002, totalizando 61 anos. Os dados, contudo, são parte integrante da função de verossimilhança e, assim, não podem ser utilizados na construção da distribuição *a priori* para o limite superior. Por outro lado, a estimativa de PMP isoladamente não é suficiente para caracterizar toda a incerteza relacionada ao limite superior. Sob essa perspectiva, a situação ideal seria aquela na

qual uma grande amostra de estimativas de PMP para uma mesma bacia, calculadas por métodos idênticos em diferentes épocas, utilizando toda a informação hidrometeorológicas, estivesse disponível. Nesse caso, seria possível avaliar objetivamente a variabilidade da PMP e elaborar modelos estatísticos capazes de descrever tal variabilidade, ou, sob o ponto de vista Bayesiano, a variabilidade da PMP constituir-se-ia na base de conhecimento do especialista sobre o limite superior. Obviamente, tal situação não existe em termos práticos, o que conduz à adoção de abordagens baseadas na construção de distribuições *a priori* a partir de estimativas de PMP de caráter regional.

Dada a pouca disponibilidade de estimativas de PMP meteorológicas em Minas Gerais, fez-se a opção de empregar o método estatístico em um conjunto de 118 estações pluviométricas distribuídas por todo o estado, cobrindo uma grande diversidade de condições climáticas e hidrometeorológicas. O conjunto de estações e as estimativas de PMP de 1 dia são apresentados na Tabela 4.1.

A Figura 4.1 apresenta o histograma de frequências das estimativas regionais de PMP. A linha contínua representa uma distribuição gama ajustada ao conjunto de estimativas, enquanto a linha tracejada mostra uma distribuição normal, sendo a aderência de ambos os modelos atestada pela não rejeição da hipótese nula no teste de Kolmogorov-Smirnov, ao nível de significância de 5%. A simples análise do histograma mostra que as duas distribuições são candidatas razoáveis para modelar o conjunto de 118 estimativas de PMP. No entanto, o uso do modelo probabilístico normal como distribuição *a priori* apresenta dois inconvenientes: (1) o histograma é ligeiramente assimétrico à direita, com coeficiente de assimetria igual a 0,25. Ainda que esse valor seja da mesma magnitude do erro padrão da estimativa da referida estatística ($s_e = 0,22$), o uso de um modelo simétrico é inconsistente com as características amostrais; (2) a distribuição normal é ilimitada também à esquerda, enquanto as estimativas de PMP são obrigatoriamente positivas. Nesse caso, seria necessário truncar o modelo distributivo para acomodar essa característica física da variável hidrológica, tornando a distribuição *a priori* especialmente complexa para as simulações do algoritmo MCMC. Diante disso, apenas o modelo gama será considerado nesta pesquisa para modelar as estimativas de PMP.

Tabela 4.1 – Estimativas de PMP estatística de 1 dia em Minas Gerais

Estação	Código	Precipitação Máxima Média Anual (mm)	Desvio Padrão (mm)	Fator k_m	PMP (mm)
Boca da Caatinga	01443000	85,6	22,08	15,65	431,20
Manga	01443001	74,6	15,17	16,09	318,67
Carinhanha	01443002	79,7	20,16	15,88	399,90
São Gonçalo	01444000	88,1	24,82	15,55	474,00
Capitânea (Várzea da Larga)	01444001	87,6	23,62	15,57	455,38
Juvenília (Escola Caio Martins)	01444004	81	22,94	15,83	444,13
Lagoa das Pedras	01444005	95,6	23,55	15,26	455,04
Fazenda Porto Alegre	01444017	92,8	30,01	15,37	554,09
Colônia do Jafba	01543002	77,6	18,62	15,97	374,93
Janaúba - Colonização II	01543013	74,9	18,12	16,08	366,27
São Francisco	01544012	90,7	19,61	15,45	393,75
Pedras de Naria da Cruz	01544017	81,5	23,08	15,81	446,48
São João da Ponte	01544019	76,3	20,45	16,02	403,89
Arinos	01546000	88	17,69	15,55	363,23
Capitão Enéas	01643020	90	24,42	15,48	467,94
São João da Vereda	01644028	82,4	19,83	15,78	395,21
São Romão	01645000	89,3	22,39	15,50	436,50
Santo Inácio	01645002	92,2	14,45	15,39	314,64
Barra do Escuro	01645003	86,3	16,03	15,62	336,65
Vila Urucula	01645005	86,6	20,18	15,61	401,58
Cachoeira da Mateiga	01645009	87,6	26,80	15,57	504,83
Fazenda Água Branca	01645013	101,8	26,70	15,04	503,32
Porto dos Poções	01646000	73,8	20,56	16,13	405,37
Unaí	01646001	90,7	16,31	15,47	342,47
Santo Antonio do Boqueirão	01646003	87,5	16,96	15,57	351,71
Várzea da Palma	01744009	82,7	19,69	15,76	393,13
Lassance	01744010	81,2	18,05	15,82	366,85
Caatinga	01745000	93,8	17,76	15,33	366,04
Cachoeira do Paredão	01745001	94,4	21,38	15,31	421,78
Porto Cavallo	01745007	98,2	25,93	15,17	491,55
Porto Extrema	01746001	106,1	27,45	14,88	514,62
Santa Rosa (Porto Buriti)	01746002	84,7	15,20	15,68	323,15
Ponte BR-040 (Rio da Prata)	01746006	78,3	14,01	15,94	301,69
Ponte BR-040 (R. Paracatu)	01746007	88,2	19,54	15,55	392,07
Guarda-Mor	01747005	95,8	22,52	15,26	439,47
Usina Paraúna	01843000	78,6	16,53	15,93	341,86
Gouveia	01843002	87	23,92	15,59	460,03
Santo Hipólito	01844001	82,4	21,60	15,78	423,22
Presidente Juscelino	01844009	76,6	12,25	16,01	272,72
Ponte do Licínio (Jusante)	01844010	84,5	25,68	15,69	487,43
Corinto	01844017	83,7	24,14	15,72	463,20
Fazenda São Félix	01845002	87,5	18,63	15,57	377,70
Lagoa do Gouveia	01845004	95,8	21,23	15,26	419,75
São Gonçalo do Abacté	01845013	99,2	24,70	15,13	473,00
Iiros	01845014	84,3	15,41	15,70	326,22
Major Porto	01846003	82,7	18,89	15,78	380,29
Presidente Olegário	01846005	98	27,45	15,18	514,51
Ponte Firme	01846016	99	21,69	15,14	427,29
Leal de Patos	01846017	87,1	21,31	15,59	419,37
MIn. Morro Velho	01943000	99,3	27,96	15,13	522,28

Sabará	01943006	93,8	24,36	15,33	467,27
Vespasiano	01943009	83,9	19,24	15,72	386,25
Caeté	01943010	97,1	24,38	15,21	467,87
Caixa de Areia	01943022	106,4	28,37	14,87	528,33
Taquaraçu	01943023	90,6	21,14	15,45	417,27
José de Melo (Nova União)	01943024	92,8	24,16	15,37	464,20
Vau da Lagoa	01943035	90,8	18,78	15,45	380,90
Ponte Nova do Paraopeba	01944004	90,5	22,91	15,46	444,72
Fazenda Escola Florestal	01944007	88	21,02	15,55	415,00
Pedro Leopoldo	01944009	80,2	17,48	15,86	357,48
Horto Florestal (Paraopeba)	01944010	89,9	27,18	15,48	510,72
Jaguaruna (Onça do Pitangui)	01944011	87,8	20,02	15,56	399,39
Velho da Taipa	01944021	97,3	23,16	15,39	448,76
Fazenda Vargem Bonita	01944024	83,9	25,55	15,72	485,42
Barro Preto	01944026	95,7	27,47	15,26	514,94
Juatuba	01944027	99,2	25,11	15,13	479,08
Ponte da Taquara	01944031	85,4	24,38	15,66	467,07
Pitangui	01944032	99,6	24,56	15,12	470,80
Mateus Leme	01944048	96,5	19,91	15,23	399,71
Papagaios	01944049	94,7	25,91	15,30	491,12
Betim	01944055	85,4	16,62	15,66	345,60
Fortuna de Minas	01944059	95,2	20,28	15,28	405,15
Barra do Funchal	01945002	84,5	20,31	15,69	403,25
Estação Álvaro da Silveira	01945004	86,9	17,51	15,60	359,95
Bom Despacho	01945008	107,2	25,23	14,84	481,66
Dores do Indaiá	01945019	95,4	22,27	15,27	435,56
Abaeté	01945035	95,5	18,15	15,27	372,62
Taquaral	01945037	94,4	16,47	15,31	346,63
Porto das Andorinhas	01945038	90,4	27,33	15,46	512,93
Martinho Campos	01945039	90,9	20,23	15,44	403,24
Tapiraí	01946000	88,2	17,16	15,55	354,97
São Gotardo	01946009	95,6	24,07	15,26	463,08
Lagoa Grande	02043002	93,2	27,60	15,36	517,02
Rio do Peixe	02043004	99,5	24,44	15,12	469,10
Conselheiro Lafaiete	02043005	87,6	17,31	15,57	357,17
Congonhas	02043013	83	21,50	15,75	421,63
Represa das Codornas	02043042	102,5	29,71	15,01	548,45
Represa do Miguelão	02043043	101,8	28,21	15,04	525,91
Itaúna	02044002	86,9	15,35	15,60	326,32
Carmo Cajuru	02044003	87,7	16,79	15,57	341,35
Divinópolis	02044006	91,1	22,30	15,44	435,23
Entre Rios de Minas	02044007	83,4	22,27	15,54	434,45
Melo Franco	02044008	84,9	14,11	15,68	306,08
Fazenda Campo Grande	02044009	88,5	14,82	15,54	318,80
Ibirité	02044012	100,3	30,23	15,09	556,48
Fazenda Benedito Chaves	02044016	89,3	20,35	15,50	404,87
Fazenda Vista Alegre	02044019	88	19,72	15,55	394,76
Calambau	02044020	89,9	14,34	15,48	311,84
Alto da Boa Vista	02044021	99,5	22,23	15,12	435,62
Fazenda Curralinho	02044024	98,2	22,28	15,17	436,07
Fazenda Coqueiros	02044026	86,8	17,53	15,60	360,26
Usina João Ribelro	02044040	88,2	24,03	15,55	461,73
Fazenda Laranjeiras	02044041	93,1	17,03	15,36	354,69
Carmo da Mata (ETA)	02044042	99,3	28,17	15,13	525,43
Estiva	02044043	106,2	20,14	14,88	405,89

Serra da Saudade	02044047	86,9	16,34	15,60	341,75
Jardim	02044052	93,5	17,67	15,34	364,56
Escola de Veterinária	02044053	99,2	26,42	15,13	499,01
BambuÍ	02045001	88,6	22,32	15,53	435,32
Iguatama	02045002	81,7	20,71	15,80	408,99
Lamounier	02045005	98,2	26,85	15,17	505,42
Arcos (COPASA)	02045010	101,8	27,34	15,04	512,91
Lagoa da Prata	02045011	107,6	29,16	14,83	539,92
Piumhi	02045012	96,7	26,32	15,22	497,43
Santo Antonio do Maonte	02045013	95,3	21,92	15,28	430,21
Fazenda Ajudas	02046007	86,5	21,72	15,61	425,67
Vargem Bonita	02046013	94,5	17,98	15,31	369,75

A variabilidade das estimativas de PMP constitui um importante indicativo das características de variação do limite superior das precipitações diárias. Contudo, a distribuição *a priori* para o limite superior certamente não é idêntica à distribuição das estimativas de PMP, uma vez que as mesmas foram realizadas em regiões com diferentes características climáticas e hidrometeorológicas. Por outro lado, é razoável supor que o histograma teria forma semelhante, possivelmente com menor variância, se tais características fossem similares. Com base nesses argumentos e assumindo que a PMP local é um estimador conveniente para α , parece plausível admitir que a distribuição gama seja adequada para modelar também a distribuição *a priori* do limite superior.

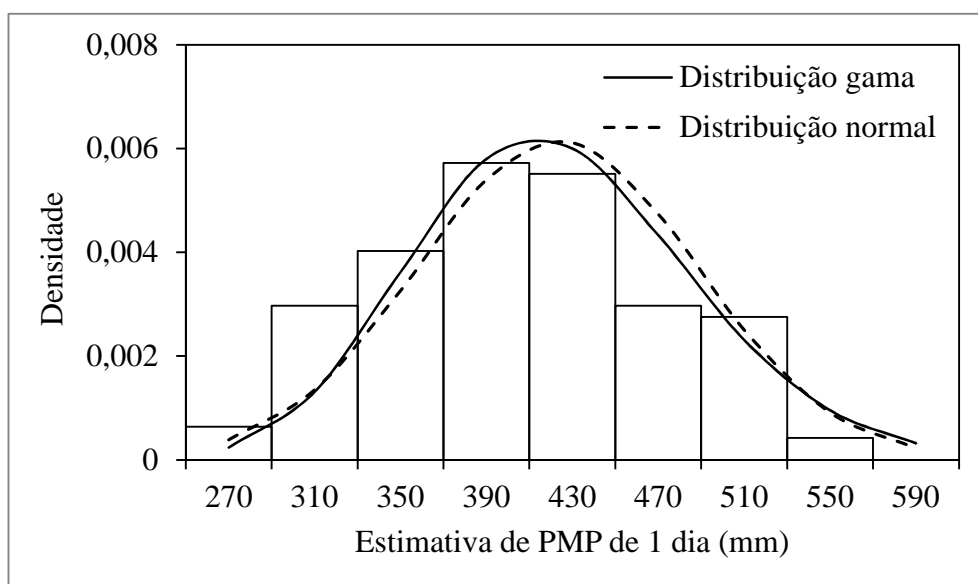


Figura 4.1 – Histograma de frequências das estimativas de PMP de 1 dia em Minas Gerais

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

$$f_X(x|\Theta) = \frac{\beta^\rho}{\Gamma(\rho)} x^{\rho-1} \exp(-\beta x), \quad x > 0 \quad (4.1)$$

Adotada a distribuição gama de dois parâmetros como distribuição *a priori* do limite superior, faz-se necessário estimar os parâmetros da mesma. Uma vez que não há momentos amostrais disponíveis para a estimação dos parâmetros, dado que não há uma série de PMP's na bacia em estudo, pode-se descrever a função gama por meio de duas características: o coeficiente de variação regional (CV) e a frequência empírica de não-excedência da PMP local.

O coeficiente de variação regional fornece a estimativa do parâmetro de forma. Pelo método dos momentos, tem-se:

$$\rho = \frac{1}{CV^2} \quad (4.2)$$

O parâmetro de escala β pode ser estimado admitindo-se uma probabilidade de não-excedência p para a estimativa local de PMP, ou seja, o parâmetro deve ser tal que $P(\alpha \leq PMP|\rho, \beta) = p$. Associar uma probabilidade de não-excedência para a PMP, contudo, é uma tarefa complexa, mesmo quando estão disponíveis a melhor informação hidrometeorológica e ferramentas adequadas para modelagem. Diante do exposto, a alternativa para se determinar uma probabilidade p de não-excedência da PMP local também pode ser fundamentada na análise regional. Nesse contexto, a ideia é se atribuir uma frequência de não-excedência empírica à estimativa local da PMP, com base em sua posição de plotagem no conjunto de 118 estimativas de PMP. As estimativas de PMP foram normalizadas pela chuva média anual, com objetivo de se extrair a influência climática da análise.

A adoção de um modelo probabilístico ilimitado para resumir a incerteza relacionada ao limite superior pode parecer incoerente. Tal expediente, contudo, permite que se considere incrementos de qualquer magnitude ao valor do referido parâmetro. Em outras

palavras, não se faz necessário ao especialista prescrever, *a priori*, um valor finito ao limite superior, o que indiretamente traduz a falta de conhecimento prévio sobre o mesmo. Ao mesmo tempo, uma distribuição com cauda exponencial, tal como a gama, atribui probabilidades bastante reduzidas àqueles valores de precipitação considerados implausíveis nas bacias em estudo, o que faz com que as estimativas *a posteriori* do limite superior se situem em uma faixa de magnitudes fisicamente mais realistas, em especial quando se tem em vista que, via-de-regra, as alturas de precipitação observadas são significativamente inferiores às estimativas de PMP e agregam pouca informação à análise quando da aplicação do teorema de Bayes.

Para os demais parâmetros, a ausência de relação direta com o fenômeno físico faz com que sejam empregadas as chamadas distribuições *a priori não informativas*, caracterizadas por atribuir massa praticamente constante ao longo do espaço paramétrico. Essas distribuições não agregam conhecimento prévio à análise. Detalhes adicionais acerca das distribuições *a priori* serão apresentados no capítulo 5.

4.2.2 Construção do limite superior para a bacia do rio American

A segunda aplicação do método proposto nesta tese será realizada na bacia do rio American, na barragem de Folsom, no estado americano da Califórnia. O interesse da referida aplicação reside no fato de que a bacia possui um conjunto de estimativas de paleocheias que podem atestar a adequação das hipóteses admitidas na presente pesquisa e constituir um indicativo da qualidade geral do método após a geração das curvas de quantis de enchentes. Também nessa bacia há uma estimativa de PMP meteorológica e um amplo banco de dados pluviométricos. Contudo, em algumas das séries temporais das 5 estações empregadas na determinação da chuva média espacializada, foram verificadas inconsistências e falhas nos registros, o que limitou o tamanho da amostra utilizada na geração estocástica de chuva a 35 anos, compreendidos entre 1961 e 1995.

A construção da distribuição *a priori* para o limite superior das precipitações na bacia do rio American é realizada de maneira semelhante àquela empregada na bacia do rio Pará. A principal distinção entre as aplicações é que, no estado da Califórnia, há um conjunto de 39 estimativas de PMP meteorológicas, as quais constituem a amostra para a determinação do coeficiente de variação regional de α . Estimativas de PMP

meteorológica são determinadas em função das áreas das respectivas bacias. Dessa forma, é necessário se empregar curvas altura-área-duração (ou DAD, do acrônimo em inglês para *Depth-Area-Duration*), para se transpor as estimativas de PMP à bacia do rio American. Foram empregadas aqui as curvas DAD contidas no “*Hydrometeorological Report n° 59 - Probable Maximum Precipitation for California*” (USACE, 1999), publicado por um conjunto de órgãos do governo americano que engloba o NOAA (*National Oceanic and Atmosphere Administration*) e o U. S. Army Corps of Engineers. Os valores das estimativas de PMP para cada bacia empregada, já corrigidos para a área da bacia do rio American, em Folsom, são apresentados na Tabela 4.2.

A Figura 4.2 apresenta o histograma de frequência das estimativas de PMP de 1 dia para a bacia do rio American, em Folsom. São mostradas também as distribuições gama (linha contínua) e normal (linha tracejada) ajustadas ao conjunto de estimativas, e cuja aderência é atestada pela não rejeição da hipótese nula no teste de Kolmogorov-Smirnov, ao nível de significância de 5%. Ainda que ambos os modelos sejam razoáveis para modelar as estimativas de PMP, somente a distribuição gama será usada nesse contexto, tendo-se em vista os argumentos apresentados no item 4.2.1.

A distribuição *a priori* para o limite superior das precipitações diárias será também uma gama nessa aplicação, e seus parâmetros serão estimados de maneira idêntica àquela descrita para a bacia do rio Pará: o parâmetro de forma será obtido por meio do coeficiente de variação regional das estimativas de PMP, e o parâmetro de escala por meio da probabilidade de não-excedência da estimativa local de PMP, adimensionalizada pela chuva média anual.

Tabela 4.2 – Estimativas de PMP meteorológica em bacias da Califórnia, reduzidas para a área da bacia do rio American, em Folsom

Região	Estação	Área (mi ²)	Área (km ²)	PMP (in)	PMP (mm)	% 16mi ²	PMP reduzida (16mi ²)	PMP Folsom
Northeast Clearlake								
	Whiskeytown	242	621	20,04	509,02	0,840	605,97	407,82
Northeast Trinity								
	SlyPark	17	44	20,52	521,21	0,990	526,47	355,57
Sierra								
	Stampede	130	335	13,24	336,30	0,865	388,78	262,43
	Shasta	3027	7816	12,29	312,17	0,639	488,52	329,75
Valley								
	Contra Loma	1,37	3,56	14	355,60	1,000	355,60	200,20
	Little Panache	82	212	11,01	279,65	0,928	301,35	169,66
	San Luis	63	164	14,56	370,53	0,926	399,04	224,67
	Los Banos	156	402	12,58	319,53	0,951	375,46	211,39
	Mammoth	346	890	19,7	500,38	0,994	721,01	403,93
	New Melones	944	2432	14,21	360,93	0,843	339,35	313,03
	Autuan	973	2516	17,72	450,09	0,857	708,57	397,80
	Front	1591	4102	15,98	405,09	0,982	610,12	345,20
Central								
	Ortega	0,0162	0,042	22	558,80	1,000	558,80	351,49
	Lamo	0,14	0,36	26	660,40	1,000	660,40	415,39
	Glenn-Arne	0,55	1,419	26,5	673,10	1,000	673,10	423,38
	Casitas	39	101	27,55	699,77	0,922	758,97	477,39
	Sutherland	50	129	19,59	492,51	0,905	544,21	342,31
	San Vicente	76	196	12,82	325,63	0,870	374,25	235,43
	Sweetwater	88	227	12,97	329,44	0,861	382,62	240,67
	Lower Otay	93	240	13,66	346,96	0,858	404,35	254,36
	Lower San	94	243	19,87	503,43	0,858	588,74	369,04
	Moresno	199	511	21,6	548,64	0,949	646,04	406,85
	Burre-Lake	121	310	18,57	471,68	0,811	560,83	352,78
Southeast								
	Sepulveda	156	402	10,60	269,21	0,827	610,83	364,25
	Hansen	157	405	29,72	759,49	0,927	802,05	554,03
	San Juan	177	457	23,71	603,03	0,819	797,30	501,54
	El Capitan	199	511	16,33	414,17	0,815	576,13	363,64
	Henshaw	293	754	17,44	442,98	0,811	546,21	343,37
	Santa Fe	2-8	640	29,90	759,96	0,796	935,01	599,82
	Lake Hodges	253	653	12,65	321,26	0,9	405,4	256,8
	Whittier Narrow	387	992	18,42	467,87	0,784	596,77	375,37
	Bradbury	477	1226	25,09	637,29	0,763	835,24	525,36
	Santa Margarita	774	1992	11,19	284,23	0,721	394,21	247,96
	Twitchell	1135	2928	10,97	278,64	0,680	409,76	257,4
	Prado Dam	2245	5792	12,19	309,63	0,607	510,05	320,85

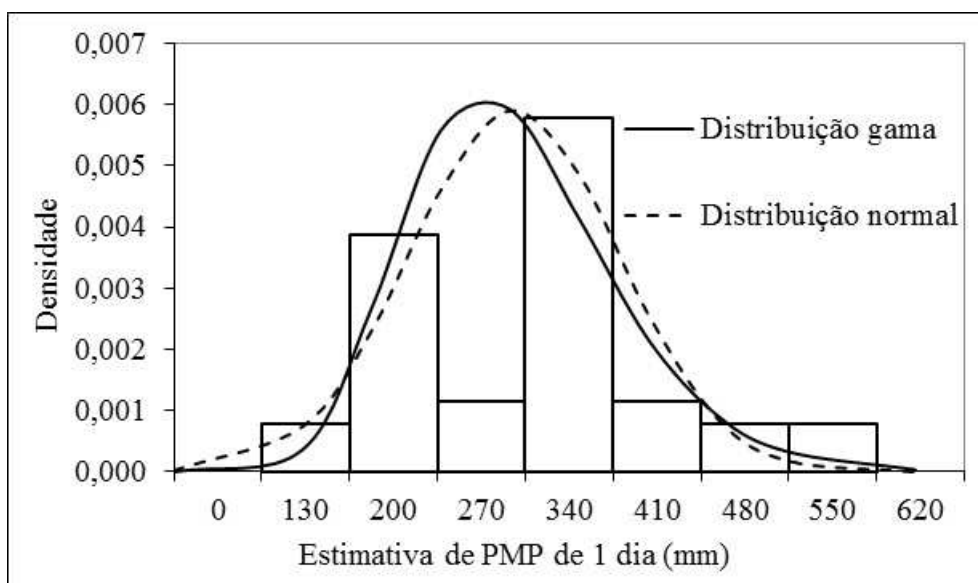


Figura 4.2 – Histograma de frequências das estimativas de PMP de 1 dia no estado da Califórnia

4.3 Calibração dos parâmetros do modelo Rio Grande de transformação chuva-vazão

O problema de calibração de parâmetros de modelos de transformação chuva-vazão tem sido objeto de grande interesse na comunidade científica, em especial devido à constatação de que os métodos mais difundidos nesse contexto, tais como o método de mínimos quadrados, são, frequentemente, incapazes de identificar corretamente os valores dos melhores conjuntos de parâmetros. De maneira geral, esses métodos baseiam-se em fortes hipóteses para o comportamento dos resíduos, tais como distribuição Gaussiana, independência e homocedasticidade, as quais são frequentemente violadas após a calibração. Além disso, é bastante usual que diferentes conjuntos de parâmetros conduzam a desempenhos similares na calibração, o que torna a determinação do “melhor” conjunto de parâmetros uma tarefa complexa e subjetiva.

O emprego de métodos Bayesianos tem trazido grandes avanços ao tratamento do problema de calibração de parâmetros de modelos hidrológicos, uma vez que os mesmos não se concentram em buscar valores ótimos para o vetor paramétrico, e sim na obtenção de distribuições *a posteriori* que sintetizam o comportamento probabilístico de cada parâmetro após a calibração e fornecem meios para quantificação das incertezas relacionadas a suas estimativas. Contudo, grande parte desses modelos ainda apresenta severas limitações relacionadas ao tratamento formal da função de verossimilhança. O método GLUE (Beven & Binley, 1992), por exemplo, realiza a quantificação da verossimilhança de maneira subjetiva, por meio de índices de qualidade de ajuste em lugar de um modelo estatístico teórico, o que, em última análise, impede que as hipóteses admitidas na modelagem sejam testadas *a posteriori* (KAVETSKI *et al.*, 2006).

Em vista do exposto, é conveniente conferir um enfoque estatístico formal à inferência dos parâmetros do modelo Rio Grande, de maneira a se verificar, *a posteriori*, a adequação do modelo de resíduos assumido na calibração. Na presente pesquisa, será adotado o modelo proposto por Schoups & Vrugt (2010) e nomeado por Silva *et al.* (2014) por função de verossimilhança generalizada. O referido modelo baseia-se na seguinte equação de regressão:

$$\tilde{y} = E + e \quad (4.3)$$

na qual \tilde{y} é um vetor com N valores observados de vazão, E é um vetor com as vazões médias correspondentes, e e é um vetor de resíduos aleatórios com média zero e que incorpora os erros relacionados à medição de vazões, aos *inputs* e à estrutura do modelo hidrológico.

Para um dado modelo de simulação chuva-vazão h , cuja estrutura é sintetizada pelo vetor paramétrico θ_h , as vazões médias correspondentes às observadas podem ser expressas, em função do tempo, da seguinte maneira:

$$E_t = Y_{h,t}(X|\theta_h)\mu_t \quad (4.4)$$

na qual $Y_{h,t}$ é uma função dos *inputs* X e dos parâmetros θ_h , e μ_t é um fator multiplicativo que sumaria o viés introduzido na vazão modelada pelas diversas fontes de erro na simulação hidrológica. A parametrização proposta para tal multiplicador é a seguinte:

$$\mu_t = \exp(\mu_h Y_{h,t}) \quad (4.5)$$

onde μ_h é um parâmetro a ser inferido dos dados e cujo intuito é amplificar a não-linearidade na resposta esperada da simulação hidrológica.

Com relação aos resíduos na equação 4.3, os mesmos são usualmente assumidos homocedásticos, independentes e identicamente distribuídos de acordo com um modelo Gaussiano $N(0, \sigma^2_e)$. No entanto, essa hipótese raramente se verifica em aplicações práticas (KAVETSKI *et al.*, 2002; KAVETSKI *et al.*, 2006a; SCHOUPS & VRUGT, 2010; SILVA *et al.*, 2013). Para contornar esse problema, Schoups & Vrugt (2010) preconizam o uso de um modelo capaz de acomodar a correlação temporal, a não normalidade e a heterocedasticidade dos resíduos ao longo da simulação. Tal modelo pode ser expresso da seguinte maneira:

$$\Phi_p(B)e_t = \sigma_t a_t, \text{ na qual } a_t \sim SEP(0, 1, \xi, \beta) \quad (4.6)$$

onde $\Phi_p(B)$ é um polinômio autoregressivo com p parâmetros ϕ_i , B é um operador de defasagem do tipo $\beta^i e_t = e_{t-i}$, σ_t é o desvio padrão no tempo t , e a_t é um erro aleatório

independente e identicamente distribuído, com média zero e desvio padrão unitário, descrito por uma distribuição exponencial potência assimétrica (SEP, do acrônimo em inglês para *Skew Exponential Power*), com parâmetros ξ (assimetria) e β (curtose), cuja função é acomodar a não normalidade dos resíduos. O modelo distributivo é expresso por:

$$p(a_t|\xi, \beta) = \frac{2\sigma_\xi}{\xi + \xi^{-1}} w_\beta \exp\left\{-c_\beta |a_{\xi,t}|^{\frac{2}{(1+\beta)}}\right\} \quad (4.7)$$

sendo $a_{\xi,t} = \xi^{-\text{sign}(\mu_\xi + \sigma_\xi a_t)} \times (\mu_\xi + \sigma_\xi a_t)$, e $\mu_\xi, \sigma_\xi, c_\beta, w_\beta$ são valores calculados a partir da assimetria ξ e da curtose β . A distribuição será simétrica se $\xi = 1$, assimétrica positivamente se $\xi > 1$ e negativamente se $\xi < 1$. Caso ocorra simetria, o modelo probabilístico se converte em uma distribuição uniforme quando $\beta = -1$, em uma distribuição Gaussiana quando $\beta = 0$ e uma distribuição de Laplace quando $\beta = 1$. Nesse contexto, os parâmetros ξ e β reduzem as limitações da hipótese de erros Gaussiano. Em particular, valores de $\beta > 0$ resultam modelos probabilísticos com caudas mais pesadas que a distribuição normal, fato esse que pode ser de grande utilidade para tornar a inferência de parâmetros mais robusta com relação à presença de *outliers* (SCHOUPS & VRUGT, 2010).

O modelo autoregressivo de ordem p assumido na equação 4.6 tem por objetivo descrever a dependência e autocorrelação entre os erros da simulação. A heterocedasticidade, por sua vez, é explicitamente incorporada ao modelo fazendo-se com que o desvio padrão varie com as vazões médias E_t da seguinte maneira:

$$\sigma_t = \sigma_0 + \sigma_1 E_t \quad (4.8)$$

onde σ_0 e σ_1 constituem, respectivamente, o intercepto e a inclinação do modelo de desvio padrão e são inferidos dos dados.

O logaritmo da função de verossimilhança generalizada é então dado por:

$$L(\theta|Y) = n \times \log \frac{2\sigma_\xi w_\beta}{\xi + \xi^{-1}} - \sum_{t=1}^n \log \sigma_t - c_\beta \sum_{t=1}^n |a_{\xi,t}|^{\frac{2}{(1+\beta)}} \quad (4.9)$$

Para obter a distribuição conjunta dos parâmetros do modelo hidrológico e dos parâmetros do modelo de resíduos foi empregado o algoritmo DREAM (do acrônimo em inglês para *Differential Evolution Adaptive Metropolis*), proposto por Vrugt *et al.* (2009). A premissa básica desse método consiste em simular várias cadeias paralelamente, com intuito de se explorar de maneira mais completa o espaço paramétrico. Durante a evolução para a distribuição *a posteriori*, o algoritmo ajusta automaticamente a escala e orientação da distribuição de referência proposta, $q(\cdot)$, o que resulta em maior eficiência na simulação de distribuições alvo complexas, não lineares ou mesmo multimodais (VRUGT *et al.*, 2009). Para detalhes adicionais acerca do algoritmo DREAM, sugere-se reportar aos trabalhos de Vrugt *et al.* (2008, 2009) e Vrugt & ter Braak (2011).

O código utilizado no procedimento de calibração foi desenvolvido por Silva *et al.* (2014), a partir de uma biblioteca implementada por Guillaume & Andrews (2012) para o software R e de rotinas em Fortran que realizam a síntese hidrológica do modelo Rio Grande (EHR/UFGM, 2001). O referido código foi empregado na presente pesquisa sem modificações. A inicialização do algoritmo de calibração requer a especificação de informações relacionadas ao modelo hidrológico e ao processo de simulação MCMC. No caso do modelo Rio Grande, é necessário fornecer as séries de precipitação, evaporação e vazões para o período de calibração especificado, além do coeficiente relacionado à forma da bacia para obtenção do hidrograma unitário de Clark. Já no caso das simulações MCMC por meio do algoritmo DREAM, as principais informações são listadas na Tabela 4.3.

Tabela 4.3 – Informações necessárias à utilização do método DREAM

Símbolo	Descrição
<i>func.type</i>	Especificação da função de verossimilhança a ser considerada
<i>pars</i>	Lista com o nome e respectivos intervalos de incerteza das variáveis aleatórias
<i>nseq</i>	Número de cadeias avaliadas em paralelo
<i>ndraw</i>	Número máximo de iterações
<i>burn-in</i>	Número de iterações desconsideradas ao início da simulação
<i>thin.t</i>	Intervalos entre avaliações da função de verossimilhança utilizados para composição da amostra
<i>Rthres</i>	Estatística de Gelman-Rubin para diagnóstico de convergência

Fonte: Silva *et al.* (2014)

Cabe ressaltar que no algoritmo empregado nesta pesquisa, todas as distribuições *a priori*, sejam elas do módulo de produção do modelo Rio Grande ou da função de verossimilhança generalizada, são uniformes. Essa é uma limitação imposta pela estrutura de construção do algoritmo DREAM. Para cada uma das aplicações previstas nesta tese, nominalmente, nas bacias do rio Pará, em Minas Gerais, e do rio American, no estado americano da Califórnia, foram avaliadas, a partir de características dos hidrogramas observados, as faixas de variação mais convenientes para os referidos parâmetros, tendo-se em vista a qualidade de calibração e a convergência do algoritmo de calibração à distribuição conjunta *a posteriori*.

Nas aplicações descritas nos Capítulos 5 e 6, foram considerados períodos de calibração e validação de 5 e 3 anos, respectivamente. A avaliação da qualidade de calibração é efetuada pelos seguintes índices:

- Coeficiente de Nash-Sutcliffe (NS), dado pela seguinte equação:

$$NS = 1 - \frac{\sum_{t=1}^n (q_{sim,t} - q_{obs,t})^2}{\sum_{t=1}^n (q_{obs,t} - \overline{q_{obs}})^2} \quad (4.10)$$

na qual $q_{sim,t}$ e $q_{obs,t}$ denotam, respectivamente, as vazões simulada e observada no intervalo de tempo t , e $\overline{q_{obs}}$ corresponde à vazão média observada;

- Raiz do erro quadrático médio (RMSE), dado pela seguinte equação:

$$RMSE = \left[\frac{1}{n} \cdot \sum_{t=1}^n (q_{sim,t} - q_{obs,t})^2 \right]^{\frac{1}{2}} \quad (4.11)$$

cujas notações são similares à da equação 4.10;

- Relação entre volumes simulados e observados (RV), dado pela seguinte equação:

$$RV = \frac{\sum_{t=1}^n q_{sim,t}}{\sum_{t=1}^n q_{obs,t}} \quad (4.12)$$

cuja notação é similar à das duas equações anteriores; e

- Coeficiente de correlação de Pearson (ρ), dado pela seguinte equação:

$$\rho = \frac{\sum_{t=1}^n (q_{sim,t} - \bar{q}_{sim})(q_{obs,t} - \bar{q}_{obs})}{\sqrt{\sum_{i=1}^n (q_{sim,t} - \bar{q}_{sim})^2} \sqrt{\sum_{i=1}^n (q_{obs,t} - \bar{q}_{obs})^2}} \quad (4.13)$$

na qual $q_{sim,t}$ e $q_{obs,t}$ denotam, respectivamente, as vazões simulada e observada no intervalo de tempo t , e \bar{q}_{sim} e \bar{q}_{obs} correspondem, respectivamente, às vazões médias simulada e observada.

O cálculo dos índices de qualidade de calibração será realizado considerando-se a moda das distribuições marginais *a posteriori* dos parâmetros do modelo Rio Grande.

Em adição aos índices de calibração, será avaliada também a adequação da função de verossimilhança adotada, por meio das seguintes análises gráficas:

- Hidrogramas simulados *versus* observados, para os períodos de calibração e validação;
- Vazões simuladas *versus* vazões observadas, ordenadas de maneira crescente;
- Resíduos *versus* vazões simuladas, cujo intuito é verificar a heterocedasticidade dos erros;
- Autocorrelação dos resíduos; e
- Ajuste dos resíduos normalizados ao modelo distributivo SEP. Para essa análise, serão empregadas as modas das distribuições *a posteriori* como estimativas dos parâmetros da função densidade de probabilidade proposta.

4.4 Construção das curvas de quantis de enchentes

A etapa final do método é relacionada à construção das curvas de quantis de cheias e avaliação das incertezas associadas aos mesmos para um conjunto de tempos de retorno especificados, a saber, 2, 5, 10, 25, 50, 100, 500, 1.000, 5.000 e 10.000 anos. Essa etapa

constitui, em última análise, o objetivo principal da presente tese. Nesse ponto serão combinadas séries de precipitação diária obtidas com o modelo misto de geração estocástica com as realizações da distribuição conjunta *a posteriori* dos parâmetros obtidas na calibração do modelo Rio Grande.

Para cada uma das bacias hidrográficas em estudo, serão geradas 1.000 séries de precipitação diária de tamanho 10.000 anos, com auxílio do modelo misto de geração estocástica. Séries com o referido tamanho, de maneira geral, contemplam as chuvas de projeto empregadas no dimensionamento de grandes vertedores e estruturas de contenção de enchentes e, assim, são suficientes para os objetivos desta pesquisa. Contudo, em função de análises mais específicas, pode-se utilizar séries de precipitação de maior tamanho.

A calibração via algoritmo DREAM fornece a descrição probabilística *a posteriori* dos parâmetros do modelo chuva-vazão. Para cada uma das séries de precipitação geradas, serão sorteados 1.000 conjuntos de parâmetros da distribuição conjunta *a posteriori*, os quais serão empregados na geração de séries sintéticas de vazão.

A seguir, serão construídas curvas de quantis para as 1.000 séries sintéticas de vazão associadas a cada série sintética de precipitação, e extraídos os valores médios para cada quantil de interesse. Essa curva média será associada à série de precipitação em consideração e assumida como uma realização particular das cheias máximas anuais na bacia. Esse processo é repetido para todas as séries sintéticas de precipitação.

De posse das 1.000 curvas médias de quantis de cheia, será então estimada uma curva média global e serão calculados os intervalos de credibilidade de 95% para os quantis de interesse, de maneira a se caracterizar as incertezas associadas a cada estimativa. A curva média será também utilizada na comparação com outros métodos de estimação de cheias, tais como análise de frequência local e regional, análise de frequência com uso de paleocheias, método pico-volume e outros estudos de estimação indireta de enchentes, com intuito de se avaliar o desempenho global do método proposto.

5 APLICAÇÃO DO MÉTODO À BACIA DO RIO PARÁ

5.1 Caracterização da bacia do rio Pará, em Ponte do Vilela

A primeira aplicação do método será realizada na bacia do rio Pará, no estado de Minas Gerais. A escolha dessa bacia é fundamentada na grande disponibilidade de dados pluviométricos e em um expressivo conjunto de estudos no contexto de geração estocástica de precipitação diária e com o modelo Rio Grande.

A bacia do rio Pará encontra-se inserida no polígono composto pelas coordenadas geográficas aproximadas de 19°10' e 20°45' de latitude sul e 44°15' e 45°20' de longitude oeste, e possui uma área de drenagem de aproximadamente 12.225 km². Os principais afluentes do rio Pará são os rios São João e do Peixe, pela margem direita, e os rios Itapecerica, Lambari e do Picão, pela margem esquerda (LIMA, 2004).

As nascentes do rio Pará localizam-se no município de Rezende Costa, nas vertentes das serras da Galga e da Cebola, em uma altitude de 1.160m. O curso d'água apresenta um desenvolvimento longitudinal de 303km, no sentido predominante SE-NW, das nascentes até a foz no rio São Francisco, em uma altitude de 590m, a localiza-se no limite entre os municípios de Pompeu, Abaeté e Martinho Campos.

A bacia de estudo, mostrada na Figura 5.1, está localizada à montante da represa de Carmo do Cajuru e abrange uma área de 1620 km². O curso d'água principal possui comprimento de 76 km, até a estação fluviométrica de Ponte do Vilela (código 40130002), a qual é operada pela Companhia Energética de Minas Gerais (CEMIG).

Uma característica marcante da bacia selecionada é a forte sazonalidade verificada ao longo do ano hidrológico. De fato, 86% da precipitação média anual se concentram na estação chuvosa, compreendida entre os meses de outubro e março. Os 14% restantes se distribuem de maneira bastante irregular ao longo dos meses da estação seca.

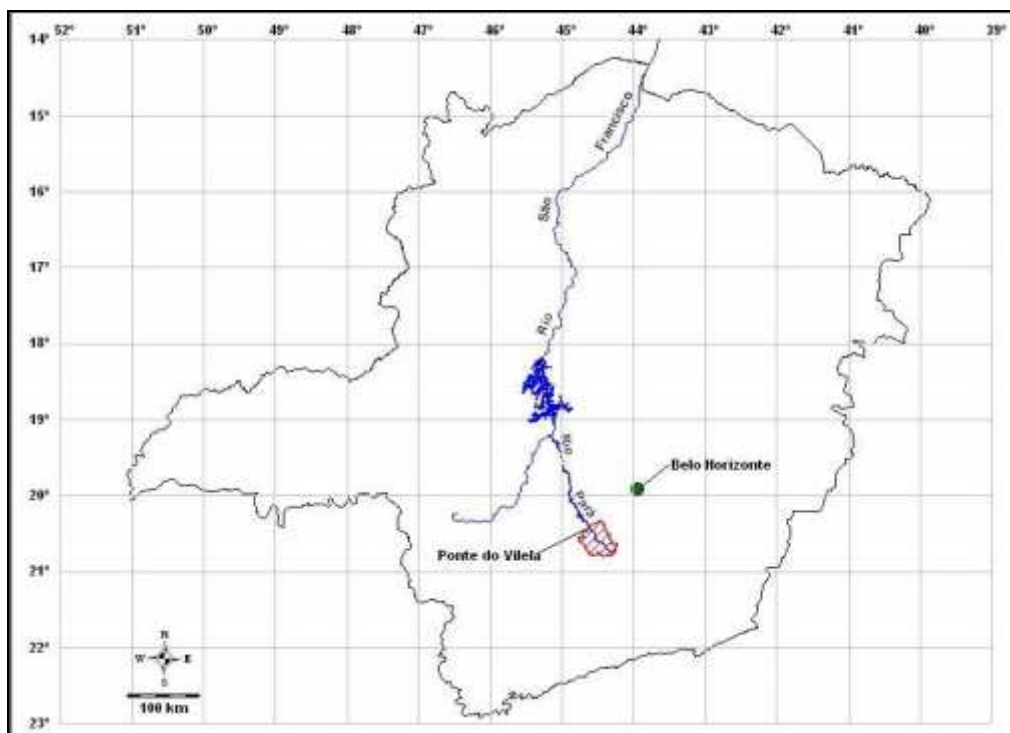


Figura 5.1 – Localização da bacia do rio Pará, na estação de Ponte do Vilela (40130002)

O perfil longitudinal do rio Pará é apresentado na Figura 5.2. Nessa figura constam também os principais tributários do curso d'água, as estações fluviométricas em operação e os reservatórios artificiais existentes.

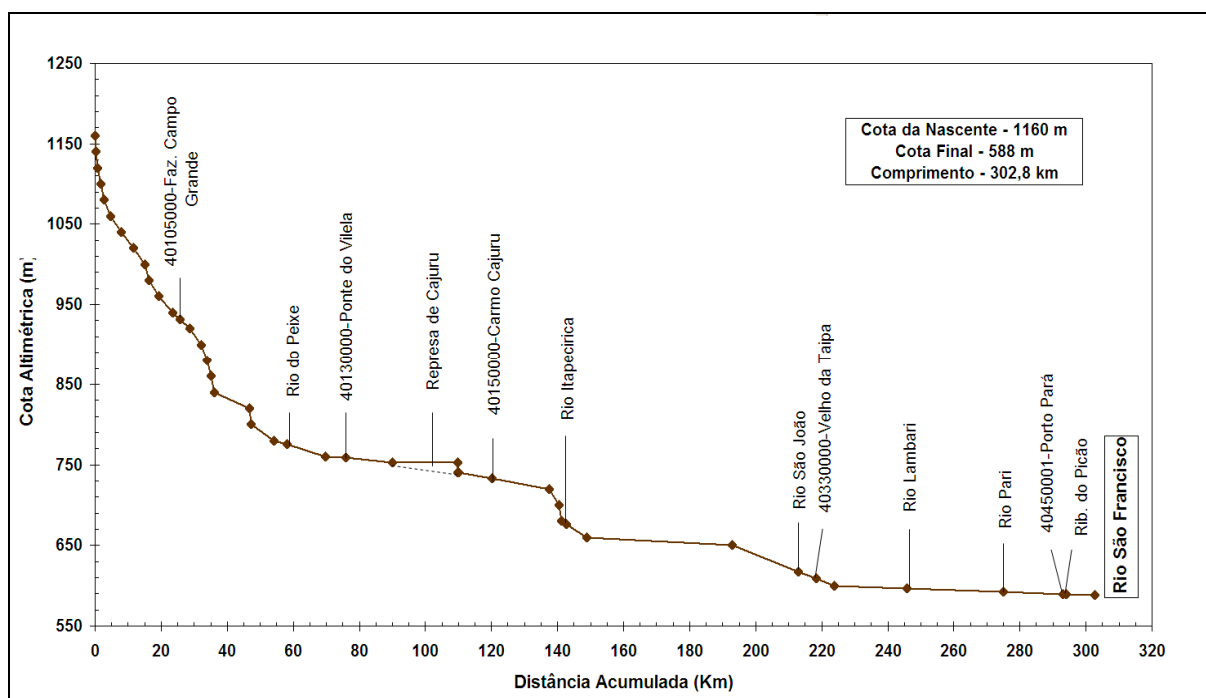


Figura 5.2 – Perfil longitudinal do rio Pará
Fonte: Pinheiro (2009)

5.2 Estações pluviométricas, fluviométricas e evaporimétricas utilizadas no estudo

O trabalho de Lima (2004) constituiu a fonte elementar de dados pluviométricos, fluviométricos e evaporimétricos para a aplicação do método proposto na bacia do rio Pará, em Ponte do Vilela. A autora realizou um extensivo procedimento de consistência e preenchimento de falhas de dados, tanto pluviométricos quanto fluviométricos. Em vista do exposto, não foram realizadas aqui avaliações mais detalhadas acerca da qualidade dos dados. Com o intuito de se realizar uma comparação com os resultados de geração estocástica de precipitação diária da referida autora, adotou-se o mesmo período de observação de dados pluviométricos, compreendido entre janeiro de 1942 e dezembro de 2002.

A determinação da chuva média espacializada na bacia do rio Pará, em Ponte do Vilela, foi realizada a partir do método dos polígonos de Thiessen. As estações utilizadas no estudo são apresentadas na Tabela 5.1, com as respectivas áreas de influência.

Tabela 5.1 – Estações pluviométricas utilizadas para determinação da chuva média espacializada na bacia do rio Pará, em Ponte do Vilela. Fonte: Lima (2004)

Estação	Código	Área de influência
Carmo da Mata	2044005/2044042	9,5
Fazenda Campo Grande	2044009	64,0
Itaguara	2044036	26,5
Total		100,0

A Tabela 5.2 apresenta as estações evaporimétricas utilizadas na presente aplicação, bem como o período de observação de dados das mesmas. As estações empregadas não se localizam no interior da área de estudo, mas em regiões próximas.

Tabela 5.2 – Estações evaporimétricas utilizadas no estudo. Fonte: Lima (2004)

Estação	Código	Período de dados
Três Marias	1845009	10/1985-05/00
Porto Pará	1945017	01/1973-11/1999

Por fim, os dados fluviométricos foram obtidos da estação de Ponte do Vilela jusante (40130002), pertencente à rede hidrométrica da CEMIG.

5.3 Distribuição a priori para o limite superior

A estimativa da PMP meteorológica de 1 dia na bacia do rio Pará, em Ponte do Vilela, é de 340 mm (LIMA, 2004). De acordo com o procedimento especificado no item 4.2.1, a distribuição gama será empregada para modelar as incertezas com relação ao limite superior. Dessa forma, torna-se necessário estimar o coeficiente de variação e a probabilidade de não-excedência da PMP local.

O coeficiente de variação estimado para o conjunto de 118 estimativas de PMP estatística é igual a 0,154. Já a frequência de não-excedência empírica, estimada com a posição de plotagem de Weibull, é igual a 0,051. Assim, a distribuição *a priori* eliciada para o limite superior é $\alpha \sim GAMA(42,166;0,094)$, denotada por modelo I.

Adicionalmente a esse modelo, foi considerada uma distribuição não informativa para o limite superior. Tal distribuição é incluída com o objetivo de se efetuar a análise sob a perspectiva de que o decisor não tem a informação *a priori* ou não compartilha da opinião que a PMP tem uma associação direta com o limite superior. O modelo não informativo adotado no presente trabalho é uma distribuição gama com grande variância, $\alpha \sim GAMA(1,0;1,0 \times 10^{-8})$, denotada por modelo II. Cabe ressaltar aqui que a parametrização do software WinBUGS (Lunn *et al.*, 2000) é distinta daquela usualmente empregada em pacotes estatísticos, uma vez que, em lugar da variância, o referido software faz uso da precisão na especificação de parâmetros de modelos.

A Tabela 5.3 apresenta as principais características das distribuições utilizadas como especificação *a priori* para o limite superior.

Tabela 5.3 – Parâmetros e características das distribuições *a priori* do limite superior

Distribuição <i>a priori</i>	ρ_α	β_α	Média	Mediana	CV_α	DP
I	42,166	0,094	448,6	445,0	0,154	69,1
II	1,000	$1,000 \times 10^{-8}$	$1,000 \times 10^8$	$6,900 \times 10^7$	1,000	$1,000 \times 10^8$

5.4 Estatísticas a posteriori

5.4.1 Modelo LN4

Para o modelo, a distribuição *a posteriori* dos parâmetros é proporcional ao produto da função de verossimilhança pela distribuição *a priori*, ou seja:

$$p(\Theta|x)\alpha p(x|\Theta)p(\alpha)p(\mu)p(\sigma) \quad (5.1)$$

onde $p(x|\Theta)$ é a função de verossimilhança, $p(\alpha)$ é a distribuição *a priori* para o limite superior dada pelos modelos I e II do item 5.3, e $p(\mu)$ e $p(\sigma)$ são, respectivamente, as distribuições *a priori* dos parâmetros μ e σ .

A interpretação dos parâmetros μ e σ , respectivamente a média e o desvio padrão da variável transformada LN4, em termos hidrometeorológicos é uma tarefa complexa. Com efeito, é extremamente difícil encontrar relações evidentes entre esses parâmetros e características das bacias de interesse. Dessa forma, serão especificadas distribuições *a priori* não informativas para os mesmos. Uma vez que μ pode assumir qualquer valor real e 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:

$$\mu \sim NORMAL(1,0;1,0 \times 10^{-6}) \text{ e } \sigma \sim GAMA(1,0;1,0 \times 10^{-8}).$$

A solução da equação 5.1 envolve o cálculo de integrais multidimensionais complexas, as quais, via de regra, são impossíveis de serem obtidas por meios analíticos. O modelo LN4 se insere nessa categoria. A alternativa à integração analítica são os algoritmos MCMC, capazes de amostrar da distribuição *a posteriori* após a convergência da cadeia.

As simulações numéricas foram realizadas com auxílio do software WinBUGS (LUNN *et al.*, 2000). Para os propósitos da presente aplicação, foi considerado que uma amostra final de tamanho 50.000, para cada um dos parâmetros, é suficiente para caracterizar a variabilidade de suas distribuições marginais *a posteriori*. Com intuito de se estabelecer o “*burn-in*” do modelo, foi empregada a estatística de Brooks-Gelman-Rubin (Brooks & Gelman, 1998), a qual se encontra disponível no WinBUGS como ferramenta de diagnóstico de convergência das cadeias de Markov construídas pelo algoritmo MCMC. Seguindo a recomendação de Cowles & Carlin (1996), 10 cadeias de Markov, com valores iniciais bastante distintos, foram empregadas na simulação. Para os 3 parâmetros do modelo distributivo LN4, a estatística de Brooks-Gelman-Rubin atingiu a convergência para a unidade após cerca de 40.000 iterações, como mostra a Figura 5.3.

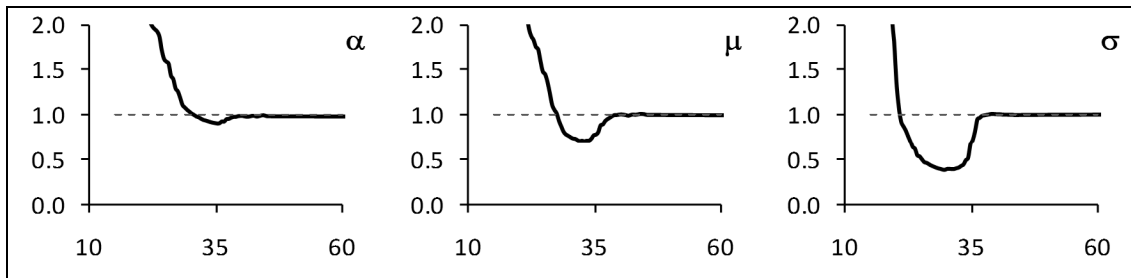


Figura 5.3 – Variação da estatística de Brooks-Gelman-Rubin ao longo da simulação. O eixo x representa o número de iterações em milhares e o eixo y os valores da referida estatística

Após a definição do “burn-in”, é necessário se estabelecer o *lag* para o qual se obtém uma amostra final não correlacionada. O procedimento foi realizado por tentativa e erro, no qual simulações foram realizadas para um conjunto de possíveis valores candidatos e a autocorrelação foi avaliada em cada situação. Foi verificado que um *lag* igual a 20 é adequado ao referido propósito. A Figura 5.4 mostra a variação dos valores dos parâmetros do modelo LN4 ao longo da simulação, após o descarte dos valores de “burn-in” e aplicação do *lag*. É possível observar que não ocorrem heterogeneidades, tais como tendências ou alterações na variância, em nenhuma região do domínio de cada parâmetro. Tal comportamento é desejável para que as amostras obtidas via algoritmo MCMC sejam adequadas para fins de inferência.

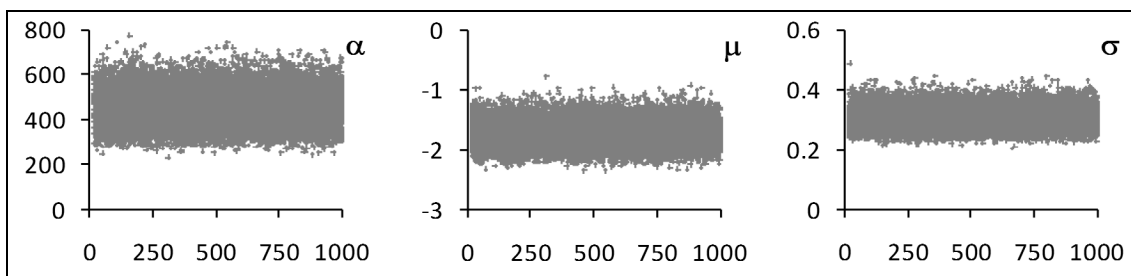


Figura 5.4 – Variação dos valores dos parâmetros do modelo LN4 ao longo da simulação. O eixo x representa o número de iterações em milhares

Estatísticas a posteriori para o parâmetro μ

A Tabela 5.4 sumaria as estatísticas *a posteriori* para o parâmetro μ . A Figura 5.5 apresenta as distribuições *a posteriori* para os dois modelos avaliados.

Tabela 5.4 – Estatísticas *a posteriori* para o parâmetro μ

Modelo	Média	DP	CV	95% HPD
I	-1,724	0,059	-0,034	(-2,085;-1,354)
II	-13,125	2,038	-0,155	(-15,810;-8,971)

95% HPD – Intervalo de credibilidade

Uma vez que não foi possível estabelecer uma distribuição *a priori* informativa para o parâmetro μ , analisar os resultados *a posteriori* constitui uma questão de grande complexidade. No entanto, é possível observar que o referido parâmetro é sensível à especificação da distribuição *a priori* do limite superior. Quando a distribuição *a priori* não informativa é utilizada, a distribuição *a posteriori* concentra-se mais à esquerda na reta real e ocorre um aumento tanto da variância quanto do intervalo de credibilidade. A interpretação física de tal comportamento é envolta em dificuldades, dada a ausência de relações físicas entre o parâmetro μ e as características hidrometeorológicas da bacia. Sob uma perspectiva estritamente matemática, contudo, a diferença de comportamento das distribuições *a posteriori* dos dois modelos avaliados se explica pela ampla faixa de variação dos valores de limite superior quando se prescreve uma distribuição *a priori* não informativa, o que provoca o deslocamento do parâmetro de posição a fim de se obter a convergência para a distribuição de equilíbrio. Pode-se verificar ainda que no modelo I a distribuição *a posteriori* é unimodal e aproximadamente simétrica. No caso do modelo II, a distribuição *a posteriori* também é unimodal. Contudo, a mesma apresenta ligeira assimetria positiva.

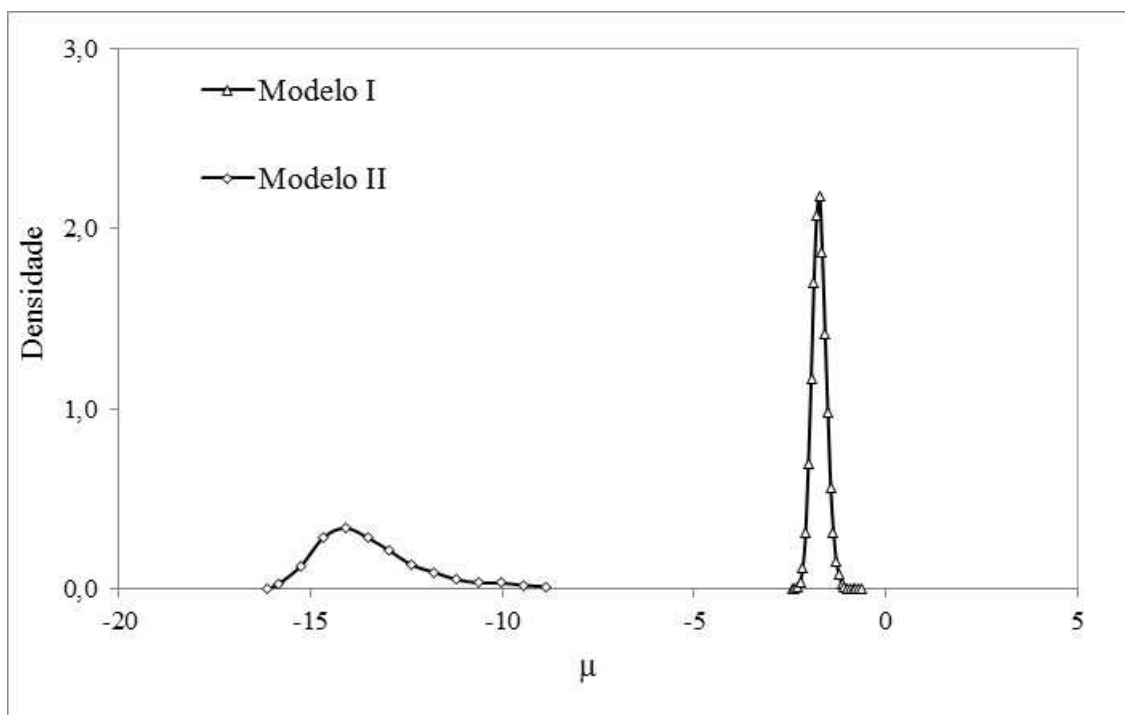


Figura 5.5 – Distribuições *a posteriori* para o parâmetro μ

Estatísticas a posteriori para o parâmetro σ

A Tabela 5.5 sumaria as estatísticas *a posteriori* para o parâmetro σ . A Figura 5.6 apresenta as distribuições *a posteriori* para os dois modelos.

Tabela 5.5 – Estatísticas *a posteriori* do parâmetro σ

Modelo	Média	DP	CV	95% HPD
I	0,304	0,031	0,102	(0,247;0,364)
II	0,254	0,071	0,280	(0,209;0,303)

95% HPD – Intervalo de credibilidade

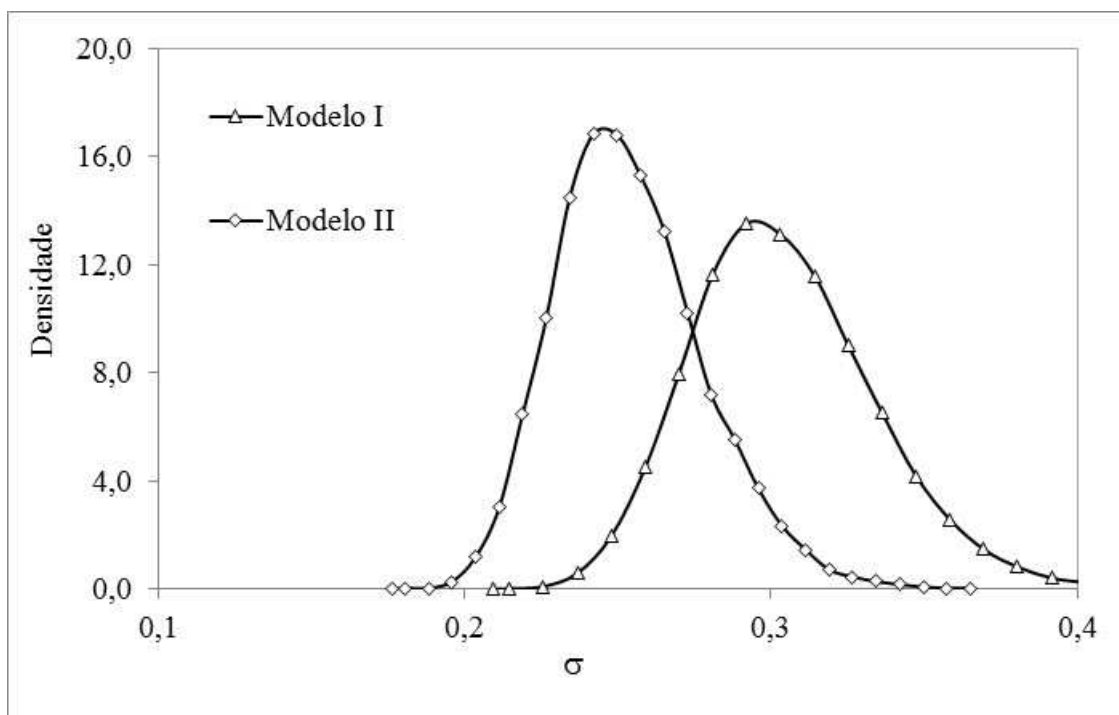


Figura 5.6 - Distribuições *a posteriori* para o parâmetro σ

De maneira similar ao parâmetro μ , o fato de não ter sido possível estabelecer uma distribuição *a priori* informativa para σ torna-se difícil a análise dos resultados *a posteriori*. Contudo, para este parâmetro as distribuições *a posteriori* apresentaram um comportamento semelhante para os dois modelos avaliados, à exceção do coeficiente de variação, bastante superior no modelo II. A relativa proximidade das médias e dos intervalos de credibilidade dos dois modelos pode ser uma evidência de sua menor sensibilidade à especificação da distribuição *a priori* para o limite superior. No caso do parâmetro σ , ambas as distribuições *a posteriori* são unimodais e aproximadamente simétricas.

Estatísticas a posteriori do parâmetro α

A Tabela 5.6 e a Figura 5.7 mostram os resultados *a posteriori* para o parâmetro α .

Tabela 5.6 – Estatísticas *a posteriori* para o parâmetro α

Modelo	Média	DP	CV	95% HPD
I	450,6	21,2	0,047	(323,3;589,8)
II	$8,3 \times 10^7$	$8,9 \times 10^7$	1	$(4,2 \times 10^4; 2,9 \times 10^8)$

95% HPD – Intervalo de credibilidade

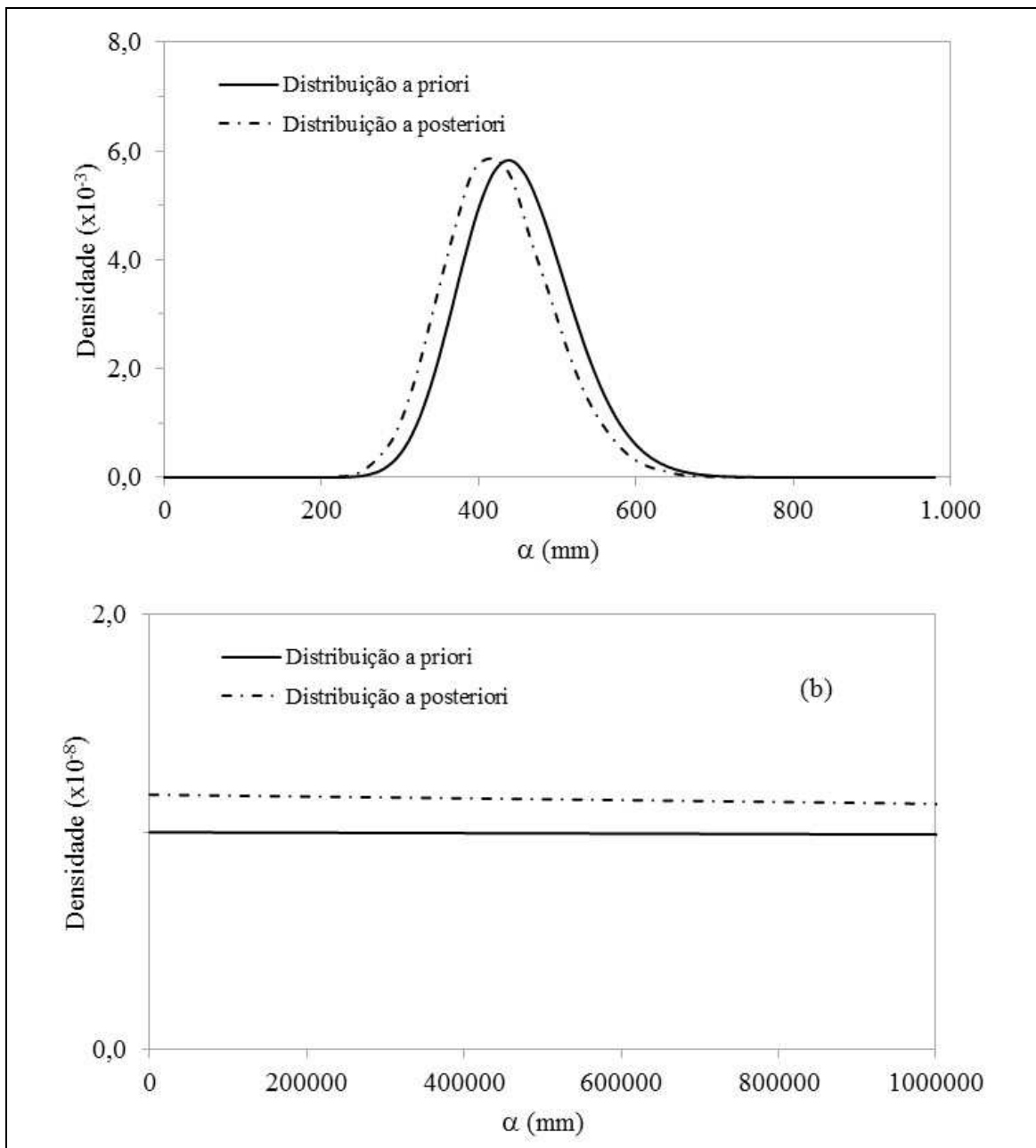


Figura 5.7 – Distribuição *a priori* e distribuição *a posteriori* do parâmetro α . (a) Modelo I; (b) Modelo II.

A Figura 5.7 mostra que as distribuições *a priori* e *a posteriori* do modelo I são relativamente próximas, com um pequeno deslocamento dessa última para a esquerda. Esse fato decorre da influência dos dados na inferência da distribuição *a posteriori* de α . Com efeito, uma vez que alturas de chuva registradas são bastante inferiores às estimativas de PMP, e a distribuição *a posteriori* do limite superior é truncada no máximo observado, a mesma tende a se mover em sua direção aos dados quando da aplicação do teorema de Bayes.

Ainda com relação ao modelo I, é possível observar que a estimativa de α é consideravelmente superior à estimativa local de PMP (aproximadamente 32%), o que ilustra a inadequação do uso dessa última como valor fixo do limite superior em modelos distributivos. Contudo, apesar de superior à estimativa local de PMP, o valor de α obtido da distribuição *a posteriori* se mostra fisicamente realista, em face do conhecimento corrente dos processos hidrometeorológicos de formação de tormentas em condições extremas na bacia do rio Pará. Além disso, é possível verificar que esse modelo atribui probabilidades extremamente reduzidas a valores de limite superior maiores que, por exemplo, 1.000mm, fato este que restringe a obtenção de estimativas implausíveis para o referido parâmetro.

A distribuição *a posteriori* obtida com o modelo II, por outro lado, apresenta um aumento ínfimo de massa na região de valores próxima aos dados, e decai de maneira especialmente lenta em direção aos maiores valores do domínio do limite superior. Esse fato é o reflexo de que, na ausência de informação prévia acerca de α , os dados isoladamente são inaptos a caracterizar a variabilidade do referido parâmetro. Como consequência, a distribuição *a posteriori* atribui probabilidades similares a qualquer valor de limite superior ao longo do domínio do parâmetro, o que inevitavelmente leva a estimativas pouco realistas, sob uma perspectiva física, para o mesmo (tais como aquela apresentada na Tabela 5.6, $8,3 \times 10^7$ mm). Dessa forma, nas etapas subsequentes do método, somente o modelo I será considerado.

5.4.2 Modelo EV4

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

$$p(\Theta|x)\alpha p(x|\Theta)p(\alpha)p(\mu)p(\sigma) \quad (5.2)$$

onde $p(x|\Theta)$ é a função de verossimilhança, $p(\alpha)$ é a distribuição *a priori* para o limite superior dada pelos modelos I e II do item 5.3, e $p(\sigma)$ e $p(\xi)$ são, respectivamente, as distribuições *a priori* dos parâmetros σ e ξ .

Tal como no modelo LN4, a interpretação dos parâmetros σ e ξ em termos hidrometeorológicos é difícil, o que inviabiliza a obtenção de relações com características da bacia. Dessa forma, serão especificadas distribuições *a priori* não informativas para os mesmos. Uma vez que ambos os parâmetros são sempre positivos, assumiram-se distribuições gama não informativas para os mesmos. Dessa forma, $\sigma \sim GAMA(1,0;1,0 \times 10^{-8})$ e $\xi \sim GAMA(1,0;1,0 \times 10^{-8})$.

Também para esse modelo distributivo foram consideradas duas distribuições *a priori* para o limite superior: a primeira, informativa, eliciada por $\alpha \sim GAMA(42,166;0,094)$ e denotada por modelo I, e a segunda, não informativa, eliciada por $\alpha \sim GAMA(1,0;1,0 \times 10^{-8})$ e denotada por modelo II.

A amostra da distribuição *a posteriori* do modelo EV4 foi obtida de maneira similar àquela prescrita para o modelo LN4, na qual a estatística de Brooks-Gelman-Rubin foi empregada na determinação do “*burn-in*” e o *lag* foi definido por tentativa e erro. Entretanto, para o modelo EV4, a convergência para a distribuição de equilíbrio se mostrou mais lenta, e o “*burn-in*” adotado foi igual a 45.000. Já com relação ao *lag*, o valor 20, definido para o modelo LN4, também foi considerado adequado ao modelo EV4. As amostras finais dos parâmetros não apresentaram heterogeneidades, tais como tendências ou alterações na variância.

Estatísticas a posteriori para o parâmetro σ

A Tabela 5.7 e a Figura 5.8 mostram os resultados *a posteriori* para o parâmetro σ .

Tabela 5.7 – Estatísticas *a posteriori* para o parâmetro σ

Modelo	Média	DP	CV	95% HPD
I	6,597	1,450	0,220	(4,322;9,014)
II	$3,3 \times 10^6$	$2,356 \times 10^6$	0,714	$(2,78 \times 10^6; 2,78 \times 10^6)$

95% HPD – Intervalo de credibilidade

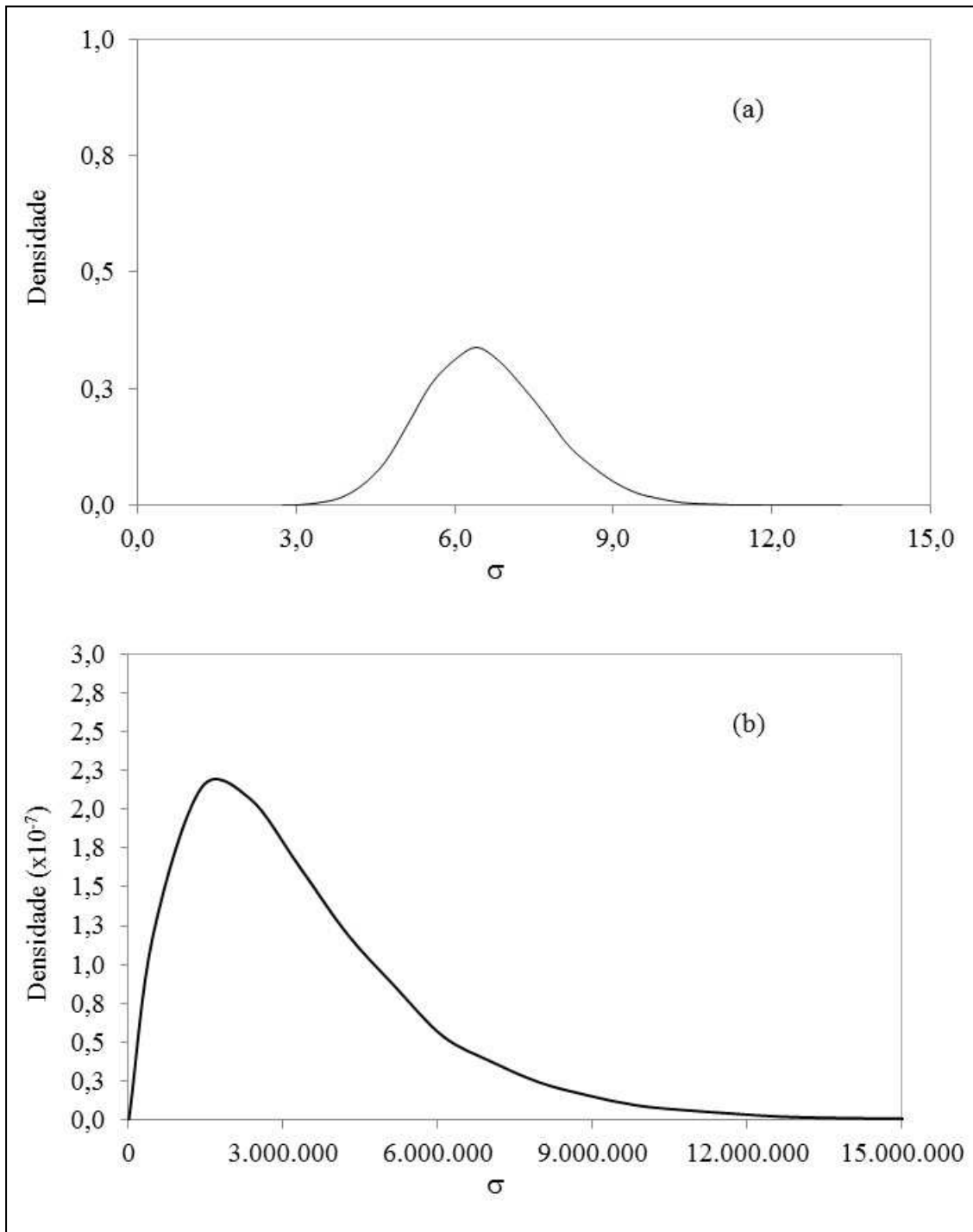


Figura 5.8 – Distribuições *a posteriori* para o parâmetro σ . (a) Modelo I; (b) Modelo II.

A análise da Figura 5.8 mostra que o parâmetro σ é extremamente sensível à especificação da distribuição *a priori* do limite superior. Com efeito, para o modelo I a distribuição *a posteriori* se concentra em uma faixa restrita do domínio do parâmetro, ao passo que, para o modelo II, a mesma se estende por várias ordens de grandeza na reta real. Além disso, no caso do modelo I, a distribuição *a posteriori* é aproximadamente simétrica, enquanto que, para o modelo II, a mesma é fortemente

assimétrica à direita. Uma vez que não foi definida uma relação clara desse parâmetro com características hidrometeorológicas da bacia, não é possível interpretar de tal comportamento em termos físicos. Contudo, de um ponto de vista estritamente matemático, pode-se explicar a ampla faixa de variação encontrada na inferência do modelo II como uma forma de compensar a também ampla faixa de variação da distribuição *a posteriori* do limite superior, uma vez que, na definição da função acumulada de probabilidade, tais parâmetros apresentam uma relação direta de proporcionalidade.

Estatísticas a posteriori para o parâmetro ξ

A Tabela 5.8 e a Figura 5.9 resumiam os resultados *a posteriori* para o parâmetro ξ .

Tabela 5.8 – Estatísticas *a posteriori* para o parâmetro ξ

Modelo	Média	DP	CV	95%HPD
I	3,801	0,383	0,101	(3,052;4,550)
II	4,429	0,427	0,096	(3,615;5,281)

95% HPD – Intervalo de credibilidade

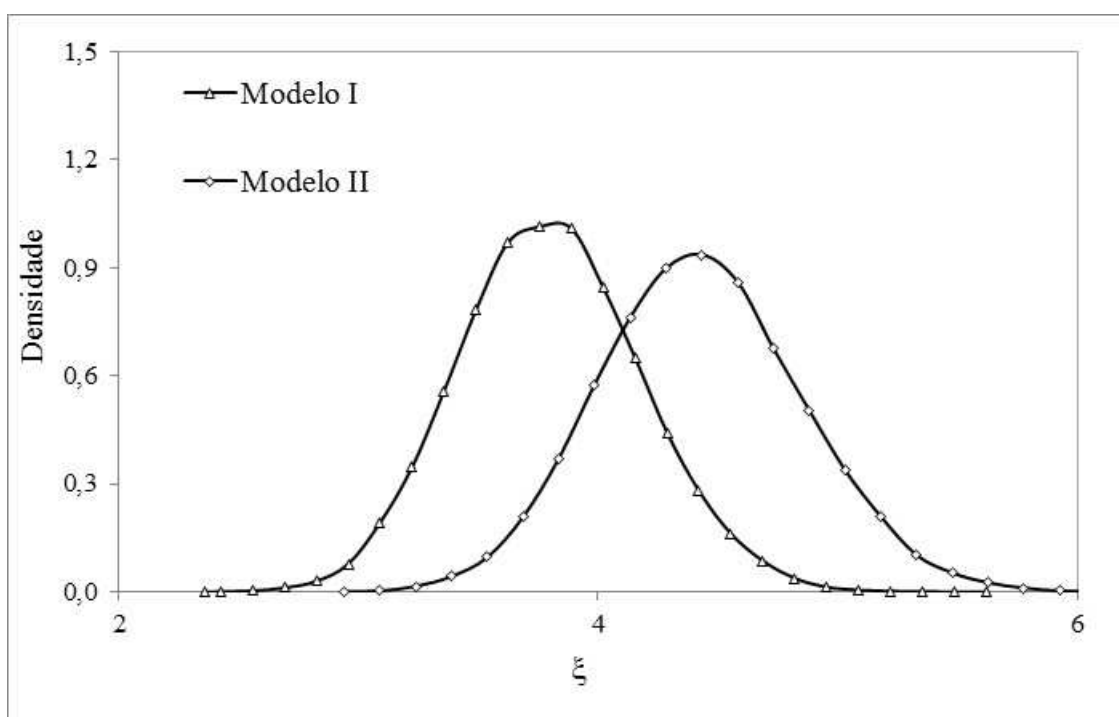


Figura 5.9 – Distribuições *a posteriori* para o parâmetro ξ

Com relação ao parâmetro de forma ξ , percebe-se que o mesmo não é significativamente afetado pela eliciação da distribuição *a priori* do limite superior. As médias e variâncias inferidas para os dois modelos são próximas, e as amplitudes dos

intervalos de credibilidade também são similares. Além disso, ambas as distribuições *a posteriori* são unimodais e aproximadamente simétricas. Contudo, também nesse caso é difícil prover uma interpretação dessas distribuições em termos físicos.

Estatísticas a posteriori do parâmetro α

A Tabela 5.9 e a Figura 5.10 mostram os resultados *a posteriori* para o parâmetro α .

Tabela 5.9 – Estatísticas *a posteriori* para o parâmetro α

Modelo	Média	DP	CV	95% HPD
I	450,8	69,7	0,155	(319,5;592,0)
II	$2,0 \times 10^8$	$1,4 \times 10^8$	0,708	$(4,0 \times 10^6; 4,7 \times 10^8)$

95% HPD – Intervalo de credibilidade

De maneira análoga ao modelo LN4, as distribuições *a priori* e *a posteriori* do modelo I são relativamente próximas, sendo verificado um pequeno deslocamento dessa última para a esquerda em função da influência dos registros observados quando da aplicação do teorema de Bayes. Mais interessante ainda é a similaridade das estimativas de α para os dois modelos probabilísticos empregados nesta tese. Esse fato pode constituir um indício de que a informação disponibilizada pela distribuição *a priori*, que fornece fundamentação física ao fenômeno modelado, é efetiva para se inferir as características de variabilidade do limite superior das precipitações diárias.

Com relação ao modelo II, também para a distribuição EV4 se verifica a inadequação de se prescrever uma distribuição *a priori* não informativa para o limite superior: as estimativas para o referido parâmetro se mostram completamente implausíveis face ao conhecimento corrente dos processos físicos de geração de tormentas severas. Parece claro, assim, que a utilização de distribuições limitadas superiormente não agrega vantagens à modelagem de precipitações diárias máximas quando não se dispõe de conhecimento prévio acerca do limite superior. Diante disso, apenas o modelo I será empregado nas etapas subsequentes do método na bacia do rio Pará.

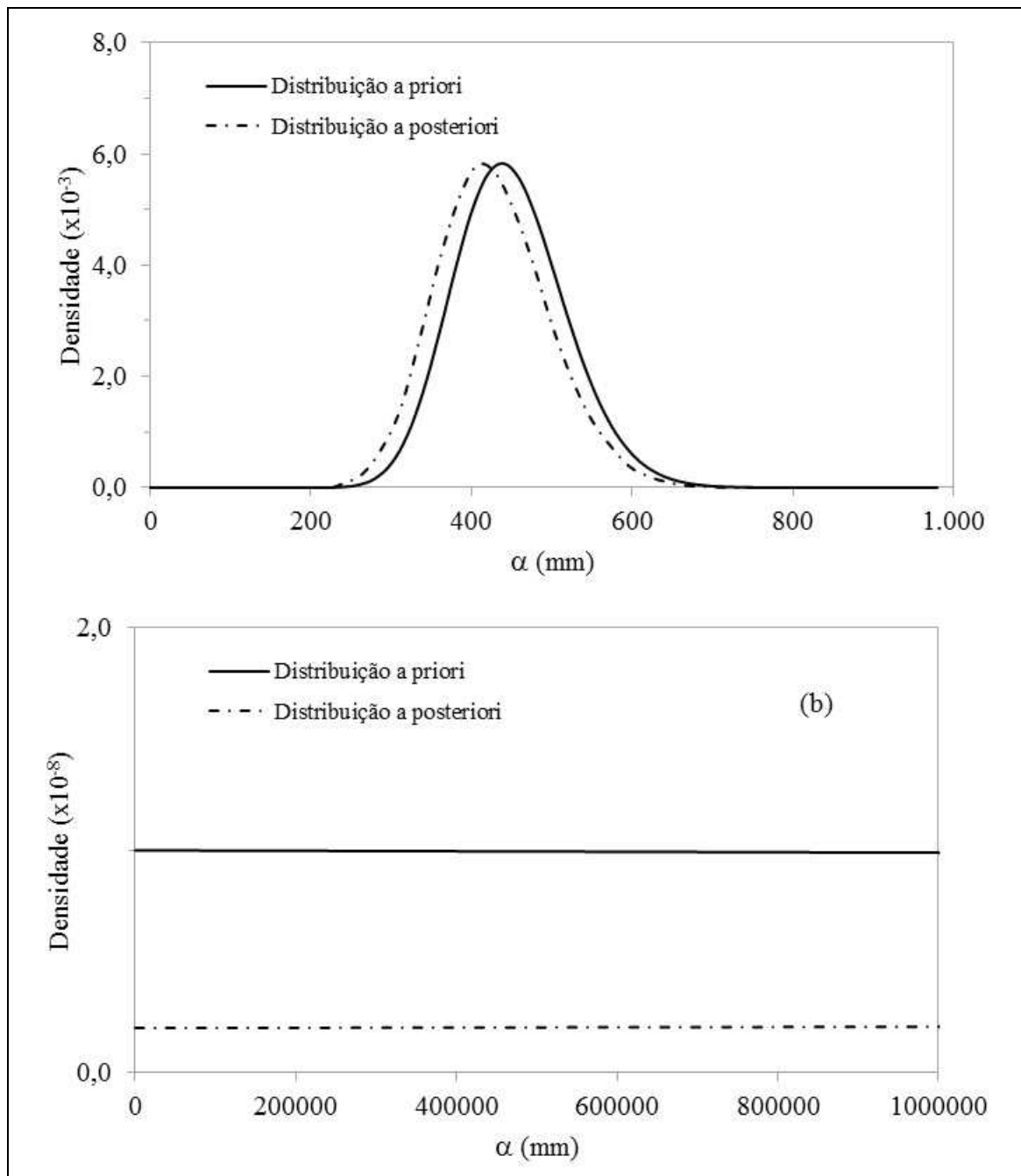


Figura 5.10 – Distribuição *a priori* e distribuição *a posteriori* do parâmetro α . (a) Modelo I; (b) Modelo II.

5.5 Geração estocástica de precipitação diária

Estimados os parâmetros dos modelos distributivos LN4 e EV4, a etapa seguinte para a aplicação do método é a geração de séries sintéticas de precipitação, as quais servirão de *input* ao modelo Rio Grande de transformação chuva-vazão. A inicialização do modelo misto para a geração estocástica de precipitação requer a especificação do limiar entre chuvas convencionais e extremas, para o módulo paramétrico, e do tamanho das janelas de reamostragem do módulo não paramétrico. Essas duas etapas são discutidas nos

próximos tópicos. Em seguida, são realizadas a análise de desempenho do modelo misto e a comparação entre modelos de geração estocástica de chuva estabelecidos na literatura, a saber, um modelo bipartido com uma distribuição gama ajustada aos dados, um modelo de matriz de probabilidade de transição que emprega a distribuição generalizada de Pareto, caracterizada pela cauda superior pesada em seu último estado, e um modelo não paramétrico, baseado na técnica de “*bootstrap*” convencional.

5.5.1 Definição do limiar entre chuvas convencionais e chuvas extremas

Conforme discutido no capítulo 4, a definição de limiares constitui um grande desafio para a correta modelagem de processos caracterizados por duas distribuições de probabilidade distintas. No entanto, tendo-se em vista a estrutura mista proposta para o gerador estocástico, a principal dificuldade associada ao estabelecimento do limiar é contornada de maneira direta, uma vez que não há necessidade de se compatibilizar estimativas de parâmetros dos dois modelos para assegurar a continuidade da distribuição híbrida obtida.

Uma vez que não é necessário se estabelecer um limiar de maneira analítica, é intuitivo se recorrer à calibração do gerador para se fixar o limiar entre chuvas convencionais e extremas. Nesse contexto, a premissa básica consiste na simulação de séries sintéticas para um conjunto de valores de limiar e na escolha daquele valor que proporciona melhor ajuste ao conjunto de dados, aferido por meio de funções-objetivo adequadas. Na presente pesquisa, foram escolhidas para esse propósito a curva de quantis de precipitações diárias máximas anuais e as precipitações médias mensais.

Para a aplicação do método na bacia do rio Pará, em Ponte do Vilela, foram testados limiares variando entre 40 e 100mm. Para cada limiar proposto, foram geradas 1.000 séries de mesmo tamanho da amostra, 61 anos, e calculados os valores médios das estatísticas de avaliação. Os resultados são apresentados separadamente para os modelos distributivos LN4 e EV4.

Modelo LN4

A Figura 5.11 apresenta as curvas de quantis de precipitações diárias máximas anuais para limiares variando entre 40 e 100mm. É possível observar que para valores entre 40 e 60 mm não ocorre ajuste ao conjunto de registros observados. Esse fato decorre do

uso demasiado da distribuição LN4 ao longo das simulações, o que ocasiona a geração de eventos extremos com frequência superior àquela verificada na amostra. Para valores entre 60 e 100 mm, por outro lado, a série de máximos gerada apresenta ajuste aos dados, o que torna qualquer valor nessa faixa de variação um candidato razoável a limiar.

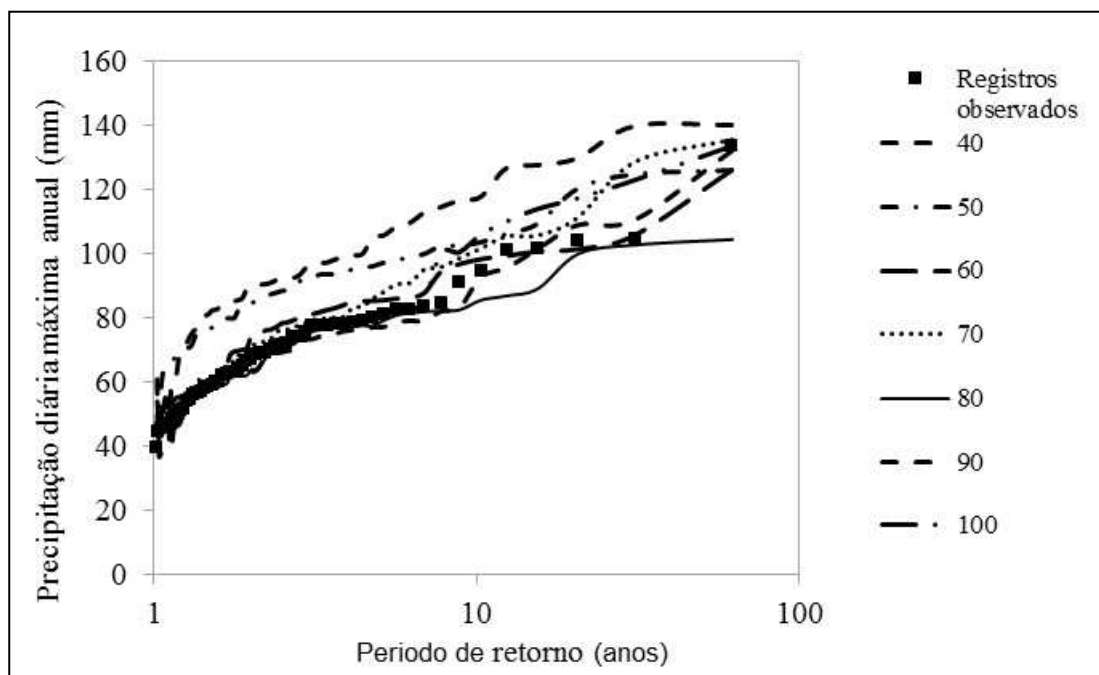


Figura 5.11 – Precipitação diária máxima anual gerada para diferentes limiares entre chuvas convencionais e extremas para o modelo LN4

A Figura 5.12 sumaria as precipitações médias mensais para os diferentes limiares aptos a representar os máximos anuais. Nessa figura, as barras em azul correspondem à média das séries simuladas e as barras em vermelho à série observada. É possível verificar que, dentre os valores testados, aquele que melhor reproduz a precipitação média mensal é 80mm. Em especial para os meses de junho, julho e agosto, a referida estatística é bastante comprometida para limiares distintos desse valor. Dessa forma, o limiar para inicialização do gerador fica estabelecido em 80 mm. Uma vantagem adicional dessa escolha é que o limiar corresponde a um valor intermediário, o qual permite que se aproveite de maneira mais efetiva a estrutura de análise Bayesiana construída para modelar os eventos extremos sem comprometer a eficiência das simulações na estação seca.

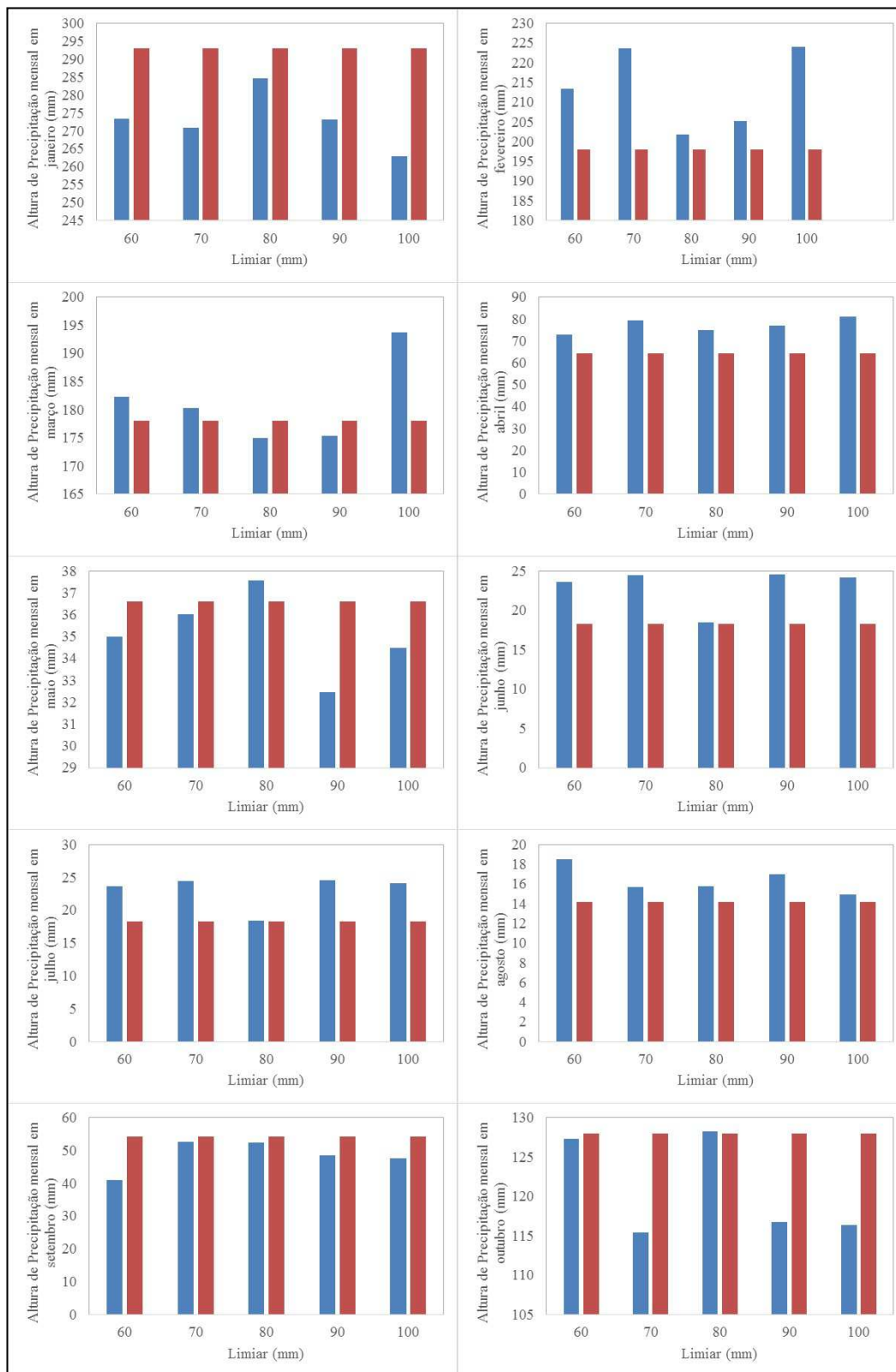


Figura 5.12 – Comparação entre a precipitação média mensal das séries simuladas com o modelo LN4 e da série observada para diferentes limiares entre chuvas convencionais e extremas

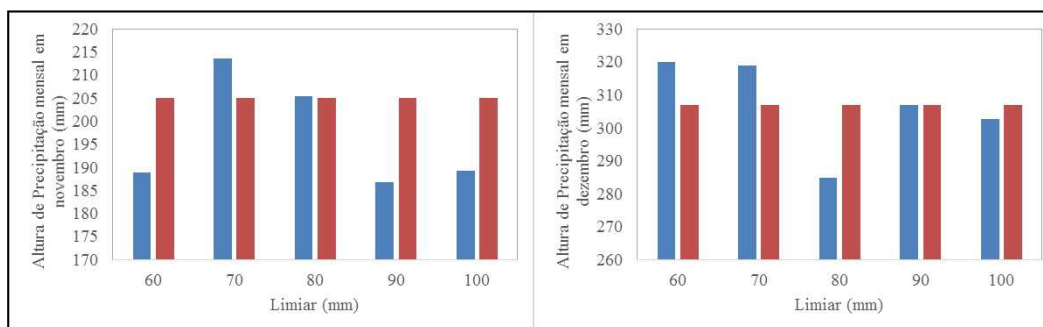


Figura 5.13 (cont.) – Comparação entre a precipitação média mensal das séries simuladas com o modelo LN4 e da série observada para diferentes limiares entre chuvas convencionais e extremas

Modelo EV4

Análises similares à do modelo LN4 foram efetuadas para a distribuição EV4. Em primeiro lugar, as curvas de quantis para diferentes limiares foram simuladas com intuito de se escolher aqueles aptos a representar os máximos anuais. Para o modelo EV4, no entanto, somente se obteve ajuste para limiares superiores a 90mm. Nesse caso, a probabilidade de geração de chuvas extremas na seca se torna nula, uma vez que não se verificam valores de alturas precipitadas dessa magnitude na amostra. Essa é uma situação indesejada de um ponto de vista de modelagem, por impedir qualquer tipo de extrapolação das alturas precipitadas nos meses mais secos do ano, especialmente quando se tem em vista que, de um ponto de vista físico, tal limitação não é coerente. Essa situação é ilustrada na Figura 5.13.

O comportamento verificado na modelagem dos máximos anuais com o modelo EV4 decorre da acentuada inclinação verificada em sua função de quantis a partir de um tempo de retorno de aproximadamente 10 anos. Esse fato faz com que os valores de chuvas extremas geradas pelo modelo EV4 sejam excepcionalmente mais altos que os do modelo LN4, por exemplo, o que força a adoção de um limiar mais elevado como alternativa para restringir o uso do módulo paramétrico do gerador, possibilitando melhor ajuste ao conjunto de dados.

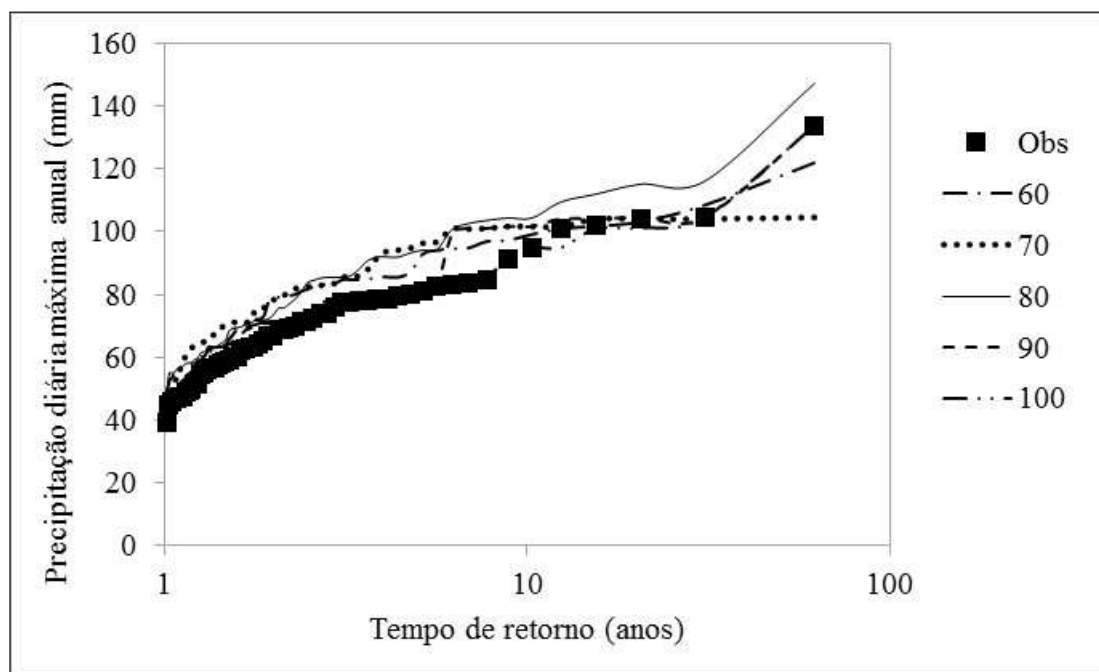


Figura 5.14 – Precipitação diária máxima anual gerada para diferentes limiares entre chuvas convencionais e extremas para o modelo EV4

Com intuito de se permitir a geração de chuvas extremas na estação seca, optou-se por fixar o limiar entre chuvas convencionais e extremas em 80 mm também para o modelo EV4. Tal expediente, por um lado, contorna o problema de ocorrência de precipitações extremas entre os meses de abril e setembro. Contudo, a fixação arbitrária do limiar introduz um problema adicional à modelagem dos máximos anuais e das precipitações médias mensais, uma vez que eventos extremos são gerados com frequência superior àquela observada na amostra. Dessa forma, as análises relacionadas ao modelo EV4 na geração estocástica de precipitação diária ficam severamente comprometidas. Ainda assim, o desempenho do gerador será avaliado nessa situação tendo-se em vista aferir a capacidade descritiva do modelo EV4.

A Figura 5.14 sumaria as precipitações médias mensais para os diferentes limiares aptos a representar os máximos anuais. Nessa figura, as barras em azul correspondem à média das séries simuladas e as barras em vermelho à série observada. É possível verificar que, o limiar assumido, 80 mm, ainda que não modele adequadamente os máximos anuais, fornece uma descrição razoável para as precipitações médias mensais. Contudo, em um âmbito global, os melhores ajustes foram obtidos para os limiares de 90 e 100mm, indicando que esses seriam mais adequados para se proceder às simulações.

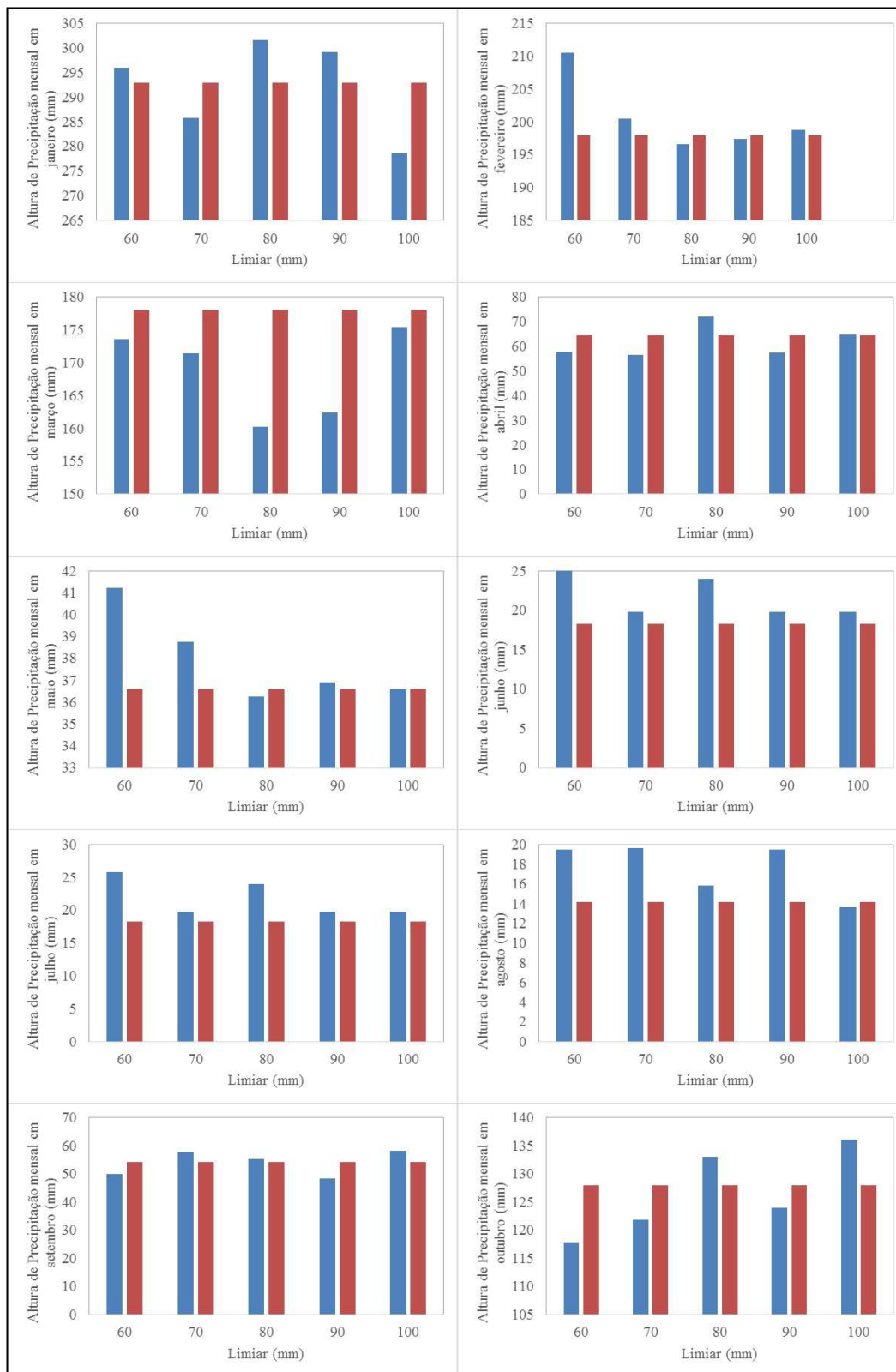


Figura 5.15 – Comparação entre a precipitação média mensal das séries simuladas com o modelo EV4 e da série observada para diferentes limiares entre chuvas convencionais e extremas

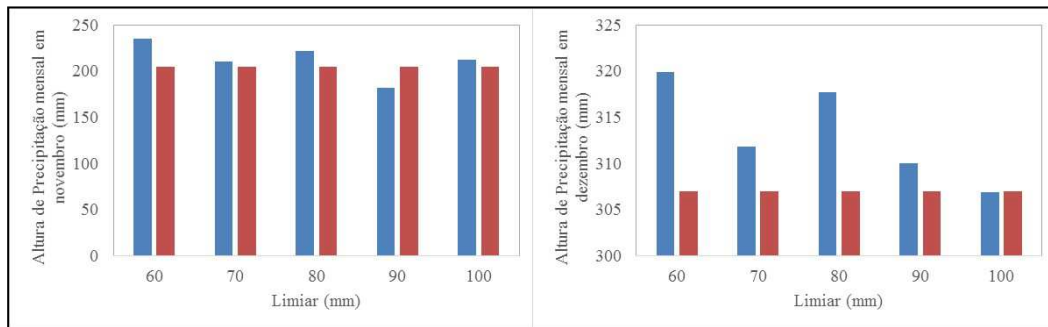


Figura 5.16 (cont.) – Comparação entre a precipitação média mensal das séries simuladas com o modelo EV4 e da série observada para diferentes limiares entre chuvas convencionais e extremas

5.5.2 Definição do tamanho das janelas de reamostragem

A etapa seguinte para a inicialização do modelo misto é a definição do tamanho das janelas de reamostragem do módulo não paramétrico de geração estocástica de precipitação diária. Foram avaliadas janelas de tamanho 8, 14 e 28 dias, com intuito de se avaliar a capacidade de reprodução da variabilidade da precipitação diária em cada uma dessas situações.

A Figura 5.15 mostra os resultados das simulações da média e do desvio padrão da precipitação diária em função do tamanho da janela de reamostragem para o modelo LN4. Pode-se observar que para qualquer tamanho de janela a média é adequadamente reproduzida. O desvio padrão, no entanto, não apresenta ajuste quando se considera a janela de 8 dias. Esse fato evidencia que uma janela desse tamanho é incapaz de capturar a variabilidade das precipitações diárias verificada na amostra. Por outro lado, as janelas de 14 e 28 dias apresentam comportamento similar e foram aptas a reproduzir ambas as estatísticas. Uma vez que a janela de 14 dias demanda esforço computacional significativamente menor na geração das séries de chuva, a mesma foi selecionada para a continuidade da aplicação.

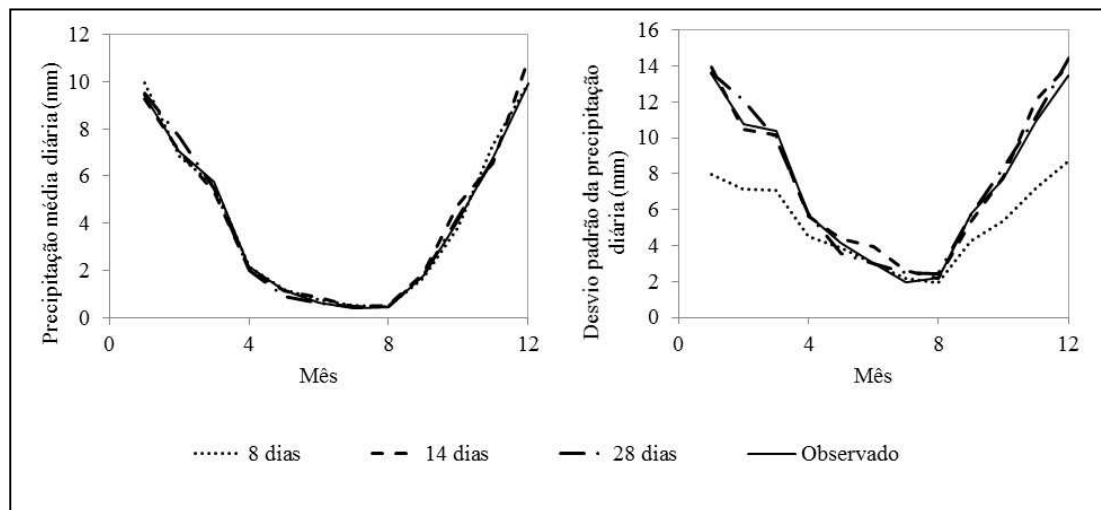


Figura 5.17 – Comparação entre estatísticas das séries geradas e observada para diferentes tamanhos de janela de reamostragem para o modelo LN4

Com relação à Figura 5.15, os cálculos foram efetuados para o limiar de 80mm, ainda que esse não tenha fornecido ajuste adequado ao conjunto de dados. Contudo, um comportamento semelhante ao do modelo LN4 foi verificado: a janela de 14 dias foi capaz de caracterizar a variabilidade das precipitações diárias observadas. Diante disso, a referida janela foi selecionada também para a distribuição EV4.

5.5.3 Avaliação de desempenho do modelo misto de geração estocástica de precipitação diária

Após a definição dos limiares e do tamanho das janelas de reamostragem para os dois modelos distributivos, procedeu-se à geração de um conjunto de 1.000 séries de precipitação diária de 61 anos. Consideraram-se como opções nessa etapa as diversas possibilidades de simulação do gerador proposto: seleção entre todos os valores não nulos dentro da janela ou somente entre pares com combinação idêntica de ocorrência de chuva e introdução de perturbações às alturas de precipitação observadas.

De maneira geral, não se verificou diferenças expressivas entre essas possibilidades, o que pode ser atribuído ao tamanho relativamente grande da amostra de precipitações diárias. Ainda assim, fez-se a opção de introduzir as perturbações nos registros observados com intuito de se aumentar a variabilidade dos valores oriundos da reamostragem.

Avaliação de desempenho do modelo LN4

A Figura 5.16 sintetiza os resultados das simulações com o modelo LN4. É possível observar que o modelo proposto reproduz de maneira adequada a média e o desvio padrão da chuva diária, para todos os meses do ano. Em outras palavras, o gerador misto é apto a simular as características de variabilidade dos dados diários de precipitação, o que implicitamente evidencia a adequação das escolhas do limiar e do tamanho das janelas. Em uma base diária, a maioria dos modelos disponíveis na literatura reproduz satisfatoriamente as referidas estatísticas. Em escalas mensais e anuais, entretanto, modelos paramétricos comumente apresentam tendência de subestimação da variância (BOUGHTON, 1999; SRIKANTHAN & McMAHON, 2000). O gerador proposto, por outro lado, foi capaz de reproduzir de forma apropriada as variâncias mensais, como mostra a Figura 5.17, o que sugere que anos excepcionalmente secos e úmidos podem ser simulados pelo modelo. Essa característica constitui uma importante vantagem do gerador misto em comparação a modelos bipartidos e de matriz de probabilidade de transição.

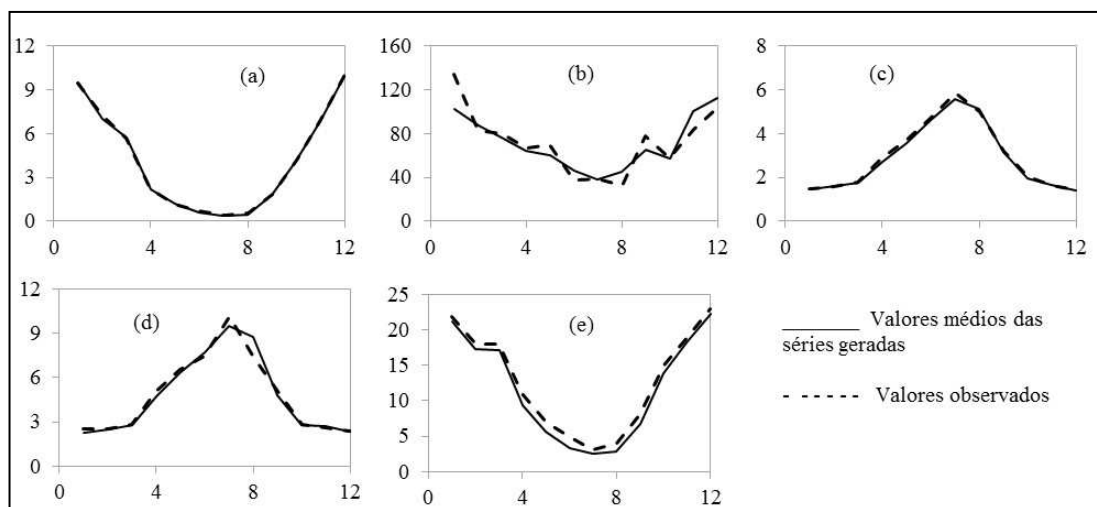


Figura 5.18 – Comparação das estatísticas das séries geradas com o modelo LN4 com a série observada para cada mês do ano. (a) chuva média diária; (b) altura máxima de chuva diária; (c) coeficiente de variação; (d) coeficiente de assimetria; (e) número médio de dias chuvosos. O eixo X representa os meses do ano.

Também é reportada na literatura a dificuldade dos modelos paramétricos em reproduzir coeficientes de assimetria superiores a 6 (SRIKANTHAN, 2004). Os resultados das simulações mostram que o gerador misto não apresenta tal limitação. Pelo contrário, coeficientes de assimetria da ordem de 10 foram razoavelmente reproduzidos pelo modelo. Esse fato também aponta as vantagens da abordagem proposta, uma vez que

um conjunto mais amplo de distribuições empíricas, com as mais variadas formas, pode ser adequadamente representado.

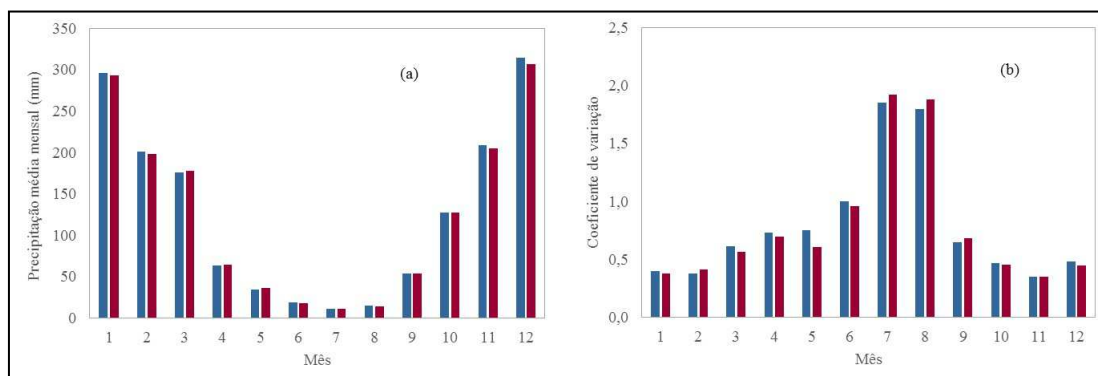


Figura 5.19 – Comparação entre as médias (a) e variâncias (b) mensais das séries simuladas com o modelo LN4 e a série observada

Com relação ao número médio de chuvosos, pode-se observar um ligeiro viés de subestimação, mais evidente nos meses da estação seca (abril-setembro). Esse fato pode estar relacionado à construção das matrizes de probabilidade de transição em uma base diária, em lugar da escala mensal usualmente empregada. Outros estudos na mesma bacia, tais como o de Lima (2004), não apontaram essa deficiência, a qual constitui uma desvantagem do gerador misto.

Por fim, as alturas máximas de precipitação diária, para cada mês do ano, são coerentes com aquelas observadas. Esse comportamento evidencia a adequação da estrutura proposta para o modelo, uma vez que os valores médios das precipitações máximas em meses secos se encontram sempre na faixa não paramétrica de simulação, o que evidencia as pequenas probabilidades de ocorrência de chuvas extremas nos mesmos. Por outro lado, na estação chuvosa, as alturas máximas podem ser relacionadas a ambos os módulos do gerador, o que mostra a consistência do modelo na reprodução dos mecanismos físicos de formação de chuva.

Particularmente para o mês de janeiro, uma forte subestimação da altura máxima de precipitação foi verificada. Entretanto, a avaliação dos outros valores no referido mês mostra que o segundo maior registro, 104 mm, é coerente com os valores simulados. Assim, o valor máximo observado, 133 mm, é provavelmente um *'outlier'*, com tempo de retorno mais elevado que aquele associado à sua posição na amostra.

A figura 5.18 apresenta a curva de quantis obtida a partir das séries geradas, empregando-se a posição de plotagem de Weibull. É possível observar que as

precipitações máximas anuais são adequadamente reproduzidas para os mais variados tempos de retorno, indicando a qualidade de ajuste do modelo LN4 ao conjunto de dados. Essa característica é de suma importância para a geração de séries mais longas, tais como aquelas de 10.000 anos a serem empregadas na estimação indireta de quantis de cheias.

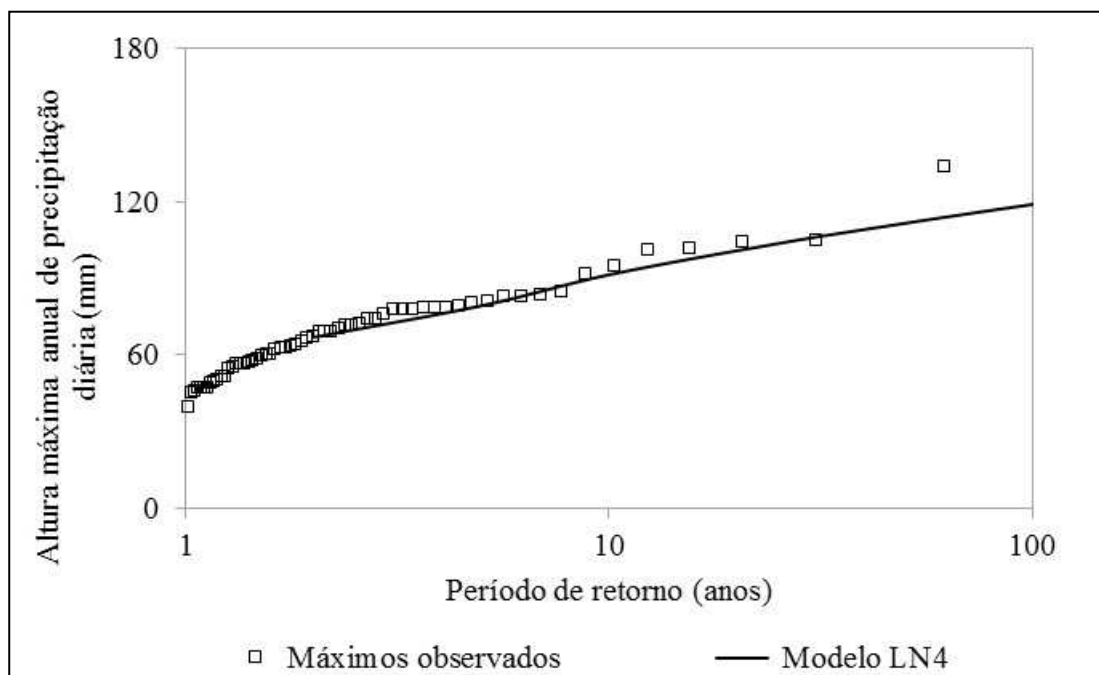


Figura 5.20 – Curva de quantis de precipitação diária obtida com o modelo LN4

Avaliação de desempenho do modelo EV4

A Figura 5.19 sintetiza os resultados das simulações com o uso do modelo EV4 para descrever os eventos extremos de precipitação. É possível observar que as estatísticas diárias, à exceção das alturas máximas de precipitação, são adequadamente reproduzidas com o referido modelo. Na simulação dos máximos anuais, no entanto, observou-se uma tendência de superestimação em boa parte da estação chuvosa, incluindo-se um valor gerado consideravelmente próximo daquele observado no mês de janeiro, o qual foi caracterizado anteriormente como um “outlier”. Esse fato pode ser atribuído ao comportamento da função de quantis descrito na seção 5.5.1. Com efeito, uma vez que o limiar de 80 mm não forneceu ajuste do modelo EV4 ao conjunto de máximos observados, eventos extremos, cujas magnitudes são consistentemente superiores às aquelas observadas na amostra para o mesmo tempo de retorno, são gerados em demasia, prejudicando a caracterização da curva de quantis.

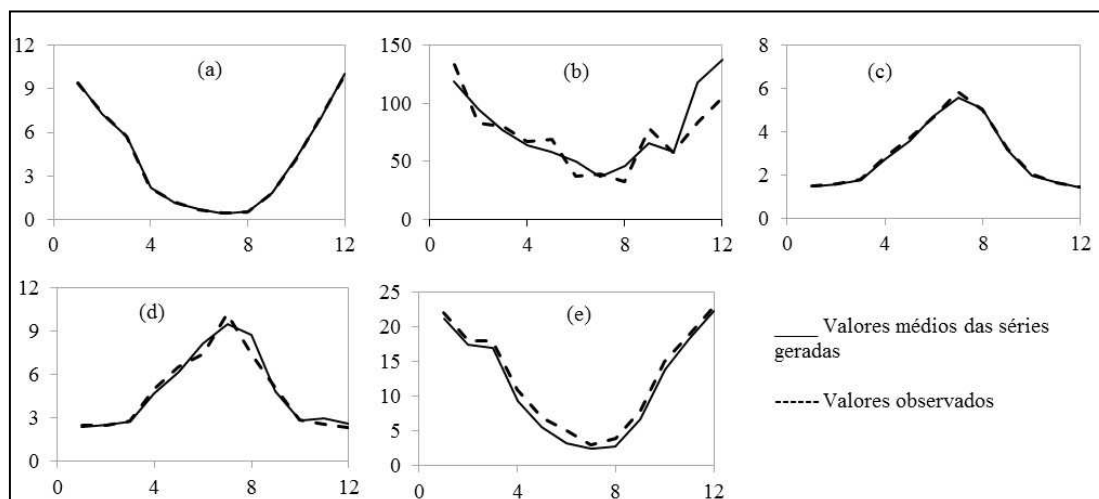


Figura 5.21 – Comparação das estatísticas das séries geradas a partir do modelo EV4 com a série observada para cada mês do ano. (a) chuva média diária; (b) altura máxima de chuva diária; (c) coeficiente de variação; (d) coeficiente de assimetria; (e) número médio de dias chuvosos. O eixo X representa os meses do ano.

Em termos mensais, verificou-se uma pequena tendência de superestimação das médias e variâncias nos meses da estação chuvosa, como mostra a Figura 5.20. Esse fato está relacionado à maior frequência de ocorrência de eventos extremos, com magnitudes por vezes extremamente elevadas na referida estação. Nos meses mais secos, as estatísticas foram bem reproduzidas, uma vez que, nesse caso, as probabilidades de geração de chuvas extremas são reduzidas. As estatísticas anuais ficaram comprometidas pela inconsistência dos resultados dos meses mais úmidos.

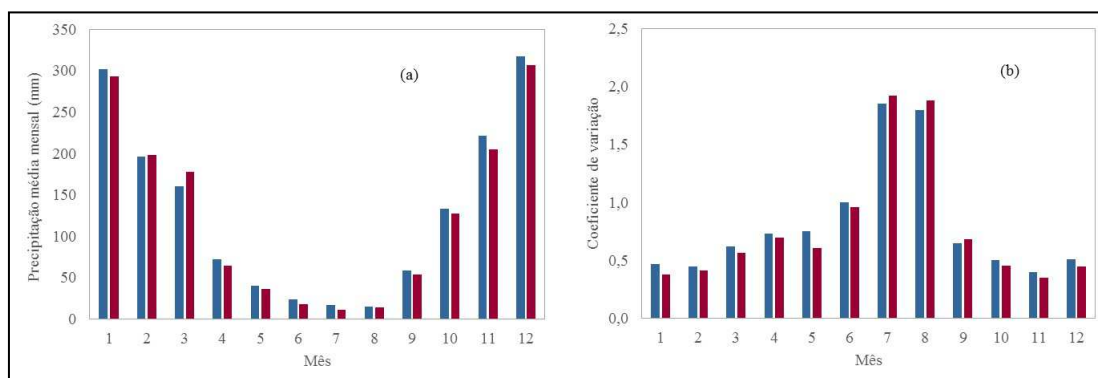


Figura 5.22 – Comparação entre as médias (a) e variâncias (b) mensais das séries simuladas com o modelo EV4 e a série observada

A Figura 5.21 apresenta a curva de quantis obtida a partir das séries geradas. Percebe-se nessa figura a ausência de ajuste do modelo EV4 para tempos de retorno superiores a 10 anos, o que evidencia que a referida distribuição é inadequada à presente aplicação. Diante disso, pode-se supor que as estimativas de cheias obtidas com as séries oriundas desse modelo distributivo também mostram pouca qualidade de ajuste aos registros

observados. Assim, somente o modelo LN4 será empregado para a estimação dos quantis de enchentes na bacia do rio Pará, em Ponte do Vilela.

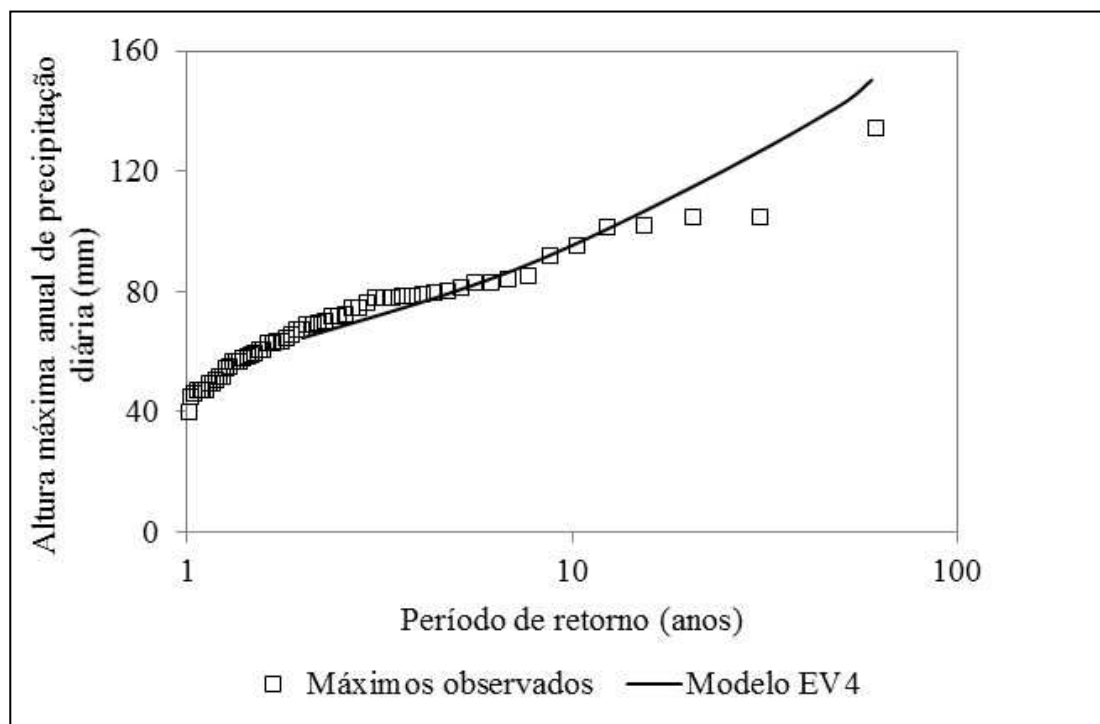


Figura 5.23 – Curva de quantis de precipitação diária obtida com o modelo EV4

5.5.4 Comparação com outros modelos de geração estocástica de precipitação diária

Conforme exposto no Capítulo 4, a principal motivação para a construção de um novo modelo de geração estocástica de precipitação diária é a possibilidade de modelar de maneira acurada tanto as chuvas convencionais quanto as extremas. Nesse contexto, uma comparação da curva de quantis obtida com aquelas oriundas de métodos difundidos na literatura é de grande utilidade para se aferir as vantagens e desvantagens do modelo misto proposto nesta pesquisa.

São considerados para fins de comparação com o gerador proposto, restrito ao uso da distribuição LN4 na descrição dos eventos extremos, os seguintes modelos: (1) um modelo bipartido, com uma distribuição gama de 2 parâmetros ajustada aos dados; (2) o modelo de matriz de probabilidade de transição (MPT) proposto por Lima (2004), o qual emprega a distribuição generalizada de Pareto na descrição dos eventos extremos; e (3) um modelo de reamostragem via “*bootstrap*” convencional.

Evidentemente, uma comparação direta entre modelos somente é possível mediante a avaliação por critérios objetivos, tais como testes de razão de verossimilhança. Entretanto, dada a diferença de fundamentação física dos modelos em avaliação, tal comparação pode ser inadequada, com a penalização daqueles modelos construídos a partir de estruturas mais complexas. Diante disso, é realizada aqui somente uma análise gráfica do comportamento de cada um dos modelos citados no parágrafo anterior, com ênfase em quantis associados a períodos de retorno mais elevados.

A Figura 5.22 apresenta as curvas de quantis para os diferentes modelos avaliados, considerando-se um tempo de retorno máximo de 10.000 anos. A análise dessas curvas requer grande cuidado, uma vez que a amostra não dispõe de valores de precipitação extremamente elevados e, assim, em uma primeira análise, qualquer modelo pode ser considerado como razoável para a extrapolação dos quantis de precipitação diária. Apesar disso, alguns comentários podem ser feitos a respeito dos resultados obtidos com os diferentes geradores. Em primeiro lugar, é possível perceber que o modelo LN4 situa-se em uma posição intermediária entre os modelos bipartido e de matriz de probabilidade de transição. Tendo-se em vista que, de maneira geral, distribuições com caudas superiores exponenciais, tais como a gama, subestimam os eventos extremos (Wilks & Wilby, 1999; Li *et al.*, 2012), e distribuições com caudas superiores pesadas, tais como a generalizada de Pareto, superestimam tais eventos (Li *et al.*, 2013; Chin & Brissette, 2014), o comportamento do referido modelo parece bastante coerente com aquele descrito por Papalexiou & Koutsouyiannis (2012) e Papalexiou *et al.* (2013) para as chuvas diárias mais extremas. Em segundo lugar, observa-se que o modelo MPT superestima os quantis de precipitação para tempos de retorno da ordem de 10 anos, indicando ausência de ajuste aos dados para valores de precipitação com razoável probabilidade de excedência, e não apenas no trecho de extrapolação da curva. Além disso, verifica-se que a estimativa de quantil de 10.000 anos do modelo MPT, 330mm, é bastante próxima da estimativa local de PMP. Dado que, em teoria, a estimativa de PMP reflete as condições hidrometeorológicas mais severas de formação de uma tormenta na bacia, é pouco razoável supor que a mesma seja associada a um tempo de retorno de 10.000 anos, tal como sugere o modelo MPT. Por fim, o modelo de reamostragem se ajusta satisfatoriamente a quantis de menor tempo de retorno, mas é incapaz de gerar valores mais elevados que o máximo observado, e, assim, atribui probabilidades de excedência distintas ao mesmo valor de altura de precipitação na simulação de séries

longas. Esse fato impossibilita a descrição correta da cauda superior das precipitações diárias.

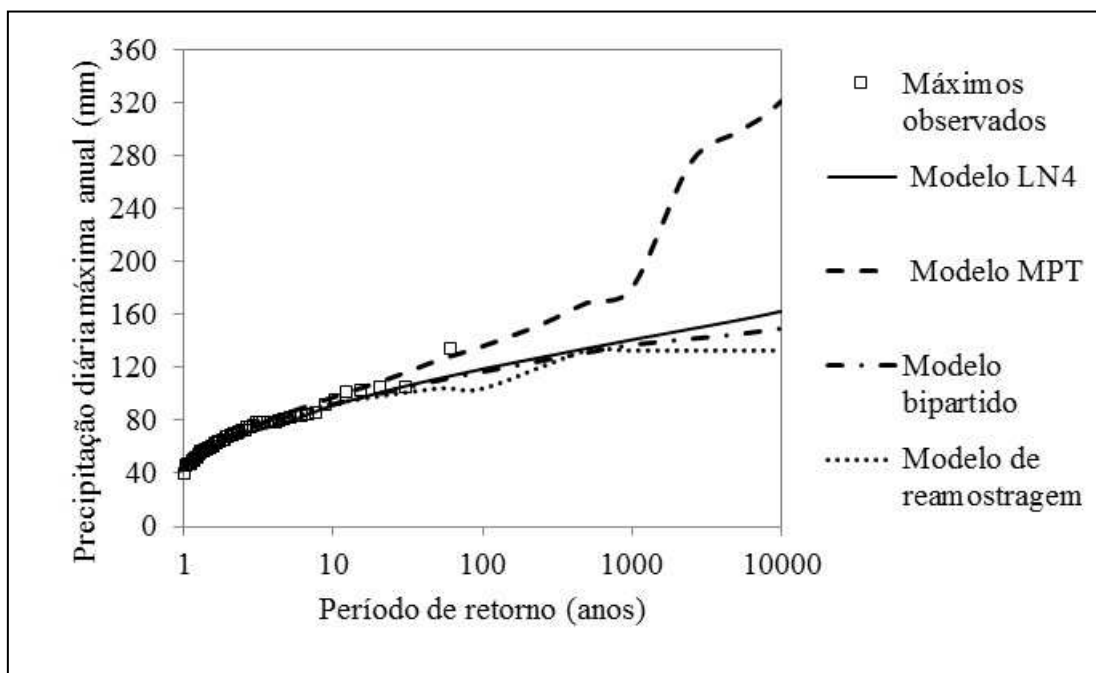


Figura 5.24 – Curvas de quantis para diferentes modelos de geração estocástica de precipitação diária

Com base nos argumentos apresentados, o uso de um modelo probabilístico limitado, que restringe os valores gerados a estimativas fisicamente plausíveis, constitui uma alternativa coerente para se abordar a questão de extrapolação no contexto da geração estocástica de precipitação diária. Entretanto, em face das limitações relacionadas a observação de precipitações extremamente elevadas, a validação do método proposto depende intrinsecamente das curvas de quantis de cheias obtidas do modelo de simulação hidrológica, as quais, por sua vez, são profundamente afetadas pelas incertezas oriundas da calibração. Esse problema é tratado no tópico seguinte.

5.6 Calibração dos parâmetros do modelo Rio Grande

A calibração de parâmetros de um modelo de simulação hidrológica, em um contexto de análise Bayesiana, requer a especificação de distribuições *a priori*, as quais descrevem faixas de variação plausíveis para a incerteza prévia acerca dos mesmos. No entanto, tão importante quanto a correta especificação das distribuições *a priori* dos parâmetros é o estabelecimento de uma função de verossimilhança adequada, capaz de identificar acuradamente a distribuição *a posteriori* dos referidos parâmetros e reproduzir o comportamento dos resíduos de calibração. Nesse contexto, a função de

verossimilhança generalizada proposta por Schoups & Vrugt (2010) é bastante interessante para fins de calibração, uma vez que a mesma acomoda a não normalidade, a heterocedasticidade e a autocorrelação dos resíduos, fornecendo assim estimativas mais confiáveis dos parâmetros do modelo chuva-vazão.

No caso do modelo Rio Grande, a calibração é restrita aos parâmetros do módulo de produção, descritos em detalhe na seção 3.5.3.1. Em uma análise inicial, admitiu-se as mesmas faixas de variação propostas por Silva *et al.* (2014) para os referidos parâmetros, na bacia do rio Pará, em Ponte do Vilela. Contudo, em função dos diferentes períodos de calibração e validação adotados na presente pesquisa, fez-se necessário assumir faixas de variação distintas do estudo citado. Os valores estabelecidos, no entanto, se encontram inseridos nas faixas de variação definidas no estudo de Pinheiro (2009), o qual reflete uma ampla gama de aplicações do modelo Rio Grande. As Tabelas 5.10 e 5.11 sumarizam as distribuições *a priori* para os parâmetros do modelo hidrológico e do modelo probabilístico assumido como função de verossimilhança, respectivamente. Tais intervalos constituem distribuições uniformes.

Tabela 5.10 – Intervalos de variação dos parâmetros do modelo Rio Grande

Símbolo	Limite inferior	Limite superior
<i>k</i>	0,5	1
<i>imp</i>	0	0,03
<i>wum</i>	5	150
<i>wlm</i>	50	250
<i>wdm</i>	5	150
<i>sm</i>	20	120
<i>b</i>	0,1	1
<i>ex</i>	0,1	2
<i>c</i>	0,01	0,4
<i>kss</i>	0,05	0,35
<i>kg</i>	0,05	0,65
<i>ci</i>	0,1	0,99
<i>cg</i>	0,8	0,999

Tabela 5.11 – Intervalos de variação dos parâmetros da função de verossimilhança generalizada

Símbolo	Limite inferior	Limite superior
σ	0	100
σ_0	-1	1
σ_1	0	1
β	-0,99	1
ξ	0,1	10
ϕ_i	0	1
μ_h	-	-

De posse dos intervalos de variação dos parâmetros do modelo Rio Grande e da função de verossimilhança, fez-se necessário definir os critérios para a estimação da distribuição *a posteriori* dos mesmos a partir da calibração via algoritmo DREAM. A Tabela 5.12 apresenta os valores adotados para a busca da distribuição de equilíbrio.

Tabela 5.12 – Parâmetros para aplicação do algoritmo DREAM

Símbolo	Valor adotado
<i>nseq</i>	10
<i>ndraw</i>	5×10^5
<i>burn-in</i>	50%
<i>thin.t</i>	50
<i>Rthres</i>	1,01

Utilizando o código desenvolvido por Silva *et al.* (2014), procedeu-se à estimação da distribuição conjunta *a posteriori* das variáveis modeladas. Adotou-se como período de calibração os 5 anos hidrológicos compreendidos entre 01/10/1992 e 30/09/1997. Para a validação, considerou-se o período entre 01/10/1997 e 30/09/2000. Após a convergência do algoritmo, foram simuladas 1.000 realizações da distribuição conjunta *a posteriori* dos parâmetros, a serem empregadas na geração das séries de vazão. A Figura 5.23 apresenta uma comparação das distribuições marginais *a priori* e *a posteriori* dos parâmetros do modelo Rio Grande após a calibração.

Avaliar a contribuição isolada de cada parâmetro na forma do hidrograma é uma tarefa de grande complexidade, em virtude da correlação entre os mesmos. Contudo, comentários gerais podem ser feitos a respeito da forma das distribuições marginais *a posteriori* dos parâmetros e das possíveis interações entre esses na modelagem de cada trecho do referido gráfico.

A análise da Figura 5.23 mostra que, à exceção do parâmetro c , para o qual a distribuição *a posteriori* guarda grande semelhança com a distribuição *a priori*, todos os parâmetros do modelo hidrológico são fortemente afetados pelos dados no processo de calibração. Os parâmetros sm , ex , que controlam a geração de escoamento superficial, e kss e kg , os quais afetam a geração de escoamento sub-superficial e subterrâneo, respectivamente, apresentaram distribuições *a posteriori* aproximadamente simétricas ou ligeiramente assimétricas à direita, as quais se distribuem em amplas faixas dos intervalos de variação assumidos *a priori*.

Por outro lado, a inferência dos parâmetros imp , wum , wlm , wdm , e b originou distribuições *a posteriori* fortemente assimétricas à esquerda, e bastante concentradas nas proximidades dos limites superiores dos intervalos admitidos *a priori*. O parâmetro b , em especial, originou estimativas *a posteriori* bastante próximas à unidade, o que sugere uma relação aproximadamente linear entre a capacidade de retenção de umidade pontual e o complemento da relação entre a área permeável e a área total da bacia.

Por fim, as distribuições *a posteriori* dos parâmetros ci e cg , os quais controlam a declividade dos trechos de recessão do hidrograma, situam-se bastante próximas aos limites máximos estabelecidos. Contudo, tais parâmetros se distribuem de maneira menos assimétrica em torno do valor modal.

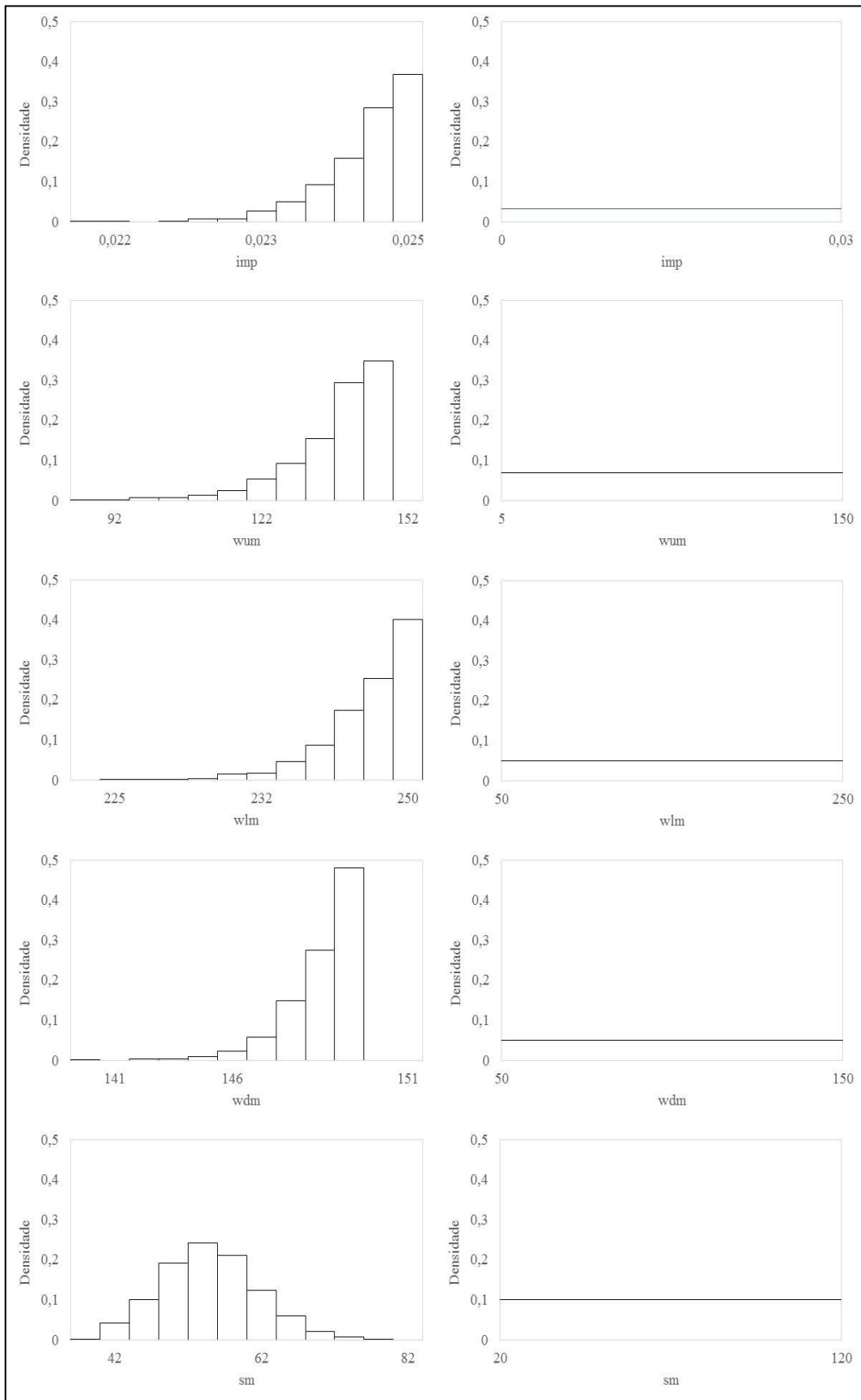


Figura 5.25 – Comparação das distribuições *a priori* e *a posteriori* dos parâmetros do modelo Rio Grande

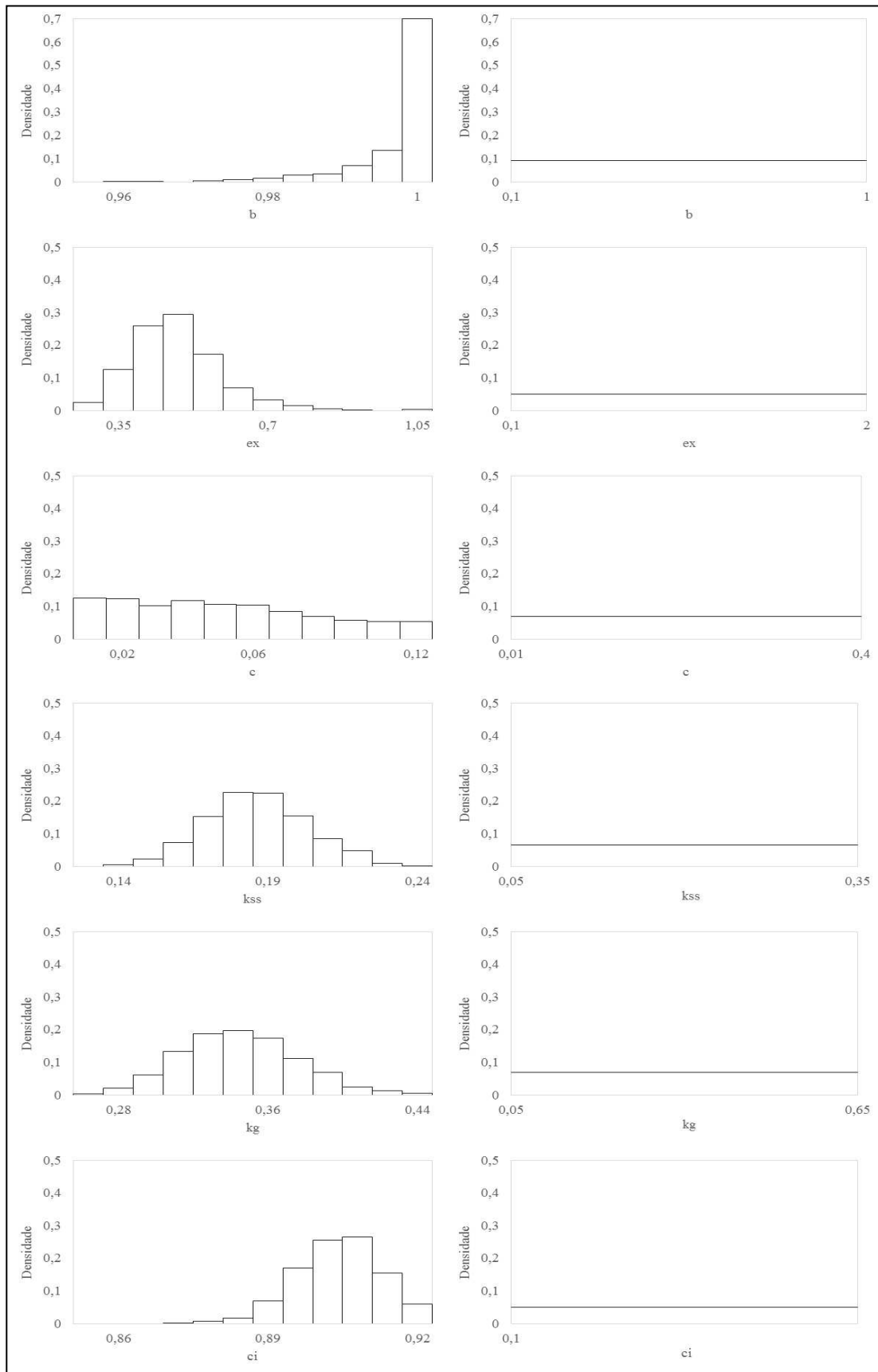


Figura 5.26 (cont.) – Comparação das distribuições *a priori* e *a posteriori* dos parâmetros do modelo Rio Grande

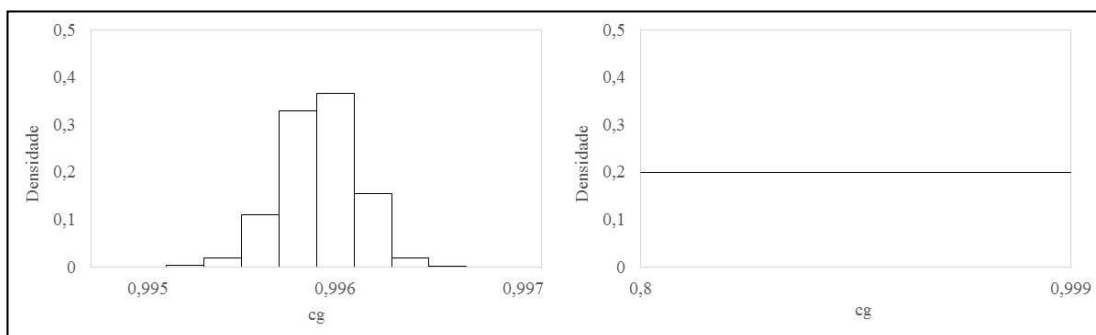


Figura 5.27 (cont.) – Comparação das distribuições *a priori* e *a posteriori* dos parâmetros do modelo Rio Grande

A Tabela 5.13 sumaria os índices de desempenho do período de calibração, calculados a partir da moda *a posteriori* dos parâmetros. Verifica-se a partir do coeficiente de Nash-Sutcliffe que o modelo Rio Grande explica aproximadamente 84% da variância natural das vazões na estação de Ponte do Vilela. Esse valor é bastante similar àquele obtido no estudo de Silva *et al.* (2014), o qual também fez uso da função de verossimilhança generalizada, e superior ao do estudo de Lima (2004), o qual emprega o procedimento tradicional de calibração. Percebe-se também que o modelo hidrológico se mostra apto a representar os volumes observados, bem como reproduzir a coerência temporal da série de vazões, fato esse evidenciado pelo elevado valor do coeficiente de Pearson.

Tabela 5.13 – Índices de desempenho do período de calibração

Índice de desempenho	Valor
Nash	0,844
RMSE	11,322
RV	0,990
Pearson	0,941

A Figura 5.24 mostra a concordância entre o hidrograma simulado e o hidrograma observado. É possível perceber que os períodos de recessão são reproduzidos de maneira bastante acurada. A modelagem das vazões de pico, por outro lado, é sujeita a erros de maior magnitude. Com efeito, alguns dos picos simulados são bastante inferiores àqueles observados. Esse fato pode decorrer da forte interação entre os parâmetros do modelo durante a síntese hidrológica. Contudo, o modelo Rio Grande se mostrou apto a reproduzir picos de grande magnitude, o que é essencial para a simulação dos quantis de cheias extremas a ser discutida na próxima seção.

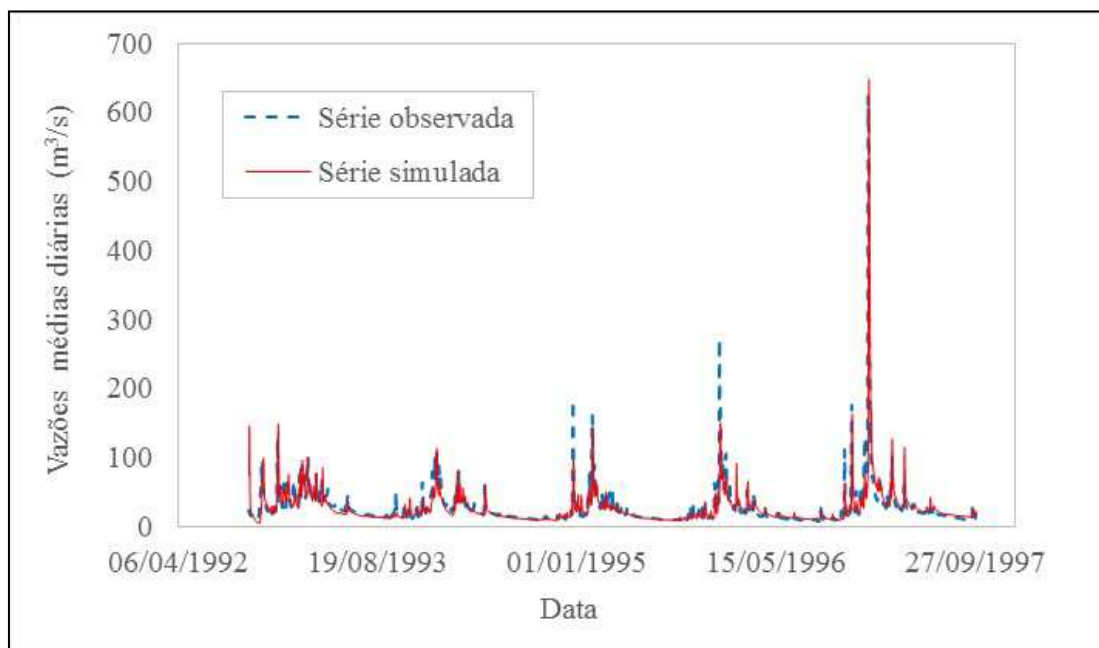


Figura 5.28 – Comparação entre os hidrogramas simulado e observado para o período de calibração

A Figura 5.25 apresenta a relação entre as vazões simuladas e aquelas observadas. É possível aferir a qualidade da calibração em face da reduzida dispersão ao redor da reta de 45°, ainda que se observe algum viés nos valores mais elevados de vazão.

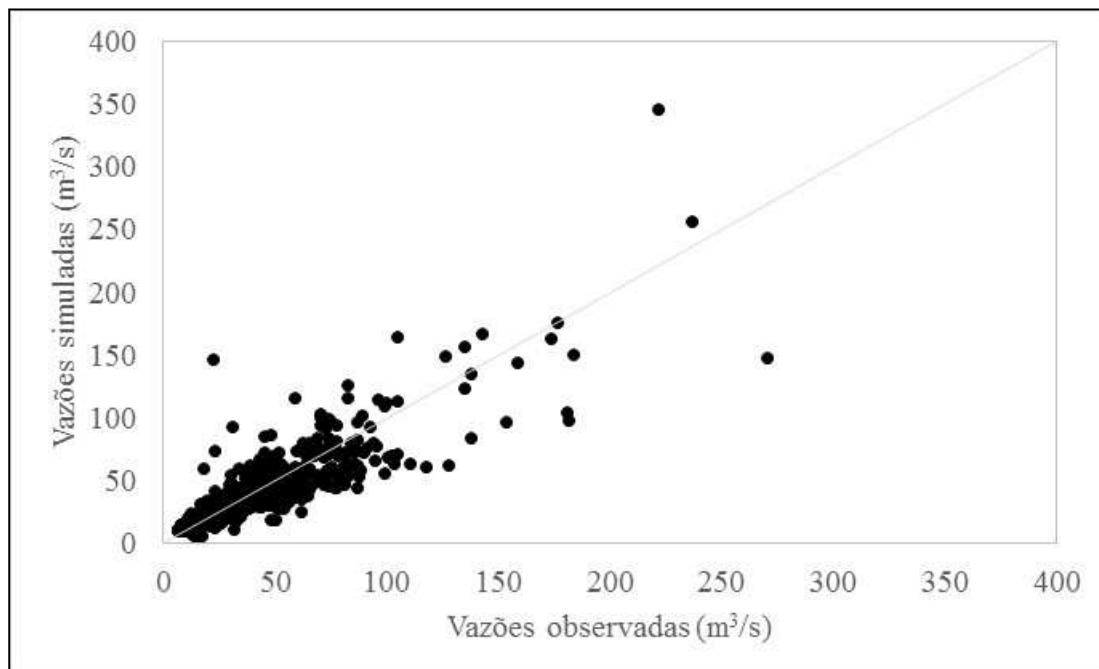


Figura 5.29 – Vazões simuladas versus vazões observadas

A Figura 5.26, por sua vez, ilustra o comportamento dos resíduos em relação às vazões observadas. Percebe-se que o uso da função de verossimilhança generalizada remove de maneira quase absoluta a heterocedasticidade dos erros de calibração. Com efeito, não

se observa o aumento dos resíduos com as vazões, comportamento esse usualmente verificado em calibrações baseadas no método dos mínimos quadrados, no qual os erros de modelagem são distribuídos segundo um modelo Gaussiano.

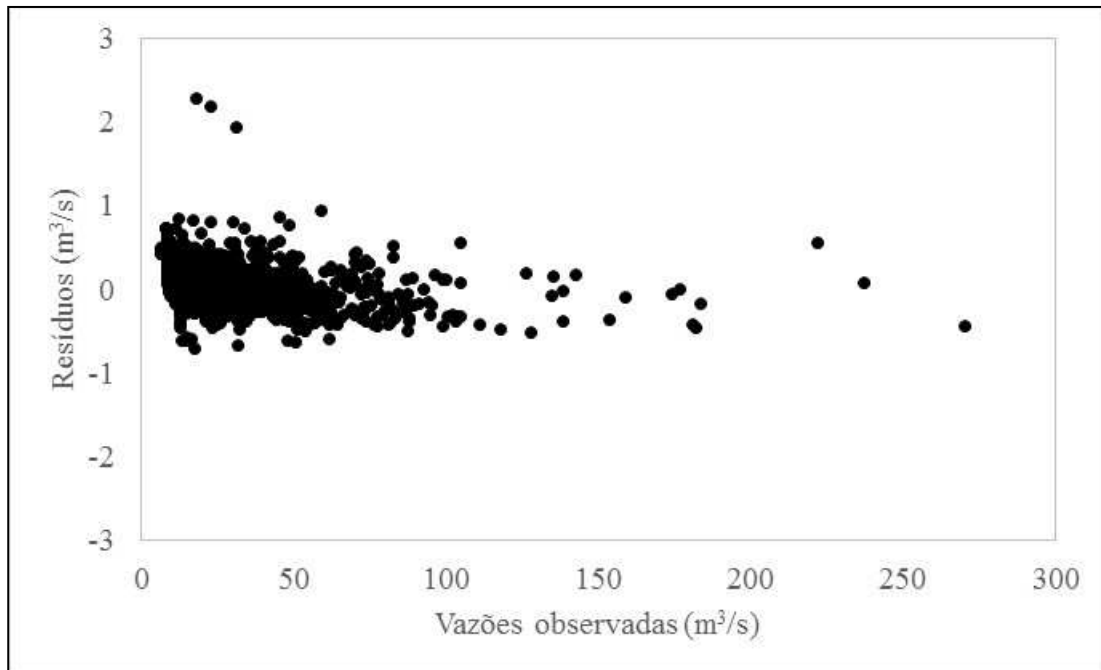


Figura 5.30 – Resíduos de calibração com a função de verossimilhança generalizada versus vazões observadas

A Figura 5.27 mostra o gráfico de autocorrelação dos resíduos de calibração. Verifica-se que, à exceção de uma pequena autocorrelação de ordem 1, os erros de simulação são essencialmente independentes. Por fim, a Figura 5.28 ilustra a distribuição empírica dos resíduos com relação ao modelo teórico proposto para os mesmos. Percebe-se a excelente aderência entre os dados e o modelo probabilístico dos resíduos. Esse conjunto de avaliações permite validar as hipóteses assumidas *a priori* acerca do comportamento dos resíduos, o que evidencia as vantagens da adoção de um modelo probabilístico teórico como função de verossimilhança.

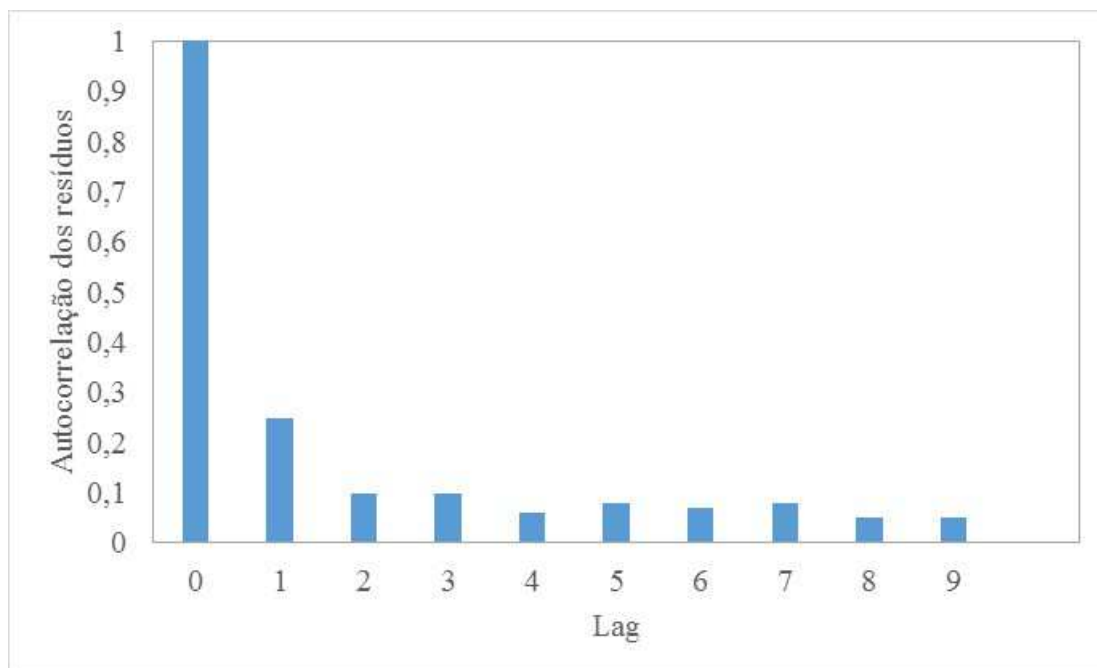


Figura 5.31 – Autocorrelação dos resíduos de calibração

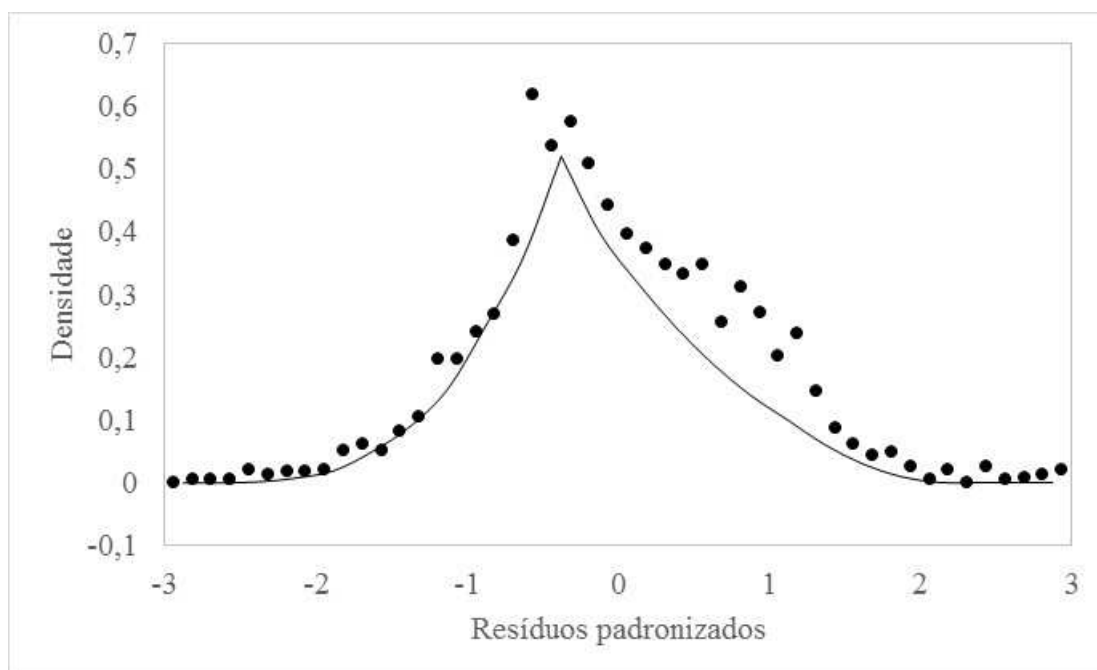


Figura 5.32 – Distribuições empírica (círculos) e teórica (linha contínua) dos resíduos de calibração padronizados

A Tabela 5.14 sumaria os índices de desempenho para o período de validação. É possível observar que o ajuste é ligeiramente inferior nessa situação. Ainda assim, o modelo Rio Grande foi capaz de explicar em torno de 77% da variância natural das vazões no período avaliado, além de reproduzir satisfatoriamente os volumes observados e a coerência temporal da série de vazões. Esse fato evidencia a adequação do conjunto de parâmetros inferido da distribuição conjunta *a posteriori* para modelar

as vazões diárias em períodos distintos daquele empregado na calibração. A Figura 5.29 ilustra a concordância entre os hidrogramas simulado e observado no período de validação.

Tabela 5.14 – Índices de desempenho para o período de validação

Índice de desempenho	Valor
Nash	0,771
RMSE	13,2
RV	0,98
Pearson	0,89

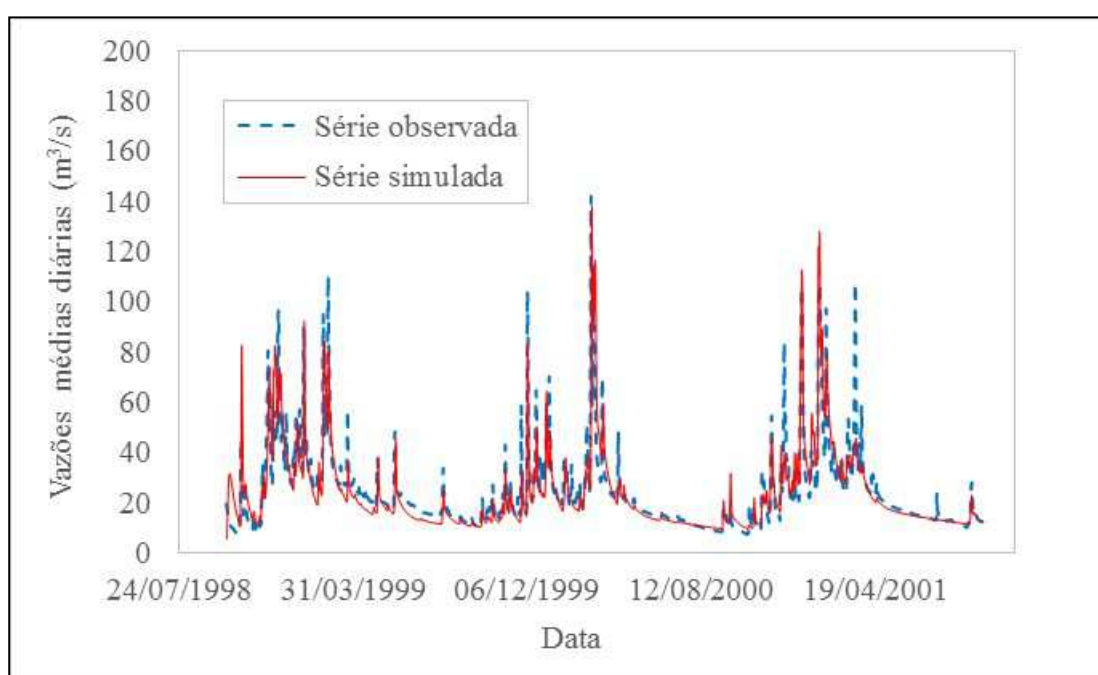


Figura 5.33 – Comparação entre os hidrogramas simulado e observado para o período de validação

5.7 Construção das curvas de quantis de enchentes

A etapa final do método proposto é a construção das curvas de quantis de enchentes para as séries de precipitação de 10.000 anos. Para tanto, foram geradas 1.000 séries de precipitação com o gerador proposto, as quais foram combinadas com um conjunto de 1.000 amostras de parâmetros obtidas aleatoriamente da distribuição conjunta *a posteriori* dos mesmos.

A Tabela 5.15 apresenta um resumo das estimativas de quantis para tempos de retorno diversos compreendidos entre 2 e 10.000 anos. A Figura 5.30, por sua vez, ilustra a comparação entre as enchentes simuladas e observadas para os referidos tempos de

retorno. Os valores observados são representados por círculos. A probabilidade empírica de excedência foi calculada com auxílio da posição de plotagem de Cunnane.

Tabela 5.15 – Estimativa de quantis de enchentes para diversos tempos de retorno

Tempo de retorno (anos)	Vazão (m ³ /s)	95% HPD
2	166,6	(162,2;170,6)
5	280,1	(272,0,6;287,7)
10	325,0	(315,1;334,7)
25	401,8	(384,2;419,2)
50	468,9	(449,5;498,2)
100	521,4	(491,7;552,6)
250	590,3	(560,3;659,1)
500	644,1	(595,9;696,1)
1000	737,6	(676,0;819,2)
2500	870,5	(742,3;958,7)
5000	946,9	(777,8;1042,5)
7500	972,6	(816,0;1126,0)
10000	996,3	(849,8;1231,2)

95% HPD – Intervalo de credibilidade

A análise da Figura 5.30 evidencia a ausência de aderência entre as cheias simuladas e observadas, mesmo para quantis com maior probabilidade de excedência. Com efeito, as enchentes observadas se encontram além do limite superior do intervalo de credibilidade de 95% já para períodos de retorno de aproximadamente 10 anos. Tal fato pode constituir um indício de que o método proposto não é apto a reproduzir as cheias máximas anuais na bacia do rio Pará, em Ponte do Vilela. Contudo, comportamento similar ao descrito foi verificado no trabalho de Lima (2004), o que sugere que a síntese hidrológica do modelo Rio Grande é a principal responsável pela caracterização pouco acurada das vazões de pico. Assim, nas análises subsequentes, essa limitação será desconsiderada, e serão enfatizadas as vazões com tempo de retorno entre 100 e 10.000 anos, as quais constituem o objeto final dessa pesquisa.

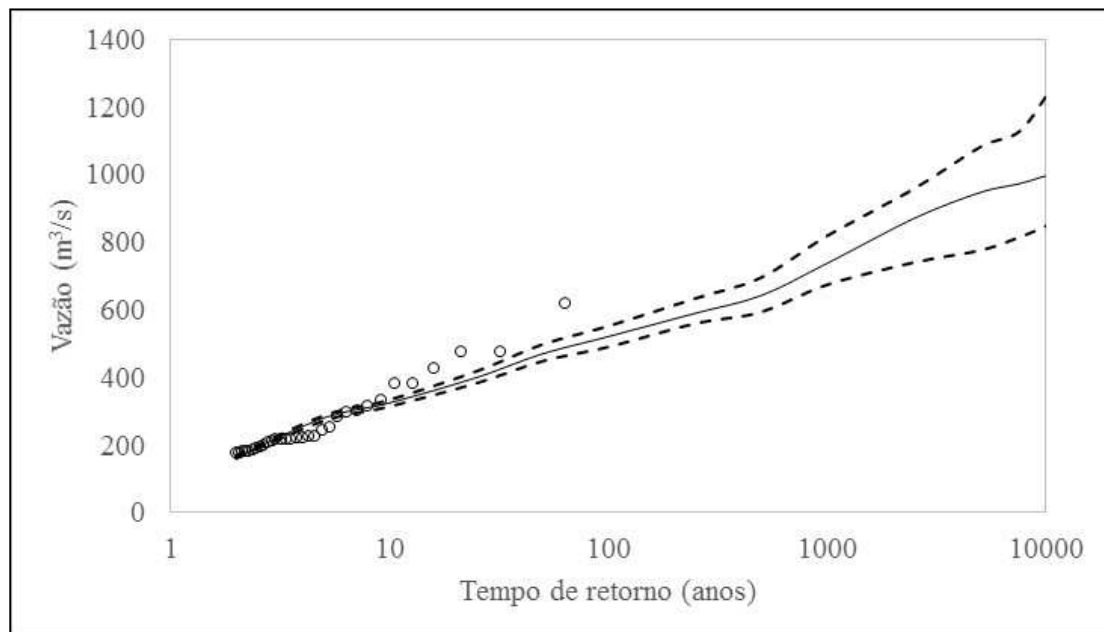


Figura 5.34 – Curva de quantis de enchentes na bacia do rio Pará, em Ponte do Vilela. As linhas tracejadas correspondem ao intervalo de credibilidade de 95%.

Na Figura 5.31 constam as curvas de quantis de enchentes obtidas por diversos métodos de simulação, os quais compreendem:

- O Método Pico-Volume-Precipitação (PVP), empregado na estação de Ponte do Vilela por Fernandes & Naghettini (2008);
- Análise de frequência convencional (AFC), empregando-se a distribuição GEV, com parâmetros estimados pelo método dos momentos-L;
- Análise de frequência regional (AFR), com quantis obtidos a partir do estudo de CPRM (2001);
- Análise de frequência Bayesiana (AFB), empregando-se a distribuição LN4 e a estimativa local de EMP como estimador do limite superior (FERNANDES *et al.*, 2010);
- O modelo de simulação (SIM) proposto por Lima (2004), o qual também envolve o modelo Rio Grande; e
- Análise de frequência por meio de séries de duração parcial (SPD), utilizando-se o modelo de Poisson-Pareto com taxa média de excedência igual a 1,5 (FERNANDES, 2009).

Tendo-se em vista as diferentes hipóteses que envolvem a construção desses modelos, os mesmos não são diretamente comparáveis. Contudo, comentários gerais podem ser feitos. A Figura 5.31 mostra que, de todos os modelos avaliados, aqueles que apresentam maior aderência aos registros observados são o método PVP e a análise de frequência convencional com a distribuição GEV. Contudo, tais modelos crescem de maneira especialmente acentuada para tempos de retorno superiores a 100 anos. Tendo-se em vista a não observação de valores de enchentes extremamente elevados, tais como paleocheias, a cauda superior da distribuição das vazões máximas anuais pode não ser adequadamente representada na análise de frequência convencional e, assim, esse método pode não ser ineficiente para a extrapolação aos quantis com maiores tempos de retorno. O método PVP, por outro lado, condensa em sua estrutura os principais fatores intervenientes na formação de cheias extremas, a saber, a hidrometeorologia local, a transformação chuva-vazão e a hidráulica fluvial (FERNANDES & NAGHETTINI, 2008). No entanto, ainda que agregue maior realismo físico à modelagem, esse método também envolve as incertezas relacionadas à estimação paramétrica e à síntese hidrológica, as quais não são consideradas de maneira explícita na análise e podem comprometer as estimativas de quantis com probabilidade de excedência reduzida.

Em uma posição intermediária, encontram-se a análise de frequência regional e o modelo de simulação proposto por Lima (2004). Diferentemente dos métodos tratados no parágrafo anterior, esses modelos não apresentam ajuste satisfatório aos dados, mesmo para tempos de retorno da ordem de 10 anos. Contudo, a incorporação de informação aos modelos, seja pela substituição do tempo pelo espaço, no caso do modelo regional, seja pelo aumento do tamanho da amostra, no caso do modelo de simulação hidrológica, melhora a caracterização da cauda superior das vazões máximas anuais, e, assim, espera-se que as estimativas de quantis associados a tempos de retorno mais elevados sejam mais confiáveis que aquelas apresentadas pelo método PVP e pela análise de frequência convencional. Ainda assim, tanto a análise de frequência regional quanto o modelo de Lima (2004) empregam distribuições ilimitadas, o que pode se traduzir em um grande viés na modelagem de eventos extremos quando da extrapolação.

O modelo de séries de duração parcial, por sua vez, se mostrou completamente inadequado ao ajuste dos dados, tendo-se em vista que todos os quantis estimados por

esse método foram inferiores àqueles observados. Ainda assim, a estimativa do quantil de 10.000 de tempo de retorno foi similar ao modelo proposto na presente pesquisa e ao modelo de análise de frequência Bayesiana proposto por Fernandes *et al.* (2010).

Por fim, a curva obtida com o método proposto apresenta comportamento bastante similar ao modelo Bayesiano proposto por Fernandes *et al.* (2010), ao longo do domínio das probabilidades de excedência avaliado. Tendo-se em vista a fundamentação física imposta a esses dois modelos, sintetizada pela especificação de modelos probabilísticos para os limites superiores de precipitação e vazão, respectivamente, torna-se bastante razoável supor que os mesmos são mais efetivos na extrapolação até os quantis mais extremos. Não obstante, ambos os modelos derivados do paradigma Bayesiano se mostraram inaptos a reproduzir os quantis de cheias observados.

É interessante comentar também que, de acordo com o estudo de Fernandes (2009), o modelo EV4, o qual foi considerado inadequado para a modelagem de precipitações extremas na bacia do rio Pará, em Ponte do Vilela, mostrou excelente aderência aos dados de enchentes.

Apesar do exposto nos parágrafos anteriores, é extremamente difícil decidir pelo uso de um modelo particular com base somente na amostra disponível. Com efeito, dados sistemáticos isoladamente, via de regra, não fornecem subsídios para a extrapolação até os quantis mais extremos. Além disso, tendo-se em vista que, à exceção da análise de frequência com a distribuição LN4, todos os métodos de estimação de cheias avaliados se baseiam no paradigma frequentista, não é possível comparar diretamente os modelos no que se refere às incertezas. A aplicação do método na bacia do rio American, a qual dispõe de um conjunto de paleocheias e estudos de enchentes baseados no paradigma Bayesiano, pode fornecer argumentos mais consistentes para a avaliação do método proposto na estimação de cheias associadas a tempos de retorno mais elevados.

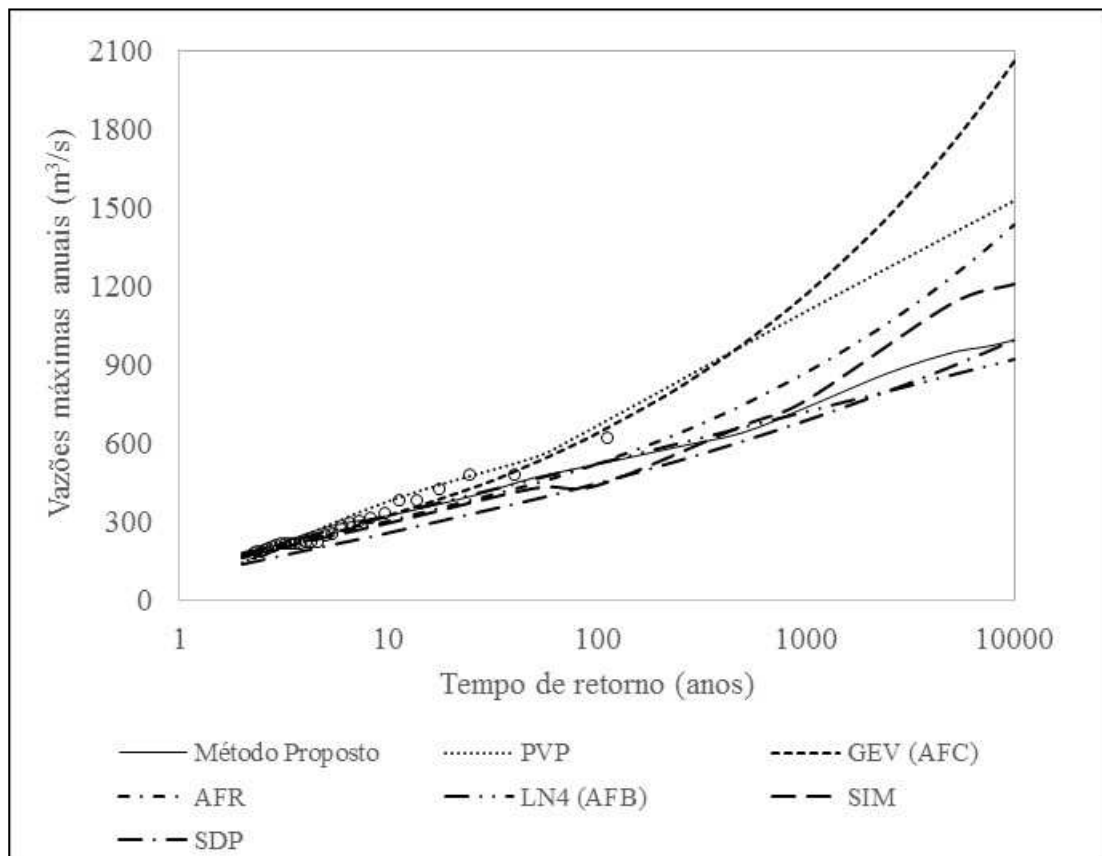


Figura 5.35 – Comparação das curvas de quantis de enchentes obtidas com diferentes métodos

6 APLICAÇÃO DO MÉTODO À BACIA DO RIO AMERICAN

A segunda aplicação do método proposto nesta pesquisa é realizada na bacia do rio American, na barragem de Folsom, no estado americano da Califórnia. Essa bacia foi selecionada em função da presença de um conjunto de informações não sistemáticas de cheias, as quais possibilitam uma melhor caracterização do comportamento da cauda superior das enchentes e, assim, permitem uma avaliação mais fundamentada da qualidade do método proposto.

6.1 Caracterização da bacia do rio American, em Folsom

O rio American nasce na cordilheira de Sierra Nevada e percorre um curso de 192 km até sua confluência com o rio Sacramento, na cidade de mesmo nome. O curso d'água possui 3 tributários principais: o chamado North Fork, o qual se desenvolve por 142 km, até receber a contribuição do trecho denominado Middle Fork, nas proximidades da cidade de Auburn. O curso d'água resultante se desenvolve por mais 50km até a junção com o terceiro tributário, South Fork, no reservatório de Folsom.

O reservatório de Folsom foi implantado em 1955 e localiza-se imediatamente a montante da cidade de Sacramento. De acordo com USBR (2002), o reservatório é parte integrante de um amplo projeto que contempla aproveitamento hidrelétrico, uso de água para abastecimento e irrigação e proteção contra cheias na região de Sacramento.

Na seção fluvial no reservatório de Folsom, a bacia hidrográfica do rio American compreende uma área de drenagem de 4.820 km², na qual se encontram 20 estações de monitoramento sistemático, operadas pelos USGS (*United States Geological Service*), e ao menos 3 pontos com indicativos de paleocheias. A região de estudo é mostrada na Figura 6.1.

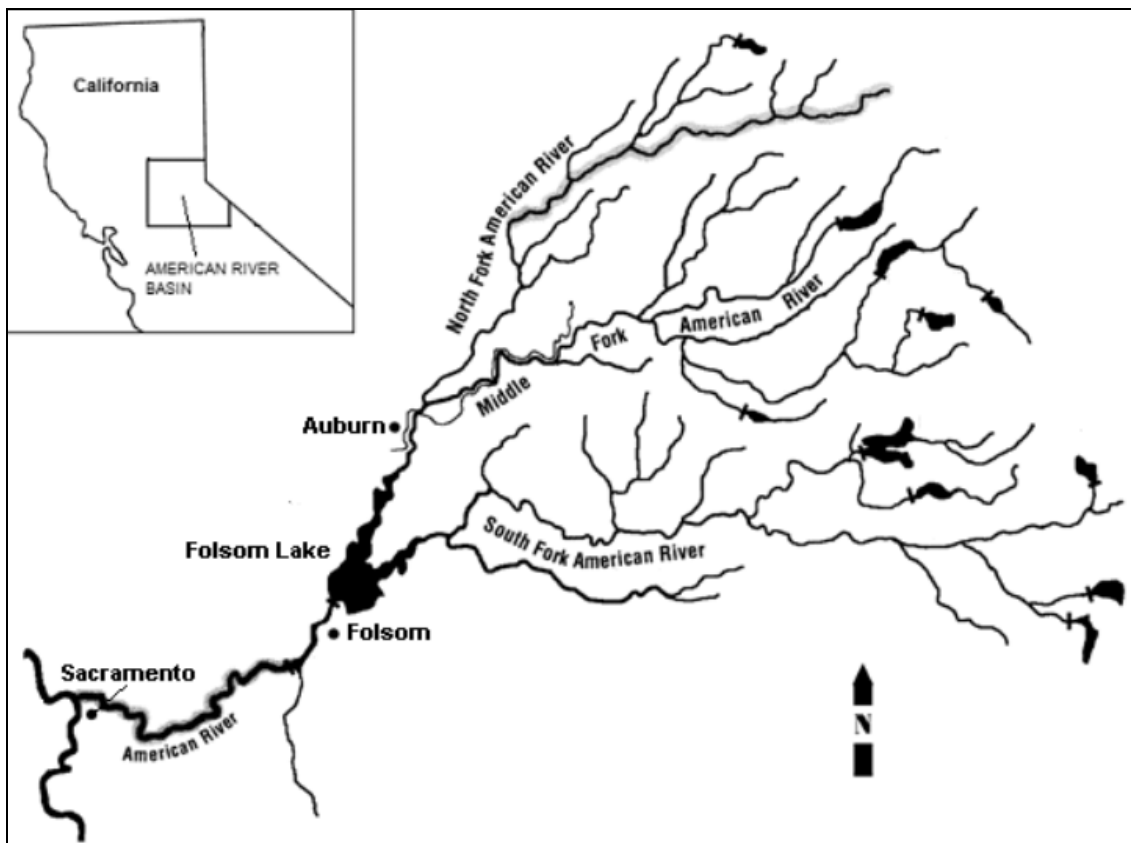


Figura 6.1 – Localização da bacia do rio American, na estação de Fair Oaks (11446500)
 Fonte: USBR (2002)

6.2 Estações pluviométricas, fluviométricas e evaporimétricas utilizadas no estudo

A seleção das estações pluviométricas para a presente aplicação teve como referência o estudo de USACE (2005). Esse trabalho, cujo propósito é a estimação indireta de cheias por meio de modelos hidrológicos, fez uso de um conjunto de 6 estações pluviométricas, distribuídas por toda a extensão da bacia hidrográfica, para estimar a chuva média espacializada. Inicialmente, tentou-se empregar as mesmas estações na presente aplicação. Contudo, uma análise expedita das séries temporais, as quais foram obtidas do site do “*National Oceanic and Atmospheric Administration*”, <http://www1.ncdc.noaa.gov>, mostrou grande inconsistência nos dados da estação Soda Springs 1E, possivelmente em decorrência de alterações nas condições de monitoramento. Diante disso, a referida estação foi excluída da análise e os polígonos de Thiessen foram novamente traçados para as 5 estações restantes, como mostra a Figura 6.2.

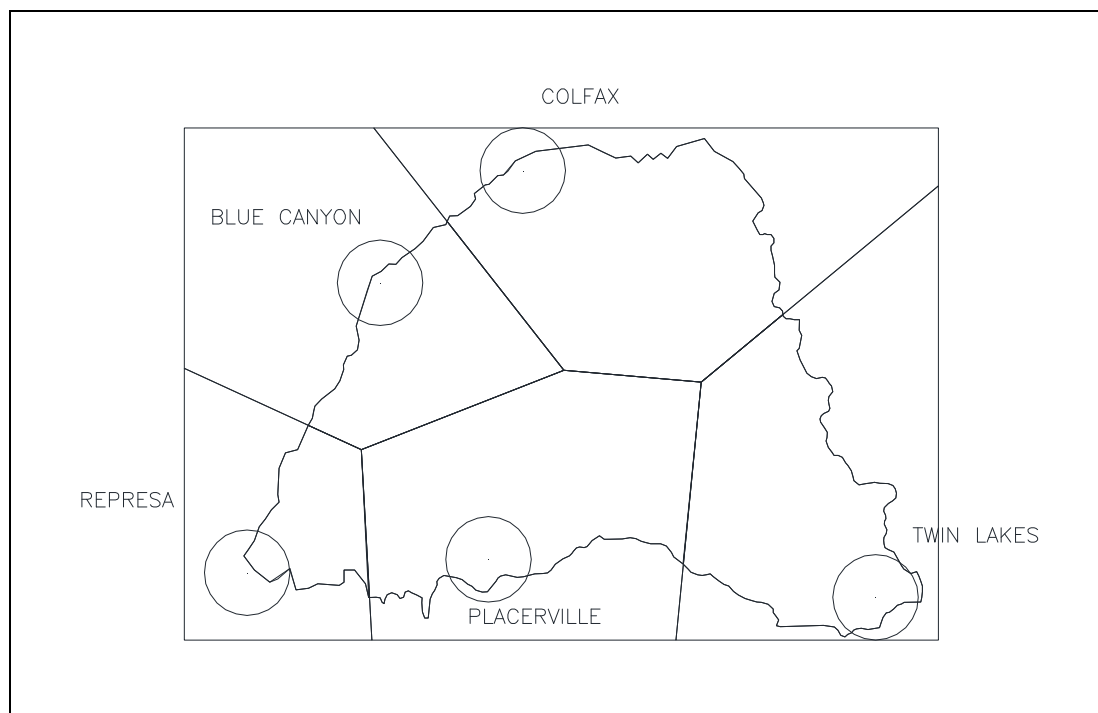


Figura 6.2 – Polígonos de Thiessen para a determinação da chuva média espacializada na bacia do rio American, em Folsom

As estações empregadas na determinação da chuva média espacializada na bacia do rio American, em Folsom, são apresentadas na Tabela 6.1, com as respectivas áreas de influência. Para os propósitos da presente aplicação, considerou-se um período de observação de dados de 35 anos, compreendidos entre 1961 e 1995.

Tabela 6.1 – Estações pluviométricas utilizadas para determinação da chuva média espacializada na bacia do rio American, em Folsom.

Estação	Código	Área de Influência
Blue Canyon	USW00023225	28,5
Colfax	CA95713	14,6
Placerville	USC00046960	28,1
Represa	USC00047370	6,7
Twin Lakes	USC00049105	22,1
Total		100,0

Os dados de evaporação, por sua vez, foram obtidos no relatório técnico TR NWS 34 (NOAA, 1982), o qual fornece estimativas da evaporação potencial mensal para diversas localidades nos Estados Unidos. Como expediente prático para estimar o *input* do modelo hidrológico, as alturas de evaporação foram divididas de maneira igualitária nos dias de cada mês do ano. Ainda que não se disponha desses dados em base diária, pode-se supor que a incerteza acerca dos mesmos é devidamente incorporada à análise por meio da distribuição *a priori* do coeficiente de tanque *k* do modelo Rio Grande.

Por fim, os dados fluviométricos foram obtidos na estação de Fair Oaks (USGS11446500), a qual se localiza imediatamente a jusante do reservatório de Folsom e cujo monitoramento sistemático remonta a 1905. As vazões registradas na referida estação são fortemente regularizadas, tendo-se em vista que, de acordo com USBR (2002), além do próprio reservatório de Folsom, existem 5 grandes reservatórios que controlam aproximadamente 90% do armazenamento a montante da mesma.

O conjunto de vazões regularizadas não se presta aos propósitos de calibração dos parâmetros do modelo de simulação hidrológica. Em vista desse fato, fez-se a opção pelo uso de registros de vazões anteriores à construção dos referidos reservatórios. Dessa forma, estabeleceu-se o período compreendido entre 01/10/1920 e 30/09/1925 para a calibração do modelo chuva-vazão e o período de 3 anos hidrológicos subsequentes para sua validação.

Cabe ressaltar que parte do escoamento na bacia é oriundo de degelo. Entretanto, de acordo com USBR (2002), tal contribuição é pouco significativa, e, assim, assume-se que a incerteza acerca da mesma pode ser coerentemente acomodada na incerteza do *output* do modelo de simulação hidrológica.

6.3 Distribuição *a priori* para o limite superior

A estimativa da PMP meteorológica de 1 dia na bacia do rio American, em Folsom, é de 370 mm (HMR59, 1999). De acordo com o procedimento especificado no item 4.2.1, a distribuição gama será empregada para modelar as incertezas com relação ao limite superior. Dessa forma, torna-se necessário estimar o coeficiente de variação e a probabilidade de não-excedência da PMP local.

O coeficiente de variação estimado para o conjunto de 39 estimativas de PMP meteorológicas na Califórnia é igual a 0,304. Já a probabilidade de não-excedência empírica, estimada com a posição de plotagem de Weibull, é igual a 0,084. Assim, a distribuição *a priori* eliciada para o limite superior é $\alpha \sim GAMA(10,821;0,018)$, denotada por modelo I.

Também nessa aplicação foi considerada uma distribuição *a priori* não informativa para o limite superior, com o intuito de se avaliar a quantidade de informação agregada à análise pelo modelo I. A distribuição *a priori* eliciada nesse caso é

$\alpha \sim GAMA(1,0;1,0 \times 10^{-8})$, a qual é denotada por modelo II. A Tabela 6.2 apresenta as principais características das distribuições utilizadas como especificação *a priori* para o limite superior.

Tabela 6.2 – Parâmetros e características das distribuições *a priori* do limite superior

Distribuição <i>a priori</i>	ρ_α	β_α	Média	Mediana	CV_α	DP
I	10,821	0,018	601,2	582,8	0,304	182,8
II	1,000	$1,000 \times 10^{-8}$	$1,000 \times 10^8$	$6,900 \times 10^7$	1,000	$1,000 \times 10^8$

6.4 Estatísticas *a posteriori*

6.4.1 Modelo LN4

Conforme exposto na aplicação anterior, a interpretação dos parâmetros μ e σ , em termos hidrometeorológicos, não é trivial e, assim, assumem-se distribuições *a priori* não informativas para os mesmos. As distribuições eliciadas são:

$$\mu \sim NORMAL(1,0;1,0 \times 10^{-6}) \text{ e } \sigma \sim GAMA(1,0;1,0 \times 10^{-8}).$$

As simulações numéricas foram realizadas com auxílio do software WinBUGS (LUNN *et al.*, 2000). A estatística de Brooks-Gelman-Rubin (Brooks & Gelman, 1998) mostrou a convergência à distribuição de equilíbrio em aproximadamente 40.000 iterações para os 3 parâmetros do modelo LN4. Também aqui um *lag* de tamanho 20 foi suficiente para se obter uma amostra final não correlacionada.

A inferência das distribuições *a posteriori* originou conclusões similares àquelas da aplicação anterior em relação ao comportamento dos parâmetros. Dessa forma, os resultados obtidos com o modelo LN4 para a bacia de Folsom serão discutidos apenas em aspectos gerais.

Estatísticas a posteriori para o parâmetro μ

A Tabela 6.3 sumaria as estatísticas *a posteriori* para o parâmetro μ . A Figura 6.3 apresenta as distribuições *a posteriori* para os dois modelos avaliados.

Tabela 6.3 – Estatísticas *a posteriori* para o parâmetro μ

Modelo	Média	DP	CV	95%HPD
I	-1,795	0,430	-0,240	(-2,518;-0,824)
II	-13,631	1,358	-0,100	(-15,55;-10,32)

95% HPD – Intervalo de credibilidade

Também nessa aplicação fica evidente a sensibilidade de μ à especificação da distribuição *a priori* do limite superior, expressa em termos da posição das duas distribuições *a posteriori* no domínio do parâmetro e das respectivas variâncias. Cabe ressaltar ainda que, para o modelo I, a distribuição *a posteriori* é unimodal e aproximadamente simétrica. Já no caso do modelo II, a distribuição *a posteriori* também é unimodal. Contudo, a mesma apresenta ligeira assimetria positiva.

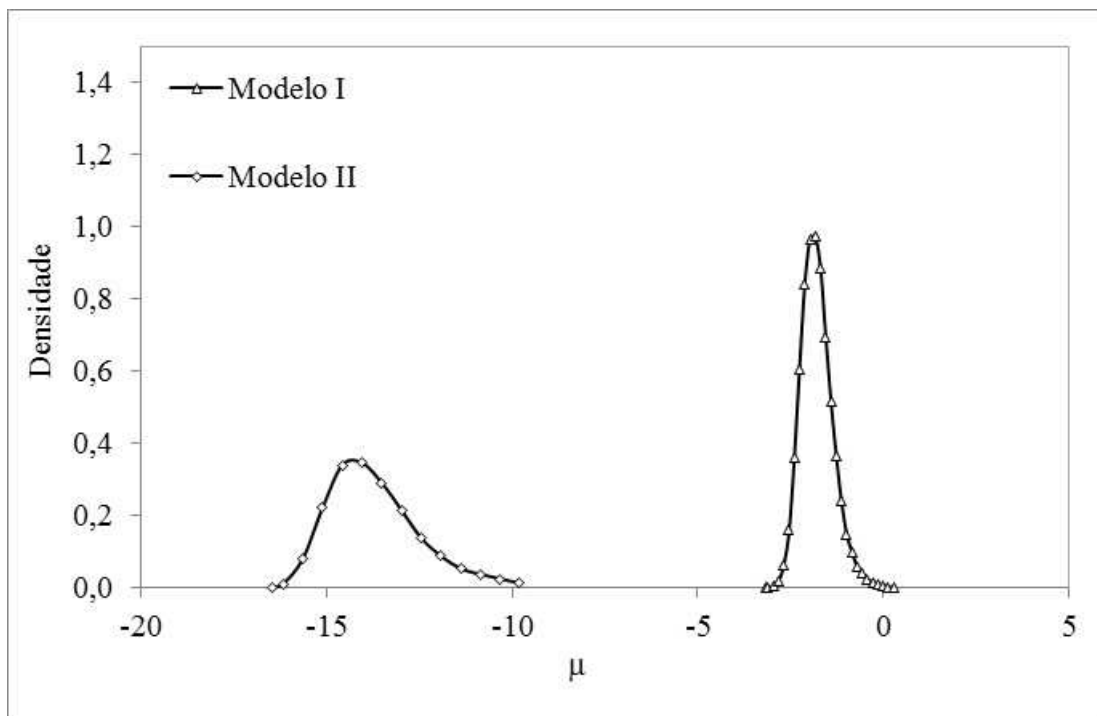


Figura 6.3 – Distribuições *a posteriori* para o parâmetro μ

Estatísticas a posteriori para o parâmetro σ

A Tabela 6.4 sumaria as estatísticas *a posteriori* para o parâmetro σ . A Figura 5.6 apresenta as distribuições *a posteriori* para os dois modelos.

Tabela 6.4 – Estatísticas *a posteriori* para o parâmetro σ

Modelo	Média	DP	CV	95%HPD
I	0,424	0,083	0,196	(0,300;0,622)
II	0,364	0,065	0,179	(0,264;0,515)

95% HPD – Intervalo de credibilidade

É possível observar a reduzida influência da distribuição *a priori* do limite superior na inferência de σ . De fato, as distribuições *a posteriori* são relativamente próximas e apresentam variâncias similares. Além disso, essas últimas são unimodais e aproximadamente simétricas.

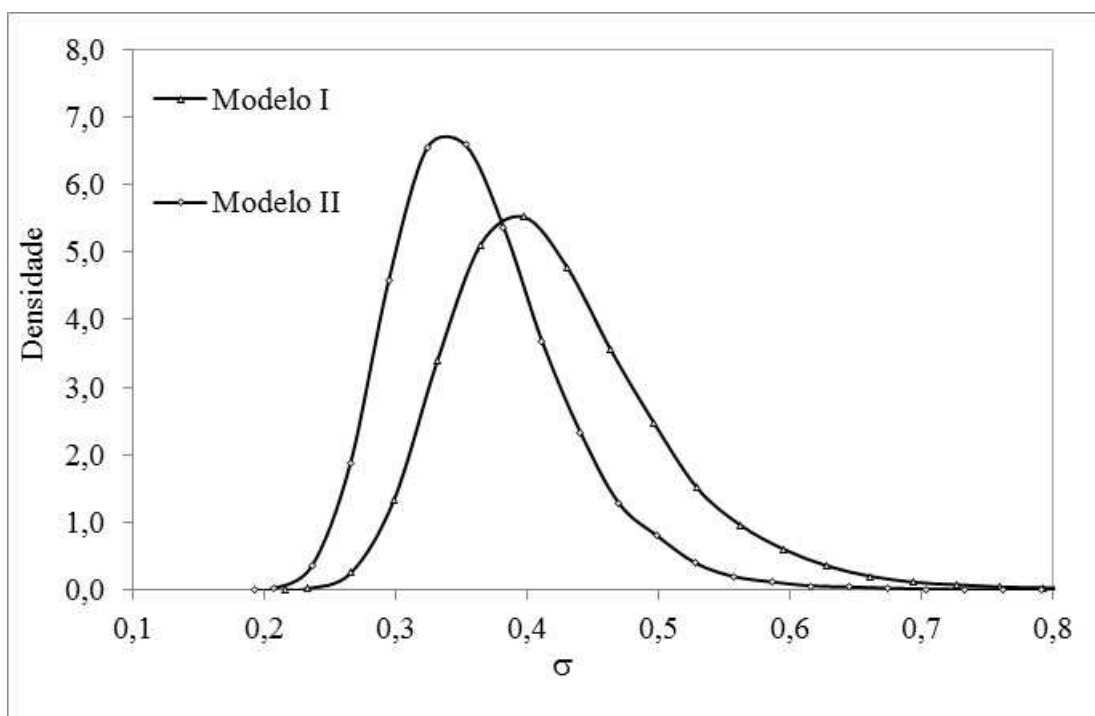


Figura 6.4 – Distribuições *a posteriori* para o parâmetro σ

Estatísticas a posteriori para o parâmetro α

A Tabela 6.5 sintetiza as estatísticas *a posteriori* para o parâmetro α . A Figura 5.6 apresenta as distribuições *a priori* e *a posteriori* para os dois modelos.

Tabela 6.5 – Estatísticas *a posteriori* para o parâmetro α

Modelo	Média	DP	CV	95%HPD
I	489,4	162,8	0,333	(218,8;852,2)
II	$9,8 \times 10^7$	$9,8 \times 10^7$	1	$(1,9 \times 10^6; 3,6 \times 10^8)$

95% HPD – Intervalo de credibilidade

Com relação ao limite superior, cabem comentários mais detalhados. Primeiramente, é possível observar que a influência dos dados parece mais significativa que a da aplicação anterior na inferência da distribuição *a posteriori* do referido parâmetro. Com efeito, a diferença entre as médias dessa distribuição e da distribuição *a priori* é significativamente maior do que aquela verificada na bacia do rio Pará. Esse fato é intrinsecamente relacionado às características amostrais da bacia do rio American, a qual se mostra isenta de “*outliers*” no período de observação de dados, e assim incorre em maior deslocamento da distribuição *a posteriori* em direção aos dados.

Além disso, percebe-se o efeito do aumento do coeficiente de variação regional das estimativas de PMP na distribuição *a posteriori* do limite superior. De fato, tanto os valores de CV_α quanto a amplitude do intervalo de credibilidade são sensivelmente maiores que aqueles estimados na bacia do rio Pará, evidenciando a maior variabilidade na descrição desse parâmetro na bacia de Folsom.

Por fim, observa-se que a eliciação de uma distribuição *a priori* não informativa para o limite superior conduz a estimativas fisicamente implausíveis para tal parâmetro, uma vez que a distribuição *a posteriori* atribui massa aproximadamente constante ao longo do domínio do mesmo quando da aplicação do teorema de Bayes. Esse fato evidencia que os registros observados agregam pouca informação à inferência de α na ausência de conhecimento prévio acerca do limite superior.

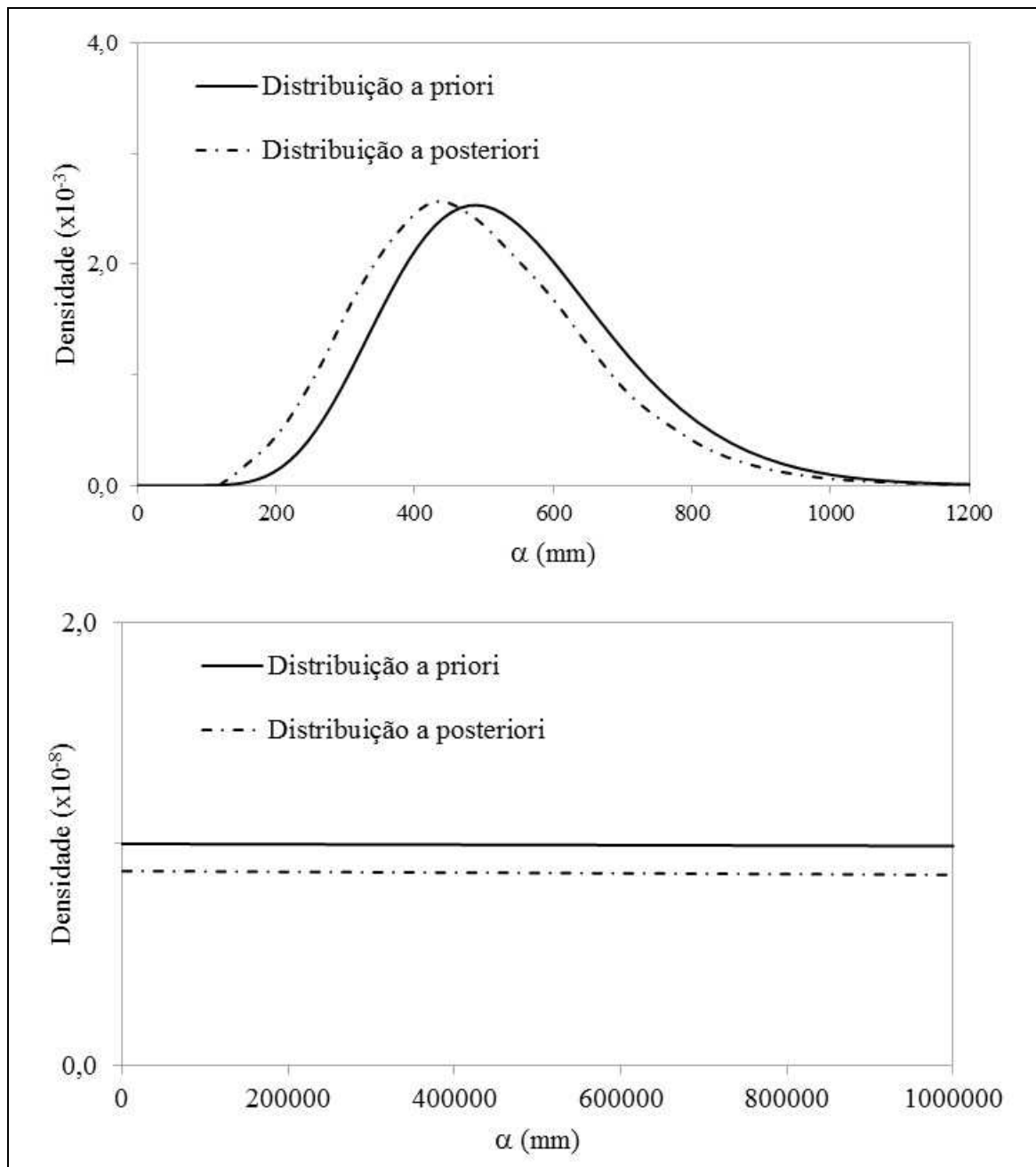


Figura 6.5 – Distribuição *a priori* e distribuição *a posteriori* do parâmetro α . (a) Modelo I; (b) Modelo II.

6.4.2 Modelo EV4

Tendo-se em vista que a interpretação física dos parâmetros σ e ξ não é trivial, são eliciadas distribuições *a priori* não informativas para os mesmos. Uma vez que ambos os parâmetros são sempre positivos, assumiram-se as seguintes distribuições: $\sigma \sim GAMA(1,0;1,0 \times 10^{-8})$ e $\xi \sim GAMA(1,0;1,0 \times 10^{-8})$.

Também para esse modelo distributivo foram consideradas duas distribuições *a priori* para o limite superior: a primeira, informativa, eliciada por $\sigma \sim GAMA(10.821;0,018)$ e

denotada por modelo I, e a segunda, não informativa, eliciada por $\alpha \sim GAMA(1,0;1,0 \times 10^{-8})$ e denotada por modelo II.

A amostra da distribuição *a posteriori* do modelo EV4 foi obtida de maneira similar àquela prescrita para o modelo LN4. A convergência à distribuição de equilíbrio ocorreu após aproximadamente 35.000 iterações. Já com relação ao *lag*, ficou estabelecido o valor 20. As amostras finais dos parâmetros não apresentaram heterogeneidades, tais como tendências ou alterações na variância.

Estatísticas a posteriori para o parâmetro σ

A Tabela 6.6 sintetiza as estatísticas *a posteriori* para o parâmetro σ . A Figura 6.6 apresenta as distribuições *a posteriori* para os dois modelos.

Tabela 6.6 – Estatísticas *a posteriori* para o parâmetro σ

Modelo	Média	DP	CV	95%HPD
I	9,495	3,416	0,360	(3,862;17,150)
II	$3,7 \times 10^6$	$2,6 \times 10^6$	0,703	$(4,2 \times 10^5; 1,0 \times 10^7)$

A influência da eliciação da distribuição *a priori* para o limite superior é clara na inferência da distribuição *a posteriori* do parâmetro de escala σ . De fato, para o modelo I, a distribuição *a posteriori* se concentra em uma faixa restrita do domínio do parâmetro, ao passo que, para o modelo II, a mesma se estende por várias ordens de grandeza na reta real. Além disso, no caso do modelo I, a distribuição *a posteriori* é aproximadamente simétrica, enquanto que, para o modelo II, a mesma é fortemente assimétrica à direita.

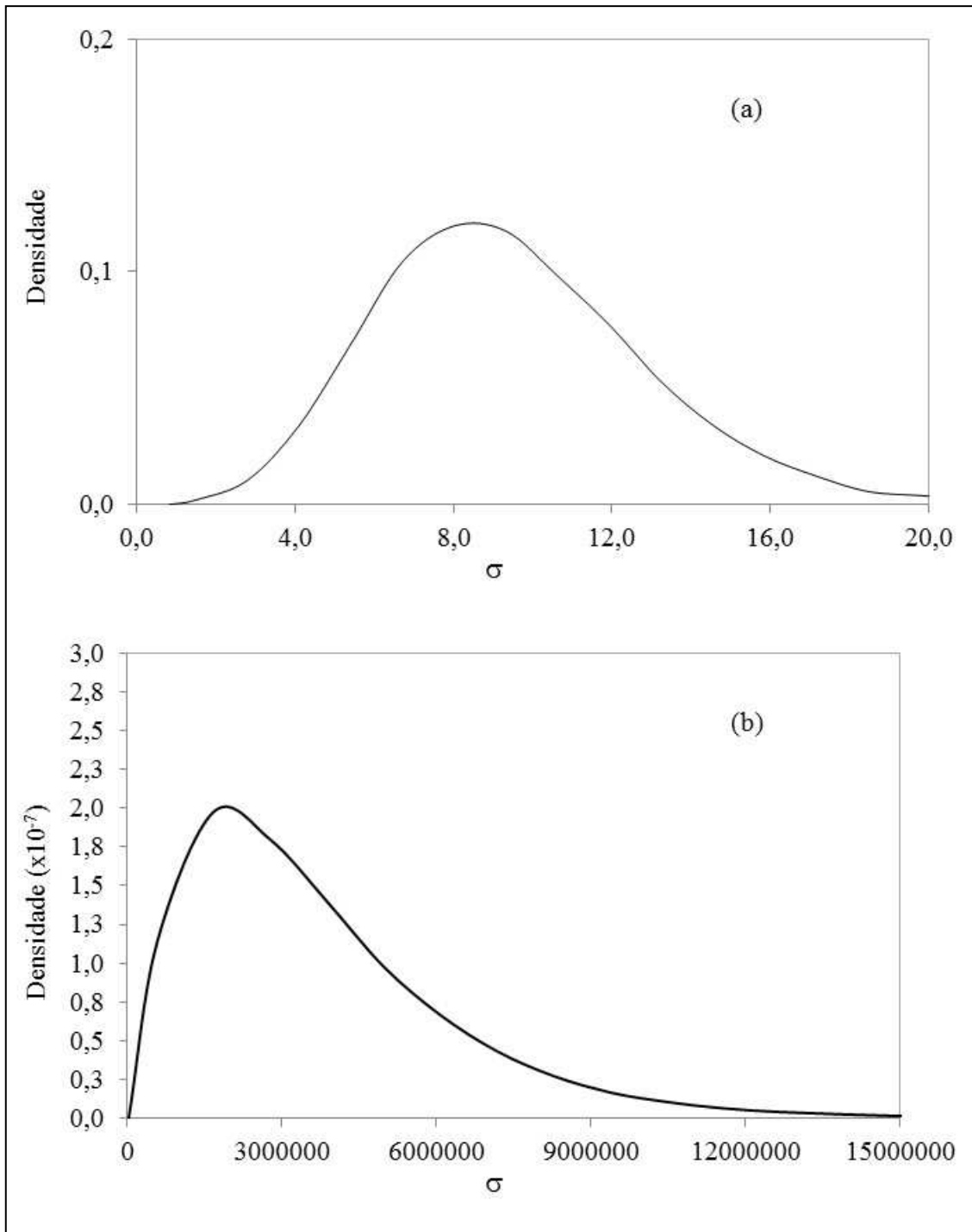


Figura 6.6 – Distribuições *a posteriori* para o parâmetro σ

Estatísticas a posteriori para o parâmetro ξ

A Tabela 6.7 sumaria as estatísticas *a posteriori* para o parâmetro ξ . A Figura 6.7 apresenta as distribuições *a posteriori* para os dois modelos.

Com relação ao parâmetro de forma ξ , percebe-se que o mesmo não é significativamente afetado pela eliciação da distribuição *a priori* do limite superior. As

médias e variâncias inferidas para os dois modelos são próximas, e as amplitudes dos intervalos de credibilidade também são similares. Em adição, ambas as distribuições *a posteriori* são unimodais e aproximadamente simétricas.

Tabela 6.7 – Estatísticas *a posteriori* para o parâmetro ξ

Modelo	Média	DP	CV	95%HPD
I	2,482	0,411	0,166	(1,720;3,324)
II	2,767	0,443	0,160	(1,940;3,674)

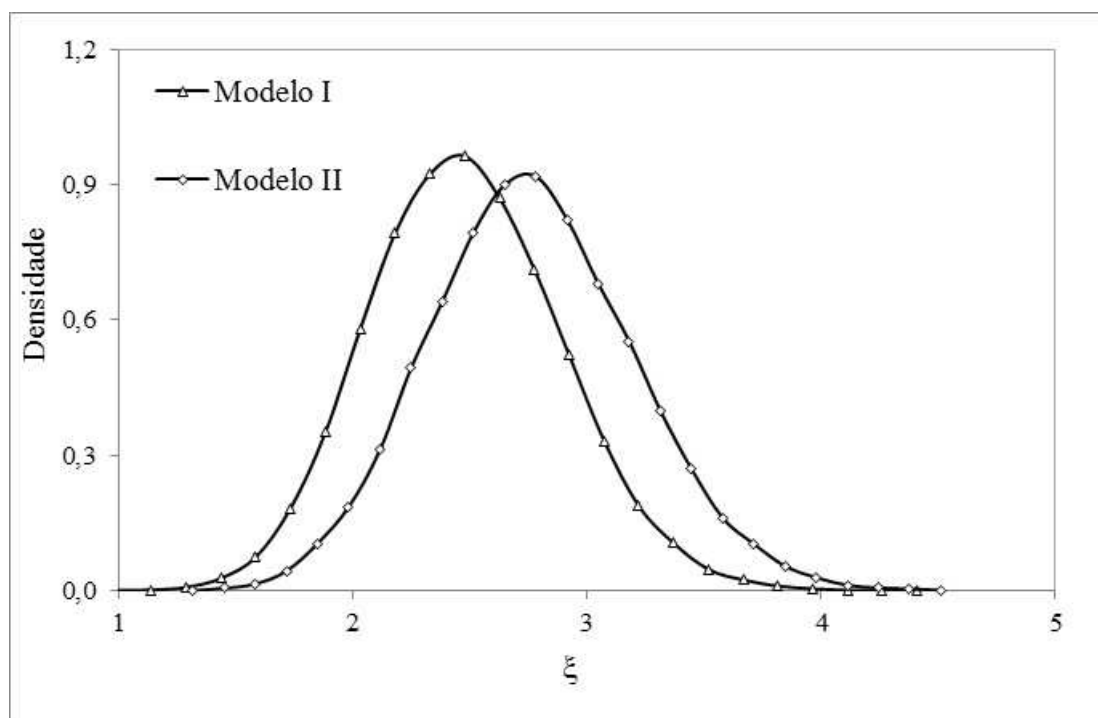


Figura 6.7 – Distribuições *a posteriori* para o parâmetro ξ

Estatísticas a posteriori para o parâmetro α

A Tabela 6.8 apresenta as estatísticas *a posteriori* para o limite superior. A Figura 6.8 apresenta as distribuições *a priori* e *a posteriori* para os dois modelos.

Tabela 6.8 – Estatísticas *a posteriori* para o parâmetro α

Modelo	Média	DP	CV	95%HPD
I	561,3	173,8	0,310	(267,7;941,8)
II	$2,0 \times 10^8$	$1,4 \times 10^8$	0,708	$(2,2 \times 10^7; 5,4 \times 10^8)$

Algumas particularidades podem ser verificadas na inferência desse parâmetro. A mais evidente delas é a diferença entre as estimativas do valor médio de α obtidas com os modelos LN4 e EV4. De fato, a estimativa oriunda do modelo EV4 é bastante superior àquela obtida com o modelo LN4. Mais do que isso, a distribuição *a posteriori* estimada

com o primeiro modelo se deslocou em sentido contrário aos dados, sugerindo que o conhecimento *a priori* exerce maior influência que esses quando da aplicação do teorema de Bayes, e atribuindo maior massa a valores elevados de α em comparação à distribuição *a priori*. Tendo-se em vista que a distribuição *a posteriori* é truncada no valor máximo observado e que não se verificam na amostra valores muito elevados, da ordem de grandeza da estimativa local de PMP, esse resultado não é intuitivo. Supõe-se que essa inconsistência esteja associada ao efeito combinado dos parâmetros do modelo distributivo no ajuste aos registros observados.

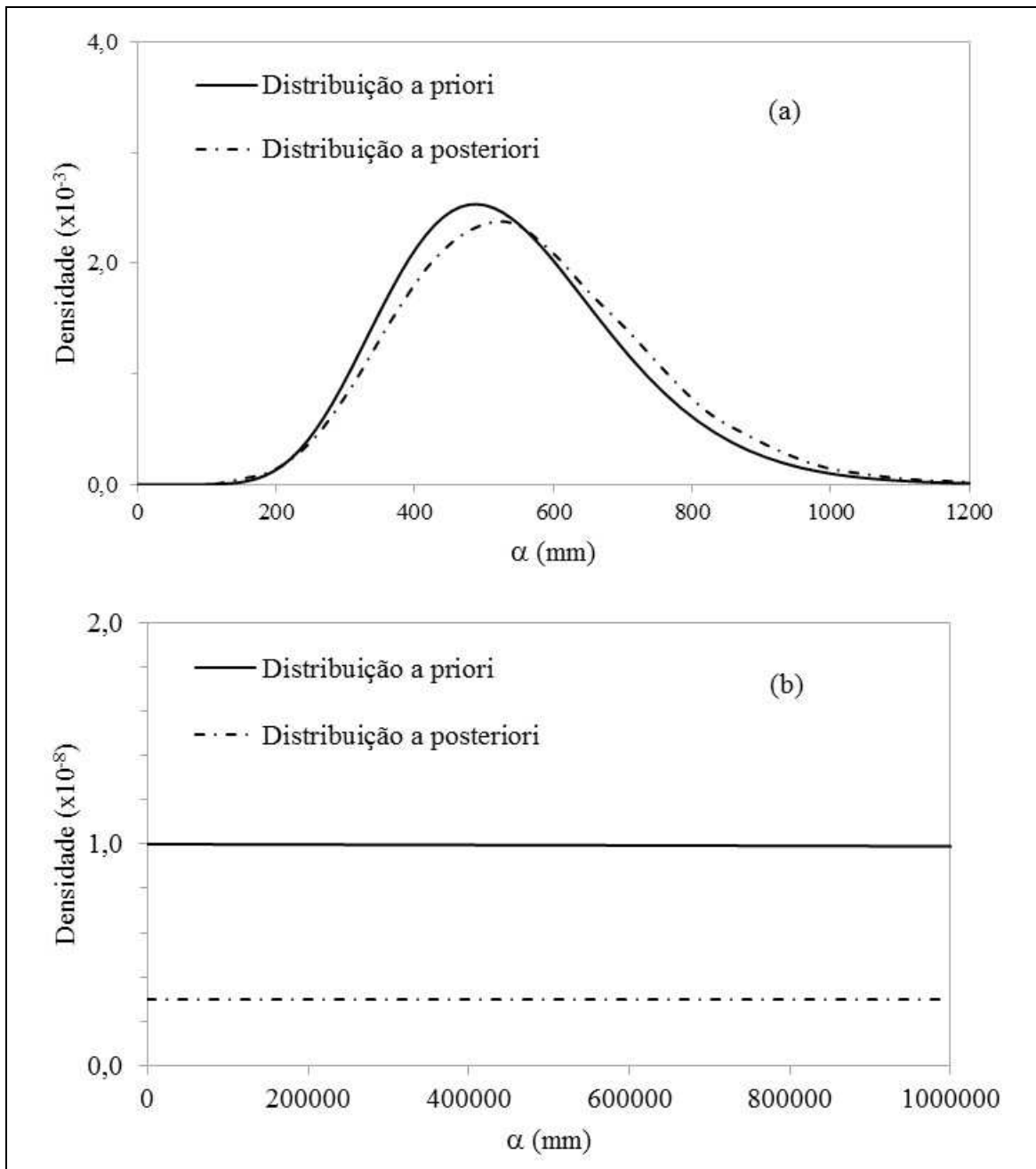


Figura 6.8 – Distribuição *a priori* e distribuição *a posteriori* do parâmetro α . (a) Modelo I; (b) Modelo II.

Já o modelo II evidencia mais uma vez que a adoção de distribuições *a priori* não informativas para o limite superior é uma prática não recomendada, uma vez que as estimativas desse parâmetro se mostram implausíveis (se não impossíveis, quando se considera a duração da chuva) em qualquer bacia hidrográfica terrestre.

6.5 Geração estocástica de precipitação diária

O procedimento de geração estocástica empregado aqui é similar àquele adotado na bacia do rio Pará. Primeiramente é estabelecido o limiar entre as chuvas convencionais e extremas, para os modelos LN4 e EV4. A seguir, é definido o tamanho das janelas de reamostragem do módulo não paramétrico do gerador. Em seguida, são geradas 1.000 séries de tamanho 35 anos e realiza-se a avaliação de desempenho do modelo misto. Por fim, é realizada uma comparação entre o gerador proposto e modelos estabelecidos na literatura, a saber, um modelo bipartido com uma distribuição gama ajustada aos dados, um modelo MPT, com o último estado modelado pela a distribuição generalizada de Pareto, e cujos parâmetros foram estimados por tentativa e erro, e um modelo não paramétrico, baseado na técnica de “*bootstrap*” convencional.

6.5.1 Definição do limiar entre chuvas convencionais e chuvas extremas

Modelo LN4

A Figura 6.9 apresenta as curvas de quantis para limiares variando entre 60 e 100 mm para o modelo LN4. É possível observar que os limiares que modelam adequadamente os máximos anuais em Folsom, entre 80 e 100 mm, são similares àqueles da bacia do rio Pará. A Figura 6.10, por sua vez, mostra a reprodução das médias mensais para os referidos limiares. Nessa figura, as barras em azul correspondem à média das séries simuladas e as barras em vermelho à série observada. A análise da Figura 6.10 fornece argumentos para a fixação do limiar em 80 mm, tendo-se em vista que tal valor é associado aos melhores ajustes na quase totalidade dos meses do ano.

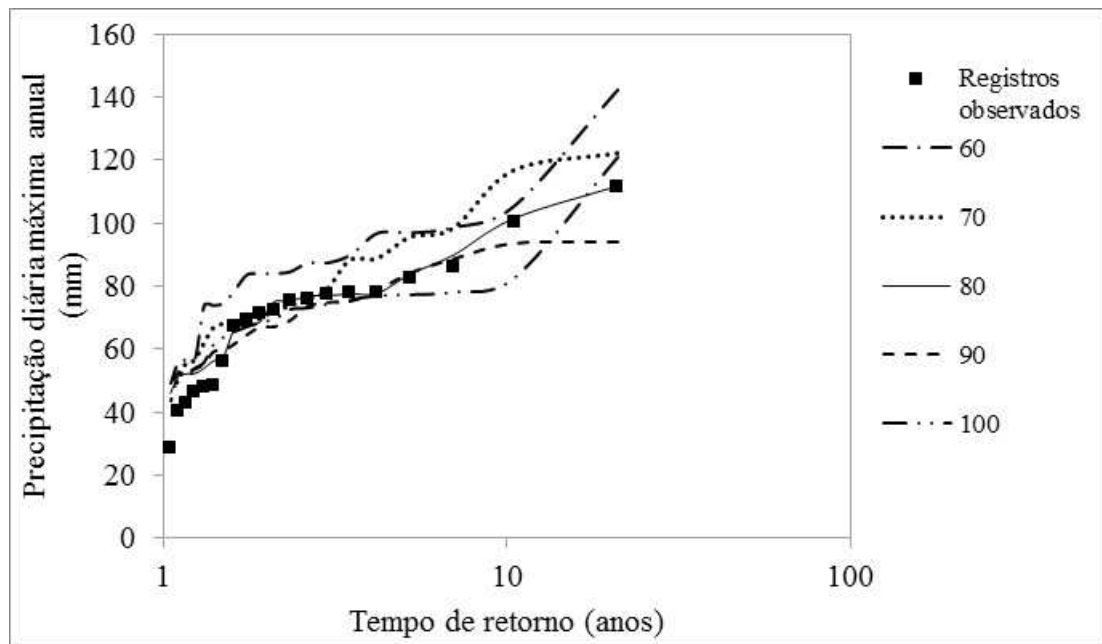


Figura 6.9 – Precipitação diária máxima anual gerada para diferentes limiares entre chuvas convencionais e extremas para o modelo LN4

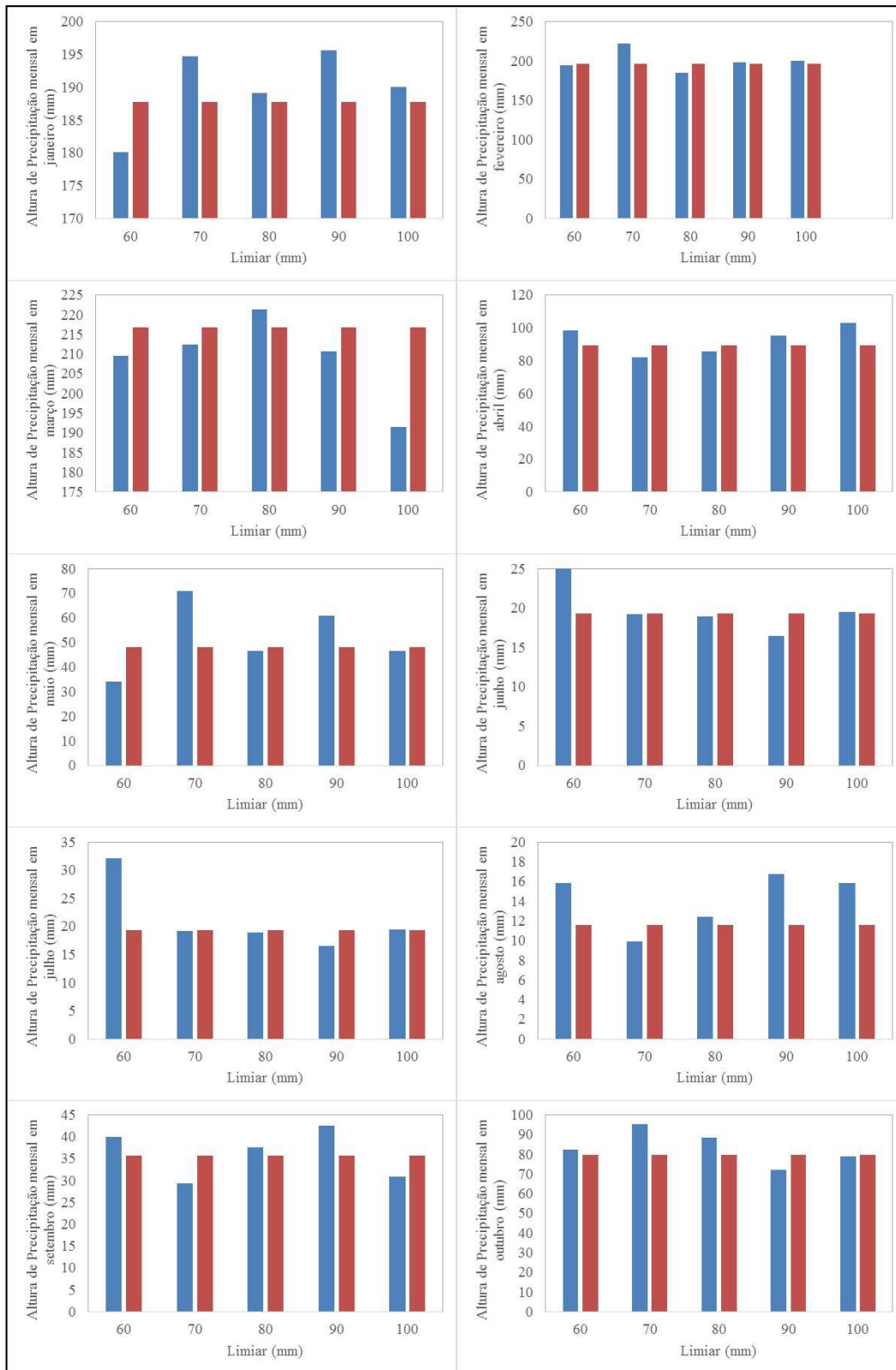


Figura 6.10 – Comparação entre a precipitação média mensal das séries simuladas com o modelo LN4 e da série observada para diferentes limiares entre chuvas convencionais e extremas

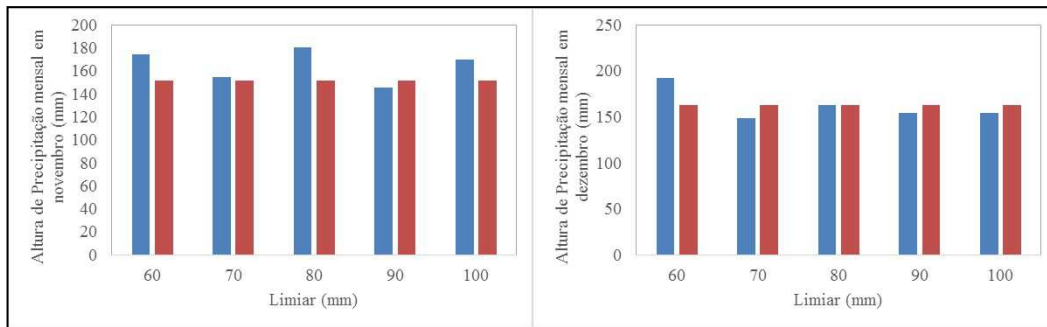


Figura 6.11 (cont.) – Comparação entre a precipitação média mensal das séries simuladas com o modelo LN4 e da série observada para diferentes limiares entre chuvas convencionais e extremas

Modelo EV4

Para o modelo EV4, somente se verificam ajustes aos máximos anuais quando limiares superiores a 90mm são adotados, como ilustra a Figura 6.11. Nesse caso, tal como na aplicação anterior, a probabilidade de geração de chuvas extremas se torna nula em meses da estação seca, o que é indesejável sob um ponto de vista de modelagem. Diante disso, fez-se a opção de fixar o limiar em 80mm, o qual permite a ocorrência de chuvas extremas em qualquer dia, e avaliar a capacidade preditiva do modelo EV4 sob tais circunstâncias.

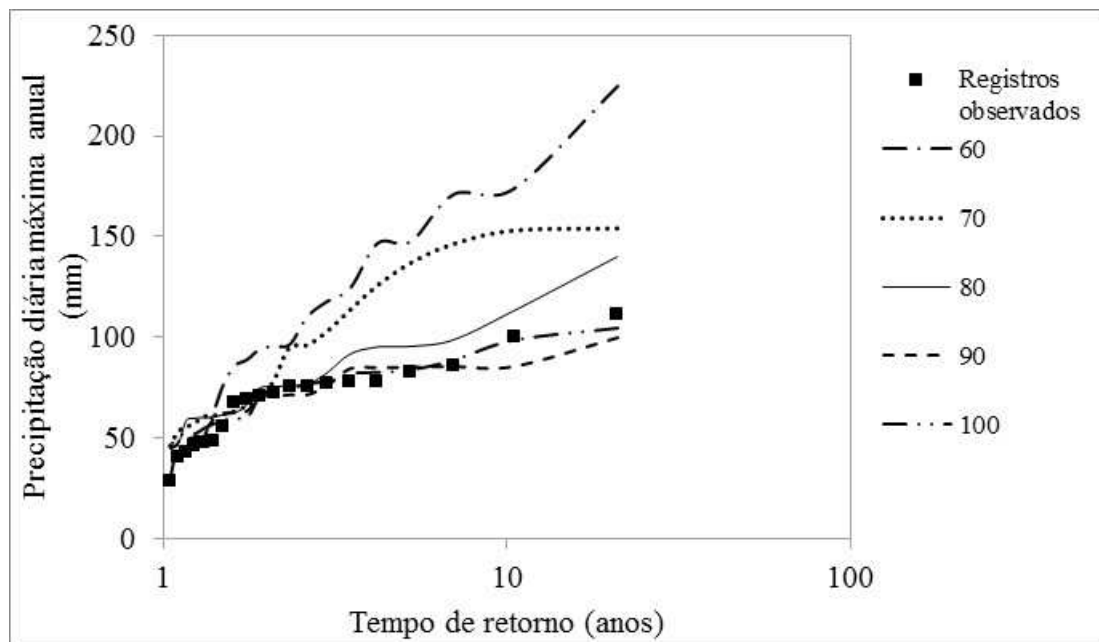


Figura 6.12 – Precipitação diária máxima anual gerada para diferentes limiares entre chuvas convencionais e extremas para o modelo EV4

Com relação às médias mensais, verifica-se na Figura 6.12 que o limiar adotado fornece ajustes razoáveis aos dados. Contudo, observa-se uma tendência de superestimação especialmente evidente nos meses entre abril e setembro, com exceção feita ao mês de

agosto. Tal fato decorre do comportamento da curva de quantis do modelo EV4, o qual resulta na geração de eventos de elevada magnitude com relativa frequência na estação seca.

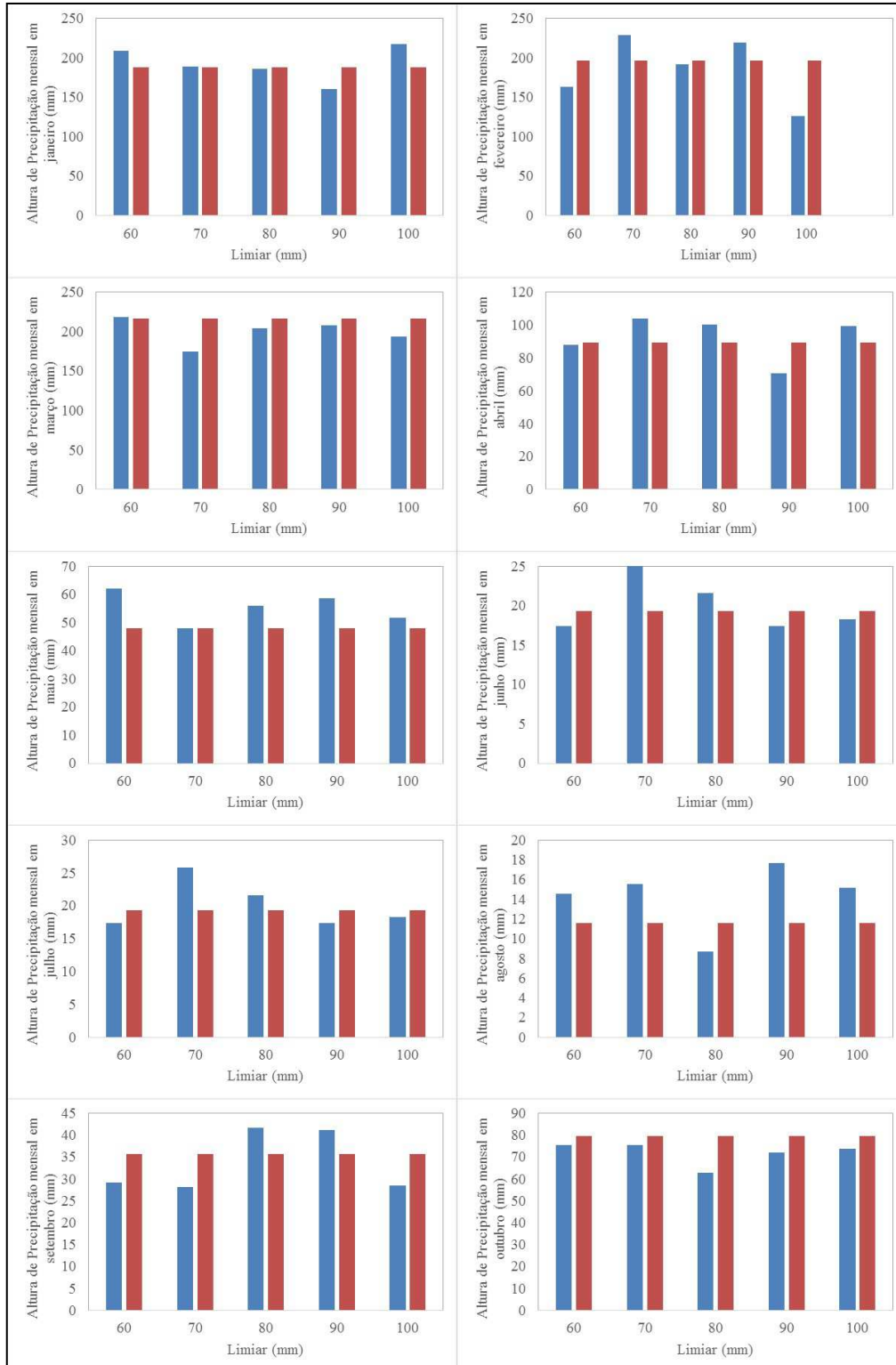


Figura 6.13 – Comparação entre a precipitação média mensal das séries simuladas com o modelo EV4 e da série observada para diferentes limiares entre chuvas convencionais e extremas

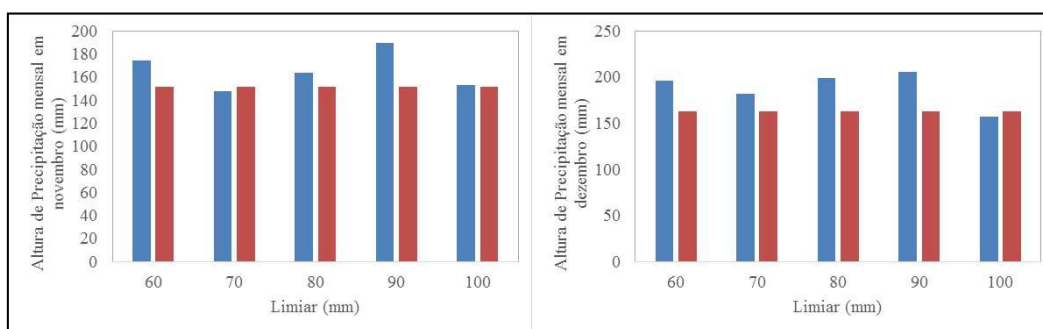


Figura 6.14 (cont.) – Comparação entre a precipitação média mensal das séries simuladas com o modelo EV4 e da série observada para diferentes limiares entre chuvas convencionais e extremas

6.5.2 Definição do tamanho das janelas de reamostragem

A Figura 6.13 resumia os resultados das simulações para janelas de reamostragem de 8, 14 e 28 dias. É possível observar que qualquer um desses intervalos se mostra apto a reproduzir a precipitação média diária, mas o intervalo de 8 dias não é efetivo na reprodução dos desvios padrão, indicando que o mesmo não captura a integralmente a variabilidade das alturas de chuva. Essa característica, contudo, é menos evidente nos dados de Folsom do que na aplicação anterior, o que pode estar relacionado ao tamanho consideravelmente maior da amostra na bacia do rio Pará.

Tendo-se em vista que a janela de 14 dias incorre em menor esforço computacional, a mesma foi selecionada para o prosseguimento das simulações.

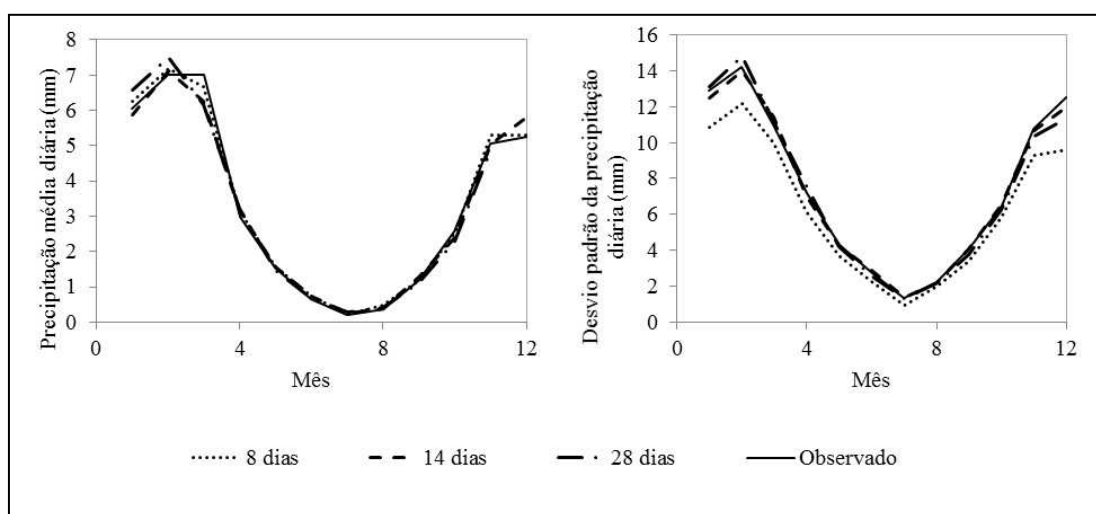


Figura 6.15 – Comparação entre estatísticas das séries geradas e observada para diferentes tamanhos de janela de reamostragem para o modelo LN4

6.5.3 Avaliação de desempenho do modelo misto de geração estocástica de precipitação diária

Estimados os parâmetros dos modelos LN4 e EV4 e definido o tamanho das janelas de reamostragem, pode-se proceder à geração estocástica de 1.000 séries de 35 anos de chuva diária, com propósito de avaliação de desempenho. Também aqui foram avaliadas as possibilidades de simulação do gerador misto e observou-se melhores resultados quando se consideram somente pares idênticos de estados de precipitação, provavelmente em decorrência do menor tamanho da amostra quando comparado àquele da aplicação anterior. Além disso, fez-se opção de introduzir perturbações nos valores oriundos da reamostragem, com intuito de se aumentar a variabilidade das alturas de chuva convencionais ao longo das simulações.

Avaliação de desempenho do modelo LN4

A Figura 6.14 sintetiza os resultados das simulações com o modelo LN4. É possível observar que o modelo é bastante efetivo na reprodução das médias e variâncias da precipitação diária ao longo do ano. Contudo, o mesmo apresentou restrições na reprodução dos coeficientes de assimetria observados na estação seca. Tal comportamento pode estar relacionado às probabilidades de geração de chuvas extremas extremamente reduzidas impostas pelo limiar selecionado, e se torna especialmente evidente quando séries pouco extensas são simuladas.

As alturas máximas de precipitação também foram razoavelmente simuladas. Exceções se manifestam nos meses de janeiro e junho. No primeiro caso, a altura máxima de precipitação foi superestimada em aproximadamente 35%. Esse fato pode ser explicado pela presença de um grande número de observações superiores a 80mm, que incorrem em grande probabilidade de geração de chuvas extremas nesse mês. Contudo, as observações se encontram contidas em uma faixa de variação restrita, com valores entre 80 e 90mm, e tal restrição não é acomodada na estrutura do gerador, o que possibilita a geração de alturas de precipitação bastante superiores àquelas observadas na amostra. Já no segundo caso, o modelo não se mostrou apto a simular uma altura de precipitação da ordem de 60mm, o que constitui um indício de que, em termos médios, a variabilidade amostral é insuficiente para que o módulo não paramétrico do gerador simule todo o espectro observado de alturas de chuva.

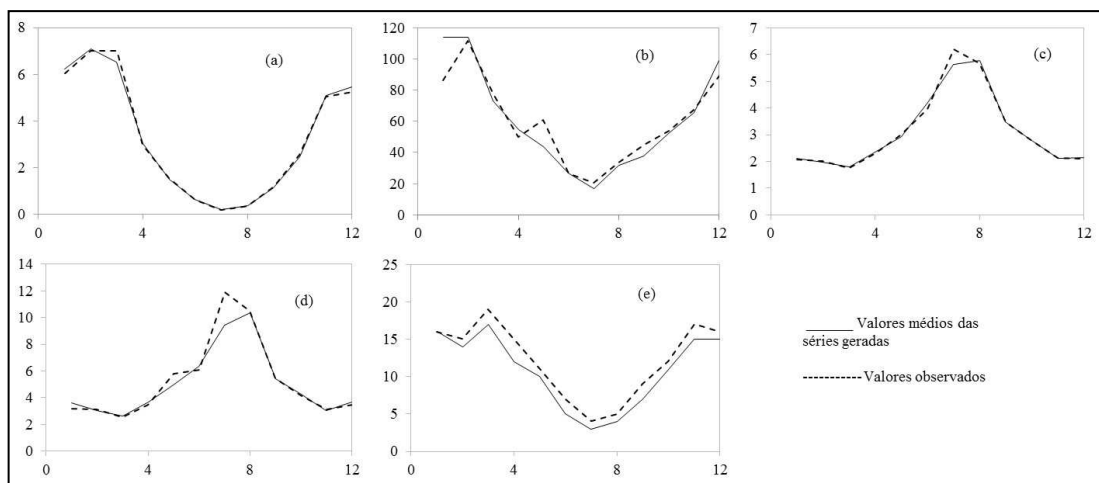


Figura 6.16 – Comparação das estatísticas das séries geradas com o modelo LN4 com a série observada para cada mês do ano. (a) chuva média diária; (b) altura máxima de chuva diária; (c) coeficiente de variação; (d) coeficiente de assimetria; (e) número médio de dias chuvosos. O eixo X representa os meses do ano.

Por fim, verifica-se na Figura 6.14 que o número médio de dias chuvosos por mês é consistentemente subestimado ao longo das simulações. Esse fato corrobora a hipótese levantada na aplicação descrita no capítulo anterior, a qual advoga que a construção de matrizes de probabilidade de transição em base diária é menos eficiente que aquela em escala mensal no que se refere a essa estatística. O uso de uma função que suavize as variações dia a dia, tal como a transformada de Fourier, sugerida por Racsko & Szeidl (1991), pode atenuar esse problema.

A Figura 6.15 mostra que as estatísticas mensais foram propriamente simuladas, o que evidencia as vantagens de se prescrever uma estrutura mista, na qual o módulo paramétrico se presta mais à extrapolação que à caracterização da variabilidade de estatísticas de longo termo.

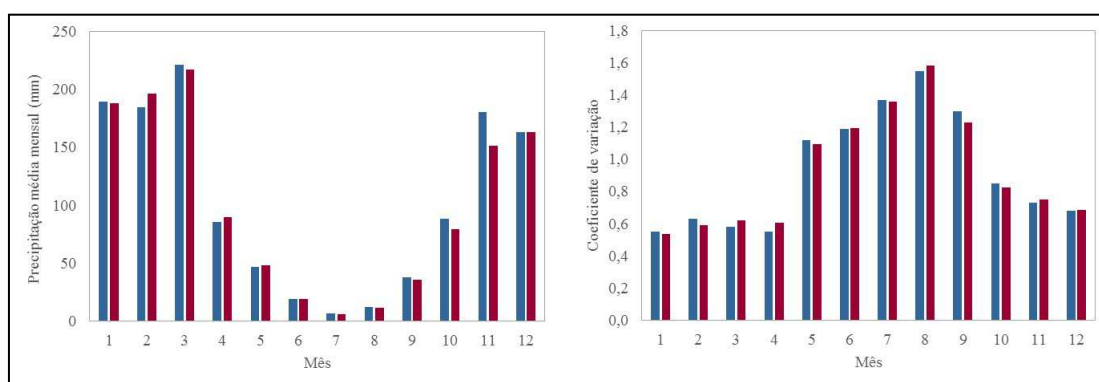


Figura 6.17 – Comparação entre as médias (a) e variâncias (b) mensais das séries simuladas com o modelo EV4 e a série observada

Por fim, a Figura 6.16 mostra a curva de quantis obtida com o modelo LN4. É possível observar que as precipitações máximas anuais são adequadamente reproduzidas para os mais variados tempos de retorno, indicando que o referido modelo distributivo oferece ajuste adequado ao conjunto de dados.

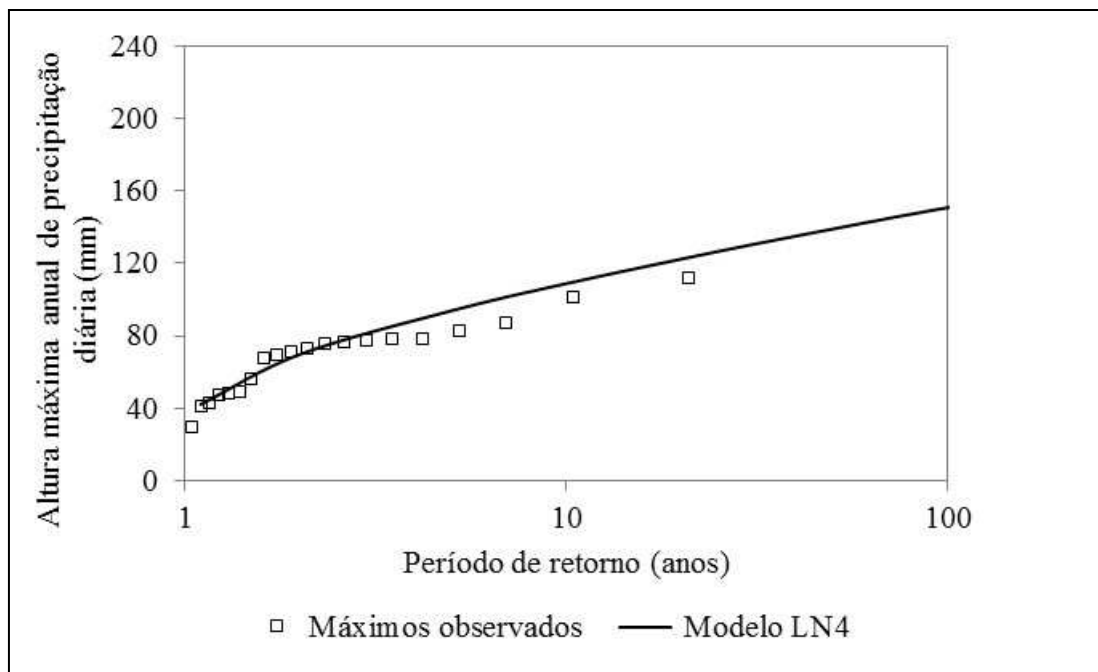


Figura 6.18 – Curva de quantis de precipitação diária obtida com o modelo LN4

Avaliação de desempenho do modelo EV4

A Figura 6.17 sumaria os resultados das simulações com o modelo EV4. Verifica-se que, nesse caso, o gerador apresentou limitações inclusive na reprodução de estatísticas como média, na estação chuvosa, e desvio padrão, na estação seca. Em adição, as alturas máximas de chuva simuladas somente são coerentes com aquelas observadas na estação seca, e as estimativas dessa estatística nos meses de janeiro e fevereiro são especialmente elevadas, em torno de 88% e 17%, respectivamente. Por fim, tal como o modelo LN4, o modelo EV4 é inapto a reproduzir aqueles coeficientes de assimetria amostrais mais elevados.

A Figura 6.18, por sua vez, mostra que as precipitações médias mensais são superestimadas ao longo de praticamente todo o ano, evidenciando que eventos de magnitude elevada são gerados com muita frequência. Além disso, as variâncias simuladas excederam aquelas observadas de maneira consistente, possivelmente em função da representação pouco acurada de eventos extremos.

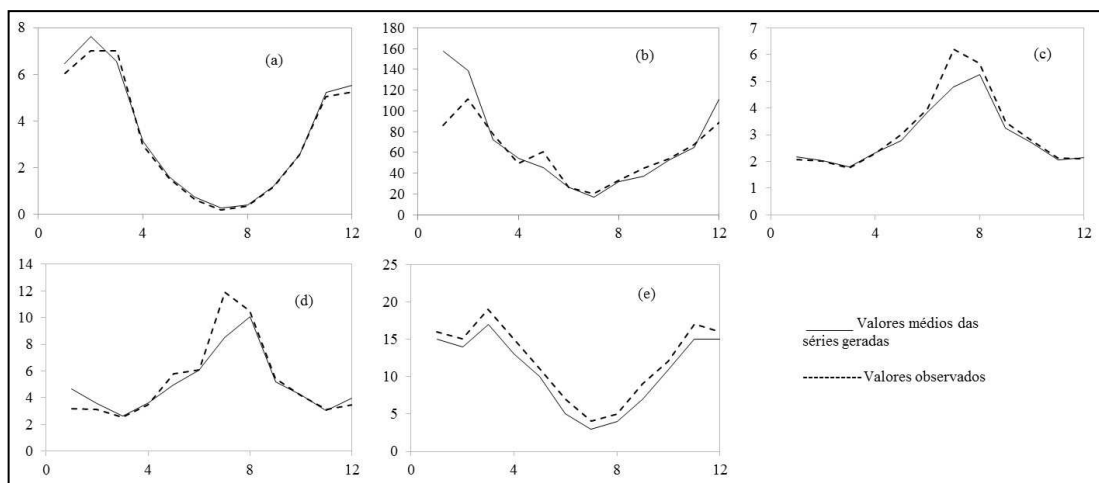


Figura 6.19 – Comparação das estatísticas das séries geradas com o modelo EV4 com a série observada para cada mês do ano. (a) chuva média diária; (b) altura máxima de chuva diária; (c) coeficiente de variação; (d) coeficiente de assimetria; (e) número médio de dias chuvosos. O eixo X representa os meses do ano.

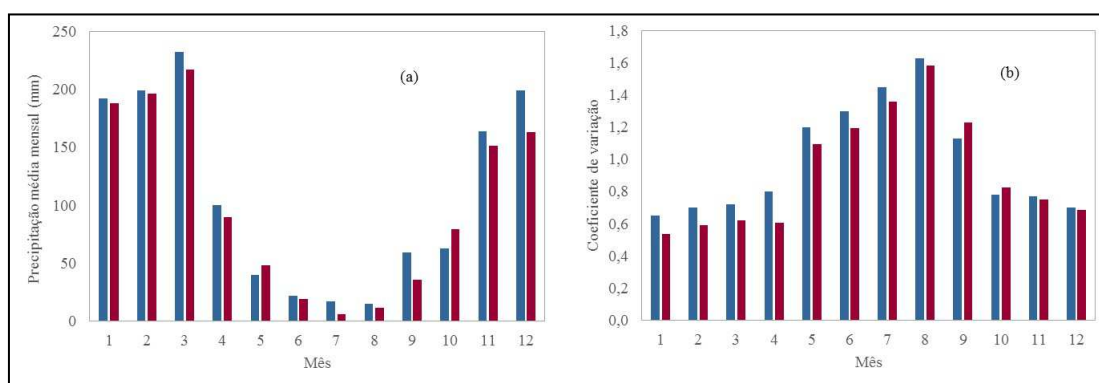


Figura 6.20 – Comparação entre as médias (a) e variâncias (b) mensais das séries simuladas com o modelo EV4 e a série observada

A Figura 6.19 apresenta a curva de quantis obtida com o modelo EV4. Percebe-se a ausência de aderência mesmo para tempos de retorno inferiores a 10 anos. Esses resultados sugerem que, ao menos para as condições hidrometeorológicas contempladas nesta tese, o referido modelo probabilístico não é adequado à modelagem de precipitações extremas. Evidentemente, amostras de 35 ou mesmo 61 anos constituem fontes de informação bastante restritas no que se refere à modelagem desse tipo de evento. Contudo, o comportamento verificado nas curvas de quantis das duas aplicações discutidas neste trabalho fornece indícios que, na ausência de valores extremamente elevados de precipitação, os quais influenciariam fortemente a estimativa do vetor paramétrico da distribuição, o modelo EV4 superestima consistentemente os quantis com menor probabilidade de superação, fato esse que restringe sua aplicação no contexto de estimação indireta de cheias.

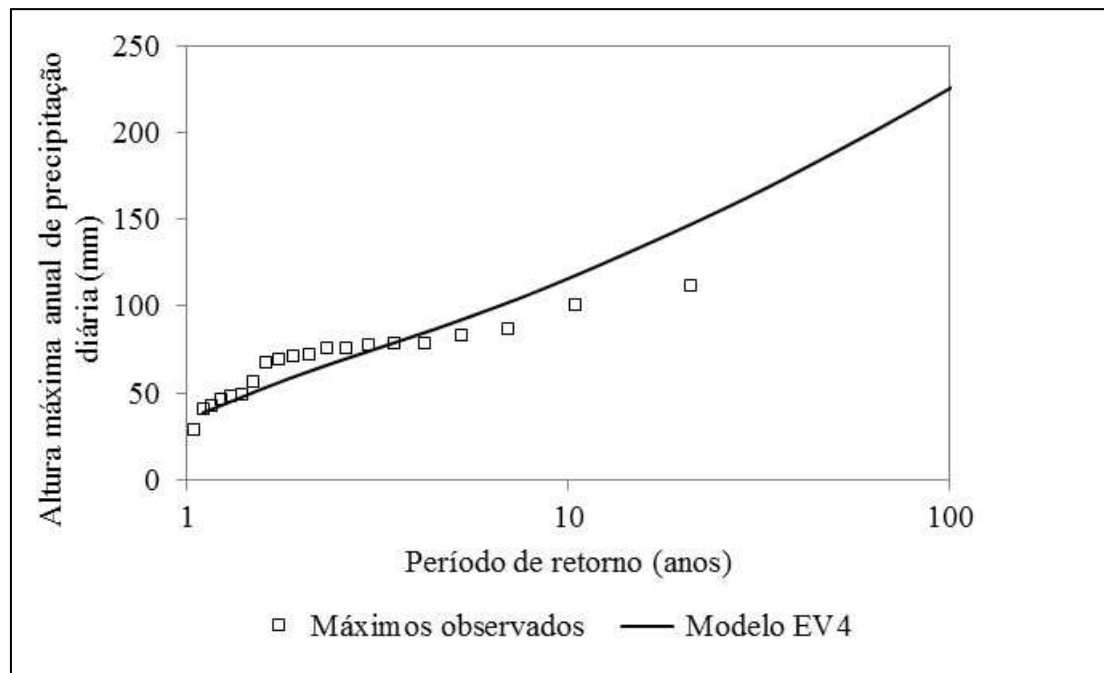


Figura 6.21 – Curva de quantis de precipitação diária obtida com o modelo EV4

6.5.4 Comparação com outros modelos de geração estocástica de precipitação diária

A Figura 6.20 apresenta as curvas de quantis para os diferentes modelos avaliados, considerando-se um tempo de retorno máximo de 10.000 anos. Tal com o exposto na aplicação anterior, a análise dessas curvas deve ser envolta em cuidados, dado que a ausência de valores extremamente elevados de alturas de precipitação faz com que qualquer modelo possa ser considerado como razoável para a extrapolação aos quantis mais raros. Ainda assim, pode-se estabelecer comentários gerais acerca das mesmas.

Primeiramente, percebe-se que os modelos com maior aderência aos dados, em especial para tempos de retorno superiores a 10 anos, são o modelo bipartido e o modelo MPT. Essa conclusão difere daquela obtida na bacia do rio Pará, para a qual esse último modelo ofereceu estimativas bastante enviesadas. Porém, quando se avalia os quantis associados a tempos de retorno superiores a 500 anos, verifica-se que o modelo MPT cresce de maneira acentuada, ao passo que o modelo bipartido evolui de maneira contrária. O modelo proposto, por outro lado, se situa em uma posição intermediária, crescendo de maneira suave em direção ao limite superior, mas ainda assim com quantis de valor mais elevado que aqueles obtidos com o a distribuição gama.

Além disso, é possível inferir a inaptidão de um modelo de reamostragem oriundo de uma amostra de tamanho reduzido para fins de extrapolação. De fato, o referido modelo

atribui ao máximo observado probabilidades de superação com ordens de grandeza de distinção, o que o torna inadequado para caracterizar a cauda superior das precipitações máximas anuais e restringe seu uso para fins de estimação indireta de quantis de cheias extremas.

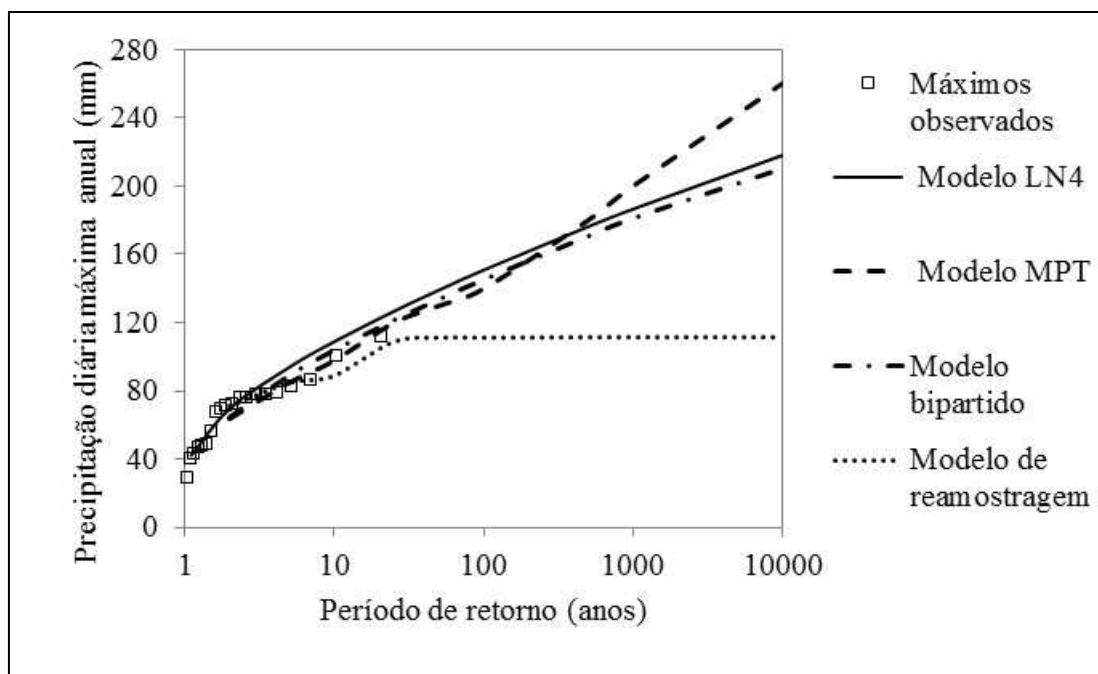


Figura 6.22 – Curvas de quantis para diferentes modelos de geração estocástica de precipitação diária

6.6 Calibração dos parâmetros do modelo Rio Grande

Em função das particularidades dos hidrogramas não regularizados avaliados da bacia do rio American, foi necessário estabelecer novos intervalos de variação dos parâmetros do modelo Rio Grande, em especial aqueles que controlam os trechos de recessão. Uma ampla gama de valores foi avaliada, tendo-se como base os limites estabelecidos nos trabalhos de Zhijia *et al.* (2013) e Lü *et al.* (2013). As Tabelas 6.9 e 6.10 sumarizam as distribuições *a priori* para os parâmetros do modelo hidrológico e do modelo probabilístico assumido como função de verossimilhança, respectivamente. Tais intervalos constituem distribuições uniformes.

Tabela 6.9 – Intervalos de variação dos parâmetros do modelo Rio Grande

Símbolo	Limite inferior	Limite superior
<i>k</i>	0,5	1
<i>imp</i>	0	0,03
<i>wum</i>	5	150
<i>wlm</i>	50	250
<i>wdm</i>	5	150
<i>sm</i>	40	120
<i>b</i>	0,1	1
<i>ex</i>	0,1	2
<i>c</i>	0,01	1
<i>kss</i>	0,05	0,35
<i>kg</i>	0,05	0,65
<i>ci</i>	0,4	0,99
<i>cg</i>	0,98	0,999

Tabela 6.10 – Intervalos de variação dos parâmetros da função de verossimilhança

Símbolo	Limite inferior	Limite superior
σ	0	100
σ_0	-2	1
σ_1	0	1
β	-0,9	1
ξ	0,1	10
ϕ_i	0	1
μ_h	-	-

De posse dos intervalos de variação dos parâmetros do modelo Rio Grande e da função de verossimilhança, fez-se necessário definir os critérios para a estimação da distribuição *a posteriori* dos mesmos a partir da calibração via algoritmo DREAM. A Tabela 6.11 apresenta os valores adotados para a busca da distribuição de equilíbrio.

Tabela 6.11 – Parâmetros para aplicação do algoritmo DREAM

Símbolo	Valor adotado
<i>nseq</i>	10
<i>ndraw</i>	5×10^5
<i>burn-in</i>	50%
<i>thin.t</i>	50
<i>Rthres</i>	1,01

Utilizando o código desenvolvido por Silva *et al.* (2014), procedeu-se à estimação da distribuição conjunta *a posteriori* das variáveis modeladas. Adotou-se como período de calibração os 5 anos hidrológicos compreendidos entre 01/10/1980 e 30/09/1985. Para a validação, considerou-se o período entre 01/10/1985 e 30/09/1988. Após a convergência do algoritmo, foram simuladas 1.000 realizações da distribuição conjunta *a posteriori* dos parâmetros, a serem empregadas na geração das séries de vazão. A Figura 6.21 apresenta uma comparação das distribuições marginais *a priori* e *a posteriori* dos parâmetros do modelo Rio Grande após a calibração.

Também aqui não cabe discutir a influência de cada parâmetro isoladamente na forma do hidrograma simulado. Contudo, algumas considerações podem ser feitas acerca da forma das distribuições *a posteriori* dessas variáveis. Primeiramente, verifica-se que as distribuições inferidas para os parâmetros k , imp , sm , c , kg , e cg são assimétricas à direita. Os limites de variação *a posteriori* dos parâmetros k , imp e c são relativamente próximos daqueles estabelecidos *a priori*. Já sm , kg e cg se distribuem em faixas restritas e são fortemente concentrados no limite inferior da distribuição *a priori*.

Por outro lado, os parâmetros wum , wlm , wdm e ci são bastante assimétricos à esquerda, e concentram-se nas proximidades do limite superior do intervalo de variação estabelecido *a priori*. Já as distribuições *a posteriori* de ex e kss apresentam assimetria à esquerda menos evidente e se estendem por todo o intervalo assumido antes da inferência. Com relação a ex , supõe-se que proximidade entre o valor modal e o limite superior do intervalo de variação seja uma maneira de se compensar o comportamento verificado para o parâmetro sm , tendo-se em vista que ambos controlam a produção de escoamento superficial.

Uma situação interessante foi verificada para o parâmetro b , uma vez que a distribuição *a posteriori* inferida para o mesmo é bimodal. Além disso, a referida distribuição se estende por quase todo o intervalo de variação estabelecido *a priori*.

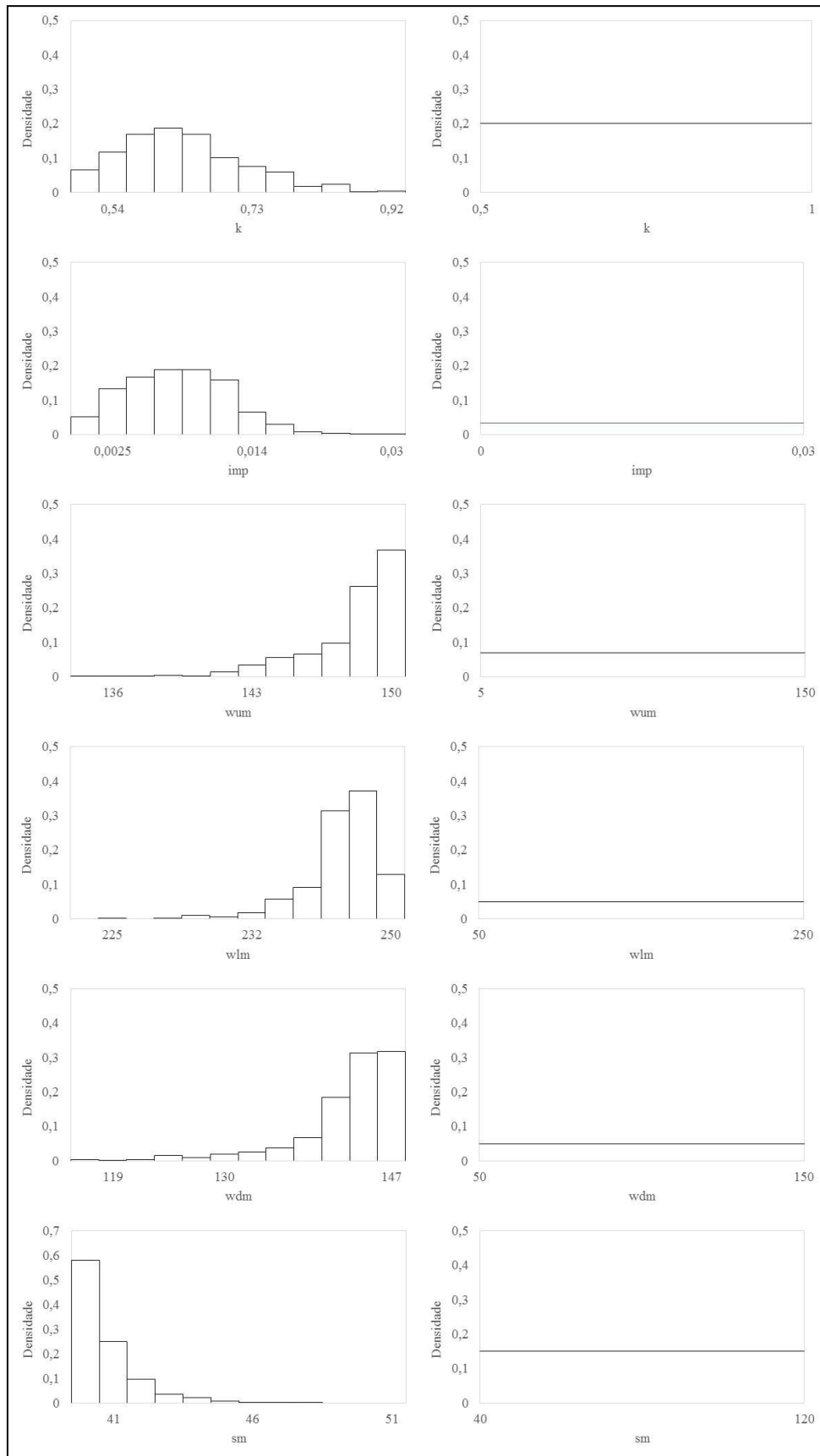


Figura 6.23 – Comparação das distribuições a priori e a posteriori dos parâmetros do modelo Rio Grande

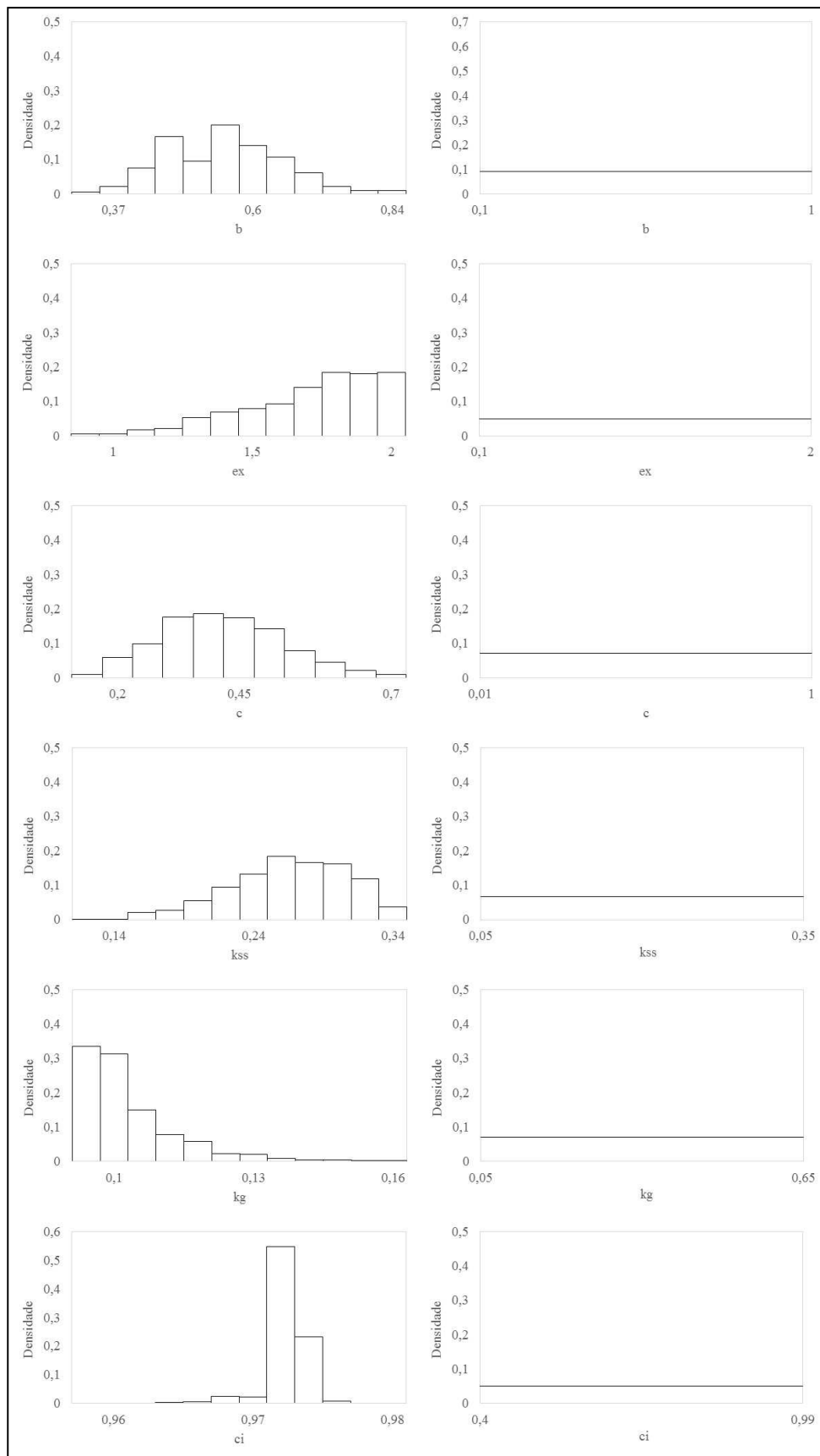


Figura 6.24 (cont.) – Comparação das distribuições a priori e a posteriori dos parâmetros do modelo Rio Grande

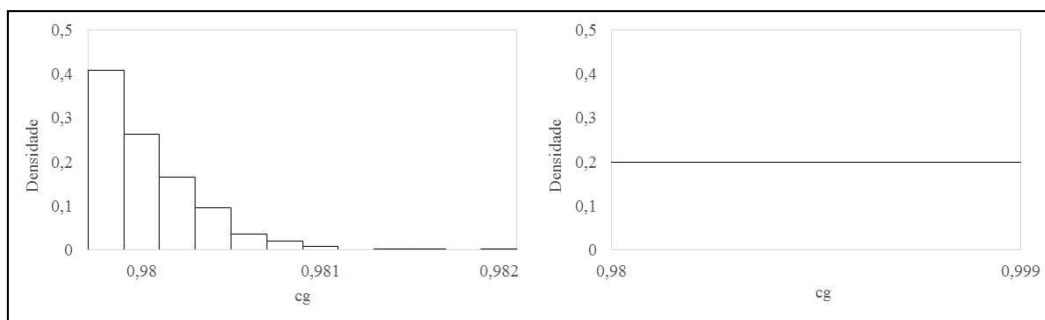


Figura 6.25 (cont.) – Comparação das distribuições a priori e a posteriori dos parâmetros do modelo Rio Grande

A Tabela 6.12 sumaria os índices de desempenho do período de calibração, calculados a partir da moda *a posteriori* dos parâmetros. Verifica-se a partir do coeficiente de Nash-Sutcliffe que o modelo Rio Grande explica aproximadamente 81% da variância natural das vazões na estação de Fair Oaks. Além disso, os picos de maior magnitude são razoavelmente reproduzidos, fato esse de grande importância para a estimação indireta de enchentes. Contudo, observa-se uma tendência de superestimação para os picos de menor valor.

Tabela 6.12 – Índices de desempenho do período de calibração

Índice de desempenho	Valor
Nash	0,814
RMSE	31,657
RV	1,207
Pearson	0,928

Com relação ao índice RV, o modelo apresentou uma tendência de superestimação, a qual pode ser explicada pela representação pouco acurada dos picos de menor magnitude. Por sua vez, a coerência temporal da série de vazões, inferida por meio do coeficiente de Pearson, foi adequadamente reproduzida. Essas considerações são ilustradas na Figura 6.22. É possível perceber, contudo, que o modelo Rio Grande introduz um viés de superestimação nas vazões em todo o espectro da simulação, tal como verificado na Figura 6.23.

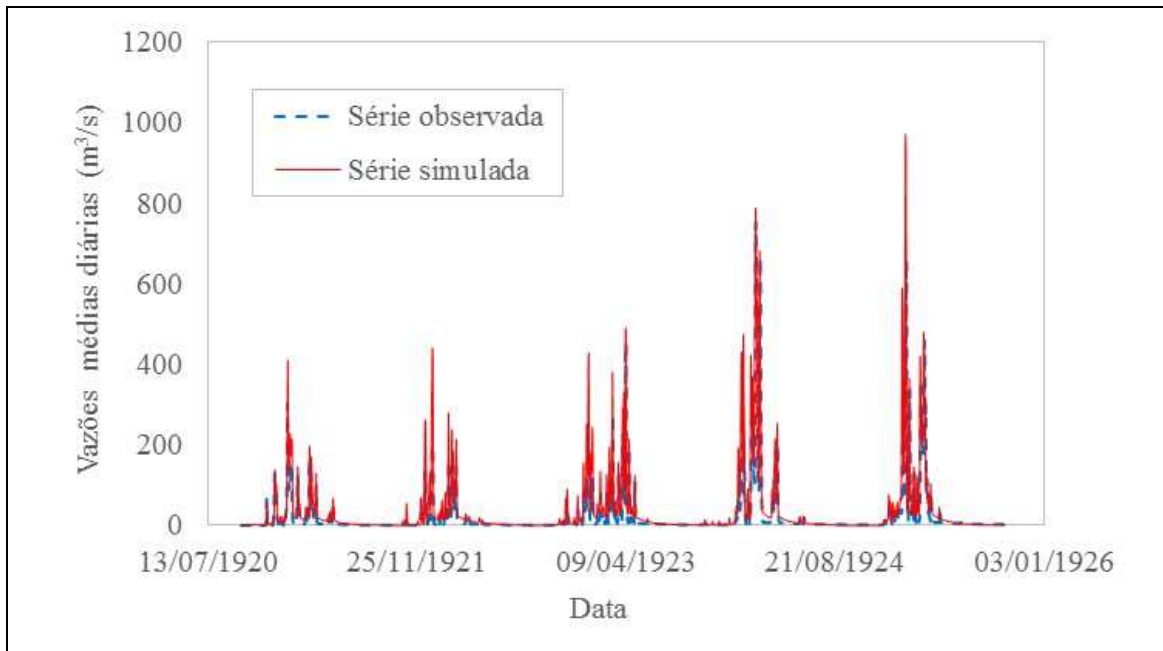


Figura 6.26 – Comparação entre os hidrogramas simulado e observado para o período de calibração

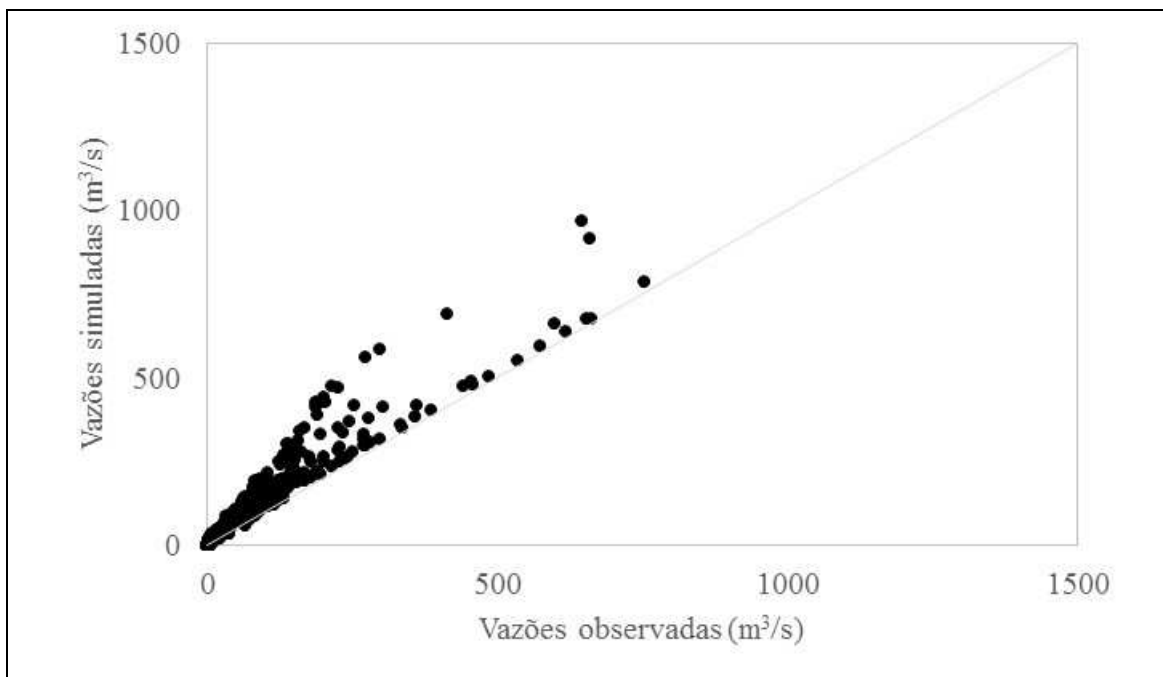


Figura 6.27 – Vazões simuladas versus vazões observadas

A Figura 6.24 ilustra o comportamento dos resíduos em relação às vazões observadas. Também aqui percebe-se que o uso da função de verossimilhança generalizada remove de maneira quase absoluta a heterocedasticidade dos erros de calibração. Contudo, nessa aplicação observam-se erros de maior valor em relação àqueles na bacia do rio Pará, especialmente para as vazões registradas no tramo de recessão do hidrograma, as quais não foram simuladas de maneira apropriada pelo modelo hidrológico. Por outro lado, os erros associados às maiores vazões são reduzidos, o que constitui um importante

aspecto para a simulação de cheias de magnitude quando da construção das curvas de quantis.

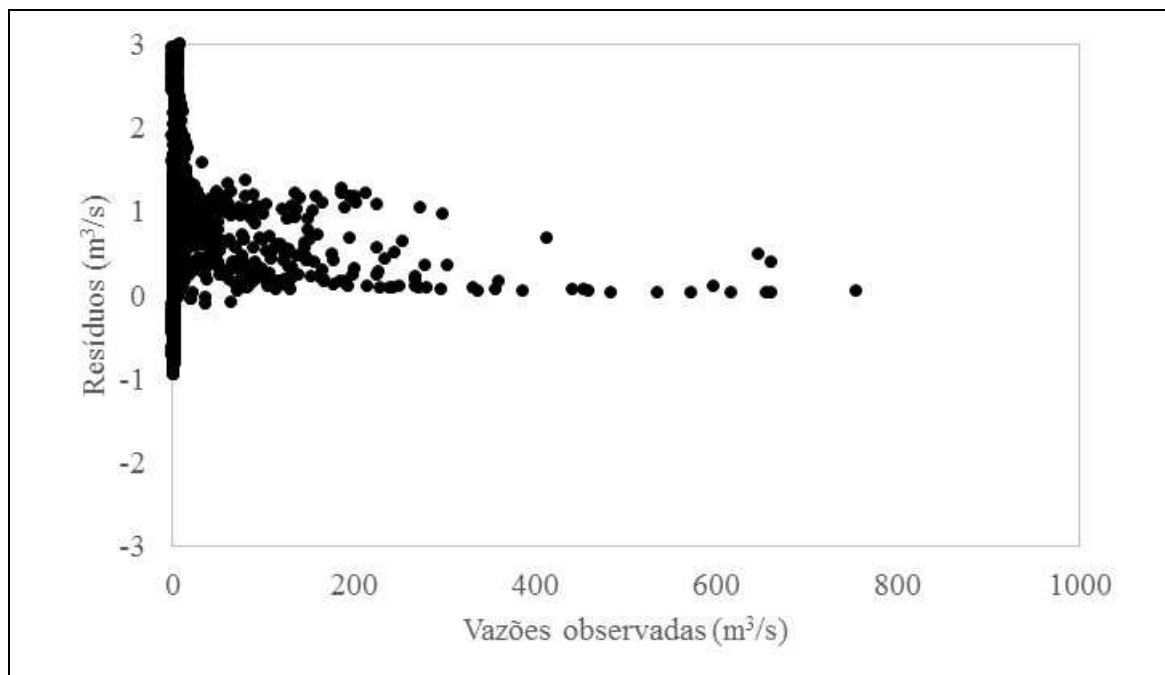


Figura 6.28 – Resíduos de calibração com a função de verossimilhança generalizada *versus* vazões observadas

A Figura 6.25 mostra a autocorrelação dos resíduos. Percebe-se que a mesma foi quase integralmente removida com o uso da função de verossimilhança generalizada, ainda que verifique pequenos valores dessa estatística para o *lag* 1. A Figura 6.26, por sua vez, ilustra a distribuição empírica dos resíduos com relação ao modelo teórico proposto para os mesmos. Percebe-se que uma distribuição empírica ainda mais assimétrica que aquela observada na bacia do rio Pará é razoavelmente modelada pela distribuição assumida para os resíduos.

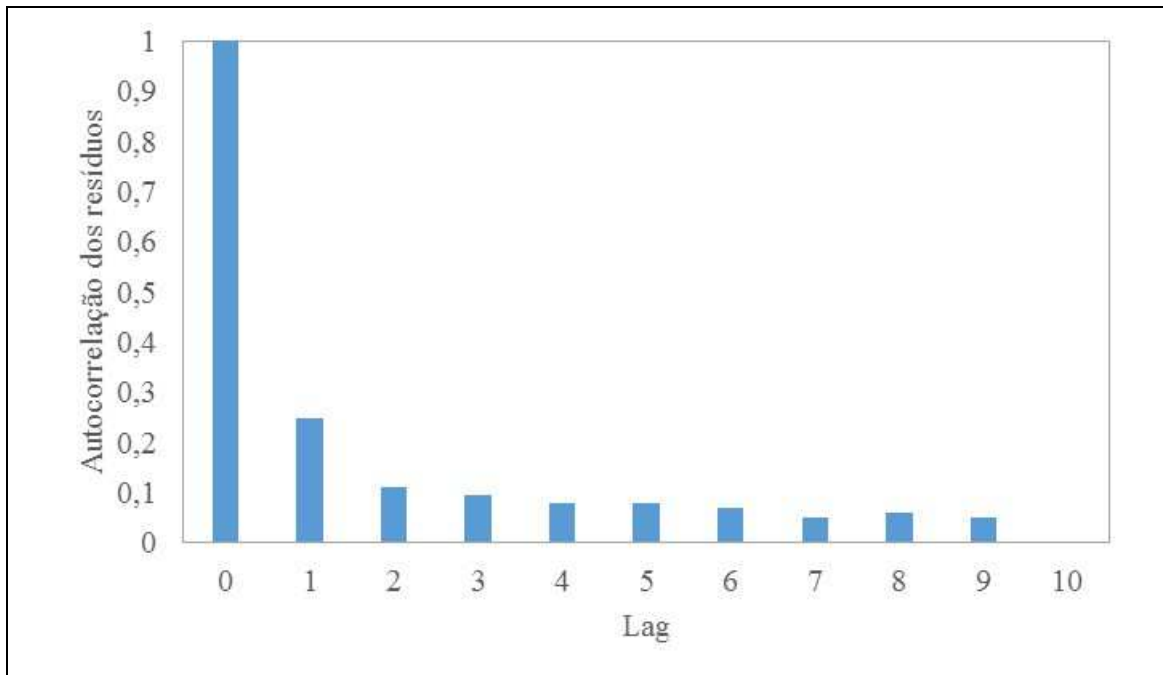


Figura 6.29 – Autocorrelação dos resíduos de calibração

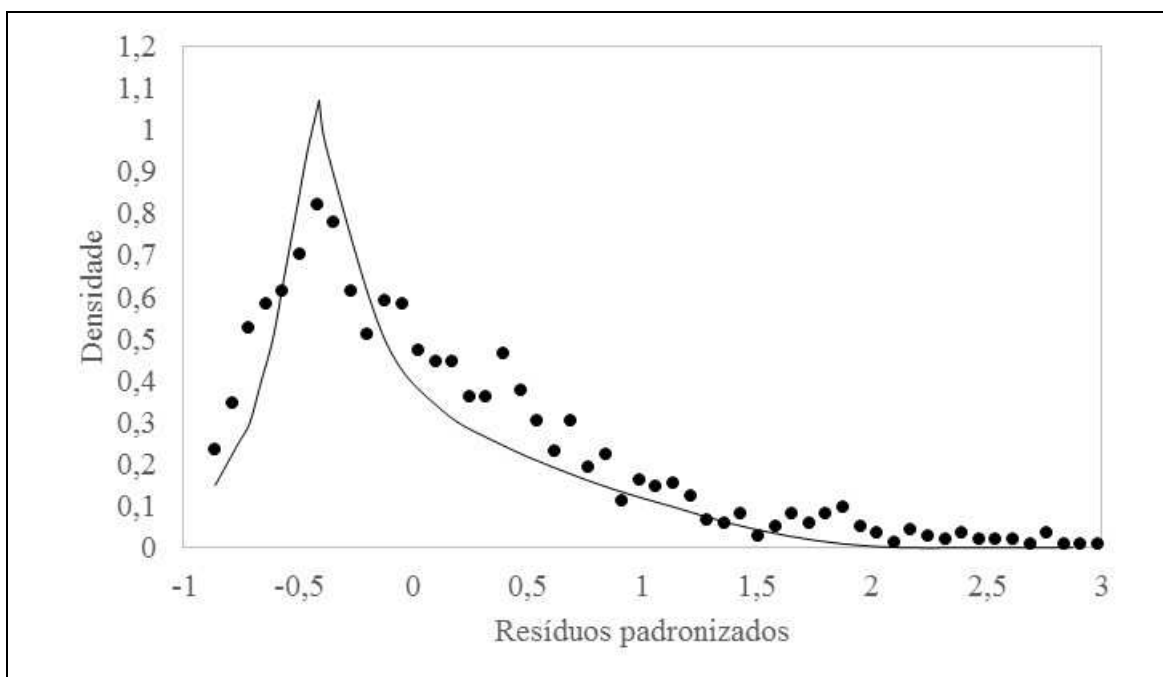


Figura 6.30 – Distribuições empírica (círculos) e teórica (linha contínua) dos resíduos de calibração padronizados

A Tabela 6.13 sumaria os índices de desempenho para o período de validação. É possível observar que o ajuste é ligeiramente inferior nessa situação. Ainda assim, o modelo Rio Grande foi capaz de explicar em torno de 75% da variância natural das vazões no período avaliado. Além disso, a coerência temporal das vazões, indicada pelo coeficiente de Pearson, é razoavelmente respeitada, ainda que se observe na Figura 6.27

um *lag* nos picos das vazões simuladas em relação às observadas. Por fim, percebe-se uma tendência de superestimação dos picos de maior magnitude.

Tabela 6.13 – Índices de desempenho para o período de validação

Índice de desempenho	Valor
Nash	0,750
RMSE	40,489
RV	1,238
Pearson	0,886

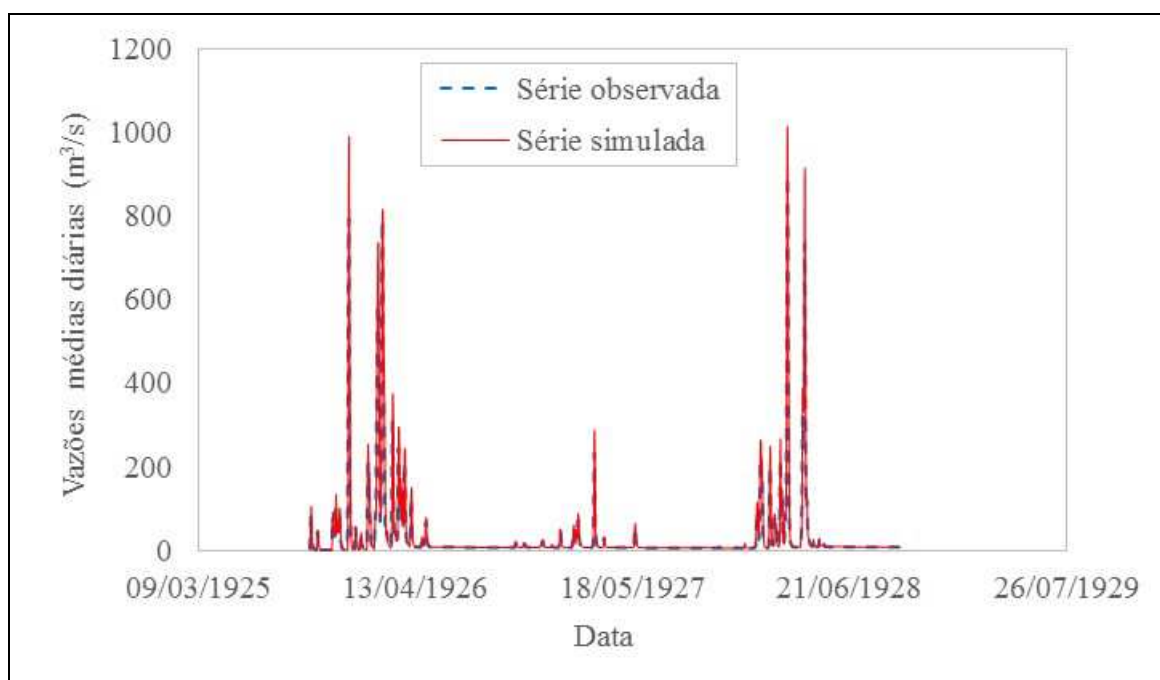


Figura 6.31 – Comparação entre os hidrogramas simulado e observado para o período de validação

6.7 Construção das curvas de quantis de enchentes

A Tabela 6.14 apresenta um resumo das estimativas de quantis para tempos de retorno diversos compreendidos entre 2 e 10.000 anos. A Figura 5.30, por sua vez, ilustra a comparação entre as enchentes simuladas e observadas para os referidos tempos de retorno. Os valores observados são representados por círculos. A probabilidade empírica de excedência foi calculada com auxílio da posição de plotagem de Cunnane.

Tabela 6.14 – Estimativa de quantis de enchentes para diversos tempos de retorno

Tempo de retorno (anos)	Vazão (m ³ /s)	95% HPD
2	1373,7	(1088,0;1713,5)
5	2927,2	(2436,4,6;3555,8)
10	4275,4	(3570,7;5178,1)
25	6282,3	(5266,1;7629,5)
50	7944,8	(6621,4;9611,8)
100	9697,0	(8087,7;11747,0)
250	11504,3	(9519,5;13919,0)
500	13911,9	(11560,7;17071,5)
1000	15704,9	(12934,3;19434,9)
2500	17445,8	(14263,9;21936,9)
5000	19653,5	(15775,2;25279,2)
7500	20593,8	(16889,9;26929,9)
10000	21249,1	(17765,1;27951,7)

95% HPD – Intervalo de credibilidade

A análise da Figura 6.28 evidencia que, também nessa aplicação, há pouca aderência entre as cheias simuladas e observadas para quantis associados a tempos de retorno compreendidos entre 10 e 50 anos. Contudo, nesse caso, as enchentes observadas estão contidas nos limites do intervalo de credibilidade de 95%. Além disso, percebe-se que a curva de quantis obtida com o método proposto fornece estimativas bastante próximas àquelas oriundas de dados não sistemáticos, o que mostra a adequação do mesmo na simulação de enchentes com magnitudes extremamente elevadas, ainda que a síntese hidrológica do modelo Rio Grande forneça resultados inconsistentes em algumas faixas de probabilidade de superação.

Na Figura 6.29 constam as curvas de quantis de enchentes obtidas por diversos métodos de simulação, os quais compreendem:

- O estudo de USBR (2002), o qual envolveu o ajuste da distribuição Log-Pearson III (LPIII) com assimetria positiva (e, portanto, ilimitada) aos dados sistemáticos e não sistemáticos da estação de Fair Oaks. O procedimento é baseado no paradigma Bayesiano e contemplou a inclusão de incertezas relacionadas aos erros de medição de vazões no período de observação sistemática, erros na estimação das paleocheias e sua data de ocorrência, e erros relacionado ao vetor de parâmetros do modelo distributivo;

- Análise de frequência convencional (AFC), empregando-se a distribuição GEV, com parâmetros estimados sob o paradigma Bayesiano (Fernandes, 2009) e função de verossimilhança construída de maneira análoga àquela empregada nos modelos superiormente limitados pelo referido autor; e
- Análise de frequência Bayesiana (AFB), empregando-se a distribuição LN4 e a estimativa local de EMP como estimador do limite superior (FERNANDES *et al.*, 2010).

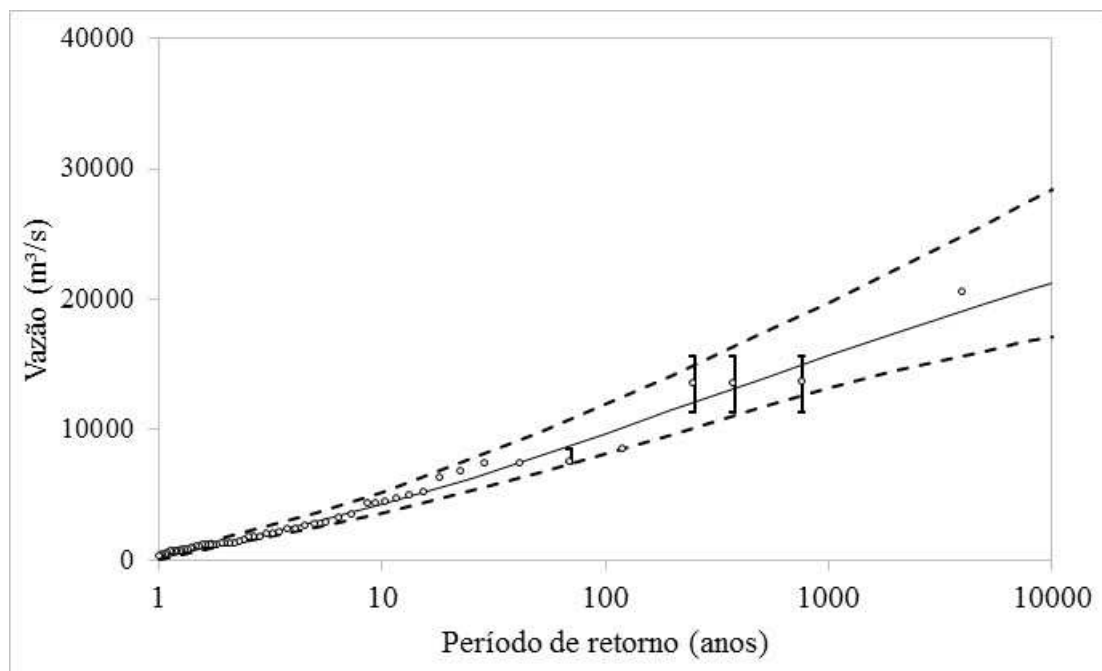


Figura 6.32 – Curva de quantis de enchentes na bacia do rio American, em Folsom. As linhas tracejadas correspondem ao intervalo de credibilidade de 95%.

Tendo-se em vista que as hipóteses inerentes à construção desses modelos são distintas, os mesmos não são diretamente comparáveis. Contudo, comentários gerais podem ser feitos acerca das curvas de quantis obtidas. Em primeiro lugar, pode-se observar que os modelos GEV e LPIII subestimam de maneira consistente os quantis de vazão com menor tempo de retorno. No entanto, o modelo LPIII fornece um ajuste razoável às paleocheias, em especial aquelas com menor probabilidade de superação. O modelo GEV, por outro lado, perde aderência para quantis associados a tempo de retorno superiores a aproximadamente 300 anos, verificando-se forte tendência de superestimação dos mesmos.

O modelo proposto nesta tese, por sua vez, apresentou resultados muito similares àqueles obtidos por Fernandes *et al.* (2010) com o modelo LN4. De fato, os quantis

estimados com esses dois métodos são equivalentes para tempos de retorno de até 500 anos, e a partir desse ponto verifica-se que os quantis provenientes da estimação indireta são ligeiramente inferiores aos da análise de frequência Bayesiana. Contudo, o modelo proposto por Fernandes *et al.* (2010) faz uso de informações não-sistemáticas na construção da função de verossimilhança, o que agrega à inferência maior conhecimento acerca da cauda superior das enchentes. Esse fato evidencia a qualidade do método proposto, uma vez que o mesmo emprega somente uma amostra reduzida de dados sistemáticos de vazões e ainda assim se mostrou apto a reproduzir cheias com reduzida probabilidade de superação.

Outro aspecto a ser mencionado com relação aos modelos avaliados diz respeito à amplitude dos intervalos de credibilidade. De fato, como ilustra a Figura 6.30 (a) e (b), os intervalos de credibilidade estimados com o método proposto são significativamente mais curtos que aqueles associados aos modelos LPIII e GEV. Em adição, a amplitude do intervalo de credibilidade é também menor que a do modelo LN4 (Figura 6.30 (c)), ainda que em menor proporção quando comparada aos modelos ilimitados. Tal situação, contudo, pode decorrer da variabilidade pouco expressiva das chuvas geradas e dos conjuntos de parâmetros inferidos da distribuição *a posteriori*, e, assim, não constituir necessariamente uma vantagem do método proposto frente à análise de frequência Bayesiana empregada por Fernandes *et al.* (2010).

Essa análise obviamente não fornece argumentos definitivos para a escolha de qualquer um desses modelos. Contudo, em função da estrutura lógica de análise imposta ao modelo proposto e ao modelo limitado de Fernandes *et al.* (2010), do bom ajuste verificado em todas as faixas de tempo de retorno e também dos intervalos de credibilidade de menor amplitude, é razoável supor que os mesmos têm melhor capacidade preditiva, fornecendo assim estimativas mais confiáveis de quantis de cheias extremas.

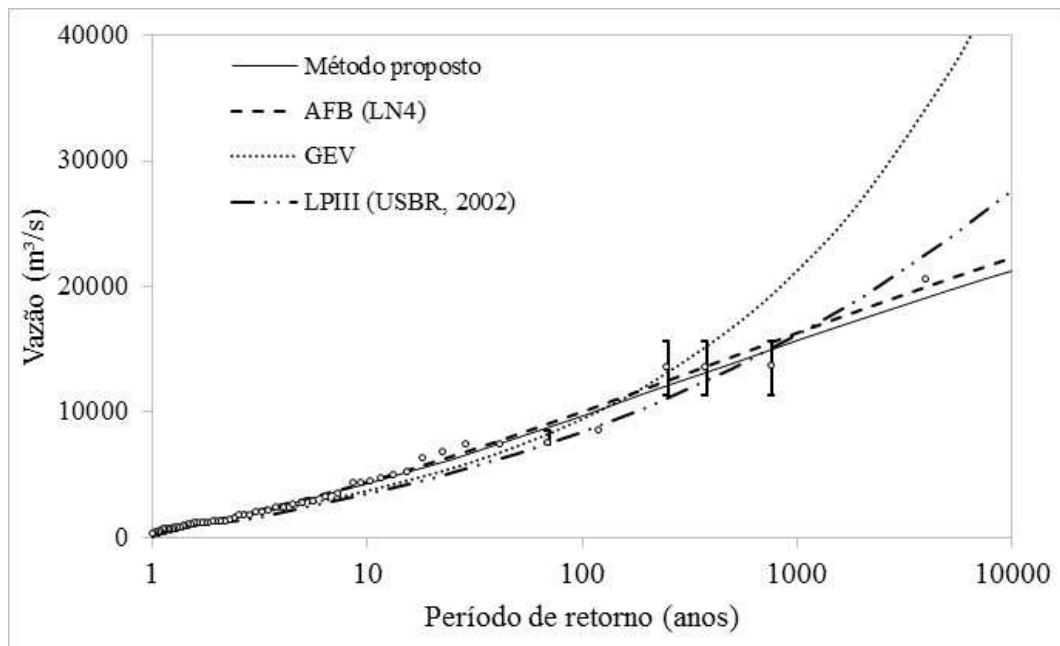


Figura 6.33 – Comparação das curvas de quantis de enchentes obtidas com diferentes métodos

A aplicação do método à bacia do rio American mostrou sua efetividade na simulação de cheias com probabilidades de superação muito reduzidas. Esse fato torna o mesmo especialmente interessante quando não se dispõe de dados não-sistemáticos de vazão, os quais constituem uma importante fonte de informação com relação à cauda superior das enchentes.

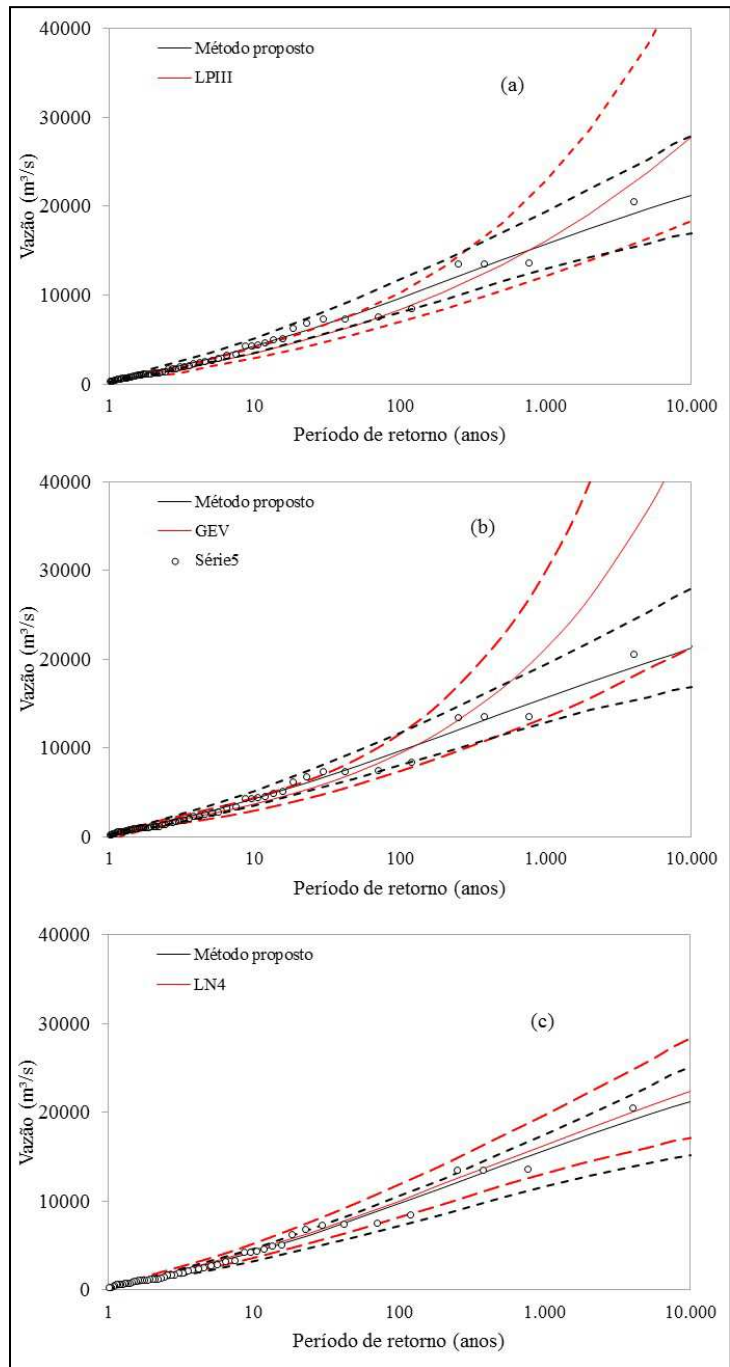


Figura 6.34 – Comparação dos intervalos de credibilidade dos métodos de estimação de cheias

7 CONCLUSÕES E RECOMENDAÇÕES

O presente trabalho teve como objetivo o estabelecimento de um método para estimação indireta de quantis de enchentes extremas a partir de um modelo de simulação hidrológica. Para tanto, fez-se uso de uma estrutura metodológica fundamentada no paradigma Bayesiano, a qual envolveu a construção de um novo modelo de geração estocástica de precipitação diária e a definição de um método de calibração de parâmetros de um modelo chuva-vazão apto a caracterizar o comportamento dos resíduos da simulação. Tal estrutura de análise oferece uma abordagem coerente para acomodar as incertezas relacionadas a essas duas etapas, o que influencia diretamente as estimativas dos quantis de cheias.

O gerador estocástico proposto se baseia em uma estrutura mista de simulação, a qual confere distinções ao tratamento dos mecanismos físicos de formação de precipitação. Nesse contexto, chuvas de alturas moderadas são simuladas em uma perspectiva não paramétrica, por meio de um modelo de reamostragem. Por sua vez, as chuvas extremas, às quais se faz necessária a extrapolação a quantis com valores bastante superiores ao máximo observado, são simuladas por modelos probabilísticos superiormente limitados, com base na hipótese de que a síntese física dos processos de formação de tormentas em condições extremas impõe um limite superior finito a essa variável.

Ainda que a existência de um limite superior para precipitações e vazões constitua uma questão controversa, há fortes evidências de que as referidas variáveis são limitadas superiormente, e, sendo assim, as mesmas devem ser modeladas sob tal perspectiva. Contudo, o mero expediente de se fixar um valor prático para o limite superior, tal como a estimativa da Precipitação Máxima Provável (PMP), tem se mostrado inadequado, uma vez que a quantificação da variabilidade espaço-temporal das variáveis que dão origem a essa grandeza é, em geral, insuficiente para se prescrever um valor inequívoco à mesma. Dessa forma, é mais coerente tratar a PMP como uma variável aleatória, a qual deve ser descrita por uma distribuição de probabilidades.

No método desenvolvido nesta tese, estabeleceu-se uma abordagem probabilística que permitiu a incorporação da PMP à estrutura do gerador estocástico. Tal inclusão se deu mediante a especificação de uma distribuição *a priori* para o limite superior, na qual a

PMP constitui um estimador para a precipitação máxima, e sua estimativa é associada a uma probabilidade de superação que certamente influenciará a estimativa do referido limite. Essa construção permite ao especialista introduzir seu conhecimento acerca da PMP local, ou, em outras palavras, avaliar o quão acurada é a estimativa de PMP na bacia de interesse e o quão próxima a mesma está do limite superior das precipitações.

Também no contexto de calibração dos parâmetros do modelo Rio Grande, o paradigma Bayesiano constituiu uma ferramenta de análise coerente para o tratamento das incertezas. Com efeito, o comportamento do conjunto de parâmetros em face da influência dos dados pôde ser sumariado por meio da forma e variância das distribuições marginais *a posteriori* dos mesmos. Ademais, o uso da abordagem estatística formal na especificação da função de verossimilhança permite que as hipóteses assumidas *a priori* quando da calibração sejam validadas *a posteriori*, indicando a adequação do modelo probabilístico teórico aos resíduos de calibração. Nesse contexto, a função de verossimilhança generalizada se mostrou uma alternativa bastante conveniente, uma vez que a mesma é apta a acomodar a não normalidade, a autocorrelação e a heterocedastidade dos erros de simulação, o que se reflete em estimativas mais consistentes dos valores do vetor paramétrico em comparação a métodos tradicionais de calibração, tais como a regressão de mínimos quadrados.

O método foi aplicado a duas bacias hidrográficas, a saber, a bacia do rio Pará, em Minas Gerais, e a bacia do rio American, na Califórnia, Estados Unidos. A etapa de geração estocástica foi abordada segundo duas perspectivas distintas: na primeira delas foi eliciada uma distribuição *a priori* informativa, baseada na variabilidade regional de estimativas de PMP diária, para o limite superior de dois modelos distributivos, a saber, a distribuição Lognormal de 4 parâmetros (LN4) e a distribuição de valores extremos do tipo IV (EV4). Na segunda, foi considerada uma distribuição *a priori* não informativa para o referido parâmetro. Já na etapa de construção de curvas de quantis de cheias somente se empregou aquele modelo distributivo limitado com melhor aderência aos dados de precipitação.

Em face das aplicações realizadas, as principais conclusões acerca da metodologia proposta nesta tese podem ser sintetizadas da seguinte forma:

- A abordagem Bayesiana permite uma avaliação mais profunda e fisicamente fundamentada das incertezas relacionadas à estimação paramétrica e à predição de valores futuros de precipitação, tendo-se em vista que a mesma incorpora à análise informações distintas daquelas provenientes da amostra, tais como a estimativa de PMP;
- A PMP, sobretudo em caráter regional, fornece informações essenciais para a eliciação de uma distribuição *a priori* informativa para o limite superior das precipitações diárias. Ademais, o método prescrito para a referida eliciação é dotado de flexibilidade suficiente para acomodar a subjetividade inerente ao conhecimento do especialista acerca dessa variável;
- Pelo menos nas aplicações aqui discutidas, o modelo limitado LN4 mostrou melhor aderência aos dados de precipitação diárias máximas anuais quando comparado ao ajuste oriundo de modelos ilimitados. Em adição, quando da extrapolação a quantis mais raros, o modelo limitado se inseriu em posição intermediária dentre aqueles dotados de cauda superior exponencial, que subestimam os eventos extremos, e aqueles com cauda superior pesada, que, por sua vez, superestimam tais eventos. Esse comportamento é mais coerente com aquele estabelecido na literatura para a modelagem de precipitações extremas em base diária;
- O modelo EV4, por outro lado, se mostrou inapto a modelar as precipitações diárias máximas anuais nas duas bacias avaliadas, com forte tendência de superestimação mesmo para quantis com maior probabilidade de superação;
- O modelo de geração estocástica proposto se mostrou adequado à reprodução de estatísticas diárias, com destaque aos coeficientes de assimetria, os quais não são propriamente simulados na maioria dos modelos discutidos na literatura. Ademais, a estrutura mista em que se baseia o gerador permitiu a reprodução acurada das variâncias em escala mensal, o que constitui uma grande vantagem do modelo proposto em relação às alternativas estritamente paramétricas;
- O paradigma Bayesiano é mais coerente que o frequentista no contexto de calibração de parâmetros de modelos de simulação hidrológica. De fato, em lugar de estimativas supostamente “ótimas”, ainda que frequentemente enviesadas,

atribuídas aos parâmetros do modelo chuva-vazão pela abordagem frequentista, o equivalente Bayesiano fornece descrições probabilísticas completas dos mesmos após a calibração. Esse fato é de grande interesse no contexto de simulação em função da maior variabilidade de realizações possíveis da distribuição conjunta *a posteriori*, o que, em última análise, oferece uma caracterização mais precisa da variabilidade de respostas hidrológicas da bacia;

- A função de verossimilhança generalizada reproduz de maneira acurada o comportamento dos resíduos de calibração;
- As curvas de quantis de enchentes obtidas com o método proposto apresentam pouca aderência aos dados sistemáticos para cheias de ocorrência mais frequente. Contudo, os quantis de cheias mais extremas, tais como aqueles associados a paleocheias, são propriamente simulados;
- As curvas de quantis estimadas nas duas aplicações são bastante próximas àquelas oriundas da análise de frequência Bayesiana proposta por Fernandes (2009) e Fernandes *et al.* (2010), possivelmente em função da fundamentação física imposta à construção desses dois modelos.

As conclusões apresentadas permitem afirmar que, de maneira geral, os objetivos da presente pesquisa foram alcançados. Contudo, alguns aspectos metodológicos podem ser melhorados e pontos não explorados nas aplicações podem ser abordados em desenvolvimentos futuros. As principais recomendações nesse sentido são:

- Avaliar uma gama mais extensa de modelos distributivos dotados de limite superior na geração estocástica de precipitação diária, tanto em termos de capacidade descritiva quanto preditiva;
- Estabelecer procedimentos alternativos para eliciação da distribuição *a priori* do limite superior, de maneira a se avaliar sua influência nas estimativas *a posteriori* do referido parâmetro;
- Buscar ferramentas adequadas, em um contexto Bayesiano, para selecionar entre os modelos testados aquele que modela mais acuradamente as precipitações de maior magnitude;

- Incorporar de maneira explícita ao método de calibração as incertezas oriundas dos *inputs* do modelo hidrológico, em especial aquelas que refletem os erros de medição pluviométrica. Nesse contexto, pode ser de grande interesse estabelecer uma distribuição *a priori* informativa para esses erros, a qual restrinja de maneira efetiva a influência de grandes variações assumidas para essa variável na inferência do vetor paramétrico;
- Avaliar de maneira mais profunda o efeito de erros estruturais e da correlação dos parâmetros do modelo hidrológico na forma dos hidrogramas simulados, em especial no que se refere às vazões de pico. Tal expediente pode incorrer em estimativas mais confiáveis das enchentes mais raras.

8 REFERÊNCIAS BIBLIOGRÁFICAS

- ANDERSON, H. L. Metropolis, Monte Carlo, and the MANIAC. *Los Alamos Science*. U.S. GOVERNMENT PRINTING OFFICE, pp. 96-108, 1986.
- APIPATTANAVIS, S.; PODESTÁ, G.; RAJAGOPALAN, B.; KATZ, R. W. A semiparametric multivariate and multisite weather generator. *Water Resources Research*, v. 43, W11401, doi: 10.1029/2006WR005714, 2007.
- BASINGER, M.; MONTALTO, F.; LALL, U. A rainwater harvesting system reliability model based on nonparametric stochastic rainfall generator. *Journal of Hydrology*, v. 392, pp. 105-118, 2010.
- BAYAZIT, M.; ÖNÖZ, B. Envelope curves for maximum floods in Turkey. *Digest 2004*, pp. 927-931, 2004.
- BERDANARA, P.; SCHERTZER, D.; SAUQUET, E.; TCHIGUIRINSKAIA, I.; LANG, M. The flood probability distribution: how heavy is it. *Stoch Environ Res Risk Assess*, v. 22, pp. 107-122, 2008.
- BERNARDO, J.; SMITH, A. *Bayesian Theory*. New York: John Wiley and Sons. 586 p., 1994.
- 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.
- BEVEN, K.; BINLEY, A. The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes*, v. 6, pp. 579-298, 1992.
- BLASONE, R. S.; MADSEN, H.; ROSBJERG, D. Uncertainty assessment of integrated distributed hydrological models using GLUE with Markov Chain Monte Carlo sampling. *Journal of Hydrology*, v. 353, pp. 18-32, 2008.
- BOTERO, B. A. Estimación de crecidas e alto período de retorno mediante funciones de distribución com limite superior e informaciónno sistemática. 223 p. Tese (Doutorado em Engenharia 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, n. 2, pp. 347-354, 1980.
- BOUGHTON, W. C. A daily rainfall generating model for water yield and flood studies. Report 99/9, CRC for Catchment Hydrology, Monash University, Melbourne, 21 p, 1999.
- BOUGHTON, W. C.; HILL, P.I. A design flood estimation procedure using data generation and a daily water balance model. Report 97/8, CRC for Catchment Hydrology, Monash University, Melbourne, 41 p, 1997.
- BROOKS, S. P. Bayesian computation: a statistical revolution. *Phil. Trans. R. Soc. Lond.*, v.361, pp. 2681-2697, 2003.
- BROOKS, S. P.; GELMAN, A. alternative methods for monitoring convergence of iterative simulations. *J Comput Gr Stat*, v. 7, pp. 434-455, 1998.
- BUIHAND, T. A. Some remarks of the use of daily rainfall models. *Journal of Hydrology*, v. 36, p. 295-308, 1978.

- BUISSHAND, T. A.; BRANDSMA, T. Multisite simulation of daily precipitation and temperature in the Rhine basin by nearest neighbor resampling. *Water Resources Research*, v. 37(11), pp. 2761-2776, 2001.
- CASTELLARIN, A.; VOGEL, E. M.; MATALAS N. C. Probabilistic behavior of a regional envelope curve. *Water Resources Research*, v. 41, doi: 10.1029/2004WR003042, 2005.
- CASTELLARIN, A.; MERZ, R.; BLÖCHL, G. Probabilistic envelope curves for extreme rainfall events. *Journal of Hydrology*, v. 378, pp. 263-271, 2009.
- CHAPMAN, T. Stochastic models for daily rainfall. In: HYDROLOGY AND WATER RESOURCES SYMPOSIUM, 1994, The Institution of Engineers, Australia, National Conference Publication 94/17, pp. 7-12, 1994.
- CHAPMAN, T. G. Refinements of the Srikanthan-McMahon stochastic model for daily rainfall. In: *International Congress on Modeling and Simulation*, Modeling and Simulation MODSIM 2001, Modeling and Simulation Society of Australia, pp. 287-292, 2001.
- CHEN, B.; BRISSETTE, F. J. Stochastic generation of daily precipitation amounts and evaluation of different models. *Climate Research*, v. 59, pp. 189-206, 2014.
- CHIN, E. H. Modeling daily precipitation occurrence process with Markov Chain. *Water Resources Research*, v. 13, n. 6, pp. 949-956, 1977.
- COWLES, M. K.; CARLIN, B. P. Markov chain Monte Carlo convergence diagnostics: a comparative review. *J Am Stat Assoc*, v. 91(434), pp. 883-904, 1996.
- CPRM - COMPANHIA DE PESQUISA DE RECURSOS MINERAIS – Superintendência Regional de Belo Horizonte. Regionalização de vazões sub-bacias 40 e 41, Convênio 015/2000 ANEEL – 013/CPRM/2000, Relatório Final. Belo Horizonte, 5v, 2001.
- CREAGER, W. P.; JUSTIN, J. D. Hydroelectric Handbook. *John Wiley and Sons*, New York, 1151p., 1950.
- DETZEL, D. H. M.. Geração de séries sintéticas de precipitação. 2009. 197 f. Dissertação de Mestrado, Departamento de Engenharia Hidráulica e Saneamento, Universidade Federal do Paraná, Curitiba, 197 p, 2009.
- DINGMAN, S. L. Physical Hydrology. Prentice Hall, Englewood Cliffs, New Jersey, 604p., 1994.
- DOULGAS, E. M.; VOGEL, R. M. The probabilistic behavior of floods of record in the United States. *Journal of Hydrologic Engineering*, v. 11(5), pp. 484-488, 2006.
- EHR/UFMG. Modelos de simulação hidrológica para previsão de vazões de curto prazo: Relatório técnico parcial 1. Projeto de pesquisa aplicada, 117 p., Belo Horizonte, 2001.
- ELÍASSON, J. A statistical model for extreme precipitation. *Water Resources Research*, v. 33, n. 3, pp. 449-455, 1997.
- ELÍASSON, J. Statistical estimates of PMP values. *Nordic Hydrology*, v. 25, pp. 301-312, 1994.

- 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, pp. 2287-2297, 1993.
- FERNANDES, W. S. Método para a estimação de quantis de enchentes extremas com o emprego conjunto de análise Bayesiana, de informações não sistemáticas e de distribuições limitadas superiormente. Tese de doutorado, Programa de Pós-Graduação em Saneamento, Meio Ambiente e Recursos Hídricos da UFMG, Belo Horizonte, 202 p, 2009.
- FERNANDES, W. S.; NAGHETTINI, M. Integrated frequency analysis of extreme flood peaks and flood volumes using the regionalized quantiles of rainfall depth as auxiliary variables. *Journal of Hydrological Engineering*, v. 13, pp. 171-179, 2008.
- FERNANDES, W. S.; NAGHETTINI, M.; LOSCHI, R. A Bayesian approach for estimating extreme flood probabilities with upper-bounded distribution functions. *Stoch Environ Res Risk Assess*, v. 24, pp. 1127-1143, 2010.
- FOUFOULA-GEORGIU, E; LEITTENMAIER, D. P. A Markov Renewal Model for Rainfall Occurrences. *Water Resources Research*, v. 23, n. 5, pp. 875-884, 1987.
- FULLER, W. E. Flood flows. *ASCE Trans.*, v. 77, pp. 564-617, 1914.
- FURRER, E. M.; KATZ, R. W. Generalized linear modeling approach to stochastic weather generators. *Climate Research*, v. 34, pp. 129-144, 2007.
- FURRER, E. M.; KATZ, R. W. Improving the simulation of extreme precipitation events by stochastic weather generators. *Water Resources Research*, v. 44, W12439, doi: 10.1029/2008WR007316, 2008.
- GABRIEL, K. R.; NEUMANN, J. A Markov chain model for daily rainfall occurrence at Tel Aviv, Q. J. R. Meteorol. Soc., 88, pp. 90-95, 1962.
- GILKS, W. R.; RICHARDSON, S.; SPIEGELHALTER, D. J. *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall/CRC, UK, 286 p., 1996.
- GONTIJO, N. T. Avaliação das relações de frequência entre precipitações e enchentes raras por meio de séries sintéticas e simulação hidrológica. Dissertação de mestrado, Programa de Pós-graduação em Saneamento, Meio Ambiente e Recursos Hídricos da UFMG, Belo Horizonte, 156 p, 2009.
- GOSH, S.; RESNICK, S. A discussion on the mean excess plots. *Stochastic Processes and their Application*, v. 120, pp. 1492-1517, 2010.
- GOYAL, M. K.; BURN, D. H.; OJHA, C. S. P. Precipitation simulation based on a k-nearest neighbor approach using Gamma kernel. *Journal of Hydrologic Engineering*, v. 18, pp. 481-487, 2013.
- GREEN, J. R. A model for rainfall occurrence. *Journal of the Royal Statistical Society Series B (Meteorological)*, v. 26(2), pp. 345-353, 1964.
- GREGORY, J. M.; WIGLEY, T. M. L.; JONES, P. D. Application of Markov models to area-average daily precipitation series and interannual variability in seasonal totals. *Climate Dynamics*, v. 8, pp. 299-310, 1993.
- GUILLAUME, J.; ANDREWS, F. DREAM: Differential Evolution Adaptive Metropolis. R package version 0.4-2, 2012. Disponível em <http://CRAN.R-project.org/package=dream>.

GUSE, B.; HOFHERR, Th.; MERZ, B. Introducing empirical e probabilistic envelope curves into a mixed bounded distribution function. *Hydro Earth Syst Scy*, v. 14, pp. 2465-2478, 2010.

HAAN, C. T., ALLEN, D. M. e STREET, J. D. A Markov Chain Model of Daily Rainfall, *Water Resources Research*, v. 12, n. 3, p. 443-449, 1976.

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.

HEC - HYDROLOGIC ENGINEERING CENTER. HEC-1 Flood Hydrograph Package - Reference Manual. U. S. Army Corps of Engineers, Davis, 1981.

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.

HITCHCOCK, D. B. A history of Metropolis-Hastings Algorithm. *The American Statistician*, v.57, n. 4, p. 254-257, doi: 10.1198/0003130032413, 2003.

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

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

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

HUNDECHA, Y.; PAHLOW, M.; SCHUMANN, A. Modeling of daily precipitation at multiple locations using a mixture of distributions to characterize the extremes. *Water Resources Research*, v. 45, W12412, doi: 10.1029/2008WR007453, 2009.

HUTCHINSON, M. F. Methods of generation of weather sequences. *Agricultural Environments: Characterization, Classification and Mapping*, A. H. Bunting, Ed., CAB International, pp. 149–157, 1987.

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

JARVIS, C. S. Flood flow characteristics. *Trans ASCE* 88, pp. 985-1032, 1925.

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.

KATZ, R. W. Computing probabilities associated with the Markov chain model for precipitation, *J. Appl. Meteor.*, 13(8), pp. 953–954, 1974.

KATZ, R. W. Precipitation as a chain-dependent process, *J. Appl. Meteor.* , 16 (7), pp. 671–676, 1977.

KAVETSKI, D.; FRANKS, S. W.; KUCZERA, G. Confronting input uncertainty in environmental modeling. In: DUAN, Q.; GUPTA, H.; SOROOSHIAN, S.;

ROUSSEAU, A.; TURCOTTE, R. (eds). Calibration of watershed models. Water Science and Application 6, American Geophysical Union, Washington, USA, 2003.

KAVETSKI, D.; KUCZERA, G.; FRANKS, S. W. Bayesian analysis of uncertainty in hydrological modeling: 1. Theory. *Water Resources Research*, v. 42, W03407, doi: 10.1029/2005WR004368, 2006a.

KAVETSKI, D.; KUCZERA, G.; FRANKS, S. W. Bayesian analysis of uncertainty in hydrological modeling: 2. Application. *Water Resources Research*, v. 42, W03408, doi: 10.1029/2005WR004376, 2006b.

KAVETSKI, D.; KUCZERA, G.; FRANKS, S. W. Calibration of conceptual hydrological models revisited: 1. Overcoming numerical artefacts. *Journal of Hydrology*, v. 320, pp. 173-186, 2006c.

KIM, Y.; KATZ, R. W.; RAJAGOPALAN, B.; PODESTÁ, G. P.; FURRER, E. M. Reducing overdispersion in stochastic weather generators using a generalized linear modeling approach. *Climate Research*, v. 53, pp. 13-24, 2012.

KLEMES, V. Hydrological and engineering relevance of flood frequency analysis. In: *Proceedings of international symposium on flood frequency risk analysis—regional flood frequency analysis*, Baton Rouge, LA. D. Reidel Publishing Company, Boston, MA, pp. 1–18, 1987.

KOUTSOYIANNIS, D. A. A probabilistic view of Hershfield's method for estimating probable maximum precipitation. *Water Resources Research*, v. 35, n. 4, pp. 1313-1322, 1999.

KOUTSOYIANNIS, D. Statistics of extremes and estimation of extreme rainfall: I. Theoretical investigation. *Hydrological Sciences Journal*, v. 49(4), pp. 575-590, 2004a.

KOUTSOYIANNIS, D. Statistics of extremes and estimation of extreme rainfall: II. Empirical investigation of long rainfall records. *Hydrological Sciences Journal*, v. 49(4), pp. 591-610, 2004b.

KOUTSOYIANNIS, D.; BALOUSTOS, G. Analysis of a long record of annual maximum rainfall in Athens, Greece, and design rainfall inferences. *Natural Hazards*, v. 29, pp. 29-48, 2000.

LALL, U.; RAJAGOPALAN, B.; TARBOTON, D. G. A non-parametric wet/dry spell model for resampling daily precipitation. *Water Resources Research*, v. 32, n. 9, pp. 2803-2893, 1996.

LALL, U.; SHARMA, A. A nearest neighbor bootstrap for resampling hydrologic time series. *Water Resources Research*, v. 32(3), pp. 679-693, 1996.

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, pp. 1339-1339, doi: 10.1029/WR019i005p01339, 1983.

LI, C.; SINGH, V. P.; MISHRA, A. K. Simulation of the entire range of daily precipitation using a hybrid probability distribution. *Water Resources Research*, v. 48, W03521, DOI: 10.1029/2011WR011446, 2012.

LI, C.; SINGH, V. P.; MISHRA, A. K. A bivariate mixed distribution with a heavy-tailed component and its application to single-site daily rainfall simulation. *Water Resources Research*, v. 49, pp. 767-789, 2013.

- 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 de mestrado, Programa de Pós-Graduação em Saneamento, Meio Ambiente e Recursos Hídricos da UFMG, Belo Horizonte, 2004.
- LIU, J. S. *Monte Carlo Strategies in Scientific Computing*. New York: Springer, 343 p., 2001.
- LONGLEY, R. W. The length of dry and wet periods. *Quarterly Journal of the Royal Meteorological Society*, v. 79, pp. 520-527, 1953.
- LÜ, H.; Hou, T.; HORTON, R.; Zhu, Y.; CHEN, X.; JIA, Y.; WANG, W.; FU, X. The streamflow estimation using the Xinanjiang rainfall runoff model and dual-state parameter estimation method. *Journal of Hydrology*, v. 480, pp. 102-114, 2013.
- LUNN, D. J.; THOMAS, A.; BEST, N.; SPIEGELHALTER, D. WinBUGS – a Bayesian modeling framework: concepts, structure, and extensibility. *Statistics and Computing*, v. 10, pp. 325-337, 2000.
- MARKOVICH, N. *Nonparametric Analysis of Univariate Heavy-Tailed Data: Research and Practice*, JohnWiley, London, 336 p., 2007.
- MEHOTRA, R.; SRIKANTHAN, R.; SHARMA, A. A comparison of three stochastic multi-site precipitation occurrence generators. *Journal of Hydrology*, v. 331, pp. 280-292, 2006.
- METROPOLIS, N. The beginning of Monte Carlo Method. *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, pp. 1087-1092, 1953.
- MIMIKOU, M. Daily precipitation occurrences modeling with Markov chain of seasonal order. *Hydrological Sciences Journal*, v. 28(2), pp. 221-233, 1983.
- MONTANARI, A. Large sample behaviors of the generalized likelihood uncertainty estimation (GLUE) in assessing the uncertainty or rainfall-runoff simulations. *Water Resources Research*, v. 41, W08406. doi:10.1029/2004WR003826, 2005.
- MORADKHANI, H.; SOROOSHIAN, S. General review of rainfall-runoff modeling: model calibration, data assimilation, and uncertainty analysis. In: SOROOSHIAN, S., HSU, K-L.; COPPOLA, E.; TOMASSETI, B.; VERDECHIA, M.; VISCONTI, G. (eds). *Hydrological modeling and the water cycle: coupling the atmospheric and hydrological models*. Springer, 2008.
- NAGHETTINI, M., NASCIMENTO, N. O., FERNANDES, W. S. Relatório Técnico 1: a estrutura funcional do Sistema Rio Grande – um sistema integrado para simulação hidrológica e previsão de vazões afluentes a reservatório de aproveitamentos hidrelétricos. Departamento de Engenharia Hidráulica e Recursos Hídricos da UFMG, Belo Horizonte, 2006, 73 p.
- NAGHETTINI, M., PINTO, E. J. A. *Hidrologia Estatística. 1a ed. Belo Horizonte: CPRM – Companhia de Pesquisa de Recursos Minerais – Superintendência Regional de Belo Horizonte*, 552 p, 2007.

NOAA – Technical Report NWS n.º 34: Mean monthly, seasonal and annual pan evaporation for the United States. US department of commerce, National Oceanic and Atmospheric Administration, National Weather Service, 1982.

NONHEBEL, S. The effects of use of average instead of daily weather data in crop growth simulation models. *Agricultural Systems*, v. 44, pp. 377-396, 1994.

OHNISHI, R.; KATO, T.; IMAIZUMI, M.; TANJI, H. A study to determine an equation of envelope curve for regional flood peaks by maximum rainfall data. *Paddy Water Environ*, v. 2, pp. 83-90, 2004.

PAPALEXIOU, S. M.; KOUTOSOYIANNIS, D. A probabilistic approach to the concept of probable maximum precipitation. *Advances in Geosciences* 7, pp. 51-54, 2006.

PAPALEXIOU, S. M.; KOUTOSOYIANNIS, D. Entropy based derivation of probability distributions: A case study to daily rainfall. *Advances in Water Resources*, v. 45, pp. 51-57, 2012.

PAPALEXIOU, S. M.; KOUTOSOYIANNIS, D.; MACROPOULOS, C. How extreme is extreme? An assessment of daily rainfall distribution tails. *Hydro Earth Syst Sci*, v. 17, pp. 851-862, 2013.

PEGRAM, G.; PARAK, M. A review of the maximum regional flood and rational formula using geomorphological information and observed floods. *Water SA*, v. 3, pp. 377-392, 2004.

PINHEIRO, V. B. Calibração de um modelo chuva-vazão em bacias sem monitoramento fluviométrico a partir de curvas de permanência sintéticas. Dissertação de mestrado, Programa de Pós-graduação em Saneamento, Meio Ambiente e Recursos Hídricos da UFMG, Belo Horizonte, 291 p, 2009.

RACSKO, P.; SZEIDL, L. A serial approach to local stochastic weather models. *Ecological Modeling*, v. 57, pp. 27-41, 1991.

RAJAGOPALAN, B.; LALL, U.; TORBOTON, D.G. A nonhomogeneous Markov model for daily precipitation simulation. *Journal of Hydrologic Engineering* 1(1), pp. 33-40, 1996.

RAJAGOPALAN, B.; LALL, U. A k-nearest neighbor simulator for daily precipitation and other weather variables. *Water Resources Research*, v. 35, n. 10, pp. 3089-3101, 1999.

RENARD, B.; KAVETSKI, D.; LeBLOIS, E.; THYER, M.; KUCZERA, G.; FRANKS, S. W. Toward a reliable decomposition of predictive uncertainty in hydrological modeling: characterizing rainfall errors using conditional simulation. *Water Resources Research*, v. 47, W11516, doi : 10.1029/2011WR010643, 2011.

RICHARDSON, C. W. Stochastic simulation of daily precipitation, temperature, and solar radiation, *Water Resources Research*, v. 17(1), pp. 182–190, 1981.

RICHARDSON, C. W. Weather simulation for crop management models. *Transactions of the American Society of Agricultural Engineers*, v. 28, pp. 1602-1606, 1985.

ROBERT, C. P.; CASELLA, G. *Monte Carlo statistical methods*. New York: Springer, 2º ed., 645 p., 2004.

SCHOUPS, G.; VRUGT, J. A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic and non-Gaussian errors. *Water Resources Research*, W10531, doi:10.1029/2009WR008933, 2010.

SELVALINGAM, S.; MIURA, M. Stochastic modeling of monthly and daily rainfall sequences. *Journal of the American Water Resources Association*, 14, pp. 1105–1120. doi: 10.1111/j.1752-1688.1978.tb02250.x, 1978.

SEMENOV, M. A.; BARROW, E. M. Use of a stochastic weather generator in the development of climate change scenarios. *Climate Research*, v. 35, pp. 397-414, 1997.

SEMENOV, M. A.; BROOKS, R. J.; BARROW, E. M.; RICHARDSON, C. W. Comparison of the WGEN and LARS-WG stochastic weather generators for diverse climates. *Climate Research*, v. 10, pp. 95-107, 1998.

SHARIF, M.; BURN, D. M. Simulating climate change scenarios using an improved K-nearest neighbor model. *Journal of Hydrology*, v. 325, pp. 179-196, 2006.

SHARMA, A.; LALL, U. A nonparametric approach for daily rainfall simulation. *Mathematics and Computers in Simulation*, v. 48, pp. 361-371, 1999.

SILVA, F.; NAGHETTINI, M., FERNANDES, W. Avaliação Bayesiana das incertezas nas estimativas dos parâmetros de um modelo chuva-vazão conceitual. *Revista Brasileira de Recursos Hídricos*, vol. 19, n. 4, 2014.

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

SMALL, M. J.; MORGAN, D. J. The Relationship Between a Continuous-Time Renewal Model and a Discrete Markov Chain Model of Precipitation Occurrence. *Water Resources Research*, v. 22, n. 10, pp. 1422-1430, 1986.

SRIKANTHAN, R. Stochastic generation of daily rainfall using a nested model. In: 57th Canadian Water Resources Association Congress, 2004, Montreal, Canada, 2004.

SRIKANTHAN, R. Stochastic generation of daily rainfall at a number of sites. Technical Report, Report 05/7, 67 p., 2005.

SRIKANTHAN, R.; McMAHON, T. Stochastic generation of rainfall and evaporation data. A WRC Technical Paper No 84, 301 p, 1985.

SRIKANTHAN, R.; McMAHON, T. Stochastic generation of climate data: a review. Report 00/16, CRC for Catchment Hydrology, Monash University, Melbourne, 34 p, 2000.

SRIKANTHAN, R.; PEGRAM, G. G. S. A nested multisite daily rainfall stochastic generation model. *Journal of Hydrology*, v. 371, pp. 142-153, 2009.

STEDINGER, J. R; VOGEL, R. M; LEE, S. U.; BATCHELDER, R. Appraisal of the generalized likelihood uncertainty estimation (GLUE) method. *Water Resources Research*, v. 44, doi: 10.1029/2008WR006822, 2008.

STERN, D.; COE, R. A model fitting analysis of daily rainfall data, *J. R. Stat. Soc., Series A*, 147, pp. 1–34, 1984.

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, pp. 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.

THOMPSON, C. S.; THOMSON, P. J.; ZHENG, X. Fitting a multisite daily rainfall model to New Zealand data. *Journal of Hydrology*, v. 340, pp. 25-39, 2007.

TODOROVIC, P.; WOOLHISER, D. A. A stochastic model of n-day precipitation. *Journal of Applied Meteorology*, v. 14, pp. 17-24, 1975.

TUCCI, C. E. M. Modelos Hidrológicos. 1ª ed. Porto Alegre: Editora Universidade (UFRGS), Associação Brasileira de Recursos Hídricos, 669 p, 1998.

USACE – United States Army Corps of Engineers. *Stochastic modeling of extreme floods on the American river at Folsom dam: flood frequency curve extension*. RD-48, USACE, 2005.

USBR – United States Bureau of Reclamation. *Flood hazard analysis, Folsom dam – Central Valley Project*, Flood Hydrology Group, Denver, USA, 2002.

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

VOGEL, R. M.; MATALAS, N. C., ENGLAND, J. F.; CASTELLARIN, A. An assessment of exceedance probabilities of envelope curves. *Water Resources Research*, v. 18(5), pp. 1461-1468, 2007.

VOGEL R. M; ZAFIRAKOU-KOULOURIS, A.; MATALAS, N. Frequency of record-breaking floods in the United States. *Water Resources Research*, v. 37(6), pp. 1723-1731, 2001.

VRUGT, J.; ter BRAAK, C.; GUPTA, H.; ROBINSON, B. Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling? *Stoch Environ Res Risk Assess*, v. 23(7), pp. 1011-1026, 2008.

VRUGT, J.; ter BRAAK, C.; DIKS, C.; ROBINSON, B.; HYMAN, J.; HIGDON, D. Accelerating Markov Chain Monte Carlo simulation by Differential Evolution with self-adaptive randomized subspace sampling. *International Journal of Nonlinear Sciences and Numerical Simulation*, v. 10(3), pp. 273-290, 2009.

VRUGT, J.; ter BRAAK, C. DREAM_(D): an adaptive Markov Chain Monte Carlo simulation algorithm to solve discrete, noncontinuous, and combinatorial posterior parameter estimation problems. *Hydro Earth Syst Sci*, v. 15, pp. 3701-3713, 2011.

VRUGT, J.; ter BRAAK, C.; DIKS, C.; SCHOUPS, G. Hydrologic data assimilation using particle Markov Chain Monte Carlo simulation: Theory, concepts and applications. *Advances in Water Resources*, v. 51, pp. 457-478, 2013.

WAGENER, T.; WHEATER, H.; GUPTA, H. V. *Rainfall-Runoff Modeling in Gauged and Ungauged Catchments*. Singapore: Imperial College Press, 306 p, 2004.

WILBY, R. L.; WIGLEY, T. M. L.; CONWAY, D.; JONES P. D.; HEWISTON, B. C.; MAIN, J.; WILKS, D. S. Statistical downscaling of general circulation model output: a comparison of methods. *Water Resources Research*, v. 34, pp. 2995-3008, 1998.

- WILLIAMS, C. B. Sequences of wet and dry days considered in the relation to the logarithmic series. *Quarterly Journal of the Royal Meteorological Society*, v. 78, pp. 91-96, 1952.
- WILKS, D. S. Adapting stochastic weather generation algorithms for climate change studies. *Climate Change*, v. 22, pp. 67-84, 1992.
- WILKS, D. S. Multisite generalization of a daily stochastic precipitation generation model, *J. Hydrol.*, 210, pp. 178–191, 1998.
- WILKS, D. S. Interannual variability and extreme-value characteristics of several stochastic daily precipitation models. *Agricultural and Forest Meteorology*, v. 93, pp. 153-169, 1999.
- WILKS, D. S.; WILBY, R. L. The weather generation game: a review of stochastic weather models. *Progress in Physical Geography*, v. 23(3), pp. 329-357, 1999.
- WOOLHISER, D. H.; ROLDÁN, J. Stochastic Daily Precipitation Models. 2. A Comparison of distribution of amounts. *Water Resources Research*, v. 18, n. 5, pp. 1461-1468, 1982.
- WOOLHISER, D. H.; ROLDÁN, J. Seasonal and regional variability of parameters for stochastic daily precipitation models. *Water Resources Research*, v. 22, n. 6, pp. 965-978, 1986.
- WMO – World Meteorological Organization. *Manual for estimation of probable maximum precipitation*, Operational Hydrologic Report No.1, WMO No. 332, 2^a ed., Geneva, 270 pp., 1986.
- YATES, D.; GANGOPADHYAY, S.; RAJAGOPALAN, B.; STRZEPEK, K. A technique for generating regional climate scenarios using a nearest-neighbor algorithm. *Water Resources Research*, v. 39(7), 1199, doi: 10.1029/2002WR001769, 2003.
- YEVJEVICH, V. Misconceptions in hydrology and their consequences. *Water Resources Research*, v 4, n. 2, pp.225–232, 1968.
- YEVJEVICH, V.; HARMANCIOGLU, N. B. Some reflections on the future of hydrology. In: *Proceedings of Rome symposium on water for the future: hydrology in perspective*, IAHS Publ. 164. International Association of Hydrological Sciences, Wallingford, UK, pp. 405-414, 1987.
- ZHIJIA, L.; PENGLEI, X.; JIAHUI, T. Study of the Xinanjiang model parameter calibration. *J. Hydrol. Eng.*, v. 18, pp. 1513-1521, 2013.
- ZHAO, R. J.; ZHUANG, L. R.; FANG, X.; LIU, R.; ZHANG, Q. S. The Xinanjiang model. In: *Hydrological Forecasting. Proceedings from the Oxford Symposium*, IAHS Publication 129, p. 351-356, 1980.