

Serviço de Pós-Graduação EESC/USP  
EXEMPLAR REVISADO  
Data de entrada no Serviço: 08/08/02  
Ass.: 

ANÁLISE DE CHEIAS ANUAIS  
SEGUNDO DISTRIBUIÇÃO  
GENERALIZADA

MANOEL MOISÉS FERREIRA DE QUEIROZ

Tese apresentada à Escola de Engenharia  
de São Carlos, da Universidade de São  
Paulo, como parte dos requisitos para  
obtenção do título de Doutor em Hidráulica  
e Saneamento

DEDALUS - Acervo - EESC



31100037012

ORIENTADOR: Prof. Tit. Fazal Hussain Chaudhry

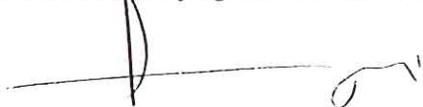


São Carlos  
2002

**FOLHA DE JULGAMENTO**

Candidato: Engenheiro **MANOEL MOISES FERREIRA DE QUEIROZ**

Tese defendida e julgada em 02-07-2002 perante a Comissão Julgadora:



APROVADO

Prof. Tit. **FAZAL HUSSAIN CHAUDHRY (Orientador)**  
(Escola de Engenharia de São Carlos/USP)



Aprovado

Prof. Dra. **LUISA FERNANDA RIBEIRO REIS**  
(Escola de Engenharia de São Carlos/USP)



Aprovado

Prof. Dr. **RODRIGO DE MELO PORTO**  
(Escola de Engenharia de São Carlos/USP)



Aprovado

Prof. Dr. **MARINHO GOMES DE ANDRADE FILHO**  
(Instituto de Ciências Matemáticas e de Computação/USP)



APROVADO

Prof. Dr. **VALTER HERNANDEZ**  
(Universidade Estadual de Campinas/UNICAMP)



Prof. Associada **MARIA DO CARMO CALIJURI**  
Coordenadora do Programa de Pós-Graduação em  
Engenharia (Hidráulica e Saneamento) e  
Presidente da Comissão de Pós-Graduação

“Porque os espelhos são rios  
os homens podem se reconhecer  
através das águas.

Se algo os impede de chegar ao mar  
é porque homens já não há  
que os possam navegar.”

*Arthur Werbeck*

À minha esposa  
Edvanina

Às minhas filhas  
Louise e Ingrid

*DEDICO*

## AGRADECIMENTO ESPECIAL

Ao Prof. Tit. Fazal Hussain Chaudhry, pela oportunidade de desenvolver este trabalho a seu lado, partilhando de sua competência, conhecimentos, orientação e apoio constantes. Agradeço pelos ensinamentos, fundamentais para a minha formação acadêmica, pelo exemplo de profissionalismo e de postura ética, pela amizade e confiança.

## AGRADECIMENTOS

À Universidade Estadual do Oeste do Paraná por ter-me concedido o afastamento para realização deste programa de Doutorado.

À Coordenadora de Aperfeiçoamento de Pessoal de Nível Superior – CAPES pela concessão de bolsa através do PICDT.

A todos os professores e funcionários do Centro de Ciências Exatas e Tecnológicas da UNIOESTE que me apoiaram.

A todos os colegas, professores e funcionários do Departamento de Hidráulica e Saneamento da EESC/USP pela colaboração.

Aos meus pais, irmãos e amigos pelo o incentivo, carinho e apoio.

## SUMÁRIO

<b>LISTA DE FIGURAS</b> .....	x
<b>LISTA DE TABELAS</b> .....	xviii
<b>LISTA DE ABREVIATURAS</b> .....	xxi
<b>LISTA DE SÍMBOLOS</b> .....	xxii
<b>RESUMO</b> .....	xxiii
<i>ABSTRACT</i> .....	xxiv
<b>1 INTRODUÇÃO</b> .....	1
<b>2 REVISÃO BIBLIOGRÁFICA</b> .....	7
2.1 ANÁLISE DE FREQUÊNCIA DE EVENTOS EXTREMOS MÁXIMOS...	8
2.2 GERAÇÃO DE SÉRIES SINTÉTICAS DE VAZÕES DIÁRIAS.....	16
2.2.1 Modelos de Vazões Diárias.....	17
<b>3 OBJETIVOS</b> .....	26
<b>4 TEORIA BÁSICA DA ESTATÍSTICA DE VALORES EXTREMOS</b> .....	29
4.1 DISTRIBUIÇÃO DE PROBABILIDADE EXATA DE EXTREMOS MÁXIMOS.....	30
4.2 DISTRIBUIÇÕES ASSINTÓTICAS DE EXTREMOS MÁXIMOS.....	32
4.2.1 Dedução das Distribuições Assintóticas de Valores Extremos.....	32
4.2.2 Aplicações.....	34
4.3 AS TRÊS FORMAS ASSINTÓTICAS DE VALORES EXTREMOS.....	37
4.3.1 A Forma Assintótica VEI.....	39
4.3.2 A Forma Assintótica VEII.....	46
4.3.3 A Forma Assintótica VEIII.....	48
4.4 DISTRIBUIÇÃO DE PROBABILIDADE GENERALIZADA DE VALORES EXTREMOS - GEV.....	50
<b>5 METODOLOGIA</b> .....	55

5.1 MÉTODOS DE ESTIMAÇÃO DE PARÂMETROS.....	56
5.1.1 Método dos Momentos com Pesos Probabilísticos.....	58
5.1.2 Método dos Momentos L.....	64
5.1.3 Método dos Momentos LH.....	69
5.2 TESTES DE QUALIDADE DE AJUSTES.....	73
5.2.1 Diagrama de Momentos L.....	73
5.2.2 Teste de Hipótese sobre o Sinal do Parâmetro de Forma k da Distribuição GEV.....	74
5.2.3 Teste de Qualidade de Ajuste da Distribuição GEV via momentos LH.....	75
5.3 DISTRIBUIÇÃO INICIAL DE PROBABILIDADE LIMITADA.....	78
5.3.1 Estimação dos Parâmetros da Distribuição Inicial Limitada.....	79
5.4 POSIÇÃO DE PLOTAGEM DOS DADOS.....	81
5.5 OUTLIERS.....	83
5.5.1 Princípios Estatísticos do Teste de Outliers.....	83
5.6 MODELOS DE VAZÕES DIÁRIAS USADOS NA HIDROLOGIA OPERACIONAL.....	85
5.6.1 Modelo Simplificado de Desagregação - MSD.....	86
5.6.2 Modelo Shot-Noise de Segunda Ordem.....	89
5.6.3 Modelo Diana.....	95
5.6.4 Modelo de Sargent.....	100
5.6.5 Modelo Proposto para Geração de Vazões Diárias.....	105
5.7 ESTUDO DA ESTIMABILIDADE DA DISTRIBUIÇÃO GEV .....	107
5.7.1 Experimento de Simulação Monte Carlo para Geração de Vazões Diárias..	107
5.7.2 Simulação de Vazões Diárias Através de Modelos Estocásticos Diários Usados na Hidrologia Operacional.....	113
5.7.3 Geração de Vazões Diárias Utilizando o Novo Modelo Estocástico Diário.....	114
5.8 AMBIENTE COMPUTACIONAL .....	118
<b>6 RESULTADOS E DISCUSSÕES.....</b>	<b>120</b>
6.1 ANÁLISE DA AJUSTABILIDADE DA DISTRIBUIÇÃO GEV.....	120
6.1.1 Estimação dos Parâmetros k, $\alpha$ e u.....	121
6.1.2 Estimativas das Cheias Decamilenares $Q_{10000}$ .....	137

---

6.1.3 Avaliação das Incertezas de Estimação de $k$ , $\alpha$ e $u$ e $Q_{10000}$ .....	145
6.1.4 Análise da Qualidade do Ajuste.....	154
6.1.5 Posição de Plotagem dos Quantis de Cheias.....	160
6.1.6 Efeito do Tamanho da Amostra na Forma da Distribuição.....	163
<b>6.2 MODELOS DE VAZÕES DIÁRIAS USADOS NA HIDROLOGIA</b>	
<b>OPERACIONAL</b> .....	170
6.2.1 Forma Assintótica das Vazões Máximas Anuais Geradas Através de Modelos Estocásticos Disponíveis na Literatura.....	170
6.2.2 Novo Procedimento de Geração de Vazões Diárias.....	173
6.2.3 Geração de Vazões Diárias e Testes Realizados.....	188
6.2.4 Análise dos Valores Extremos Máximos.....	202
<b>7 CONCLUSÕES</b> .....	209
<b>8 REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	214
<b>9 APÊNDICES</b> .....	229
9.1 APÊNDICE I.....	229
9.2 APÊNDICE II.....	230
9.3 APÊNDICE III.....	231

## LISTA DE FIGURAS

FIGURA 1: Funções densidade e distribuição das formas exata e assintótica de valores extremos máximos de uma distribuição inicial exponencial.....	35
FIGURA 2: Distribuições das 3 formas de valores extremos em função da variável reduzida de Gumbel.....	54
FIGURA 3: Ilustração da determinação da altura de um pulso de vazão.....	102
FIGURA 4: Esquema representativo do novo modelo de geração de vazões diárias proposto neste trabalho.....	106
FIGURA 5: Razão entre os valores de $u$ e $\alpha$ que rendem vazões não negativas através da distribuição limitada para diferentes valores de probabilidade.....	109
FIGURA 6: Efeito do valor real de $k$ sobre sua estimação, em função do tamanho da amostra, mantido $\alpha = 120$ e $u = 3000$ .....	122
FIGURA 7: Efeito do valor real de $\alpha$ sobre a estimação de $k$ , em função do tamanho da amostra, mantido $k = 0,20$ e $u = 3000$ .....	123
FIGURA 8: Efeito do valor real de $u$ sobre a estimação de $k$ , em função do tamanho da amostra, mantido $\alpha = 120$ e $k = 0,20$ .....	124

FIGURA 9: Dispersão da estimação de $k$ verificada através do desvio máximo e do coeficiente de variação, em função do tamanho da amostra.....	126
FIGURA 10: Efeito do valor real de $\alpha$ sobre sua estimação, em função do tamanho da amostra, mantido $k = 0,20$ e $u = 3000$ .....	128
FIGURA 11: Efeito do valor real de $k$ sobre a estimação de $\alpha$ , em função do tamanho da amostra, mantido $\alpha = 120$ e $u = 3000$ .....	129
FIGURA 12: Efeito do valor real de $u$ sobre a estimação de $\alpha$ , em função do tamanho da amostra, mantido $k = 0,20$ e $\alpha = 120$ .....	130
FIGURA 13: Dispersão da estimação de $\alpha$ verificada através do desvio máximo e do coeficiente de variação, em função do tamanho da amostra.....	131
FIGURA 14: Efeito do valor real de $u$ sobre sua estimação, em função do tamanho da amostra, mantido $k = 0,20$ e $\alpha = 120$ .....	133
FIGURA 15: Efeito do valor real de $k$ sobre a estimação de $u$ , em função do tamanho da amostra, mantido $\alpha = 120$ e $u = 3000$ .....	134
FIGURA 16: Efeito do valor real de $\alpha$ sobre a estimação de $u$ , em função do tamanho da amostra, mantido $k = 0,20$ e $u = 3000$ .....	135
FIGURA 17: Dispersão da estimação de $u$ verificada através do desvio máximo e do coeficiente de variação, em função do tamanho da amostra.....	136
FIGURA 18: Efeito do valor real de $k$ sobre a estimação da cheia decamilenar, em função do tamanho da amostra, mantido $\alpha = 120$ e $u = 3000$ .....	138

FIGURA 19: Efeito do valor real de $\alpha$ sobre a estimação da cheia decamilenar, em função do tamanho da amostra, mantido $k = 0,20$ e $u = 3000$ .....	139
FIGURA 20: Efeito do valor real de $u$ sobre a estimação da cheia decamilenar, em função do tamanho da amostra, mantido $k = 0,20$ e $\alpha = 120$ .....	140
FIGURA 21: Comportamento teórico dos valores de cheias decamilenares em função da variação dos parâmetros $k$ , $\alpha$ e $u$ da distribuição GEV.....	141
FIGURA 22: Dispersão da estimação de $Q_{10000}$ verificada através do desvio máximo e do coeficiente de variação, em função do tamanho da amostra.....	143
FIGURA 23: Raiz do erro quadrático médio da estimação dos parâmetros: (a) $k$ , (b) $\alpha$ , (c) $u$ e da (d) cheia decamilenar ( $Q_{10000}$ ), em função do tamanho da amostra, para os diferentes valores de $k$ , $\alpha = 120$ e $u = 3000$ .....	146
FIGURA 24: Raiz do erro quadrático médio da estimação dos parâmetros: (a) $k$ , (b) $\alpha$ , (c) $u$ e da (d) cheia decamilenar ( $Q_{10000}$ ), em função do tamanho da amostra, para os diferentes valores de $\alpha$ , $k = 0,20$ e $u = 3000$ .....	147
FIGURA 25: Raiz do erro quadrático médio da estimação dos parâmetros: (a) $k$ , (b) $\alpha$ , (c) $u$ e da (d) cheia decamilenar ( $Q_{10000}$ ), em função do tamanho da amostra, para os diferentes valores de $u$ , $\alpha = 120$ e $k = 0,20$ .....	148

---

FIGURA 26: Erro máximo de estimação dos parâmetros $k$ , $\alpha$ , $u$ e da cheia decamilenar para os diferentes valores reais de $k$ e tamanhos de amostras.....	151
FIGURA 27: Erro máximo de estimação dos parâmetros $k$ , $\alpha$ , $u$ e da cheia decamilenar para os diferentes valores reais de $\alpha$ e tamanhos de amostras.....	152
FIGURA 28: Erro máximo de estimação dos parâmetros $k$ , $\alpha$ , $u$ e da cheia decamilenar para os diferentes valores reais de $u$ e tamanhos de amostras.....	153
FIGURA 29a: Esquema do diagrama de momentos $L$ da distribuição GEV para diferentes valores de $k$ .....	155
FIGURA 29b: Diagrama de momentos $L$ de vazões máximas anuais de 107 rios mundiais.....	155
FIGURA 30: Diagramas de momentos $L$ , referentes aos ajustes da GEV às séries de vazões máximas com vários tamanhos, obtidas de vazões diárias geradas com diferentes valores de $k$ , $\alpha = 120$ e $u = 3000$ .....	156
FIGURA 31: Diagramas de momentos $L$ , referentes aos ajustes da GEV às séries de vazões máximas com vários tamanhos, obtidas de vazões diárias geradas com diferentes valores de $\alpha$ , $k = 0,2$ e $u = 3000$ .....	157
FIGURA 32: Diagramas de momentos $L$ , referentes aos ajustes da GEV às séries de vazões máximas com vários tamanhos, obtidas de vazões diárias geradas com diferentes valores de $u$ , $\alpha=120$ e $k=0,20$ .....	158

FIGURA 33a: Ajuste da GEV à série de vazões máximas anuais ( $N = 50$ ) utilizando diferentes fórmulas de Posição de Plotagem.....	161
FIGURA 33b: Ajuste da GEV à série de vazões máximas anuais ( $N = 50$ ) utilizando diferentes fórmulas de Posição de Plotagem.....	162
FIGURA 34: Distribuições GEV teórica e ajustada através de momentos LH, usando cheias máximas anuais obtidas de série de vazões diárias geradas com $k = 0,0001$ , $\alpha = 120$ e $u = 3000$ (posição de plotagem APL).....	164
FIGURA 35: Distribuições GEV teórica e ajustada através de momentos LH, usando cheias máximas anuais obtidas de série de vazões diárias geradas com $k=0,05$ , $\alpha = 120$ e $u = 3000$ (posição de plotagem APL).....	165
FIGURA 36: Distribuições GEV teórica e ajustada através de momentos LH, usando cheias máximas anuais obtidas de série de vazões diárias geradas com $k = 0,10$ , $\alpha = 120$ e $u = 3000$ (posição de plotagem APL).....	166
FIGURA 37: Distribuições GEV teórica e ajustada através de momentos LH, usando cheias máximas anuais obtidas de série de vazões diárias geradas com $k = 0,20$ , $\alpha = 120$ e $u = 3000$ (posição de plotagem APL).....	167
FIGURA 38: Distribuições GEV teórica e ajustada através de momentos LH, usando cheias máximas anuais obtidas de série de vazões diárias geradas com $k = 0,30$ , $\alpha = 120$ e $u = 3000$ (posição de plotagem APL).....	168

<p>FIGURA 39: Distribuições GEV teórica e ajustada através de momentos LH, usando cheias máximas anuais obtidas de série de vazões diárias geradas com <math>k = 0,40</math>, <math>\alpha = 120</math> e <math>u = 3000</math> (posição de plotagem APL).....</p>	169
<p>FIGURA 40: Ajuste da GEV às séries de vazões máximas geradas através dos modelos de vazões diárias: Diana, Shot-noise, Fragmentação e Sargent ajustados às vazões do rio dos Patos e rio Paraná.....</p>	171
<p>FIGURA 41a: Ajustes das distribuições de probabilidades exponencial e exponencial modificada dos pulsos de vazões simulados, ajustados à série dos pulsos observados em cada mês, durante 73 anos de registro.....</p>	178
<p>FIGURA 41b: Ajustes das distribuições de probabilidades exponencial e exponencial modificada dos pulsos de vazões simulados, ajustados à série dos pulsos observados em cada mês, durante 73 anos de registro.....</p>	179
<p>FIGURA 42: Ajuste da distribuição de probabilidade limitada às três diferentes séries de pulsos históricos, referentes aos meses de janeiro e junho.....</p>	183
<p>FIGURA 43: Média e Desvio Padrão das vazões diárias, históricas e simuladas, referente ao ajuste das três diferentes séries de pulsos de cheias observadas.....</p>	184
<p>FIGURA 44: Média e Desvio Padrão das vazões mensais, históricas e simuladas, referentes ao ajuste das três diferentes séries de pulsos de cheias observadas.....</p>	185

FIGURA 45:	Vazões extremas, históricas e simuladas, referentes ao ajuste das três diferentes séries de pulsos de cheias observadas.....	186
FIGURA 46a:	Ajuste da distribuição de probabilidade limitada aos pulsos históricos de cheias, referentes aos meses de janeiro à junho, dos 73 anos de dados.....	189
FIGURA 46b:	Ajuste da distribuição de probabilidade limitada aos pulsos históricos de cheias, referentes aos meses de julho à dezembro, dos 73 anos de dados.....	190
FIGURA 47:	Hidrógrafas representativas das vazões diárias históricas e simuladas geradas através dos três modelos de geração de pulsos de vazões.....	192
FIGURA 48:	Média e Desvio Padrão das vazões diárias, históricas e simuladas, referentes às 15 séries geradas por cada um dos três modelos de pulsos.....	193
FIGURA 49:	Média e Desvio Padrão das vazões mensais, históricas e simuladas, referentes às 15 séries geradas por cada um dos três modelos de pulsos.....	194
FIGURA 50:	Média e Desvio Padrão das vazões anuais, históricas e simuladas, referentes às 15 séries geradas por cada um dos três modelos de pulsos.....	195
FIGURA 51:	Vazões extremas, históricas e simuladas, referentes às 15 séries de vazões geradas por cada um dos três modelos de pulsos.....	196
FIGURA 52:	Diagrama de momentos amostrais das vazões diárias, mensais e anuais referentes às 15 séries simuladas e à série histórica....	197

---

FIGURA 53: Distribuição das vazões máximas observadas, simuladas e dos respectivos ajustes da GEV, considerando os três modelos de pulsos de vazões.....	204
FIGURA 54: Ajustes da GEV às séries de vazões máximas históricas e simuladas com distribuição limitada, exponencial e exponencial modificada para os pulsos de vazões.....	206
FIGURA 55: Ajuste da distribuição GEV às séries de vazões máximas formadas pelos primeiros 20, 35, 50, 100, 500 e 1095 valores da série total de vazões geradas através do modelo proposto.....	208

## LISTA DE TABELAS

TABELA 1:	Algumas distribuições mais utilizadas em análise de frequência de cheia.....	15
TABELA 2:	As três formas de distribuição assintóticas determinadas por FISHER & TIPPET (1928) e classificada por GUMBEL (1958).....	38
TABELA 3:	Valores dos coeficientes da equação (61).....	72
TABELA 4:	Valores dos coeficientes da equação (70).....	77
TABELA 5:	Valores dos coeficientes da equação (71).....	77
TABELA 6:	Alternativas de posição de plotagem e suas motivações.....	82
TABELA 7:	Parâmetros da GEV e valores dos testes de qualidade de ajuste e forma da distribuição ajustada às séries de vazões máximas geradas através dos modelos: Diana, Shot noise, Fragmentação e Sargent.....	172
TABELA 8a:	Probabilidades de transição histórica de um dia seco ser seguido por outro seco e de um dia seco ser seguido por um dia chuvoso.....	174

TABELA 8b:	Probabilidades de transição histórica de um dia chuvoso ser seguido por outro chuvoso e de um dia chuvoso ser seguido por um dia seco.....	174
TABELA 9:	Parâmetros dos modelos de recessão da vazão diária.....	175
TABELA 10:	Dados referentes às séries históricas de pulsos de vazões dos doze meses.....	176
TABELA 11:	Valores dos parâmetros e testes dos ajustes das distribuições exponencial e exponencial modificada às séries históricas de pulsos de vazões.....	177
TABELA 12a:	Parâmetros da distribuição limitada referentes às diferentes séries de pulsos de vazões ordenados de forma crescente.....	181
TABELA 12b:	Parâmetros da distribuição limitada referentes às séries de picos de vazões dos 12 meses utilizadas no ajuste do modelo proposto.....	188
TABELA 13:	Testes da qualidade de ajustes dos modelos utilizados.....	198
TABELA 14a:	Avaliação de desempenho dos três modelos de distribuição de pulsos, considerando os momentos das vazões diárias e mensais.....	200
TABELA 14b:	Avaliação de desempenho dos três modelos de distribuição de pulsos, considerando os momentos das vazões anuais e máximas anuais e a vazão máxima.....	201
TABELA 15:	Índice de avaliação global.....	202

---

TABELA 16: Resultados do ajuste da distribuição GEV às 15 séries de vazões máximas com 73 anos, referentes aos três modelos de pulsos de vazões.....	203
TABELA 17: Resultados do ajuste da distribuição GEV às séries de vazões máximas observadas e simuladas através dos três modelos de pulsos de vazões.....	206
TABELA 18: Valores de coeficiente de correlação $\rho'$ em uma distribuição normal bivariada com ambas variáveis truncadas $\beta$ .....	230

## LISTA DE ABREVIATURAS

VEI	- Valores extremos tipo 1
VEII	- Valores extremos tipo 2
VEIII	- Valores extremos tipo 3
GEV	- Distribuição de probabilidade generalizada de valores extremos
USWRC	- United States Water Resources Council
USA	- United States of American
I.E.A.	- Institution of Engineers Australian
PWMs	- Momentos de pesos probabilísticos
reqm	- Raiz do erro quadrático médio
SOSN	- Modelo shot-noise de Segunda ordem
DSON	- Modelo shot-noise duplo
MSOSN	- Modelo shot-noise modificado de segunda ordem

## LISTA DE SÍMBOLOS

$u_n$	- Maior valor característico da variável inicial X
$v_n$	- Maior valor característico da variável inicial Y
$w_n$	- Maior valor característico da variável inicial Z
$\alpha_n$	- Medida inversa da dispersão dos valores extremos de uma variável inicial
$\omega$	- Limite superior da variável inicial
$h$	- Função de risco
$G_s$	- Função geratriz de momentos
$\Gamma$	- Função gama
$\mu$	- Média
$\sigma$	- Desvio padrão
$\gamma$	- Número de Euler
$k$	- Parâmetro de forma da distribuição GEV
$\alpha$	- Parâmetro de escala da distribuição GEV
$u$	- Parâmetro de posicionamento da distribuição GEV
$M_{p,r,s}$	- Momentos de pesos probabilísticos
$B[.]$	- Função beta
$\beta_s$	- Momentos de pesos probabilísticos de ordem s igual a $M_{1,0,s}$
$\beta_r$	- Momentos de pesos probabilísticos de ordem r igual a $M_{1,r,0}$
$bs$	- Estimador não viesado de $\beta_s$
$\lambda_r$	- Momentos de combinações lineares das estatísticas de ordem r
$\lambda_r''$	- Momentos de combinações lineares das estatísticas $\eta$ de alta ordem r
$\theta$	- Parâmetro de forma da distribuição limitada
$Q_{10000}$	- Vazão decamilenar
$Q_T$	- Quantil de cheia com período de retorno igual a T anos

## RESUMO

QUEIROZ, M. M. F. de (2002). *Análise de cheias anuais segundo distribuição generalizada*. São Carlos, 2002. 236p. Tese (Doutorado) – Escola de Engenharia de São Carlos – Universidade de São Paulo.

A análise de frequência de cheias através da distribuição de probabilidade generalizada de valores extremos-GEV tem crescido nos últimos anos. A estimação de altos quantis de cheias é comumente praticada extrapolando o ajuste, representado por uma das 3 formas inversas da distribuição GEV, para períodos de retorno bem superiores ao período dos dados observados. Eventos hidrológicos ocorrem na natureza com valores finitos, tal que, seus valores máximos seguem a forma assintótica da GEV limitada. Neste trabalho estuda-se a estimabilidade da distribuição GEV através de momentos LH, usando séries de cheias anuais com diferentes características e comprimentos, obtidas de séries de vazões diárias geradas de diversas formas. Primeiramente, seqüências estocásticas de vazões diárias foram obtidas da distribuição limitada como subjacente da distribuição GEV limitada. Os resultados da estimação dos parâmetros via momentos-LH, mostram que o ajuste da distribuição GEV a amostras de cheias anuais com menos de 100 valores, pode indicar qualquer forma de distribuição de valores extremos e não somente a forma limitada como seria esperado. Também, houve grande incerteza na estimação dos parâmetros obtidos de 50 séries geradas da mesma distribuição. Ajustes da distribuição GEV às séries de vazões anuais, obtidas de séries de fluxo diários gerados com 4 modelos estocásticos disponíveis na literatura e calibrados aos dados dos rios Paraná e dos Patos, resultaram na forma da Gumbel. Propõe-se um modelo de geração diária que simula picos de vazões usando a distribuição limitada. O ajuste do novo modelo às vazões diárias do rio Paraná reproduziu as estatísticas diárias, mensais, anuais, assim como os valores extremos da série histórica. Além disso, a série das cheias anuais com longa duração, foi adequadamente descrita pela forma da distribuição GEV limitada.

**Palavras-chave:** distribuição limitada, estimabilidade da distribuição GEV, modelos de vazões diárias, estatística de extremos, momentos LH, simulação estocástica.

## *ABSTRACT*

QUEIROZ, M. M. F. de (2002). *Analysis of annual floods by Generalized Distribution*. São Carlos, 2002, 236p. Doctoral Thesis – São Carlos of Engineering, University of São Paulo.

Frequency analysis of floods by Generalized Extreme Value probability distribution has multiplied in the last few years. The estimations of high quantile floods is commonly practiced extrapolating the adjustment represented by one of the three forms of inverse GEV distribution for the return periods much greater than the period of observation. The hydrologic events occur in nature with finite values such that their maximum values follow the asymptotic form of limited GEV distribution. This work studies the identifiability of GEV distribution by LH-moments using annual flood series of different characteristics and lengths, obtained from daily flow series generated by various methods. Firstly, stochastic sequences of daily flows were obtained from the limited distribution underlying the GEV limited distribution. The results from the LH-moment estimation of parameters show that fitting GEV distribution to annual flood samples of less than 100 values may indicate any form of extreme value distribution and not just the limited form as one would expect. Also, there was great uncertainty noticed in the estimated parameters obtained for 50 series generated from the same distribution. Fitting GEV distribution to annual flood series, obtained from daily flow series generated by 4 stochastic model available in literature calibrated for the data from Paraná and dos Patos rivers, indicated Gumbel distribution. A daily flow generator is proposed which simulated the high flow pulses by limited distribution. It successfully reproduced the statistics related to daily, monthly and annual values as well as the extreme values of historic data. Further, annual flood series of long duration are shown to follow the form of asymptotic limited GEV distribution.

**Keywords:** limited distribution, identifiability of GEV distribution, daily flow models, statistics of extremes, LH-moments, stochastic simulation.

## 1 INTRODUÇÃO

Em planejamento e elaboração de projeto estrutural de obras hidráulicas, geralmente, recorre-se à análise de frequência para estimação da cheia de projeto, ou seja, busca-se estimar a cheia que corresponde a hidrógrafa afluente ao local do aproveitamento, tal que, a capacidade estrutural da obra hidráulica seja definida de forma a garantir a sua segurança na hipótese de ocorrência deste evento. Desta forma, a estimação da cheia de projeto baseia-se na análise probabilística dos registros fluviométricos, onde a cada ano está associado uma cheia máxima anual. Isto resulta num conjunto de valores  $(q_1, q_2, \dots, q_n)$  que pode ser visto como uma amostra da variável aleatória  $Q$ , máxima vazão anual. Como o interesse é o de determinar apenas a cheia máxima de projeto  $Q_T$ , o problema é formulado como sendo o de estimar o valor de  $Q_T$ , tal que,  $P[Q \geq Q_T] = 1/T$ , sendo  $T$  o período de retorno preestabelecido.

Em projeto de grande porte, como por exemplo, os de vertedores de grandes barragens, a estimação da cheia de projeto é feita para tempos de retorno milenares, decamilenares ou maiores. Contudo, as séries históricas de cheias observadas não passam de 70 anos de registros. Assim, a estimação da cheia é, essencialmente, baseada na extrapolação dos registros de dados de vazões, utilizando o ajuste de uma função distribuição de probabilidade conhecida, que melhor justifique os mesmos (KELMAN, 1983).

A estimação de cheias com base em registros de vazões tem sido objeto de intensa pesquisa nas últimas 5 décadas. Vários modelos de distribuição de probabilidade e métodos de estimação de seus parâmetros foram propostos através da literatura científica. Apesar do esforço despendido, muitos dos procedimentos existentes são caracterizados pelo baixo grau de precisão das cheias estimadas (BOBÉE et al., 1993), notadamente, quando há necessidade de extrapolação dos dados amostrais para grandes períodos de retornos.

Acrescenta-se às incertezas de estimação de cheias anuais, o fato de as cheias observadas estarem baseadas no uso de curvas chaves construídas com carência de observações de vazões elevadas, devido as dificuldades experimentais de medições das mesmas associadas ao efeito de histerese presente nas fases de cotas crescentes e decrescentes. Os erros sistemáticos de medição que têm afetados as estimativas de descargas na secção de medição das estações fluviométricas têm sido estudados em trabalhos como os de POTTER & WALKER (1982, 1985), KUCZERA (1992, 1996) e outros.

A grande dificuldade encontrada na análise de frequência preditiva de cheias, além dos erros de medição, decorre, principalmente, da limitação dos dados de vazões disponíveis, ademais, não há qualquer conexão dos modelos de distribuição de frequência utilizados com a realidade dos processos físicos que ocorrem dentro da bacia hidrográfica. Daí, o grande desafio é dispor de uma metodologia consistente e capaz de descrever os dados de cheias corretamente, e mais importante, extrapolá-los de forma adequada e segura para grandes períodos de retornos, sem no entanto, desrespeitar os limites físicos de ocorrência dos eventos hidrológicos naturais.

Recentemente, a distribuição de probabilidade generalizada de valores extremos (GEV), introduzida por JENKINSON (1955), tem encontrado muitas aplicações em hidrologia, com crescente aceitação para descrição dos eventos máximos naturais, principalmente dados de cheias máximas anuais. Decorrente da consideração teórica de que a distribuição de valores extremos de uma amostra, converge assintoticamente para uma das três formas de valores máximos reunidas através da GEV, independente de sua distribuição subjacente (WANG, 1997b).

Na prática, a distribuição GEV tem sido usada para modelar uma extensa variedade de máximos naturais, incluindo, além de cheias, chuvas, velocidade do vento, temperaturas, picos de ondas do mar e outros máximos (MARTINS & STEDINGER, 2000). Seu uso associado aos métodos de estimação de parâmetros como: momentos de pesos probabilísticos (GREENWOOD et al., 1979; HOSKING et al., 1985a); momentos de combinações lineares das estatísticas de ordens - *momentos L* (HOSKING, 1990); momentos de altos pesos probabilísticos e momentos de combinações lineares das estatísticas de altas ordens - *momentos LH* (WANG, 1996; 1997a; 1997b) tem produzido bons resultados de ajuste aos dados de vazões máximas anuais de várias localidades.

A modelação de cheias máximas anuais através da GEV tem como resultado a obtenção de uma das três formas assintóticas de valores extremos (VE) denominadas do tipo 1 (VEI), que corresponde a distribuição de Gumbel, do tipo 2 (VEII) ou do tipo 3 (VEIII). A função de quantis de cheias definida através do inverso da GEV possui limite superior infinito para as formas VEI e VEII e apresenta-se limitada superiormente para a forma VEIII, porém a forma VEII é a única com limite inferior finito.

No âmbito da análise de frequência preditiva de vazões máximas, o maior interesse está voltado para cauda superior da distribuição. Assim, quando obtém-se a forma VEI ou VEII como resultado do ajuste da GEV aos dados amostrais de cheias, a extrapolação desses dados para períodos de retorno consideráveis torna-se duvidosa e a conseqüente estimação de altos quantis de cheias resulta em valores extremamente elevados ou infinitos (valores irrealis), pressupondo a existência de eventos hidrológicos naturais ilimitados.

Por outro lado, quando o ajuste recai na forma VEIII, a análise preditiva torna-se mais realística, onde o limite superior da função de quantis pode ter como referencial o valor da cheia máxima provável. Logo, essa forma de distribuição apresenta-se como a mais adequada para descrever cheias máximas anuais sobre o aspecto preditivo, pois o seu uso permite levar em consideração os limites reais dos processos hidrológicos fisicamente possíveis de ocorrer numa bacia. GARCIA (1997) propôs o método conjugado que considera o limite superior da GEV tipo VEIII igual a cheia máxima provável, obtida deterministicamente, na análise de frequência de cheias.

Os registros de vazões diárias observados em diferentes locais de vários países, que servem de base para a obtenção das séries anuais de cheias, possuem comprimentos que variam em torno dos 50 a 70 anos. Ajustes da GEV a esses dados tem revelado as três formas assintóticas. Nesses casos, a estimação da cheia milenar ou decamilenar necessita da extrapolação da amostra de dados.

Como é muito difícil validar qualquer técnica de extrapolação, na prática, tem ocorrido um esforço considerável na tentativa de desenvolver procedimentos de estimação de parâmetros eficientes, quando utiliza-se amostra com número reduzido de dados observados, na expectativa de que o bom desempenho obtido para períodos de retorno relativamente curtos, se mantenha para eventos menos frequentes.

Metodologias voltadas para a análise preditiva têm sido propostas, em que a estimação de altos quantis de cheias é feita seguindo a tendência de seus valores mais elevados. Este feito é obtido quando o ajuste ocorre sobre a faixa dos valores mais altos dos dados, através da censura amostral ou através de métodos que, embora considere a amostra na íntegra, os dados com valores mais elevados recebem maiores pesos durante o processo de estimação dos parâmetros da distribuição. Esses procedimentos por si só, não asseguram a obtenção do modelo mais adequado sob o ponto de vista da extrapolação dos dados para grandes períodos de retorno, principalmente, porque a quantidade de dados disponíveis não garante a convergência assintótica à forma mais apropriada.

Diante das questões colocadas, parece imprescindível o conhecimento do comportamento e das características da distribuição GEV, frente a um processo de ajuste com extrapolação para períodos de retornos consideráveis, antes de se propor alguma metodologia de análise de frequência de cheias, com base em registros históricos. Porquanto, efeitos de fatores como tamanho da amostra; presença de outliers; posição de plotagem dos dados; incerteza dos parâmetros estimados relacionados a faixa de variação de seus valores; influência da distribuição subjacente, além de outros, precisam ser investigados. Bem como, o tipo de cenário utilizado para avaliação da performance dos ajustes da GEV associado ao modelo de estimação dos parâmetros, onde, geralmente, a própria GEV tem sido usada como distribuição mãe. Além disso, é necessário definir cenários em cima de um plano de simulação mais realístico para que a comparação dos resultados tenha consistência (BOBÉE et al., 1993).

Sobre o entendimento dos processos hidrometeorológicos reinantes na natureza, é intuitivo de que mesmo nas regiões mais úmidas onde a quantidade de água precipitável presente na atmosfera é enorme, e, conseqüentemente, há ocorrência de índices pluviométricos extremamente elevados, mesmo assim, o volume de água precipitado e parte dele que se transforma em deflúvio superficial é limitado. Embora, seja, comparativamente, muito superior àqueles volumes precipitados em regiões menos úmidas ou semi-áridas.

A definição da precipitação máxima provável dada pela WORLD METEOROLOGICAL ORGANIZATION (1973) que corresponde a maior altura pluviométrica fisicamente possível de ocorrer sobre uma bacia hidrográfica,

preconiza a idéia de que a quantidade de água precipitada numa bacia é limitada e mensurável; apesar das incertezas presentes na sua estimação e da não existência de uma demonstração física da ocorrência de limites extremos de fatores meteorológicos que produzam a precipitação máxima provável, como coloca YEVEVICH (1968).

Por conseguinte, a resposta da bacia a qualquer evento de precipitação, através da vazão na sua secção de saída e dos outros componentes do ciclo hidrológico envolvidos no processo, ocorre, também, numa faixa de valores finitos e quantificáveis, por mais rara que seja a chuva. Assim, é razoável esperar que a distribuição de valores máximos de variáveis hidrológicas naturais seja limitada na sua parte superior. JENKINSON (1955) observou que a distribuição de valores extremos do tipo VEIII é a mais freqüentemente encontrada na natureza.

De outro lado, valores máximos de uma variável aleatória inicial com distribuição de probabilidade limitada convergem assintoticamente para a GEV com forma VEIII (GUMBEL, 1958), onde, neste caso, existe uma relação teórica entre as duas distribuições, definida matematicamente através de seus parâmetros. Este fato permite conhecer a distribuição teórica dos valores extremos de uma variável a partir da estimação dos parâmetros da sua distribuição inicial.

Porquanto, a utilização da distribuição limitada como subjacente, no processo de geração de vazões diárias via simulação Monte Carlo, além de produzir séries sintéticas de cheias com diferentes tamanhos e características diversas, fornece a distribuição teórica dos valores extremos esperada com os ajustes da GEV às referidas séries de cheias. Desta forma, é possível comparar as estimativas dos parâmetros da GEV e dos quantis de cheias, provenientes da modelação das séries sintéticas, com seus respectivos valores teóricos. Isto possibilita proceder à análise da estimabilidade da GEV de forma mais consistente.

Sobre o aspecto prático da análise de freqüência, as dificuldades e incertezas impostas pela restrição dos dados no processo de ajuste da distribuição e extrapolação da amostra para grandes períodos de recorrência, podem ser contornados com a utilização de séries sintéticas de vazões diárias longas o suficiente para permitir a estimação de eventos com tempo de retorno elevado.

Neste contexto, a geração de vazões diárias é mais apropriada quando adotam-se modelos estocásticos de simulação diária. Pois, esses modelos possuem na sua formulação estruturas eficazes de modulação das acessões brucas e recessões

lentas, características das hidrógrafas de vazões diárias. Ademais, utilizam a totalidade das informações históricas existentes no arranjo seqüencial das vazões diárias observadas, permitindo aproximar-se, de maneira mais adequada, da complexidade do fenômeno natural. Porém, este procedimento não fornece nenhuma informação a respeito da verdadeira distribuição dos valores máximos das vazões diárias, como acontece com o uso da distribuição limitada. Por outro lado, a distribuição limitada não reproduz a sazonalidade que é peculiar em registros de vazões diárias, tão bem quanto os diversos modelos estocásticos de simulação diária existentes.

Como os modelos estocásticos diários utilizam procedimentos próprios para a geração da componente aleatória presente no valor final da vazão diária simulada, a priori, não se sabe que tipo de distribuição de valores extremos pode-se obter com a geração de vazões diárias, utilizando os diferentes modelos estocásticos disponíveis na literatura.

É preciso desenvolver uma metodologia de geração de vazões diárias que, além de representar adequadamente os dados históricos de vazões diárias observadas, consiga gerar valores máximos de cheias que, quando ajustados a um modelo de distribuição de valores extremos, apresentem convergência assintótica para um valor superior limitado, mostrando-se coerente com a realidade dos processos hidrológicos que acontecem sob um determinado limite físico factível.

## 2 REVISÃO BIBLIOGRÁFICA

Sistemas hidrológicos são, às vezes, impactados através de eventos extremos, tal como cheias extremas, provocando inundações de áreas de interesse social ou secas drásticas. A magnitude de um evento extremo está relacionada com o inverso da sua frequência de ocorrência. Os eventos severos ocorrem com menores frequência do que eventos mais moderados. O objetivo da análise de frequência de dados hidrológicos é relacionar a magnitude de eventos extremos à sua frequência de ocorrência através do uso de distribuição de probabilidade apropriada. Os dados hidrológicos analisados são assumidos como sendo independentes e identicamente distribuídos e o sistema hidrológico que os produz é considerado estocástico. Os resultados da análise de frequência de cheias podem ser usados para vários propósitos de engenharia como, projetos de barragens, pontes e estruturas de controle de cheias, por exemplo, para determinar o valor econômico de projetos de controle de cheias, para delinear planos de cheias e o efeito de se transgredir esses planos, etc. (CHOW et al., 1988). TUCCI (1997) apresenta em detalhes os tipos de obras e correspondente períodos de retornos recomendados.

Por outro lado, é cada vez mais freqüente a utilização de séries sintéticas no planejamento de sistemas de aproveitamento de recursos hídricos, decorrente da limitação dos dados históricos de cheias máximas anuais que impõem dificuldades na escolha e ajuste da distribuição de frequência e maiores incertezas nos resultados obtidos. Neste sentido, vários modelos estocásticos vêm sendo propostos e utilizados em estudos hidrológicos, como por exemplo, para cálculo de volume de espera de reservatórios utilizados em controle de cheias; cálculo da confiabilidade de reservatórios de abastecimentos; planejamento de geração de energia; análise de segurança de barragens, bem como na modelagem de vazões extremas etc.

Nesta seção, apresenta-se, primeiro, uma revisão de literatura sobre a análise de frequência de cheias observadas e, em seguida, descreve-se sobre a geração de séries sintéticas de vazões, utilizando modelos estocásticos de geração de vazões diárias.

## 2.1 ANÁLISE DE FREQUÊNCIA DE EVENTOS EXTREMOS MÁXIMOS

Na análise de frequência de cheias, uma determinada distribuição de probabilidade é ajustada aos valores disponíveis, visando, principalmente, estimar magnitude de cheias na cauda superior da distribuição, correspondente a faixa de extrapolação dos dados. Neste sentido, estudos vêm sendo desenvolvido na busca de determinar a distribuição que melhor se ajusta aos dados e possibilita uma análise preditiva eficaz.

O emprego dos modelos de análise de frequência de cheias, que começou por volta de 1913 com o uso da distribuição normal (HORTON, 1913) e da distribuição log-normal com dois parâmetros (HAZEN, 1914), tem desenvolvido ao longo do tempo linhas divergentes, conduzindo a métodos de análise e, conseqüentemente, resultados não uniformes.

O Water Resources Council dos Estados Unidos (U.S. WATER RESOURCES COUNCIL – USWRC, 1967), através do seu comitê de hidrologia, criou um grupo de trabalho para estudar os métodos de frequência de cheias, com o objetivo de estabelecer uma técnica de análise que fosse uniforme. O grupo ajustou as distribuições gama, gumbel, log-gumbel, log-normal (todas com dois parâmetros), log-Pearson tipo III (com 3 parâmetros) e o método de Hazen aos registros de cheias de 10 estações de várias partes dos USA. Os registros variaram entre 44 e 97 anos, dos quais, somente 2 excederam os 60 anos. Nenhum teste estatístico foi aplicado para verificar a qualidade dos ajustes. Contudo, valores de cheias com períodos de retornos entre 2 e 50 anos foram obtidos, através de gráficos de probabilidades e comparados aos correspondentes valores obtidos através das distribuições utilizadas. As distribuições gama, gumbel e log-gumbel apresentaram desvio médio padrão alto, enquanto a log-normal, log-Pearson tipo III e Método de Hazen forneceram valores baixos. Não obstante, a distribuição log-Pearson III foi preferida, devido ao uso comum já existente e por ser capaz de ajustar-se a dados assimétricos. As conclusões

e recomendações deste estudo conduziram, em grade escala, a adoção da log-Pearson III nos USA, muito embora, o grupo de trabalho tenha feito as seguintes recomendações: 1) que a log-Pearson III (com a log-normal como caso especial) fosse adotada como método base para análise de frequência de cheias; 2) que fossem utilizadas outras distribuições ou técnicas com melhor desempenho e, 3) que a escolha do método base não fosse considerado como resultado final (BENSON, 1968).

Visando justificar a recomendação do comitê de hidrologia do US WRC, BENSON (1968) ajustou várias distribuições a 10 séries de dados de cheias retirados de postos fluviométricos que dispunham de longo registro histórico. Usando um critério de adequação de ajuste, baseado no desvio relativo entre vazões prevista e observadas, para determinados períodos de retorno, sua conclusão foi favorável ao uso da distribuição log-Pearson tipo III, apesar da distribuição log-normal e o método de Hazen, também, apresentassem bons ajustes. Anos mais tarde, em 1976, O US WRC recomendou o uso da distribuição log-Pearson III com coeficiente de assimetria regional, para os casos de amostras curtas.

BEARD (1974), usando 14200 dados de cheias anuais de 300 estações dos USA, verificou o número de estações, nas quais, as cheias de 1000 anos estimadas em cada local, foram excedidas no registro histórico. Para tanto, ele ajustou 8 diferentes modelos (log-normal, com 2 parâmetro; Gumbel, com 2 diferentes estimativas de parâmetros; log-Gumbel; Pearson III; log-Pearson III e log-Pearson III, regional) aos referidos dados. Isto foi feito sob o argumento de que, de um total de 14200 dados, teria aproximadamente 14 (14200/1000) excedências da verdadeira cheia de 1000 anos. Dos métodos utilizados, somente as distribuições log-Pearson-III e log-normal com 2 parâmetros, reproduziram as 14 excedências. Com este teste, BEARD (1974) conseguiu convencer os hidrólogos de que, a distribuição log-Pearson III e a log-normal com 2 parâmetros, são os modelos que melhor se aproximam dos dados de cheias observados nos Estados Unidos.

Na Austrália, um extensivo e comparativo estudo feito por CONWER (1971), utilizando dados de cheias de New South Wales, e por KOPITTKKE et al. (1976), usando dados de cheias de Queensland, concluiu que a distribuição log-Pearson III teve o melhor desempenho entre as distribuições examinadas. O INSTITUTION OF ENGINEERS AUSTRALIAN - I.E.A. (1972) publicou Australian Rainfall and

Runoff, em que a distribuição log-Pearson III é recomendada para a Austrália. MCMAHON & SRIKANTHAN (1981) usando diagramas de momentos do coeficiente de assimetria-Cs com o coeficiente de variação-Cv e da curtose-k com coeficiente de assimetria-Cs ajustaram várias distribuições de probabilidades (log-Pearson III, normal, retangular, gumbel, Pearson-V, Weibull, gama, exponencial e log-normal-2 parâmetros) aos picos de descargas anuais de 172 cursos d'água da Austrália. Os resultados indicam que a distribuição log-Pearson III é apropriada para análise de frequência de cheias.

CICIONI et al. (1973) consideraram as distribuições: log-normal (com 2 parâmetros), log-normal, Pearson tipo III e a distribuição GEV (todas com 3 parâmetros) para o ajuste de dados de cheias de 108 estações da Itália, com registros acima de 27 anos. A qualidade dos ajustes foi avaliada com a aplicação dos testes qui-quadrado, Kolmogorov-Smirnov (KS), Cramer-Von Mises e Anderson-Darling. A distribuição log-normal com 2 parâmetros foi o melhor ajuste segundo o teste qui-quadrado, enquanto a GEV e log-normal com 3 parâmetros foram os melhores ajustes de acordo com os demais testes utilizados.

ROSSI (1984) com base na análise estatística de 35 séries de cheias anuais da Itália sugeriu que a distribuição com duas componentes de valores extremos (TCEV) poderia ser assumida como a distribuição mãe de cheias, isto é, a distribuição TCEV seria representativa da experiência de cheias reais.

O NATIONAL ENVIRONMENT RESEARCH COUNCIL - NERC (1975) do Reino Unido analisou dados de cheias de aproximadamente 30 estações do Reino Unido, com registros acima de 30 anos, dos quais, apenas quatro excederam os 60 anos. 7 distribuições foram testadas, 4 com 2 parâmetros (log-normal, Gumbel, gama e log-normal) e o restante com 3 parâmetros (GEV, Pearson III e log-Pearson III). Os testes qui-quadrado e Komogorov-Smirnov mostraram-se insensíveis aos afastamentos das distribuições adotadas para os tamanhos das amostras utilizadas. O estudo foi repetido usando as diferenças entre os pontos plotados e as distribuições teóricas. Verificou-se que os resultados se mostraram sensíveis às fórmulas de posição de plotagem e, embora, a distribuição log-Pearson III tenha se revelado melhor com o uso da fórmula de posição de plotagem de Weibull, a distribuição GEV foi a mais estável para todas as fórmulas. Com base neste resultado, a distribuição GEV foi recomendada para ser usada em análise de frequência de cheias

no Reino Unido. Na Alemanha Ocidental as distribuições Pearson III e log-Pearson III foram recomendada para uso geral (BOBÉE et al., 1993).

HOUGHTON (1978) introduziu o uso da distribuição Wakeby para análise de frequência de cheias. Esta distribuição é bem mais complexa que as demais utilizadas em análise de frequência, pois não permite escrever sua função de verossimilhança, além disso, possui 5 parâmetros a serem ajustados. Segundo CLARKE (1994) esta distribuição ainda não tem emprego em hidrologia, sendo assunto de pesquisa, pois não está claro se o seu uso terá benefícios práticos que justifique sua complexidade.

PINHEIRO et al. (1981) realizaram análise de frequência das vazões diárias máximas anuais, para 316 postos fluviométricos das regiões Sul e Sudeste do Brasil, utilizando várias distribuições de probabilidade usualmente empregadas em análise de frequência de cheias. Para medir a eficiência do ajuste das distribuições, foi adotado um método de amostras particionadas. Na estimação dos parâmetros das distribuições: log-normal com 2 parâmetros, log-normal com 3 parâmetros, Gumbel, GEV, Pearson III e gama com 2 parâmetros utilizaram-se os métodos dos momentos e da máxima verossimilhança. Para a distribuição log-Pearson III foram utilizados os momentos dos logaritmos e máxima verossimilhança. No caso da distribuição normal usou-se, apenas, o método dos momentos. Adotou-se ainda os momentos com peso probabilísticos para a distribuição Gumbel. Em média, a distribuição log-Pearson III apresentou o melhor ajuste, sendo recomendado o seu uso para as regiões estudadas, para estimativa de vazões com tempo de recorrência da ordem do comprimento dos registros.

A ELETROBRAS (1987) desenvolveu pesquisa com o objetivo de definir os métodos mais robustos de análise de frequência de cheias, tendo em vista a determinação de cheias de projeto de vertedores no Brasil. O estudo indicou preferencialmente a distribuição exponencial de 2 parâmetros, baseada na hipótese de que a assimetria de população situa-se entre 1,0 e 2,5. Porém, a distribuição Gumbel pode ser utilizada quando a assimetria da população se encontrar próxima ao limite inferior do citado intervalo.

WALLIS & WOOD (1985) criticaram o uso do procedimento da log-Pearson III. Através de simulação no computador, eles mostraram que a distribuição log-Pearson III regional tem desempenho pior do que Wakeby regional e a GEV, sobretudo, quando a distribuição da população foi log-Pearson III. Em experimento

semelhante, HAKTANIR (1992) gerou amostras de 30 e 100 anos, utilizando a log-Pearson III, GEV e Wakeby como população, sobre as quais, ajustou várias distribuições de probabilidade. Comparando as cheias de 100, 1000 e 10000 anos, estimadas através de vários métodos com seus valores populacionais, ele encontrou que a log-normal, com 3 parâmetros, teve melhor desempenho em relação as três distribuições mãe, seguido da log-normal com 2 parâmetros, Gumbel, log-Pearson III e Pearson III.

HOSKING et al. (1985a) mostraram que estimadores de momentos de pesos probabilísticos da distribuição GEV, equivalentes aos estimadores de momentos L (HOSKING, 1990), são melhores do que estimadores de máxima verossimilhança em termos de tendência de estimação e variância para amostras, variando de 15 a 100 anos.

Usando diagrama de momentos L, PEARSON et al. (1991) e NATHAN & WEINMANN (1991) observaram que a distribuição GEV foi o modelo que melhor se aproximou da distribuição de fluxos de cheias do sudoeste da Austrália e Victoria central respectivamente.

GUNASEKARA E CUNNANE (1992) repetiram o experimento de BEARD (1974) com dados sintéticos, referentes as amostras com 40 eventos. Eles experimentaram o ajuste de 13 métodos aos dados gerados, através de 7 diferentes distribuições populacionais, concluindo que a distribuição GEV com estimação dos parâmetros através do método dos momentos de pesos probabilísticos, foi a que melhor se ajustou, no local, para estimar cheias de 100 e 1000 anos, utilizando amostras de 20 eventos. A distribuição log-Pearson III com coeficiente de assimetria regional gerou resultados compatíveis. Esses dois métodos foram os mais robustos sobre as distribuições populacionais consideradas, embora, para um mesmo local, a log-Pearson III tenha se mostrado tendenciosa.

VOGEL et al. (1993a) usando diagrama de momentos L mostraram que dados de cheias de 383 locais no sudoeste dos Estados Unidos, foram igualmente bem aproximados através das distribuições: log-Pearson III, log-normal com 3 parâmetros e GEV. Enquanto que na Austrália, VOGEL et al. (1993b) mostraram, através do diagrama de momentos L, que dados de cheias de 61 estações com registros maiores ou iguais a 20 anos, foram bem aproximados através da distribuição GEV, nas regiões da Austrália que receberam mais chuvas durante os meses de

inverno e, através da distribuição geral de Pareto, nas regiões que receberam mais chuvas durante os meses de verão.

ÖNÖZ E BAYAZIT (1995) avaliaram registros de cheias de 19 estações, variando entre 60 e 165 anos, dos quais, 9 pertencem ao Reino Unido e o restante foram fornecidas pela UNESCO. Neste estudo, foram consideradas 7 distribuições de probabilidades (normal, log-normal e Gumbel, com 2 parâmetros; log-normal, GEV, Pearson III e log-Pearson III, com 3 parâmetros). Para selecionar a distribuição mais conveniente, foram usados os seguintes procedimentos estatísticos: teste qui-quadrado, teste de Komogorov-Smirnov, teste de coeficiente de correlação do gráfico de probabilidade, teste Anderson-Darling, diagrama de momentos-L e o número esperado de excedência. Embora tenha ocorrido diferenças com respeito ao local e procedimento, a distribuição GEV foi superior às demais distribuições testadas.

As distribuições usadas por ÖNÖZ E BAYAZIT (1995), com exceção da normal, têm sido selecionadas devido ao uso corrente em vários países. A distribuição normal foi incluída para efeito de comparação na avaliação do poder dos testes, pois como é bem conhecido, distribuições de fluxos de cheias estão longe de serem simétricas.

HAKTANIR E BOZDUMAN (1995) investigaram, experimentalmente, o efeito do uso das fórmulas de posição de plotagem: Landwehr [ $P_i=(i-1)/(n-1)$ ]; APL [ $(P_i=i-0.35)/n$ ]; Cunnane [ $P_i=(i-0.40)/(n+0.2)$ ] e Weibull [ $P_i=i/(n+1)$ ] na determinação dos parâmetros das distribuições de probabilidade: log-normal com 3 parâmetros, log-Pearson III, GEV e Wakeby, através dos momentos de pesos probabilísticos. O método da máxima verossimilhança também foi aplicado às 3 primeiras distribuições que têm três parâmetros. No total, 19 métodos foram aplicados. Com base na estimação das cheias para períodos de retornos de 100, 1000 e 10000 anos, concluíram que a GEV e a log-Pearson III ajustadas através do método dos momentos de pesos probabilísticos, usando a fórmula de posição de plotagem de Landwehr foram os melhores modelos. Contudo, a log-normal com 3 parâmetros, ajustada através da máxima verossimilhança ou dos momentos de pesos probabilísticos, utilizando a fórmula de plotagem APL ou Landwehr, se mostrou tão boa quanto as duas anteriores. Os autores não encontraram nenhuma evidência convincente da superioridade da Wakeby sobre as demais distribuições utilizadas.

WANG (1996) mostrou através de simulação Monte Carlo que o ajuste da distribuição GEV, utilizando momentos de pesos probabilísticos parciais, em amostras de dados censurados na faixa de valores mais baixos, não reduz a eficiência da estimação de altos quantis de cheias, sobretudo, se a amostra vem de uma distribuição GEV verdadeira.

MKHANDI et al. (1996) avaliaram o desempenho das distribuições: normal, exponencial, Gumbel, log-normal, Pearson III, log-Pearson III, gama com 2 parâmetros, GEV, log logístico e Wakeby, utilizando os métodos de estimação de parâmetros: momentos, máxima verossimilhança e momentos de pesos probabilísticos para ajustar dados de cheias de 53 estações da Tanzânia. Eles concluíram que a log-Pearson III, GEV e log logístico através dos momentos de pesos probabilísticos descreveram igualmente bem os dados analisados.

WANG (1997a,b) demonstrou a eficiência dos métodos de estimação de parâmetros: momentos de altos pesos probabilísticos e momentos LH, propostos por ele, no ajuste da distribuição GEV aos dados de cheias da Austrália. Em trabalhos como DUPUIS (1997), HOYBYE & IRITZ (1997), MEIGH et al. (1997), MULLER & FILL (1997) percebe-se o grande interesse no uso da distribuição GEV em estudos de cheias máximas anuais. MADSEN et al., (1997a,b) descrevem a relativa precisão da estimação de quantis de cheias, obtidos com o ajuste da GEV às séries parcial e total de cheias máximas, utilizando momentos, momentos L e máxima verossimilhança para estimação dos parâmetros.

A distribuição GEV tem, também, recebido especial atenção para a análises de frequência regional desde a introdução do procedimento index flood baseado em momentos de pesos probabilísticos por GREIS & WOOD (1981). Desde então, vários trabalhos têm empregado a distribuição GEV neste tipo de análise (HOSKING et al., 1985b; WALLIS & WOOD, 1985; LETTENMAIER et al., 1987; CUNNANNE, 1988; HOSKING & WALLIS, 1988; CHOWDHURY et al., 1991; STEDINGER & LU, 1995, MADSEN et al., 1987b)

MARTINS & STEDINGER (2000) citam o emprego da GEV no ajuste de vários máximos naturais como: cheia anuais, temperatura, chuva, velocidade do vento, altura de ondas do mar, profundidade de camadas de neve e outros máximos. Além de mostrar a substancial qualidade de estimação de quantis da GEV, em

amostras pequenas, utilizando a máxima verossimilhança generalizada para a faixa de valores do parâmetro de forma  $k$  entre  $-0.4$  e  $0$ .

Segundo LAWLESS (1982), quando as inferências são feitas baseadas em um modelo estatístico, é importante satisfazer-se da conveniência do modelo. No mínimo, o modelo deve ser coerente com os dados, indicados através de testes de qualidades de ajuste e outros procedimentos de avaliação. Deve-se, também, reconhecer até que ponto as inferências dependem do modelo assumido, até mesmo se o modelo se ajusta aos dados adequadamente. Quando o tamanho da amostra é pequeno ou moderado é quase sempre impossível discriminar, entre vários modelos alternativos, tais evidências. Daí, a importância de se saber os efeitos do afastamento das suposições nas inferências realizadas.

Apresenta-se a seguir através da Tabela 1, um resumo da várias distribuições mais utilizadas na análise de frequência de cheias segundo BOBÉE et al. (1993).

TABELA 1 - Algumas distribuições mais utilizadas em análise de frequência de cheias.\*

Distribuição	Comentário	Referência
Normal		Horton (1913)
Log-normal (2 parâmetro)		Hazen (1914)
Pearson tipo I e III	Curvas estabelecidas em convenientes formas para AFC	Foster (1924)
Gumbel		Gumbel (1941)
Log-normal (3 parâmetros)	A distribuição de EVI é virtualmente um caso especial da LN3	Chow (1954)
Valor extremo generalizada	Fórmula única das 3 distribuições de valores extremos recomendado pelo NERC(1975)	Jenkinson (1955)
Log-Pearson tipo III	Recomend. por USWRC (1967)	Beard (1962) Benson (1968)
Gama generalizada	Conhecida pelos hidrologistas como a dist. Kritsky-Menkel	Kritsky-Menkel(1969)
Wakeby	5 parâmetros	Houghton (1978)
Dois componentes de valores extremos	4 parâmetros	Rossi et al. (1984)

\* BOBÉE et al. (1993)

## 2.2 GERAÇÃO DE SÉRIES SINTÉTICAS DE VAZÕES DIÁRIAS

A notável capacidade da distribuição GEV em descrever eventos máximos naturais tem despertado grande interesse no seu uso, para modelação de eventos extremos de variáveis hidrológicas, especialmente, dados de cheias.

A estimação de altos quantis cheias é comumente praticado, extrapolando o ajuste obtido sobre uma série de cheias anuais, para tempos de retorno bem superiores ao período dos dados observados. Como a distribuição GEV apresenta três formas assintóticas de distribuição de valores extremos e o número de cheias anuais utilizado na análise de frequência não é suficiente para assegurar tal condição, sempre haverá incerteza sobre a forma assintótica obtida com o ajuste, principalmente, em se tratando de análise preditiva.

Uma maneira de resolver este impasse é através da geração de séries sintéticas de vazões diárias, para a qual já existe alguns modelos, com formas distintas de abordagem do assunto, disponíveis na literatura científica (QUIMPO, 1968; PAYNE et al., 1969; TRIBER & PLATE, 1975; WEISS, 1977; O'CONNELL & JONES, 1978; SARGENT, 1979; YAKOWITZ, 1979, SVANIDZE, 1980; KELMAN, 1980; KELMAN et al., 1983; KELMAN, 1987; PORTELA & SANTOS, 1989; MÜLLER, 1998; etc.). Há, no entanto, necessidade de se estabelecer metodologias apropriadas para simular vazões diárias, com base nos registros históricos, condizentes com a realidade dos processos físicos hidrológicos que ocorrem na natureza. Assim, existe interesse em comparar os resultados de vazões máximas obtidos das várias formas de composição das vazões diárias, desenvolvidas com base na teoria de processos estocásticos.

A geração de séries sintéticas de vazões, largamente utilizadas em estudos de planejamento, operação e aproveitamento de sistemas de recursos hídricos, busca reproduzir um conjunto de propriedades estatísticas que são encontradas nos registros históricos de vazões observadas. Logo, uma série sintética de vazão pode ser vista como uma série de vazões simulada que não ocorreu, mas que, baseada em certas considerações estatísticas, poderia ter sido um registro real. Sob este ponto de vista, FIERING (1961) definiu hidrologia sintética ou hidrologia operacional.

Por outro lado, uma série de vazões observadas em um determinado local, pode ser vista como sendo o resultado de um "sorteio" efetuado pela Natureza, cujos

valores não serão exatamente repetidos no futuro. Contudo, os valores futuros ocorrerão com a mesma probabilidade, desde que sejam mantidas as mesmas condições hidrológicas da bacia hidrográfica. Assim, as técnicas de modelagem estocástica de vazões diárias buscam reproduzir o mecanismo pelo qual a Natureza realiza esses “sorteios”.

Segundo KELMAN (1983), desenvolver hipóteses sobre as equações que regem este mecanismo é o mesmo que construir um modelo estocástico para as vazões. Neste sentido, se o modelo estocástico estiver bem caracterizado, será possível gerar séries sintéticas de vazões. Estas séries sintéticas serão diferentes entre si e também diferente da série histórica. No entanto, a hipótese subjacente a esta abordagem é que, em lugar da série histórica, qualquer uma das séries sintéticas poderia ter sido sorteada no passado. O mesmo pode ser dito com relação ao futuro. Desta forma, supõe-se que a série histórica e as séries sintéticas formam um conjunto de elementos igualmente prováveis.

Sobre essas premissas, os modelos estocásticos de vazões diárias tem-se desenvolvidos, além de considerar, que as vazões diárias, consideradas como variáveis aleatórias, têm distribuição de probabilidade assimétrica, usualmente com assimetria positiva; seus valores são fortemente correlacionados, e, em séries de sucessivos valores, numerosos máximos locais podem ser encontrados, os quais são precedidos de períodos de rápidas ascensões com curta duração seguida de recessão lenta.

### 2.2.1 Modelos de Vazões Diárias

A simulação de vazões diárias pode ser feita, utilizando-se de dados observados de vazão ou de precipitação diária. Os modelos que utilizam diretamente de vazões observadas compõem os chamados métodos de modelagem direta ou métodos diretos. Já os modelos que utilizam séries de precipitações diárias como dados de entrada de modelos conceituais de transformação de chuva em vazões, constituem os chamados métodos de modelagem indireta, ou simplesmente, método indiretos. Como as proposições de realização deste trabalho têm como base a utilização de dados de vazões observadas ou simuladas a partir das observadas, são tratados, apenas, os métodos diretos.

Os modelos de geração de vazões diárias que fazem parte dos métodos diretos podem ser classificados como (KELMAN, 1987a, MÜLLER, 1998):

- a) Modelos de geração de vazões diárias baseados em simuladores de vazões mensais;
- b) Modelos de desagregação ou de fragmentos;
- c) Modelos não-paramétricos;
- d) Modelos de pulsos, que são subdivididos em:
  - Modelos “shot – noise”;
  - Modelos de pulsos empíricos;
  - Modelos baseados na transição de probabilidade “dia seco” “dia chuvoso”.

- a) Modelos de geração de vazões diárias baseados em simuladores mensais:

A utilização de abordagens empregada na modelação de vazões mensais para geração de vazões diárias, corresponde as primeiras tentativas de modelagem de vazões diárias (KELMAN, 1987a; WEISS, 1977). Enquadram-se neste cenário os modelos desenvolvidos por HALTER & MILER (1967) e BEARD (1967), bem como, os modelos auto-regressivos propostos por QUIMPO (1967) e por PAYNE et al., (1969).

HALTER & MILLER (1967) desenvolveram um simulador de vazão diária, usando um modelo de regressão linear que gerava 30 valores de vazões para cada mês, com base na média e desvio padrão das vazões mensais. As hidrógrafas simuladas não são adequadas porque o gerador não incorpora a correlação serial entre as vazões sucessivas, com exceção das curvas de recessão. As vazões seguem uma curva de recessão quando uma vazão gerada é maior do que um valor assumido para ser o mais alto. A grande inconveniência deste modelo é a grande variação entre as vazões diárias geradas.

O modelo proposto por BEARD (1967) gera vazões diárias durante as estações chuvosas, utilizando uma cadeia de Markov de segunda ordem e as características das frequências históricas das vazões diárias dentro do mês. O simulador proposto é uma extensão de um simulador desenvolvido pelo autor em 1965, com a diferença de que as vazões históricas são rearranjadas e a simulação de vazões mensais não é necessária.

QUIMPO (1968) propôs um modelo estocástico que corresponde a uma adaptação de modelos auto-regressivos de ordem 1, 2 e 3 para a simulação de vazões diárias. A utilização de modelos auto-regressivos para modelagem de vazões diárias, esbarra na necessidade da definição da distribuição de probabilidade para o ruído Gaussiano. Além disso, as séries são geradas com média nula e desvio padrão unitário, onde posteriormente serão multiplicadas pelo desvio padrão e somadas às médias de longo período. Isto tornou inviável a aplicação deste tipo de modelo para o caso da modelação de vazões diárias, devido aos valores diversos da média e desvio padrão para cada dia do ano. Para resolver este problema, os valores da média e do desvio padrão foram calculados, ajustando-se séries de Fourier respectivamente aos valores da média e desvio padrão amostrais para todos os dias do ano.

Como inconveniências deste modelo, destacam-se as dificuldades de ajuste das distribuições empíricas do ruído, decorrente da alta assimetria e curtose das distribuições históricas. Isto torna difícil conseguir uma função de transformação que mapeie vazões diárias numa variável normalmente distribuída. Além disso, as hidrógrafas simuladas não apresentam ascensão abrupta acompanhada de recessão suave, como normalmente ocorre com as séries históricas (KELMAN, 1983; 1987a).

O modelo de simulação diária proposto por PAYNE et al., (1969) é baseado no rearranjo das vazões históricas, onde as ascensões e recessões da série histórica são retiradas do lugar de origem e recolocadas nos dias de maior probabilidade de ocorrência. Em seguida é utilizado um modelo log-Pearson do tipo III para gerar novas séries de vazões. O modelo de PAYNE et al., (1969) foi originalmente testado no rio Calapooia – USA, que apresenta forte sazonalidade. Em geral, o modelo é considerado de aplicação específica (KOTTEGODA, 1972), com maior dificuldade de aplicação em rios de sazonalidade fraca. O modelo apresenta hidrógrafas mais irregulares que as do histórico (MÜLLER, 1998).

#### b) Modelos de desagregação ou de fragmentos

Essa classe de modelos geram vazões diárias a partir de séries de vazões discretizadas em intervalos de tempo maiores do que um dia. Geralmente, utilizam-se séries mensais ou anuais geradas por outro modelo qualquer. Os modelos de

desagregação foram introduzidos para preservar propriedades estatísticas em mais de um nível de desagregação (HARMS & CAMPBELL, 1967).

HARMS & CAMPBELL (1967) apresentaram um modelo, considerado o primeiro modelo de desagregação, que é na verdade uma extensão do algoritmo de geração seqüencial de vazões utilizado por THOMAS & FIERING (1962). No processo de modelagem, o modelo de THOMAS & FIERING (1962), um auto-regressivo de ordem 1 mensal, é usado para gerar vazões mensais e um modelo auto-regressivo de ordem 1 para gerar vazões anuais. O modelo de HARMS & CAMPBELL (1967) multiplica as vazões mensais geradas por THOMAS & FIERING (1962) por um fator de escala, transformando-as em vazões anuais com o uso do AR(1) anual. Na verdade este método faz o processo inverso da desagregação, onde a agregação de fluxos mensais gerados rende fluxos anuais simulados.

Assim, o modelo de VALENCIA & SCHAAKE (1973) de desagregação de vazões anuais em mensais é aceito unanimemente como o primeiro modelo de desagregação proposto. O modelo tem a seguinte forma geral:

$$Y = AX + B\epsilon$$

Onde:

**A**, **B** são matrizes de parâmetros;

**X** – vazão anual gerada com média nula;

**Y** – vetor com as vazões mensais desagregadas com média nula das vazões previamente geradas;

$\epsilon$  - vetor de variáveis aleatórias com distribuição normal padrão.

As notações em negritos representam formas matriciais.

A partir da proposição de desagregação de cheias feita por VALENCIA & SCHAAKE (1973), vários modelos de desagregação de fluxo anual em mensal têm sido propostos (MIJIA & ROUSSELLE, 1976; TAO & DELLER, 1976; SRIKTHAN, 1978; SALAS et al., 1980), além dos modelos desenvolvidos para dividir fluxos anuais ou mensais agregados a uma determinada bacia hidrográfica em fluxos referentes a locais específicos (SALAS et al., 1980). Outros modelos de desagregação com várias melhorias e redução do número de parâmetros foram

sugeridos, como STEDINGER et al., (1985), GRYGIER & STEDINGER (1988), SANTOS & SALAS (1992), KOUTSOYIANNIS (1992), BARTOLINI & SALAS (1993) e KOUTSOYIANNIS & MANETAS (1996). Contudo, esses modelos têm seus enfoques na desagregação, no tempo ou no espaço, de vazões anuais para vazões mensais ou semanais.

Para simulação de vazões diárias, KOTTEGODA (1972) propôs um modelo de geração de séries de vazões médias de cinco dias. Com base neste modelo, GREEN (1973) desenvolveu um modelo que desagrega essas vazões médias através de procedimentos sofisticados de interpolações. GREEN (1973) coloca que seu modelo não reproduz adequadamente as vazões máximas.

SVANIDZE (1980) desenvolveu um método de fragmentos, que desagrega vazões anuais em mensais ou em intervalos menores, que pode ser até de horas. O método realiza uma dupla amostragem das vazões, onde a vazão média  $X(i)$  corresponde a primeira amostragem para o ano  $i$  e os fragmentos  $y_{adm}(i,k)$ , que representam os fluxos observados no passado com intervalo menor que um ano adimensionalizados, a segunda amostragem para o ano  $i$  e o mês  $k$ . Multiplicando-se a vazão média anual  $X(i)$ , gerada por um outro modelo, pelos fragmentos  $y_{adm}(i,k)$  obtém-se um novo hidrograma das vazões média com períodos menores que um ano. Para o modelo de fragmentos diários (SVANIDZE, 1980), primeiro são calculados os fragmentos adimensionalizados  $q_{adm}(l,k)$ , onde  $l$  é o ano e  $k$  o dia da série histórica. Esses fragmentos são obtidos dividindo a série diária histórica de vazões  $q(l,k)$  pela média anual  $x(l)$  do ano em que as vazões  $q(l,k)$  ocorreram. Deste modo, para cada ano  $l$  do histórico tem-se um vetor fragmento  $q_{adm}(l,k)$  com 365 valores adimensionalizados.

PORTELA & SANTOS (1989) apresentaram um modelo que procede a desagregação de vazões anuais em mensais e depois realiza a desagregação das vazões mensais geradas em intervalos diários, seguindo um esquema de desagregação que corresponde a uma simplificação do modelo de MIJÍA & ROUSSELLE (1976). A complexidade deste modelo não é compatível com os resultados obtidos, além disso, o mesmo não leva em consideração as características próprias encontradas na geração de vazões diárias (MÜLLER, 1998).

MÜLLER (1998) propôs os modelos de fragmentos MSD (Modelo Simplificado de Desagregação) e MMD (Modelo Modificado de Desagregação) que

são uma variação do modelo dos fragmentos de SVANIDZE (1980), com as seguintes diferenças básicas: a série  $X(i)$  é mensal e não anual; o método MMD é indicado para que não ocorram saltos na hidrógrafa do fim de um mês para o início do outro. Além disso, o modelo proposto gera uma maior variabilidade de séries intra-ano do que no caso do método dos fragmentos original. Outro fato é que os fragmentos correspondente a segunda amostragem, podem ter como origem uma série sintética gerada por outro modelo.

### c) Modelos não-paramétricos

Nesta classe de modelos enquadra-se o modelo de geração de vazões diárias proposto por YAKOWITZ (1979). Para a geração de séries sintéticas de vazões diárias, YAKOWITZ (1979) adotou um procedimento baseado na utilização da distribuição de probabilidade empírica da vazão diária  $q_t$  condicionada aos valores imediatamente anteriores,  $\mathbf{q} = (q_{t-1}, q_{t-2}, \dots, q_{t-\tau})$ , onde  $\tau$  é a memória arbitrada para o processo. Para isso, YAKOWITZ (1979) define um conjunto finito de vetores representativos  $\mathbf{q}(1), \mathbf{q}(2), \dots, \mathbf{q}(m)$  de tal forma que qualquer vetor  $\mathbf{q}$ , da série histórica ou gerada, possa ser classificado numa das  $m$  categorias de acordo com a distância entre  $\mathbf{q}$  a cada um dos  $\mathbf{q}(i)$ ,  $i=1, m$ . Desta forma:

$$\| \mathbf{q}(j) - \mathbf{q} \| \leq \| \mathbf{q}(i) - \mathbf{q} \|, \quad \forall i$$

diz-se que o vetor  $\mathbf{q}$  tem distância mínima para o  $j$ -ésimo vetor representativo. Neste caso, a vazão do próximo dia  $q_t$  é dita pertencer ao conjunto  $C_j$ ,  $j=1, m$ . Classificando-se todos os vetores  $\mathbf{q}_t$  da série histórica de acordo com o procedimento acima é possível caracterizar a distribuição de probabilidade empírica associada a cada conjunto  $C_j$ , definida por  $F_j(x) = k_j/n_j$ , onde  $k_j$  é o número de elementos do conjunto  $C_j$  menores ou iguais a  $x$  e  $n_j$  é o número total de elementos de  $C_j$ .

A geração de uma série sintética se processa sorteando-se sucessivamente variáveis aleatórias independentes e uniformemente distribuídas entre 0 e 1,  $w_t$ , onde  $q_t$  é dado pela forma inversa de  $F_j(w_t)$ , sendo  $j$  definido como descrito acima, ou seja, dependente de  $q_{t-1}, q_{t-2}, \dots, q_{t-\tau}$ .

Os modelos de desagregação apresentados na secção anterior, bem como o modelo Diana, mostrado a seguir, também constituem modelos não-paramétricos.

d) Modelos de pulsos:

Os modelos que fazem parte desta classificação geram ascensões rápidas da hidrógrafa a partir da simulação de um pulso, utilizando um procedimento qualquer, a qual representa um pico de cheia provocado por uma chuva, estimado unicamente através dos registros de vazões. A forma como são gerados os pulsos no processo de modelagem de vazões dá origem a cada tipo de modelo desta classe, como por exemplo, o modelo *shot-noise* gera os pulsos, utilizando funções exponenciais.

*i – modelos shot-noise*

Os modelos shot-noise, segundo WEISS(1977), KELMAN (1987a) e MINE (1990), compõem-se de uma série de eventos, obedecendo a uma lei de Poisson, onde são produzidos pulsos de altura aleatória, superposto a uma equação exponencial ajustada às recessões. Isto constitui um processo regressivo de primeira ordem contínuo no tempo, com distribuição gama dos valores instantâneos. O modelo pode ser ajustado aos dados históricos de tal forma a preservar a média, o desvio padrão e as correlações seriais da série de dados observados. WEISS (1977) propôs ainda os modelos de segunda ordem, que divide os fluxos em duas parcelas, correspondente ao escoamento básico e ao direto ou de superfície.

*ii – modelos de pulsos empíricos*

Nesta categoria enquadram-se o “modelo estocástico de vazões diárias” (KELMAN, 1980), os modelos Diana (KELMAN et al., 1983) e Xadrez (KELMAN, 1987b) que têm como gerador de pulsos nos dias considerados chuvosos a distribuição empírica de  $Q(t)/Q(t-1)$  ou  $Q(t)-Q(t-1)$ , conforme o caso, onde  $Q(t)$  é a vazão diária observada no dia  $t$ . O modelo Diana representa um aperfeiçoamento do modelo ‘estocástico de vazões diárias’, onde as diferenças entre os mesmos estão descritas em KELMAN(1987a,b). O modelo Xadrez tem como componentes

elementares os “saltos de vazão”, definidos como a diferença de vazão entre dois dias consecutivos. Este método foi proposto com o propósito de prover um mecanismo gerador de séries sintéticas diárias mais rápida, sob o ponto de vista computacional.

*iii – modelos baseados na probabilidade de transição “dia seco - dia chuvoso”*

Os modelos baseados na probabilidade de transição *dia seco - dia chuvoso* tiveram início através do trabalho de TREIBER & PLATE (1975). Este autores desenvolveram um modelo com função de transferência baseada na probabilidade de transição dia seco – dia chuvoso calculada a partir das vazões observadas, em que o pulso pode ou não ocorrer. TREIBER & PLATE imaginaram que a seqüência de dias com e sem pulso pode ser modelada por uma cadeia de Markov de dois estados. Para cada dia chuvoso, um pulso é gerado, utilizando uma função exponencial modificada que corresponde a distribuição de Weibull. O pulso é então adaptado a função de transferência para então obter a seqüência de vazões para os dias onde não ocorrem pulsos. Esta função de transferência é determinada, ajustando-se uma função polinomial às recessões, identificada na série histórica, pelo método dos mínimos quadrados. Neste modelo os pulsos não são independentes e compõe um processo AR(1), onde a componente aleatória tem distribuição exponencial. A ocorrência de um dia sem pulso interrompe a realização do processo autoregressivo. Para assegurar o comportamento sazonal das séries sintéticas, os parâmetros são ajustados mês a mês.

O’CONNELL & JONES (1978) consideraram que a vazão do dia  $t$  poderia ser bem representada por uma dependência na vazão do dia anterior expressa por uma função não-linear do tipo:

$$Q_t = g(q_{t-1}) + x_t$$

onde  $g(\cdot)$  representa o contorno inferior aos pares de pontos  $(q_{t-1}, q_t)$  e  $x_t$  é uma variável aleatória não negativa. Para a função  $g(q_{t-1})$  O’Connell e Jones sugeriram a equação mostrada a seguir, cujos parâmetros  $a$ ,  $b$  e  $c$  devem ser estimados a partir de um ajuste com os dados da série histórica.

$$g(q_{t-1}) = (1-c)a \cdot \text{arc Senh}(q_{t-1}/a) + bc \ln(1 + q_{t-1}/b)$$

Para representar a sucessão de dias em que  $x_t$  é nulo ou positivo, os autores recomendaram a adoção de uma cadeia de Markov de dois estados, cujas probabilidades de transição são estimadas da série histórica. Tal qual o modelo de TREIBER & PLATE, os pulsos  $x_t$  compõem um processo AR(1), em que a componente aleatória tem distribuição exponencial (com uma transformação de potência). Também a ocorrência de um dia sem pulso interrompe a realização do processo autorregressivo e os parâmetros são estimados mês a mês, a partir de amostras obtidas da série histórica

SARGENT (1979) apresentou um modelo que utiliza a mesma estrutura básica usada por TREIBER & PLATE (1975), com diferença na função de transferência, onde a forma da recessão é representada por duas funções: uma polinomial, que descreve as recessões com descargas mais elevadas e outra exponencial, representando a parte da recessão com valores de descarga baixos. O modelo gera inicialmente os dias onde deve ou não ocorrer pulsos, usando uma cadeia de Markov de dois estados. Para cada dia chuvoso, um pulso é gerado utilizando uma distribuição exponencial modificada. O pulso é então adaptado a uma das funções de transferência para obter a seqüência de vazões para os dias onde não ocorrem pulsos. Posteriormente é feito um ajuste, segundo o modelo de HARMS & CAMPBELL (1967), das vazões diárias geradas às vazões mensais previamente geradas.

### 3 OBJETIVOS

A distribuição GEV tem encontrado muitas aplicações em hidrologia desde sua recomendação para a análise de frequência de cheias no Reino Unido (NERC, 1975). Para a análise de frequência regional, a referida distribuição tem recebido especial atenção desde a introdução do procedimento index-flood baseado em momentos pesos probabilísticos por GREIS & WOOD (1981). Na prática, a mesma vem sendo usada para modelar uma extensa variedade de extremos naturais, incluindo cheias anuais, chuvas, velocidade do vento, alturas de ondas do mar, profundidade da camada de neve e outros eventos naturais (MARTINS & STEDINGER, 2000).

O grande interesse na aplicação da distribuição GEV para descrever máximos naturais, decorre da consideração teórica de que a distribuição de valores extremos de uma variável, corresponde a uma das três formas assintóticas descritas pela distribuição GEV (WANG, 1997). Do ponto de vista matemático, ela é muito atraente devido sua função inversa (função de quantis) apresentar-se numa forma bem definida, além disso, seus parâmetros são facilmente estimados através de momentos com combinação linear das estatísticas de ordem chamado de momentos L (HOSKING et al., 1985a, HOSKING, 1990), ou mais genericamente, por meio de momentos com combinação linear das estatísticas de ordens mais elevadas ou simplesmente momentos LH (WANG, 1997).

Na modelação de cheias anuais através da distribuição GEV, os altos quantis de cheias são estimados, extrapolando-se a função inversa da distribuição ajustada a uma série de cheias anuais, para tempos de retorno bem superiores ao período de observação dos dados. Quando o ajuste resulta na forma assintótica da distribuição de valores do tipo I (Gumbel) ou do tipo II, a função de quantis de ambas formas, apresenta-se ilimitada na faixa dos valores mais elevados. Já a função de quantis

referente a forma assintótica da distribuição dos valores extremos do tipo III, possui limite superior finito.

Como os eventos hidrológicos que ocorrem dentro da bacia hidrográfica são originários de fenômenos meteorológicos, que acontecem na natureza dentro de valores fisicamente limitados. Espera-se que a distribuição dos máximos de eventos hidrológicos sigam a forma assintótica da distribuição GEV limitada. Segundo JENKINSON (1955) esta é a distribuição de valores extremos mais encontrada na natureza.

Sobre o aspecto teórico da distribuição de frequência preditiva, a distribuição GEV limitada se apresenta como a mais apropriada para descrever cheias anuais. GUMBEL (1958) definiu um modelo de distribuição de probabilidade subjacente da distribuição GEV limitada, denominando-a de distribuição limitada.

Do ponto de vista prático, como o proceder ajuste da distribuição GEV, que é assintótica, a partir de séries de cheias com poucos anos de observação dos dados? A estimação de parâmetros da distribuição assintótica de uma variável aleatória, utilizando amostra pequenas, é duvidosa, por mais eficiente que seja o método de estimação empregado. Neste caso, qualquer procedimento para fazer previsões a longo prazo (eventos raros), tendo em mãos uma série de dados com poucos registros, não é confiável.

Como existe uma relação teórica entre os parâmetros da distribuição inicial e os parâmetros da distribuição de extremos, significa que a distribuição inicial pode ser inferida com base em um número muito mais abrangente de dados (número de dias da amostra) do que a distribuição de extremos que é estimada a partir de poucos dados (número de anos da amostra). Desta forma, a utilização da relação teórica entre os parâmetros da distribuição GEV limitada e a de sua distribuição subjacente (distribuição limitada), num processo de simulação Monte Carlo de vazões diárias, pode revelar os erros e incertezas já que a resposta via teoria é conhecida.

A geração estocástica de vazões diárias é mais adequada quando são utilizados modelos estocásticos diários, disponíveis na literatura, que simulam seqüências de vazões mais próximas da realidade dos eventos hidrológicos. Assim, a sazonalidade presente nos registros históricos diários é reproduzida de maneira mais eficiente do que com o uso da distribuição limitada. No entanto, com este

procedimento não é possível estabelecer uma relação teórica entre as distribuições inicial e de extremos através de seus parâmetros, uma vez que a distribuição verdadeira de extremos é desconhecida, não se sabendo, inclusive, o tipo de distribuição de extremos que pode ocorrer.

Diante do que foi colocado, fica evidente a potencialidade da distribuição GEV para modelação de cheias, com a possibilidade de utilização da sua forma limitada para descrever os máximos de eventos hidrológicos, realisticamente. Entretanto, existe a necessidade de mais informações a respeito da estimabilidade da distribuição, considerando as incertezas e os erros de estimação decorrentes da limitação dos dados disponíveis, usados para ajustar uma distribuição que é assintótica. O emprego da hidrologia operacional resolve o problema da extensão da série de cheias anuais, a partir de séries sintéticas de vazões diárias obtidas com base nos registros históricos. Não obstante, necessita-se de metodologia que assegure a construção de seqüências de cheias anuais condizentes com a realidade dos processos hidrológicos naturais. Portanto, este trabalho foi desenvolvido com os objetivos de:

- 1- Analisar a estimabilidade da distribuição GEV através de momentos LH, considerando as incertezas e erros da estimação de seus parâmetros e de altos quantis de cheias, utilizando séries de cheias anuais com diferentes características e comprimentos, obtidas de séries de vazões diárias geradas de diversas maneiras, via simulação Monte Carlo, adotando-se a distribuição limitada como subjacente da distribuição GEV limitada.
- 2- Identificar a forma do ajuste da distribuição GEV às séries de cheias anuais obtidas com a geração de seqüências diárias de longo período, utilizando diversos modelos de simulação de vazões diárias usados na hidrologia operacional.
- 3- Propor um modelo de simulação de vazões diárias capaz de gerar cheias anuais que apresentem convergência assintótica para a forma de valores extremos limitada.

## 4 TEORIA BÁSICA DA ESTATÍSTICA DE VALORES EXTREMOS

O alvo principal da teoria estatística de valores extremos é a análise de extremos observados para prever valores extremos no futuro. Em muitas aplicações de engenharia, valores extremos de uma variável são de especial importância, como por exemplo, em estudos de segurança estrutural de uma obra ou na confiabilidade de uma estrutura; em programas de controle de cheias, onde a cheia máxima de um rio é o fluxo mais significativo, enquanto em trabalhos de irrigação e programas de suplementação de água, o menor fluxo seria o mais relevante.

A previsão de condições futuras, muitas vezes requerida em planejamento e elaboração de projetos de engenharia, pode envolver a estimação do maior ou menor valor de uma variável. Neste caso, a extrapolação de valores extremos previamente observados é invariavelmente necessária. Para este propósito a teoria assintótica da estatística de extremos constitui numa base poderosa para a obtenção da informação de engenharia requerida.

A repetição da série de observações (amostra de tamanho  $n$ ) da variável aleatória inicial resulta em novos valores extremos, os quais são, também, variáveis aleatórias que podem ser modeladas através de suas respectivas distribuições de probabilidades. Essas distribuições e seus parâmetros têm características especiais que são próprias dos valores extremos. O entendimento dos conceitos e características da estatística de extremos é fundamental para sua aplicação em problemas, especificamente, de engenharia. A descrição abaixo segue de perto àquela apresentada em GUMBEL (1958) e ANG & TANG (1984) com alguns detalhamentos para facilitar a compreensão do assunto.

Neste capítulo são apresentados tópicos da teoria básica da estatística de extremos, considerando as distribuições de probabilidades: exatas, assintóticas e de valores extremos generalizadas (GEVs) relacionadas aos valores máximos.

#### 4.1 DISTRIBUIÇÃO DE PROBABILIDADE EXATA DE EXTREMOS MÁXIMOS

Seja  $X$  uma variável aleatória absolutamente contínua independente e identicamente distribuída com função distribuição  $F_X(x)$  e função densidade  $f_X(x)$ . Seja a amostra de tamanho  $n$  tomada, independente e aleatoriamente, de  $X$  corresponde a  $\{x_1, x_2, \dots, x_n\}$ , onde cada valor  $x_i$  observado é uma variável aleatória. Seja a variável aleatória  $y_1, y_2, \dots, y_n$  uma realização amostral da variável aleatória  $X$  dada por

$$\begin{aligned} Y_1 &= \text{menor valor de } \{x_1, x_2, \dots, x_n\} \\ Y_2 &= \text{segundo menor valor de } \{x_1, x_2, \dots, x_n\} \\ &\vdots \\ Y_j &= \text{j-ésimo menor valor de } \{x_1, x_2, \dots, x_n\} \\ &\vdots \\ Y_n &= \text{maior de } \{x_1, x_2, \dots, x_n\} \end{aligned}$$

as observações  $\{x_1, x_2, \dots, x_n\}$  referem-se a uma realização da variável aleatória  $\{Y_1, Y_2, \dots, Y_n\}$ . Definem-se os valores extremos, máximo e mínimo  $Y_n$  e  $Y_1$  respectivamente, os quais são, também, variáveis aleatórias e, portanto, têm distribuição de probabilidade própria. A ordenação  $Y_1 \leq Y_2 \leq \dots \leq Y_j \leq \dots \leq Y_n$  é conhecida como estatísticas de ordem correspondentes às observações da variável aleatória  $X$ ,  $\{x_1, x_2, \dots, x_n\}$ , onde a variável aleatória  $Y_j$  é chamada de  $j$ -ésima estatística de ordem. A função densidade de  $Y_j$  é dada pela seguinte equação:

$$f_{Y_j}(y) = \frac{n!}{(n-j)!(j-1)!} [F_X(y)]^{j-1} [1 - F_X(y)]^{n-j} f_X(y) \quad (1)$$

A integração da equação (1) resulta na função distribuição acumulada de  $Y_j$ , como segue:

$$F_{Y_j}(y) = \frac{n!}{(n-j)!(j-1)!} \int_{-\infty}^y [F_X(y)]^{j-1} [1 - F_X(y)]^{n-j} f_X(y) dy \quad (2)$$

O problema de extremos é encontrar as respectivas funções densidade e distribuição de  $Y_n$  e de  $Y_1$  e dar a conveniente forma assintótica para essas funções, quando  $n$  torna-se grande. Porém, como  $F_{Y_j}(y)$  pode ser deduzida através da noção de que a  $j$ -ésima estatística de ordem é menor ou igual a  $y$ , se e somente se, existir  $j$  ou mais  $x_i$ 's que são menores ou iguais a  $y$ . Portanto, como o número de  $x_i$ 's que são menores ou iguais a  $y$  é uma variável aleatória binomial com parâmetros  $n$  e  $p = F_X(y)$ , segue-se que:

$$F_{Y_j}(y) = P[Y_j \leq y] = P[j \text{ ou mais dos } x_i\text{'s sejam } \leq y]$$

logo

$$F_{Y_j}(y) = \sum_{k=j}^n \binom{n}{k} [F_X(y)]^k [1 - F_X(y)]^{n-k} \quad (3)$$

Em particular, pode-se escrever com algumas simplificações as funções distribuição de  $Y_n$  e de  $Y_1$ . Porém, como o maior interesse neste trabalho é a distribuição dos valores extremos máximos, este estudo fica restrito apenas a estatística  $Y_n$ , o maior valor da amostra, onde sua função distribuição acumulada resulta em:

$$F_{Y_n}(y) = [F_X(y)]^n \quad (4)$$

e a correspondente função densidade de probabilidade, obtida através da derivação da equação (4), em:

$$f_{Y_n}(y) = n f_X(y) [F_X(y)]^{n-1} \quad (5)$$

As equações (4) e (5) descrevem as funções distribuição e a densidade de probabilidade exata de extremos máximos de uma amostra de tamanho  $n$ , as quais, são dependentes de  $F_X(x)$  e do valor de  $n$ .

## 4.2 DISTRIBUIÇÕES ASSINTÓTICAS DE EXTREMOS MÁXIMOS

As características das distribuições de extremos foram estudadas por FISHER & TIPPETT (1927), GNEDENKO (1943), CRAMER (1946) e GUMBEL (1958) que constituem a teoria assintótica da estatística de extremos.

A teoria assintótica das distribuições de extremos diz respeito às formas limites de  $F_{Y_n}(y)$  ou  $f_{Y_n}(y)$  quando o tamanho da amostra  $n$  aumenta, onde  $F_{Y_n}(y)$  pode convergir, em distribuição, para uma particular forma funcional (GUMBEL, 1958; ANG & TANG, 1984).

### 4.2.1 Dedução das Distribuições Assintóticas de Valores Extremos:

A dedução analítica da distribuição assintótica de extremos, dada a distribuição de uma variável  $X$ , considerando a distribuição de valores extremos máximos de  $X$ , é facilitada com o uso da técnica de CRAMER (1946) (GUMBEL, 1958; EPSTEIN, 1960; ANG & TANG, 1984).

CRAMER (1946), considerando a probabilidade elementar  $[g_{Y_n}(y)dy]$  numa distribuição amostral de valores extremos  $y$ , introduziu uma variável aleatória transformada:

$$\xi_n = n[1 - F_X(Y_n)] \quad (6)$$

onde a densidade de probabilidade de  $\xi_n$  converge para um valor maior que zero quando  $n$  tende a infinito. Assim, levando-se em conta  $\xi_n$ , temos que:

$$F_{\xi_n} = P(\xi_n \leq \xi) = P\{n[1 - F_X(Y_n)] \leq \xi\} = P\left[F_X(Y_n) \geq 1 - \frac{\xi}{n}\right]$$

$$F_{\xi_n} = 1 - F_{Y_n}\left[F_X^{-1}\left(1 - \frac{\xi}{n}\right)\right] = 1 - \left\{F_X\left[F_X^{-1}\left(1 - \frac{\xi}{n}\right)\right]\right\}^n = 1 - \left(1 - \frac{\xi}{n}\right)^n$$

Tomando o logaritmo de  $\left(1 - \frac{\xi}{n}\right)^n$  e aplicando o limite quando  $n$  tende a infinito, tem-se:

$$\lim_{n \rightarrow \infty} \ln \left(1 - \frac{\xi}{n}\right)^n = \lim_{n \rightarrow \infty} n \ln \left(1 - \frac{\xi}{n}\right) = \lim_{n \rightarrow \infty} \ln \frac{1 - \frac{\xi}{n}}{\frac{1}{n}} = -\xi$$

$$\lim_{n \rightarrow \infty} \left(1 - \frac{\xi}{n}\right)^n = e^{-\xi}$$

e conseqüentemente, quando  $n \rightarrow \infty$

$$F_{\xi_n}(\xi) = 1 - e^{-\xi} \quad (7)$$

onde a correspondente função densidade de probabilidade é

$$f_{\xi_n}(\xi) = e^{-\xi}. \quad (8)$$

Invertendo a equação (6), tem-se

$$Y_n = F_X^{-1} \left(1 - \frac{\xi_n}{n}\right) \quad (9)$$

Logo, quando o tamanho  $n$  da amostra aumenta, a distribuição assintótica de  $Y_n$  pode ser determinada a partir das equações (7) e (8) com base na relação funcional da equação (9). Observa-se através da equação (6) que quando  $\xi_n$  decresce  $Y_n$  aumenta, por conseguinte:

$$P(Y_n \leq y) = P[\xi_n > g(y)]$$

onde  $g(y)$  é a expressão do lado direito da equação (6), assim a função distribuição acumulada de  $Y_n$  pode ser obtida através da distribuição de  $\xi_n$ , como segue:

$$F_{Y_n}(y) = 1 - F_{\xi_n}[g(y)] = 1 - [1 - e^{-g(y)}] = e^{-g(y)} \quad (10)$$

e, conseqüentemente, a função densidade de probabilidade, da forma seguinte:

$$f_{Y_n}(y) = -\frac{dg(y)}{dy} e^{-g(y)} \quad (11)$$

#### 4.2.2 Aplicações:

Para melhor compreensão do que foi apresentado, seguem-se três aplicações de distribuições de variáveis iniciais:

a) *Considere que a variável  $X$  tem função densidade de probabilidade exponencial:*

$$f_X(x) = \lambda e^{-\lambda x}, \quad \text{logo } F_X(x) = 1 - e^{-\lambda x} \quad \text{com } x \geq 0 \quad (12)$$

que através das equações (6), (10) e (11) conduz a

$$\xi_n = n[1 - (1 - e^{-\lambda x_n})] = ne^{-\lambda x_n} \quad (12a)$$

logo

$$F_{Y_n}(y) = \exp(-ne^{-\lambda y}) \quad (12b)$$

$$f_{Y_n}(y) = n\lambda e^{-\lambda y} \exp(-ne^{-\lambda y}) \quad (12c)$$

Por outro lado, as correspondentes funções distribuição e densidade exatas, obtidas com a aplicação das equações (4) e (5), são:

$$F_{Y_n}(y) = (1 - e^{-\lambda y})^n \quad (13)$$

$$f_{Y_n}(y) = \lambda n(1 - e^{-\lambda y})^{n-1} e^{-\lambda y} \quad ; y \geq 0 \quad (13a)$$

Neste caso, comparando a distribuição assintótica com a distribuição exata, através de seus gráficos apresentados na Figura 1, observa-se que, quando  $n$  cresce, as duas distribuições convergem para mesma forma assintótica.

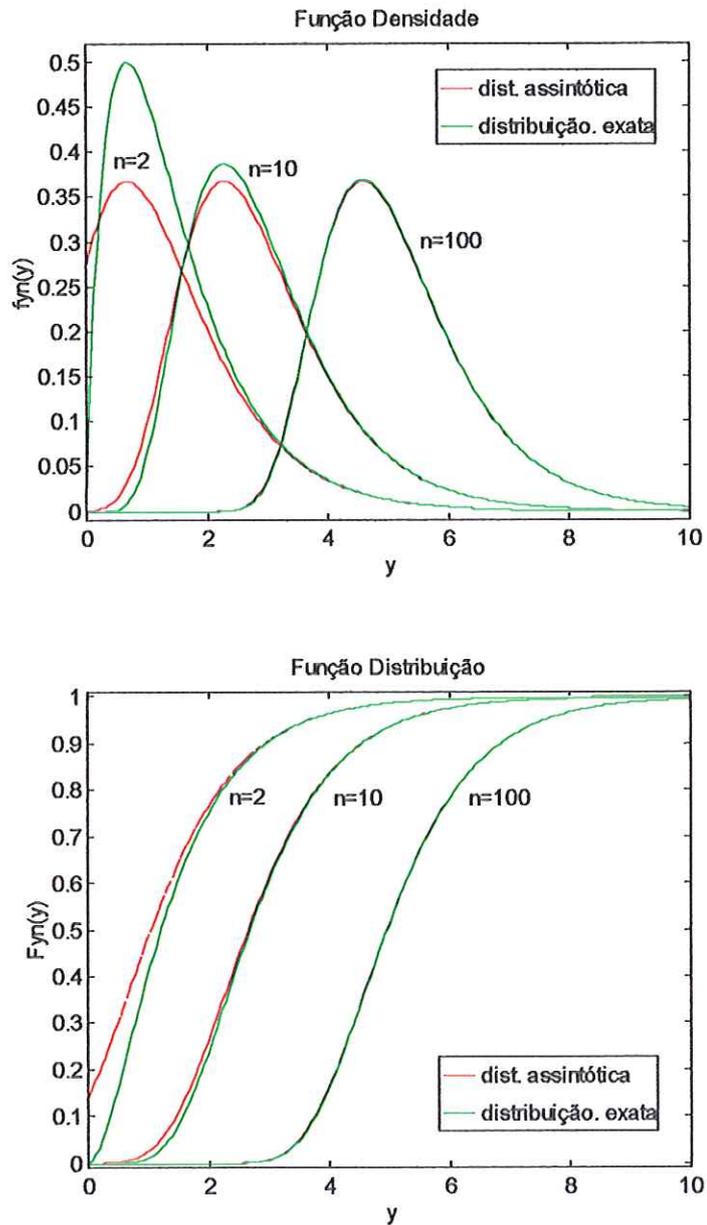


FIGURA 1- Funções densidade e distribuição das formas exata e assintótica de valores extremos máximos de uma distribuição inicial exponencial.

b) Considere a variável  $X$  com a seguinte função densidade de probabilidade:

$$f_X(x) = \frac{k}{x^{k+1}}; \text{ se } x \geq 1 \quad \text{e} \quad f_X(x) = 0 \text{ se } x < 1$$

que resulta na seguinte função distribuição:

$$F_X(x) = 1 - \frac{1}{x^k} \quad \text{se } x \geq 1 \quad (14)$$

Partindo das equações (6), (10) e (11), têm-se a variável transformada e as respectivas funções distribuição e densidade:

$$\xi_n = n \left[ 1 - \left( 1 - \frac{1}{Y_n^k} \right) \right] = \frac{n}{Y_n^k}; \quad (14a)$$

$$F_{Y_n}(y) = \exp\left(-\frac{n}{y^k}\right); \quad y \geq 1 \quad (14b)$$

$$f_{Y_n}(y) = \frac{nk}{y^{k+1}} \exp\left(-\frac{n}{y^k}\right); \quad y \geq 1 \quad (14c)$$

Neste exemplo a função distribuição acumulada é da forma:  $e^{-b/y^k}$ .

c) Considerando agora uma variável aleatória com limite superior finito. Assim, seja  $X$  uma variável aleatória com função distribuição uniforme entre 0 e  $a$ ; logo:

$$f_X(x) = \frac{1}{a}; \quad 0 \leq x \leq a$$

$$f_X(x) = 0; \quad \text{caso contrário}$$

a correspondente função distribuição é:

$$F_X(x) = \frac{x}{a}; \quad 0 \leq x \leq a \quad (15)$$

A variável transformada, a função distribuição e a função densidade de  $Y_n$  são:

$$\xi = n \left( 1 - \frac{Y_n}{a} \right) \quad (15a)$$

$$F_{Y_n}(y) = \exp \left[ -\frac{n}{a}(a-y) \right]; \quad 0 \leq y \leq a \quad (15b)$$

e

$$f_{Y_n}(y) = \frac{n}{a} \exp \left[ -\frac{n}{a}(a-y) \right]; \quad 0 \leq y \leq a \quad (15c)$$

observe que a função distribuição acumulada neste caso é da forma exponencial  $e^{-c(\omega-y)}$ , sendo  $\omega$  o valor superior da variável  $X$ .

Analisando os 3 exemplos dados, percebe-se que a distribuição assintótica de extremos tende a convergir para certa forma limite, quando  $n$  torna-se grande; especificamente, para a forma dupla exponencial como na equação (12b) ou para as duas diferentes formas exponenciais, apresentadas nas equações (14b) e (15b).

#### 4.3 AS TRÊS FORMAS ASSINTÓTICAS DE VALORES EXTREMOS:

A existência das formas assintóticas de distribuições de extremos é baseada na premissa de que a forma de uma distribuição de probabilidade é invariante sob uma transformação linear (*estabilidade postulada*). Assim, a distribuição dos valores máximos da variável  $X$  pode ser expressa como uma função linear de  $X$ , tal como:

$$[F_X(x)]^n = F_X(a_n x + b_n) \quad (16)$$

Onde  $a_n$  e  $b_n$  são funções de  $n$ . Soluções da equação (16) rendem todas as possíveis formas limite de  $F_X(x)$ , quando  $n$  tende para infinito (FISHER & TIPPETT, 1928). Dependendo das funções  $a_n$  e  $b_n$  que satisfazem a equação (16) diferentes formas limites da função  $F_X(x)$  podem ocorrer, onde cada uma assume um comportamento específico para os valores extremos da variável inicial.

Três formas foram determinadas por FISHER & TIPPET, as quais, posteriormente, GUMBEL (1958) considerou como três tipos de formas assintóticas, classificando-as de valores extremos do tipo I, II e III, denotados por VEI, VEII e VEIII.

Na Tabela 2 são apresentados as três formas de distribuições assintóticas com os correspondentes critérios de convergência das funções densidade e distribuição de probabilidade da variável inicial, além de exemplos de distribuição inicial, associados a cada forma de distribuição assintótica.

TABELA 2 - As três formas de distribuição assintóticas determinadas por FISHER & TIPPET (1928) e classificada por GUMBEL(1958)\*.

Distribuição inicial	Forma da distribuição assintótica	Critério de convergência	Tipo de distribuição assintótica
<i>Exponencial</i> $F_X(x) = 1 - e^{-\alpha x}$ $x \geq 0; \alpha > 0$	$F_{X_n}(x) = \exp[e^{-\alpha_n(x-u_n)}]$	$\lim_{x \rightarrow \infty} \frac{d}{dx} \left[ \frac{1 - F_X(x)}{f_X(x)} \right] = 0$	VEI <i>Exponencial dupla</i>
<i>Pareto</i> $F_Y(y) = 1 - (y - \varepsilon)^{-k}$ $y > 1 + \varepsilon; k \geq 1; \varepsilon \geq 0$	$F_{Y_n}(y) = \exp \left[ - \left( \frac{y - \varepsilon}{v_n - \varepsilon} \right)^k \right]$	$\lim_{y \rightarrow \infty} \frac{y f_Y(y)}{1 - F_Y(y)} = a$ $a > 0 (a = const.)$	VEII <i>Exponencial</i>
<i>Limitada</i> $F_Z(z) = 1 - (\omega - z)^k$ $\omega > 0; z \leq \omega; k > 1$	$F_{Z_n}(z) = \exp \left\{ - \left[ \frac{\omega - z}{\omega - w_n} \right]^k \right\}$ $z \leq \omega$ $k = \alpha_n (\omega - w_n)$	$\lim_{z \rightarrow \infty} \frac{(\omega - z) f_Z(z)}{1 - F_Z(z)} = b$ $b > 0 (b = const.)$	VEIII <i>Exponencial com limite superior <math>\omega</math></i>

\* MITSIOPOULOS et al. (1991) e GUMBEL (1958)

As funções de quantis (função inversa) das distribuições assintóticas do tipo VEI e VEII, que correspondem às formas limites de distribuições de valores extremos  $X_n$  e  $Y_n$  das distribuições iniciais  $X$  e  $Y$  respectivamente, são ilimitadas em relação aos seus extremos máximos ou mínimos, dependendo dos valores que estão sendo considerados. Em contraste, a função de quantis da forma assintótica tipo VEIII, que representa a distribuição limite dos extremos  $Z_n$  da distribuição inicial  $Z$ , tem limite superior ou inferior finito, quando se trata dos respectivos valores máximos ou mínimos da variável inicial  $Z$ .

Valores extremos de uma variável aleatória  $X$  estão invariavelmente associados com a forma da cauda da função densidade de probabilidade de  $X$ . A convergência da função distribuição de seus valores extremos para uma particular forma, por esta razão, dependerá, primordialmente, do comportamento da cauda da distribuição  $X$  em relação aos seus extremos. Em particular, valores extremos de uma distribuição  $X$  com decaimento exponencial da cauda em direção aos extremos, convergem assintoticamente para a forma VEI. Ao passo que valores extremos de uma variável inicial  $Y$ , cuja cauda decai polinomialmente quando  $y$  se aproxima de seus extremos, convergem assintoticamente para a forma VEII. Por fim, se o valor extremo de uma distribuição inicial  $Z$  é limitado, ou seja, se os valores máximos ou mínimos de  $Z$  tem respectivamente limite superior ou inferior finito, a correspondente distribuição de extremos converge para a forma assintótica VEIII, quando o tamanho da amostra  $n$  é grande (GUMBEL, 1958; ANG & TANG, 1984; MITSIPOULOS et al., 1991). A descrição dos valores máximos que se apresentam a seguir, referem-se aos extremos máximos de uma população inicial, por isso as diversas variáveis associadas aos mesmos estão indexadas com a letra  $n$  subscrita.

#### 4.3.1 A Forma Assintótica VEI

Na função distribuição  $F_{X_n}(x)$  da forma assintótica VEI para distribuição de valores extremos  $X_n$ ,  $u_n$  e  $\alpha_n$  são, respectivamente, parâmetros de posicionamento e escala, onde  $u_n$  é o valor máximo característico da variável inicial  $X$  e  $\alpha_n$  é uma medida inversa da dispersão de  $X_n$ .

O valor máximo característico,  $u_n$ , é uma conveniente medida de posicionamento ou tendência central dos possíveis valores máximos. Numa amostra de tamanho  $n$  de uma variável inicial  $X$ , o número esperado de valores amostrais que são maiores do que  $x$  é  $n[1 - F_x(x)]$  (GUMBEL, 1958). Logo,  $u_n$  é definido como o particular valor de  $X$ , tal que, em uma amostra com  $n$  valores da população  $X$ , o número esperado de valores amostrais maior do que  $u_n$  é 1 (um), ou seja:

$$n[1 - F_x(x)] = 1;$$

logo

$$F_x(u_n) = 1 - \frac{1}{n} \quad (17)$$

portanto

$$u_n = F_x^{-1}\left(1 - \frac{1}{n}\right) \quad (18)$$

O valor máximo característico,  $u_n$ , é o valor modal (isto é, o valor mais provável) da variável extrema  $X_n$ . Isto significa que  $u_n$  é o mais provável maior valor de amostras da variável  $X$ . Isto pode ser mostrado da seguinte forma:

Seja  $\tilde{x}$  o valor modal de  $X_n$ , então  $\tilde{x}$  pode ser a solução de:

$$f'_{X_n}(\tilde{x}) = 0 \quad (19)$$

Tomando a derivada da equação (5), obtém-se:

$$f'_{X_n}(\tilde{x}) = n[F_x(\tilde{x})]^{n-1} f'_x(\tilde{x}) + n(n-1)[F_x(\tilde{x})]^{n-2} f_x^2(\tilde{x}) \quad (19a)$$

A equação (19) combinada com (19a) rende (suprimindo o subscrito  $x$  em  $F_x$  e  $f_x$ )

$$\frac{f'(\tilde{x})}{f(\tilde{x})} + (n-1)\frac{f(\tilde{x})}{F(\tilde{x})} = 0 \quad (19b)$$

Considerando que, para uma variável inicial  $X$  com a cauda decaindo de forma exponencial em relação aos extremos, seus valores extremos convergem para a forma assintótica VEI. GUMBEL (1958) definiu uma distribuição inicial  $X$  com cauda exponencial para ser ilimitada na direção dos extremos, onde para valores elevados positivos de  $X$ , as ordenadas de  $f_X(x)$  e  $1 - F_X(x)$  são pequenas, ao passo que,  $f'_X(x)$  é pequena e negativa, especificamente, para  $x$  grande, assim:

$$\frac{f(x)}{1 - F(x)} = -\frac{f'(x)}{f(x)} \quad (20)$$

Combinando a equação (20) com a (19b), obtém-se:

$$-\frac{f(\tilde{x})}{1 - F(\tilde{x})} + (n-1)\frac{f(\tilde{x})}{F(\tilde{x})} = 0$$

logo

$$F_X(\tilde{x}) = 1 - \frac{1}{n} = F_X(u_n)$$

Portanto,  $u_n = \tilde{x}$  é o valor modal, para o caso de uma distribuição com cauda exponencial.

Para encontrar a relação que define  $\alpha_n$ , é necessário recorrer a função de risco. Na teoria da confiabilidade, a probabilidade condicional de falha no tempo  $(t, t+dt)$  dada nenhuma falha em  $(0, t)$  envolve a função de risco (CROWDER et al., 1991):

$$h(t) = \frac{f(t)}{1 - F(t)} \quad (21)$$

Tal função é também envolvida na distribuição de valores extremos. Assim, para o valor mais elevado, a função de risco de  $X$  pode ser definida como:

$$h_n(t) = \frac{f_X(t)}{1 - F_X(t)} \quad (22)$$

Para o valor máximo característico,  $u_n$ , a função de risco (22) combinada com a equação (17), torna-se:

$$h_n(u_n) = nf_n(u_n) \quad (23)$$

Considerando o maior valor de uma distribuição  $f_x(x)$  com cauda exponencial e assumindo que todas as derivadas de  $f_x(x)$  exista, pode-se Expandir  $F_x(x)$  em uma série de Taylor sobre  $u_n$  para obter:

$$F_x(x) = F(u_n) + (x - u_n)f(u_n) + \frac{1}{2}(x - u_n)^2 f'(u_n) + \frac{1}{3!}(x - u_n)^3 f''(u_n) + \dots$$

Da equação (23), obtém-se

$$f(u_n) = \frac{1}{n}h(u_n)$$

Por outro lado, a combinação da equação (20) com a (22), rende:

$$f'(u_n) = -f(u_n)h(u_n) = -\frac{1}{n}h^2(u_n)$$

Tomando a derivada de  $f'(x)$  da equação (20) e fazendo  $x = u_n$ , encontra-se:

$$f''(x) = -f'(x)h(x) = \frac{1}{n}h^3(x)$$

$$f''(u_n) = \frac{1}{n}h^3(u_n)$$

Altas derivadas podem ser obtidas de forma semelhante. Assim, com a equação (17) e retomando a série de Taylor, obtém-se:

$$F_x(x) = 1 - \frac{1}{n} \left[ 1 - (x - u_n)h(u_n) + \frac{1}{2}(x - u_n)^2 h^2(u_n) - \frac{1}{3!}(x - u_n)^3 h^3(u_n) + \dots \right]$$

$$F_x(x) = 1 - \frac{1}{n} \exp[-h(u)(x - u_n)]$$

desta forma, a distribuição do maior valor, de acordo com a equação (4), torna-se:

$$F_{X_n}(x) = \left[ 1 - \frac{1}{n} e^{-h(u_n)(x - u_n)} \right]^n$$

e

$$\ln F_{X_n}(x) = \frac{\ln \left[ 1 - \frac{1}{n} e^{-h(u_n)(x - u_n)} \right]}{\frac{1}{n}}$$

como  $n \rightarrow \infty$

$$\lim_{n \rightarrow \infty} \ln F_{X_n}(x) = -e^{-h(u_n)(x - u_n)}$$

Portanto, quando  $n$  for grande

$$F_{X_n}(x) = \exp[-e^{-h(u_n)(x - u_n)}] \quad (24)$$

Comparando a última equação com a equação da forma assintótica do tipo VEI, apresentada na Tabela 2, observa-se que a mesma corresponde a distribuição assintótica de valor máximo do tipo VEI com o seguinte parâmetro de escala:

$$\alpha_n = h_n(u_n) = n f_x(u_n) \quad (25)$$

$\alpha_n$  é uma medida inversa de dispersão de  $X_n$ , que pode ser verificado através do seguinte procedimento:

Para valores máximos do tipo I, pode-se introduzir a variável extremo padrão, como sendo:

$$S = \alpha_n (X_n - u_n) \quad (26)$$

onde suas funções distribuição e densidade são:

$$F_S(S) = \exp(-e^{-S}) \quad \text{e} \quad f_S = e^{-S} \exp(-e^{-S})$$

Neste caso, a função geratriz de momentos, torna-se:

$$G_S(t) = E(e^{tS}) = \int_{-\infty}^{\infty} e^{tS} e^{-S} e^{-e^{-S}} dS, \quad \text{fazendo } r = e^{-S} \quad \text{e} \quad dS = -\frac{dr}{r}$$

$$G_S(t) = \int_0^{\infty} e^{tS} e^{-r} dr = \int_0^{\infty} e^{-t \ln r} e^{-r} dr = \int_0^{\infty} r^{-t} e^{-r} dr = \Gamma(1-t) \quad (27)$$

Logo, a função geratriz de momentos de S é igual a função gama. As derivadas de  $G_S(t)$ , avaliada em  $t = 0$ , rendem os respectivos momentos de S. A k-ésima derivada de  $\Gamma(1-t)$ , segundo CRAMER (1946), corresponde a:

$$\Gamma^{(k)}(1-t) = \int_0^{\infty} r^{-t} e^{-r} (\ln r)^k dr \quad (28)$$

Em  $t = 0$ , as várias derivadas de (28) têm sido avaliadas por GUMBEL (1958), onde as primeiras quatro derivadas são:

$$\frac{d\Gamma(1)}{dt} = \gamma = 0,577216\dots \quad (\text{número de Euler})$$

$$\frac{d^2\Gamma(1)}{dt^2} = \gamma^2 + \frac{\pi^2}{6}$$

$$\frac{d^3\Gamma(1)}{dt^3} = \gamma^3 + \frac{\pi^2}{2}\gamma + 2S_3$$

$$\frac{d^4\Gamma(1)}{dt^4} = \gamma^4 + \pi^2\gamma^2 + 8\gamma S_3 + \frac{3\pi^4}{20}$$

Sendo  $S_3 = 1.20205690$ . Portanto, a média e a variância de  $S$ , são:

$$\mu_s = \gamma = 0,57716$$

$$\sigma_s^2 = \frac{\pi^2}{6}.$$

O momento de uma variável extremo geral tipo I pode ser avaliada através do momento de  $S$ , usando a equação (26). Em particular, obtém-se a média e a variância de  $X_n$ , como segue:

$$X_n = u_n + \frac{S}{\alpha_n}$$

Consequentemente, o valor médio e a variância de  $X_n$  são:

$$\mu_{X_n} = u_n + \frac{\gamma}{\alpha_n}$$

$$\sigma_{X_n}^2 = \frac{1}{\alpha_n^2} \sigma_s^2 = \frac{\pi}{6\alpha_n^2}$$

O que mostra que  $\alpha_n$  é uma medida inversa da dispersão dos valores extremos de uma variável inicial  $X$ .

### 4.3.2 A Forma Assintótica VEII

Na função distribuição  $F_{Y_n}(y)$  da forma assintótica de valores extremos de uma distribuição inicial  $Y$  com cauda polinomial,  $v_n$  é o valor máximo característico da variável  $Y$  e  $k$  é o parâmetro de forma.  $1/k$  representa uma medida de dispersão de  $Y_n$ .

$v_n$  é igual ao valor modal de  $Y_n$  e é definido, tal como  $u_n$ , a partir da equação (17), ou seja, para uma variável inicial  $Y$ :

$$F_Y(v_n) = 1 - \frac{1}{n} \quad \text{logo} \quad v_n = F_Y^{-1}\left(1 - \frac{1}{n}\right)$$

É possível mostrar que  $1/k$  é uma medida de dispersão de  $Y_n$ , partindo do fato de que as formas assintóticas VEI e VEII se relacionam através da seguinte transformação logarítmica.

Se  $Y_n$  tem distribuição assintótica do tipo VEII com parâmetros  $v_n$  e  $k$ , a distribuição de  $\ln Y_n$  terá a forma assintótica tipo VEI com parâmetros  $u_n = \ln(v_n)$  e  $\alpha_n = k$ .

Desta maneira:

$$X_n = \ln Y_n \quad \text{donde} \quad Y_n = e^{X_n}$$

Logo, a distribuição de  $X_n$  torna-se:

$$F_{X_n}(x) = F_{Y_n}(e^x) = \exp\left[-\left(\frac{v_n}{e^x}\right)^k\right]$$

Adotando  $v_n = e^{u_n}$  e  $k = \alpha_n$ , a função de distribuição de  $X_n$  resulta em:

$$F_{X_n}(x) = \exp\left[-e^{-\alpha_n(x-u_n)}\right]$$

Por conseguinte, a variável extremo padronizada para VEII e a sua função distribuição são as seguintes:

$$S = \alpha_n(X_n - u_n) = k(\ln Y_n - \ln v_n) = k \ln \left( \frac{Y_n}{v_n} \right)$$

$$F_{Y_n}(y) = F_S \left( k \ln \frac{y}{v_n} \right)$$

Assim,  $Y_n = v_n e^{S/k}$  é o  $t$ -ésimo momento de  $Y_n$

$$E(Y_n^t) = v_n^t E(e^{(t/k)S}) = v_n^t G_S(t/k)$$

Onde  $G_S(t/k)$  é a função geratriz de momentos de  $S$ , dada pela equação (27). Logo, o  $t$ -ésimo momento de  $Y_n$  corresponde a:

$$E(Y_n^t) = v_n^t \Gamma \left( 1 - \frac{t}{k} \right), \quad t < k \quad (29)$$

A partir da equação (29) obtém-se a média e a variância de  $Y_n$ , como segue:

$$\mu_{Y_n} = E(Y_n) = v_n \Gamma \left( 1 - \frac{1}{k} \right), \quad \text{para } k > 1$$

$$E(Y_n^2) = v_n^2 \Gamma \left( 1 - \frac{2}{k} \right), \quad \text{para } k \geq 2$$

$$\sigma_{Y_n}^2 = E(Y_n^2) - [E(Y_n)]^2 = v_n^2 \Gamma \left( 1 - \frac{2}{k} \right) - \left[ v_n \Gamma \left( 1 - \frac{1}{k} \right) \right]^2$$

logo

$$\sigma_{Y_n}^2 = v_n^2 \left[ \Gamma \left( 1 - \frac{2}{k} \right) - \Gamma^2 \left( 1 - \frac{1}{k} \right) \right]$$

Portanto,  $1/k$  é uma medida da dispersão de  $Y_n$ .

### 4.3.3 A Forma Assintótica VEIII

A forma assintótica VEIII com função distribuição  $F_{Z_n}(z)$ , representa a distribuição limite dos valores extremos das distribuições inicial  $Z$  e tem um limite superior ou inferior finito, quando referir-se aos valores máximos ou mínimos, respectivamente. Diferente das formas assintóticas VEI e VEII, que são ilimitadas.

O parâmetro  $k$ , presente na função distribuição da forma assintótica VEIII, é um parâmetro de forma e seu inverso ( $1/k$ ) é uma medida de dispersão de  $Z_n$ ;  $\omega$  é o valor limite superior de  $Z$  [ $F_Z(\omega) = 1$ ] e  $w_n$  é o valor máximo característico de  $Z$ , o qual, representa, também, o valor modal de  $Z_n$ . O mesmo é definido através da equação (17), isto é:

$$F_Z(w_n) = 1 - \frac{1}{n}$$

A verificação de que  $1/k$  é uma medida de dispersão de  $Z_n$  é feita através da definição da variância de  $Z_n$ , usando o mesmo procedimento empregado para definir  $\sigma_{Z_n}^n$ , considerando que a forma assintótica VEIII pode ser relacionada com a forma VEI e supondo que:

$$X_n = a - \ln\left(\frac{\omega - Z_n}{\omega - w_n}\right) \quad \text{onde} \quad Z_n = \omega - (\omega - w_n)e^{-(X_n - a)}$$

então, se a distribuição de  $Z_n$  é da forma VEIII, a distribuição de  $X_n$  torna-se:

$$F_{X_n}(x) = \exp\left\{-\left[\frac{\omega - \omega + (\omega - w_n)e^{-(x-a)}}{\omega - w_n}\right]^k\right\} = \exp[-e^{-k(x-a)}]$$

De tal modo que  $X_n$  tem distribuição assintótica VEI com parâmetros  $u_n = a$  e  $\alpha_n = k$ . Assim, pode-se definir a correspondente variável extremo padronizada tipo VEIII como:

$$S = \alpha_n(X_n - u_n) = k \left[ a - \ln \left( \frac{\omega - Z_n}{\omega - w_n} \right) - a \right] = -k \ln \left( \frac{\omega - Z_n}{\omega - w_n} \right)$$

$$\frac{\omega - Z_n}{\omega - w_n} = e^{-S/k} \quad \text{donde} \quad \omega - Z_n = (\omega - w_n) e^{-S/k}$$

Elevando à  $t$  os dois lados da equação acima e em seguida tomando a esperança em ambos os lados, tem-se:

$$E(\omega - Z_n)^t = (\omega - w_n)^t E(e^{-tS/k}) = (\omega - w_n)^t G_S(-t/k)$$

Onde  $G_S(-t/k)$  é a função geratriz de momentos de  $S$ , dada pela equação (27). Portanto, o  $t$ -ésimo momento, o valor médio e a variância de  $Z_n$  são:

$$E(\omega - Z_n)^t = (\omega - w_n)^t \Gamma(1 + t/k) \quad (t\text{-ésimo momento de } Z_n)$$

logo

$$E(\omega - Z_n) = (\omega - w_n) \Gamma(1 + 1/k)$$

e

$$\mu_{Z_n} = \omega - (\omega - w_n) \Gamma(1 + 1/k) \quad (\text{valor médio de } Z_n)$$

enquanto que a variância de  $Z_n$  é dada por:

$$\sigma_{Z_n}^2 = \text{var}(\omega - Z_n) = E(\omega - Z_n)^2 - [E(\omega - Z_n)]^2$$

como

$$E(\omega - Z_n)^2 = (\omega - w_n)^2 \Gamma(1 + 2/k)$$

$$E(\omega - Z_n) = (\omega - w_n) \Gamma(1 + 1/k)$$

tem-se

$$\sigma_{Z_n}^2 = (\omega - w_n)^2 \Gamma(1 + 2/k) - (\omega - w_n)^2 \Gamma^2(1 + 1/k)$$

$$\sigma_{Z_n}^2 = (\omega - w_n)^2 [\Gamma(1 + 2/k) - \Gamma^2(1 + 1/k)] \quad (\text{variância de } Z_n)$$

o que mostra que  $1/k$  é uma medida de dispersão.

Observa-se que a forma assintótica de distribuição de extremos é largamente independente da forma exata da distribuição inicial  $X$ , porém, está associada a forma de decaimento da cauda da função distribuição acumulada de  $X$ , quando  $x$  se aproxima dos extremos. Além disso, os parâmetros  $u$  e  $\alpha$  das três formas de distribuição assintótica são, também, dependentes da distribuição de  $X$ .

d) Critérios de convergência:

Os critérios de convergência, apresentados na Tabela 1, são critérios matemáticos para identificar a forma assintótica de distribuições em situações onde o comportamento da cauda não é tão claro, como no caso de variáveis  $X$  com função densidade exponencial ou normal, em que as caudas apresentam claramente decaimento exponencial na direção dos extremos, conseqüentemente, a distribuição dos grandes valores de uma ou outra, dessas variáveis iniciais, converge para a forma assintótica tipo EVI.

#### 4.4 DISTRIBUIÇÃO DE PROBABILIDADE GENERALIZADA DE VALORES EXTREMOS (GEV)

A observação de dados meteorológicos mostrava que, na maioria dos casos, seus valores máximos cresciam linearmente com o logaritmo do período de retorno. GUMBEL (1935) proveu justificativa teórica para esta aproximação. Defeitos nesta aproximação foram apontados por BARRICELLI (1943) e por BROOKS & CARRUTHERS (1953), onde para dados de temperaturas a aproximação de Gumbel superestimava e para dados de chuva subestimava os valores máximos observados em longos períodos.

FISHER & TIPPETT (1928) mostraram que valores extremos de uma variável aleatória satisfazem a uma equação funcional referida na página 37, equação (16), como estabilidade postulada. JENKINSON (1955) encontrou a solução geral desta equação, a qual descreve todos os tipos de distribuições de extremos encontradas em dados meteorológicos. A aproximação (distribuição) de Gumbel é

incluída nesta solução como um caso especial.

JENKINSON (1955) utilizando valores máximos de chuvas, pressão de vapor, pressão, temperatura e velocidade do vento, todas representadas pela variável  $x$ , e tomando  $P = \exp(-\exp(-y))$ ,  $P=F$ , observou que  $y$  é sempre uma função crescente de  $x$ , e que três tipos de curvas  $y = f(x)$  foram distingüidas através de  $dy/dx$ , quando  $dy/dx$  crescia ou diminuía ou permanecia constante com o aumento de  $x$ . A teoria apresentada por GUMBEL (1935) infere que  $dy/dx$  é constante, já a teoria de BARRICELLI (1943) determina que  $dy/dx$  ou é constante ou aumenta com o crescimento de  $x$ .

A equação geral para as curvas  $y = f(x)$  propostas por JENKINSON (1955), que inclui os três tipos de curvas por ele observados, é expressa através da seguinte equação:

$$x = a(1 - e^{-ky}), \quad ak > 0 \quad (30)$$

onde

$$\frac{dy}{dx} = \frac{1}{ak} e^{ky}$$

Para  $k$  positivo (forma VEIII):

$y = -\infty,$	$x = -\infty$	e	$dy/dx = 0;$
$y = 0,$	$x = 0$	e	$dy/dx = 1/ak;$
$y = +\infty,$	$x = a$	e	$dy/dx = +\infty;$

Para  $k$  negativo (forma VEII):

$y = -\infty,$	$x = -a$	e	$dy/dx = +\infty;$
$y = 0,$	$x = 0$	e	$dy/dx = 1/ak;$
$y = +\infty,$	$x = +\infty$	e	$dy/dx = 0;$

Para  $k \rightarrow 0,$   $x \rightarrow aky$  (forma VEI – Gumbel)

GUMBEL (1958), utilizando a transformação de  $x$  para  $y$  da equação (30), através da expressão:

$$x = u + \frac{1}{k\alpha} \left( 1 - e^{-\frac{y}{k}} \right) \quad (31)$$

descreveu a solução geral encontrada por JENKINSON (1955) da seguinte maneira:

$$F_x(x) = \exp \left\{ - \left[ 1 - \alpha k (x - u) \right]^{1/k} \right\}, \quad \alpha > 0 \quad (32)$$

Se  $k > 0$  há um limite superior  $\omega = u + 1/k\alpha$  e obtém-se a forma assintótica VEIII, se  $k < 0$  há um limite inferior  $\varepsilon = u + 1/k\alpha$  que corresponde a forma assintótica VEII. Quando  $k \rightarrow 0$  conduz-se para a forma VEI (distribuição Gumbel).

Como o valor do inverso de  $\alpha$  representa uma medida de dispersão dos valores máximos, muitos autores preferem usar a distribuição GEV, equação (32), da forma mostrada a seguir, onde na referida equação é usado  $1/\alpha$  em vez de  $\alpha$ :

$$F_x(x) = \exp \left[ - \left( 1 - \kappa \frac{x - u}{\alpha} \right)^{1/\kappa} \right]; \quad \kappa \neq 0 \quad (32a)$$

sendo

$$\begin{aligned} \varepsilon \leq x < +\infty, \quad k < 0 & \quad \text{- distribuição VEII} \\ -\infty < x \leq \omega, \quad k > 0 & \quad \text{- distribuição VEIII} \end{aligned}$$

onde  $u$  é um parâmetro de posicionamento com  $-\infty < u < +\infty$ ,  $\alpha$  é um parâmetro de escala com  $0 < \alpha < +\infty$  e  $k$  é um parâmetro de forma com  $-\infty < k < +\infty$ . Por quanto, quando  $k > 0$  o limite superior da distribuição assintótica VEIII torna-se  $\omega = u + \alpha/k$  e quando  $k < 0$  o limite inferior da distribuição assintótica VEII torna-se  $\varepsilon = u + \alpha/k$ .

A expressão da distribuição GEV quando  $k \rightarrow 0$  é obtida, partindo-se da equação (32a). Fazendo  $q = 1/\kappa$ , quando  $k \rightarrow 0$  significa que  $q \rightarrow \infty$ . Logo,

aplicando o logaritmo em ambos os membros da equação (32a), a mesma se transforma em:

$$\ln F_X(x) = -\left(1 - \frac{1}{q} \frac{x-u}{\alpha}\right)^q$$

Aplicando o logaritmo novamente, obtém-se:

$$\ln[-\ln F_X(x)] = \ln\left(1 - \frac{1}{q} \frac{x-u}{\alpha}\right)^q = q \ln\left(1 - \frac{1}{q} \frac{x-u}{\alpha}\right) = \frac{\ln\left(1 - \frac{1}{q} \frac{x-u}{\alpha}\right)}{\frac{1}{q}}$$

Aplicando o limite quando  $q \rightarrow \infty$ , a expressão resultante é

$$\lim_{q \rightarrow \infty} \ln[-\ln F_X(x)] = -\frac{x-u}{\alpha}$$

logo

$$F_X(x) = \exp\left[-\exp\left(-\frac{x-u}{\alpha}\right)\right]; \quad \kappa = 0 \quad (32b)$$

onde  $u$  é o parâmetro de posicionamento,  $\alpha$  é o parâmetro de escala e  $-\infty < x < +\infty$  (distribuição VEI). Nos três casos  $\alpha$  é assumido positivo. Na prática, quando  $|\kappa| < 0.03$  a forma geral da distribuição GEV é similar a distribuição Gumbel (STEDINGER et al., 1993), que corresponde a forma assintótica VEI.

Em consequência, o p-ésimo quantil de uma distribuição GEV é dado pelas seguintes relações:

$$x_p = u + \frac{\alpha}{k} \left\{1 - [-\ln(p)]^k\right\} \quad \text{para } k \neq 0, \quad 0 < p < 1 \quad (33a)$$

$$x_p = u - \alpha \ln[-\ln(p)] \quad \text{para } k = 0, \quad 0 < p < 1 \quad (33b)$$

Considerando a equação (32b), define-se uma variável reduzida  $z$ , chamada variável reduzida de Gumbel, como  $z = (x - u)/\alpha$ . substituindo  $z$  em (32b), obtém-se:

$$F(x) = \exp[-\exp(-z)]$$

Logo

$$z = \ln[-\ln(F(x))] \quad (34)$$

onde, neste caso, a variável reduzida de Gumbel,  $z$ , também relaciona-se com o período de retorno ( $T$ ),  $T = 1/F(x)$ . Além disso, a equação (34) pode ser usada para definir  $z$  com respeito as distribuições VEI, VEII e VEIII. Assim, em um gráfico  $x$  versus  $z$  define-se o comportamento das três formas de distribuições de valores extremos, com relação a posição de plotagem de  $x$ , como é mostra na Figura 2.

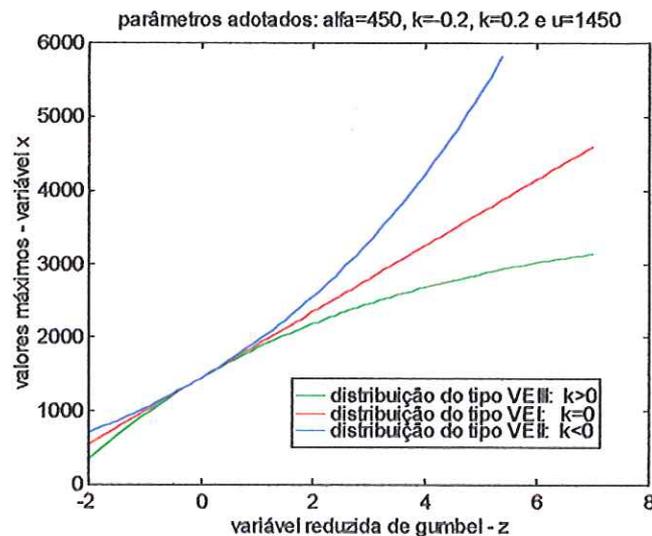


FIGURA 2 - Distribuições das 3 formas de valores extremos em função da variável reduzida de Gumbel.

Com o aumento dos valores de  $z$  os valores de  $x$  do tipo VEI aumentam segundo uma linha reta, ao passo que os valores de  $x$  do tipo VEII crescem mais rapidamente seguindo uma curva côncava, ambos aumentado de forma ilimitada na parte superior. Já os valores de  $x$  do tipo VEIII crescem seguindo uma curva convexa para um limite superior finito.

## 5 METODOLOGIA

O estudo da análise de frequência de cheias através da distribuição GEV desenvolvido neste trabalho, considerou três situações distintas. Primeiro, analisou-se a estimabilidade da distribuição GEV, considerando as incertezas e erros na estimação dos parâmetros e dos altos quantis de cheias decorrentes da limitação dos dados de cheias disponíveis. Esta investigação baseou-se na utilização de séries de cheias anuais com diferentes características e comprimentos, obtidas de seqüências sintéticas de vazões diárias. Deste modo, procedeu-se à geração de vazões diárias, utilizando simulação Monte Carlo com a distribuição limitada como subjacente, conforme pressuposto da distribuição de extremos VEIII. Neste caso, a relação teórica existente entre as duas distribuições permitiu comparar as estimativas das diversas variáveis envolvidas no processo, com seus respectivos valores teóricos.

Num segundo momento, verificou-se a possibilidade da utilização de modelos estocásticos de simulação de vazões diárias disponíveis na literatura, para produzir séries sintéticas de vazões diárias de longos períodos, como forma de resolver as inconveniências resultantes das limitações das séries históricas. Para tanto, foram simuladas seqüências estocásticas de vazões diárias, utilizando diversos modelos diários, ajustados aos dados de vazões observadas em diferentes rios. A partir das referidas seqüências diárias, obtiveram-se séries de cheias anuais que foram submetidas ao ajuste da distribuição GEV.

Por fim, Propôs-se um novo modelo estocástico de geração de vazões diárias, que utiliza a distribuição inicial limitada para simular pulsos de vazões durante o processo de simulação. Os pulsos gerados são transformados em valores de vazões diárias por meio de uma função de transferência adotada no modelo. Para avaliação do modelo, as séries de vazões diárias, mensais, anuais e de valores extremos, obtidas com a simulação de seqüências diárias através do método proposto, foram

comparadas às respectivas séries históricas. A série de cheias anuais foi modelada através da distribuição GEV.

Assim, neste capítulo são apresentados: modelos de estimação de parâmetros, testes de qualidade de ajustes e de forma da distribuição ajustada, distribuição inicial de probabilidade limitada, fórmulas de posição de plotagem, descrição dos modelos estocásticos utilizados e do modelo proposto e, por fim, o experimento de geração de vazões diárias, dividido em três partes, de acordo com cada situação estudada.

## 5.1 MÉTODOS DE ESTIMAÇÃO DE PARÂMETROS

Um dos principais objetivos da inferência estatística é a estimação, com confiança, das características da população de uma variável aleatória. Os procedimentos utilizados são chamados de métodos ou técnicas de estimação.

As propriedades populacionais de uma variável aleatória são, geralmente, caracterizadas através de seus parâmetros, quantis ou outro número similar. Como populações são raramente conhecidas em ciências como hidrologia, estas propriedades da variável aleatória podem ser estimadas através de amostras disponíveis de dados. O procedimento usual para estimação dos parâmetros é obter uma amostra aleatória  $x_1, x_2, \dots, x_n$  de uma população  $X$ , a qual é usada para estimativa dos parâmetros. Assim,  $\hat{\theta}_i$ , uma estimativa ou estimador do parâmetro  $\theta_i$ , é uma função das observações ou das variáveis aleatórias. Consequentemente,  $\hat{\theta}_i$  é uma variável aleatória possuindo média, variância e distribuição de probabilidade.

Intuitivamente, percebe-se que quanto mais observações das variáveis aleatórias estão disponíveis para a estimação do parâmetro, mais  $\hat{\theta}$  se aproxima de  $\theta$ . Também, se muitas amostras são utilizadas para obtenção de  $\hat{\theta}$ , pode-se observar que o valor médio de  $\hat{\theta}$  se aproxima de  $\theta$ . Essas duas colocações são condizentes com duas propriedades conhecidas como consistência e não viesamento.

Um estimador  $\hat{\theta}$  de  $\theta$  é dito não viesado se  $E(\hat{\theta}) = \theta$ . Se  $P(|\hat{\theta} - \theta| > \epsilon)$  tende a zero quando o tamanho da amostra se aproxima de infinito, para todo  $\epsilon > 0$ , o estimador  $\hat{\theta}$  de um parâmetro  $\theta$  é chamado de consistente. Porém, quando o estimador  $\hat{\theta}$  é não viesado e sua variância é, no mínimo, tão pequena quanto a de

qualquer outro estimador não viesado de  $\theta$ , diz-se que ele é o mais eficiente estimador de  $\theta$ . Agora, se  $\hat{\theta}$  usa todas as informações relevantes de  $\theta$  que estão contidas na amostra, ele é dito ser um estimador suficiente (HAAN, 1977).

O uso em hidrologia de estimadores simples como o método gráfico ou o método dos momentos, considerando que estes são suficientemente precisos, em virtude da escassez dos dados hidrológicos, é um grande engano. Para dados disponíveis como os da hidrologia, com grandes variações, distribuições altamente assimétricas e séries curtas de dados observados, a estimativa não viesada e eficiente parece ser o caminho (YEVJEVICH, 1972). Resta, selecionar o método de estimação que resulte em estimadores não viesados e eficientes.

Os parâmetros da distribuição GEV têm sido estimados por diversos métodos. Apesar da forte motivação e boa qualidade estatística do método da máxima verossimilhança, devido à necessidade da utilização de processos numéricos para resolver as expressões implícitas contendo os três parâmetros da GEV, o mesmo cedeu lugar ao uso do método dos momentos ordinários, durante bastante tempo, motivado pelas facilidades dos cálculos.

Procedimentos computacionais para estimação dos parâmetros da GEV, utilizando estimadores de máxima verossimilhança foram propostos por OTTEN & VAN MONTFORT (1980), PRESCOTT & WALDEN (1980, 1985) e HOSKING (1985). Contudo, esses estimadores são muito instáveis quando aplicados a pequenas amostras (HOSKING et al., 1995a) e podem gerar valores absurdos do parâmetro de forma ( $k$ ) da GEV (MARTINS & STEDINGER, 2000).

O método de máxima verossimilhança para estimação de parâmetros é assintoticamente ótimo, não viesado, rende mínima variância, mas pode produzir estimativas ruins em amostras pequenas, especialmente quando a variável aleatória está restrita ao intervalo do qual dependem os parâmetros (BOBÉE et al., 1993).

Alternativamente, têm sido utilizados os momentos L que são equivalentes aos momentos com peso probabilístico, cujos estimadores são combinações lineares das estatísticas de ordem das observações. Por isso, estão sujeitos a menores erros de estimação do que momentos ordinários, que envolvem o quadrado ou o cúbico das observações durante a estimação da variância e da assimetria amostral. Por outro lado, estimadores de momentos L da distribuição GEV são melhores do que

estimadores de máxima verossimilhança, em termos de tendência (erro) de estimação e variância, para amostras com tamanhos entre 15 e 100 valores (HOSKING et al., 1985a; HOSKING, 1990).

5.1.1 Método dos Momentos com Pesos Probabilísticos:

Seja  $X_1, X_2, \dots, X_n$  uma amostra aleatória de uma população com função densidade de probabilidade  $f(x)$  e função distribuição  $F(x)$ , e seja  $x_{1:n} \leq x_{2:n} \leq \dots \leq x_{n:n}$  as estatísticas de ordem obtidos da amostra acima. Denota-se o momento  $E(X_{i:n}^k)$  através  $M_{i:n}^k$ . Então, considerando-se a função densidade  $f_{i:n}(x)$  de  $x_{i:n}$ , tal como mostrado na equação (1), tem-se (BALAKRISHNAN & COHEN, 1991; DAVID, 1981):

$$M_{i:n}^k = E(X_{i:n}^k) = \int_{-\infty}^{\infty} x^k f_{i:n}(x) dx$$

ou

$$M_{i:n}^k = \frac{n!}{(n-i)!(i-1)!} \int_0^1 x^k [F(x)]^{i-1} [1-F(x)]^{n-i} dF(x) \tag{35}$$

$$i = 1, 2, \dots, n, k \geq 1$$

Uma função distribuição acumulada  $F \equiv F(x) = P(X \leq x)$  pode ser caracterizada através dos momentos de pesos probabilísticos (PWMs), os quais são definidos como (GREENWOOD et al., 1979):

$$M_{p,r,s} = E\{X^p [F]^r [1-F]^s\} = \int_0^1 [x(F)]^p F^r (1-F)^s dF \tag{36}$$

onde  $p, r$  e  $s$  são números reais. GREENWOOD et al., (1979) notaram que, no caso especial onde  $p, r$  e  $s$  são inteiros não negativos,  $M_{p,r,s}$  é proporcional ao  $p$ -ésimo momento em torno da origem da  $(r+1)$ -ésima estatística de ordem, para uma amostra de tamanho  $(s+r+1)$ . Especificamente:

$$M_{p,r,s} = B[r+1, s+1] E[X_{(r+1, s+r+1)}^p] \quad (37)$$

onde  $B[\cdot, \cdot]$  é a função beta.

Se  $r = s = 0$  e  $p$  é um inteiro não negativo, então  $M_{p,0,0}$  representam os momentos convencionais não centrais de  $X$ . Se  $M_{p,0,0}$  de ordem  $p$  existe e  $x$  é uma função contínua de  $F$ , então  $M_{p,r,s}$  existe para todo número inteiro não negativo  $r$  e  $s$ . Se  $r$  e  $s$  são não negativos inteiros, então:

$$M_{p,0,s} = \sum_{r=0}^s \binom{s}{r} (-1)^r M_{p,s,0} \quad (38a)$$

$$M_{p,r,0} = \sum_{s=0}^r \binom{r}{s} (-1)^s M_{p,0,s} \quad (38b)$$

onde, se  $M_{p,0,s}$  existe e  $X$  é uma função contínua de  $F$ ,  $M_{p,r,0}$  existe. Para estimar os parâmetros da distribuição de  $X$  é preferível usar  $M_{l,r,s}$ , onde a ocorrência, somente, da primeira potência de  $X$  na expressão de  $M_{l,r,s}$ , significa que a relação entre parâmetros e momentos, toma, muitas vezes, uma forma mais simples. Neste caso, então, quando usam-se momentos convencionais. Quando  $r$  e  $s$  são inteiros,  $F^r(1-F)^s$  pode ser expresso como uma combinação linear da potência de  $F$  ou  $(1-F)$ . Assim, é natural exprimir a distribuição através dos momentos  $M_{l,r,0}$ , ( $r = 0, 1, 2, \dots$ ) ou utilizando  $M_{l,0,s}$ , ( $s = 0, 1, 2, \dots$ ).

GREENWOOD et al. (1979) utilizaram a última aproximação, enquanto que HOSKING et al. (1985a), preferiram a primeira. Portanto, considerando a equação (37), os momentos definidos pela segunda aproximação, denotados por  $\beta_s$ , correspondem a:

$$\beta_s = M_{1,0,s} = \frac{1}{s+1} E[X_{1,s+1}] = \int_0^1 x(1-F)^s dF \quad (39)$$

onde

$$B[r+1, s+1] = \int_0^1 x^{(r+1)-1} (1-x)^{(s+1)-1} dx$$

$$B[1, s + 1] = \int_0^1 x^0 (1 - x)^{(s+1)-1} dx = \frac{1}{s + 1}$$

A estimativa não viesada de  $\beta_s$ ,  $b_s$ , de uma amostra de tamanho  $n$ , onde  $s$  é um inteiro não negativo, é obtida como segue (LANDWEHR et al., 1979).

Dada uma amostra aleatória de tamanho  $n$  de uma distribuição  $F$ , a estimação de  $\beta_s$  é mais convenientemente baseada na amostra ordenada  $x_1 \leq x_2 \leq \dots \leq x_n$ . O número total de possíveis caminhos para formar uma subamostra de tamanho  $(s+1)$  é:

$$N = \binom{n}{s + 1} \tag{40a}$$

Dessas  $N$  subamostras, o número que contém  $x_1$ , que corresponde ao número de possíveis subamostras em que  $x_1$  pode ser a menor observação, é:

$$N_1 = \binom{n - 1}{s} \tag{40b}$$

Similarmente, das  $N$  subamostras, o número que pode conter  $x_2$  mas não  $x_1$ , que corresponde ao número de possíveis subamostras em que  $x_2$  é a menor observação, é:

$$N_2 = \binom{n - 2}{s} \tag{40c}$$

Em geral, das  $N$  possíveis subamostras de tamanho  $(s+1)$ , retirada da amostra original de tamanho  $n$ , o número de subamostras que contém a  $i$ -ésima estatística de ordem  $x_i$ , mas, nenhuma das menores observação  $x_1, \dots, x_{i-1}$ , é:

$$N_i = \binom{n - i}{s} \tag{40d}$$

Note que se  $i \geq n - s + 1$ ,  $N_i = 0$ .

Dado que existem  $N$  total possíveis subamostras, e,  $N_i$  em que  $x_i$  pode ser a menor observação, uma estimativa da probabilidade que  $x_i$  é a menor observação em uma amostra de tamanho  $(s+1)$ , retirada de uma amostra de tamanho  $n$ , é:

$$P = \frac{N_i}{N} = \frac{\binom{n-i}{s}}{\binom{n}{s+1}} \quad (40e)$$

Enquanto que, uma estimativa da primeira estatística de ordem de uma subamostra de tamanho  $(s+1)$ , retirada de uma amostra de tamanho  $n$ , corresponde a:

$$E[X_{1,s+1} | n] = \sum_{i=1}^n x_i \frac{\binom{n-i}{s}}{\binom{n}{s+1}} \quad (40f)$$

Contudo, da relação estabelecida em (39), obtém-se:

$$b_s = \frac{1}{s+1} \sum_{i=1}^n x_i \frac{\binom{n-i}{s}}{\binom{n}{s+1}} = \frac{1}{n} \sum_{i=1}^n x_i \frac{\binom{n-i}{s}}{\binom{n-1}{s}}$$

ou

$$b_s = \frac{1}{n} \sum_{i=1}^{n-s} \frac{(i-1)(i-2)\dots(i-s)}{(n-1)(n-2)\dots(n-s)} x_i \quad (41)$$

onde  $b_s$  é um estimador não viesado  $\beta_s$  (Ver Apêndice I).

Portanto, para alguma distribuição, pode-se realizar estimativas amostrais não viesadas através dos PWMs, como segue:

$$b_0 = \bar{x} = \frac{1}{n} \sum_{j=1}^n x_j; \quad (42a)$$

$$b_1 = \sum_{j=1}^{n-1} \left[ \frac{(n-j)}{n(n-1)} \right] x_{(j)}; \quad (42b)$$

$$b_2 = \sum_{j=1}^{n-2} \left[ \frac{(n-j)(n-j-1)}{n(n-1)(n-2)} \right] x_{(j)}; \quad (42c)$$

$$b_3 = \sum_{j=1}^{n-3} \left[ \frac{(n-j)(n-j-1)(n-j-2)}{n(n-1)(n-2)(n-3)} \right] x_{(j)} \quad (42d)$$

onde  $x_{(j)}$  representa a ordenação dos extremos com  $x_{(1)}$  sendo a maior e  $x_{(n)}$  a menor observação.

a) *Estimativa dos parâmetros da GEV através dos PWMs*

Os momentos de pesos probabilísticos da distribuição GEV, para  $k \neq 0$ , são obtidos aplicando as equações (33a) e (36), logo, os PWMs da GEV são dados como segue:

$$\beta_r = M_{1,r,0} = \int_0^1 \left[ u + \alpha (1 - (-\ln F)^k) / k \right] F^r dF$$

fazendo  $z = -\ln F$ , onde  $F = e^{-z}$  e  $dF = -e^{-z} dz$ , a equação acima se transforma em:

$$\beta_r = - \int_0^{\infty} \left[ u + \alpha (1 - z^k) / k \right] e^{-(r+1)z} dz = (u + \alpha / k) \int_0^{\infty} e^{-(r+1)z} dz - \alpha / k \int_0^{\infty} z^k e^{-(r+1)z} dz$$

Desta forma

$$\beta_r = (u + \alpha / k)(r+1)^{-1} - (\alpha / k)(r+1)^{-1-k} \Gamma(1+k)$$

para valores de  $k > -1$ , tem-se:

$$\beta_r = (r+1)^{-1} \left[ u + \alpha \frac{1 - (r+1)^{-k} \Gamma(1+k)}{k} \right] \quad (43)$$

Quando  $k \leq -1$ ,  $\beta_0$  (média da distribuição) e os demais  $\beta_r$  não existem. Da equação (43), tem-se:

$$\beta_0 = u + \alpha \frac{1 - \Gamma(1+k)}{k} \quad (44a)$$

$$2\beta_1 - \beta_0 = \alpha \frac{\Gamma(1+k)(1 - 2^{-k})}{k} \quad (44b)$$

$$\frac{3\beta_2 - \beta_0}{2\beta_1 - \beta_0} = \frac{1 - 3^{-k}}{1 - 2^{-k}} \quad (44c)$$

Os estimadores de PWMs  $\hat{u}, \hat{\alpha}$  e  $\hat{k}$  dos parâmetros  $u, \alpha$  e  $k$ , são as soluções das equações (44a), (44b) e (44c) quando os  $\beta_r$  são substituídos através de seus estimadores  $b_r$ . Para obter  $k$ , é preciso resolver a equação:

$$\frac{3b_2 - b_0}{2b_1 - b_0} = \frac{1 - 3^{-k}}{1 - 2^{-k}} \quad (45)$$

Neste caso, a solução requer métodos iterativos. Porém, como a função  $(1-3^{-k})/(1-2^{-k})$  é praticamente linear no intervalo  $-1/2 < k < 1/2$ , o qual corresponde a faixa de valores de  $k$ , geralmente, encontrados na prática. Um estimador,  $\hat{k}$ , para  $k$  é dado com boa precisão, através da aproximação polinomial de baixa ordem (HOSKING et al., 1985a).

$$\hat{k} = 7.8590c + 2.9554c^2 \quad (46a)$$

onde

$$c = \frac{2b_1 - b_0}{3b_2 - b_0} - \frac{\log 2}{\log 3}$$

Obtido  $\hat{k}$ , os parâmetros de escala e de posicionamento podem ser estimados sucessivamente através das equações (44b) e (44a), resultando em:

$$\hat{\alpha} = \frac{(2b_1 - b_0)\hat{k}}{\Gamma(1 + \hat{k})(1 - 2^{-\hat{k}})} \quad (46b)$$

$$\hat{u} = b_0 + \hat{\alpha} \frac{\Gamma(1 + \hat{k}) - 1}{\hat{k}} \quad (46c)$$

Uma vez encontrados  $\hat{u}, \hat{\alpha}$  e  $\hat{k}$ , os quantis da distribuição GEV são estimados, usando a função inversa da distribuição, definida através da equação (33a).

O método de estimação dos momentos de pesos probabilísticos da distribuição GEV tem revelado possuir menores erros de estimação (tendência) e melhor eficiência do que a estimação através da máxima verossimilhança, quando o tamanho da amostra é pequeno (HOSKING et al., 1985a). O método dos momentos de pesos probabilísticos é, também, computacionalmente muito mais simples e mais robusto do que o da máxima verossimilhança.

Segundo HOSKING et al. (1986), os momentos com pesos probabilísticos são capazes de caracterizar teoricamente uma distribuição, entretanto, seus valores não possuem significado prático para identificar e comparar o posicionamento, a escala e forma de distribuições. Logo, para esta finalidade é mais conveniente utilizar combinações lineares dos PWMs, denominados momentos lineares ou momentos L.

### 5.1.2 Método dos Momentos L

Muitas técnicas estatísticas são baseadas no uso de combinações lineares de estatísticas de ordem, como pode ser visto em DAVID (1981), por exemplo. Contudo, não tem sido desenvolvida uma aproximação unificada do uso das estatísticas de ordem na análise estatística de distribuições de probabilidade univariada.

HOSKING (1990), com base em resultados teóricos e técnicos descritos, principalmente, por SILLITO (1951, 1969), DOWNTON (1966), CHAN (1967), KONHEIN (1971), MALLOWS (1973) e GREENWOOD et al. (1979) propôs o método dos momentos L, o qual representa a base para esta aproximação que faltava, como segue.

Seja, então,  $X$  uma variável aleatória real com função de distribuição  $F(x)$  e função de quantis  $x(F)$ , e sejam  $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$  as estatísticas de ordem de uma amostra aleatória de tamanho  $n$ , tomada através da distribuição de  $X$ . Define-se momentos L de  $X$ , como sendo (HOSKING, 1990):

$$\lambda_r \equiv r^{-1} \sum_{j=0}^{n-1} (-1)^j \binom{r-1}{j} EX_{r-j:n} \quad (r = 1, 2, \dots) \quad (47)$$

Onde  $\lambda_r$  é uma **função linear** das estatísticas de ordem esperadas. Daí, o uso do L em “momento L”. SILLITO (1969) deduziu momentos L, sem essa denominação, como coeficientes na aproximação de uma função de quantil através de polinômios. HOSKING (1990) preferiu olhar a equação (47) como uma definição fundamental, expressando as funções de quantis em termos de momentos L.

O valor esperado de uma estatística de ordem correspondente ao momento de ordem 1, é definido pela equação (35). Substituindo (35) em (47), procedendo a expansão binomial em  $F(x)$  e somando os coeficientes de cada potência de  $F(x)$ , obtém-se:

$$\lambda_r = \int_0^1 x(F) P_{r-1}^*(F) dF, \quad (r = 1, 2, \dots) \quad (48)$$

onde

$$P_r^*(F) = \sum_{j=0}^r p_{r,j}^* F^j \quad (49a)$$

e

$$p_{r,j}^* = (-1)^{r-j} \binom{r}{j} \binom{r+j}{j} = \frac{(-1)^{r-j} (r+j)!}{(j!)^2 (r-j)!} \quad (49b)$$

Logo, os quatros primeiros momentos L são:

$$\lambda_1 = EX = \int_0^1 x(F) dF; \quad (50a)$$

$$\lambda_2 = \frac{1}{2} E(X_{2:2} - X_{1:2}) = \int_0^1 x(F)(2F - 1) dF; \quad (50b)$$

$$\lambda_3 = \frac{1}{3} E(X_{3:3} - 2X_{2:3} + X_{1:3}) = \int_0^1 x(F)(6F^2 - 6F + 1) dF; \quad (50c)$$

$$\lambda_4 = \frac{1}{4} E(X_{4:4} - 3X_{3:4} + 3X_{2:4} - X_{1:4}) = \int_0^1 x(F)(20F^3 - 30F^2 + 12F - 1) dF \quad (50d)$$

onde  $X_{k:n}$  é a  $k$ -ésima menor estatística de ordem de uma amostra de tamanho  $n$ , retirada da distribuição de  $X$ .

Uma distribuição pode ser especificada através de seus momentos L, mesmo se alguns de seus momentos convencionais não exista. Assim, esta especificação é sempre única. O contrário não é verdade para momentos convencionais.  $\lambda_1$  é o primeiro momento L que coincide com a média ou valor esperado da variável aleatória. O segundo momento L,  $\lambda_2$ , é uma medida da escala ou dispersão da variável aleatória  $X$ . Ela é, muitas vezes, conveniente para padronizar os altos momentos  $\lambda_r$ ,  $r \geq 3$ , de modo a obter valores independentes da variável aleatória. Por isso, eles são independentes das unidades de medidas de  $X$ . Definem-se então as chamadas razões de momentos L de  $X$ , como sendo:

$$\tau_r \equiv \frac{\lambda_r}{\lambda_2}, \quad r = 3, 4, \dots \quad (51)$$

É possível definir uma função de momentos L que seja análogo ao coeficiente de variação (cv), isto é,  $\tau_2 = \lambda_2 / \lambda_1$  (coeficiente de variação L). Assim como, as razões de momentos L,  $\tau_3 = \lambda_3 / \lambda_2$ , corresponde ao coeficiente de assimetria L, e  $\tau_4 = \lambda_4 / \lambda_3$ , ao coeficiente de curtose L (HOSKING, 1990; STEDINGER et al., 1993; VOGEL & FENNESSY, 1993).

Os limites dos valores numéricos das razões de momentos L e coeficiente de variação L são dados por MALLOWS (1973) e HOSKING (1990), onde verifica-se que:

$$\lambda_2 > 0; \quad -1 < \tau_3 < 1; \quad \frac{1}{4}(5\tau_3^2 - 1) \leq \tau_4 < 1.$$

A vantagem de conhecer os limites das razões de momentos L é que, intuitivamente, é fácil interpretar uma medida tal como  $\tau_3$ , que é obrigado a ficar dentro do intervalo  $[-1,1]$ , do que a assimetria convencional, que pode tomar, arbitrariamente, grandes valores.

A mais simples aproximação para descrever momentos L é através da definição de momentos de pesos probabilísticos, onde momentos L são combinações lineares de PWMs (GREENWOOD et al., 1979; HOSKING, 1990). Todas as ordens mais altas de PWMs são, simplesmente, combinações lineares das estatísticas de ordem. Certas combinações lineares de PWMs podem ser interpretadas como medidas de posicionamento, escala e forma da distribuição de probabilidade. Esses são os momentos L definidos por HOSKING (1990), conforme equações (48), (49a) e (49b). Assim, momentos L são facilmente calculados em termos de PWMs, usando (HOSKING, 1990; VOGEL & FENNESSY, 1993, HOSKING & WALLIS, 1995):

$$\lambda_{r+1} = (-1)^r \sum_{k=0}^r P_{r,k}^* \beta_k = \sum_{k=0}^r \beta_r (-1)^{r-k} \binom{r}{k} \binom{r+k}{k} \quad (52)$$

onde  $P_{r-1}^*$  é definido conforme equação (49b).

Para alguma distribuição, os 4 primeiros momentos L, calculados a partir dos PWMs, utilizando a equação (52), são:

$$\lambda_1 = \beta_0; \quad (53a)$$

$$\lambda_2 = 2\beta_1 - \beta_0; \quad (53b)$$

$$\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0; \quad (53c)$$

$$\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0; \quad (53d)$$

logo, através das equações (43) e (53a), (53b), (53c) e (53d) obtém-se os quatro primeiros momentos L da distribuição GEV, quais sejam:

$$\lambda_1 = u + \frac{\alpha}{k} [1 - \Gamma(1+k)] \quad (54a)$$

$$\lambda_2 = \frac{\alpha}{k} [\Gamma(1+k)(1 - 2^{-k})] \quad (54b)$$

$$\lambda_3 = \frac{\alpha}{k} [\Gamma(1+k)(3 \times 2^{-k} - 2 \times 3^{-k} - 1)] \quad (54c)$$

$$\lambda_4 = \frac{\alpha}{k} [\Gamma(1+k)(1 - 6 \times 2^{-k} + 10 \times 3^{-k} - 5 \times 4^{-k})] \quad (54d)$$

Por outro lado, os 4 primeiros estimadores amostrais de momentos L não viesados são obtidos, substituindo  $\beta_r$  nas equações (53a), (53b), (53c) e (53d) pelos estimadores amostrais  $b_r$ , definidos através da equação (41).

a) *Estimativas dos parâmetros da distribuição GEV através dos momentos L*

HOSKING et al. (1985a) deduziram as expressões dos estimadores dos parâmetros da distribuição GEV,  $\hat{u}, \hat{\alpha}, \hat{k}$ . Como  $k$  não pode ser escrito explicitamente em função dos  $\beta_r$ , utiliza-se a seguinte solução aproximada para estimativa deste parâmetro, válida para  $-1/2 < k < 1/2$ :

$$\hat{k} = 7,8590c + 2,9554c^2 \quad (55a)$$

onde

$$c = \frac{2}{\hat{t}_3 + 3} - \frac{\ln 2}{\ln 3} = \frac{2}{\hat{t}_3 + 3} - 0,6309;$$

$$\hat{\alpha} = \frac{\hat{\lambda}_2 \hat{k}}{\Gamma(1+\hat{k})(1 - 2^{-\hat{k}})}; \quad (55b)$$

e

$$\hat{u} = \hat{\lambda}_1 + \frac{\hat{\alpha}[\Gamma(1+\hat{k}) - 1]}{\hat{k}}. \quad (55c)$$

### b) Razões de momentos e coeficiente de variação L

As razões de momentos L e o coeficiente de variação L da distribuição GEV são descritos, relacionando as equações (54a), (54b), (54c) e (54d) de acordo com as respectivas definições acima e, considerando a equação (51), como segue:

$$\tau_2 = \frac{\alpha(1-2^{-k})\Gamma(1+k)}{ku + \alpha[1-\Gamma(1+k)]} \quad (56a)$$

$$\tau_3 = \frac{2(1-3^{-k})}{(1-2^{-k})} - 3 \quad (56b)$$

$$\tau_4 = \frac{1-6 \times 2^{-k} + 10 \times 3^{-k} - 5 \times 4^{-k}}{1-2^{-k}} \quad (56c)$$

#### 5.1.3 Método dos Momentos LH

Momentos L foram introduzidos por HOSKING (1990) como uma alternativa aos momentos ordinários para caracterizar distribuições e dados (WANG, 1997b). A análise estatística de extremos é, muitas vezes, conduzida para prever eventos de grande período de retorno, onde o mais relevante para a análise é a cauda superior da distribuição e os eventos amostrais mais extremos. Neste sentido, momentos e razões de momentos têm sido criticados, por serem, também, sensíveis a cauda superior da distribuição e, portanto, aos outliers amostrais (VOGEL & FENNESSY, 1993). O problema é que os momentos L são sensíveis à parte inferior dos dados e dá insuficiente peso aos valores amostrais mais altos, faixa que contém informação útil sobre a parte superior da cauda da distribuição (BOBÉE et al., 1995).

Com base nessas questões e seguindo o mesmo raciocínio empregado na definição de momentos de altos pesos probabilísticos, WANG (1997b) propôs o método baseado na combinação linear das mais elevadas estatísticas de ordem, que representa uma generalização dos momentos L, chamado Momentos LH. Neste método, tal como nos momentos de altos pesos probabilísticos, busca-se a caracterização da parte superior da distribuição e dos maiores eventos amostrais.

Dada uma amostra de tamanho  $m$  retirada de uma distribuição  $F(x)=P(X\leq x)$ . Baseado na combinação linear das mais elevadas estatísticas de ordem e na equação (35), os momentos LH são definidos como:

$$\lambda_1^\eta = E[X_{(\eta+1)(\eta+1)}] \quad (57a)$$

$$\lambda_2^\eta = \frac{1}{2} E[X_{(\eta+2)(\eta+2)} - X_{(\eta+1)(\eta+2)}] \quad (57b)$$

$$\lambda_3^\eta = \frac{1}{3} E[X_{(\eta+3)(\eta+3)} - 2X_{(\eta+2)(\eta+3)} + X_{(\eta+1)(\eta+3)}] \quad (57c)$$

$$\lambda_4^\eta = \frac{1}{4} E[X_{(\eta+4)(\eta+4)} - 3X_{(\eta+3)(\eta+4)} + 3X_{(\eta+2)(\eta+4)} - X_{(\eta+1)(\eta+4)}] \quad (57d)$$

onde  $\lambda_1^\eta$ , maior valor esperado na amostra de tamanho  $\eta + 1$ , corresponde a uma medida de posicionamento da distribuição;  $\lambda_2^\eta$ , metade da diferença entre o maior e segundo maior valor esperado na amostra de tamanho  $\eta + 2$ , caracteriza a expansão da parte superior da distribuição;  $\lambda_3^\eta$ , reflete como está a assimetria da parte superior da distribuição, através dos três maiores valores esperados na amostra de tamanho  $\eta + 3$ ; e  $\lambda_4^\eta$  prover uma medida da pontiagudez da parte superior da distribuição através dos quatros maiores valores esperados na amostra de tamanho  $\eta + 4$ .

Quando  $\eta = 0$ , momentos LH tornam-se iguais aos momentos L. Como  $\eta$  aumenta, os momentos LH refletem mais e mais as características da parte superior da distribuição e dos valores extremos máximos dos dados. Momentos LH são chamados momentos  $L_1$ , momentos  $L_2$ , .... para  $\eta = 1, 2, \dots$  respectivamente. Normalizando os momentos LH, obtém-se o coeficiente de variação LH ( $\tau_2^\eta$ ), assimetria ( $\tau_3^\eta$ ) e curtose ( $\tau_4^\eta$ ), respectivamente, como:

$$\tau_2^\eta = \frac{\lambda_2^\eta}{\lambda_1^\eta}; \quad \tau_3^\eta = \frac{\lambda_3^\eta}{\lambda_2^\eta}; \quad \tau_4^\eta = \frac{\lambda_4^\eta}{\lambda_2^\eta} \quad (58)$$

ordenando a amostra em  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ , a estimativa dos momentos LH é feita como segue:

$$\hat{\lambda}_1^\eta = \frac{1}{n C_{\eta+1}} \sum_{i=1}^n {}^{i-1}C_\eta x_{(i)} \quad (59a)$$

$$\hat{\lambda}_2^\eta = \frac{1}{2} \frac{1}{n C_{\eta+2}} \sum_{i=1}^n ({}^{i-1}C_{\eta+1} - {}^{i-1}C_\eta {}^{n-1}C_1) x_{(i)} \quad (59b)$$

$$\hat{\lambda}_3^\eta = \frac{1}{3} \frac{1}{n C_{\eta+3}} \sum_{i=1}^n ({}^{i-1}C_{\eta+2} - 2{}^{i-1}C_{\eta+1} {}^{n-i}C_1 + {}^{i-1}C_\eta {}^{n-i}C_2) x_{(i)} \quad (59c)$$

$$\hat{\lambda}_4^\eta = \frac{1}{4} \frac{1}{n C_{\eta+4}} \sum_{i=1}^n ({}^{i-1}C_{\eta+3} - 3{}^{i-1}C_{\eta+2} {}^{n-i}C_1 + 3{}^{i-1}C_{\eta+1} {}^{n-i}C_2 - {}^{i-1}C_\eta {}^{n-i}C_3) x_{(i)} \quad (59d)$$

onde

$${}^m C_j = \binom{m}{j} = \frac{m!}{j!(m-j)!} \quad \text{e} \quad {}^m C_j = 0 \quad \text{quando} \quad j > m.$$

a) *Estimativas dos parâmetros da distribuição GEV:*

Dada uma amostra, os três parâmetros  $u, \alpha, \kappa$  da distribuição GEV podem ser estimados, considerando a estimativa dos momentos LH amostrais através das equações (59a), (59b), (59c) e (59d), para um valor selecionado de  $\eta$  e  $\kappa \neq 0$ , como segue (WANG, 1977b):

$$\lambda_1^\eta = u + \frac{\alpha}{\kappa} \left[ 1 - \Gamma(1 + \kappa)(\eta + 1)^{-\kappa} \right] \quad (60a)$$

$$\lambda_2^\eta = \frac{(\eta + 2)\alpha\Gamma(1 + \kappa)}{2!\kappa} \left[ -(\eta + 2)^{-\kappa} + (\eta + 1)^{-\kappa} \right] \quad (60b)$$

$$\lambda_3^\eta = \frac{(\eta + 3)\alpha\Gamma(1 + \kappa)}{3!\kappa} \left[ -(\eta + 4)(\eta + 3)^{-\kappa} + 2(\eta + 3)(\eta + 2)^{-\kappa} \right. \\ \left. - (\eta + 2)(\eta + 1)^{-\kappa} \right] \quad (60c)$$

$$\lambda_4^\eta = \frac{(\eta + 4)\alpha\Gamma(1 + \kappa)}{4!\kappa} \left[ -(\eta + 6)(\eta + 5)(\eta + 4)^{-\kappa} + 3(\eta + 5)(\eta + 4)(\eta + 3)^{-\kappa} \right. \\ \left. - 3(\eta + 4)(\eta + 3)(\eta + 2)^{-\kappa} + (\eta + 3)(\eta + 2)(\eta + 1)^{-\kappa} \right] \quad (60d)$$

Assim, os parâmetros  $u$ ,  $\alpha$  e  $k$  da distribuição GEV podem estar ser estimados, substituindo os três primeiros momentos LH nas equações (60a), (60b) e (60c) pelo os seus respectivos estimadores amostrais nas equações (59a), (59b) e (59c), para cada valor de  $\eta$  selecionado.

Para facilitar o procedimento computacional, WANG, (1977b) propôs uma equação aproximada para o cálculo de  $k$ , tomando como base as equações (60b) e (60c) e a equação (58) que define  $\tau_3^\eta$ , a qual corresponde a:

$$k = \alpha_0 + \alpha_1 [\tau_3^\eta] + \alpha_2 [\tau_3^\eta]^2 + \alpha_3 [\tau_3^\eta]^3 \quad (61)$$

onde os coeficientes  $\alpha_0$ ,  $\alpha_2$  e  $\alpha_3$  variam em função de  $\eta$  (Tabela 3). Uma vez que  $k$  é obtido de (61), as equações (60b) e (60a) fornecem respectivamente  $\alpha$  e  $u$ .

TABELA 3 - Valores dos coeficientes da equação 61\*.

$\eta$	$\alpha_0$	$\alpha_1$	$\alpha_2$	$\alpha_3$
0	0,2849	-1,8213	0,8140	-0,2835
1	0,4823	-2,1494	0,7269	-0,2103
2	0,5914	-2,2351	0,6442	-0,1616
3	0,6618	-2,4548	0,5733	-0,1273
4	0,7113	-2,5383	0,5142	-0,1027

\*WANG (1997)

Análises de dados observados e dados obtidos via simulação Monte Carlo, mostraram que momentos LH reduzem as influências indesejáveis que os menores eventos amostrais podem exercer na estimação de eventos com grandes períodos de retorno, comparado ao uso de momentos L (WANG, 1997b).

#### b) Razões de momentos e coeficiente de variação LH

As razões de momentos e coeficiente de variação LH da distribuição GEV são calculadas de acordo com as relações em (58) e equações (60a), (60b), (60c) e

(60d), resultando nas seguintes expressões:

$$\tau_2^\eta = \frac{1}{2} \frac{(\eta+2)\alpha\Gamma(1+k)\left[-(\eta+2)^{-k} + (\eta+1)^{-k}\right]}{ku + \alpha\left[1 - \Gamma(1+k)(\eta+1)^{-k}\right]} \quad (62a)$$

$$\tau_3^\eta = \frac{1}{3} \frac{(\eta+3)\left[-(\eta+4)(\eta+3)^{-k} + 2(\eta+3)(\eta+2)^{-k} - (\eta+2)(\eta+1)^{-k}\right]}{(\eta+2)\left[-(\eta+2)^{-k} + (\eta+1)^{-k}\right]} \quad (62b)$$

$$\tau_4^\eta = \frac{1}{12} \frac{(\eta+4)\left[-(\eta+6)(\eta+5)(\eta+4)^{-k} + 3(\eta+5)(\eta+4)(\eta+3)^{-k}\right]}{(\eta+2)\left[-(\eta+2)^{-k} + (\eta+1)^{-k}\right]} + \left(\frac{-3(\eta+4)(\eta+3)(\eta+2)^{-k} + (\eta+3)(\eta+2)(\eta+1)^{-k}}{-(\eta+2)^{-k} + (\eta+1)^{-k}}\right) \quad (62c)$$

onde  $\tau_2^\eta, \tau_3^\eta, \tau_4^\eta$  são, respectivamente, o coeficiente de variação LH, assimetria LH e curtose LH. As razões de momentos e coeficiente de variação LH amostrais são obtidas, relacionando os momentos LH amostrais, indicados através das equações (59a), (59b), (59c) e (59d) seguindo as fórmulas em (58), onde os momentos LH da distribuição GEV são substituídos pelo os seus respectivos estimadores.

## 5.2 TESTES DE QUALIDADE DE AJUSTES

### 5.2.1 Diagrama de Momentos L

O diagrama de momentos L ilustra graficamente a relação entre a curtose L e a assimetria L, tanto da distribuição de probabilidade como da amostra em questão, ou seja, ele compara estimativas de razões de momentos L teóricos com suas contrapartes amostrais. Vários estudos recentes documentam que diagrama de momentos L são muito úteis para avaliar a qualidade de ajuste da distribuição de probabilidade à séries de dados amostrais (HOSKING, 1990; CHOWDHURY et al., 1991; VOGEL & FENNESSEY, 1993; GUTTMAN et al., 1993; VEGEL & WILSON, 1995; GUTTMAN, 1996; HOSKING & WALLIS, 1997) e muitos outros.

O diagrama de momentos L permite comparar várias distribuições, usando o mesmo gráfico. As mesmas considerações são válidas para diagrama de momentos LH. Porém, como momentos LH refletem as características da parte superior de distribuições mais do que momentos L, diagrama de momentos LH pode ser usado para checar se os eventos mais altos numa amostra estão bem descritos através de uma distribuição particular (WANG, 1997b).

### 5.2.2 Teste de Hipótese Sobre o Sinal do Parâmetro de Forma k da Distribuição GEV

A distribuição de valores extremos VEI é um caso especial da distribuição GEV. Ela é, muitas vezes, útil para testar se uma série de dados é gerada através de uma Gumbel ou através de uma GEV. Isto é equivalente a testar se o parâmetro de forma k é zero na distribuição GEV. Assim, um teste de hipótese pode ser baseado na estimativa de k feita através de momentos de peso probabilísticos ou de momentos L. Na hipótese nula,  $H_0: k = 0$ ,  $\hat{k}$  estimado é assintoticamente distribuído através da distribuição normal com média zero e variância  $0,5633/n$ , onde n é o tamanho da amostra. Portanto, o teste pode ser realizado, comparando a seguinte estatística:

$$Z_H = \hat{k} \left( \frac{n}{0,5633} \right)^{\frac{1}{2}} \quad (63)$$

com valores críticos de uma distribuição normal padrão. Valores significativamente positivos de  $Z_H$  implica numa rejeição de  $H_0$  em favor da hipótese alternativa de  $k > 0$ , que corresponde a forma assintótica VEIII. Quando  $Z_H$  for significativamente negativo, a rejeição de  $H_0$  se faz em favor de  $k < 0$ , que corresponde a forma VEII (HOSKING, et al., 1985a; HOSKING, 1990).

HOSKING (1990) mostrou que a distribuição amostral do coeficiente de variação, da assimetria e curtose estimadas através de momentos L, é assintoticamente normal. Ele encontrou que para uma população normal a estatística  $\tau_3$  (assimetria L) é assintoticamente normal com média zero e variância  $0,1866/n$ . Quando n é pequeno, a melhor aproximação para a variância é  $\nu_n = 0,1866/n + 0,8/n^2$ . Assim, HOSKING (1990) propôs:

$$N_n = \nu_n \hat{\tau}_3 \quad (64)$$

como um teste para normalidade contra alternativas assimétricas, onde os valores críticos são obtidos através de referência com a distribuição normal padrão.

CHOWDHURY et al., (1991) propuseram alguns testes mais gerais que avaliam se o parâmetro de forma  $k$  de uma distribuição GEV ajustada é significativamente diferente de um dado valor conhecido. Entre eles estão um teste de Kolmogorov-Smirnov, um teste de correlação de plotagem de probabilidade e um teste baseado na assimetria L.

### 5.2.3 Teste de Qualidade de Ajuste da Distribuição GEV Via Momentos LH

A distribuição GEV pode ser ajustada para uma série de dados, igualando os seus três primeiros momentos LH aos respectivos momentos LH amostrais, como mostrado anteriormente (WANG, 1997b). A curtose LH da população ( $\tau_4^n$ ) é uma função da assimetria LH populacional ( $\tau_3^n$ ), onde ambos dependem apenas do parâmetro de forma  $k$ . Como o valor estimado da curtose LH amostral ( $\hat{\tau}_4^n$ ) não é usado no ajuste da distribuição, para determinar qual é a chance, estatisticamente pequena, de  $\hat{\tau}_4^n$  ocorrer se a verdadeira distribuição subjacente é uma distribuição GEV. Dado um particular estimador amostral  $\hat{\tau}_3^n$ , precisa-se conhecer  $p(\hat{\tau}_4^n | \hat{\tau}_3^n)$ . Porém, não é tão simples encontrar  $p(\hat{\tau}_4^n | \hat{\tau}_3^n)$  quando a mesma depende de  $\tau_3^n$  da população. Contudo, dado  $\hat{\tau}_3^n$ , é possível inferir  $\tau_3^n$  populacional, usando o teorema de Bayes, mostrado a seguir:

$$p(\tau_3^n | \hat{\tau}_3^n) \propto p(\hat{\tau}_3^n | \tau_3^n) p(\tau_3^n) \quad (65)$$

onde  $p(\tau_3^n)$  é uma distribuição a priori que pode ser informativa ou não informativa. Assim, dado  $p(\tau_3^n | \hat{\tau}_3^n)$  encontra-se:

$$p(\hat{\tau}_4^n | \hat{\tau}_3^n) = \int P(\hat{\tau}_4^n | \tau_3^n, \hat{\tau}_3^n) p(\tau_3^n | \hat{\tau}_3^n) d\tau_3^n \quad (66)$$

onde  $p(\hat{\tau}_3^\eta | \tau_3^\eta)$  em (65) e  $p(\hat{\tau}_4^\eta | \tau_3^\eta, \hat{\tau}_3^\eta)$  em (66) podem ser derivada de  $p(\hat{\tau}_4^\eta, \hat{\tau}_3^\eta | \tau_3^\eta)$ , principalmente, utilizando simulação Monte Carlo. Portanto, teoricamente, é possível comparar  $\hat{\tau}_4^\eta$  com  $p(\hat{\tau}_4^\eta | \hat{\tau}_3^\eta)$  para inferir se a distribuição subjacente é significativamente diferente da distribuição GEV. Não obstante, proceder por tal inferência requer muito esforço computacional que, em geral, não é prático (WANG, 1998).

WANG (1998) desenvolveu um teste de qualidade de ajuste da GEV com base em  $p(\hat{\tau}_4^\eta | \tau_3^\eta = \hat{\tau}_3^\eta)$  como uma aproximação de  $p(\hat{\tau}_4^\eta | \tau_3^\eta, \hat{\tau}_3^\eta)$ , além de assumir que, as distribuições de  $\hat{\tau}_3^\eta$  e de  $\hat{\tau}_4^\eta$  da GEV, seguem uma distribuição conjunta normal. Para descrição completa da distribuição conjunta normal, precisa-se conhecer a média, desvio padrão e coeficiente de correlação dos estimadores amostrais  $\hat{\tau}_3^\eta$  e  $\hat{\tau}_4^\eta$ . Suas médias são assumidas para serem os valores populacionais de  $\tau_3^\eta$  e  $\tau_4^\eta$ , respectivamente, onde é negligenciado algum erro de estimação. Os desvios padrão e coeficiente de correlação, denotados como  $\sigma(\hat{\tau}_3^\eta)$ ,  $\sigma(\hat{\tau}_4^\eta)$  e  $\rho(\hat{\tau}_3^\eta, \hat{\tau}_4^\eta)$ , respectivamente, ambos são funções de  $\tau_3^\eta$  e do tamanho da amostra e podem ser encontrados através de simulação Monte Carlo.

A distribuição condicional amostral de  $\hat{\tau}_4^\eta$  quando  $\tau_3^\eta = \hat{\tau}_3^\eta$  é normalmente distribuída com média  $\tau_4^\eta$  e desvio padrão dado como segue (WANG, 1998):

$$\sigma(\hat{\tau}_4^\eta | \hat{\tau}_3^\eta = \tau_3^\eta) = \sigma(\hat{\tau}_4^\eta) \left[ 1 - \rho^2(\hat{\tau}_3^\eta, \hat{\tau}_4^\eta) \right]^{1/2} \quad (67)$$

Um teste de hipótese de que uma série de dados vem da distribuição ajustada pode ser conduzido na base da estimativa amostral  $\hat{\tau}_4^\eta$  através da comparação da seguinte estatística (WANG, 1998):

$$Z_W = \frac{\hat{\tau}_4^\eta - \tau_4^\eta}{\sigma(\hat{\tau}_4^\eta | \hat{\tau}_3^\eta = \tau_3^\eta)} \quad (68)$$

com valores críticos de uma distribuição normal padrão.

O desvio padrão em (68),  $\sigma(\hat{\tau}_4^\eta | \hat{\tau}_3^\eta = \tau_3^\eta)$ , é função de  $\tau_3^\eta$  e do tamanho da amostra e pode ser calculado com  $\sigma(\hat{\tau}_4^\eta)$  e  $\rho(\hat{\tau}_3^\eta, \hat{\tau}_4^\eta)$  através da equação (67), usando simulação Monte Carlo. Para evitar o enorme esforço computacional envolvido nos várias fases do teste, WANG (1998) propôs a seguinte aproximação:

$$\sigma^2(\hat{\tau}_4^\eta | \hat{\tau}_3^\eta = \tau_3^\eta) = \frac{b}{n} + \frac{c}{n^2} \quad (69)$$

onde

$$b = b_0 + b_1[\tau_3^\eta] + b_2[\tau_3^\eta]^2 + b_3[\tau_3^\eta]^3 + b_4[\tau_3^\eta]^4 \quad (70)$$

$$c = c_0 + c_1[\tau_3^\eta] + c_2[\tau_3^\eta]^2 + c_3[\tau_3^\eta]^3 + c_4[\tau_3^\eta]^4 \quad (71)$$

os coeficientes em (70) e (71) variam com os valores de  $\eta$  e estão apresentados nas Tabelas 4 e 5.

TABELA 4 - Valores dos coeficientes da equação (70)\*

$\eta$	$b_0$	$b_1$	$b_2$	$b_3$	$b_4$
0	0.0745	0.0555	0.0067	-0.3090	0.2240
1	0.0579	-0.0328	0.1524	-0.4102	0.2672
2	0.0488	-0.0527	0.1620	-0.3856	0.2566
3	0.0380	-0.0309	0.0354	-0.1233	0.0878
4	0.0241	0.0024	-0.0813	0.0733	-0.0210

\* WANG (1998)

TABELA 5 - Valores dos coeficientes da equação (71)\*

$\eta$	$c_0$	$c_1$	$c_2$	$c_3$	$c_4$
0	1.0100	-0.0282	-2.9336	4.0801	-1.0874
1	1.3403	-0.8291	-3.8777	9.5371	-5.7866
2	1.8800	-2.2233	-2.5825	10.4350	-7.3887
3	2.6784	-4.8418	3.5255	2.3736	-3.2076
4	3.7793	-8.3485	11.5170	-7.9095	1.9459

\* WANG (1998)

A equação (69) tem a mesma forma daquela utilizada por CHOWDHURY et al.,(1991) para coeficiente de variação L e assimetria L. A mesma assegura que a variância é assintótica e inversamente proporcional ao tamanho da amostra. O segundo termo assegura tal efeito em amostras pequenas.

### 5.3 DISTRIBUIÇÃO INICIAL DE PROBABILIDADE LIMITADA

Uma função distribuição inicial limitada pode ser expressa como (GUMBEL, 1958):

$$F_x(x) = 1 - \frac{1}{n} \left( \frac{\omega - x}{\omega - w_n} \right)^\theta \quad (72)$$

onde,  $\omega$  é o valor limite superior de X, tal que,  $F_x(\omega) = 1$ . Os valores extremos  $X_n$  de uma variável inicial X, convergem para a distribuição assintótica VEIII (Tabela 1). Aplicado a equação (4), a função distribuição de valores extremos  $X_n$  torna-se:

$$F_{X_n}(x) = \left[ 1 - \frac{1}{n} \left( \frac{\omega - x}{\omega - w_n} \right)^\theta \right]^n$$

aplicando o logaritmo em ambos os lados da equação acima, obtém-se:

$$\ln[F_{X_n}(x)] = n \ln \left[ 1 - \frac{1}{n} \left( \frac{\omega - x}{\omega - w_n} \right)^\theta \right] = \frac{\ln \left[ 1 - \frac{1}{n} \left( \frac{\omega - x}{\omega - w_n} \right)^\theta \right]}{\frac{1}{n}}$$

aplicando o limite quando  $n \rightarrow \infty$ , a expressão resultando é

$$\lim_{n \rightarrow \infty} \ln[F_{X_n}(x)] = - \left( \frac{\omega - x}{\omega - w_n} \right)^\theta$$

logo

$$F_{x_n}(x) = \exp \left[ - \left( \frac{\omega - x}{\omega - w_n} \right)^\theta \right] \quad (73)$$

que corresponde a forma assintótica VEIII. Comparando a equação(73) com a forma assintótica VEIII da distribuição GEV (equação 32a), obtém-se as seguintes relações matemáticas entre seus parâmetros:

$$\omega = u + \frac{\alpha}{k}, \quad \theta = \frac{1}{k} \quad \text{e} \quad W_n = u \quad (74)$$

onde  $u$ ,  $\alpha$  e  $k$  são, respectivamente, parâmetros de posicionamento, escala e forma da distribuição GEV com  $k > 0$ , enquanto que  $n$  é o tamanho da amostra com distribuição de probabilidade inicial limitada.

Com a substituição das equações (73) em (74), após proceder alguns mudanças, obtém-se a equação da distribuição limitada em função dos parâmetros da GEV, como segue:

$$F(x) = 1 - \frac{1}{n} \left[ 1 - k \frac{x - u}{\alpha} \right]^{\frac{1}{k}} \quad (75)$$

Tomando a forma inversa da equação (75), tem-se:

$$x = u + \frac{\alpha}{k} - \frac{\alpha}{k} [n(1 - F(x))]^k, \quad x \leq u + \frac{\alpha}{k} \quad (76)$$

### 5.3.1 Estimação dos Parâmetros da Distribuição Inicial Limitada

Não há registro do uso da distribuição de probabilidade limitada em hidrologia, a não ser como exemplo de distribuição inicial, onde seus eventos extremos convergem assintoticamente para a distribuição GEV VEIII (GUMBEL, 1958; ANG & TANG, 1984). Logo, para proceder a estimação dos parâmetros da

distribuição inicial limitada, o autor deste trabalho desenvolveu um estimador não viesado de momentos com pesos probabilísticos para referida distribuição, como segue:

Os momentos de pesos probabilísticos introduzidos por GREENWOOD et al., (1979) são definidos de acordo com a equação (36). Um estimador não viesado do  $r$ -ésimo PWMs da distribuição inicial de probabilidade limitada é obtido, combinando as equações (36) e (76), que resulta na expressão:

$$\beta_r = \int_0^1 \left[ u + \frac{\alpha}{k} - \frac{\alpha}{k} \{n(1-F)\}^k \right] F^r dF \quad (77)$$

Resolvendo a equação (77), obtém-se o citado estimador, como mostrado a seguir:

$$\beta_r = (r+1)^{-1} \left( u + \frac{\alpha}{k} \left[ 1 - \frac{\Gamma(r+2)n^k}{(k+1)(k+2)\cdots(k+r+1)} \right] \right), \quad r = 0, 1, 2, \dots \quad (78)$$

Uma estimativa não viesada de  $\beta_r$ ,  $b_r$ , de uma amostra de tamanho  $n$ , onde  $r$  é um inteiro não negativo, é dada através da equação (41) proposta por LANDWEHR et al., (1979). Substituindo os quatro primeiros momentos amostrais, definidos através das equação (42a), (42b), (42c) e (42d), na equação (78) para  $r = 0, 1, 2$  e  $3$ , obtém-se as expressões dos estimadores dos parâmetros  $u$ ,  $\alpha$  e  $k$  da distribuição inicial limitada (equação 75), como segue:

$$\hat{k} = \frac{2(2b_1 - b_0)}{3b_2 - 2b_1} - 3 \quad (79)$$

$$\hat{\alpha} = (2b_1 - b_0) \frac{(k+1)(k+2)}{n^k} \quad (80)$$

$$\hat{u} = b_0 + \frac{\alpha}{k} \left( \frac{n^k}{(k+1)} - 1 \right) \quad (81)$$

onde  $\hat{k}$ ,  $\hat{\alpha}$  e  $\hat{u}$  correspondem as estimativas dos parâmetros  $k$ ,  $\alpha$  e  $u$  da distribuição inicial de probabilidade limitada, os quais representam, teoricamente, os parâmetros da distribuição GEV que descreve os valores máximo da referida distribuição limitada para um  $n$  grande.

#### 5.4 POSIÇÃO DE PLOTAGEM DOS DADOS

A análise gráfica da adequação de um ajuste de distribuição é, geralmente, realizada através da plotagem das observações juntamente com quantis obtidos através da distribuição em estudo, na mesma faixa de frequência das observações. Isto pode ser feito com o uso de papel de probabilidade disponíveis para algumas distribuições, ou utilizando a técnica mais geral descrita a seguir.

Para uma série de valores observados  $\{X_i\}$ , onde  $X_{(i)}$  é o  $i$ -ésimo maior valor de uma amostra com  $X_{(n)} \leq X_{(n-1)} \leq \dots \leq X_{(1)}$ . Define-se a variável aleatória  $U_i$  como:

$$U_i = 1 - F_X[X_{(i)}] \quad (82)$$

que corresponde a probabilidade de excedência associada a  $i$ -ésima maior observação. Se  $\{X_i\}$  for independente, em repetidas amostragens,  $U_i$  tem distribuição de probabilidade beta com média:

$$E[U_i] = \frac{i}{n+1} \quad (83)$$

e variância

$$Var[U_i] = \frac{i(n-i+1)}{(n+1)^2(n+2)} \quad (84)$$

Conhecendo a distribuição de probabilidade de excedência  $U_i$ , é possível desenvolver estimadores  $q_i$  de seus valores, que podem ser usados para plotar cada valor  $X_{(i)}$  numa escala de probabilidade. Assim, se  $G(X)$  é a distribuição de  $X$ , o valor de  $X_{(i)} = G^{-1}(1-U_i)$  se aproxima de  $G^{-1}(1-q_i)$ , onde a posição de plotagem  $q_i$  é a

estimativa de  $U_i$ .

Diferentes posições de plotagem que sempre buscam alcançar quantis não viesados para diferentes distribuição, podem ser escritos como:

$$q_i = \frac{i - a}{n + 1 - 2a} \quad (85)$$

diferentes valores de  $a$  são propostos por vários autores, dando origem a seus modelos, como resumido na Tabela 6 (STEDINGER, 1993). Nestes casos, o período de retorno para a maior observação é  $T_1=1/q_1$ .

TABELA 6 - Alternativas de posição de plotagem e suas motivações\* .

Autor	Fórmula	$\alpha$ eq. (85)	Período de Retorno: $T_1$	Motivação
Weibull	$\frac{i}{n+1}$	0	$n+1$	Probabilidade de excedência não viesada para todas distribuições
Median	$\frac{i - 0.3175}{n + 0.365}$	0.3175	$1.57n+0.5$	Probabilidade de excedência mediana para todas distribuições
APL	$\frac{i - 0.35}{n}$	$\sim 0.35$	$1.54n$	Usada com momentos de pesos probabilísticos
Blom	$\frac{i - 3/8}{n + 1/4}$	0.375	$1.6n+0.4$	Quantis normais não viesados
Cunnane	$\frac{i - 0.40}{n + 0.2}$	0.40	$1.67n+0.3$	Quantis não viesados aproximadamente
Gringorten	$\frac{i - 0.44}{n + 0.12}$	0.44	$1.79n+0.2$	Otimizada para distribuição Gumbel
Hazen	$\frac{i - 0.5}{n}$	0.5	$2n$	Uma escolha tradicional

\* STEDINGER, 1993

## 5.5 OUTLIERS

Em análise de freqüência de cheias o termo outliers é comumente usado para denotar grandes valores de cheias presentes nos registros observados que encontram-se majoritariamente acima dos demais valores amostrais. A análise de freqüência de cheias torna-se mais complicada na presença de valores atípicos e quando os mesmos são simplesmente descartados, os resultados de estimação podem ser distorcidos (KIRBY, 1974). Para evitar alguma subjetividade na detecção e tratamento de outliers, métodos baseados em princípios de teste de hipóteses estatísticas tem sido utilizados para determinação de freqüência de cheias.

### 5.5.1 Princípios estatísticos do teste de outlier

Em uma série de observações  $(X_1, X_2, \dots, X_n)$  em que  $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$  denota a série de observações ordenadas, onde  $n$  é o tamanho da amostra. Se o valor  $X_{(n)}$  é muito alto, o mesmo é suspeito para ser um outlier. Contudo, para fazer um julgamento objetivo, uma análise é feita usando teste de hipóteses estatísticas. Para construir um teste estatístico que expresse a diferença entre uma observação extraordinária e um outro valor em uma amostra, considera-se que sob a hipótese nula não há outlier nessa amostra. A distribuição desse teste estatístico é derivado analiticamente ou numericamente para determinar um limite para um nível de significância preestabelecido. Se o valor da estatística calculada é maior do que este limite, a observação extraordinária é, então, declarada para ser um outlier. Alguns dos testes estatísticos utilizados são:

a) *A estatística da região studentizada,  $T_n$ :*

$$T_n = \frac{X_{(n)} - \bar{X}}{S}$$

b) *A estatística de Dixon,  $D_n$ :*

$$D_n = \frac{X_{(n)} - X_{(n-1)}}{X_{(n)} - X_{(1)}}$$

c) *A estatística de assimetria,  $G_n$ :*

$$G_n = \frac{\sqrt{n} \sum_{i=1}^n (X_i - \bar{X})^3}{\left[ \sum_{i=1}^n (X_i - \bar{X})^2 \right]^{3/2}}$$

onde  $\bar{X}$  e  $S$  são, respectivamente, a média e o desvio padrão da amostra. Diferentes métodos de teste outlier podem ser obtidos, construindo diferentes teste estatísticos e avaliando sua distribuição. Por exemplo, a U.S. WATER RESOURCES COUNCIL (1981) usa a estatística  $T_n$  para detectar valores extraordinários em registro de cheias.

Para explicar o aparecimento de outlier em amostras é necessário especificar uma hipótese alternativa, que corresponde a um modelo de geração do mesmo. Algumas hipóteses alternativas usadas para explicar a existência de outlier são descritas como segue.

- *Alternativa distribucional:* sob esta alternativa a hipótese nula  $H$  é expressa como  $H:F$  que significa que todas as observações surgem de uma distribuição  $F$ . A hipótese alternativa, escolhida adequadamente, explica que  $\hat{H}:G$ , o que significa que todas as observações surgem de uma distribuição  $G$ , em que observações extremamente altas não são suspeitas de ser outliers.

- *Alternativa mista:* é expressa como  $H:F$  versus  $\hat{H}:(1-\lambda)F+\lambda G$ . Na hipótese nula todas as observações surgem de uma distribuição  $F$ , enquanto que na hipótese alternativa, o aparecimento de outlier é devido ao fato de que uma pequena fração das observações surge de uma distribuição  $G$ , que difere um pouco da distribuição inicial  $F$ .

- *Alternativa slippage:* esta é usada, principalmente, para descrever a presença de outlier quando há mudança de valores de parâmetros sem, no entanto, alterar a forma da distribuição. Supondo a hipótese nula  $H$  com uma alternativa  $\hat{H}$ . O teste assume

$X_{(n)}$  como sendo um outlier se  $T_n > T_{n,\alpha}$  onde  $T_{n,\alpha}$  é o valor limite para um preestabelecido nível de significância definido como:

$$\alpha = \Pr[T_n > T_{n,\alpha} / H]$$

onde  $\alpha$  é chamado erro do tipo I se a hipótese nula for verdadeira, isto é, nenhum outlier está presente na amostra. Com a consideração de  $\hat{H}$  como uma alternativa, a seguinte função pode ser definida:

$$PF = \Pr[T_n > T_{n,\alpha} / \hat{H}]$$

onde PF se relaciona com o erro tipo II  $\beta$ , com  $\beta = 1 - PF$  (SIYI, 1987). Informações úteis sobre o desempenho de alguns destes testes são encontradas em BARNETT E LEWIS (1978).

A análise dos outliers neste trabalho procedeu-se através dos gráficos de posição de plotagens, cujo objetivo, foi verificar o comportamento dos dados atípicos de cheias em relação as suas posições de plotagens, quando o tamanho da amostra aumentava, sem, no entanto, descartá-los durante o processo de ajuste da GEV.

## 5.6 MODELOS DE VAZÕES DIÁRIAS USADOS NA HIDROLOGIA OPERACIONAL

A geração de vazões diárias, através de modelos estocásticos, foi utilizada como alternativa de disponibilizar séries sintéticas de vazões diárias como base para compor séries de cheias anuais com quantidades de dados suficientemente capaz de eliminar ou minimizar as dificuldades e incertezas impostas pela restrição, em termos de número de anos e em relação aos problemas de extrapolação da curva chave para a obtenção dos valores mais elevados, de dados das séries históricas à análise de frequência preditiva. Além disso, buscou-se analisar a forma da distribuição dos valores máximos anuais, gerados a partir de séries de vazões diárias simuladas por diversos modelos estocásticos, tendo em vista, a hipótese de convergência assintótica de eventos hidrológicos naturais máximos para a forma VEIII da GEV, bem como, os

procedimentos estatísticos adotados, em cada tipo de modelo, para gerar as alturas de picos de cheias.

Neste sentido, foram escolhidos os modelos simplificado de desagregação, shot-noise de 2ª ordem, Diana e modelo de SARGENT (1979), representativos entre as classes de modelos estocásticos existentes, para gerar séries sintéticas de vazões diárias e, por conseguinte, séries de vazões máximas anuais com 1000 dados.

Os modelos baseados em simuladores de vazões mensais, como o modelo de QUIMPO (1968) e PAYNE (1969), apesar de serem bastante citados na literatura e representarem as primeiras tentativas de reproduzir, nas séries sintéticas, as hidrógrafas das séries históricas, não foram selecionados por apresentarem dificuldades na geração de vazões diárias, semelhantes às apresentadas pelos modelos ARMA(p,q) (MÜLLER, 1978).

Entre os modelos de desagregação, escolheu-se o modelo simplificado de desagregação por constituir uma proposta recente, com referência ao modelo de SVANIDZE (1980), além dos bons resultados apresentados na modelagem de vazões de rio dos Patos (Paraná). Da mesma forma, o modelo shot-noise de 2ª ordem foi selecionado para representar sua classe, por ser bastante difundido e apresentar resultados satisfatórios de modelação das vazões de rios do litoral do Paraná (MINE, 1990). Já o modelo Diana foi escolhido devido a sua sofisticação e aperfeiçoamento em relação aos modelos “estocásticos de vazões diárias” e Xadrez, além da boa adequação apresentada na modelação de vazões diárias de rios brasileiros (KELMAN et al., 1983) e constituir-se também, um modelo não-paramétrico. Por fim, selecionou-se o modelo de SARGENT (1979), que representa uma simplificação do modelo de TREIBER & PLATE (1975), por utilizar uma estrutura de modelação bastante simples, além de apresentar uma solução diferente para preservar as características mensais e anuais através da série de vazões diárias.

#### 5.6.1 Modelo Simplificado de Desagregação – MSD

O modelo simplificado de desagregação, proposto por MÜLLER, 1998), é uma extensão do modelo dos fragmentos na desagregação de vazões mensais em diárias, onde as vazões mensais podem ser as próprias vazões históricas ou vazões geradas por um outro modelo de geração de vazões mensais. A geração de séries por

este modelo é baseada no seguinte princípio básico: dadas as vazões diárias geradas de um mês  $j$ , calcula-se por tentativas, qual fragmento adimensionalizado que, multiplicado pela vazão mensal do mês  $j+1$ , resulta na menor distância entre o último dia do mês  $j$  e o primeiro dia do mês  $j+1$ . Esta geração envolve os seguintes passos:

1º) Geração da série de vazões mensais  $Y(i,j)$ , onde  $i$  representa o ano e  $j$  o mês, utilizando um modelo de vazões mensais.

2º) Preparação dos registros históricos com meses de 28, 30 e 31 dias, ou seja, das séries  $q_{28}(t,\xi,\kappa)$ ,  $q_{30}(t,\xi,\kappa)$   $q_{31}(t,\xi,\kappa)$ , dada a série  $q(t,\xi,\kappa)$ , chamada de série original.  $q_{28}(t,\xi,\kappa)$  representa as séries formadas com as vazões do dia 1 até o dia 28 de cada mês  $\xi$  ( $\xi=1\dots12$ ) do ano  $t$  ( $t$ =ano inicial até ano final da série histórica) e dia do mês  $\kappa$  ( $\kappa=1\dots28$ ).  $q_{30}(t,\xi,\kappa)$  representa as séries formadas com as vazões do dia 1 até dia 30 de cada mês  $\xi$  ( $\xi=1\dots12$ ) do ano  $t$  ( $t$ =ano inicial até ano final da série histórica) e dia do mês  $\kappa$  ( $\kappa=1\dots30$ ) e  $q_{31}(t,\xi,\kappa)$  representa as séries formadas com as vazões do dia 1 até dia 31 de cada mês  $\xi$  ( $\xi=1\dots12$ ) do ano  $t$  ( $t$ =ano inicial até ano final da série histórica) e dia do mês  $\kappa$  ( $\kappa=1\dots31$ ). Para os meses com 30 dias,  $q_{31}(t,\xi,\kappa)$  é formada com os 30 dias do mês  $\xi$  mais o primeiro dia do mês  $\xi+1$ , assim como, para os meses de fevereiro,  $q_{30}(t,\xi,\kappa)$  é formada com os 28 ou 29 dias do mês  $\xi$  mais o primeiro ou o primeiro e o segundo dia do mês  $\xi+1$  e  $q_{31}(t,\xi,\kappa)$  é formada com os 28 ou 29 dias do mês  $\xi$  mais o primeiro e segundo dia ou primeiro, segundo e terceiro dia do mês  $\xi+1$ , conforme for o caso. Isto é feito para aumentar o número de fragmentos na série histórica. Assim, cada mês do ano tem agora 3 novas séries com 28, 30 e 31 valores de vazões.

3º) obter as séries adimensionalizadas, dividindo cada vazão diária do mês  $m$  ( $m=28, 30$  e  $31$ ) pela correspondente vazão média, utilizando a seguinte expressão:

$$\tilde{q}_m(t,\xi,\kappa) = \frac{q_m(t,\xi,\kappa)}{y_m(t,\xi)}, \quad m = 28, 30 \text{ e } 31 \quad (86)$$

onde

$$y_m(t,\xi) = \frac{1}{m} \sum_{\kappa=1}^m q_m(t,\xi,\kappa), \quad m = 28, 30 \text{ e } 31 \quad (87)$$

4º) Quando forem gerados meses de janeiro, buscam-se os “meses” adimensionalizados com 31 dias; quando forem gerados fevereiro, utilizam-se “meses” adimensionalizados de 28 dias e assim por diante. O procedimento de geração consiste em multiplicar o valor da vazão mensal  $Y(i,j)$ , obtido no primeiro passo, pelas vazões adimensionalizadas, obtidas no terceiro passo, para obtenção das vazões diárias.

Para o primeiro mês a ser gerado, multiplica-se o valor  $Y(i,j)$  obtido para o ano  $i$  igual ao ano 1 da série a ser gerada e  $j = 1$ , pelas vazões de uma das séries  $\check{q}_{31}(t, \xi, \kappa)$ ,  $k = 1, \dots, 31$  com  $t$  e  $\xi$  quaisquer para a obtenção do primeiro mês com vazões diárias geradas  $Q(\text{ano}i, 1, k)$ ,  $k = 1, \dots, 31$ . O próximo mês ( $j+1$ ) gerado, que corresponde ao mês de fevereiro, é obtido respeitando-se a condição:

$$\min |Y(i, j+1)q_{28}(t, \xi, 1) - Q(i, j, 31)|, \quad (88)$$

onde  $t =$  ano 1 até o total de anos simulados e  $\xi = 1, \dots, 12$ .

esta condição foi imposta de modo que a distância entre o último dia de um mês e o primeiro dia do próximo seja a menor possível. Raciocínio semelhante é empregado na obtenção do mês de março subsequente. Neste caso, como março tem 31 dias (mês  $j+1$ ) e fevereiro tem 28 dias (mês  $j$ ), a condição em (88) passa a ser:

$$\min |Y(i, j+1)q_{31}(t, \xi, 1) - Q(i, j, 28)| \quad (89)$$

onde  $t =$  ano 1 até o total de anos simulados e  $\xi = 1, \dots, 12$ .

Aplicando o mesmo raciocínio, procede a geração dos meses seguintes, lembrando que na passagem de um ano para outro tem-se  $j=12$  e  $j+1=1$ . A diferença entre a vazão no último dia de um mês e o início do outro, feita como em (88) ou (89), ou de forma semelhante é possível, considerando que, as vazões diárias normalmente tem alta correlação lag 1; os valores de tal diferença nem sempre é pequena, assemelhando-se o caso gerado com o simulado e, por fim, o fato de não estar respeitando fidedignamente a correlação entre a última vazão de um mês e a primeira do próximo é compensada pela facilidade de programação de cálculos.

### 5.6.2 Modelo Shot-Noise de Segunda Ordem

O modelo shot-noise foi derivado sob a hipótese de representação da bacia hidrográfica por um reservatório linear. A resposta de um sistema linear é caracterizada pela função resposta de impulso. Se um sistema recebe um entrada unitária aplicada instantaneamente (impulso unitário) no tempo  $\tau$ , a resposta do sistema mais tarde no tempo  $t$  é descrita através de uma função resposta de impulso unitário  $u(t-\tau)$ . Contínuas entradas no sistema podem ser tratadas como uma soma infinitesimal de impulsos. Assim, seja  $I(\tau)$  a intensidade de precipitação (mm/h) e  $I(\tau)d\tau$  a quantidade que entra no sistema entre  $\tau$  e  $\tau+d\tau$ , logo:

$$Q(t) = \int_0^t I(\tau).u(t-\tau)d\tau$$

corresponde à integral de convolução, que é fundamental na solução de um sistema linear numa escala contínua.

Em geral, para uma entrada na bacia hidrográfica (reservatório linear)  $Y(t)$ , contínua no tempo, tem-se a vazão de saída  $X(t)$  dada pela integral de convolução, como segue:

$$X(t) = \int_0^t I(\tau).u(t-\tau)d\tau \quad (90)$$

Supondo a bacia hidrográfica representada por um único reservatório linear, tem-se:

$$X(t) = b.s(t) \quad (91)$$

onde  $X(t)$  é a vazão instantânea na saída do reservatório;  $s(t)$  é o armazenamento instantâneo e  $b$  é o coeficiente de armazenamento.

Por outro lado, através a equação da continuidade tem-se:

$$\frac{d}{dt}s(t) = Y(t) - X(t) \quad (92)$$

onde  $Y(t)$  é a vazão de entrada instantânea no reservatório linear. Assim, combinando as equações (91) e (92), chega-se a:

$$\frac{d}{dt}X(t) + bX(t) = bY(t) \quad (93)$$

Integrando a equação (93) para um pulso instantâneo (função delta Dirac)  $Y(0) = 1$ ,  $Y(t)=0$ , para qualquer  $t \neq 0$  resulta em:

$$u(t - \tau) = e^{-b(t-\tau)}$$

que deve determinar as configurações das recessões em  $X(t)$ . Portanto, a equação (90) torna-se:

$$X(t) = \int_0^t Y(\tau) e^{-b(t-\tau)} d\tau \quad (94)$$

Caso a entrada na bacia hidrográfica até o instante  $t$  ocorra como uma sucessão de  $N(t)$  pulsos instantâneos de precipitação  $Y(\tau_1)$ ,  $Y(\tau_2)$ , ...,  $Y(\tau_n)$ ,  $\tau_1 < \tau_2 < \dots < \tau_n < t$ , a integral (95) degenera-se no somatório:

$$X(t) = \sum_{m=N(-\infty)}^{N(t)} Y(\tau_m) e^{-b(t-\tau_m)} \quad (95)$$

onde  $Y(\tau_m) = y_m$  corresponde ao volume de chuva do episódio pluvial com início em  $\tau_m$ .

O processo tem três parâmetros  $\nu$ ,  $\theta$  e  $b$ . Supõe-se que a ocorrência de pulsos seja um processo de Poisson, logo  $N(t)$  é um processo de Poisson com taxa  $\nu$ .

Pesquisadores (TODOROVIC & WOOLHISER, 1975) consideram que a altura de precipitação medida num intervalo de discretização muito pequeno é uma variável aleatória com distribuição exponencial. Logo,  $Y$  segue uma distribuição exponencial com média  $\theta$  e função distribuição de probabilidade dada por:

$$f_y(y) = \frac{1}{\theta} e^{-\frac{y}{\theta}}, \quad y > 0 \quad (96)$$

O processo estocástico  $X(t)$  dado pela equação (95) segue uma distribuição gama com parâmetros  $(1/\theta, \nu/b)$  e  $X(t)$  é não negativa e positivamente assimétrica com função distribuição de probabilidade dada por:

$$f_x(x) = \frac{\left(\frac{1}{\theta}\right)^{\frac{1}{b}} \left(\frac{\nu}{b}\right)^{-1}}{\Gamma\left(\frac{\nu}{b}\right)} e^{-\frac{x}{\theta}} \quad (97)$$

Os parâmetros  $\nu$ ,  $\theta$  e  $b$  são obtidos através das fórmulas:

$$E[X(t)] = \nu\theta/b \quad (98a)$$

$$\text{Var}[X(t)] = \nu\theta^2/b \quad (98b)$$

$$\rho[X(t), X(t+s)] = \exp(-bs), \quad s > 0 \quad (98c)$$

onde  $E[.]$  é o valor esperado,  $\text{Var}[.]$  a variância e  $\rho[.]$  o coeficiente de correlação. Para estimar os valores de  $\nu$ ,  $\theta$  e  $b$  basta substituir nas equação (98a), (98b) e (98c) os momentos amostrais.

A equação (95) pode ser reescrita para  $X(t + s)$  como:

$$X(t + s) = e^{-bs} \sum_{m=N(-\infty)}^{N(t)} Y(\tau_m) e^{-b(t-\tau_m)} + \sum_{m=N(t)}^{N(t+s)} Y(\tau_m) e^{-b(t+s-\tau_m)} \quad (99)$$

Os dois termos em (99) são independentes. O primeiro representa o efeito dos eventos anteriores ao tempo  $t$  e é  $e^{-bs}X(t)$ . O segundo inclui os eventos em  $(t, t + s)$  e é o termo de inovação. Denotando a inovação por  $\varepsilon_s(t + s)$  tem-se:

$$X(t + s) = e^{-bs}X(t) + \varepsilon_s(t + s) \tag{100}$$

Desta forma, vê-se que o shot-noise é um processo auto-regressivo de 1ª ordem. Difere do AR(1) Gaussiano porque  $\varepsilon_s(t + s)$  não é um ruído branco.

Admite-se que as vazões seguem um processo shot-noise  $X(t)$  contínuo no tempo e que os dados observados são médias deste processo sobre o período de tempo  $T = 1$  dia. Os dados  $X_1, X_2, \dots$  são então definidos como:

$$X_j = \frac{1}{T} \int_{(j-1)T}^{jT} X(t) dt \tag{101}$$

Os momentos de  $X_j$ , que são diferentes daqueles de  $X(t)$  em (98a), (98b) e (98c), são dados por:

$$E(X_j) = \frac{\nu\theta}{b} \tag{102a}$$

$$Var(X_j) = \frac{\nu\theta}{b} \frac{2[b - (1 - e^{-b})]}{b^2} \tag{102b}$$

$$\rho(X_j, X_{j+s}) = \frac{(1 - e^{-b})^2}{2[b - (1 - e^{-b})]} e^{-b(s-1)} \quad s \geq 1 \tag{102c}$$

Em adição, as recessões para valores médios serão

$$u(t) = \frac{1}{b} (1 - e^{-bt}) \quad 0 < t \leq 1 \tag{103a}$$

$$u(t) = \frac{1 - e^{-b}}{b} e^{-b(t-1)} \quad t > 1 \tag{103b}$$

Como no caso discreto, basta substituir as estatísticas amostrais nas equações (102a), (102b) e (102c) para obter os estimadores dos parâmetros  $\nu$ ,  $\theta$  e  $b$ .

a) *Geração de Vazões Diárias através do Shot-Noise de Primeira Ordem*

Denotando por  $X_t$ ,  $t = 1, 2, \dots$ , o shot-noise médio a ser gerado e por  $X(t)$ ,  $t = 0, 1, 2, \dots$ , os valores discretizados, obtém-se de (99) e (103a):

$$X_{t+1} = \frac{1}{b}(1 - e^{-b})X(t) + \sum_{m=N(t)}^{N(t+1)} \frac{1}{b} [1 - e^{-b(t+1-\tau_m)}] Y(\tau_m) \quad (104a)$$

$$X(t+1) = e^{-b} X(t) + \sum_{m=N(t)}^{N(t+1)} e^{-b(t+1-\tau_m)} Y(\tau_m) \quad (104b)$$

As primeiras parcelas do lado direito das equações (104a) e (104b) são as contribuições dos eventos precedentes  $t$  e as segundas parcelas são as contribuições em  $(t, t+1)$ . A geração então se dá através dos seguintes passos:

- 1º) Assume-se um valor inicial para  $X_t$  e  $X(t)$ ;
- 2º) Os primeiros termos do lado direito de (104a) e (104b), aqueles que correspondem ao decaimento em  $t+1$  se não houver pulsos) são calculados a partir de  $X(t)$  e  $X_t$  e obtém-se um valor inicial para  $X(t+1)$  e  $X_{t+1}$ ;
- 3º) Faça  $m = 1$  e  $\tau_m = 0$ ;
- 4º) O próximo evento,  $\tau_{m+1}$  é gerado fazendo  $\tau_{m+1} = \tau_m + E$ , onde  $E$  é gerado aleatoriamente a partir de uma distribuição exponencial com média  $1/\nu$ . Neste caso,  $E = 1/\nu(-\ln(u))$ , onde  $u \sim U(1/2, 1/12)$ ;
- 5º) Se  $\tau_{m+1} > 1$ , então não haverá pulsos no dia  $t+1$  e os valores de  $X_t$  e  $X(t)$  são iguais àqueles gerados no 2º passo. Terminando a geração no dia  $t+1$  retorna-se ao 2º passo;
- 6º) Se  $\tau_{m+1} < 1$ , então haverá pulso no dia  $t+1$ . Gera-se  $Y(\tau_{m+1})$  de uma exponencial com média  $\theta$ ;
- 7º) Calcula-se a contribuição de  $Y(\tau_{m+1})$  em (104a) e (104b) através da segunda parcela das equações e adiciona-se aos valores de  $X_{t+1}$  e  $X(t+1)$  respectivamente;
- 8º) Faz-se  $m = m+1$  e repete os passos do 4º ao 8º.

b) *Geração de Vazões Diárias através do Modelo Shot-Noise de Segunda Ordem*

Os modelos shot-noise de segunda ordem se baseia na dupla amostragem, que constitui a soma de dois processos, o rápido, representando o escoamento superficial com decaimento exponencial  $b_1$  e o lento, que se refere ao escoamento básico com decaimento exponencial  $b_2$ . No processo rápido deve-se obter os parâmetros  $b_1$ ,  $v_1$  e  $\theta_1$ , da mesma forma no processo lento são determinados  $b_2$ ,  $v_2$  e  $\theta_2$ . O modo como estas duas amostragens e, conseqüentemente, seus parâmetros relacionam-se entre si separa os modelos de segunda ordem em três tipos distintos: SOSN – shot-noise de segunda ordem, onde há grande dependência entre os fluxos lento e rápido; DSON – shot-noise duplo, os dois fluxos são totalmente independentes e o MSOSN – shot-noise modificado de segunda ordem, que é um caso intermediário entre os dois anteriores.

Para este estudo foi escolhido o modelo SOSN onde os processos rápido e lento ocorrem simultaneamente. Neste caso,  $v_1 = v_2 = v_{12}$ ,  $b_1 > b_2$ ,  $\theta = \theta_1 + \theta_2$  e a função de transferência corresponde a:

$$u(t - \tau) = \frac{\theta_1}{\theta} e^{-b_1(t-\tau)} + \frac{\theta_2}{\theta} e^{-b_2(t-\tau)} \quad (105)$$

Para estimação dos parâmetros  $b_1$ ,  $v_1$ ,  $\theta_1$ ,  $b_2$ ,  $v_2$  e  $\theta_2$ , Inicialmente obtém-se os valores de  $b_1$  e  $b_2$  através do seguinte sistema de equações:

$$\rho_{T_1} \sigma_{T_1}^2 = \frac{\sigma_{11}^2 d(2,2) - \sigma_{r2}^2 d(2,1)}{d(2,2)d(1,1) - d(1,2)d(2,1)} e(1,1) + \frac{\sigma_{12}^2 d(1,1) - \sigma_{r1}^2 d(1,2)}{d(2,2)d(1,1) - d(1,2)d(2,1)} e(2,1) \quad (106a)$$

$$\rho_{T_2} \sigma_{T_2}^2 = \frac{\sigma_{11}^2 d(2,2) - \sigma_{r2}^2 d(2,1)}{d(2,2)d(1,1) - d(1,2)d(2,1)} e(1,2) + \frac{\sigma_{12}^2 d(1,1) - \sigma_{r1}^2 d(1,2)}{d(2,2)d(1,1) - d(1,2)d(2,1)} e(2,2) \quad (106b)$$

onde  $d(i,j) = d(b_i, T_j)$  e  $e(i,j) = e(b_i, T_j)$ ,  $(i,j = 1, 2)$  sendo  $T_1 = 1$  dia e  $T_2 = 1$  mês

$$d(b,T) = \{2[bT - (1 - \exp(-bT))]/(bT)^2\}$$

$$e(b,T) = [(1 - \exp(-bT))^2/(bT)^2]$$

Uma vez obtidos  $b_1$  e  $b_2$  calculam-se:

$$ss_1 = \frac{\sigma_{T1}^2 d_{22} - \sigma_{T2}^2 d_{21}}{d_{22} d_{11} - d_{12} d_{21}} \quad \text{e} \quad ss_2 = \frac{\sigma_{T2}^2 d_{11} - \sigma_{T1}^2 d_{12}}{d_{22} d_{11} - d_{12} d_{21}}$$

Para obter as taxas dos eventos e alturas média do salto tem-se

$$ss_1 = (v_1 \theta_1^2 / b_1) + [2v_{12} \theta_1 \theta_2 / (b_1 + b_2)]$$

$$ss_2 = (v_2 \theta_2^2 / b_2) + [2v_{12} \theta_1 \theta_2 / (b_1 + b_2)]$$

$$\mu = v_1 \theta_1 / b_1 + v_2 \theta_2 / b_2$$

As equação (106a) e (106b) devem ser resolvidas numericamente. O procedimento de geração de vazões diária é semelhante ao do modelo shot-noise de 1ª ordem, devendo, apenas, fazer a modificação das equações (104a) e (104b) em virtude da nova função de transferência como apresenta a equação (105).

### 5.6.3 Modelo Diana

No modelo Diana de geração de séries sintéticas diárias (KELMAN et al., 1983) a vazão no dia  $t$  da hidrógrafa de um posto fluviométrico é considerada como a soma de dois componentes:

$$Q(t) = U(t) + O(t), \quad t = 1, 2, \dots \quad (107)$$

Onde  $U(t)$  depende de fatores externos, tais como precipitações, que pela natureza intermitente pode ocasionar na hidrógrafa pulsos irregulares espaçados no tempo. Já

$O(t)$  representa o contínuo esvaziamento da água armazenada na bacia hidrográfica. Como na prática dispõe-se apenas de  $Q(t)$ , é preciso separá-lo em  $U(t)$  e  $O(t)$ .

Para tanto, o modelo Diana considera a ocorrência de  $U(t)$  positivo sempre que  $Q(t) > \lambda Q(t-1)$ , onde  $\lambda$  é uma taxa de recessão característica do posto fluviométrico e varia entre 0 e 1. Assim, tem-se:

$$U(t) = 0 \quad \text{se } Q(t) \leq \lambda Q(t-1) \quad (108a)$$

$$U(t) = Q(t) - \lambda Q(t-1) \quad \text{se } Q(t) > \lambda Q(t-1) \quad (108b)$$

Nos dias em que atuam fatores externos [ $U(t) > 0$ ], a vazão total é dada através do rearranjo da equação (108b), logo:

$$Q(t) = \lambda Q(t-1) + U(t), \quad U(t) > 0 \quad (109)$$

portanto, comparando as equações (109) e (107), segue-se que:

$$O(t) = \lambda Q(t-1), \quad U(t) > 0$$

Nos dias em que não atuam fatores externos [ $U(t) = 0$ ], a vazão total é dada simplesmente por  $O(t)$ . Considerando  $O(t)$  numa abordagem de representação da bacia através de um reservatório linear de comportamento estocástico, tem-se que, quando  $U(t) = 0$ ,  $O(t) = 0$  e, portanto,  $Q(t)$  é uma fração  $k(t) \leq \lambda$  da vazão anterior, logo:

$$Q(t) = O(t) = k(t).Q(t-1), \quad k(t) \leq \lambda \text{ e } U(t) = 0 \quad (110)$$

Combinando as equações (107), (109) e (110) o modelo Diana pode ser descrito por:

$$Q(t) = U(t) + k(t).Q(t-1) \quad (111)$$

$$U(t) = 0 \Rightarrow k(t) \leq \lambda$$

$$U(t) > 0 \Rightarrow k(t) = \lambda$$

Na modelagem do processo  $U(t)$ , dois aspectos devem ser considerados: definir sua distribuição de probabilidade marginal, considerando que existe uma probabilidade finita  $p$  de que  $U(t) = 0$   $\{P[U(t)=0] = p\}$  e, preservar a dependência entre valores sucessivos de  $U(t)$ , visto que os fatores externos que influenciam os ramos ascendentes da hidrógrafa podem resultar de fenômenos meteorológicos persistentes sobre a bacia hidrográfica.

Para resolver estes aspectos, o modelo Diana mapeia o processo  $U(t)$  através de uma transformação do tipo  $U(t) = [Y(t)]^\alpha$ . O processo  $Y(t)$  é suposto resultante de uma censura imposta a um processo auto-regressivo de ordem 1 de distribuição normal, em que todas as observações negativas foram “perdidas”. O modelo Diana adota a própria distribuição empírica,  $F_U(\cdot)$ , uma vez que, usualmente, dispõe-se de um grande número de observações de  $U(t)$ . Quanto a dependência entre  $U(t-1)$  e  $U(t)$ , o modelo assume a existência de um processo auto-regressivo normal de ordem 1 com censura, cujo coeficiente de autocorrelação  $\rho$  deve ser preservado. Este processo é mapeado em  $U(t)$  através de uma transformação não-paramétrica que deve preservar a distribuição  $F_U(\cdot)$ , como exposto a seguir:

Seja o processo markoviano  $Z(t)$  definido por:

$$Z(t) = \rho Z(t-1) + \sqrt{1-\rho^2} \varepsilon(t) \quad (112)$$

onde  $\varepsilon(t)$  é um ruído normal padrão e  $\rho$  a correlação lag 1 do processo markoviano.

Seja o processo  $Y(t)$  resultado de uma censura imposta ao processo  $Z(t)$  definida por:

$$\begin{aligned} Y(t) &= Z(t) & \text{se} & \quad Z(t) > \beta \\ Y(t) &= \beta & \text{se} & \quad Z(t) \leq \beta \end{aligned} \quad (113)$$

$$\beta = \Phi^{-1}(p)$$

onde  $\beta$  define o intervalo de censura  $(-\infty, \beta)$ ,  $\Phi(\cdot)$  é a distribuição acumulada normal e  $p = P[U(t) = 0]$ . A relação entre  $U(t)$  e  $Y(t)$  é obtida resolvendo-se  $F_U[U(t)] = \Phi[Y(t)]$ .

a) *Geração de Vazões Através do Modelo Diana*

Os passos para a geração de vazões diárias pelo modelo Diana são os seguintes:

1º) Seja a série histórica de vazões diárias  $q(\tau)$ ,  $\tau = 1, \dots, h$ , onde  $h$  é o número de dias do histórico. Necessita-se inicialmente determinar a função empírica de distribuição acumulada  $F_d[\delta(\tau')]$ , onde:

$$\delta(\tau') = q(\tau)/q(\tau-1) \quad \text{se } \delta(\tau') < 1, \quad \tau = 1, \dots, \Lambda$$

onde  $\Lambda$  é o número de vezes no histórico em que  $q(\tau)/q(\tau-1) < 1$ . Este processo serve para fazer uma estimativa inicial do valor de  $\lambda$ .

2º) Cálculo da distribuição empírica acumulada  $F_k[k(\tau'')]$ , onde:

$$k(\tau'') = q(\tau)/q(\tau-1) \quad \text{se } k(\tau'') < \lambda, \quad \tau'' = 1, \dots, m$$

onde  $m$  é o número de vezes no histórico em que  $q(\tau)/q(\tau-1) < \lambda$ .

Para calcular as funções empíricas acumuladas no 1º e 2º passo, deve-se ordenar a amostra de  $\delta$  e de  $k$ , e então, para cada valor ordenado em cada amostra, associa o correspondente valor  $i/N$ , onde  $i$  é o número de ordenação e  $N$  o tamanho da amostra.

3º) Calcula-se a distribuição empírica acumulada  $F_U[U(\tau)]$ , onde:

$$\begin{aligned} U(\tau) &= Q(\tau) - \lambda Q(\tau-1) & \text{se } Q(\tau)/Q(\tau-1) > \lambda \\ U(\tau) &= 0 & \text{se } Q(\tau)/Q(\tau-1) < \lambda \end{aligned}$$

Neste caso, para calcular a distribuição empírica acumulada, primeiro ordena-se a amostra  $U(\tau)$ , de modo que  $U(1) \leq U(2) \leq \dots \leq U(h)$ , chamando-a de  $U(j)$ . Depois, verifica o número de zeros em  $U(j)$ , denotando por  $m$ , e calcula a probabilidade de  $U(j)$  ser zero como sendo  $P[U(j) = 0] = m/h$ . Para cada Valor  $U(j) \neq 0$  corresponderá um par  $[D(j), E(j)]$  e seu valor médio  $\psi(j)$ , onde:

$$D(j) = (j-1)/h \quad j = m+1, m+2, \dots, h$$

$$E(j) = j/h \quad j = m+1, m+2, \dots, h$$

$$\psi(j) = [D(j) + E(j)]/2$$

A distribuição empírica acumulada de  $U(\tau)$  é dada finalmente por  $F_U[U(\tau)] = \psi(j)$ .

4º) Sejam as séries  $Z(\tau)$  e  $Y(\tau)$  determinadas de acordo com (112) e (113) respectivamente.

O valor de  $\rho$ , que corresponde ao coeficiente de correlação da população  $Z(\tau)$ , em (112) é calculado através da distribuição histórica de  $Y(\tau)$ , da seguinte maneira:

Seja o processo  $U(\tau)$  histórico não ordenado. A série  $\psi(\tau)$  histórica não ordenada é obtida através de  $\psi(\tau) = F_U[U(\tau)]$ , isto é, relacionando-se um valor  $\psi(\tau)$  a cada  $U(\tau)$  do histórico através da relação mostrada no 3º passo. Para cada valor  $U(\tau)$  igual a zero o valor correspondente de  $\psi(\tau)$  será  $m/h$ . A série  $Y(\tau)$  é obtida pela transformação  $Y(\tau) = \phi[\psi(\tau)]$ , sendo que  $Y(\tau) = \beta$  quando  $\psi(\tau) = m/h$ . A estimativa inicial de  $\rho$  ( $\hat{\rho}'$ ), que corresponde ao coeficiente de correlação na distribuição truncada devido a censura imposta, é obtida, utilizando-se de todos os pares contíguos [ $Y(\tau) > \beta$ ,  $Y(\tau+1) > \beta$ ], aplicando a definição de coeficiente de correlação. Finalmente o coeficiente correlação  $\rho$  é obtido a partir de  $\hat{\rho}'$  e de  $\beta$ , utilizando valores de autocorrelação de uma distribuição bivariada normal truncada fornecidos por REIGIER & HAMDAM (1971), como mostrado no Apêndice II.

5º) Inicia-se a geração das vazões diárias  $Q(t)$ , sorteando  $Q(0)$  da distribuição empírica e  $Z(0)$  da distribuição normal padrão  $\phi(\cdot)$ .

6º) Faça  $t = 1$ .

7º) Sorteia-se  $\varepsilon(t)$  da distribuição normal padrão  $\phi(\cdot)$ .

8º) Calcula  $Z(t)$  através da equação (112). Aqui o processo  $Z(t)$  sofre a censura conforme (113), tomando  $Y(t) = \max[\beta, Z(t)]$ .

9º) Obtém-se  $U(t)$ , fazendo  $\psi(t) = \phi[Y(t)]$  e tomando  $U(t) = F_U^{-1}[\psi(t)]$ , onde  $F_U$  é a distribuição empírica acumulada da variável  $U(t)$ , conforme o 3º passo.

10º) Se  $U(t) > 0$ , faça  $Q(t) = U(t) + \lambda Q(t-1)$ . Se  $U(t) = 0$ , obtenha  $k(t)$  da função empírica de distribuição acumulada de  $k(\tau)$ ,  $F_k[k(\tau)]$ , através da função  $k(\tau) = F_k^{-1}(u)$ , onde  $u$  é um número aleatório entre 0 e 1. Então faça  $Q(t) = k(t)Q(t-1)$ .

11º) Faça  $t = t + 1$  e retorne ao 7º passo.

#### 5.6.4 Modelo de Sargent

O modelo de SARGENT (1979) visa preservar características importantes dos dados históricos através da determinação: de uma seqüência de dias, utilizando uma Cadeia Markoviana, onde um aumento de vazões deve ou não ocorrer; da distribuição das alturas dos pulsos e da inclinação da curva de recessão. O método se fundamenta no modelo de HARMS & CAMPBELL (1967) como procedimento para preservar estatísticas mensais e anuais, além das diárias.

##### a) Estrutura Funcional do Modelo

Inicialmente é feita uma análise dos dados históricos, fazendo as seguintes considerações:

Seja um dia “chuvoso” aquele em que a vazão é superior àquela do dia anterior, e um dia “seco”, o caso contrário. O primeiro passo a ser considerado no estudo das características da série histórica é o cálculo das seguintes probabilidade de transição, mês a mês:

$P_{SS}$  - probabilidade de que um dia seco seja seguido por outro dia seco;

$P_{SC}$  – probabilidade de que um dia seco seja seguido por um dia chuvoso;

$P_{CS}$  – probabilidade de que um dia chuvoso seja seguido por um dia seco;

$P_{CC}$  – probabilidade de que um dia chuvoso seja seguido por outro dia chuvoso.

Estas probabilidades, mês a mês, são estimadas através das seguintes fórmulas:

$$P_{SS} = \frac{n_{SS}}{n_{SS} + n_{SC}} \quad (114a)$$

$$P_{SC} = \frac{n_{SC}}{n_{SS} + n_{SC}} \quad (114b)$$

$$P_{CS} = \frac{n_{CS}}{n_{CS} + n_{CC}} \quad (114c)$$

$$P_{CC} = \frac{n_{CC}}{n_{CC} + n_{CS}} \quad (114d)$$



onde nss, nsc, ncs e ncc são os números de dias referente a cada uma das 4 possibilidades de ocorrência. O passo seguinte é a análise da curva de recessão característica. Esta curva é dividida em duas partes: recessão “alta” (upper) e recessão “baixa” (lower).

A recessão “alta” é assumida como tendo a forma:

$$Q(t) = Q_p T^{B_1} \quad (115)$$

onde  $Q(t)$  é a vazão no dia  $t$ ,  $Q_p$  é a vazão de pulso (pulso) precedente,  $T$  o número de dias depois do pulso e  $B_1$  é um parâmetro de recessão.  $B_1$  é estimado por regressão após proceder a transformação  $\ln[Q(t)/Q_p] = B_1 \ln(T)$ .

A recessão “baixa” assume a forma

$$Q(t) = Q'_p \exp(a + bT') \quad (116)$$

onde  $Q(t)$  é a vazão no dia  $t$ ,  $Q'_p$  a vazão inicial na parte exponencial, onde inicia-se a recessão baixa,  $T'$  é o tempo contado a partir do início da recessão baixa e  $a$  e  $b$  são parâmetros da recessão baixa. Os parâmetros  $a$  e  $b$  são estimados por regressão após proceder a transformação  $\ln[Q(t)/Q'_p] = a + bT'$ .

O modelo simplificado proposto por SARGENT (1979) adota um procedimento arbitrário para separar as vazões pertencente às recessões “baixa” ou “alta”. A razão  $Q(t)/Q(t-1)$  foi calculada para pares de dias sucessivos no início de cada recessão. Quando este valor é maior do que 0,9; então estas vazões compõem a parte “baixa” da recessão; se for menor do que 0,9; pertencem a parte “alta”. Este procedimento mostrou-se realístico na inspeção de uma série de seqüências de recessões (SARGENT, 1979).

Prossegue-se com a identificação das alturas dos pulsos de vazões ou alturas dos pulsos. Neste caso, a série histórica de vazões diárias é analisada para que se obtenham os valores dos pulsos de vazões correspondente a cada mês. A altura de pulso de vazão (pulso) é calculada em cada dia em que a vazão seja maior ou igual à

vazão do dia precedente, tomando não a diferença entre os dois valores diretamente, mas, levando em consideração o efeito da recessão de um dia anterior em que não ocorreu nenhum pulso, ou seja, a altura de pulso corresponde à diferença entre o valor da vazão naquele dia que ocorreu o pulso e o valor da vazão que ocorreria naquele dia, caso não houvesse pulsos, obtido através da curva de recessão “baixa” ajustada, como está representado na Figura 3.

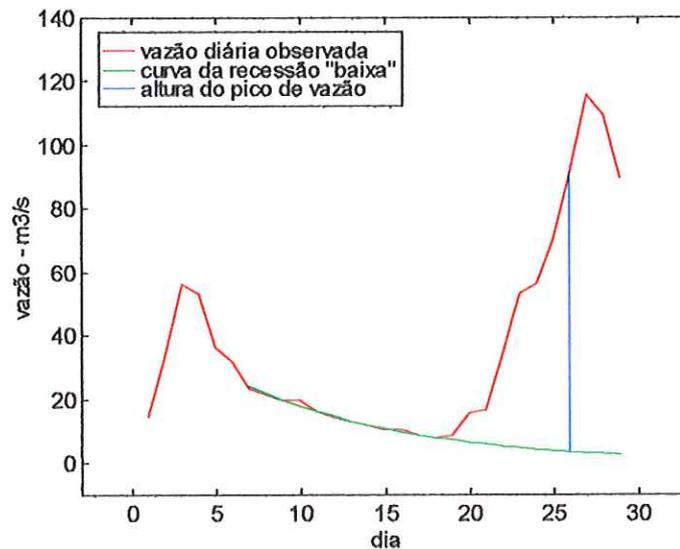


FIGURA 3 - Ilustração da determinação da altura de um pulso de vazão.

Procedendo-se desta forma, determina-se a série histórica de pulsos de vazão para cada mês do ano.

No processo de modelagem de vazões diárias, as alturas de pulsos de vazão são geradas a partir de uma distribuição de probabilidade exponencial transformada, ajustada à série histórica de pulsos, onde a altura de pulso  $z(i)$  corresponde a:

$$z(i) = \varepsilon(i)^\eta \quad (117)$$

e

$$\varepsilon(i) = (1/\delta) \cdot [-\ln(u)]$$

onde  $\varepsilon(i)$  é uma variável aleatória obtida através de uma distribuição exponencial com parâmetro  $\delta$ ;  $\eta$  é o parâmetro de transformação da exponencial e  $u$  é uma

variável aleatória com distribuição uniforme entre 0 e 1 [ $u \sim U(1/2, 1/12)$ ]. A estimação de  $\delta$  e  $\eta$  é feita através de solução numérica da seguinte equação:

$$\frac{\sigma_z}{\mu_z} = \frac{2\Gamma(2\eta)}{\eta\Gamma^2(\eta)} - 1 \quad (117a)$$

onde  $\sigma_z$  e  $\mu_z$  são o desvio padrão e a média dos pulsos de vazão e  $\Gamma(\cdot)$  a função gama. SARGENT (1979) sugeriu, por simplificação, que  $\eta$  assumisse apenas valores inteiro, isto é:

$$\Gamma(\eta) = (\eta-1)!$$

O parâmetro  $\delta$  da distribuição exponencial é dado por:

$$\frac{1}{\delta} = \frac{\Gamma(\eta+1)}{\mu_z} \quad (117b)$$

Em geral, quando ocorrem pulsos em dias sucessivos há que se levar em conta a correlação serial entre os mesmos. Isto é feito utilizando a seguinte fórmula:

$$X(i) = a(j).X(i-1-j) + z(i) \quad (118)$$

onde:

$j$  – número de dias entre pulsos sucessivos;

$X(i)$  – magnitude do pulso no dia  $i$ ;

$X(i-1-j)$  – magnitude de pulso no dia anterior a  $i$ ;

$z(i)$  – altura de pulso gerado no dia  $i$ ;

$a(i)$  – parâmetro da Cadeia de Markov, onde  $a(j) = \rho\rho$  quando  $j = 0$  (onde  $\rho\rho$  é a correlação serial lag 1 entre pulsos que ocorrem em dias sucessivos), e  $a(j) = 0$  para  $j > 0$ .

### b) Procedimento de Geração de Dados de Vazões Diárias

A seqüência de dias secos ou chuvosos é construída da seguinte forma: gera-se uma seqüência de variáveis aleatórias  $E(i)$ , uniformemente distribuída no intervalo (0,1). Os valores da variável aleatória  $E(i)$  representam as possíveis transições:

- ii) Se o dia precedente é seco [ $q(i-1) < q(i-2)$ ] e se  $E(i) > P_{ss}(j)$  o dia corrente é chuvoso;
- ii) Se o dia precedente é seco [ $q(i-1) < q(i-2)$ ] e se  $E(i) \leq P_{ss}(j)$  o dia corrente é seco;
- iii) Se o dia precedente é chuvoso [ $q(i-1) \geq q(i-2)$ ] e se  $E(i) > P_{cs}(j)$  o dia corrente é chuvoso;
- iv) Se o dia precedente é chuvoso [ $q(i-1) \geq q(i-2)$ ] e se  $E(i) \leq P_{cs}(j)$  o dia corrente é seco;

onde  $j$  representa o mês  $j$  gerado.

Durante a geração de vazões em dias “secos”, a recessão alta e a recessão baixa é identificada, utilizando-se como limiar aproximado, a média mensal de vazões (SARGENT, 1979). Assim, para valores acima do limiar próximo a vazão mensal, a recessão é do tipo alta e neste caso gera-se vazões através da equação (115). Em caso contrário, é do tipo baixa e as vazões são geradas, utilizando a equação (116). Para os dias “chuvosos” são gerados pulsos de vazão de acordo com a equação (117), os quais são acrescidos às vazões dos dias precedentes, sempre levando-se em conta a correlação serial, quando ocorrem pulsos em dias sucessivos, obtidas conforme equação (118).

### c) Esquema de Ajuste para Reproduzir Estatísticas Mensais e Anuais

O esquema de ajuste para reproduzir as estatísticas mensais e anuais é feito, utilizando um processo inverso de desagregação através do uso do modelo de HARMS & CAMPBELL (1967), onde o ajuste ocorre para que as vazões mensais resultem nas vazões anuais. Num processo de desagregação, as vazões anuais são

primeiro geradas e depois desagregadas em vazões mensais. Neste caso, procede-se ao ajuste com a geração das vazões anuais e mensais separadamente, para depois uma ser adaptada ao resultado da outra e, por fim, adaptadas às vazões diárias geradas através do modelo de SARGENT .

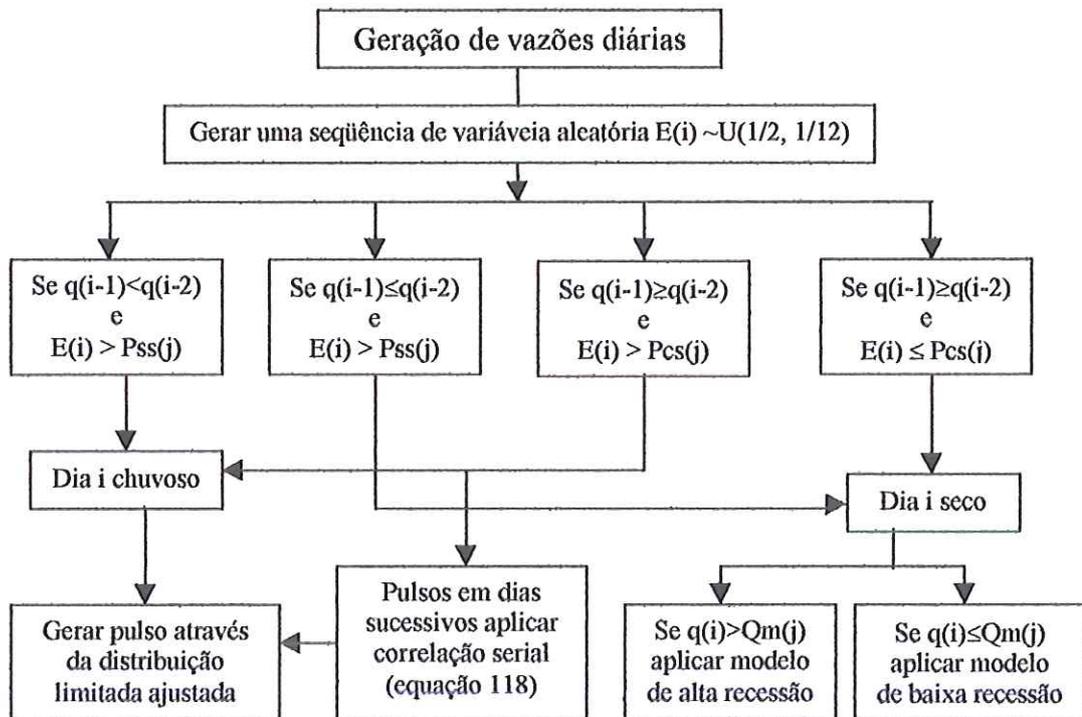
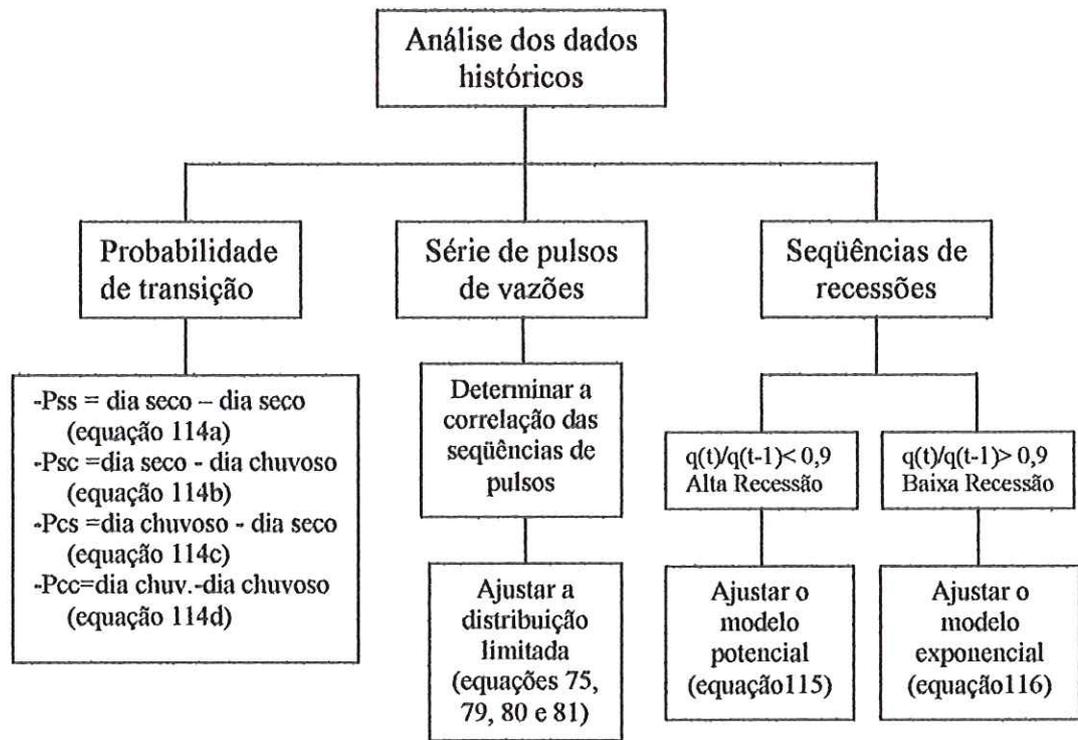
Seja, portanto,  $x(i)$  a vazão anual gerada por um modelo qualquer num ano  $i$  e  $y(i,j)^0$  a vazão original mensal gerada por outro modelo num ano  $i$  e mês  $j$ , e seja  $m(j)$  o número de dias num mês  $j$ ,  $j=1, \dots, 12$ . A vazão ajustada pelo modelo de HARMS & CAMPBELL tem a forma:

$$y(i, j) = 365 \frac{y(i, j)^0 x(i)}{\sum_{j=1}^{12} m(j) y(i, j)^0} \quad (119)$$

O ajuste aplicado por SARGENT originalmente gerava vazões anuais por um modelo auto-regressivo e vazões mensais pelo modelo de THOMAS & FIERING mensal. O ajuste através da equação (119) era então aplicado. Um procedimento semelhante é utilizado para ajustar as vazões diárias, previamente geradas, às vazões mensais.

#### 5.6.5 Modelo Proposto para Geração de Vazões Diárias

Propõe-se, neste trabalho, um modelo de geração de vazões diárias com base no modelo de SARGENT (1979), considerando a hipótese de convergência assintótica dos valores extremos gerados, para a forma VEIII da distribuição GEV. Para tanto, o modelo utiliza, como gerador de pulsos de vazão, a distribuição inicial de probabilidade limitada (equação 75) ajustada à série histórica de pulsos de vazões mês a mês. Para o ajuste da distribuição limitada foi desenvolvido um estimador de momentos de pesos probabilísticos para a referida distribuição, cujos parâmetros são estimados utilizando as equações 79, 80 e 81. Além desta implementação, o novo modelo não utiliza nenhum artifício para reproduzir as estatísticas da série histórica, nem mesmo aquele usado no modelo de SARGENT. A Figura 4 apresenta um diagrama esquemático da estrutura do novo modelo e de sua forma de geração de vazões diárias.



j representa o mês gerado;  $Q_m$  é a vazão média mensal

FIGURA 4 – Esquema representativo do novo modelo de geração de vazões diárias proposto neste trabalho.

## 5.7 ESTUDOS DA ESTIMABILIDADE DA DISTRIBUIÇÃO GEV

O presente trabalho foi desenvolvido com base no estudo de séries de vazões geradas através de três processos distintos, conforme os objetivos estabelecidos. Para tanto, inicialmente procedeu-se à geração de vazões diárias, utilizando simulação Monte Carlo com a distribuição inicial limitada como subjacente da distribuição GEV limitada (forma assintótica VEIII). As séries de cheias anuais resultantes da simulação diária foram utilizadas na análise da estimabilidade da distribuição GEV e das incertezas e erros envolvidas no processo.

Depois, a geração de séries de vazões diárias foi conduzida, empregando alguns modelos estocásticos diários existentes na literatura, buscando resolver o problema da baixa disponibilidade de cheias anuais na análise de frequência, com a utilização de metodologias já aprovadas no meio científico.

Finalmente, produziram-se séries de vazões diárias utilizando um novo método, proposto neste trabalho, para simulação de seqüências estocásticas diárias, com a perspectiva da obtenção de séries de cheias anuais com forma de distribuição, tal qual, a descrita pela GEV VEIII, e portanto, condizente com os eventos hidrológicos naturais. Além da série de cheias anuais, obtiveram-se as séries de vazões diária, mensal, anual e de extremos, todas utilizadas para avaliação do modelo.

### 5.7.1 Experimento de Simulação Monte Carlo para Geração de Vazões Diárias

Em análise de frequência de cheias, experimentos de simulação Monte Carlo são bastante utilizados, principalmente, nas análises de performance, qualidade de ajuste e sensibilidade dos modelos de distribuições de frequência usados, bem como, para avaliação de testes e métodos de estimação de parâmetros propostos. Neste sentido, tem sido prática comum, a geração de séries sintéticas de vazões máximas, empregando diretamente as distribuições de probabilidade, usualmente, utilizadas na análise de frequência de cheias. Em trabalhos sobre a distribuição GEV, ou envolvendo a mesma, como os de LANDWEHR & MATALAS (1979); HOSKING et al., (1985), WANG (1990a,b); CHOWDHURY et al., (1991); HOSKING & WALLIS (1995); WANG (1996), MAKHANDI et al., (1996); WANG (1997a,b);

WANG (1998); MARTINS & STEDINGER (2000) etc., a própria GEV tem sido utilizada para gerar séries de valores extremos, tomando  $-0,5 < k < 0,5$ ;  $\alpha = 1$  e  $u = 0$ .

A execução desta parte do trabalho foi realizada, com base em dados de vazões diárias, obtidas através de simulação numérica, onde a distribuição inicial de probabilidade limitada foi utilizada como distribuição subjacente no processo de geração das vazões diárias. A partir das séries de vazões diárias simuladas, obtiveram-se as séries de cheias anuais, além disso, a relação teórica entre a distribuição inicial e a de valores extremos é conhecida.

#### a) *Simulação de Dados de Vazões Diárias*

A simulação das vazões diárias foi feita, utilizando o gerador de vazões definido de acordo com a equação (76), cuja expressão é a seguinte:

$$Q_y = u + \frac{\alpha}{k} - \frac{\alpha}{k} [n(1 - F(Q_y))]^k, \quad Q_y \leq u + \frac{\alpha}{k} \quad (120)$$

onde  $Q_y$  é a vazão diária;  $u$ ,  $\alpha$  e  $k$  são respectivamente os parâmetros de posicionamento, escala e forma da distribuição GEV com  $k > 0$ ;  $n$  é o número de dados da amostra (tamanho da amostra) e  $F(Q_y) = p$ , onde  $p$  é uma variável aleatória com distribuição uniforme entre 0 e 1, ou seja,  $p \sim U(1/2, 1/12)$ .

Para a utilização da equação (120) como um gerador de vazões diárias, onde os valores de  $Q_y$  fossem sempre não negativos, isto é,  $0 \leq Q_y \leq u + \alpha/k$ , foi preciso determinar as condições de contorno em relação aos valores de  $\alpha$ ,  $u$  e  $p$ , para  $k > 0$  e  $n = 365$ . Assim, tomando a equação (120) igual a zero, obteve-se a razão  $R$  entre o parâmetro  $u$  e  $\alpha$  ( $R = u/\alpha$ ), para valores mínimos de  $p = F(Q_y)$ , expressa através da seguinte equação:

$$R = \frac{1}{k} \{ [n(1 - F(Q_y))]^k - 1 \}, \quad R = \frac{u}{\alpha} \quad (121)$$

Construiu-se a curva de  $R$  versus  $k$ , para valores de  $k$  entre 0 e 0,5 e diferentes faixas de valores de  $p = F(Q_y)$ , como mostra a Figura 5. Para valores de  $p < 0,01$  as curvas de  $R$  vs  $k$  são coincidentes.

Os valores de  $\alpha$  e  $u$  que rendem valores de  $R$  acima da curva com  $p < 0,01$ , satisfazem as condições de fronteiras da equação (121) para os correspondentes valores de  $k$ . Assim, a partir de valores de  $R$  e de  $k$ , definiram-se valores de  $\alpha$  e de  $u$ , para a geração de vazões diárias, através da referida equação.

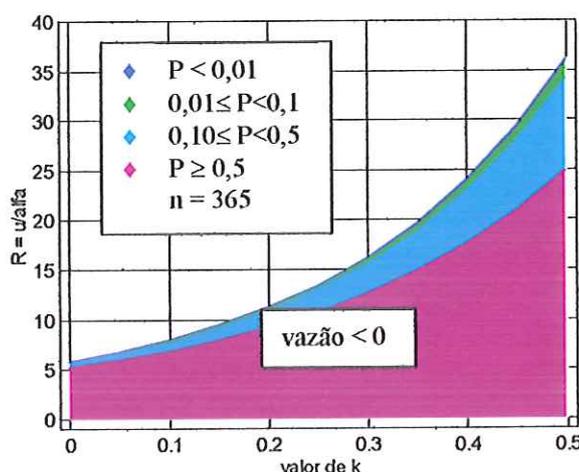


FIGURA 5 - Razão entre os valores de  $u$  e  $\alpha$  que rendem vazões não negativas através da distribuição limitada para diferentes valores de probabilidade.

#### *a<sub>1</sub>) Valores dos parâmetros $k$ , $\alpha$ e $u$ utilizados:*

Para a obtenção de séries de vazões máximas com comportamento e características distintas, utilizaram-se diferentes combinações de valores de  $k$ ,  $\alpha$  e  $u$ , para valores de  $k = 0,0001; 0,05; 0,1; 0,2; 0,3; 0,4; p < 0,01$  e  $n = 365$ . Os valores de  $\alpha$  e  $u$  foram estabelecidos, considerando os valores de  $R$ , para cada valor de  $k$  acima da curva  $R$  vs  $k$  com  $p < 0,01$ , que possibilite, sempre, gerar  $Q_y > 0$ . Os valores selecionados de  $k$  atendem aos critérios teóricos dos métodos de estimação de parâmetros utilizados, os quais utilizam  $-1/2 < k < 1/2$ , além disso os ajustes da distribuição GEV aos dados hidrológicos de vazões observados na prática, rendem valores de  $k$  que se enquadram na referida faixa (GREENWOOD et al., 1979;

HOSKING et al., 1985a; STEDINGER et al., 1993). Foram adotados valores de  $k > 0$  em virtude do estudo sobre a distribuição GEV limitada, conforme os objetivos deste trabalho.

Portanto, foram geradas séries de vazões diárias, considerando seis diferentes combinações de  $k$ ,  $\alpha$  e  $u$ , a partir dos valores de:

1º caso:  $k = 0,0001; 0,05; 0,1; 0,2; 0,3$  e  $0,4$  com  $\alpha = 120$  e  $u = 3000$ ;

2º caso:  $\alpha = 60, 100, 140, 180, 220, 260$  com  $k = 0,2$  e  $u = 3000$ ;

3º caso:  $u = 1400, 1900, 1400, 2900, 3400$  e  $3900$  com  $k = 0,2$  e  $\alpha = 120$ .

#### b) *Obtenção das Séries de Cheias Anuais*

Para cada uma das 6 combinações de parâmetros  $u$ ,  $\alpha$  e  $k$  adotadas nos três casos, foram geradas 50 séries de vazões máximas anuais contendo 1000 valores cada uma delas. O processo de geração dos dados foi realizado através de simulação Monte Carlo, que utilizou a equação (120) para gerar as vazões diárias, sendo  $F$  gerado aleatoriamente por meio da distribuição uniforme, ou seja,  $F \sim U(1/2, 1/12)$ .

Para a obtenção de cada série de vazões anuais foram geradas 1000 amostras consecutivas de vazões diárias com 365 valores e retirada, de cada uma das séries de vazões diárias seu valor máximo. Este processo foi repetido 50 vezes para cada combinação de valores  $u$ ,  $\alpha$  e  $k$  adotada, nos três casos considerados.

Em cada uma das séries de 1000 vazões anuais, foram tomados para análise de frequência, os seus 20, 35, 50, 100 e 500 primeiros valores e por fim os 1000 valores, obtendo-se, assim, 5 novas séries mais a série original de variados comprimentos. Desta forma, cada série de vazões anuais simulada foi transformada em 6 séries com 20, 35, 50, 100, 500 e 1000 anos de duração. Logo, para cada uma das 6 combinações de valores de  $u$ ,  $\alpha$  e  $k$  adotadas, em cada caso, foram obtidos 6 conjuntos de 50 séries de vazões anuais com 20, 35, 50, 100, 500 e 1000 dados. Com isto, foi possível estudar o comportamento do ajuste da GEV em função do tamanho da amostra para diferentes valores de seus parâmetros. Além disso, permitiu-se verificar como os dados atípicos (mesmo seguindo a distribuição subjacente)

presentes nas séries menores, intervém no processo de ajuste a medida que o tamanho da mesma aumenta.

c) *Ajuste da Distribuição GEV às Séries de Cheias Anuais*

Todas as séries de cheias anuais, obtidas das séries de vazões diárias geradas a partir das diversas combinações de valores  $u$ ,  $\alpha$  e  $k$ , foram submetidas ao ajuste da distribuição GEV. A modelação da distribuição GEV aos dados de cheias anuais foi realizada, utilizando-se o método dos momentos LH, com  $\eta=0$ , que é equivalente ao método dos momentos L. Desta forma, todos os dados de cada série, foram submetidos as mesmas condições de ajustes, a partir dos quais, obtiveram-se 50 conjuntos de valores de  $u$ ,  $\alpha$  e  $k$  correspondente a cada combinação de parâmetros adotada. A adequação dos ajustes realizados foi analisada, conjuntamente, através do diagrama de momentos L.

Para analisar a influência do tamanho da amostra na qualidade de ajuste da distribuição, procedem-se ao ajuste da GEV, utilizando momentos LH, a algumas séries de cheias anuais referentes as seis combinações de parâmetros adotadas no primeiro caso, com 20, 35, 50, 100, 500 e 1000 dados, para cada conjunto de  $u$ ,  $\alpha$  e  $k$  adotado. As distribuições ajustada e teórica, referente a cada combinação de parâmetros, foram comparadas graficamente e a qualidade do ajuste aferida através do teste proposto por WANG (1998).

d) *Estimação de Cheias Decamilenares*

Valores de cheias decamilenares foram estimados através da equação (33a), utilizando todos os conjuntos de valores de  $k$ ,  $\alpha$  e  $u$  estimados, obtidos com o ajuste da distribuição GEV a todas as séries de cheias anuais geradas, conforme descrito anteriormente. Além disso, calculou-se o valor da cheia decamilenar para cada combinação de valores de  $k$ ,  $\alpha$  e  $u$  adotados, nos três casos, usando a citada equação.

e) *Estimação de Séries de Quantis de Cheias Anuais*

Séries de cheias anuais com 20, 35, 50, 100, 500 e 1000 dados foram estimadas para as diversas combinações de parâmetros da GEV estimados, usando a equação (33a). Da mesma forma, foram construídas séries de vazões anuais com as mesmas extensões, cujos valores foram calculados a partir das combinações de valores de  $k$ ,  $\alpha$  e  $u$  adotados nas três situações, conforme já descrito, utilizando a mesma equação. Este procedimento possibilitou a comparação gráfica das distribuições ajustadas com as respectivas distribuições teóricas.

f) *Posição de Plotagem de Dados de cheias Anuais*

Na verificação da qualidade de ajuste da distribuição aos dados amostrais, valores de quantis de cheias anuais estimados e calculados (valores teóricos) foram plotados em gráficos semelhantes ao da Figura 2. A posição que cada dado assumiu em relação aos demais foi calculada, utilizando os modelos de posição de plotagem de: Weibull, Mediam, AP, Blom, Cunnane, Gringorten e Hazen para escolha do modelo mais adequado, levando-se em consideração a presença de valores atípicos em cada série de dados. Assim, observou-se, também, o comportamento dos outliers em relação a sua posição de plotagem quando o tamanho da amostra variou.

g) *Análise dos Resultados*

A análise comparativa entre os valores estimados e adotados de  $u$ ,  $\alpha$  e  $k$  e das cheias decamilenares estimadas e calculadas através dos parâmetros adotados, foi feita graficamente, em função da quantidade de dados utilizados. Onde são apresentados, ainda, a média e o desvio padrão em cada gráfico.

O grau de dispersão de cada variável estimada, em estudo, foi avaliado através do coeficiente de variação e do desvio máximo em relação ao seu valor teórico. Já o grau de incerteza de estimação dos parâmetros da GEV e da cheia decamilenar foi analisado, conjuntamente, através da raiz do erro quadrático médio e do erro máximo de estimação, conforme as expressões seguintes:

$$dm_e = Vm_e - V_c \quad (121)$$

$$Em_e = \frac{Vm_e - V_c}{V_c} 100 \quad (122)$$

$$reqm = \frac{\left\{ E[\hat{V} - E(\hat{V})]^2 \right\}^{\frac{1}{2}}}{V_c} 100 \quad (123)$$

onde  $dm_e$  e  $Em_e$  são, respectivamente, o desvio máximo e o erro máximo de estimação da variável em questão ( $u$ ,  $\alpha$  e  $k$  e  $Q_{10000}$ ),  $Vm_e$  valor máximo ou mínimo da variável em estudo,  $V$  valor estimado e  $V_c$  o seu valor conhecido.

### 5.7.2 Simulação de Vazões Diárias Através de Modelos Estocásticos Diários Usados na Hidrologia Operacional

A necessidade de séries de vazões máximas com extensões bem além dos registros históricos, levou ao desenvolvimento de vários modelos estocásticos de geração de vazões diárias, cuja eficácia tem sido comprovada através da modelagem de vazões observadas em algumas situações práticas.

Entretanto, sob a hipótese de convergência assintótica de valores extremos máximos de variáveis hidrológicas naturais e, considerando que a parte aleatória desses modelos estocásticos é reproduzida, utilizando a distribuição padrão normal ou a distribuição exponencial, onde as mesmas constituem exemplos de distribuições iniciais da forma assintótica VEI, procedeu-se um estudo sobre a forma assintótica das vazões máximas anuais geradas por alguns modelos estocásticos de vazões diárias.

Para tanto, foram produzidas séries sintéticas de cheias anuais com 1000 valores, a partir da geração de vazões diárias, através dos seguintes modelos: Modelo Simplificado de Desagregação, Modelo Shot-Noise de Segunda Ordem, Modelo Diana e Modelo de Sargent. No caso do modelo de desagregação, foi utilizado um modelo PAR(1) para geração de vazões mensais.

Os quatro modelos foram ajustados a duas séries de vazões diárias; uma com 66 anos de registros referente ao rio dos Patos, estação fluviométrica do rio dos Patos (código do DNAEE 64620000), na Bacia do Rio Ivaí, estado do Paraná; a outra,

referente ao Rio Paraná, com 73 anos de registros, na estação fluviométrica de Guaira (código do DNAEE 64843000), na Bacia do Rio Paraná, estado do Paraná.

As séries de cheias anuais geradas através dos quatro modelos foram submetidas ao ajuste da distribuição GEV, utilizando o método dos momentos LH para estimação dos parâmetros. A qualidade de cada ajuste foi avaliada através do teste de WANG (1998) e para definição da forma da distribuição ajustada foi aplicado o teste proposto por HOSKING et al.,(1985a).

### 5.7.3 Geração de Vazões Diárias Utilizando Novo Modelo Estocástico Diário

Desenvolveu-se um modelo de geração de vazões diárias através da modificação do modelo de SARGENT (1979), onde a geração de picos de vazões é feita através da distribuição inicial de probabilidade limitada [equação (75)] ajustada à série histórica de picos de vazões mês a mês.

O ajuste da distribuição limitada à séries de picos de vazões, para cada mês, foi feito através do método dos momentos com pesos probabilísticos, utilizando as equações de estimação dos  $\mu, \alpha$  e  $k$  de acordo com as equações (79), (80) e (81).

O modelo proposto foi ajustado à série histórica de vazões diárias do Rio Paraná, que possui 73 anos de registros, conforme descrição feita no item anterior. Para efeito de comparação, a série histórica de alturas de picos de vazões de cada mês foi ajustada, utilizando os seguintes modelos de distribuição de picos:

1º caso: distribuição exponencial;

2º caso: distribuição exponencial modificada (usada por SARGENT, 1979);

3º caso: distribuição inicial limitada (nova proposição).

A adequação de ajuste de cada modelo, aos dados de picos de vazões de cada mês, foi verificada através do teste de  $\chi^2$ .

Seguindo os procedimentos de geração de vazões diárias descrito no modelo de SARGENT, fazendo distinção na forma de geração de picos de vazões utilizado pelo modelo, foram simuladas 15 séries de vazões diárias com 73 anos de registro, para cada um dos três modelos de distribuições de picos ajustados, correspondendo, em cada caso, a uma série total com 1095 anos de registros. A transformação das 15 séries em uma única foi possível, pois os últimos valores de uma série foram

utilizados para gerar os primeiros de outra, num processo de simulação contínuo. As 15 séries foram transformada em uma só para obter-se a série de cheias anuais com 1000 valores, como nos dois casos anteriores, possibilitando analisar estes dados sob o ponto de vista da convergência assintótica da distribuição dos valores extremos.

Para avaliação do desempenho do modelo em reproduzir as estatísticas diárias, mensais, anuais e dos valores extremos de vazões da série histórica, comparado ao modelo de SARGENT e quando utiliza-se a distribuição exponencial para geração de picos de vazões, foram utilizados os seguintes procedimentos:

a) *Estimativa dos Momentos*

Foram calculados os 4 primeiros momentos das seqüências de vazões diárias, mensais, anuais e de valores máximos anuais referentes à série histórica e às 15 séries simuladas por meio dos 3 modelos de geração de picos adotados. Os resultados foram comparados graficamente e utilizados para cálculo de índice de desempenho do modelo, descrito adiante.

b) *Coefficiente Correlação*

Os coeficientes de correlação diária, mensal e anual com defasagem 1 (um) foram determinadas para a série histórica e para cada uma das 15 séries sintéticas de vazões, geradas através de cada um dos três modelos de pulsos. Para cada caso, compararam-se os coeficientes de correlação da série histórica com a média dos coeficientes de correlação das 15 séries simuladas.

c) *Índice de Comparação das Séries Geradas com a Histórica*

Dois índices de ajuste estatístico que comparam a distribuição das vazões diárias históricas com as geradas foram aplicados: o índice de igualdade de duas distribuições multinomiais que corresponde ao teste de  $\chi^2$  aplicado para comparação de duas amostras e o índice ou teste de Kolmogorov-Smirnov para duas amostras. Estes índices serviram apenas para comparação entre os modelos, tendo em vista que

não existe valores críticos para amostras correlacionadas, como ocorre com vazões diárias.

d) *Índice de Avaliação Global Desenvolvido por O'Connel (KELMAN et al., 1983)*

Este índice consiste em calcular para a série histórica e para diversas séries sintéticas com a mesma extensão da histórica, estatísticas específicas ou índices. Como resultado, tem-se uma distribuição de probabilidade para os índices onde a coerência entre as séries geradas pelo modelo e a séries histórica pode ser medida pela posição do índice histórico nesta distribuição. Um índice  $I$  qualquer é então definido como uma função de uma série de vazões,  $I = f[Q(1), Q(2), \dots, Q(h)]$ . Por exemplo,  $I = \max[Q(1), Q(2), \dots, Q(h)]$ . Aplicando-se a função à série de vazões históricas obtém-se o índice histórico  $I(\text{hist})$ . Da mesma forma, com  $n$  séries geradas com o mesmo comprimento da série histórica, obtém-se  $n$  índices gerados  $I(j)$ ,  $j = 1, 2, \dots, n$ . O desempenho do modelo é então medido pela proporção de índices gerados maiores do que o histórico,  $P[i(j) > i(\text{hist})]$ . Se esta proporção for muito pequena ou muito grande, isto é uma indicação de que as observações históricas é atípica para o modelo. Em geral, deseja-se testa  $m$  índices,  $I(i)$ ,  $i = 1, \dots, m$ , (ex: média, desvio padrão, assimetria, ...) arranjados no vetor

$$r = \begin{bmatrix} I(1) \\ I(2) \\ \vdots \\ \vdots \\ I(m) \end{bmatrix} \quad I(i) = f[Q(1), \dots, Q(h)]$$

O vetor obtido com o histórico é denotado  $r_{\text{hist}}$  e existem  $n$  vetores  $r_j$ ,  $j=1, \dots, n$ , cada um obtido com uma série gerada. Define-se, portanto, uma distribuição de probabilidade multivariada. Para cada vetor  $r_{\text{hist}}$  e para  $r_j$ , define-se a distância  $d$  ao centro da distribuição pelo escalar positivo:

$$d_j = (r_j - \mu)^T \Sigma^{-1} (r_j - \mu)$$

$$d_{\text{hist}} = (\mathbf{r}_{\text{hist}} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{r}_{\text{hist}} - \boldsymbol{\mu})$$

onde

$$\boldsymbol{\mu} = \frac{1}{n} \sum_{j=1}^n \mathbf{r}_j$$

$$\boldsymbol{\Sigma} = \frac{1}{n} \sum_{j=1}^n (\mathbf{r}_j - \boldsymbol{\mu}) \boldsymbol{\Sigma}^{-1} (\mathbf{r}_j - \boldsymbol{\mu})^T$$

(os símbolos em negrito significam notação matricial)

A variável  $d_{\text{hist}}$  é então comparada com as distâncias  $d_j$  através de duas medidas de desempenho: a proporção de distâncias geradas maiores que a distância histórica,  $P[d_j > d_{\text{hist}}]$ , e a distância padronizada (nata -  $\psi$ ) definida por:

$$\psi = \frac{d_{\text{hist}} - m}{s}$$

onde  $m$  é a média dos valores de  $d_j$ ,  $j=1, \dots, n$  e  $s$  é definido como:

$$s = \sqrt{\frac{1}{n} \sum_{j=1}^n (d_j - m)^2}$$

Quanto mais  $P[d_j > d_{\text{hist}}]$  se aproximar de 0.5 e quanto menor o valor absoluto de  $\psi$ , melhor é o desempenho do modelo.

Os índices escolhidos para serem testados na avaliação global foram: média da vazão diária, desvio padrão da vazão diária, média da vazão mensal, desvio padrão da vazão mensal, média da vazão anual, desvio padrão da vazão anual, média da vazão máxima anual, desvio padrão da vazão máxima anual e vazão máxima anual.

#### e) Análise das Vazões Máximas Anuais Simuladas

As séries sintéticas de vazões diárias referentes aos 3 modelo de distribuição de picos de vazões utilizados, renderam 15 séries de cheias anuais com 73 anos cada

uma, para cada modelo de pulso. De forma similar, os dados históricos forneceram uma série de cheia anual com a mesma duração. Estas seqüências de vazões máximas anuais foram submetidas ao ajuste da distribuição GEV através de momentos LH, sendo a qualidade do ajuste avaliada por meio do teste de WANG (1998) e a forma da distribuição verificada com a aplicação do teste de HOSKING et al. (1985a).

As 15 séries de cheias anuais, correspondentes a cada modelo de pico, foram transformada em uma única série, resultando em 3 diferentes amostras com 1095 valores máximos anuais, cada uma delas. Isto foi possível porque durante o processo de simulação, as vazões foram produzidas de forma consecutiva. Assim, cada bloco de 15 séries com 73 anos, transformou-se em uma série longa, suficiente para permitir a análise da convergência assintótica da distribuição de seus valores extremos máximos.

As referidas séries foram, também, submetidas ao ajuste da distribuição GEV, onde a qualidade de ajuste e a forma da distribuição ajustada foram avaliadas através dos teste de WANG (1998) e HOSKING et al. (1985a), respectivamente. Os três modelos de distribuição GEV ajustados foram comparados, juntamente com os dados máximos amostrais, num gráfico de posição de plotagem.

Por fim, construíram-se seis séries com os primeiros 20, 35, 50, 100, 500 e 1095 valores da série de vazões máximas com 1095 valores gerados com o uso do modelo de distribuição limitada de picos. As novas séries foram ajustadas através da distribuição GEV, usando momentos LH para análise do comportamento da forma da distribuição em função do tamanho da amostra.

## 5.8 AMBIENTE COMPUTACIONAL

Todo processo de simulação numérica via Monte Carlo e através da modelagem estocástica realizado neste trabalho, estimação de parâmetros da distribuição GEV, bem como, toda apresentação gráfica dos resultados obtidos, foram desenvolvidos num ambiente MATLAB, versão 5.2, da Math Works. Para tanto, desenvolveram-se rotinas específicas para cada caso. No Apêndice III são apresentadas rotinas desenvolvidas para estimação de parâmetros da distribuição

GEV, utilizando: momentos com pesos probabilístico, momentos L e momentos LH, para qualquer tamanho de amostra.

As referências bibliográficas e citações, bem como a estrutura geral do presente trabalho seguem as Diretrizes Para Elaboração de Dissertações e Teses na EESC-USP – São Carlos, 1996, 2ª Edição Revista e Ampliada.

## 6 RESULTADOS E DISCUSSÕES

A apresentação e discussão dos resultados segue a mesma estrutura empregada nos capítulos anteriores. Primeiro, discorre-se sobre os resultados provenientes da modelação de vazões diárias, utilizando a distribuição de probabilidade limitada inicial como subjacente e a análise de frequência de seus valores extremos empregados no estudo da estimabilidade da distribuição GEV. Em seguida, prossegue-se com a apresentação e discussão dos resultados referentes às análises sobre as distribuições das séries de cheias anuais com longos períodos, obtidas a partir da modelação estocástica com o uso de modelos de vazões diárias existentes na literatura, bem como através de novo procedimento de geração de vazões proposto e discutido neste trabalho.

### 6.1 ANÁLISE DA ESTIMABILIDADE DA DISTRIBUIÇÃO GEV

A utilização da distribuição inicial limitada como subjacente da distribuição de valores extremos, possibilitou a geração de séries de vazões de forma mais realística, onde o processo natural de reprodução das séries de máximos anuais foi respeitado. Além disso, a convergência assintótica dos seus valores extremos máximos para a forma VEIII, que corresponde a GEV com  $k > 0$ , permitiu conhecer a distribuição GEV teórica a partir dos valores de  $k$ ,  $\alpha$  e  $u$  utilizados na geração de vazões diárias. Assim, foi possível proceder à análise comparativa das estimativas dos parâmetros  $k$ ,  $\alpha$  e  $u$  e dos quantis de cheias com seus verdadeiros valores, para cada uma das diferentes combinações de valores de  $k$ ,  $\alpha$  e  $u$  e de tamanho de amostra adotados. Bem como, verificar a partir de que tamanho de amostra ocorre a convergência à verdadeira forma assintótica.

Neste cenário a adoção de valores de  $k$  entre 0 e 0,5 é justificada, considerando que o interesse da pesquisa sobre a hipótese de convergência assintótica da variáveis hidrológicas máximas para a forma VEIII, pressupõe  $k > 0$ ; A faixa de valores de  $k$  adotada, corresponde à parte não negativa da faixa linear de  $k$ , sobre a qual, HOSKING et al., (1985a) e HOSKING (1990) definiram os estimadores de  $k$ ,  $\alpha$  e  $u$  quando propuseram os métodos de momentos de pesos probabilísticos e momentos  $L$ , respectivamente; Os valores de  $k > 0$  resultantes do ajuste da GEV às séries de vazões máximas observadas na prática, enquadram-se na referida faixa de valores de  $k$ .

Os valores de  $\alpha$  e  $u$  foram estabelecidos de acordo com os limites de contorno do gerador de vazões diárias (equação 121) definidos através da Figura 5 com os diferentes valores de  $k$  adotados. Com este procedimento, garantiu-se a geração de vazões diárias não negativas, respeitando, portanto, os limites físicos de ocorrência dos processos hidrológicos, já que a distribuição subjacente utilizada, fornece valores extremos limitados.

#### 6.1.1 Estimação dos Parâmetros $k$ , $\alpha$ e $u$

As séries de vazões anuais, obtidas das simulações de vazões diárias a partir de cada combinação de valores de  $k$ ,  $\alpha$  e  $u$ , foram submetidas ao ajuste da distribuição GEV, onde a estimação de seus parâmetros foi realizada através do método de momentos  $L$ .

Os resultados da estimação dos parâmetros da GEV estão apresentados de forma gráfica, em função dos seis diferentes valores de tamanhos das séries de cheias anuais ( $N$ ), considerando os diferentes valores reais de cada um dos 3 parâmetros, adotados no processo de simulação de vazões diárias.

##### a) *Parâmetro de Forma $k$*

Os resultados da estimação de  $k$  sobre as 50 séries sintéticas com diferentes tamanhos, referentes aos seis valores reais de  $k$ ,  $\alpha$  e  $u$ , encontram-se, respectivamente, nas Figuras 6, 7 e 8. A dispersão de sua estimação, decorrente da

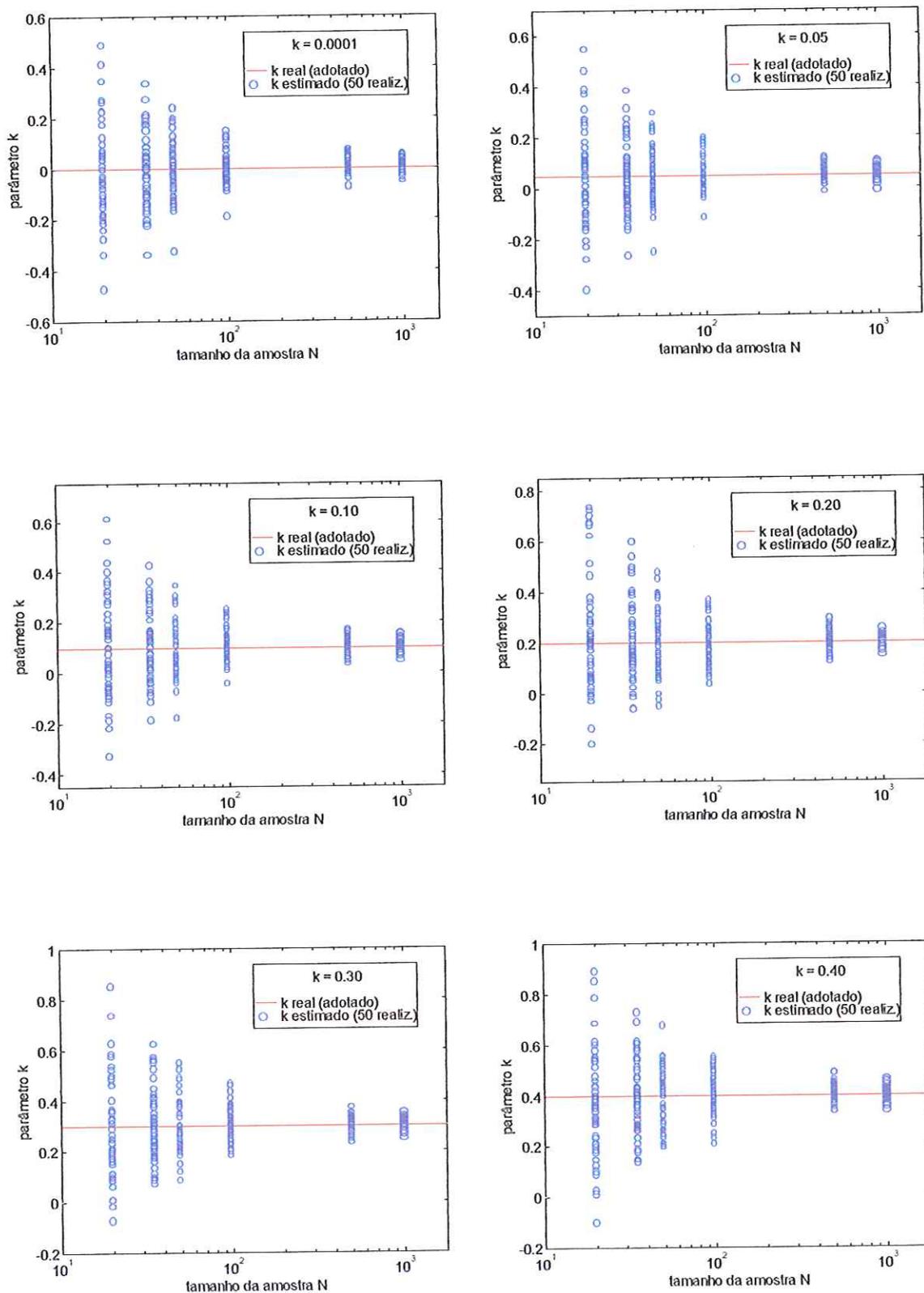


FIGURA 6 - Efeito do valor real de  $k$  sobre sua estimação, em função do tamanho da amostra, mantido  $\alpha = 120$  e  $u = 3000$ .

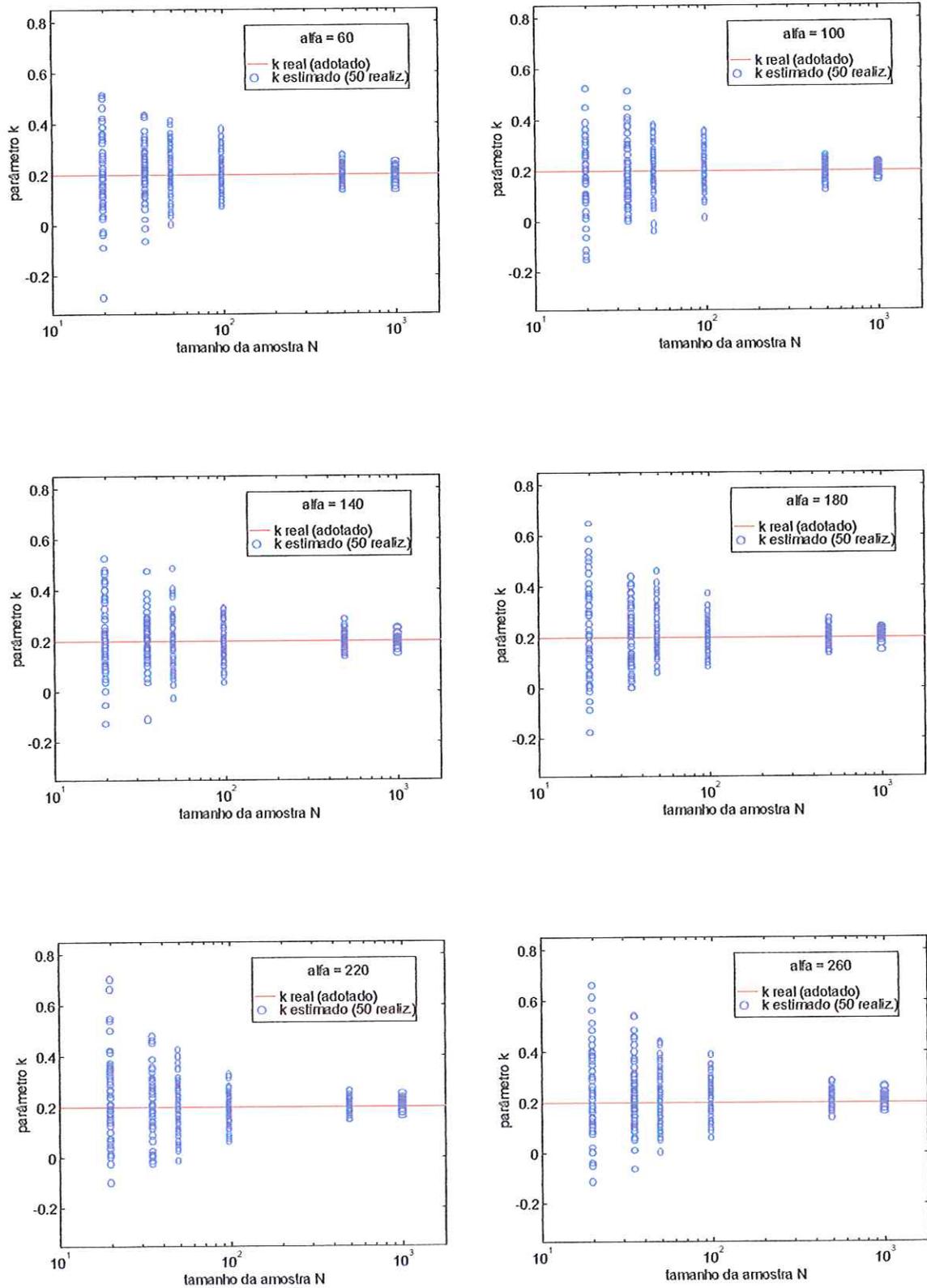


FIGURA 7 - Efeito do valor real de  $\alpha$  sobre a estimação de  $k$ , em função do tamanho da amostra, mantido  $k = 0,20$  e  $u = 3000$ .

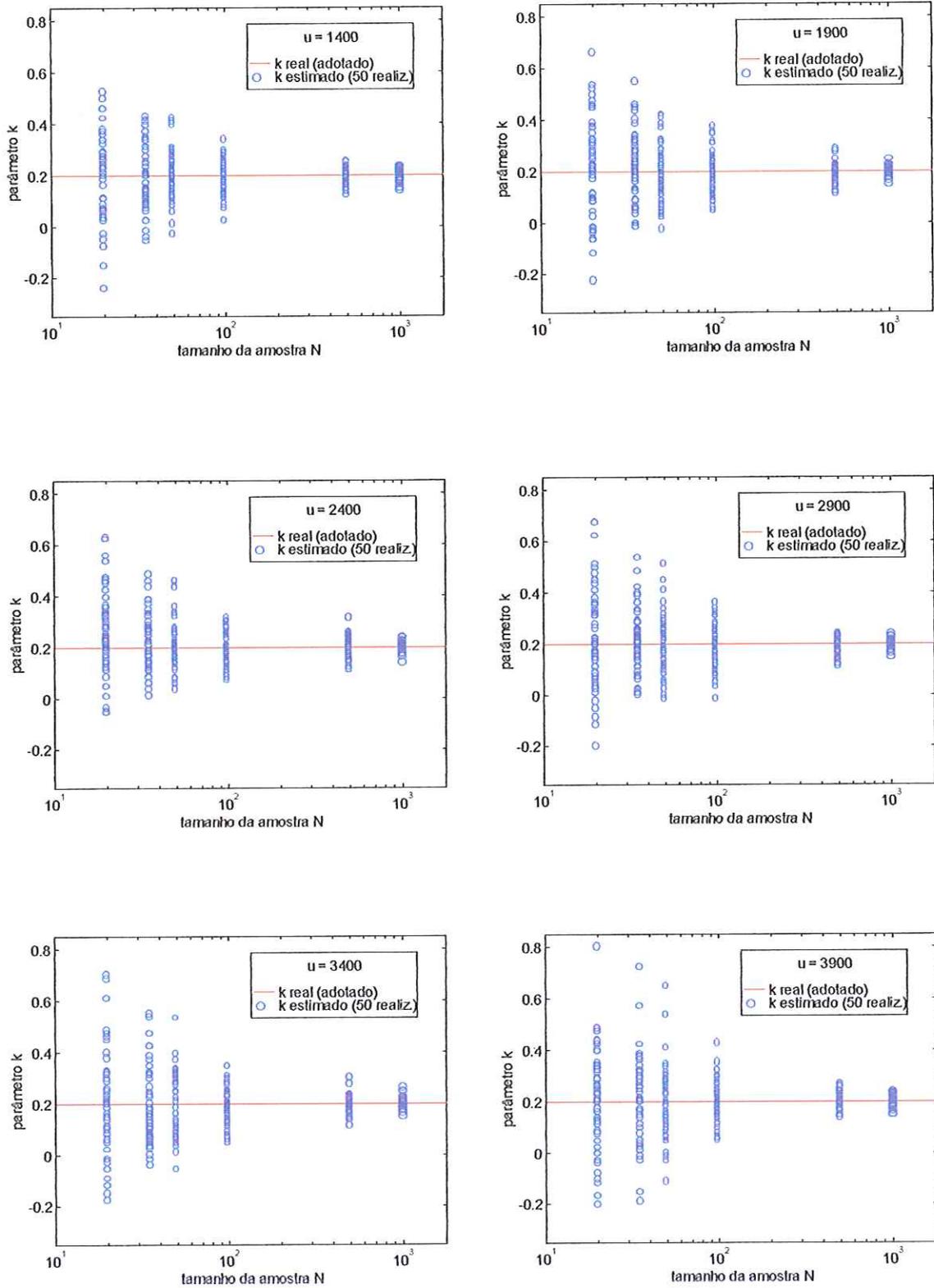


FIGURA 8 - Efeito do valor real de  $u$  sobre a estimação de  $k$ , em função do tamanho da amostra, mantido  $\alpha = 120$  e  $k = 0,20$ .

variação de cada parâmetro, determinada através dos desvios máximos e dos coeficientes de variação das suas estimativas, estão apresentados na Figura 9, em função dos diversos comprimentos de amostras.

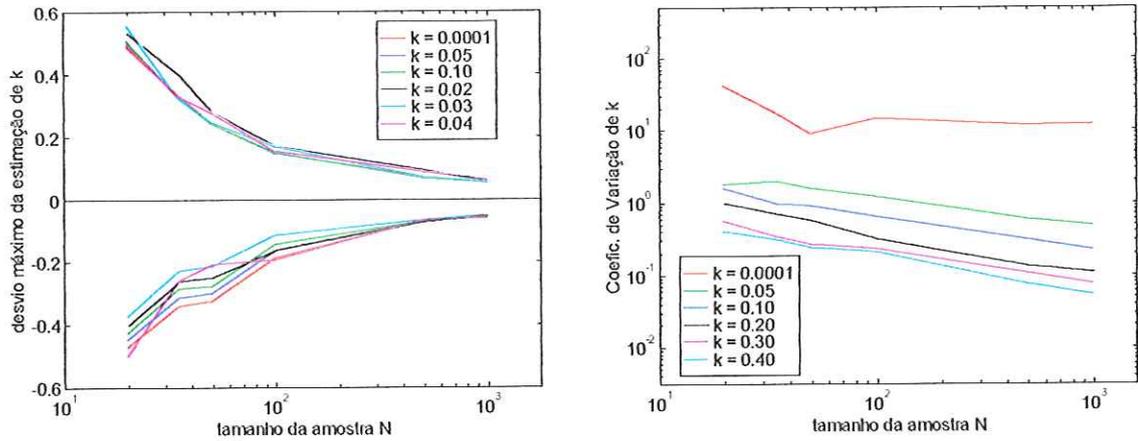
As Figuras 6, 7 e 8 mostram que as estimativas de  $k$  distribuem-se de forma aproximadamente simétrica em torno dos seus valores reais, que correspondem aos valores esperados com a estimação. Onde a variabilidade dos dados estimados aumenta de forma significativa à medida que diminui-se o tamanho da amostra, mantendo-se o mesmo padrão de dispersão, para os diversos valores reais de  $k$ ,  $\alpha$  e  $u$ .

Note que há ocorrência de valores nulos e negativos de  $k$ , além dos positivos, para todos os tamanhos de amostras se  $k_{\text{real}} \leq 0,05$ , para os seis valores reais de  $k$  quando o tamanho da amostra é 20 e para  $N < 100$ , quando  $k_{\text{real}} \leq 0,2$ . Para  $k_{\text{real}} > 0,2$  e  $N > 20$  todas as estimativas do parâmetro de forma são positivas. Isto significa que o ajuste da distribuição GEV sobre amostras com menos de 100 dados de cheias pode resultar em qualquer uma das três formas de distribuição de valores extremos, principalmente, quando o verdadeiro valor de  $k$  encontra-se entre 0 e 0,2.

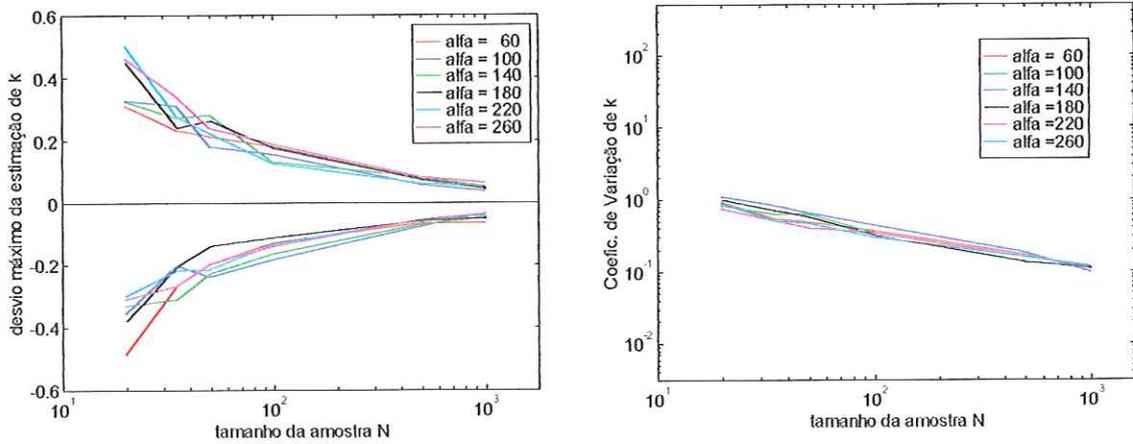
O efeito da variação de  $k$ ,  $\alpha$  ou  $u$  sobre os desvios máximos da estimação de  $k$  não ocorre claramente (Figura 9). A pouca distinção existente nos 3 casos, especificamente, para  $N < 100$ , não segue nenhum padrão passível de ser atribuído ao efeito de um ou outro parâmetro. Neste caso, tais diferenças parecem estar mais associadas à influência do tamanho da amostra.

A variação de  $k$  real afeta a variabilidade de suas estimativas, como mostram os coeficientes de variação de  $k$ , apresentados na Figura 9a. A dispersão dos dados estimados de  $k$ , cresce acentuadamente, quando  $k$  se aproxima de zero. Embora, o coeficiente de variação para  $k=0,0001$  tenha pouco efeito prático, devido o mesmo ser praticamente zero, as curvas referentes aos demais valores reais de  $k$  exibem sua influência sobre a variabilidade dos dados estimados. Tais efeitos não acontecem quando os valores reais de  $\alpha$  e de  $u$  variam, como ilustram os respectivos Gráficos de coeficiente de variação da referida figura. Nestes dois casos, os dados estimados do parâmetro de forma, cujo valor esperado com a estimação é de 0,2, apresentam a mesma consistência.

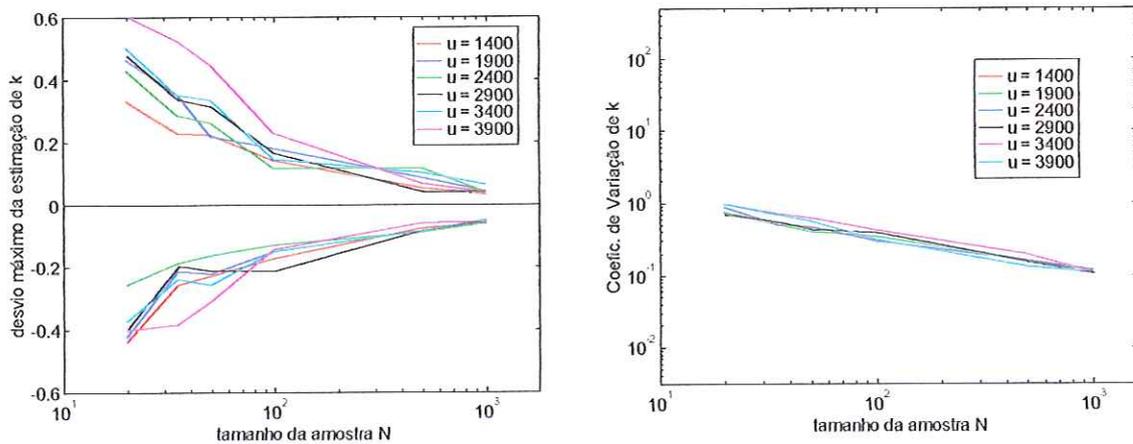
Observa-se, por exemplo, que para  $k_{\text{real}} = 0,2$ , os coeficientes de variação das estimativas de  $k$  ocorrem na mesma faixa de variação para os diferentes valores reais



a) estimação de  $k$ , usando vazões simuladas com diferentes valores de  $k$ ,  $\alpha=120$  e  $u=3000$



b) estimação de  $k$ , usando vazões simuladas com diferentes valores de  $\alpha$ ,  $k=0,2$  e  $u=3000$



c) estimação de  $k$ , usando vazões simuladas com diferentes valores de  $u$ ,  $k=0,2$  e  $\alpha=120$

FIGURA 9 - Dispersão da estimação de  $k$  verificada através do desvio máximo e do coeficiente de variação, em função do tamanho da amostra.

de  $\alpha$  e  $u$ , ou seja, entre 1 e 0,1 (100 e 10 %) para  $N$  entre 20 e 1000 respectivamente. Esses valores passam para a faixa de 300 a 40 % para  $k = 0,05$  e para o intervalo da ordem de 3000 % quando  $k \rightarrow 0$ .

Apesar dos altos desvios de estimação do parâmetro de forma  $k$ , relacionados aos tamanhos reduzidos de amostras e aos valores reais de  $k$  próximo de zero, um resultado extremamente importante é o sinal do parâmetro, o qual define a forma da distribuição dos valores extremos. Como descrito anteriormente, para  $N = 20$  e valor qualquer de  $k$  real ou para  $N < 100$  e  $k \leq 0,2$ , a modelação de cheias anuais pode conduzir a valores diversos de  $k$  e, conseqüentemente, tipos quaisquer de distribuição de extremos. Resultados de valores quaisquer de  $k$  com a estimação de parâmetros sobre séries de cheias relacionadas aos diferentes valores reais de  $\alpha$  ou  $u$  são devidos ao efeito de  $N$  que impõe comportamento similar.

#### *b) Parâmetro de Escala $\alpha$*

As estimativas de  $\alpha$ , referentes aos seis valores reais de  $k$ ,  $\alpha$  e de  $u$ , estão exibidas nas Figura 10, 11 e 12, respectivamente, em função dos diferentes tamanhos de amostras. A dispersão da estimação de  $\alpha$  verificada através dos seus desvios máximos e dos coeficientes de variação, decorrente dos diversos valores reais dos três parâmetros, encontra-se na Figura 13, plotada em função dos valores de  $N$ .

A Figura 10, percebe-se, nitidamente, o efeito da variação de  $\alpha$  sobre sua estimação, além da interferência do tamanho da amostra. O aumento da variabilidade dos dados estimados do parâmetro de escala ocorre diretamente com o acréscimo da magnitude de  $\alpha$  real. Já o tamanho da amostra intervém de maneira similar ao efeito exercido sobre o parâmetro de forma, independentemente da variação de cada parâmetro adotado.

Os diferentes valores reais de  $k$  e  $u$  não afetam as estimativas de  $\alpha$ , como pode-se constatar através das Figuras 11 e 12, onde os resultados apresentam o mesmo padrão, cuja variabilidade decorre do efeito do comprimento da amostra.

Os desvios máximos da estimação de  $\alpha$  ilustrados na Figura 13, corroboram o que foi descrito antes, onde o Gráfico *a* da citada figura, evidencia claramente, o efeito dos diversos valores de  $\alpha$  e  $N$ , enquanto nos Gráficos *b* e *c* da referida variável,

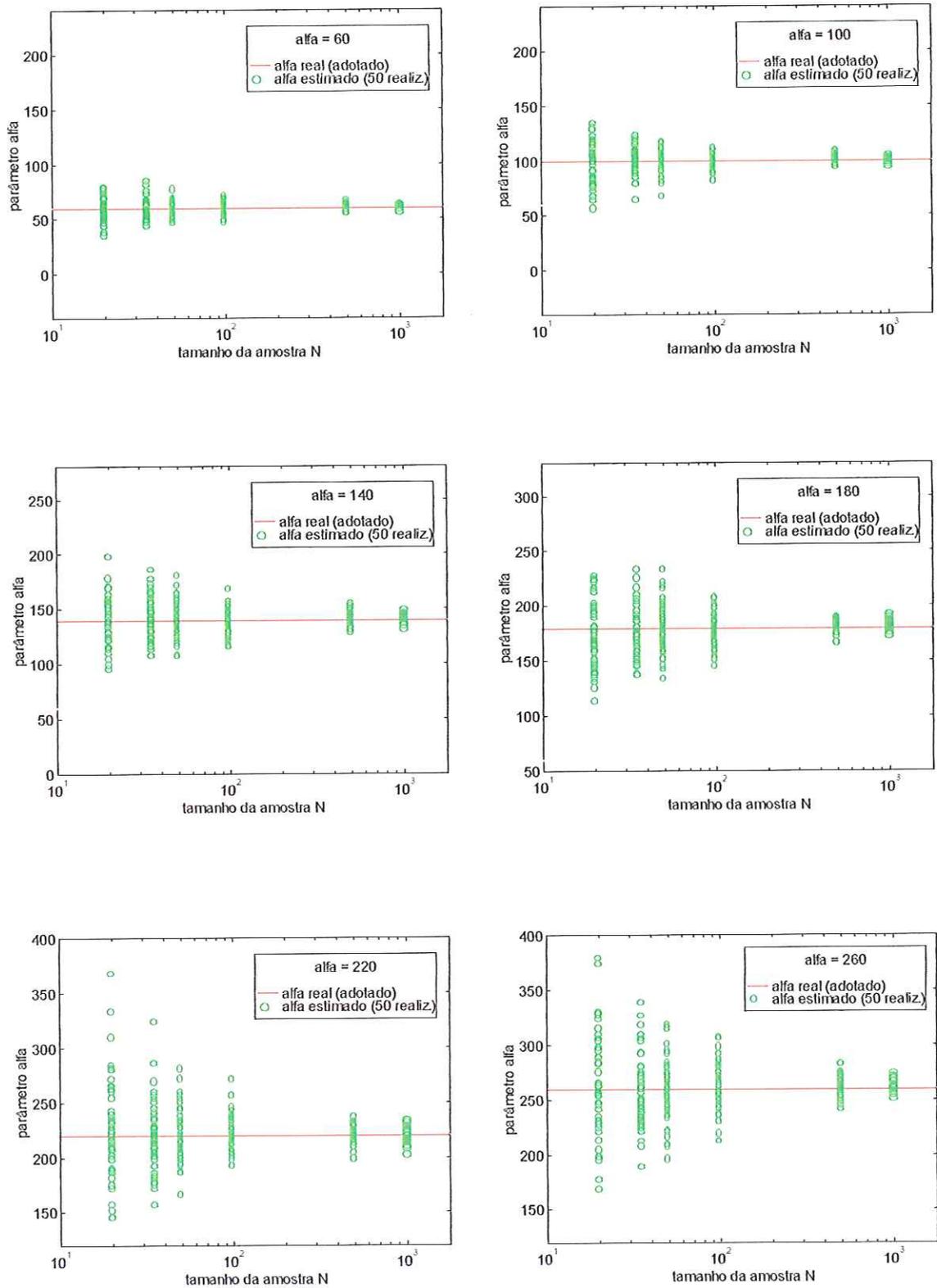


FIGURA 10 - Efeito do valor real de  $\alpha$  sobre sua estimação, em função do tamanho da amostra, mantido  $k = 0,20$  e  $u = 3000$ .

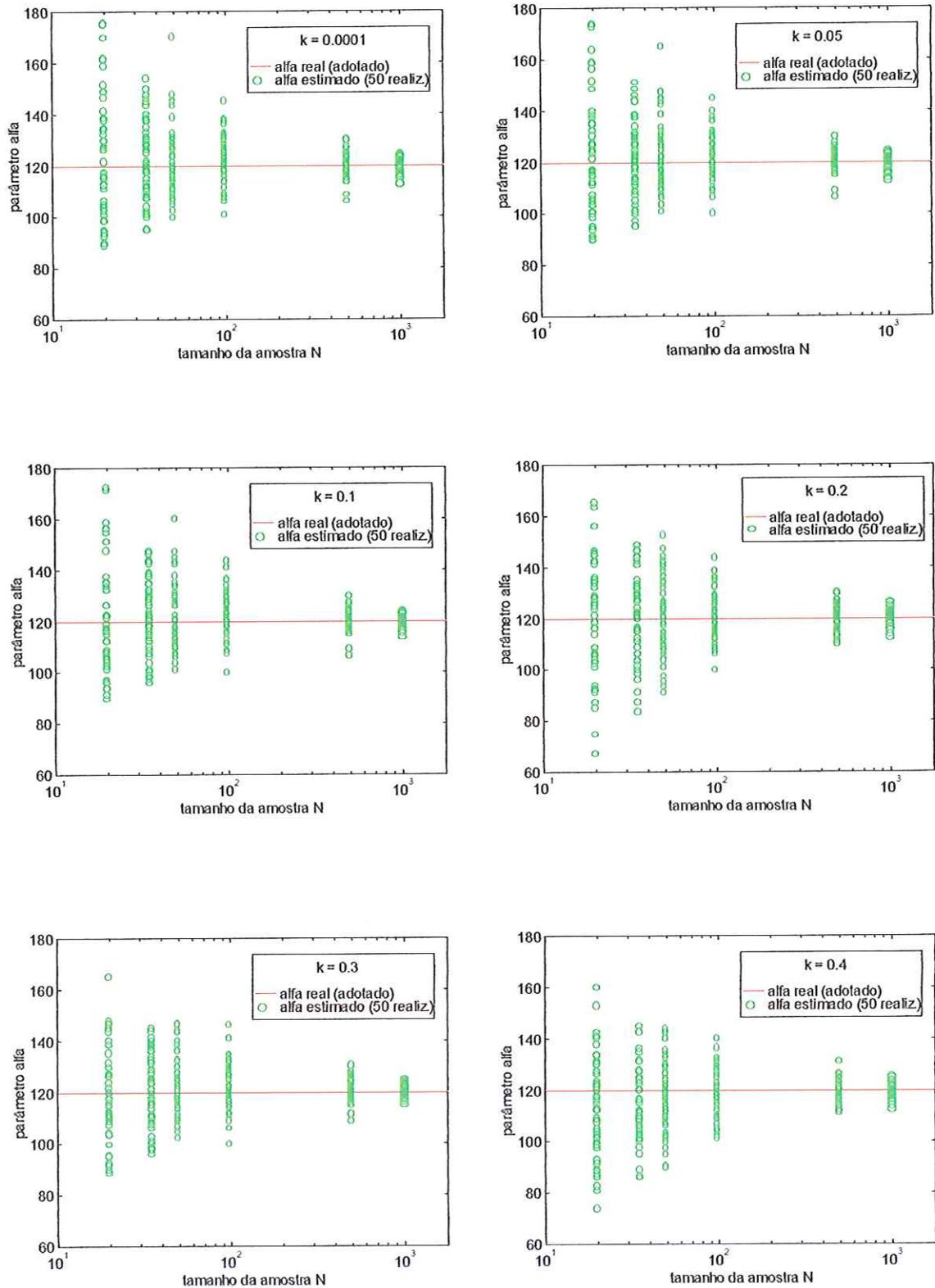


FIGURA 11 - Efeito do valor real de k sobre a estimação de  $\alpha$ , em função do tamanho da amostra, mantido  $\alpha = 120$  e  $u = 3000$ .

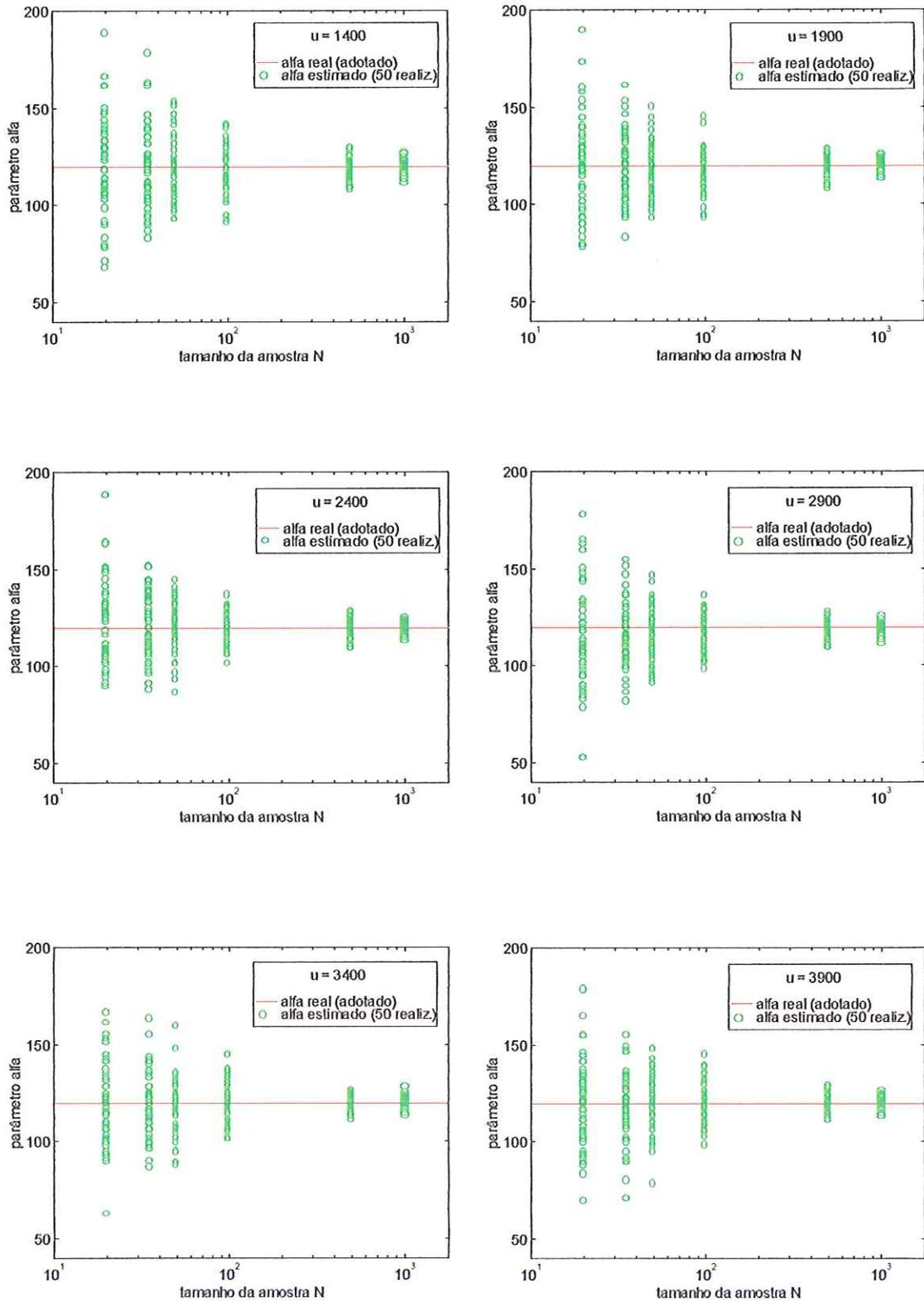
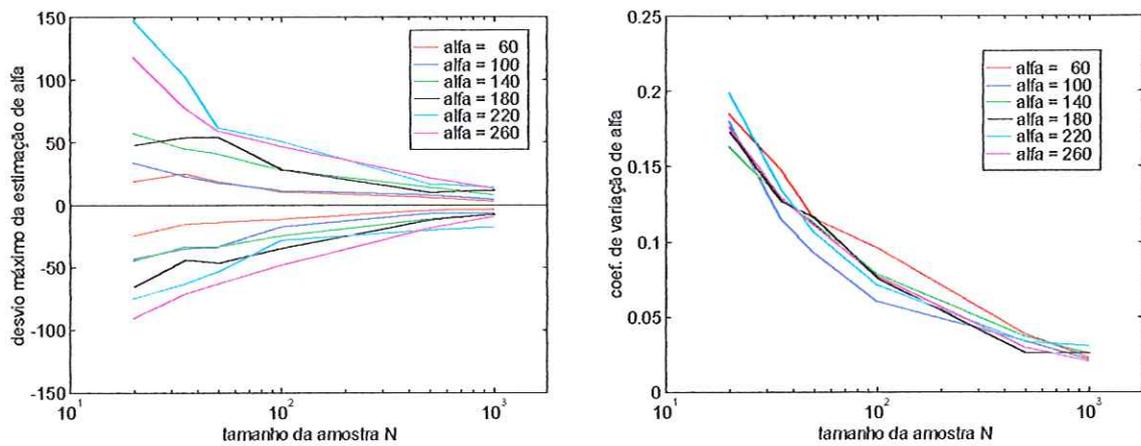
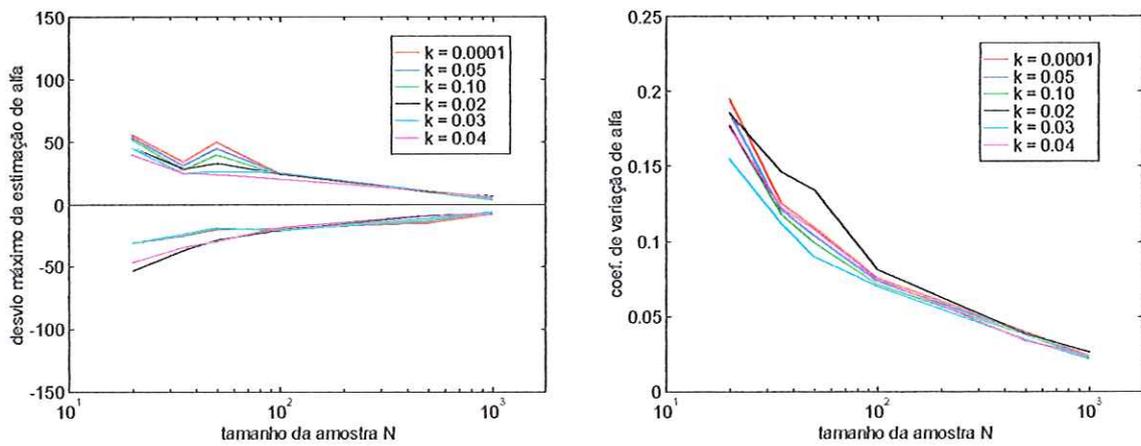


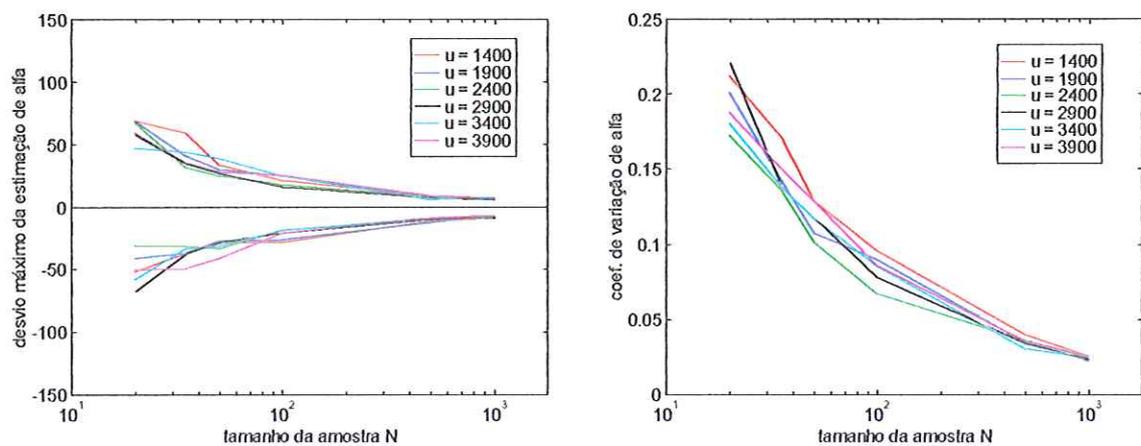
FIGURA 12 - Efeito do valor real de  $u$  sobre a estimação de  $\alpha$ , em função do tamanho da amostra, mantido  $k = 0,20$  e  $\alpha = 120$ .



a) estimação de  $\alpha$ , usando vazões simuladas com diferentes valores de  $\alpha$ ,  $k=0,20$  e  $u=3000$



b) estimação de  $\alpha$ , usando vazões simuladas com diferentes valores de  $k$ ,  $\alpha=120$  e  $u=3000$



c) estimação de  $\alpha$ , usando vazões simuladas com diferentes valores de  $u$ ,  $k=0,2$  e  $\alpha=120$

FIGURA 13 - Dispersão da estimação de  $\alpha$  verificada através do desvio máximo e do coeficiente de variação, em função do tamanho da amostra.

identifica-se, apenas, o efeito de  $N$ . Não obstante, a interferência dos valores reais de  $k$ ,  $\alpha$  e  $u$  sobre a variabilidade dos dados estimados de  $\alpha$  não é perceptível, como ocorre com o tamanho da amostra, em que os coeficientes de variação, apresentados nos Gráfico *a*, *b* e *c* da Figura 13, seguem o mesmo padrão. Nos três casos, o coeficiente de variação encontra-se na faixa de 0,0025 a 0,2 (0,25 a 20 %) para  $N$  igual a 1000 e a 20 respectivamente, revelando uma consistência das estimativas relativamente boa.

### c) *Parâmetro de Posicionamento $u$*

Os resultados da estimação de  $u$  são apresentados nas Figuras 14, 15 e 16, em função dos valores de  $N$ , referentes aos seis valores reais de  $u$ ,  $k$  e  $\alpha$  respectivamente. Na Figura 17, apresentam-se as estimativas de dispersão dos valores estimados de  $u$ , feita através dos desvios máximos e dos coeficientes de variação, considerando a influência da variação dos três parâmetros.

Os dados das Figuras 14, 15 e 16 mostram que o efeito do tamanho da amostra sobre a estimação do parâmetro  $u$  segue o mesmo padrão de interferência apresentado com as estimativas de  $k$  e  $\alpha$ . Ademais, apenas a variação do parâmetro de escala, afeta de forma direta a magnitude das estimativas de  $u$ , cujo efeito ocorre de proporcionalmente ao valor de  $\alpha$ .

Os desvios máximos da estimação de  $u$ , apresentam-se de forma similar, para os diferentes valores reais de  $k$  e  $u$ . Porém, são visivelmente distinguível com os diversos tamanhos de amostras, como pode-se notar através dos Gráfico *a* e *b* da Figura 17. O Gráfico *c* da mesma figura referente aos desvios máximos, mostra com nitidez o efeito direto do parâmetro  $\alpha$  sobre a estimação de  $u$ , associado ao efeito do valor de  $N$ .

Todavia, a variabilidade dos dados de  $u$  estimado, é afetada com a variação do tamanho da amostra e dos parâmetros  $\alpha$  e  $u$  reais, como evidenciam os Gráficos dos coeficientes de variação *a* e *c* da Figura 17. Este efeito não é verificado com a variação de  $k$  real, conforme o Gráfico *b* da mesma figura.

Percebe-se, no entanto, que a variabilidade de  $u$ , decorrente da variação do seu valor real, acontece de forma inversa ao do efeito exercido pelos diversos valores

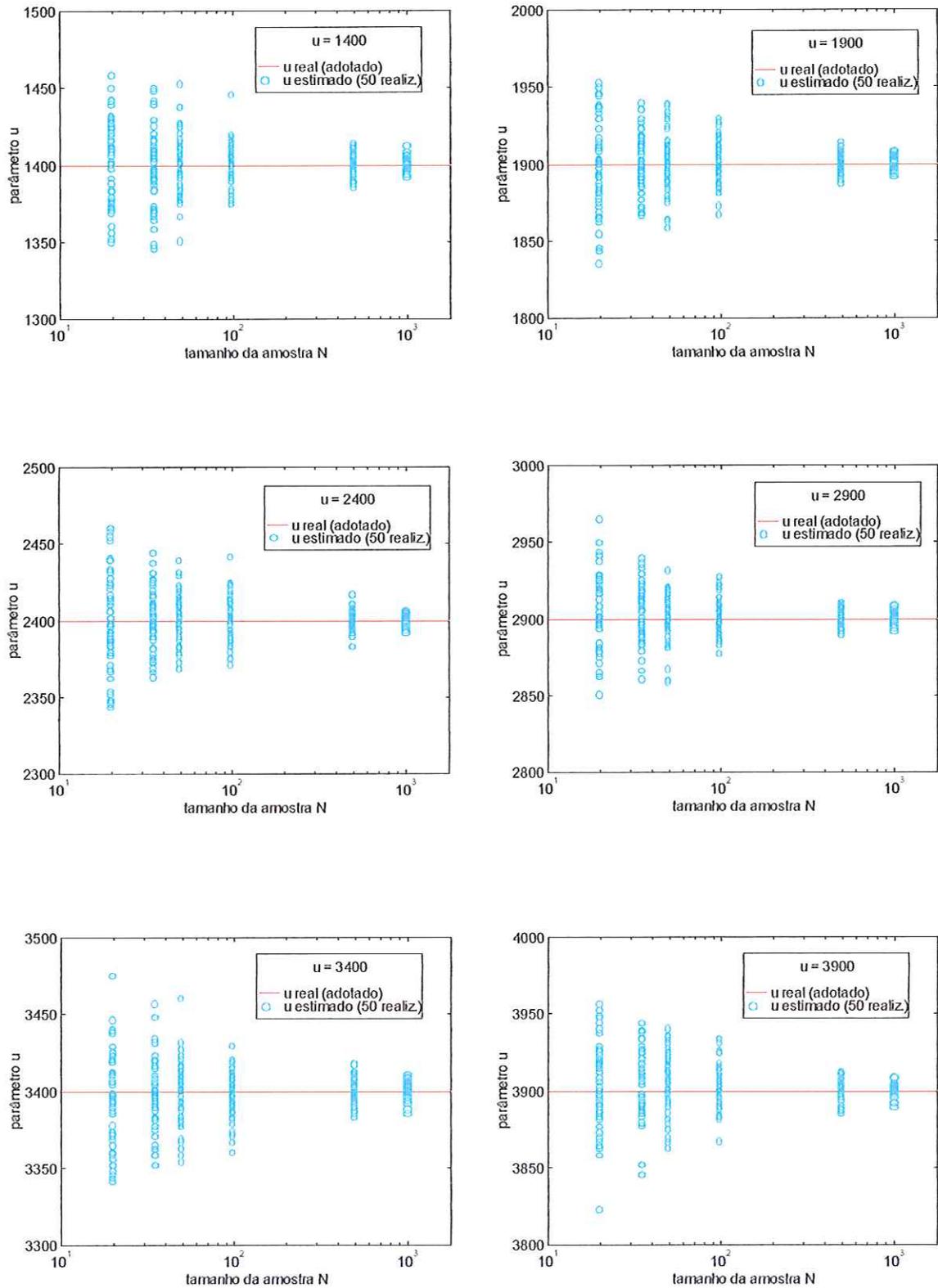


FIGURA 14 - Efeito do valor real de  $u$  sobre sua estimação, em função do tamanho da amostra, mantido  $k = 0,20$  e  $\alpha = 120$ .

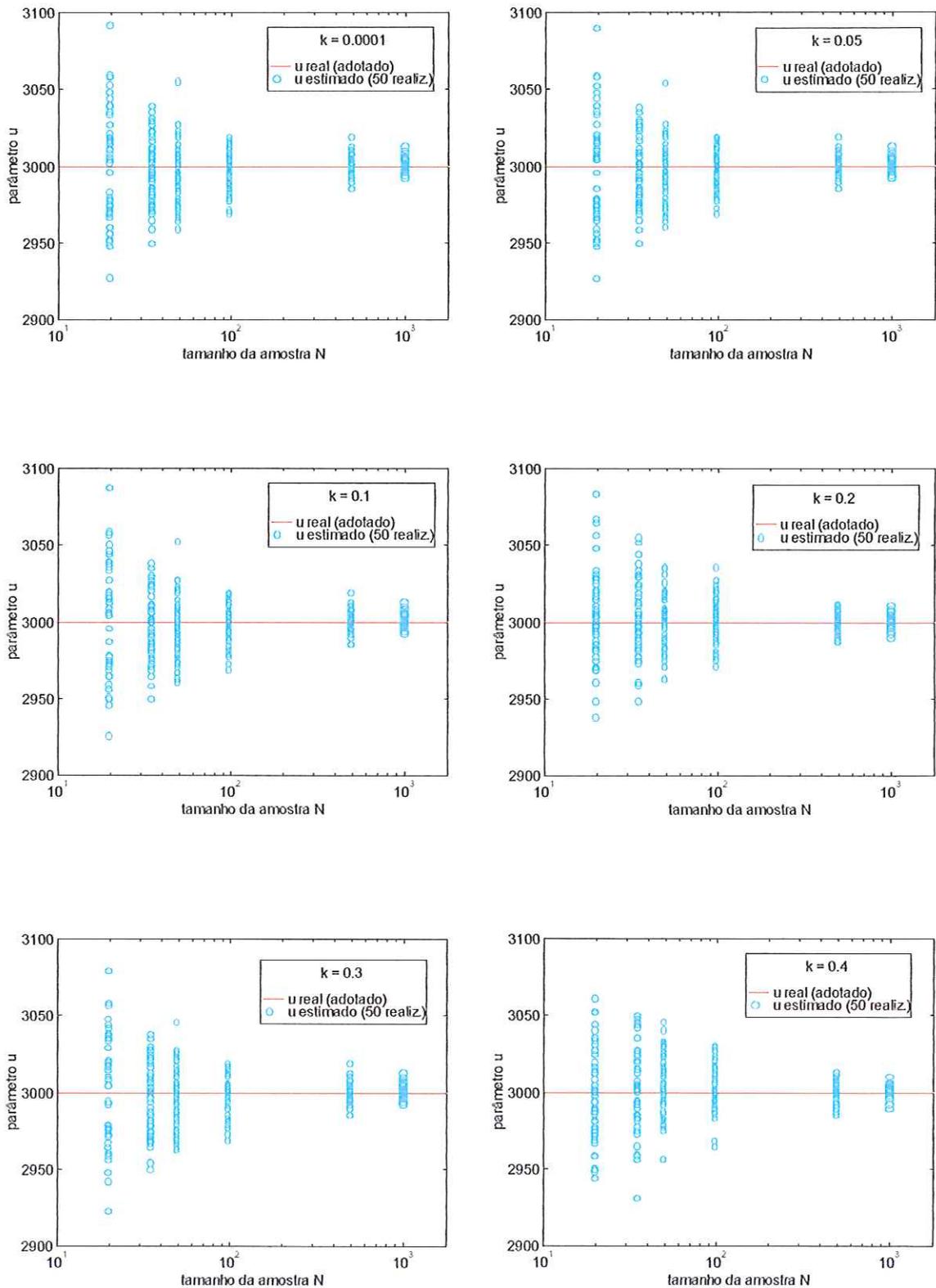


FIGURA 15 - Efeito do valor real de  $k$  sobre a estimação de  $u$ , em função do tamanho da amostra, mantido  $\alpha = 120$  e  $u = 3000$ .

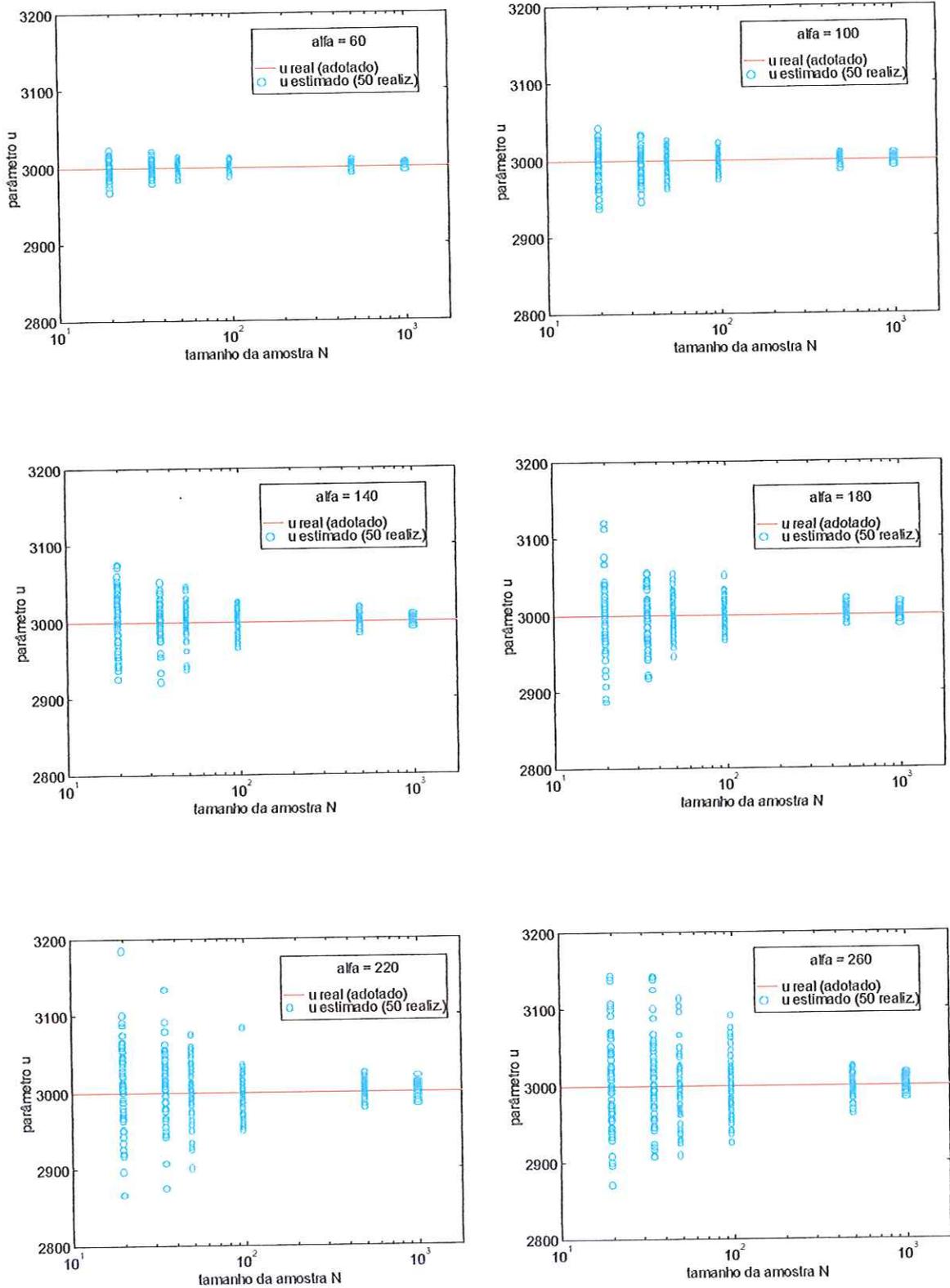
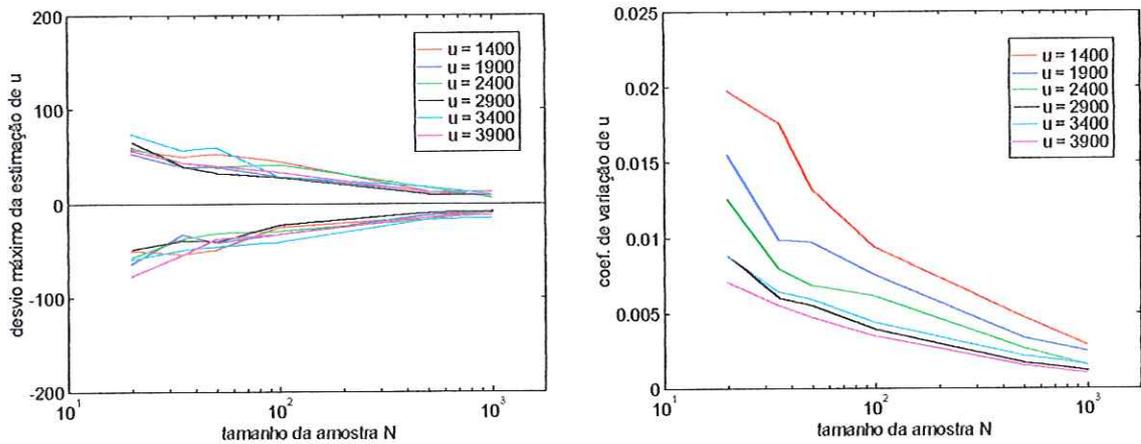
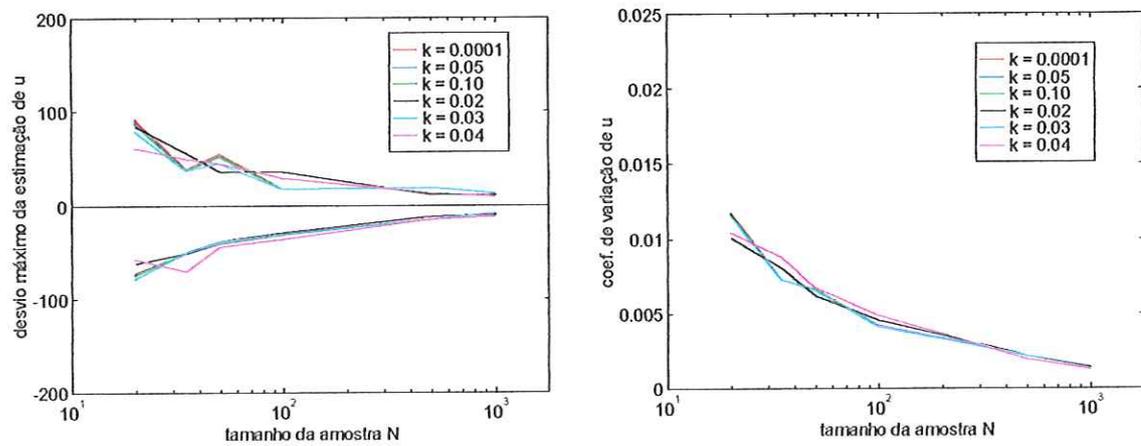


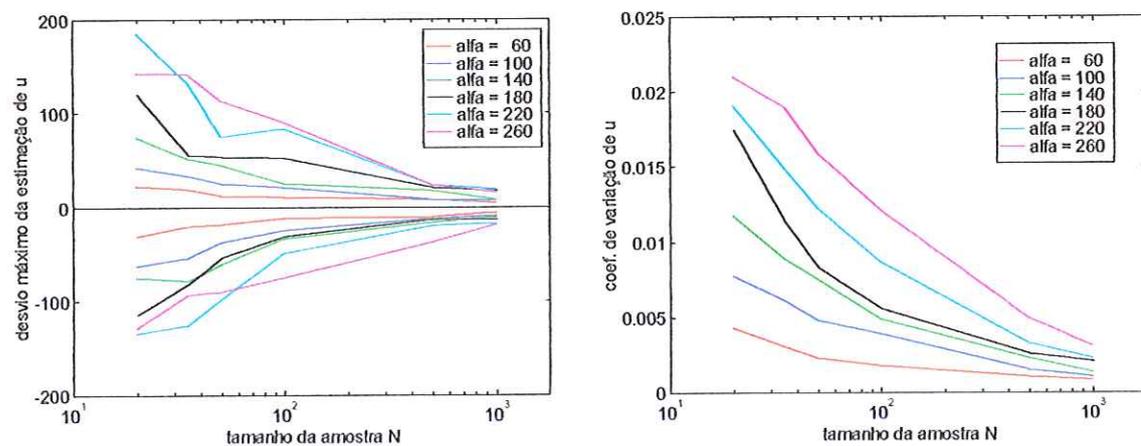
FIGURA 16 - Efeito do valor real de  $\alpha$  sobre a estimação de  $u$ , em função do tamanho da amostra, mantido  $k = 0,20$  e  $u = 3000$ .



a) estimação de  $u$ , usando vazões simuladas com diferentes valores de  $u$ ,  $k=0,20$  e  $\alpha=120$



b) estimação de  $u$ , usando vazões simuladas com diferentes valores de  $k$ ,  $\alpha=120$  e  $u=3000$



c) estimação de  $u$ , usando vazões simuladas com diferentes valores de  $\alpha$ ,  $k=0,2$  e  $u=3000$

FIGURA 17 - Dispersão da estimação de  $u$  verificada através do desvio máximo e do coeficiente de variação, em função do tamanho da amostra.

reais de  $\alpha$ . Contudo, a boa consistência dos seus resultados é comprovada com os baixos valores de coeficiente de variação, mesmo para  $N = 20$ , onde os mesmos não excedem de 2 %.

De modo geral, a precisão da estimação dos parâmetros  $k$ ,  $\alpha$  e  $u$  é mais fortemente afetada com o tamanho da amostra do que com a variação de um ou outro parâmetro. Este efeito é mais marcante no caso do parâmetro de forma, especificamente, quando seus valores esperados com a estimação, encontram-se próximos de zeros, que devido a ocorrência simétrica dos dados em torno do valor real, a estimação resulta em valores negativos de  $k$  na mesma proporção e magnitude dos positivos.

#### 6.1.2 Estimativas das Cheias Decamilenares $Q_{10000}$

A estimação das cheias decamilenares refere-se aos resultados obtidos com a aplicação da forma inversa da distribuição GEV (equação 33a) para determinação dos quantis  $Q_{10000}$ , utilizando as estimativas dos parâmetros  $k$ ,  $\alpha$  e  $u$  apresentados no item 6.1.1. Assim, as cheias decamilenares são estimadas, considerando o efeito da variação de cada parâmetro sobre sua modelação.

Os dados conhecidos da cheia decamilenar, usados como referência neste estudo, compreendem àquelas determinadas através da equação (33a), utilizando as diversas combinações de valores reais de  $k$ ,  $\alpha$  e  $u$ , empregadas no processo de simulação das vazões diárias.

Os resultados de cheias decamilenares mostrados na Figura 18, são resultantes do uso dos parâmetros  $k$ ,  $\alpha$  e  $u$  estimados a partir de diferentes valores reais de  $k$ , sendo mantidos  $\alpha = 120$  e  $u = 3000$ . Na Figura 19, são exibidos os dados de cheias decamilenares determinados com as estimativas de  $k$ ,  $\alpha$  e  $u$ , decorrentes dos diferentes valores reais de  $\alpha$ ,  $k = 0,2$  e  $u = 3000$ . A Figura 20 apresenta as cheias decamilenares obtidas com os dados estimados de  $k$ ,  $\alpha$  e  $u$  referentes aos diversos valores reais de  $u$ ,  $k = 0,2$  e  $\alpha = 120$ .

A Figura 21 exhibe o comportamento exato (teórico) da cheia decamilenar resultante da combinação de vários valores reais de dois dos três parâmetros  $k$ ,  $\alpha$  e  $u$ , sendo mantido o um deles constante. Os Gráficos da referida figura possibilitam a

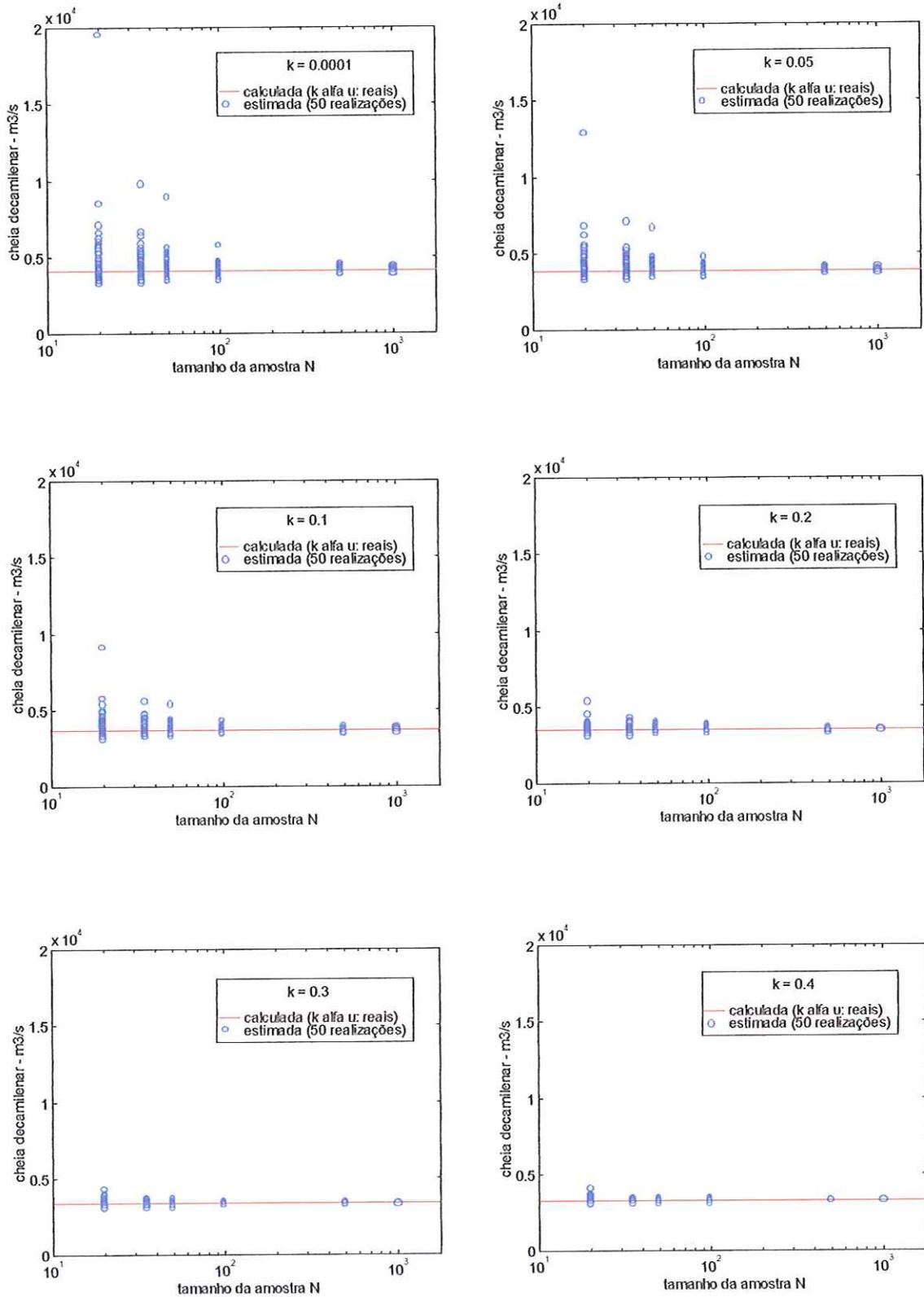


FIGURA 18 - Efeito do valor real de  $k$  sobre a estimação da cheia decamilenar, em função do tamanho da amostra, mantido  $\alpha = 120$  e  $u = 3000$ .

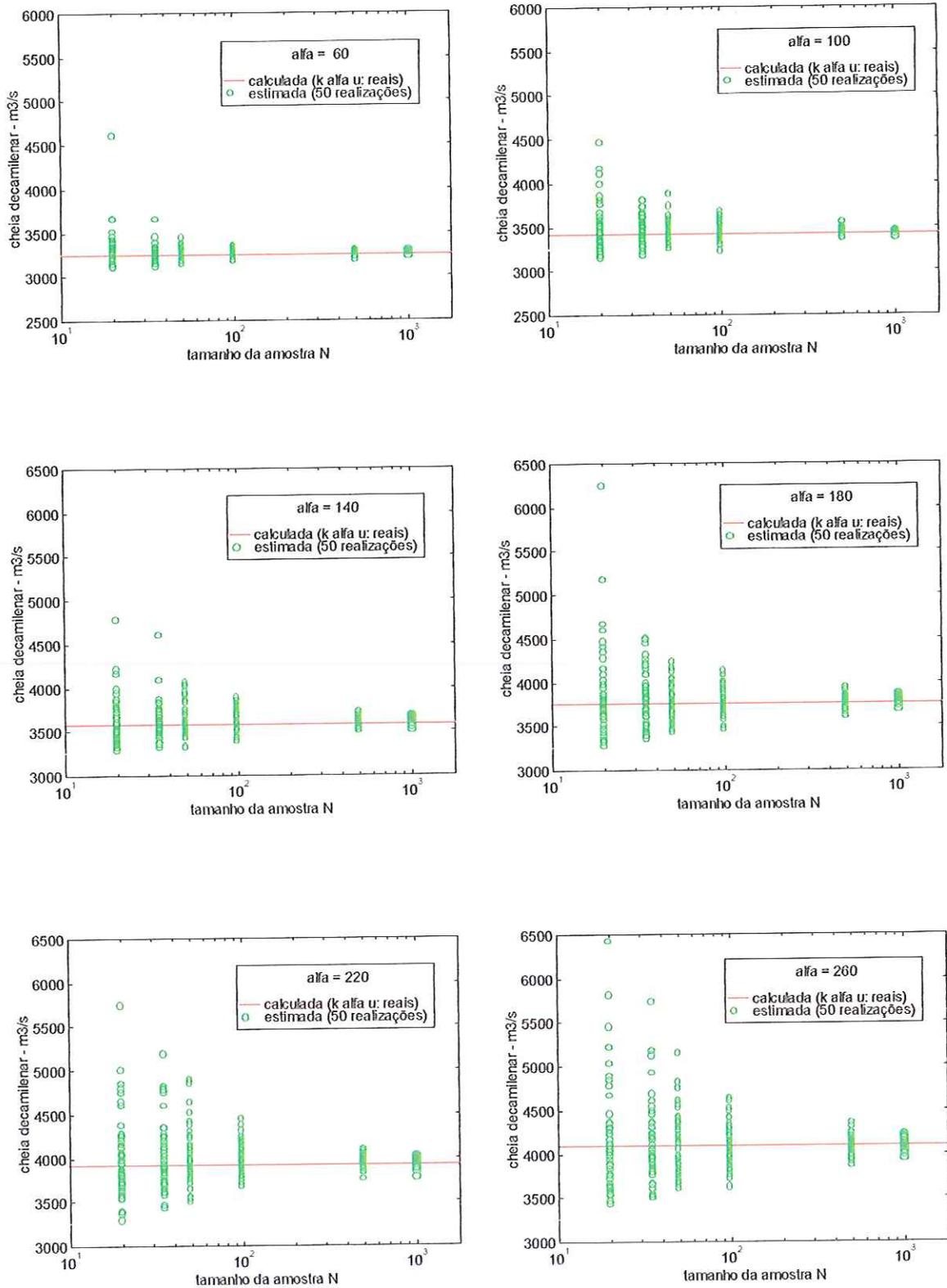


FIGURA 19 - Efeito do valor real de  $\alpha$  sobre a estimação da cheia decamilenar, em função do tamanho da amostra, mantido  $k = 0,20$  e  $u = 3000$ .

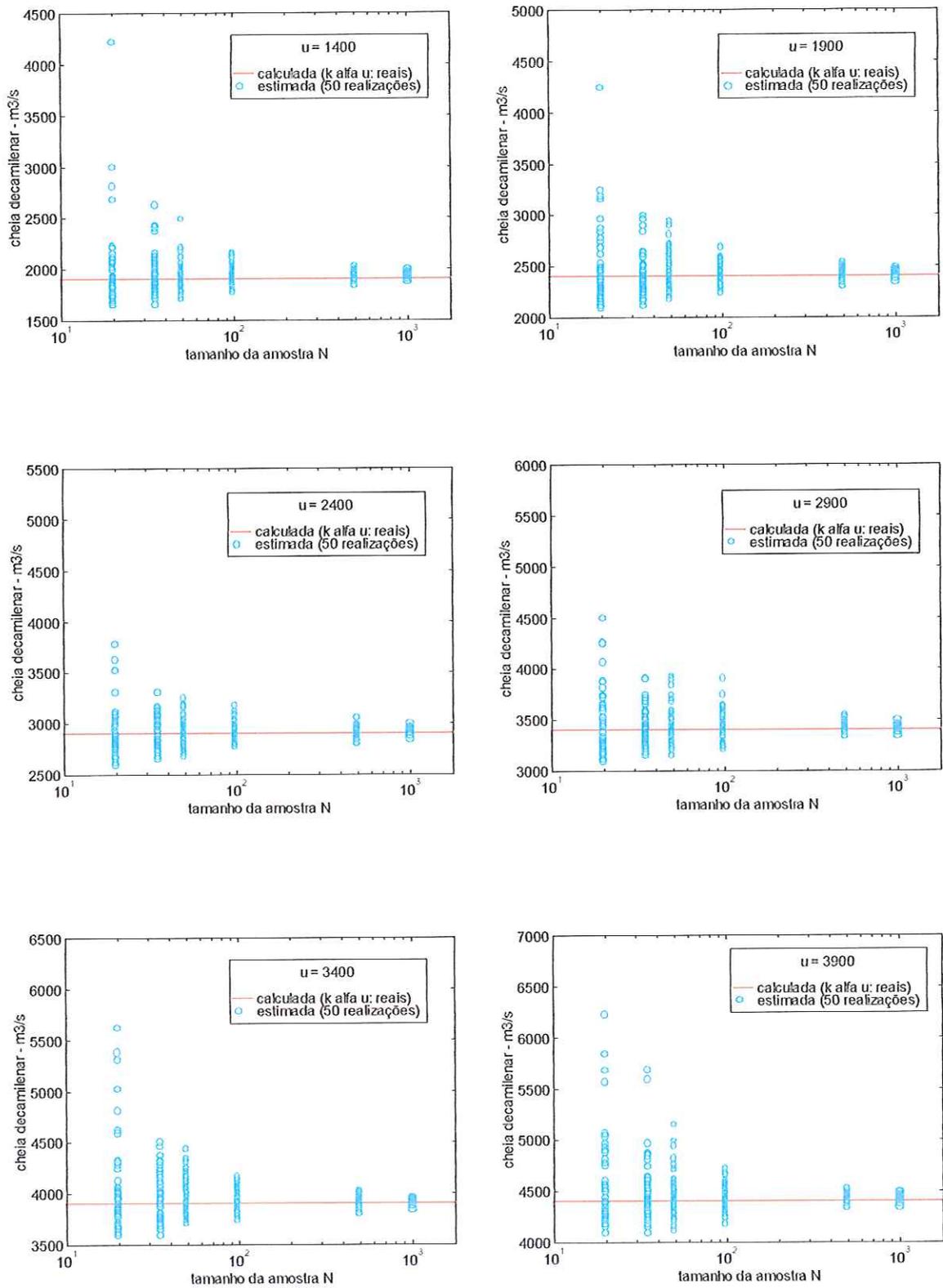


FIGURA 20 - Efeito do valor real de  $u$  sobre a estimação da cheia decamilenar, em função do tamanho da amostra, mantido  $k = 0,20$  e  $\alpha = 120$ .

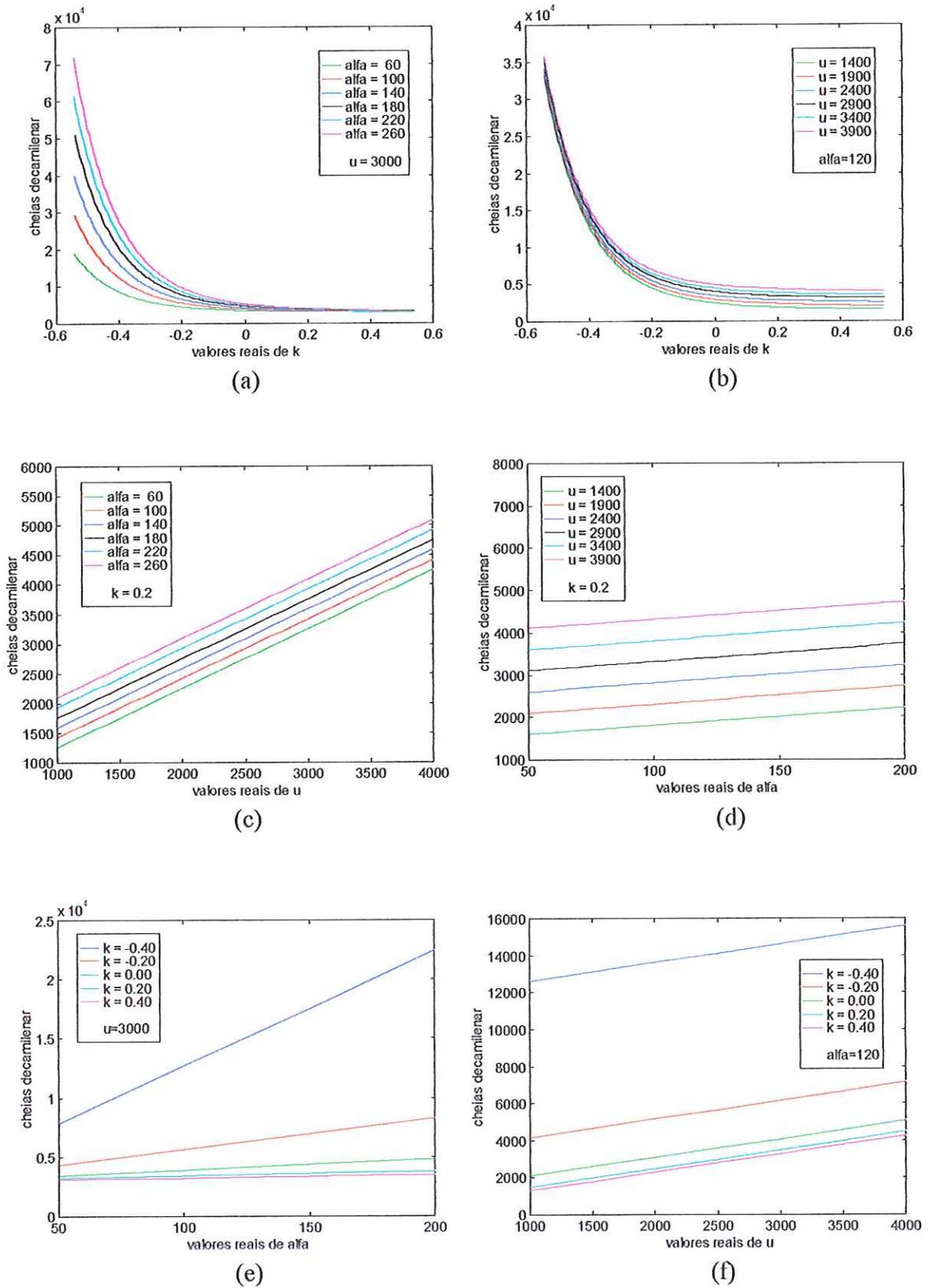


FIGURA 21 - Comportamento teórico dos valores de cheias decamilenares em função da variação dos parâmetros  $k$ ,  $\alpha$  e  $u$  da distribuição GEV.

obtenção de uma melhor compreensão do efeito da variabilidade das estimativas de cada parâmetro, na magnitude da cheia decamilenar estimada.

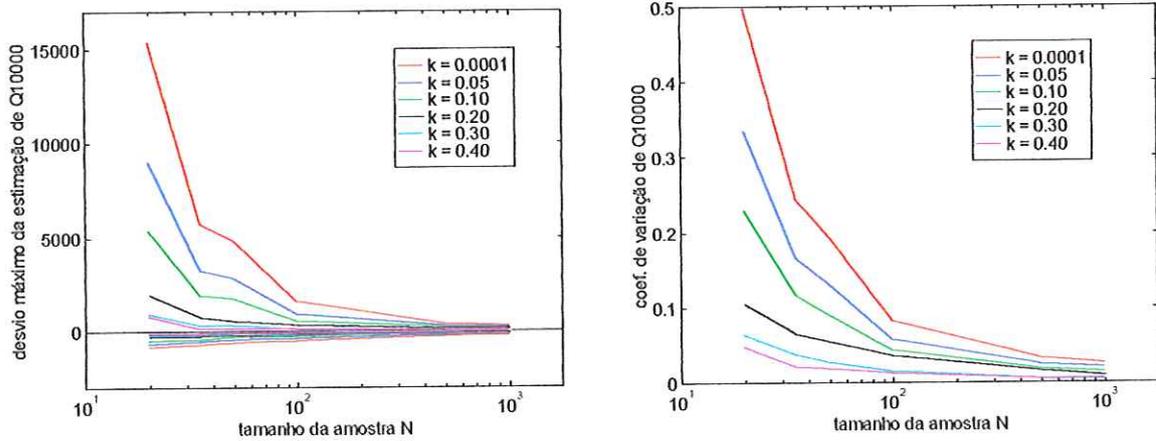
A dispersão da cheia decamilenar estimada, em relação ao seu valor calculado, determinada através do desvio máximo de estimação e do coeficiente de estimação de  $Q_{10000}$ , referentes aos seis diferentes valores reais de  $k$ ,  $\alpha$  e  $u$ , está apresentada na Figura 22, em função do tamanho da amostra.

Observa-se através dos Gráficos *a*, *b*, *e* e *f* da Figura 21 que o aumento acentuado da magnitude da cheia decamilenar sempre ocorre na faixa de valores de  $k < 0$ . Em que para valores mais negativos de  $k$ , a variação de  $\alpha$  altera de forma significativa a elevada magnitude de  $Q_{10000}$  (Gráficos *a* e *e* da referida figura). Com respeito ao parâmetro  $u$ , não há grandes diferenças resultantes da sua variação, mas, o valor de  $Q_{10000}$  torna-se extremamente elevado (Gráficos *b* e *f*, Figura 21) à medida que o valor negativo de  $k$  cresce.

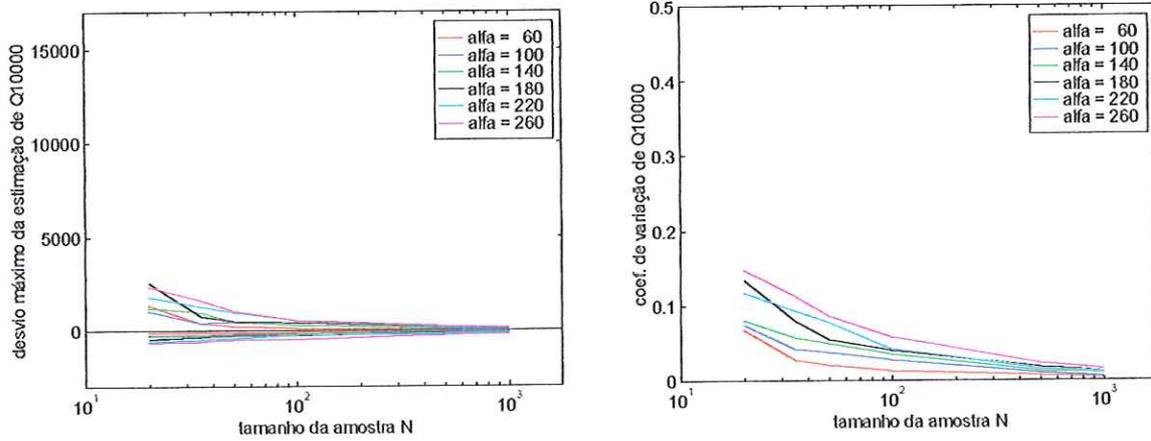
Os dados de cheia decamilenar apresentados na Figura 18 tornam evidente a influência da variabilidade do parâmetro de forma, notadamente, o efeito dos valores negativos de  $k$  estimado sobre a magnitude de  $Q_{10000}$ . Nota-se, que para os seis diferentes valores reais de  $k$ , as estimativas da cheia decamilenar posicionadas abaixo do valor calculado, são resultantes dos dados estimados de  $k$  situados acima do seu valor real (Figura 6). Já as cheias decamilenares que encontram-se acima da cheia calculada, provém dos valores estimados de  $k$  inferiores ao seu valor real, inclusive as estimativas de  $k < 0$ . Neste caso, os maiores desvios ocorrem para valores reais de  $k \leq 0,2$  e tamanhos de amostras inferiores a 100 anos, os quais correspondem aos intervalos de ocorrência de estimativas negativas de  $k$ .

As vazões decamilenares são superestimadas, em relação aos seus valores reais (calculados), com maior frequência do que subestimada. As maiores magnitudes de  $Q_{10000}$  estão relacionadas às amostras com menor número de dados, provenientes de simulação com valores de  $k$  próximos de zero. A subestimação da mesma, decorrente da superestimação, ocorre bem próximo da cheia decamilenar calculada.

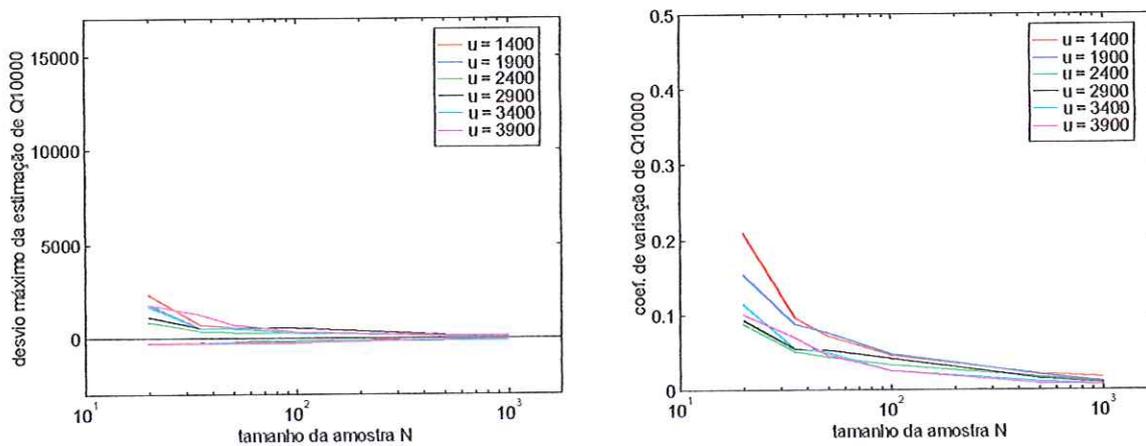
Do ponto de vista prático, isto significa que, quando as vazões máximas anuais seguem a distribuição Gumbel, a estimação da cheia decamilenar através da modelação de séries de cheia anuais composta com baixo número de dados, especial-



a) estimação de  $Q_{10000}$  usando vazões simuladas com diferentes valores de  $k$ ,  $\alpha = 120$  e  $u = 3000$



b) estimação de  $Q_{10000}$  usando vazões simuladas com diferentes valores de  $\alpha$ ,  $k=0,2$  e  $u = 3000$



c) estimação de  $Q_{10000}$  usando vazões simuladas com diferentes valores de  $u$ ,  $k=0,2$  e  $\alpha = 120$

FIGURA 22 - Dispersão da estimação de  $Q_{10000}$  verificada através do desvio máximo e do coeficiente de variação, em função do tamanho da amostra.

mente, amostra com menos de 50 anos, é passível da obtenção de valores muito superestimados de  $Q_{10000}$ . Contudo, se as cheias anuais são descritas através da distribuição GEV limitada (forma VEIII) com parâmetro de forma superior a 0,1, a estimação da cheia decamilenar resulta em valores próximos da verdadeira cheia, mesmo quando utilizam-se amostras pequenas.

O efeito do parâmetro de escala sobre a estimação da cheia decamilenar, descrito através dos dados da Figura 19, é evidenciado nas amostras inferiores a 100 anos, com maior predominância da superestimação de  $Q_{10000}$  relacionados aos valores reais mais elevados de  $\alpha$ . Este efeito está associado à ocorrência de estimativas negativas de  $k$  na mesma faixa de valores de  $\alpha$  e  $N$  (Figura 7) e é condizente com os resultados mostrados no Gráfico *a* da Figura 20. No caso do parâmetro  $\alpha$ , tanto a subestimação como a superestimação da cheia decamilenar, são resultantes das estimativas de  $k$  que encontram-se, respectivamente, acima e abaixo do valor real.

Os dados de vazões decamilenares plotados na Figura 20, referentes aos diferentes valores reais de  $u$ , apresentam um padrão de dispersão que é indistingüível entre os seis valores de  $u$ . Para amostras com  $N < 100$  anos, há superestimação da cheia decamilenar, com maior intensidade para  $N = 20$ . A superestimação ou a subvalorização de  $Q_{10000}$  é estabelecida pela variabilidade das estimativas do parâmetro de forma, relacionada aos diferentes valores de  $u$  e tamanho de amostra. Porquanto, os resultados mais elevados de  $Q_{10000}$  decorrem das estimativas negativas de  $k$  mais distante de zero.

O efeito de cada parâmetro sobre a estimação da cheia decamilenar, só acontece de maneira significativa, quando em qualquer um dos casos, os valores do parâmetro de forma são menores do que zero. Para valores de  $k > 0$  não há influência marcante das estimativas de  $k$ ,  $\alpha$  ou  $u$  sobre a magnitude de  $Q_{10000}$ , cujo resultado, neste caso, ocorre bem próximo do verdadeiro valor da cheia decamilenar.

A evidência do efeito de cada parâmetro sobre a cheia decamilenar torna-se mais contundente, quando se verifica a dispersão de suas estimativas apresentada na Figura 22. Os desvios máximos de  $Q_{10000}$  demonstram claramente o efeito de cada valor real de  $k$  em amostras com  $N < 100$  anos. Os desvios exagerados relacionados a  $N = 20$  e  $k < 0,2$  são resultantes da influência dos valores negativos de  $k$  mais distantes de zero, que ocorreram durante o processo de estimação. Para as estimativas

de  $k > 0$  e superiores ao seu valor real, os desvios máximos da cheia decamilenar são desprezíveis. Neste caso, é impossível atribuir qualquer efeito aos diferentes valores de  $k$  (Gráfico *a* da Figura 22).

Os desvios máximos de  $Q_{10000}$  referentes aos valores de  $\alpha$  e  $u$  mostrados nos Gráficos *b* e *c* da Figura 22, respectivamente, apresentam o mesmo padrão, onde não é possível separar o efeito de cada parâmetro. Os maiores desvios estão relacionados às menores amostras, com valores expressivos crescente para  $N \leq 50$ .

Os coeficientes de variação da cheia decamilenar decorrentes dos diversos valores de  $k$  (Gráfico *a* da Figura 22) evidencia a distinta variabilidade das estimativas de  $Q_{10000}$  para cada valor real de  $k$  e tamanho de amostra, que cresce acentuadamente com a diminuição dos mesmos. Os coeficiente de variação da cheia decamilenar relacionados aos valores de  $\alpha$  e de  $u$  (Gráficos *b* e *c* Figura 22) encontram-se aproximadamente na mesma faixa de variação, onde percebe-se uma pequena intervenção dos valores de  $\alpha$  na variabilidade de  $Q_{10000}$ . Fato que não acontece com o parâmetro  $u$ , cujo efeito perceptível diz respeito ao comprimento da amostra.

Fica evidente que a dispersão da cheia decamilenar considerada relevante, decorrem da presença de valores elevados de  $k < 0$  no processo de estimação, os quais estão relacionados às amostras de cheias utilizadas na modelação, com menores disposições de dados.

### 6.1.3 Avaliação das Incertezas de Estimação de $k$ , $\alpha$ , $u$ e $Q_{10000}$

Para avaliar as incertezas de estimação de  $k$ ,  $\alpha$  e  $u$  e da correspondente cheia decamilenar determinaram-se a raiz do erro quadrático médio percentual ( $reqm$ ) e o erro máximo de estimação das referidas variáveis, considerando o efeito das variações de cada parâmetro e do tamanho da amostra utilizada na modelação.

As Figuras 23, 24 e 25 exibem a  $reqm$  de  $k$ ,  $\alpha$  e  $u$  e de  $Q_{10000}$ , em função do tamanho da amostra, respectivamente, para os diversos valores reais de  $k$ ,  $\alpha$  e  $u$ , usados na geração das vazões diárias.

Os dados da Figura 23 (Gráfico *a*) mostram que a incerteza da estimação de  $k$  torna-se extremamente elevada quando as cheias máximas seguem, assintoticamente,

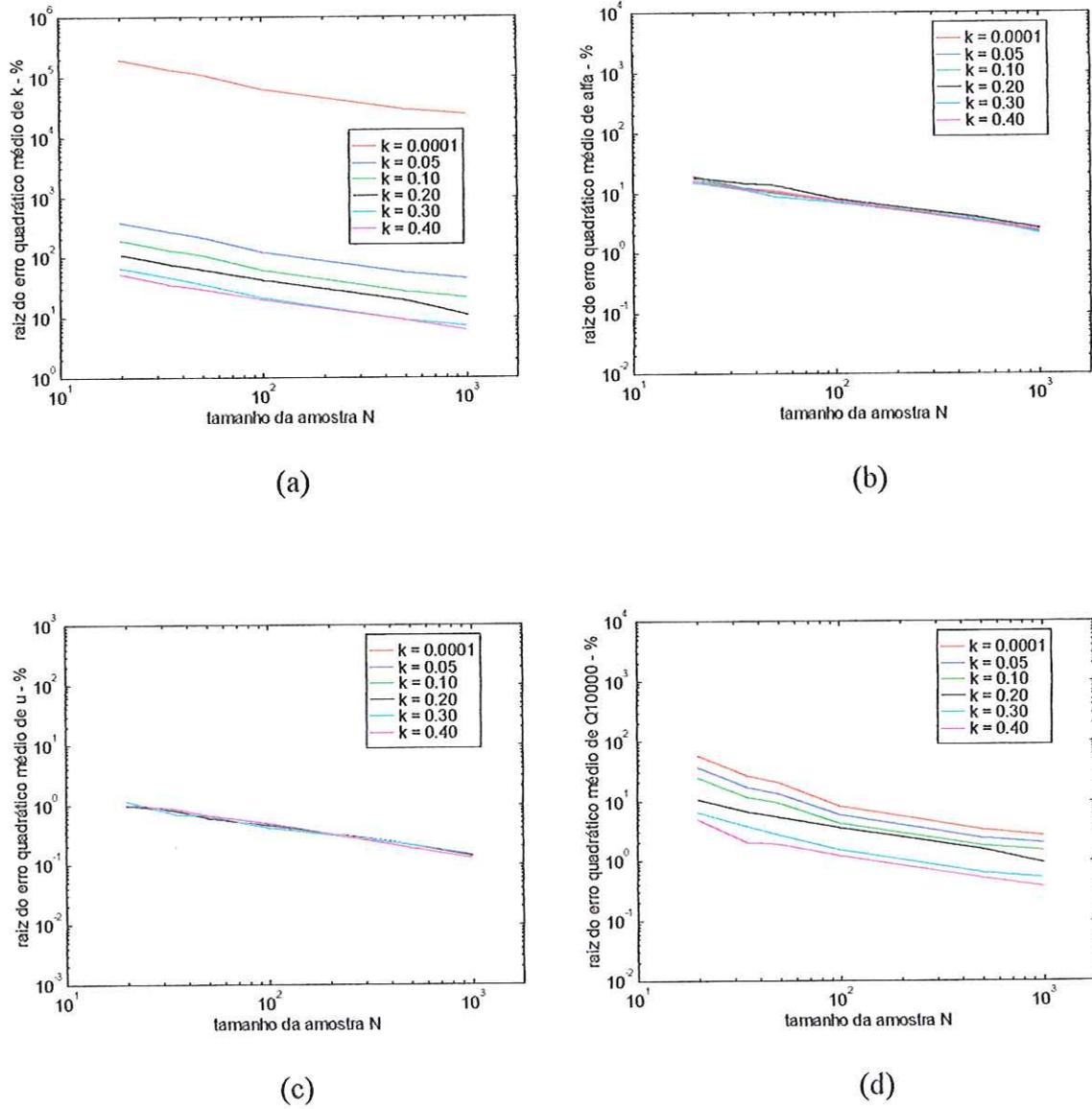


FIGURA 23 - Raiz do erro quadrático médio da estimação dos parâmetros: (a)  $k$ , (b)  $\alpha$ , (c)  $u$  e da (d) cheia decamilenar ( $Q_{10000}$ ), em função do tamanho da amostra, para os diferentes valores reais de  $k$ ,  $\alpha = 120$  e  $u = 3000$ .

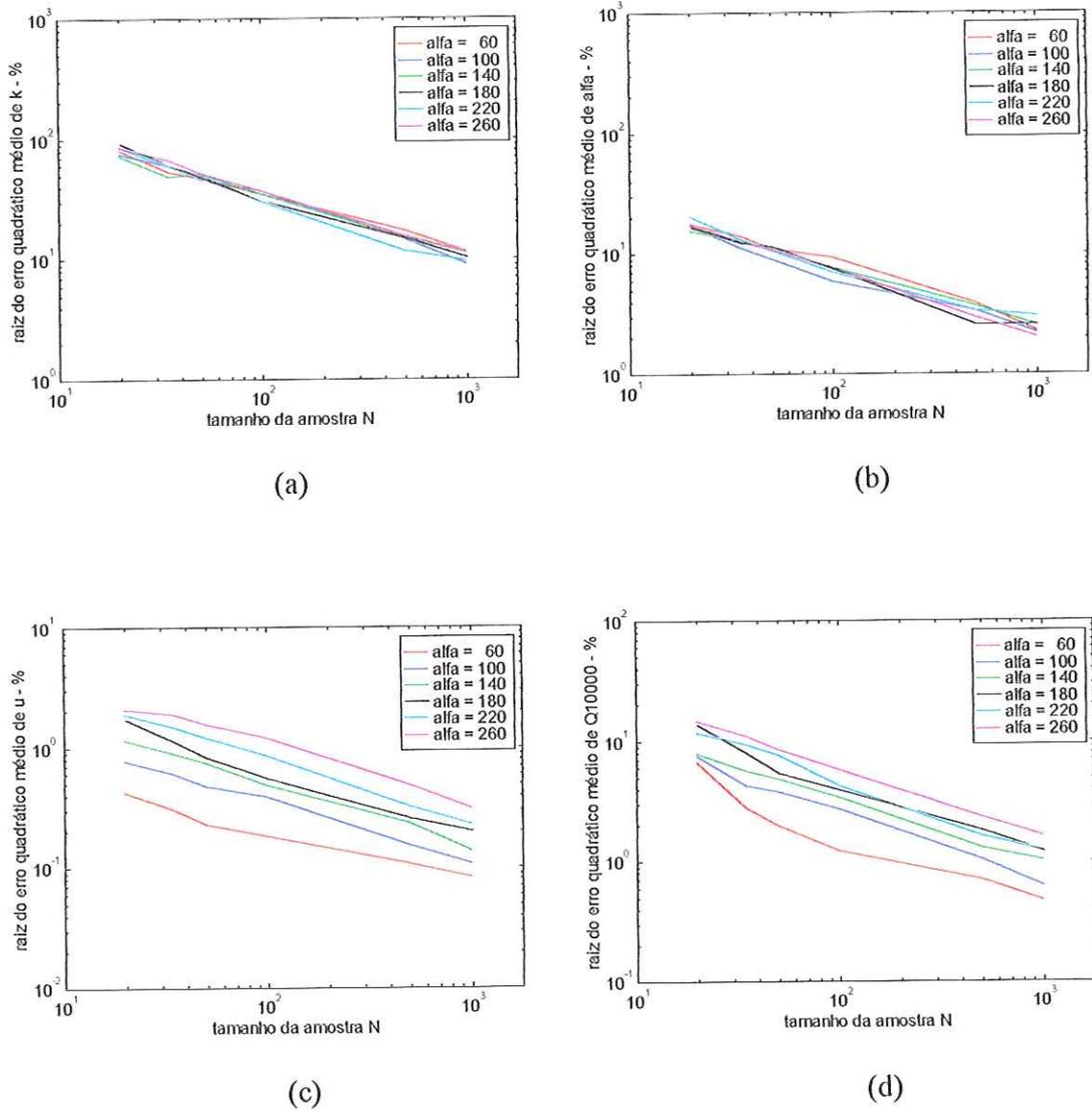


FIGURA 24 - Raiz do erro quadrático médio da estimação dos parâmetros: (a)  $k$ , (b)  $\alpha$ , (c)  $u$  e da (d) cheia decamilenar ( $Q_{10000}$ ), em função do tamanho da amostra, para os diferentes valores reais de  $\alpha$ ,  $k = 0,20$  e  $u = 3000$ .

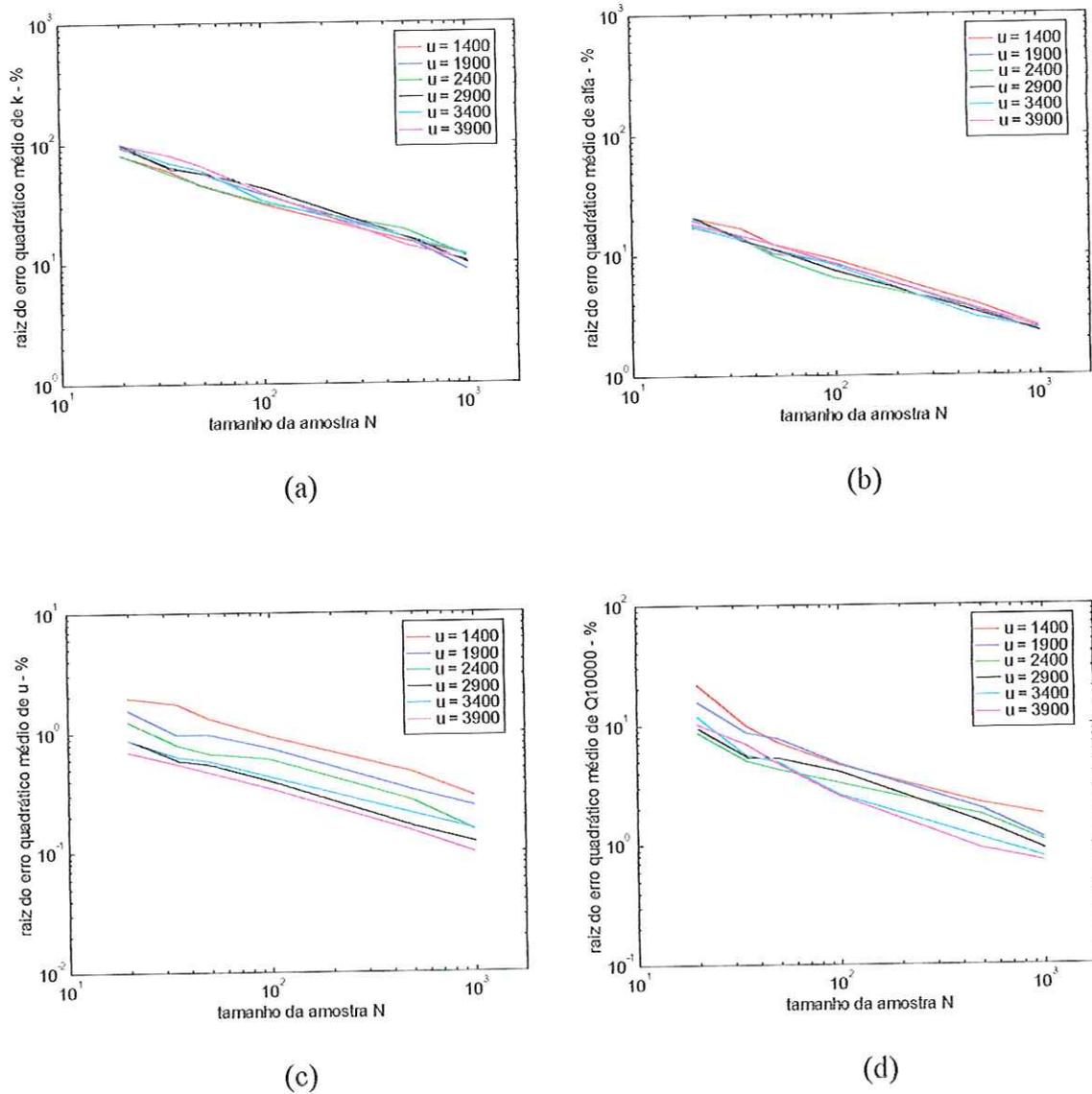


FIGURA 25 - Raiz do erro quadrático médio da estimação dos parâmetros: (a)  $k$ , (b)  $\alpha$ , (c)  $u$  e da (d) cheia decamilenar ( $Q_{10000}$ ), em função do tamanho da amostra, para os diferentes valores reais de  $u$ ,  $\alpha = 120$  e  $k = 0,20$ .

a distribuição Gumbel, notadamente, para amostras pequenas. Não obstante, mesmo para valores reais de  $k$  superiores a 0,5, a incerteza da estimação de  $k$  ainda é elevada. Ademais, há uma relação direta entre a precisão da estimação e o tamanho da amostra. Em amostras pequenas, como  $N=20$ , a  $reqm$  varia de 50 para 500 % quando  $k$  assume valores entre 0,4 e 0,05, atingindo a ordem de  $10^5$  % quando  $k$  se aproxima de zero (distribuição Gumbel). Nota-se também, que para um mesmo valor de  $k$ , o  $reqm$  da estimação de  $k$  é 10 vezes maior quando o tamanho da amostra muda de 1000 para 20 valores, independente da ordem de grandeza da mesma.

Amostras que apresentam, com o ajuste da distribuição GEV, parâmetro de forma entre 0 e 0,03 advém de uma distribuição Gumbel (STEDINGER et al., 1993). Como para essas amostras há uma pobreza de informações à respeito do parâmetro  $k$ , pois para a Gumbel o mesmo não existe, quando ajusta-se a GEV às referidas amostras, obtém-se valores de  $k$  com erros de estimação extremamente elevados a exemplo do gráfico *a* da Figura 23.

Os Gráficos *b* e *c* da Figura 23 demonstram que a variação do parâmetro de forma não intervém na precisão da estimação de  $\alpha$  e  $u$ , respectivamente. O  $reqm$  decorrente do tamanho da amostra ocorre na mesma proporção referente ao parâmetro  $k$ . Todavia, a incerteza presente na estimação de  $\alpha$  é 10 vezes superior àquela que acontece com a estimação de  $u$ .

O Gráfico *d* da Figura 23 indica que o grau de incerteza da estimação de  $Q_{10000}$  é bem inferior àquele ocorrido com o parâmetro de forma e não segue nenhum padrão que possa associar-se à grandeza do erro referente à  $k = 0,0001$ . Observa-se um relativo efeito de  $k$  sobre o erro de estimação da cheia decamilenar com a  $reqm$  10 vezes menor quando comparada à  $reqm$  ocorrida com os diferentes valores de  $k$ , excluído  $k = 0,0001$ . A incerteza proveniente da variação do tamanho da amostra sobre  $k$ ,  $\alpha$ ,  $u$  e  $Q_{10000}$  mantém a proporção de 10 vezes da  $reqm$  de  $N=1000$  quando tem-se  $N=20$ .

De modo geral, a  $reqm$  da cheia decamilenar varia de 5 para 50 % quando  $k$  passa de 0,4 para 0,0001 em amostra com 20 anos. Se  $N = 1000$  anos, esses valores mudam de 0,5 para 5 % em relação ao mesmo valor de  $k$ .

Os Gráficos *a* e *b* da Figura 24 apresentam a  $reqm$  dos parâmetros  $k$  e  $\alpha$ , respectivamente, para os diferentes valores de  $\alpha$ , nos quais não é possível identificar

o efeito dos mesmos. A reqm de  $k$  varia de 100 para 10 % quando  $N$  passa de 20 para 1000 anos. Já a reqm de  $\alpha$  varia de 2 à 0,2 % para os referidos valores de  $N$ . A estimação de  $u$  é afetada com a variação de  $\alpha$  real, embora seja perceptível tal efeito, a diferença das reqm's referentes aos diversos valores de  $\alpha$  é bem pequena, onde para  $N = 20$ , a reqm varia de 0,4 até 0,1 % quando  $\alpha$  passa de 260 para 60 (Gráfico *c* da Figura 24). Efeito similar ocorre com a cheia decamilenar, cuja ordem de grandeza da reqm é 10 vezes menor à ocorrida com o parâmetro  $k$  (Gráfico *d* da Figura 24).

O comportamento da reqm de  $k$ ,  $\alpha$ ,  $u$  e de  $Q_{10000}$  relacionado aos diferentes valores reais de  $u$ , observado através da Figura 25, segue o mesmo padrão da Figura 24. Em que o efeito de  $u$  sobre a reqm de sua estimação, embora distintos, ocorre discretamente (Gráfico *c* da Figura 25).

A reqm da cheia decamilenar acontece na mesma ordem de grandeza do parâmetro  $\alpha$ , 10 vezes menor que o parâmetro  $k$  e 10 vezes superior ao parâmetro  $u$ . Essa proporcionalidade da reqm entre os parâmetros e a cheia decamilenar também se repete nos casos das Figuras 23 e 24 para  $k = 0,2$ .

Os erros máximos de estimação apresentados nas Figuras 26, 27 e 28, mostram as incertezas de estimação de  $k$ ,  $\alpha$ ,  $u$  e  $Q_{10000}$ , em função da variação de cada parâmetro, para os diferentes valores de  $N$ , através dos quais constata-se as mesmas incertezas já discutidas, vistas de forma diferente.

A Figura 26 mostra claramente que as incertezas da estimação de  $\alpha$  e  $u$  provém, unicamente, do tamanho da amostra. Já o parâmetro  $k$  e a  $Q_{10000}$  sofrem o efeito da variação do valor real de  $k$  e do tamanho da amostra, destacando-se a mudança abrupta na magnitude do erro máximo de estimação de  $k$ , quando o seu valor real aproxima-se de zero. Isto decorre do fato de que para  $k$  entre 0 e 0,03 as amostras provem de uma distribuição Gumbel, em que a ausência de informações sobre o parâmetro de forma  $k$  resulta em erros de estimação muito elevados.

Quando se analisa o erro máximo de estimação mostrado na Figura 27, para os diferentes valores de  $\alpha$ , nota-se o seu efeito, agindo de forma direta sobre a estimação de  $u$  e de  $Q_{10000}$ . Enquanto, o efeito do tamanho da amostra, prossegue-se com o mesmo padrão de interferência em relação aos parâmetros  $k$ ,  $\alpha$  e  $u$  e à cheia decamilenar.

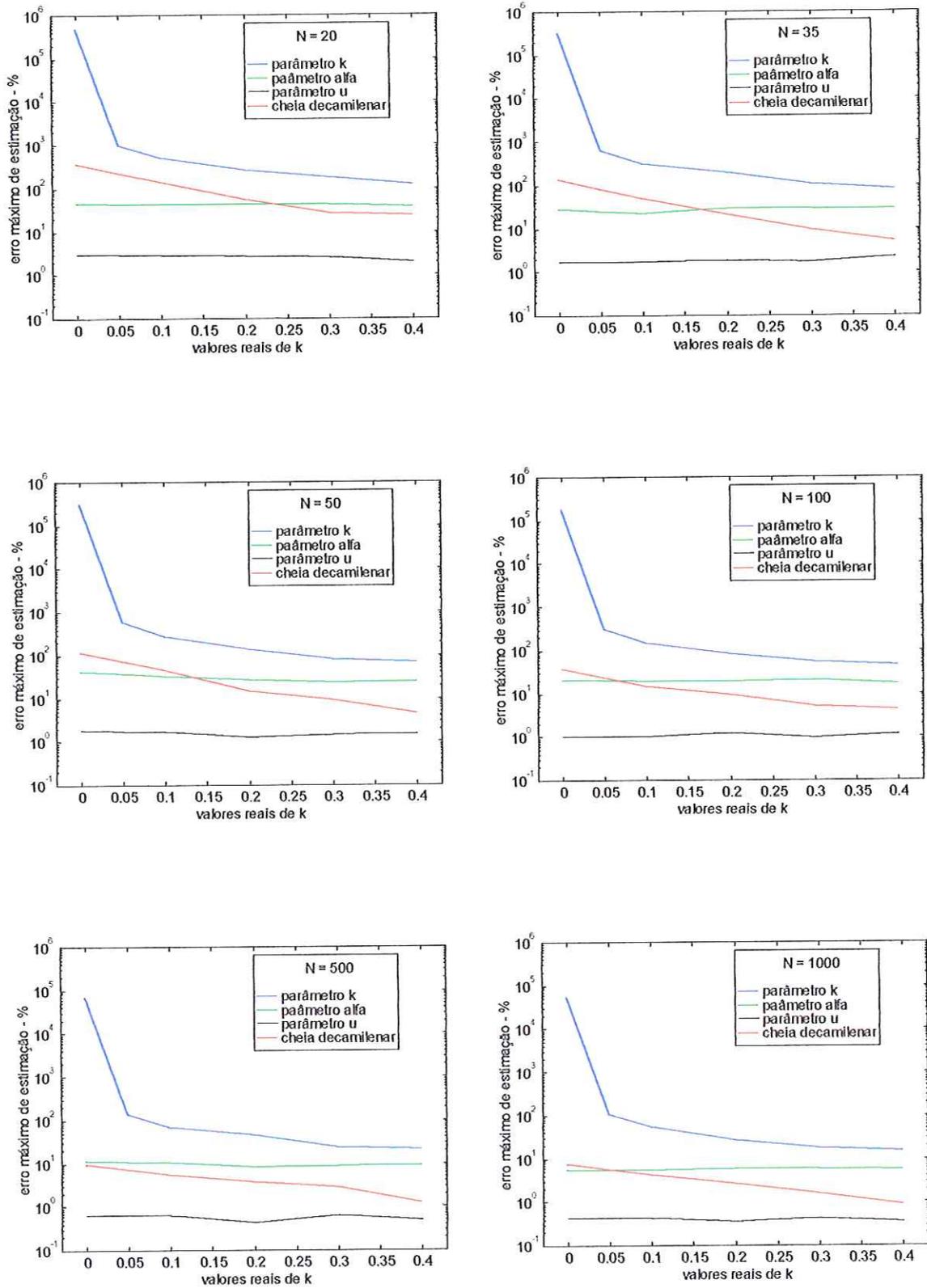


FIGURA 26 - Erro máximo de estimação dos parâmetros  $k$ ,  $\alpha$ ,  $u$  e da cheia decamilenar para os diferentes valores reais de  $k$  e tamanhos de amostras.

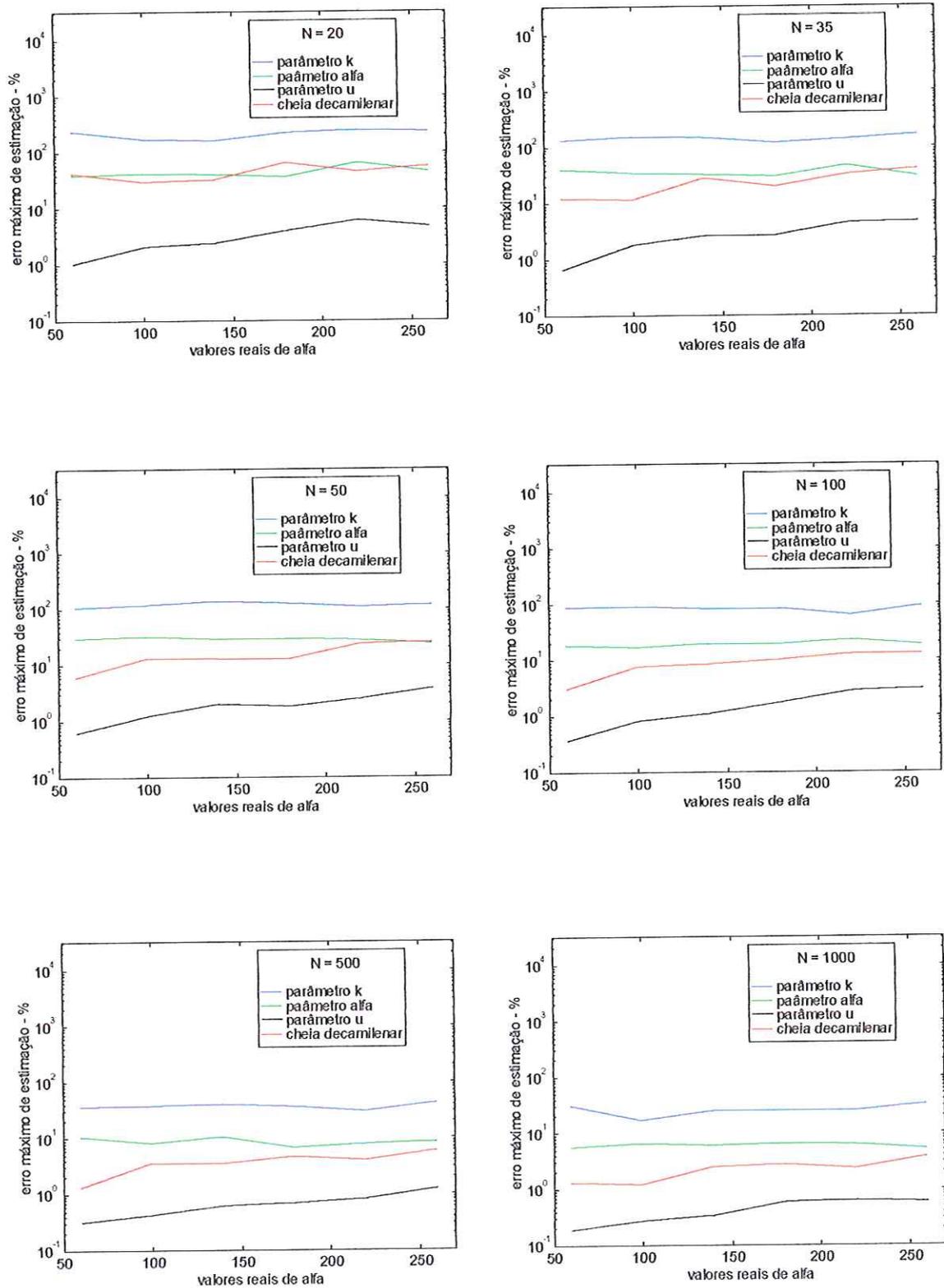


FIGURA 27 - Erro máximo de estimação dos parâmetros  $k$ ,  $\alpha$ ,  $u$  e da cheia decamilenar para os diferentes valores reais de  $\alpha$  e tamanhos de amostras.

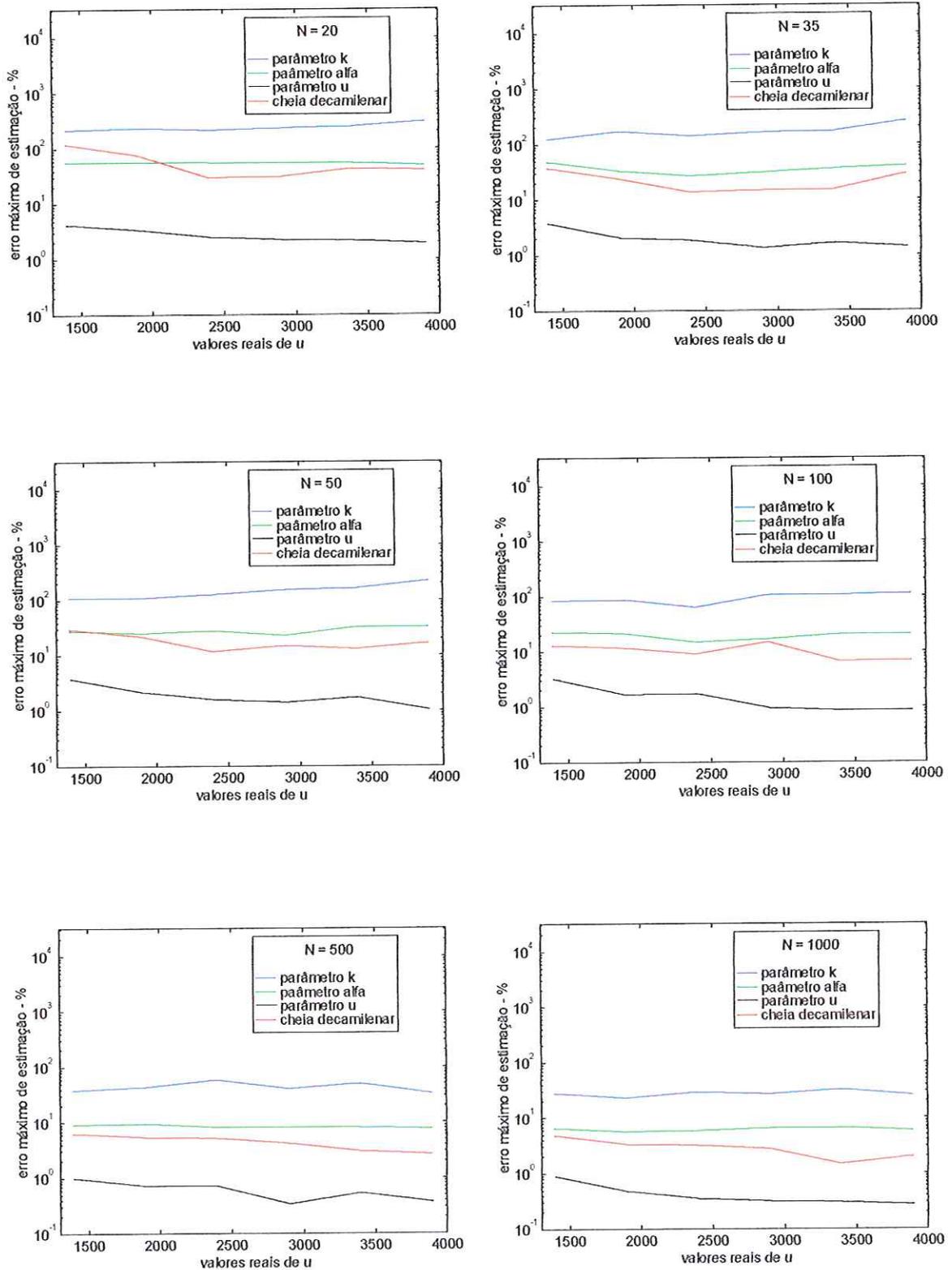


FIGURA 28 - Erro máximo de estimação dos parâmetros  $k$ ,  $\alpha$ ,  $u$  e da cheia decamilenar para os diferentes valores reais de  $u$  e tamanhos de amostras.

Na Figura 28, é perceptível o efeito inverso do valor real de  $u$  sobre sua estimação e sobre a estimação de  $Q_{10000}$ . Entretanto, não há nenhum efeito sobre  $k$  e  $\alpha$ , além do efeito de  $N$ , que ocorre sobre todos eles, seguindo o mesmo padrão dos casos anteriores.

Resumidamente, as incertezas de estimação dos três parâmetros da GEV ocorrem, de modo geral, bem reduzidas para  $\alpha$  e  $u$  e elevada para o parâmetro de forma  $k$ , notadamente, quando as séries de cheias anuais pertencem a uma população que segue a distribuição Gumbel ou a distribuição GEV limitada com parâmetro de forma  $k < 0,2$  ou as amostras apresentam tamanhos inferiores a 100 anos. Assim, em análise de frequência preditiva através da GEV, a incerteza da estimação de  $Q_{10000}$  decorre, preponderantemente, do efeito da alta variabilidade dos valores de  $k$  estimados, imposta pelo tamanho da amostra, para valores menores de 100 anos, do que das incertezas dos demais parâmetros.

#### 6.1.4 Análise da Qualidade do Ajuste

O diagrama de momentos L de assimetria e curtose tem sido usado com um significado de diagnóstico de quão bem uma distribuição descreve os dados numa amostra. Em distribuições com três parâmetros, como a GEV, a distribuição pode ser estimada, utilizando estimativas de posicionamento, escala e assimetria. A comparação da curtose amostral com a a curtose da distribuição infere sobre a qualidade do ajuste e sobre a habilidade da distribuição em descrever os dados.

As razões de momentos  $\tau_{3-L}$  e  $\tau_{4-L}$  que definem, respectivamente, a assimetria L e a curtose L são funções, apenas, do parâmetro de forma  $k$  (equações 56b e 56c). A Figura 29a mostra a relação entre a assimetria L,  $\tau_{3-L}$ , e a curtose L,  $\tau_{4-L}$ , estabelecida para diferentes valores de  $k$ . Vários pares ordenados  $(\tau_3, \tau_4)$  L referentes a valores específicos de  $k$ , inclusive aqueles utilizados no processo de simulação de vazões, são plotados com destaque sobre a curva teórica  $\tau_{4 \times \tau_3}$  L, para evidenciar como a assimetria e a curtose variam com os diversos valores de  $k$ .

A Figura 29b refere-se ao diagrama de momentos L de assimetria e curtose de 107 séries de vazões máxima anuais observadas em rios de várias partes do mundo (WANG, 1997). As Figuras 29a e 29b são usadas como base de discussão para os

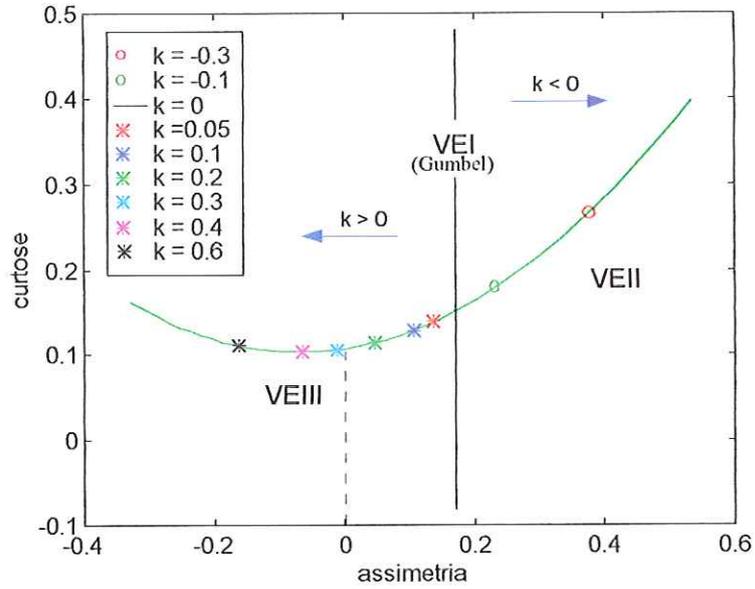


FIGURA 29a - Esquema do diagrama de momentos L da distribuição GEV para diferentes valores de k.

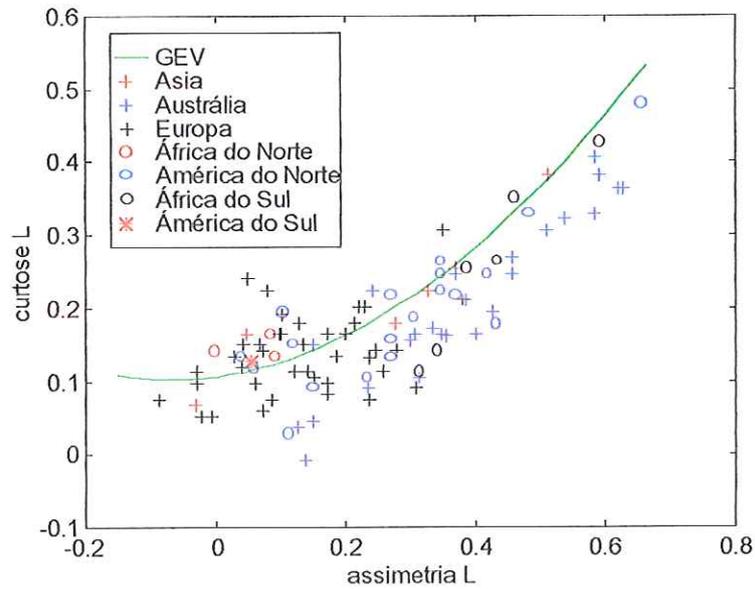


FIGURA 29b – Diagrama de momentos L de vazões máximas anuais de 107 rios mundiais (WANG, 1997).

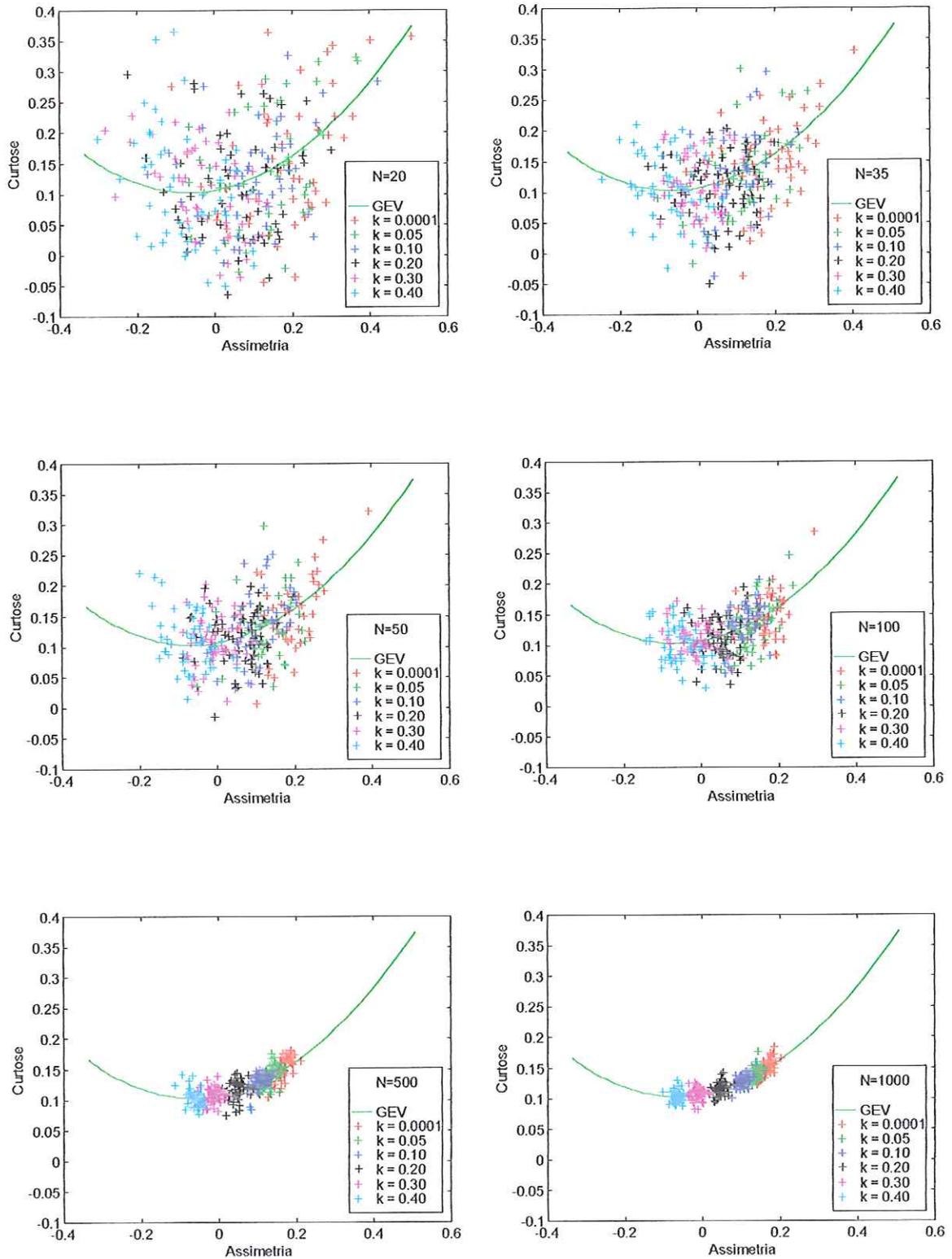


FIGURA 30 - Diagramas de momentos L, referentes aos ajustes da GEV às séries de vazões máximas com vários tamanhos, obtidas de vazões diárias geradas com diferentes valores de  $k$ ,  $\alpha = 120$  e  $u = 3000$ .

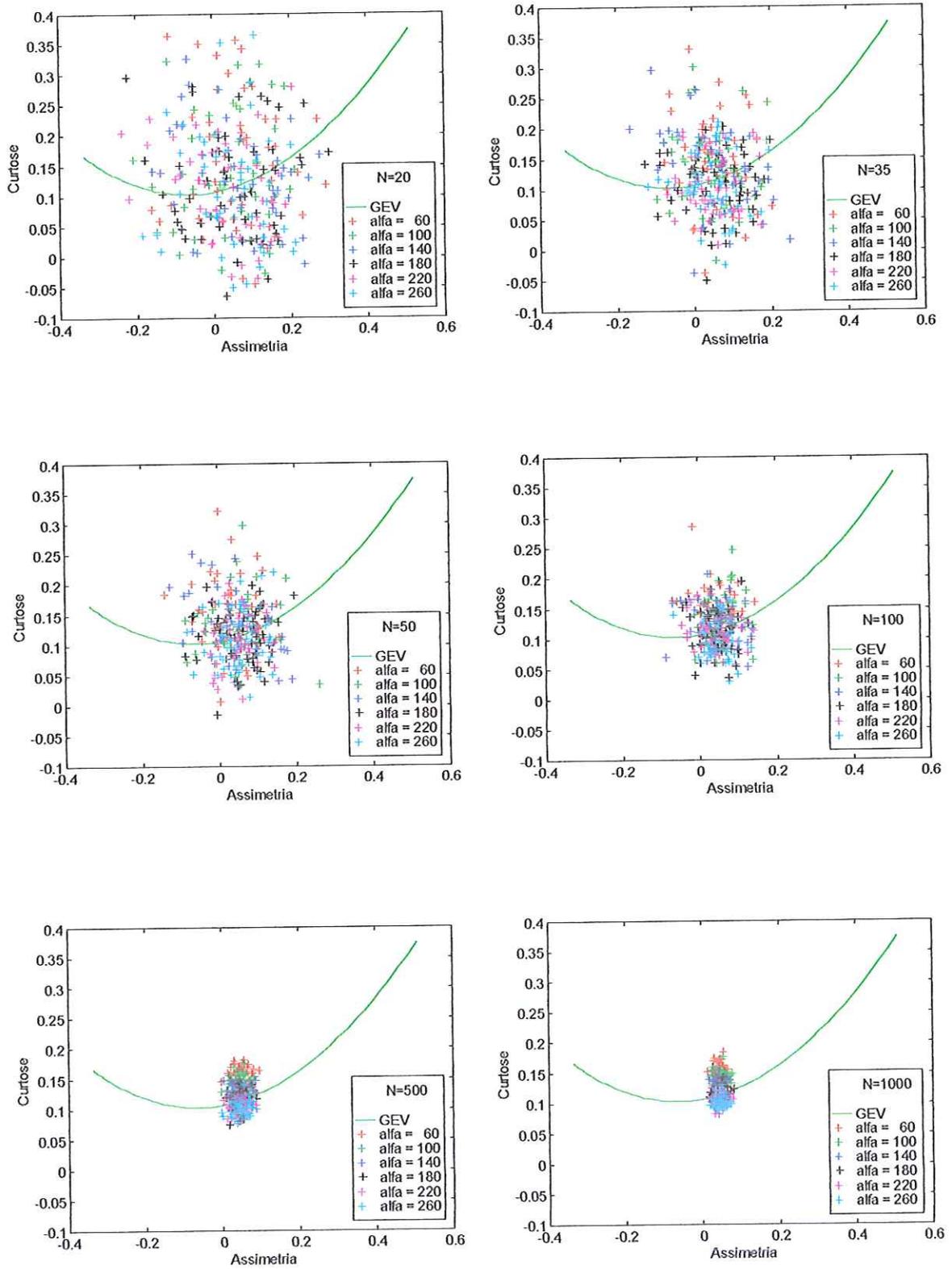


FIGURA 31 - Diagramas de momentos L, referentes aos ajustes da GEV às séries de vazões máximas com vários tamanhos, obtidas de vazões diárias geradas com diferentes valores de  $\alpha$ ,  $k = 0,20$  e  $u = 3000$ .

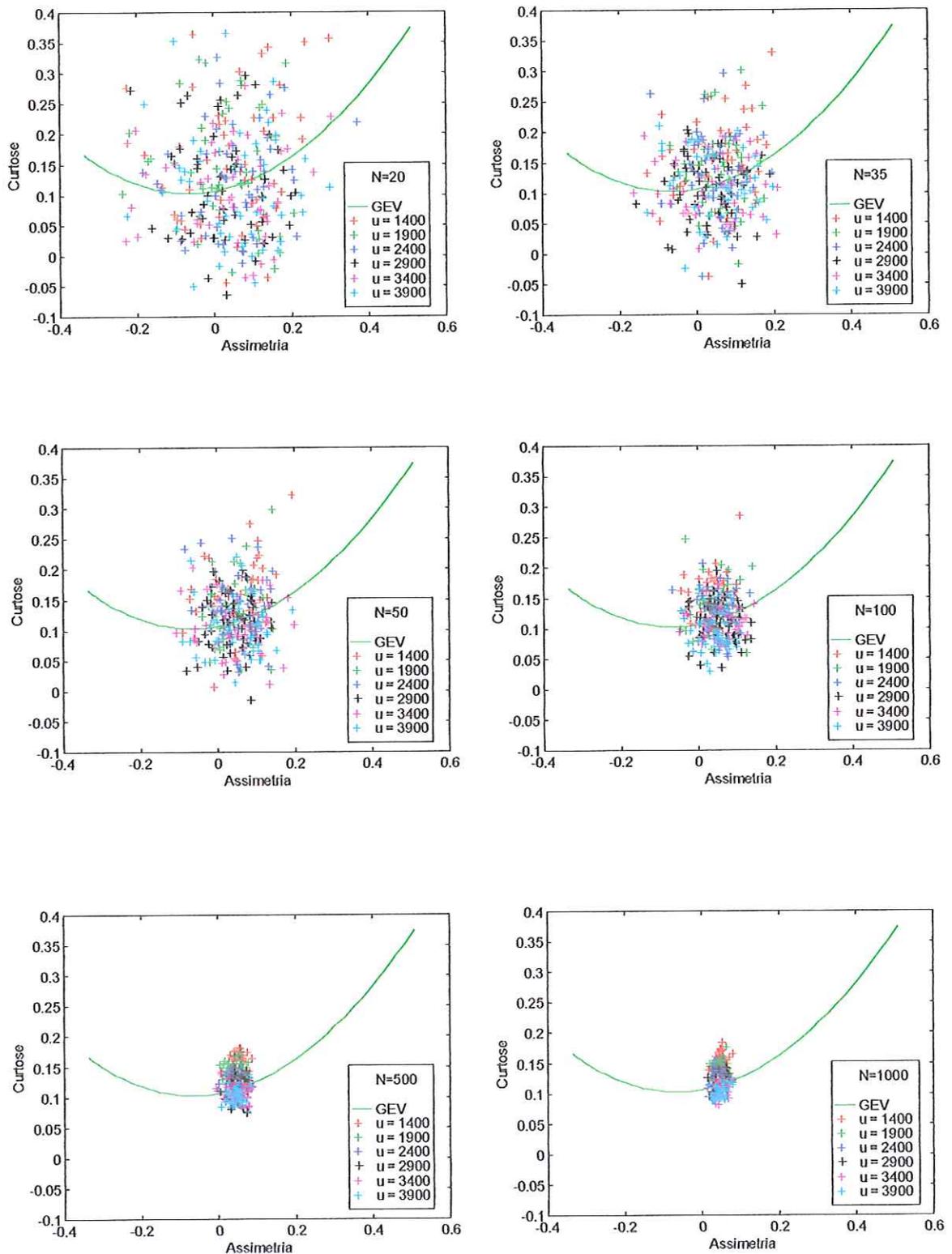


FIGURA 32 - Diagramas de momentos L, referentes aos ajustes da GEV às séries de vazões máximas com vários tamanhos, obtidas de vazões diárias geradas com diferentes valores de  $u$ ,  $\alpha = 120$  e  $k = 0,20$ .

resultados das relações entre a assimetria e curtose, apresentados nos diagramas de momentos L, referentes às 50 séries sintéticas de cheias anuais provenientes da simulação de vazões diárias com as diferentes combinações de  $k$ ,  $\alpha$  e  $u$  reais.

Os ajustes referentes às 50 séries de vazões máximas, obtidas de vazões diárias simuladas com diferentes valores de  $k$ ,  $\alpha = 120$  e  $u = 3000$ ; diversos valores de  $\alpha$ ,  $k = 0,2$  e  $u = 3000$  e vários valores de  $u$ ,  $\alpha = 120$  e  $k = 0,2$ , são avaliados, conjuntamente, para cada tamanho de amostra, através dos diagramas de momentos L das Figuras 30, 31 e 32, respectivamente.

A Figura 29a mostra que amostras de cheias anuais com assimetria em torno de 0,17 advém da distribuição Gumbel. Amostras com assimetria superior a 0,17 são descritas através da distribuição de valores extremos ilimitada (GEV com  $k < 0$ ). Enquanto assimetrias amostrais inferiores a 0,17 são decorrentes de amostras de populações que seguem a GEV limitada ( $k > 0$ ). Nota-se que as assimetrias teóricas referentes aos seis valores reais de  $k$  variam entre  $-0,06$  ( $k = 0,4$ ) e  $0,17$  ( $k \cong 0$ ).

A assimetria das séries de cheias anuais observadas em rios de diferentes locais do mundo variam entre  $-0,1$  e  $0,7$  e as correspondentes curtoses entre  $-0,01$  e  $0,5$  (Figura 29b). Tais séries possuem comprimentos que não ultrapassam 75 – 80 anos de registros.

A influência do tamanho da amostra na qualidade do ajuste fica bem caracterizada através dos diagramas de momentos L apresentados nas Figuras 30, 31 e 32. Em amostras curtas, com  $N \leq 100$ , a dispersão da curtose amostral em torno da curva teórica resulta da ocorrência das três formas de distribuições de valores extremos, quando esperava-se, apenas, a forma da GEV limitada, com considerável afastamento das distribuições da respectiva distribuição teórica. Todavia, quando o tamanho da amostra torna-se grande o suficiente para assegurar a convergência assintótica à forma de valores extremos, os dados estimados de assimetria e curtose aproximam-se de seus verdadeiros valores, como acontece com  $N = 1000$ .

Observa-se através das Figuras 30, 31 e 32, que a dispersão dos dados de  $\tau_4$  referentes às amostras mais curtas de cheias anuais, provenientes das séries com 1000 anos, é similar àquela apresentada pelos dados da Figura 29b, que diz respeito às séries de cheias anuais, cuja extensão não ultrapassam 80 anos de observações.

Assim, é possível imaginar que a distribuição da população, da qual foi retirada cada amostra de vazões máximas observadas nos diversos rios, pode muito bem ser descrita por uma das distribuições assintóticas de valores extremos, definida através de umas das combinações de parâmetros com  $k$  variando de 0.0001 até 0,4, utilizadas na simulação de vazões diárias. Este fato reforça a idéia de convergência assintótica dos eventos máximos de variáveis hidrológicas, como é o caso das vazões de rios.

#### 6.1.5 Posição de Plotagem dos Quantis de Cheias

As Figuras 33a e 33b resumem o efeito das diferentes fórmulas de plotagem dos quantis de cheias, em função da variável reduzida de Gumbel, verificado através da plotagem das várias séries de cheias anuais utilizadas neste trabalho. O referido efeito está associado à probabilidade de excedência ou ao período de retorno do evento em questão. A fórmula de posição de plotagem não participa do processo de estimação de parâmetros da distribuição adotada neste estudo e, por sua vez, não intervém na qualidade do ajuste da distribuição.

A posição de plotagem, através de cada fórmula proposta, busca a melhor adequação, numa apresentação gráfica, de um ajuste de distribuição juntamente com os dados utilizados, na mesma faixa de frequência das observações. No presente estudo, a posição de plotagem dos quantis de cheias, feita através das várias fórmulas, diferem entre si, apenas, através dos últimos 3 ou 4 maiores valores (Figuras 33a e 33b). No caso dos modelos de CUNANE e de BLOM, as posições de plotagem assumidas pelos dados foram visualmente coincidentes e, por isso, são apresentados em um único gráfico.

Observa-se que, quando as séries de dados apresentam valores atípicos máximos, a fórmula de posição de plotagem APL apresenta melhor adequação gráfica da distribuição em relação aos dados, principalmente, porque esta fórmula associa ao valor do evento mais elevado uma probabilidade menor do que as demais fórmulas. Isto significa uma posição mais avançada da plotagem do dado mais elevado, em relação a posição determinada pelas outras fórmulas, como pode ser observado nas Figuras 33a e 33b.

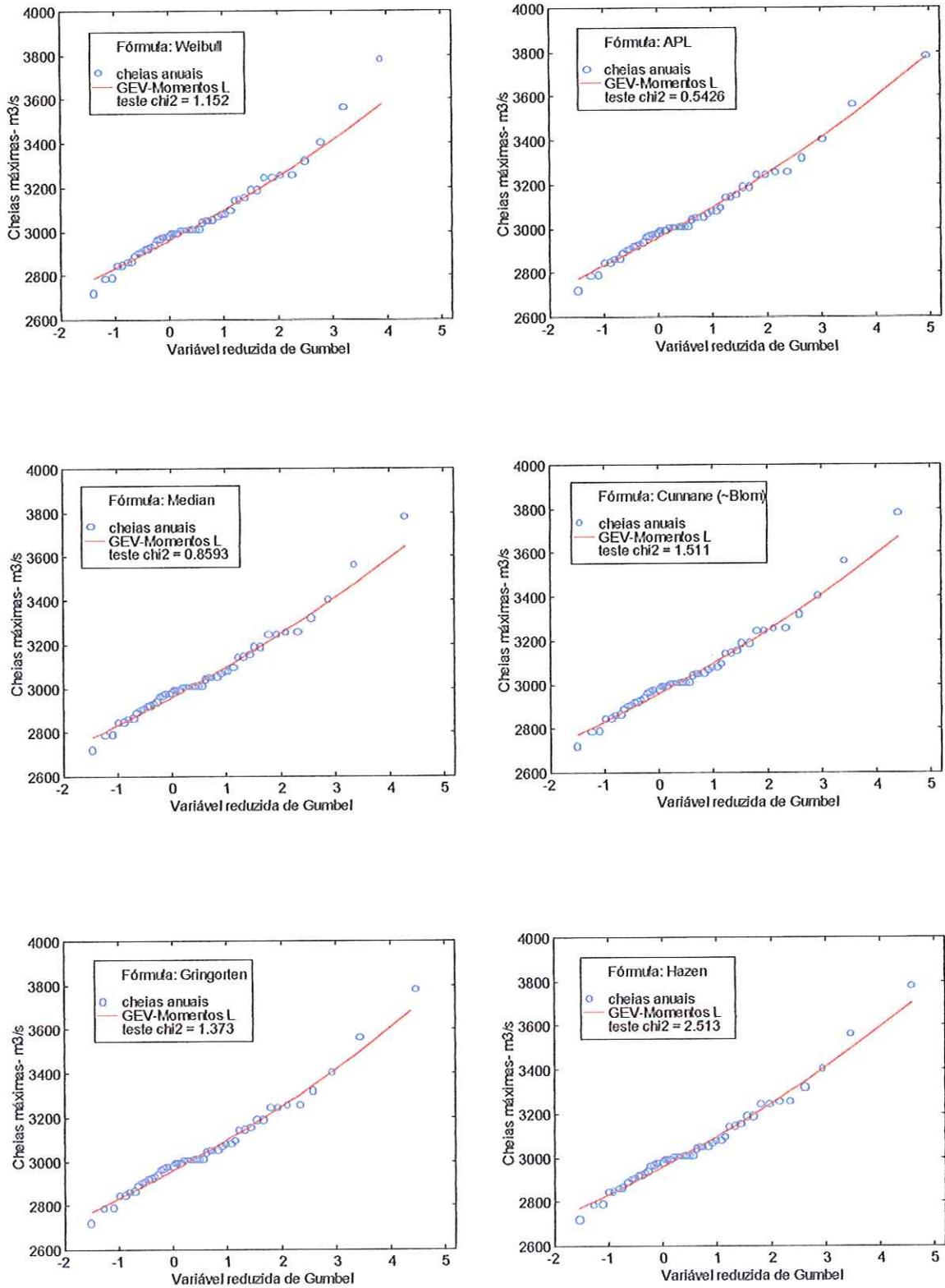


FIGURA 33a - Ajuste da GEV à série de vazões máximas anuais (N = 50) utilizando diferentes fórmulas de Posição de Plotagem.

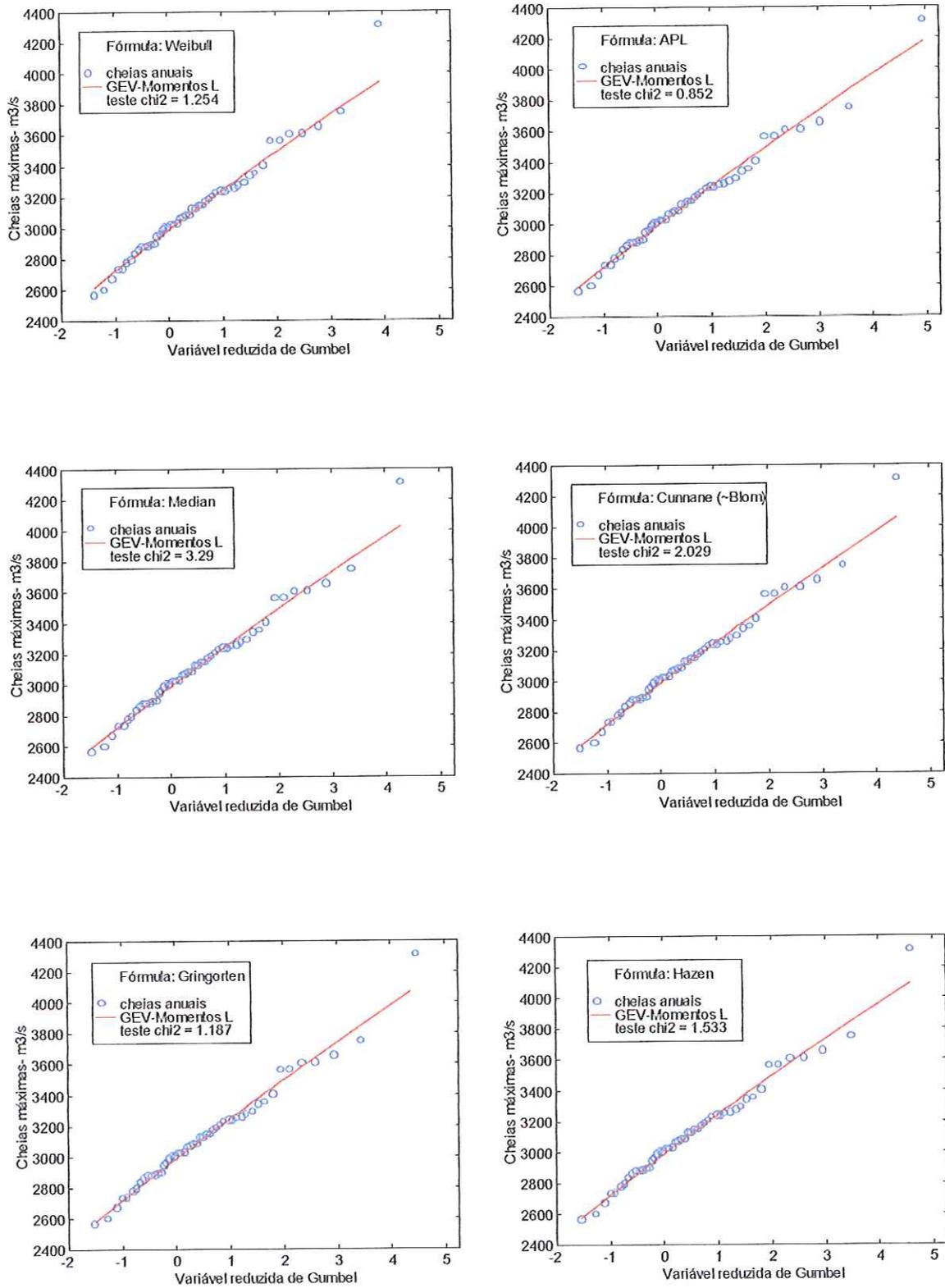


FIGURA 33b - Ajuste da GEV à série de vazões máximas anuais (N = 50) utilizando diferentes fórmulas de Posição de Plotagem.

A fórmula de WEIBULL apresenta os maiores afastamentos entre os dois dados de cheias mais elevados da série e a distribuição teórica, quando esses dados se apresentam como atípicos. Neste caso, as fórmulas de GRINGOTEN e de HAZEN são as que mais aproximam os dados da distribuição, depois da APL. Embora as mesmas, apresentem resultados bem aproximados aos das fórmulas MEDIAN e de CUNNANE. Em geral, quando a fórmula de WEIBULL consegue plotar os dados adequadamente, as outras fórmulas também apresentam boa adequação.

#### 6.1.6 Efeito do tamanho da Amostra na Forma da Distribuição

Nas Figuras de 34 a 49, apresentam-se os resultados dos ajustes da distribuição GEV às séries com 20, 35, 50, 100, 500 e 1000 anos, resultantes da simulação de vazões diárias com os seis diferentes valores reais de  $k$ , respectivamente, mantido  $\alpha = 120$  e  $u = 3000$ . Para estimação da distribuição foi utilizado momentos LH, cuja ordem foi definida em função do melhor ajuste, segundo o teste de hipótese proposto por WANG (1998). Estes resultados estão exibidos na legenda de cada gráfico das referidas figuras.

As várias Figuras ilustram, de forma gráfica, todos os aspectos discutidos antes. Verifica-se a ocorrência das três formas de distribuição de valores extremos, para amostras com menos de 100 anos. Em amostras curtas, mesmo quando a modelação resulta em  $k > 0$ , a distribuição ajustada apresenta grande distanciamento da distribuição teórica. Este efeito é mais marcante em amostras de populações descritas através da distribuição Gumbel ( $k=0$ ), devido a pobreza de informações sobre o parâmetro de forma, para este tipo de distribuição, observada nas amostras, sobre as quais ajustou-se a distribuição GEV que possui o parâmetro  $k$ .

Amostras provenientes de populações que seguem a distribuição GEV limitada com parâmetro de forma  $k > 0,2$ , apresentaram resultados de quantis de cheias bem mais próximo aos valores da distribuição verdadeira, mesmo quando as amostras têm menos de 100 anos.

Há ocorrência de outliers em amostras mais curtas ( $N \leq 50$ ) e  $k_{real} < 0,2$  (Figuras 34 a 36), os quais foram considerados durante o processo de ajuste da GEV. Esses valores atípicos, presentes nas séries mais curtas, perdem esta característica

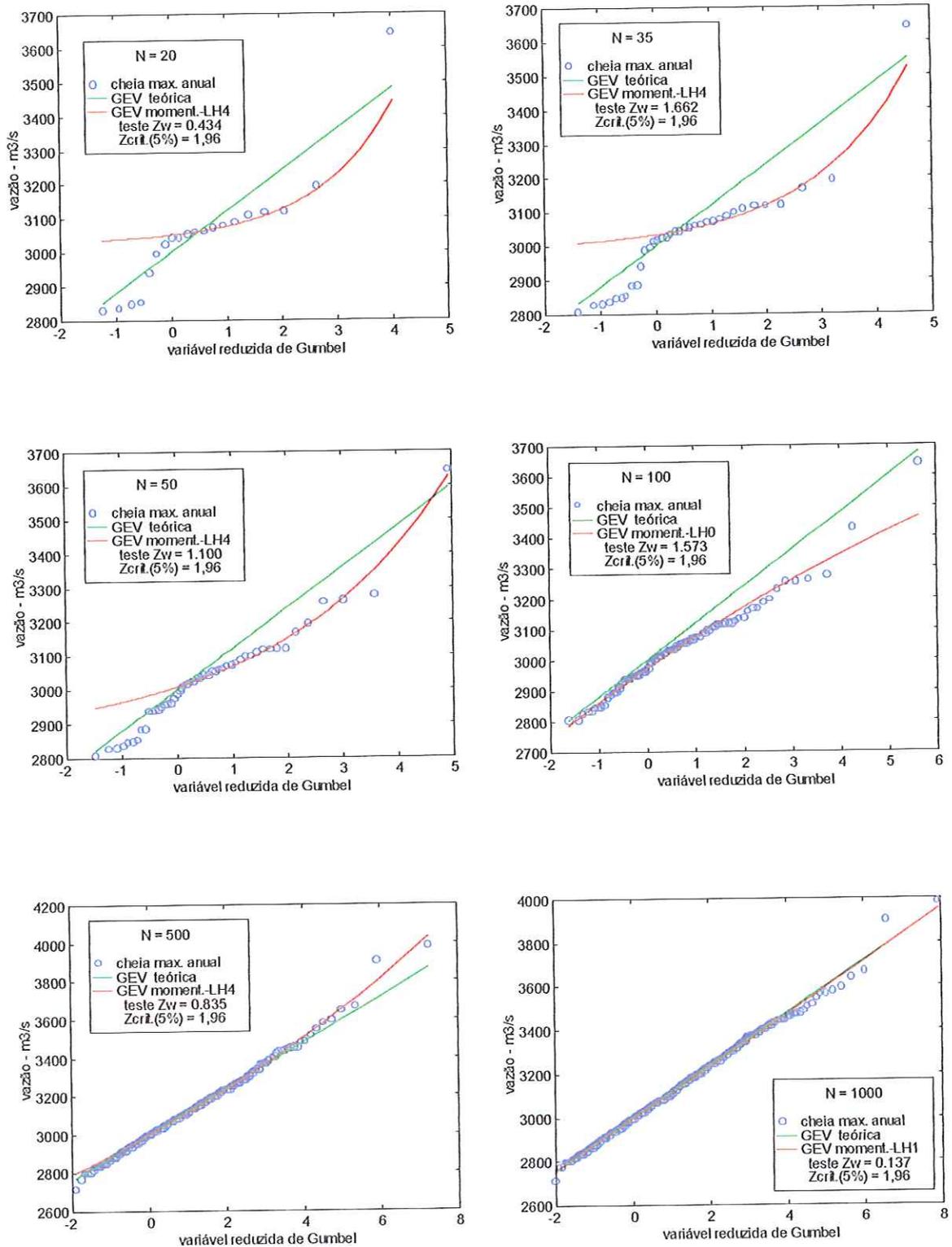


FIGURA 34 - Distribuições GEV teórica e ajustada através de momentos LH, usando cheias máximas anuais obtidas de série de vazões diárias geradas com  $k = 0,0001$ ,  $\alpha = 120$  e  $u = 3000$  (posição de plotagem APL).

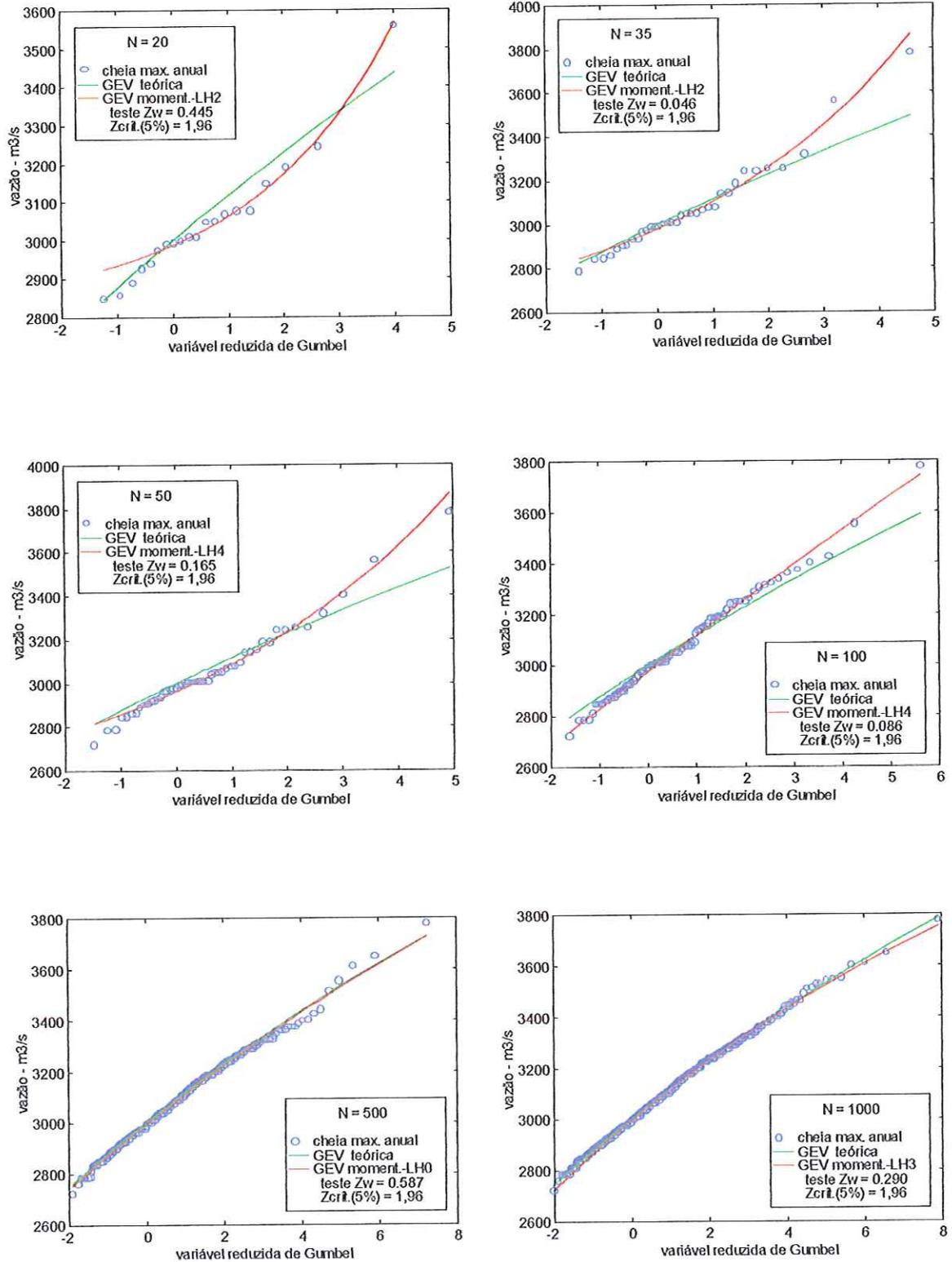


FIGURA 35 - Distribuições GEV teórica e ajustada através de momentos LH, usando cheias máximas anuais obtidas de série de vazões diárias geradas com  $k = 0,05$ ,  $\alpha = 120$  e  $u = 3000$  (posição de plotagem APL).

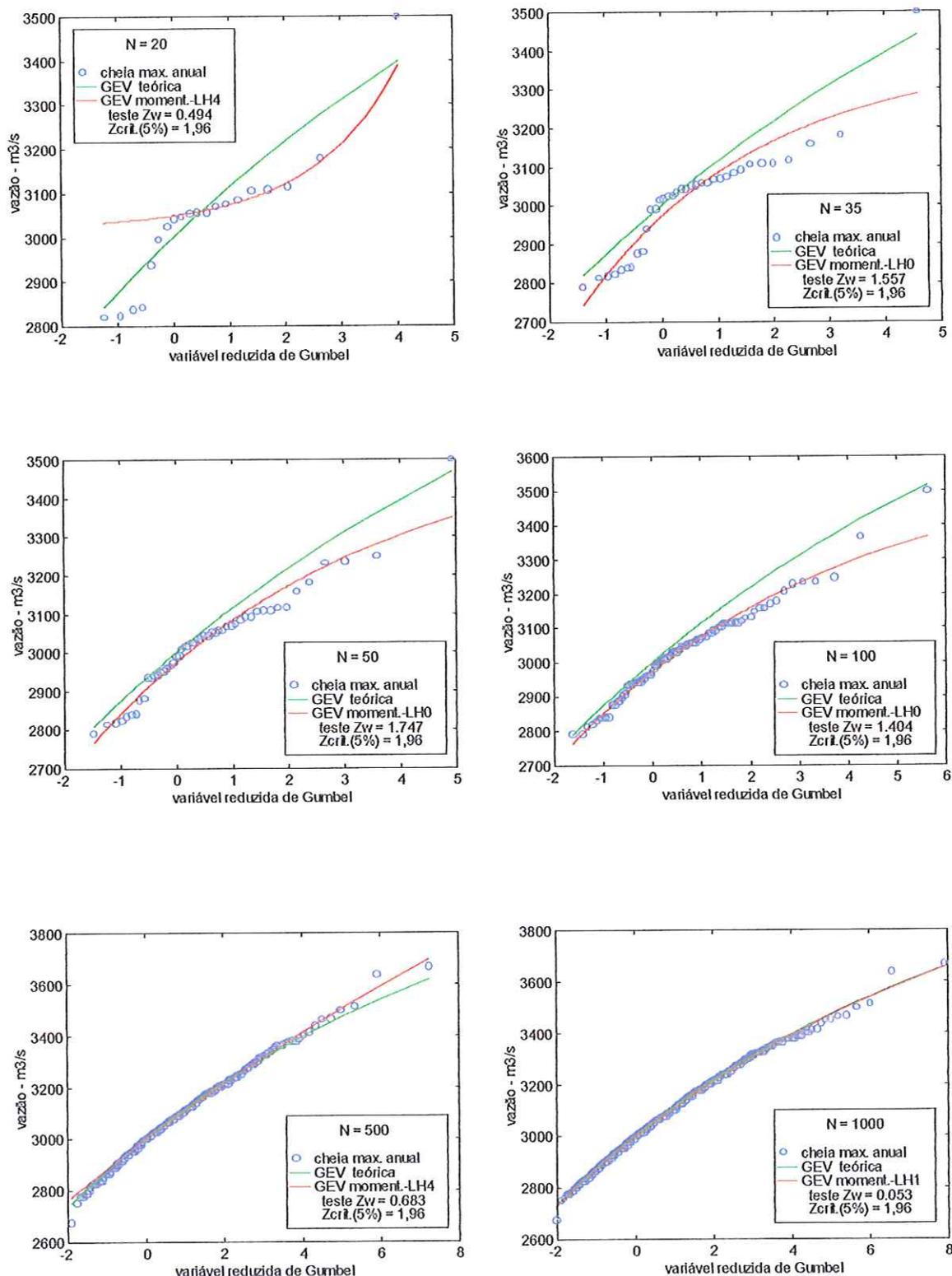


FIGURA 36 - Distribuições GEV teórica e ajustada através de momentos LH, usando cheias máximas anuais obtidas de série de vazões diárias geradas com  $k = 0,10$ ,  $\alpha = 120$  e  $u = 3000$  (posição de plotagem APL).

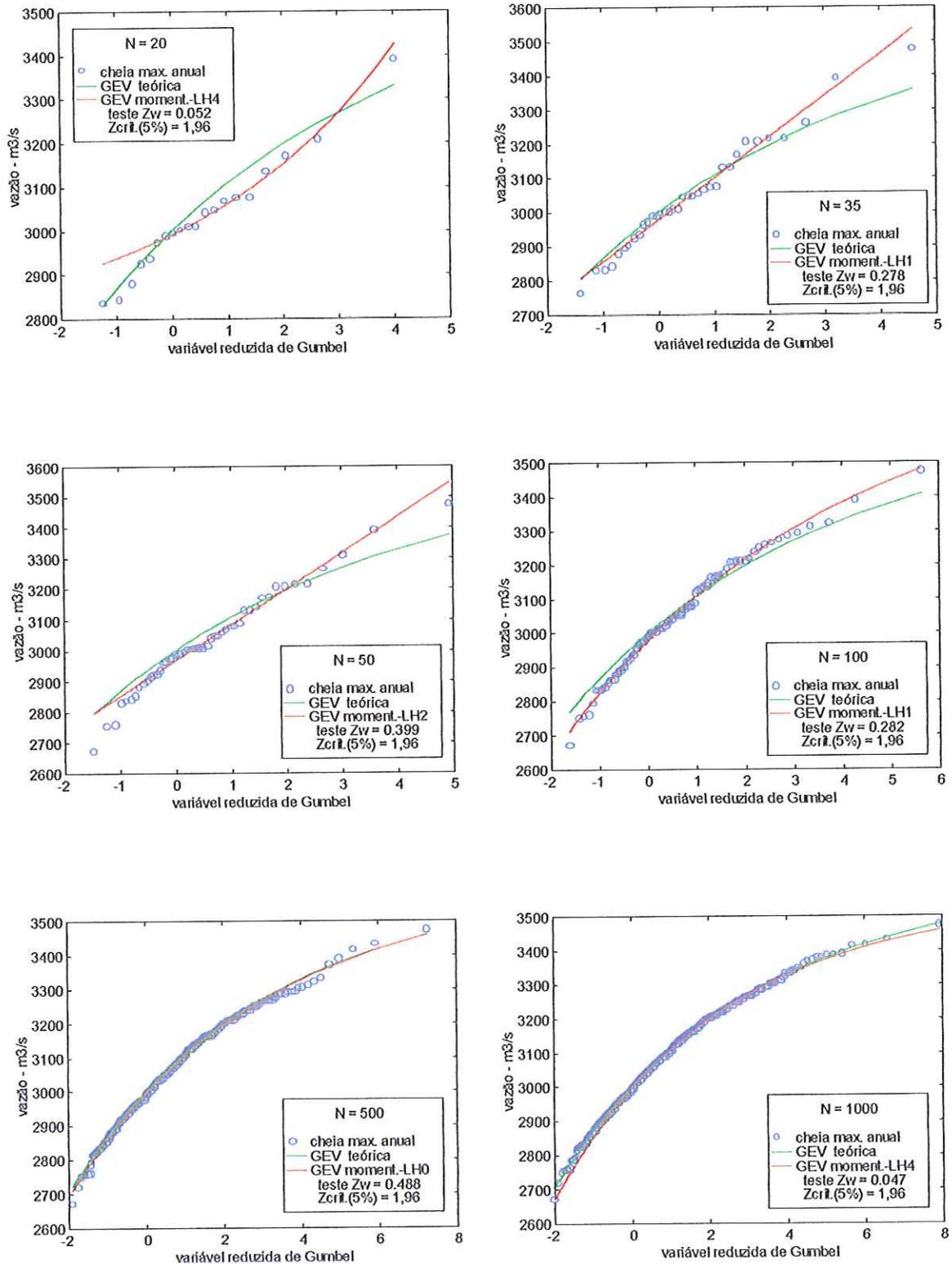


FIGURA 37 - Distribuições GEV teórica e ajustada através de momentos LH, usando cheias máximas anuais obtidas de série de vazões diárias geradas com  $k = 0,20$ ,  $\alpha = 120$  e  $u = 3000$  (posição de plotagem APL).

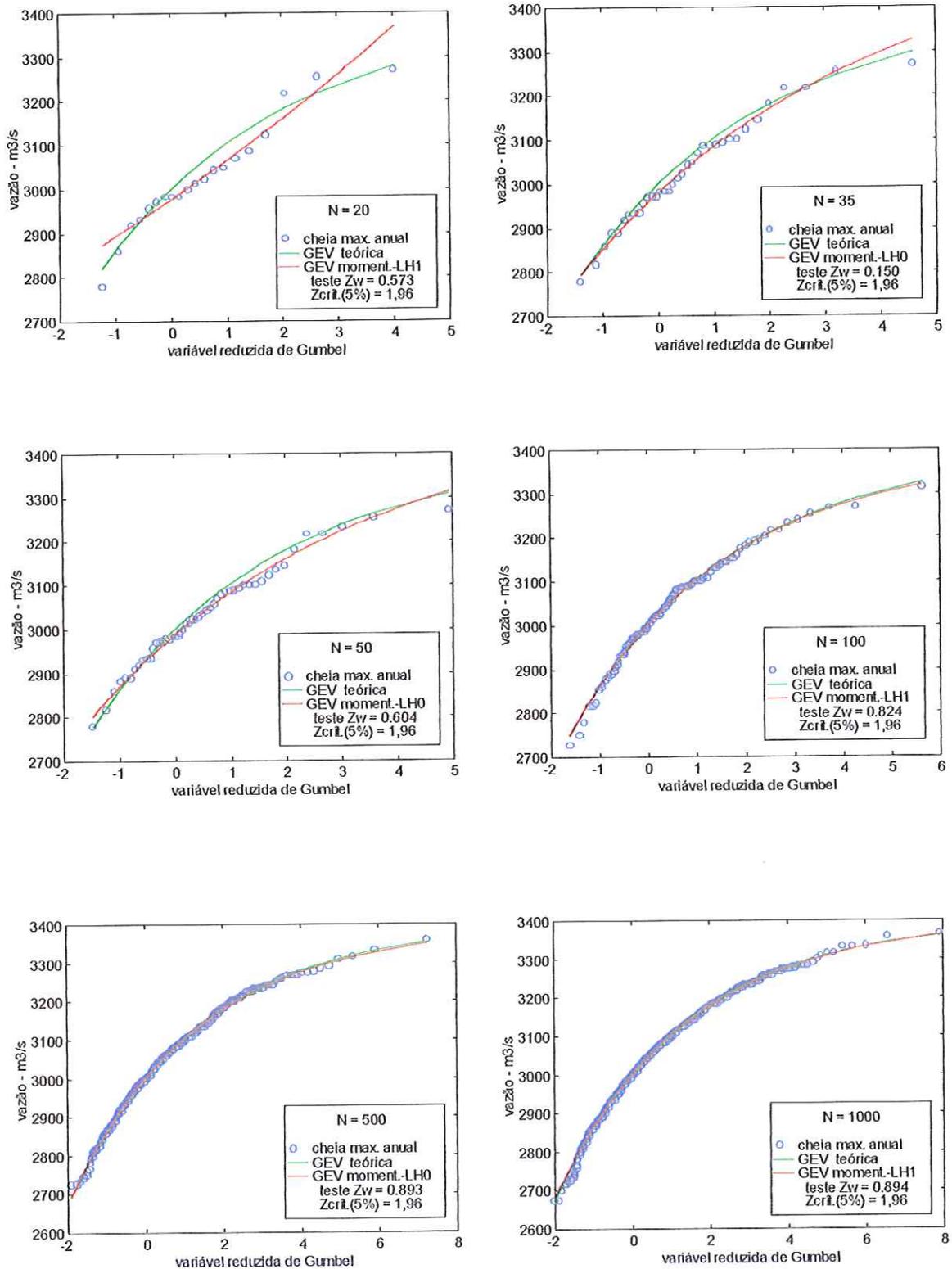


FIGURA 38 - Distribuições GEV teórica e ajustada através de momentos LH, usando cheias máximas anuais obtidas de série de vazões diárias geradas com  $k = 0,30$ ,  $\alpha = 120$  e  $u = 3000$  (posição de plotagem APL).

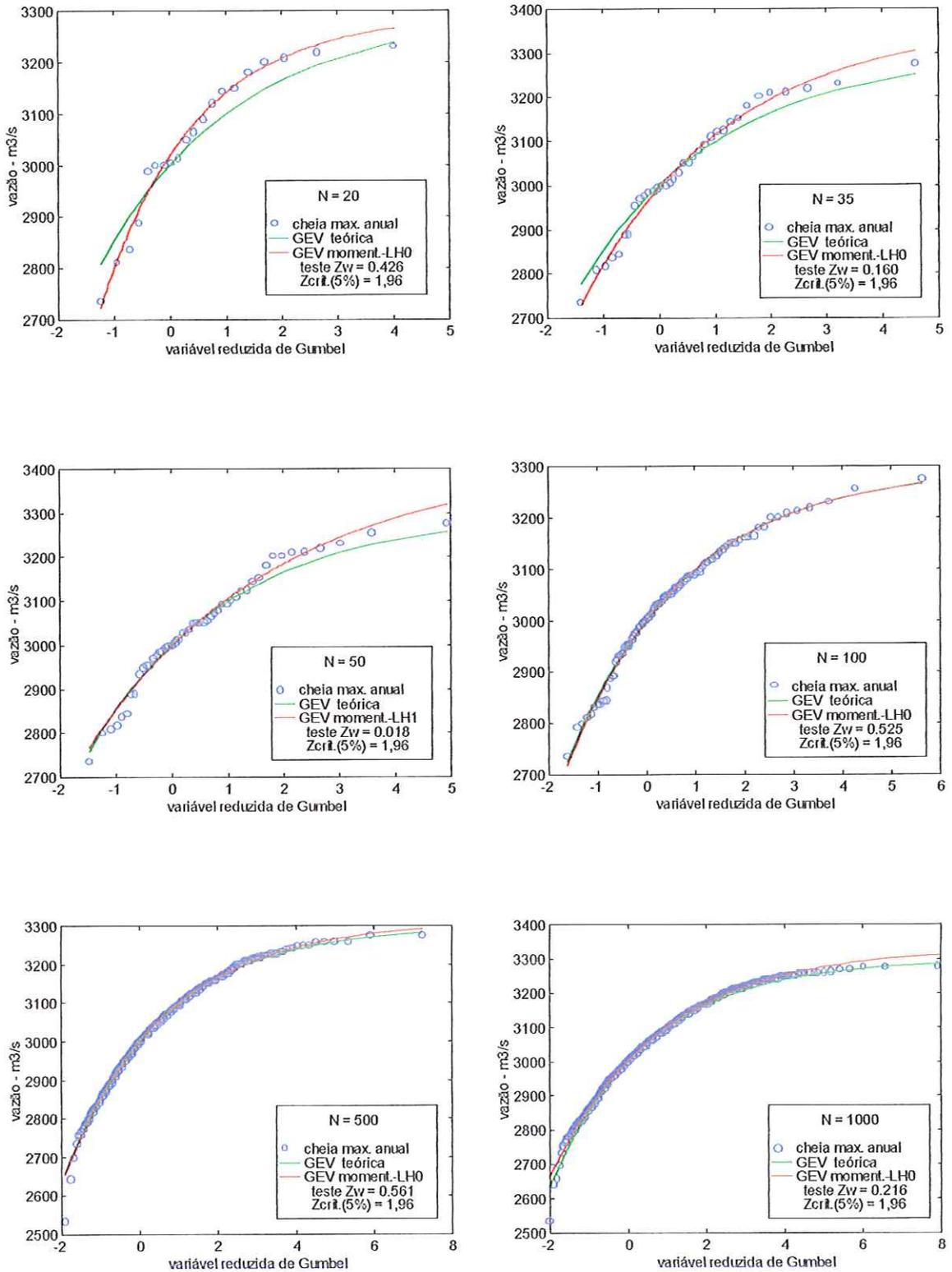


FIGURA 39 - Distribuições GEV teórica e ajustada através de momentos LH, usando cheias máximas anuais obtidas de série de vazões diárias geradas com  $k = 0,40$ ,  $\alpha = 120$  e  $u = 3000$  (posição de plotagem APL).

quando mais dados são incorporados à amostra. Assim, um dado valor, apresenta-se como outlier nas séries com os primeiros 20 ou 35 valores, porém, quando consideram os primeiros 50 ou 100 valores, o mesmo deixou de ser um valor atípicos, em decorrência da nova posição de plotagem assumida.

Torna-se evidente, que o problema do outlier está relacionado a sua posição de plotagem, a qual é desfavorável, quando a amostra corresponde a um período reduzido. Outra observação que merece destaque, é a boa adequação dos ajustes aos dados utilizados, evidenciado nas citadas figuras, obtida através do uso da fórmula APL no cálculo das posições de plotagens.

Todo os ajustes realizados em amostras com  $N > 500$  anos, apresentam resultados similares aos da distribuição verdadeira. Isto mostra que o efeito do tamanho da amostra na qualidade do ajuste torna-se insignificante, quando a quantidade de dados observados for suficientemente grande, com mais de 500 anos. Neste caso, ocorre a condição de convergência assintótica, em que as incertezas dos ajustes são desprezíveis.

## 6.2 MODELOS DE VAZÕES DIÁRIAS USADOS NA HIDROLOGIA OPERACIONAL

Nesta seção, são apresentadas e discutidas as distribuições dos valores de vazões máximas obtidas de modelagem estocástica de vazões diárias através de 4 modelos disponíveis na literatura científica. Apresenta-se, também, um novo modelo de geração de vazões diárias, sugerido neste trabalho, juntamente com a análise de desempenho do mesmo.

### 6.2.1 Forma Assintótica das Vazões Máximas Anuais Geradas Através de Modelos Estocásticos Disponíveis na Literatura

Com os modelos diários Diana, Shot-Noise de Segunda Ordem, Modelo Simplificado de Desagregação e Modelo de Sargent modelaram-se os dados de vazões históricas diárias observadas no rio Paraná, que apresentam média diária de 9410,3 m<sup>3</sup>/s e desvio padrão de 4615,4 m<sup>3</sup>/s. Porém, para verificar o comportamento das distribuições de vazões máximas com magnitudes menores, modelaram-se os

dados de vazões do rio dos Patos, que possui média diária de  $20,57 \text{ m}^3/\text{s}$  com desvio padrão de  $34,00 \text{ m}^3/\text{s}$ .

As séries de vazões máximas anuais com 1000 valores, obtidas com aplicação dos 4 modelos nos dois rios, foram ajustadas através da distribuição GEV. Os ajustes foram avaliados com a aplicação do teste de WANG (1998) e a forma das distribuições ajustadas foi verificada através do teste de HOSKING et al. (1985).

As distribuições de vazões máximas histórica e ajustadas, referentes aos quatro modelos diários aplicados aos dois rios, estão mostrados na Figura 40.

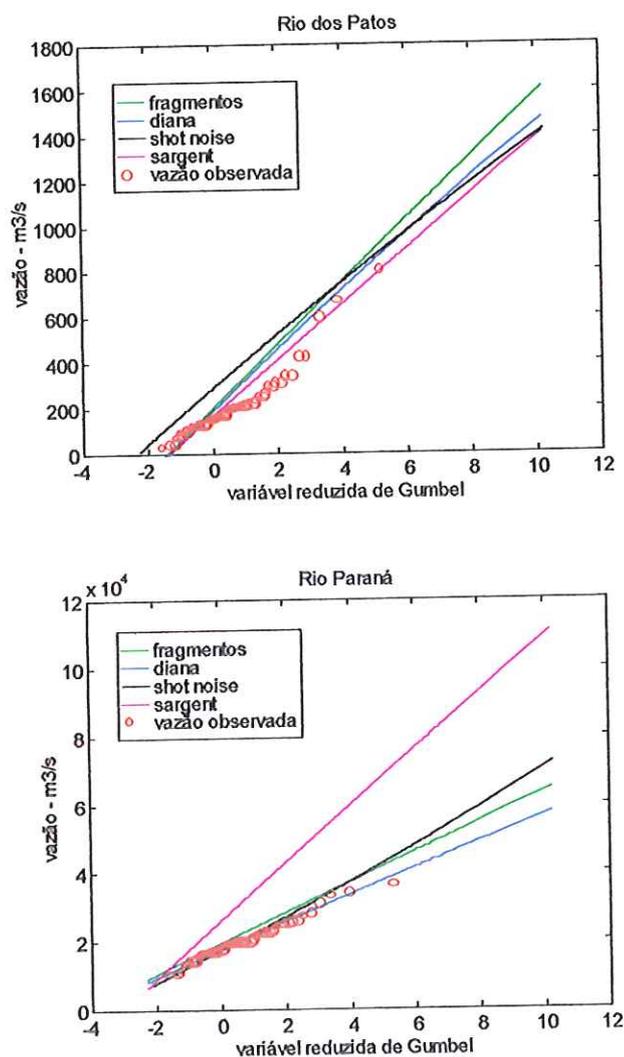


FIGURA 40 - Ajuste da GEV às séries de vazões máximas geradas através dos modelos de vazões diárias: Diana, Shot noise, Fragmentação e Sargent ajustados às vazões do rio dos Patos e rio Paraná.

Enquanto, os resultados das estimativas dos parâmetros  $k$ ,  $\alpha$  e  $u$ , juntamente com os resultados dos testes aplicados, encontram-se nas Tabelas 7.

TABELA 7 - Parâmetros da GEV e valores dos testes de qualidade de ajuste e forma da distribuição ajustada às séries de vazões máximas geradas através dos modelos: Diana, Shot noise, Fragmentação e Sargent.

Modelos	Distribuição GEV			Testes	
	$K$	$\alpha$	$u$	$Z_{Wang}$	$Z_{Hosking}$
Rio Paraná					
Fragmentos	0,003	4480	19510	0,0071*	0,0446*
Diana	0,009	4078	17660	0,3615*	0,4016*
Shot-Noise	-0,021	4793	17400	0,3688*	0,2927*
Sargent	0.005	5483	20960	0.4500*	0.1012*
Rio dos Patos					
Fragmentos	0,004	140,3	198,9	0,921*	0,1589*
Diana	0,019	138,1	186,8	0,410*	0,8282*
Shot-Noise	0,017	120,0	289,7	0,516*	0,7333*
Sargent	0.0018	122,5	166,1	0.198*	0.0137*

\* Significativo ao nível de 5% ( $Z_{0,025}=1,96$ )

A boa adequação dos ajustes da GEV às séries sintéticas de dados é verificada através dos resultados do teste de WANG, mostrados nas duas tabelas. Assim como, os resultados do teste de HOSKING indicam evidências amostrais de aceitação da hipótese de que o parâmetro de forma  $k$  é igual a zero, para todas distribuições testadas, com nível de significância estatística de 5 %. Isto indica que os valores de vazões máximas anuais das séries de vazões diárias geradas pelos 4 modelos apresentam convergência assintótica para a forma VEI (distribuição Gumbel), com chances de mais de 95 % de probabilidade disto acontecer.

Sobre os ajustes realizados com os 4 modelos, o único interesse foi o de verificar a forma das distribuições de seus valores extremos máximos, portanto, não foi avaliado qual dos modelos ajustou-se melhor aos dados de cada rio.

### 6.2.2 Novo Procedimento de Geração de Vazões Diárias

O novo método de gerações de vazões diária que está sendo proposto, considera que os dados de pulsos de vazões de cada mês seguem distribuição de probabilidade limitada. Assim, espera-se que em modelagem estocástica de vazões diárias com base na geração de pulsos de vazões, a utilização da distribuição de probabilidade limitada como gerador de alturas de pulsos, reproduzirá vazões diárias com valores máximos anuais convergindo assintoticamente à forma VEIII.

Dada a simplicidade estrutural do modelo de SARGENT (1979), onde a parte que modela as recessões é ajustada, separadamente, da parte que simula as alturas de pulsos, o referido método foi utilizado como base para estruturação do novo modelo de geração de vazões diárias devido à facilidade de adequação do novo procedimento de geração de alturas de pulsos de vazões.

Logo, a forma de ajuste e modelagem das recessões e o procedimento de construção das séries de pulsos históricos e da geração de vazões do novo modelo, são idênticos aos do modelo de SARGENT. Enquanto, para geração de pulsos, utiliza-se a função distribuição de probabilidade limitada ajustada aos dados de pulsos históricos de cada mês, através do método de momentos de pesos probabilísticos, considerando a correlação serial dos seus valores nos períodos de ocorrências consecutivas de pulsos durante o processo de simulação.

Para o ajuste e avaliação do modelo proposto foram utilizados os dados de vazões diárias do rio Paraná, por apresentar uma série de considerável tamanho, com 73 anos de registros. No processo de geração de vazões diárias as alturas de pulsos de vazões foram simulada utilizando, além do modelo de distribuição limitada, o modelo exponencial e o modelo exponencial modificado para análise comparativa.

#### a) *Probabilidades de transição*

A partir das séries históricas de vazões diárias determinaram-se as probabilidades de transição históricas de um dia seco ser seguido de outro dia seco; de um dia seco ser seguido de um dia chuvoso; de um dia chuvoso ser seguido de um dia seco e de um dia chuvoso ser seguido por outro chuvoso. Os resultados encontram-se nas Tabelas 8a e 8b.

TABELA 8a - Probabilidades de transição histórica de um dia seco ser seguido por outro seco e de um dia seco ser seguido por um dia chuvoso.

Mês	Probabilidade dia seco – dia seco			Probabilidade dia seco – dia chuvoso		
	Média	Mínima	Máxima	Média	Mínima	Máxima
jan.	0,7825	0,4167	0,9524	0,2175	0,0476	0,5833
fev.	0,8058	0,0000	0,9583	0,1942	0,0417	1,0000
mar.	0,8393	0,5000	0,9655	0,1607	0,0345	0,5000
abr.	0,8707	0,5833	1,0000	0,1293	0,0000	0,4167
maio	0,8594	0,5385	1,0000	0,1406	0,0000	0,4615
jun.	0,8397	0,5714	1,0000	0,1603	0,0000	0,4286
jul.	0,8704	0,4167	1,0000	0,1296	0,0000	0,5833
ago.	0,8635	0,5625	1,0000	0,1365	0,0000	0,4375
set.	0,7945	0,5000	1,0000	0,2055	0,0000	0,5000
out.	0,7871	0,4545	0,9643	0,2129	0,0357	0,5455
nov.	0,7884	0,4167	1,0000	0,2116	0,0000	0,5833
dez.	0,7325	0,2857	0,9474	0,2675	0,0526	0,7143

TABELA 8b - Probabilidades de transição histórica de um dia chuvoso ser seguido por outro chuvoso e de um dia chuvoso ser seguido por um dia seco.

Mês	Prob. dia chuvoso – dia chuvoso			Probabilidade dia chuvoso - dia seco		
	Média	Mínima	Máxima	Média	Mínima	Máxima
jan.	0,8310	0,5000	1,0000	0,1690	0,0000	0,5000
fev.	0,7821	0,2000	1,0000	0,2179	0,0000	0,8000
mar.	0,7390	0,0000	0,9524	0,2610	0,0476	1,0000
abr.	0,5636	0,0000	0,9231	0,3816	0,0000	1,0000
maio	0,5805	0,0000	0,9167	0,3373	0,0000	1,0000
jun.	0,6373	0,0000	0,9333	0,3217	0,0000	1,0000
jul.	0,4908	0,0000	1,0000	0,3859	0,0000	1,0000
ago.	0,5138	0,0000	1,0000	0,4040	0,0000	1,0000
set.	0,6130	0,0000	0,9333	0,3733	0,0000	1,0000
out.	0,7345	0,2857	1,0000	0,2655	0,0000	0,7143
nov.	0,7336	0,0000	0,9500	0,2527	0,0000	0,6667
dez.	0,7709	0,2500	1,0000	0,2291	0,0000	0,7500

b) *Recessão Alta e Recessão Baixa*

As vazões diárias nos períodos de recessões históricas ocorridos em cada mês, durante 73 anos de registros, foram identificadas e diferenciadas em recessão alta e recessão baixa, quando era o caso. As vazões de cada período de recessão foram submetidas às análises de regressões lineares, após as devidas transformações logarítmicas dos seus valores, obtendo-se, assim, os parâmetros dos modelos polinomial da recessão alta ( $B_1$ ) e exponencial da recessão baixa ( $a$  e  $b$ ) por meio das equações (115) e (116), respectivamente. A média dos valores dos parâmetro obtidos para cada mês, durante os 73 anos, encontram-se na Tabela 9.

Observa-se que, nos 73 meses de janeiro, as recessões ocorreram de forma mais suave, caracterizando-se como recessão baixa. Por isso, a Tabela 9 não apresenta o valor do parâmetro  $B_1$ , do modelo polinomial da recessão alta, no mês de janeiro.

TABELA 9 - Parâmetros dos modelos de recessão da vazão diária.

Mês	Recessão Alta		Recessão Baixa		
	$B_1$	$r^2$	$a$	$b$	$r^2$
jan.	-	-	0,0070	-0,0247	0,9419
fev.	-0,2050	0,9721	0,0055	-0,0198	0,9832
mar.	-0,2236	0,9892	0,0048	-0,0210	0,9509
abr.	-0,2476	0,9903	0,0062	-0,0250	0,9712
maio	-0,1731	0,9910	-0,0021	-0,0213	0,9673
jun.	-0,2395	0,9705	-0,0012	-0,0253	0,9654
jul.	-0,1610	0,9885	-0,0117	-0,0216	0,9856
ago.	-3,2690	0,9890	-0,0020	-0,0229	0,9812
set.	-0,1709	0,9911	0,0015	-0,0239	0,9803
out.	-0,2912	0,9510	0,0017	-0,0243	0,9669
nov.	-0,2308	0,9659	0,0055	-0,0236	0,9757
dez.	-0,2035	0,9798	0,0011	-0,0222	0,9645

$r^2$  – coeficiente de determinação da análise de regressão linear

c) *Série Histórica de Pulsos*

A série histórica de pulsos de vazões de cada mês foi formada, conforme Figura 3, reunindo numa mesma amostra todos os valores de pulsos observados naquele mês, durante os 73 anos de registros. Os valores referentes à média, desvio padrão, coeficiente correlação e número de pulsos da série histórica de cada mês estão relacionados na Tabela 10.

TABELA 10 - Dados referente às séries históricas de pulsos de vazões de cada mês.

Mês	Média	Desvio Padrão	Coef. de Correlação	Nº Total de Pulsos
janeiro	4676	4161	0,7333	1328
fevereiro	4453	3850	0,8050	1048
março	3696	3504	0,7841	997
abril	2192	2216	0,7687	665
maio	2107	2325	0,7528	753
junho	1754	1890	0,7319	777
julho	1332	1247	0,6939	630
agosto	1178	1215	0,6715	700
setembro	1323	1327	0,7728	960
outubro	1762	1711	0,7668	1111
novembro	1990	1754	0,7734	1077
dezembro	2849	2720	0,8035	1330

Nota-se que os valores da média e desvio padrão da série de pulsos de cada mês são bem próximos, indicando aproximação com a distribuição exponencial. Observa-se, que o mês de janeiro apresenta a maior média e agosto a menor, já os maiores números de pulsos ocorreram em dezembro e janeiro e o menor em agosto.

d) *Ajustes das Distribuição Exponencial e Exponencial Modificada Para Pulsos*

Os parâmetros  $\eta$  e  $\delta$  da distribuição exponencial modificada foram estimados utilizando a média e o desvio padrão dos dados de pulsos referente a cada mês. Já o

único parâmetro da distribuição exponencial corresponde ao valor médio dos pulsos de cada mês. A estimação de  $\eta$  e  $\delta$  foi realizada através de procedimentos numéricos, utilizando-se o método da secante para resolver as equações (117a) e (117b). Os valores dos parâmetros dos dois modelos de distribuição encontram-se descritos na Tabela 11. O valor de  $\eta$  nos 12 meses encontram-se em torno de um, indicando uma aproximação dos dois modelos.

TABELA 11 - Valores dos parâmetros e testes dos ajustes das distribuições exponencial e exponencial modificada às séries históricas de pulsos de vazões.

Mês	Distribuição Exponencial Modificada			Distribuição Exponencial		
	$\eta$	$\delta$	Teste $\chi^2$	$\eta$	$\delta$	Teste $\chi^2$
jan.	0,8891	4906,8	477,99	1,0	4676	14,46*
fev.	0,8620	4729,9	527,05	1,0	4453	19,53
mar.	0,9477	3779,7	92,91	1,0	3696	10,97*
abr.	1,0109	2181,9	7,44*	1,0	2192	3,95*
maio	1,1019	2020,2	162,26	1,0	2107	14,33*
jun.	1,0766	1699,1	111,51	1,0	1754	10,40*
jul.	0,9355	1369,4	60,53	1,0	1332	8,89*
ago.	1,0312	1162,7	24,60	1,0	1178	6,78*
set.	1,0030	1321,3	5,81*	1,0	1323	4,93*
out.	0,9709	1783,9	25,97	1,0	1762	5,22*
nov.	0,8795	2092,2	392,82	1,0	1990	8,15*
dez.	0,9544	2905,0	88,65	1,0	2849	13,46*

\* Significativo ao nível de 5%  $-\chi^2_{(0,95, 8)} = 15,5$

Com os dois modelos ajustados, geraram-se quantis de pulsos de vazões na mesma faixa dos valores observados em cada mês, e com mesmo número de ocorrências, utilizando simulação Monte Carlo. Compararam-se as séries sintéticas e históricas através da aplicação do teste  $\chi^2$  (ver Tabela 11) e graficamente, como mostram as Figuras 41a e 41b.

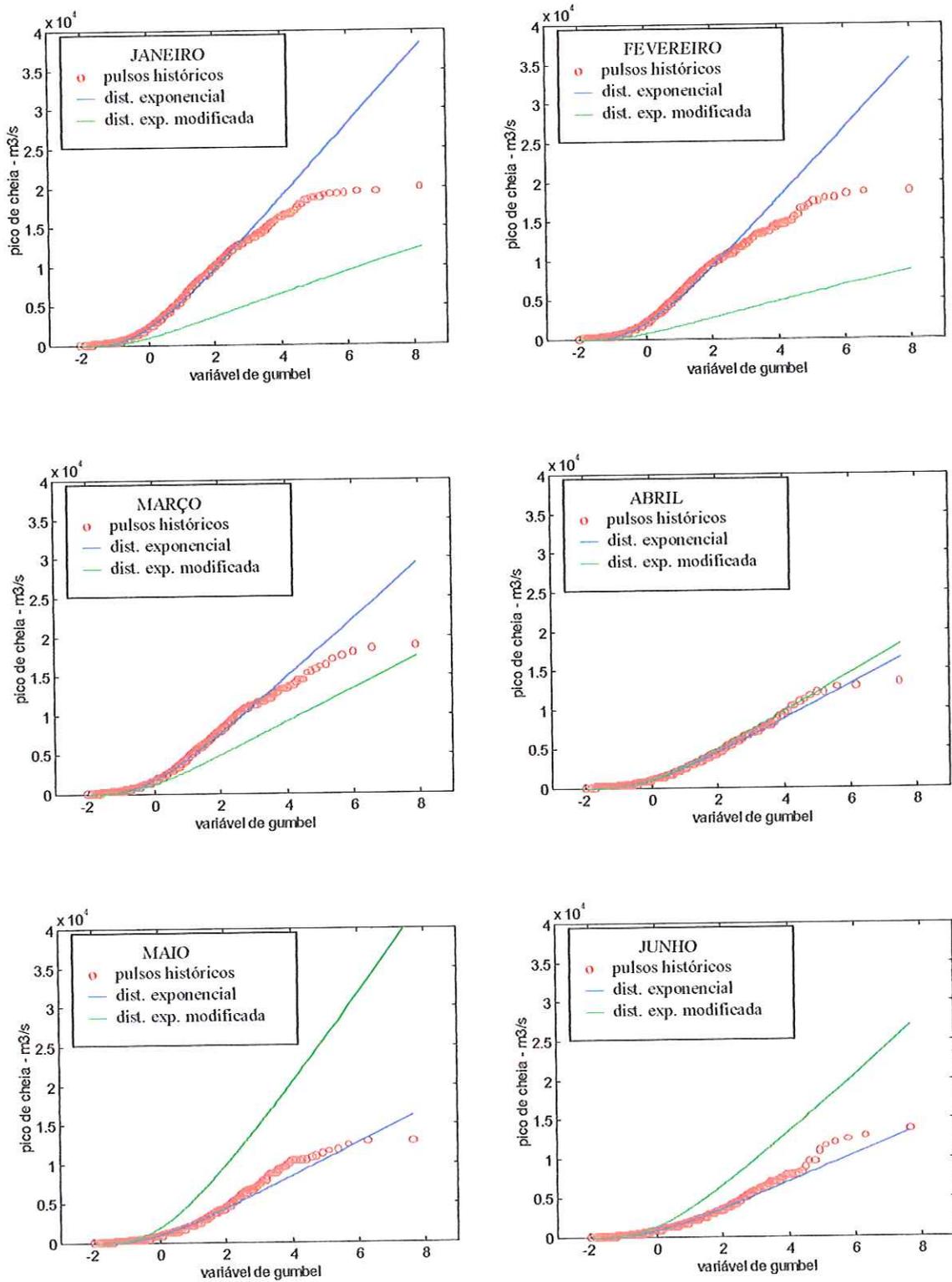


FIGURA 41a - Ajustes das distribuições de probabilidades exponencial e exponencial modificada dos pulsos de vazões simulados, ajustados à série dos pulsos observados em cada mês, durante 73 anos de registro.

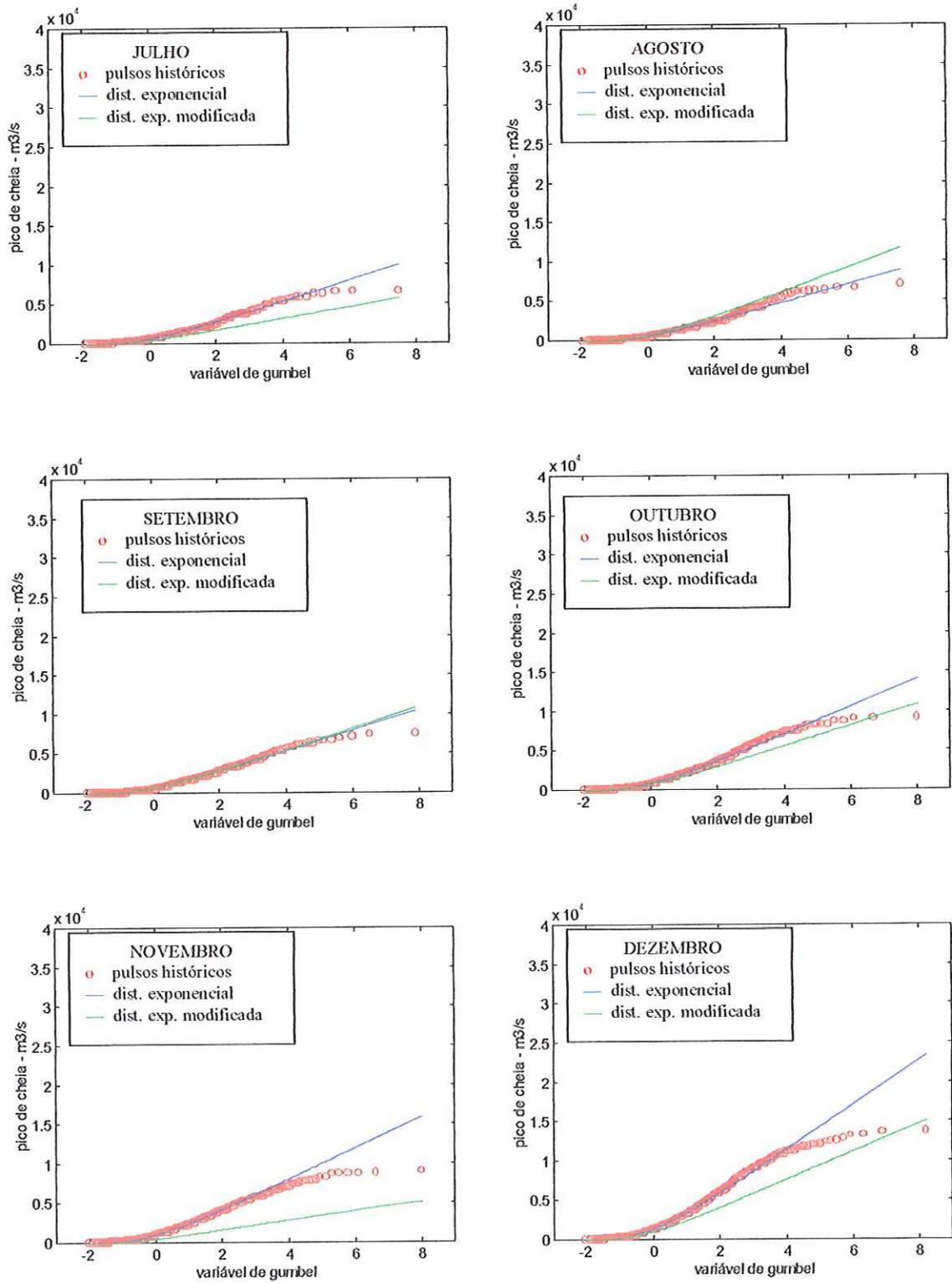


FIGURA 41b - Ajustes das distribuições de probabilidades exponencial e exponencial modificada dos pulsos de vazões simulados, ajustados à série dos pulsos observados em cada mês, durante 73 anos de registro.

A análise gráfica da duas figuras mostra que nenhum dos dois modelos descreve apropriadamente os dados históricos de pulsos de vazões, embora os resultados do teste  $\chi^2$  referentes ao modelo exponencial, sejam significativos na maioria dos meses. O modelo de distribuição exponencial modificada descreve inadequadamente os dados de pulso de vazões. Enquanto o modelo exponencial consegue uma maior aproximação aos dados, na faixa mais baixa dos valores observados. Todavia, os dados de pulsos de vazões, em todos os meses, indicam que os mesmos seguem uma distribuição limitada em vez de uma distribuição ilimitada, que é o caso da distribuição exponencial. Tal comportamento é visualizado em todos os gráficos das Figuras 41a e 41b, até mesmo, nos meses em que o resultado do teste  $\chi^2$  referente ao ajuste da distribuição exponencial, sugere adequação do modelo.

e) *Ajuste do Modelo de Distribuição Inicial Limitada Para Pulsos de Vazões*

O ajuste da distribuição inicial limitada à uma série de dados é feita através de um método de estimação dos parâmetros  $k$ ,  $\alpha$  e  $u$ , utilizando o conjunto de valores da série de forma ordenada, diferentemente dos modelos anteriores que usam, apenas, o valor da média e desvio padrão amostral.

No caso da série de pulsos, a distribuição limitada pode ser ajustada, considerando todo o conjunto de valores ou apenas parte dele. Assim, consideraram-se as série de valores de pulsos históricos ordenados de forma crescente até um valor limite igual a: metade do valor médio; valor médio e o dobro do valor médio dos pulsos de cada mês.

As novas séries de pulsos ordenados, foram ajustadas através da distribuição de probabilidade limitada, onde foi utilizado o método dos momentos de pesos probabilísticos para estimação dos parâmetros, empregando as equações (79), (80) e (81). A qualidade dos ajustes foi avaliada, aplicando-se o teste  $\chi^2$  em todos os casos.

As estimativas dos parâmetros  $k$ ,  $\alpha$  e  $u$ , o número de pulsos e o resultado do teste aplicado, para cada série de pulsos de vazões, encontram-se descritos na Tabela 12a. Na Figura 42 são apresentadas as distribuições histórica e de probabilidade limitada ajustadas às novas séries de pulsos referentes aos meses de janeiro e junho.

TABELA 12a - Parâmetros da distribuição limitada referentes às diferentes séries de pulsos de vazões ordenados de forma crescente.

Mês	Nº de Pulsos	$K$	$\alpha$	$u$	Teste $\chi^2$
SÉRIES DE PULSOS ORDENADOS ATÉ A METADE DE SEU VALOR MÉDIO					
jan.	514	0,85	9,37	2340	0,042*
fev.	405	0,82	12,29	2226	0,042*
mar.	405	0,83	9,43	1827	0,021*
abr.	272	0,86	6,36	1093	0,037*
maio	326	0,78	8,06	1073	0,036*
jun.	307	0,85	5,04	880	0,086*
jul.	239	0,89	3,87	666	0,083*
ago.	282	0,73	6,23	593	0,035*
set.	398	0,74	5,34	670	0,053*
out.	442	0,80	4,82	876	0,071*
nov.	392	0,82	5,35	1014	0,049*
dez.	530	0,88	4,61	1431	0,033*
SÉRIES DE PULSOS ORDENADOS ATÉ SEU VALOR MÉDIO					
jan.	813	0,70	29,91	4725	0,162*
fev.	633	0,68	37,40	4516	0,229*
mar.	634	0,65	35,31	3723	0,057*
abr.	443	0,70	20,70	2220	0,105*
maio	520	0,66	21,71	2148	0,036*
jun.	525	0,75	11,53	1788	0,087*
jul.	404	0,76	10,16	1354	0,047*
ago.	458	0,70	11,08	1195	0,032*
set.	606	0,63	14,48	1342	0,159*
out.	711	0,73	10,29	1764	0,508*
nov.	680	0,80	8,31	2029	0,168*
dez.	858	0,75	13,30	2890	0,090*
SÉRIES DE PULSOS ORDENADOS ATÉ O DOBRO DO SEU VALOR MÉDIO					
jan.	1135	0,57	99,24	9517	0,135*
fev.	890	0,60	93,11	9064	0,241*
mar.	847	0,52	119,1	7470	0,183*
abr.	577	0,45	118,4	4488	0,124*
maio	650	0,44	113,6	4305	0,141*

TABELA 12a - Parâmetros da distribuição limitada referentes às diferentes séries de pulsos de vazões ordenados de forma crescente.

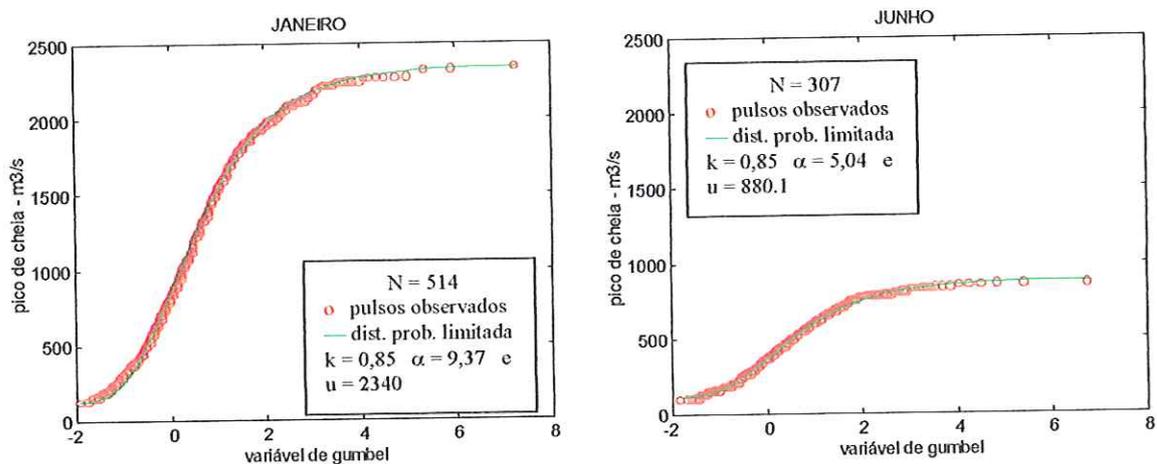
Mês	Nº de Pulsos	$K$	$\alpha$	$u$	Teste $\chi^2$
jun.	679	0.44	91,50	3540	0,142*
jul.	554	0.48	63,06	2664	0,147*
ago.	615	0.50	48,47	2385	0,267*
set.	830	0.50	47,68	2689	0,035*
out.	968	0.54	47,42	3558	0,116*
nov.	934	0.55	52,14	4078	0,117*
dez.	1136	0.45	112,4	5807	0,265*

\* Significativo ao nível de 5% -  $\chi^2_{(0,95, 7)} = 14,07$

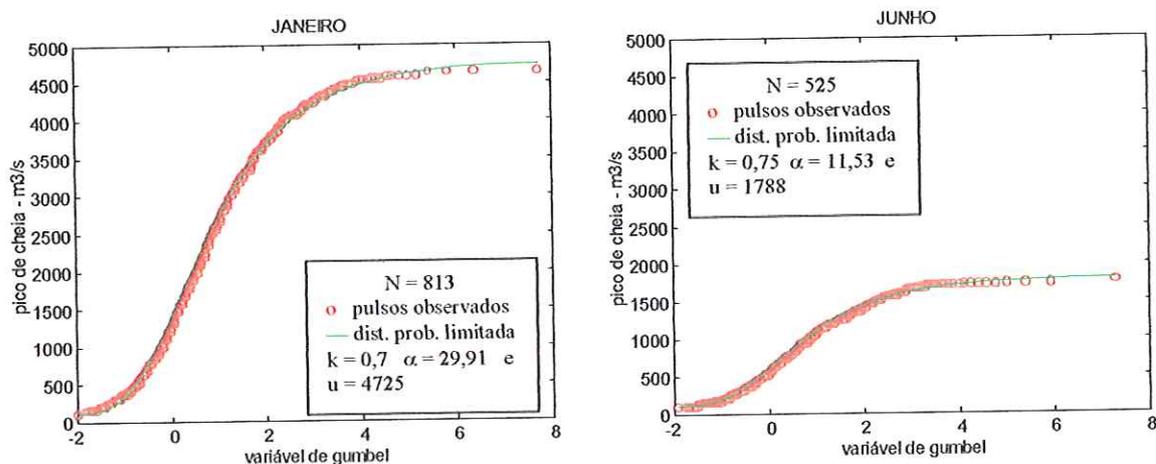
Os resultados do teste  $\chi^2$  evidenciam a ótima adequação dos ajustes da distribuição limitada aos dados de pulsos de vazões, em todos os casos, com nível de significância de 5 % de probabilidade. Os gráficos das distribuições ajustadas referentes aos meses de janeiro e junho, para as três extensões de séries, mostrados na Figura 42, representam a ótima qualidade dos ajustes e a adequabilidade da distribuição inicial limitada em descrever os dados de pulsos, obtidos em todos os meses.

Com os valores de  $k$ ,  $\alpha$  e  $u$  descritos na Tabela 12a, procederam-se simulações de vazões diárias através do novo modelo, obtendo-se 3 séries sintéticas de vazões diárias com 73 anos de extensão, correspondentes às 3 diferentes séries de pulsos ordenados. As hidrógrafas das vazões média e dos desvios padrão das séries de vazões diárias e mensais, histórica e simuladas, estão exibidas, respectivamente, nas Figuras 43 e 44. Enquanto a Figura 45 mostra as hidrógrafas das vazão extremas, históricas e simuladas, para os 3 diferentes comprimentos de séries de pulsos.

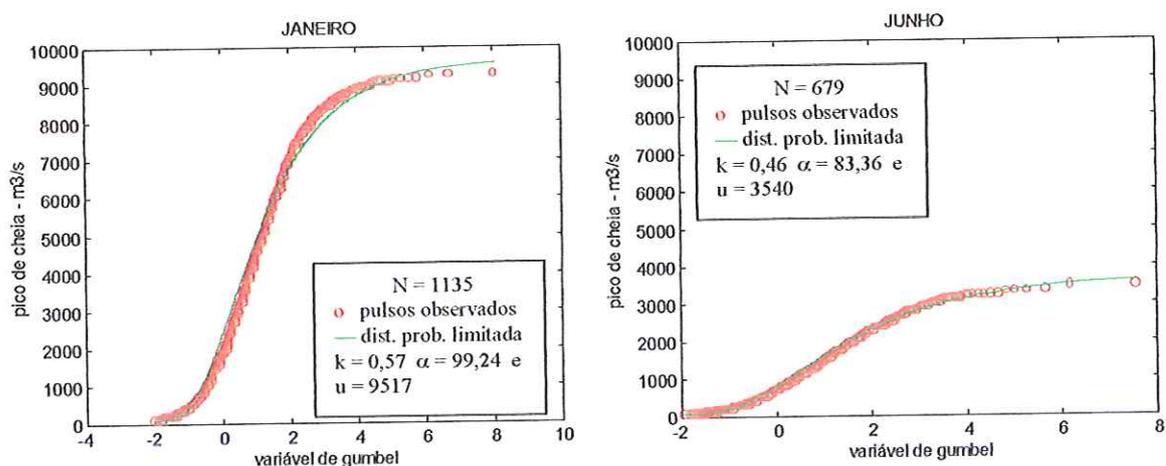
Observam-se através dos valores de  $k$ ,  $\alpha$  e  $u$  (Tabela 12a), bem como dos gráficos apresentados nas Figuras 43, 44 e 45 que, à medida em que aumenta-se a faixa de valores da série de pulsos ordenados, obtém-se dados de vazões simuladas com maiores desvios e valores extremos mais altos. Assim, para as séries de pulsos ordenados até a metade do seu valor médio, obtém-se ótimos ajustes de vazões média



a) série de pulsos históricos ordenados crescentemente até a metade de seu valor médio

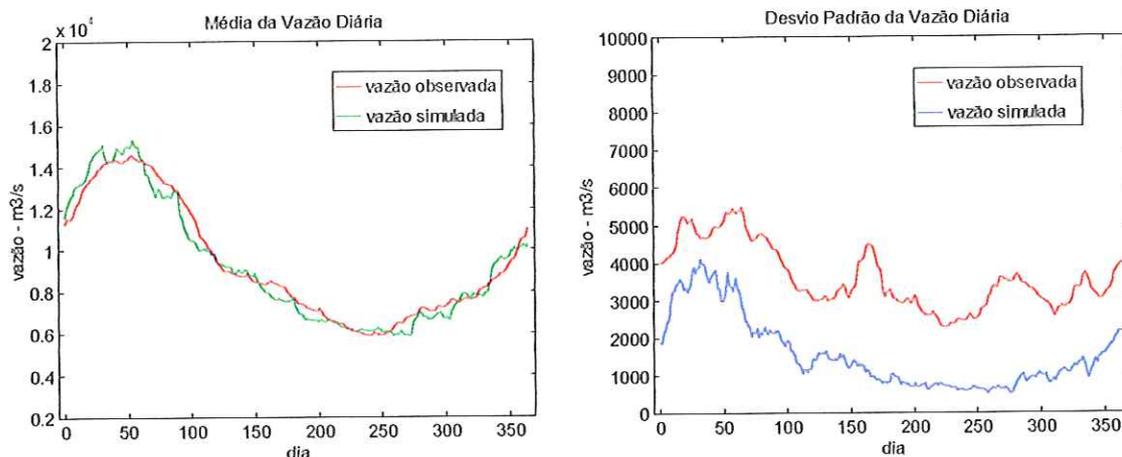


b) série de pulsos históricos ordenados crescentemente até o seu valor médio

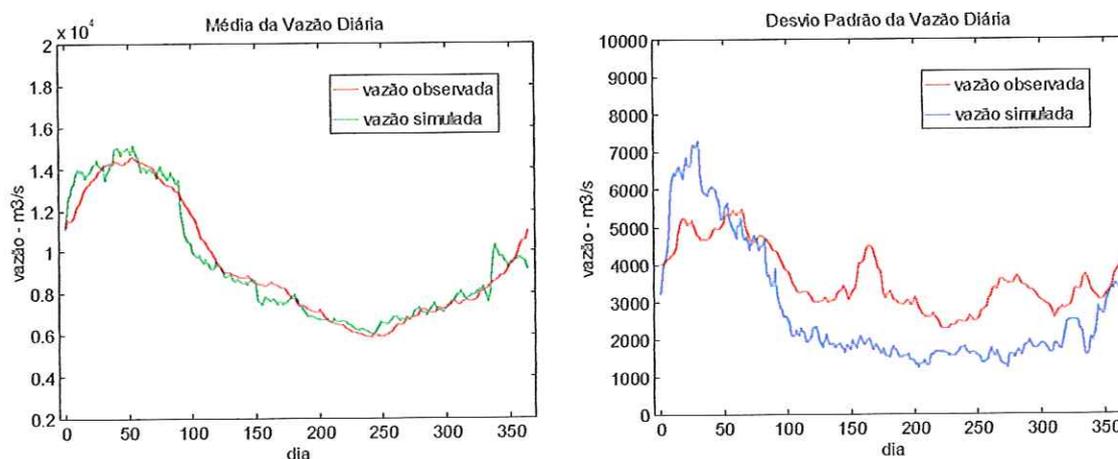


c) série de pulsos históricos ordenados crescentemente até o dobro de seu valor médio

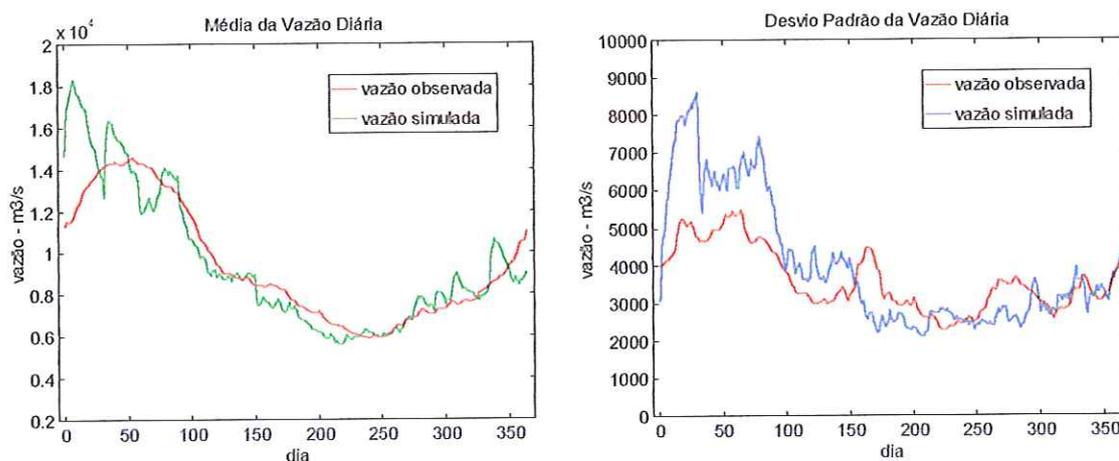
FIGURA 42 - Ajuste da distribuição de probabilidade limitada às três diferentes séries de pulsos históricos, referentes aos meses de janeiro e junho.



a) vazões geradas utilizando o ajuste a série de pulsos ordenados até metade do seu valor médio

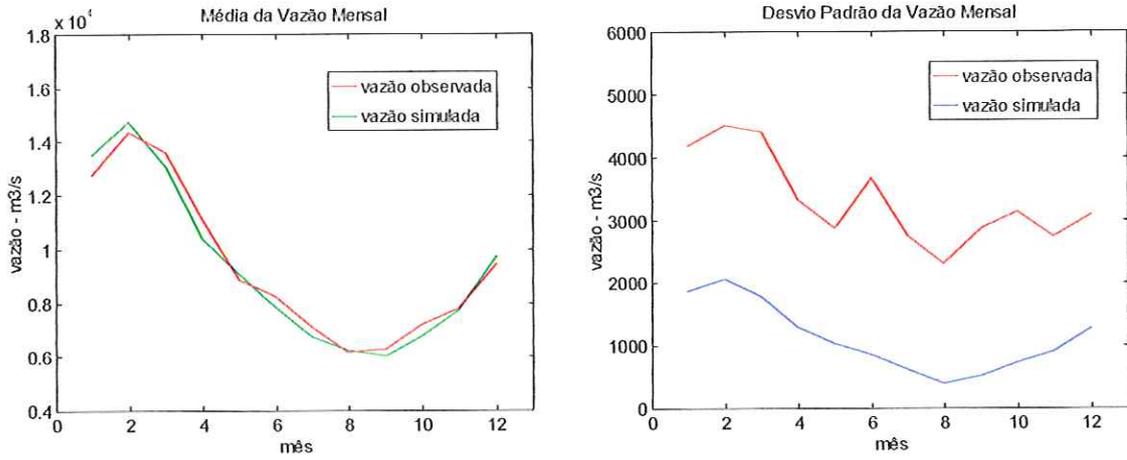


b) vazões geradas utilizando o ajuste a série de pulsos ordenados até seu valor médio

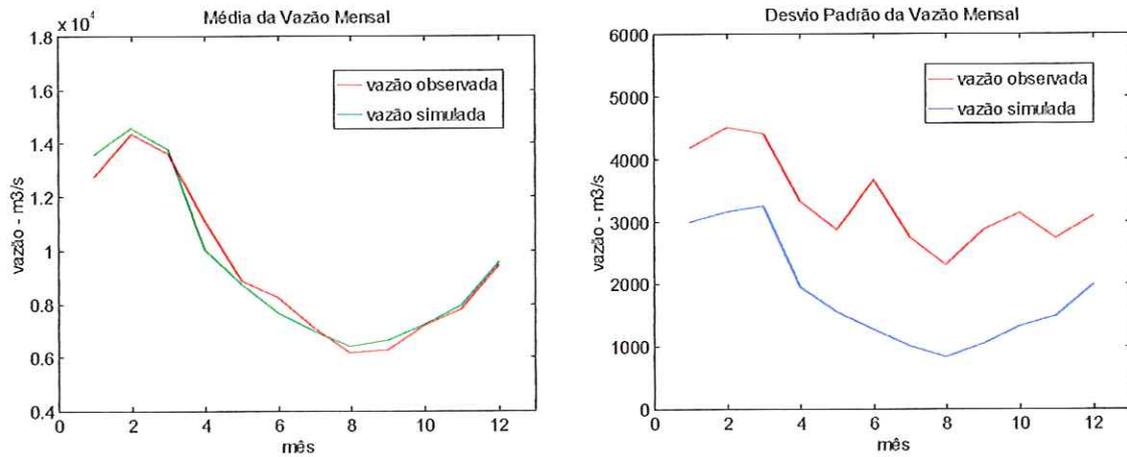


c) vazões geradas utilizando o ajuste a série de pulsos ordenados até o dobro de seu valor médio

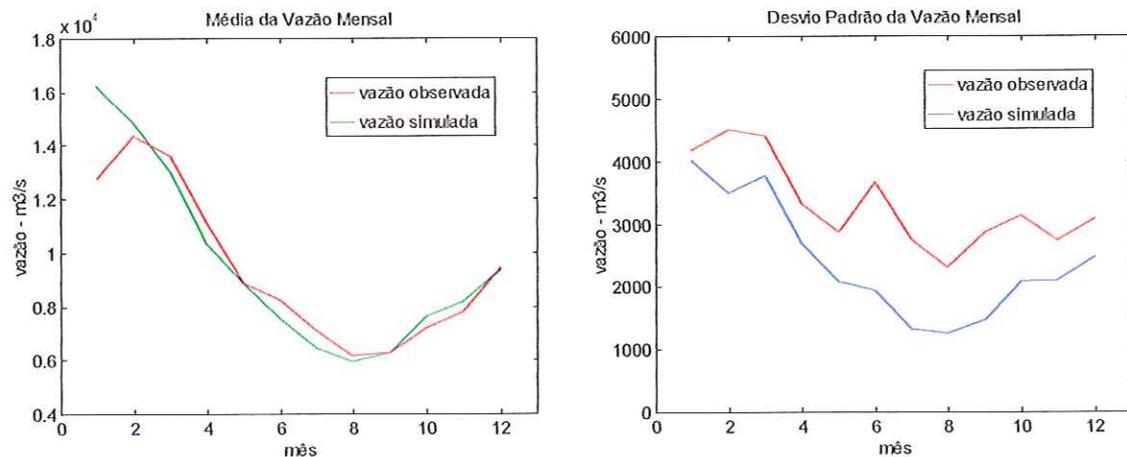
FIGURA 43 - Média e Desvio Padrão das vazões diárias, históricas e simuladas, referentes ao ajuste das três diferentes séries de pulsos de cheias observadas.



a) vazões geradas utilizando o ajuste a série de pulsos ordenados até metade do seu valor médio

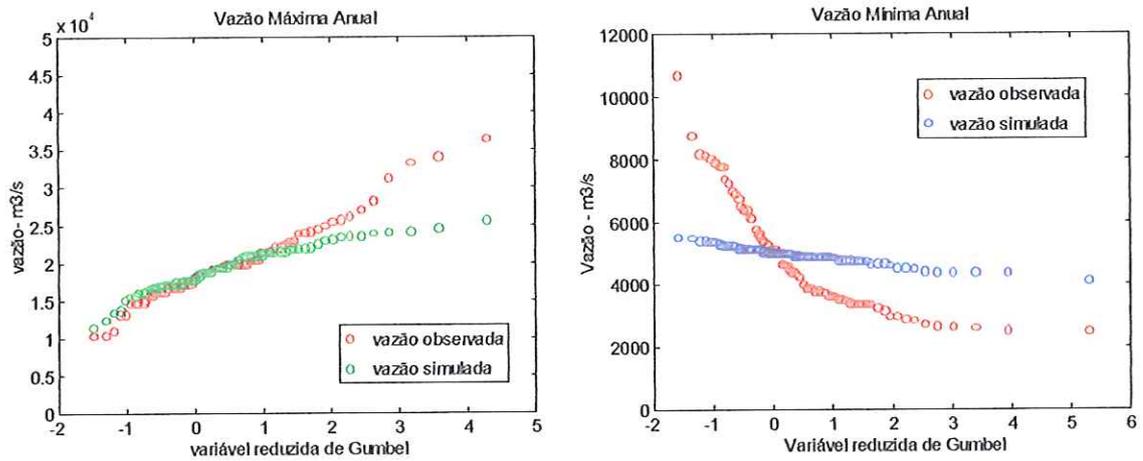


b) vazões geradas utilizando o ajuste a série de pulsos ordenados até seu valor médio

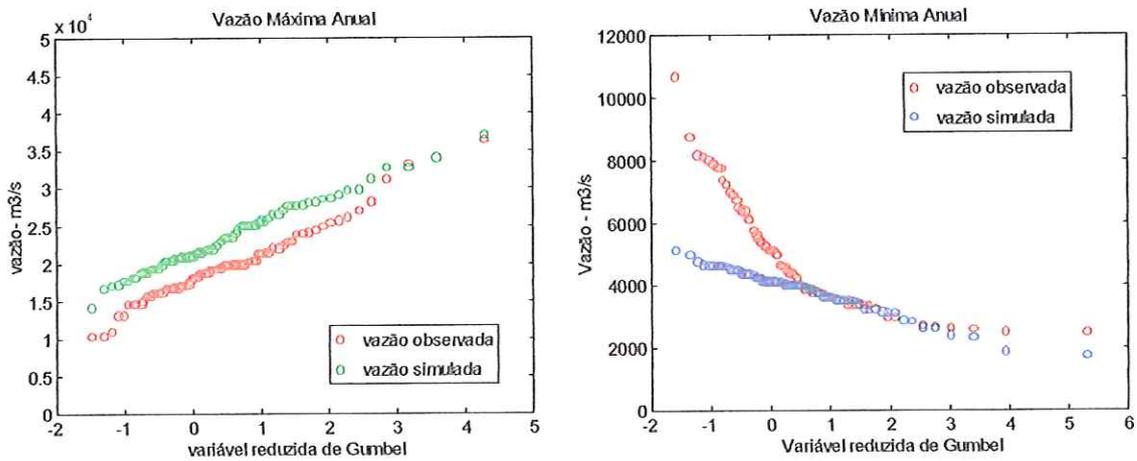


c) vazões geradas utilizando o ajuste a série de pulsos ordenados até o dobro de seu valor médio

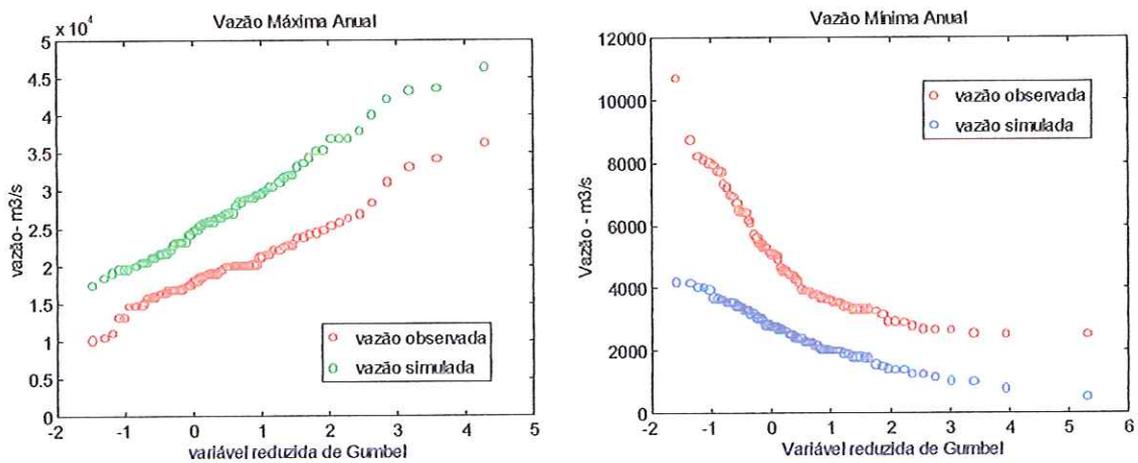
FIGURA 44 - Média e Desvio Padrão das vazões mensais, históricas e simuladas, referentes ao ajuste das três diferentes séries de pulsos de cheias observadas.



a) vazões geradas utilizando o ajuste a série de pulsos ordenados até metade do seu valor médio



b) vazões geradas utilizando o ajuste a série de pulsos ordenados até seu valor médio



c) vazões geradas utilizando o ajuste a série de pulsos ordenados até o dobro de seu valor médio

FIGURA 45 - Vazões extremas, históricas e simuladas, referentes ao ajuste das três diferentes séries de pulsos de cheias observadas.

diária e mensal, no entanto, os respectivos desvios padrão são subestimados. No caso da média e o desvio padrão da vazão anual, os resultados para a série simulada são 9318,6 e 514,8 e, para a série histórica, 9264,8 e 1958,6, respectivamente. Com relação aos valores extremos, os dados simulados não conseguiram superar os valores mais extremos da série histórica.

À proporção em que se aumenta a faixa de valores das séries de pulsos ordenados, os ajustes da vazão média diária e mensal mantêm-se adequadamente, enquanto os desvios padrão diário e mensal tornam-se mais próximos aos dados históricos (Figuras 43 e 44). A média da vazão anual permanece no mesmo patamar, 9301,8 e 9518,8, já o desvio padrão aumenta para 860,1 e 1303,4, respectivamente, quando os pulsos são ordenados até seu valor médio e, depois, até o dobro de seu valor médio. Para as referidas séries de pulsos, os valores extremos simulados se aproximam dos seus respectivos valores históricos e depois ultrapassam esses valores, como mostra a Figura 45.

Com isto, constata-se que a faixa dos valores de pulsos ordenados mais apropriada para o ajuste do modelo, considerando as estatísticas diária, mensal, anual e a série de valores extremos, encontra-se entre o valor médio e o dobro da valor médio. Com duas tentativas de ajustes, estabeleceu-se o limite superior da série de pulsos ordenados para cada mês, que resultou na melhor adequação do modelo. As séries de pulsos ordenados foram modeladas através da distribuição inicial limitada e utilizadas no processo de geração de vazões diárias através do novo modelo.

As estimativas dos parâmetros da distribuição limitada, resultante do seu ajuste às séries históricas de pulsos utilizadas para a calibração do modelo, bem como, os resultados do teste  $\chi^2$  aplicado para avaliar a qualidade dos referidos ajustes, estão mostrados na Tabela 12b. As Figuras 46a e 46b exibem as distribuições histórica e limitada ajustada à série de pulsos de cada mês.

O desempenho notável da distribuição limitada na modelação das séries de pulsos de vazões utilizadas na calibração do novo modelo, é comprovado através da excelente qualidade dos ajustes, verificada nos gráficos das Figuras 46a e 46b. A adequabilidade do modelo é corroborada pelos baixíssimos valores do teste  $\chi^2$  aplicado, como pode ser verificado na Tabela 12b.

TABELA 12b - Parâmetros da distribuição limitada referentes às séries de pulsos de vazões dos 12 meses utilizadas no ajuste do modelo proposto.

Mês	Nº de Pulsos	$K$	$\alpha$	$u$	Teste $\chi^2$
jan.	686	0,83	12,27	3487	0,057*
fev.	650	0,68	38,69	4775	0,214*
mar.	638	0,65	36,14	3786	0,095*
abr.	560	0,49	92,69	4129	0,095*
maio	625	0,45	98,64	3862	0,083*
jun.	708	0,35	168,37	4428	0,173*
jul.	567	0,46	74,85	2955	0,058*
ago.	603	0,55	35,87	2224	0,069*
set.	874	0,43	84,14	3436	0,130*
out.	950	0,54	44,80	3359	0,071*
nov.	950	0,52	63,79	4294	0,075*
dez.	970	0,65	26,94	3682	0,045*

\* Significativo ao nível de 5%  $-\chi^2_{(0,95, 7)} = 14,07$

### 6.2.3 Geração de Vazões Diárias e Testes Realizados

Com os valores dos parâmetros da distribuição limitada descritos na Tabela 12b e os da distribuição exponencial e exponencial modificada apresentados na Tabela 11, procederam-se simulações de vazões diárias através do novo modelo, com geração de pulsos de vazões, utilizando as três distribuições ajustadas.

As recessões das cheias simuladas foram determinadas, nos três casos, usando as equações (115) e (116) e os respectivos parâmetros relacionados na Tabela 9. A seqüência de dias com e sem pulsos foram modeladas por uma cadeia de Markov, utilizando as probabilidades de transição apresentadas nas Tabelas 8a e 8b, conforme o procedimento descrito no item b da secção 5.6.4.

Foram geradas 15 séries de vazões diárias, com 73 anos de extensão cada uma, num processo contínuo de simulação passando de um ano para o seguinte, utilizando cada um dos três modelo de simulação de pulsos. Assim, obtiveram-se 3 séries sintéticas de vazões diárias, com 1095 anos cada uma.

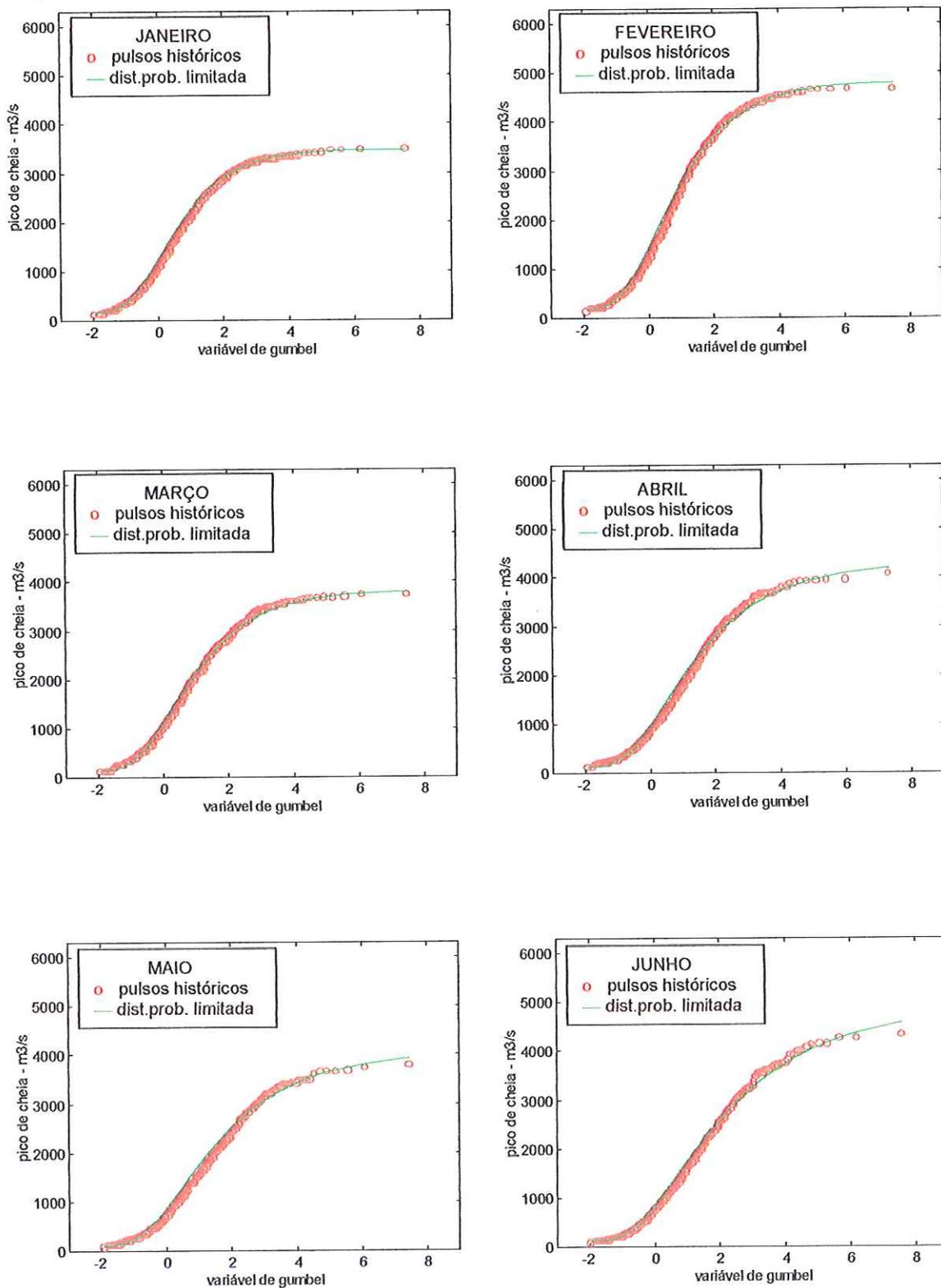


FIGURA 46a - Ajuste da distribuição de probabilidade limitada aos pulsos históricos de cheias, referentes aos meses de janeiro à junho, dos 73 anos de dados.

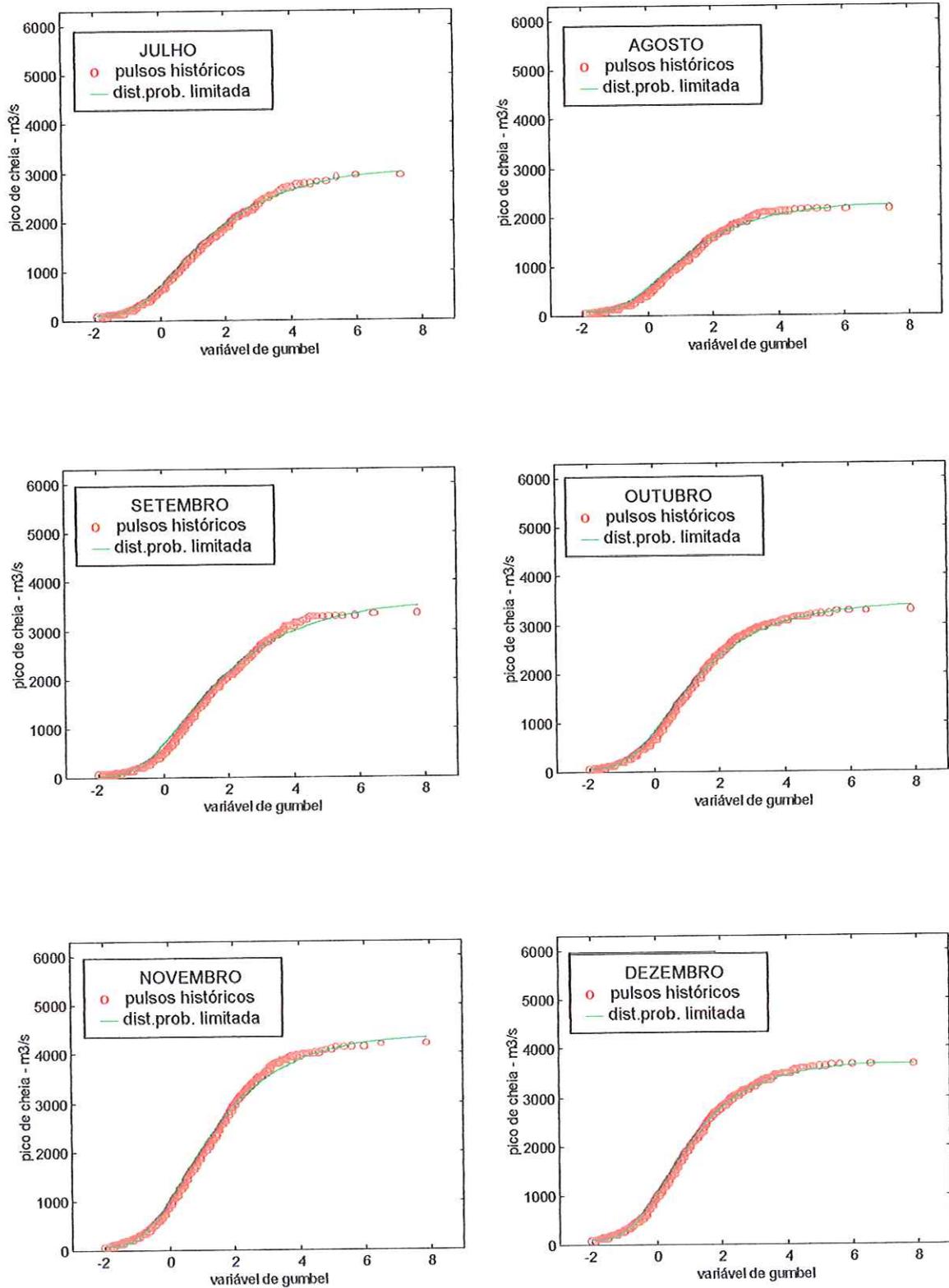


FIGURA 46b - Ajuste da distribuição de probabilidade limitada aos pulsos históricos de cheias, referentes aos meses de julho à dezembro, dos 73 anos de dados.

### a) Análise Comparativa

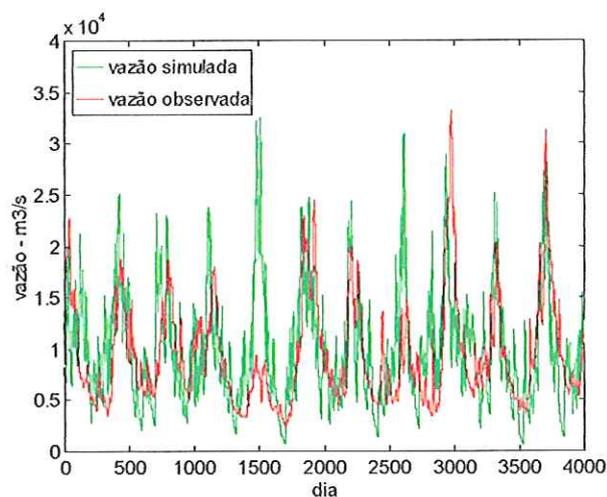
Para efeito de comparação, utilizaram-se as estatísticas referentes às 15 séries sintéticas e à série histórica de vazões diárias, todas com 73 anos de extensão, considerando cada modelo de geração de pulsos. As hidrógrafas da vazão diária e os gráficos da média e desvio padrão das séries de vazões diárias, mensais e anuais, históricas e simuladas, estão apresentadas nas Figuras 47, 48, 49 e 50, respectivamente. Enquanto os valores extremos de vazões simuladas e observadas são exibidos na Figura 51. Na Figura 52 são apresentadas as relações entre a assimetria e curtose, através de diagramas de momentos, das séries sintéticas e histórica de vazões diárias, mensais e anuais, referentes aos três modelos de pulsos.

Os gráficos apresentados nas Figuras de 48 a 52, mostram, nitidamente, a superioridade do modelo proposto, em reproduzir adequadamente as estatísticas nos níveis diário, mensal e anual, além dos valores extremos, em relação aos dados históricos. A simulação de vazões diárias baseada na geração de pulsos de vazões através da distribuição limitada, adotada no novo modelo, reproduz as estatísticas da série histórica sem a necessidade de utilizar o esquema de ajuste adotado por SARGENT (1979), conforme descrição do item c na secção 5.6.4.

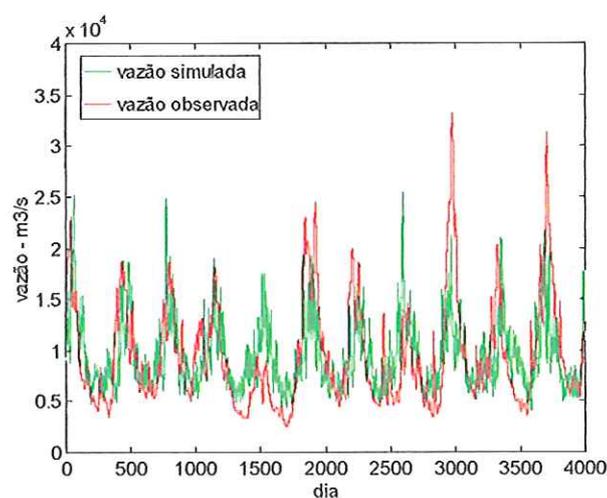
Os diagramas de momentos da Figura 50, evidencia uma melhor aproximação dos valores de assimetria e curtose das 15 séries de vazões diárias, mensais e anuais resultantes da simulação de vazões diárias com o novo modelo, utilizando a distribuição limitada na modelação dos pulsos de vazões, com os respectivos valores históricos.

O modelo de geração de vazões diárias apresenta desempenho regular, quando utiliza a distribuição exponencial para simulação dos pulsos, reproduzindo de forma satisfatória, apenas as vazões médias diárias e mensais e os valores extremos máximos. Contudo, quando emprega o modelo exponencial modificado para modelar os pulsos de vazões, apresenta desempenho insatisfatório, subestimando as estatísticas da série histórica, nos níveis diário, mensal e anual, bem como as vazões máximas.

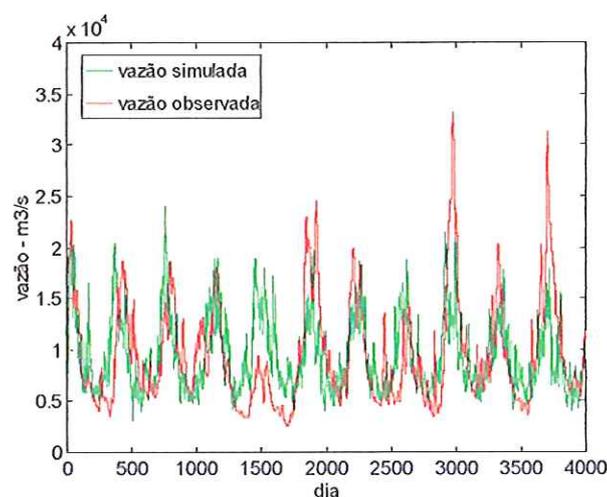
Os resultados dos testes de qualidade de ajuste  $\chi^2$  e Kolmogorov-Smirnov aplicados aos dados das 15 séries de vazões diárias simuladas, em relação aos dados históricos, estão relacionados na Tabela 13, para cada modelo de pulsos de vazões.



a) Vazões geradas utilizando o modelo de distribuição limitada de pulsos

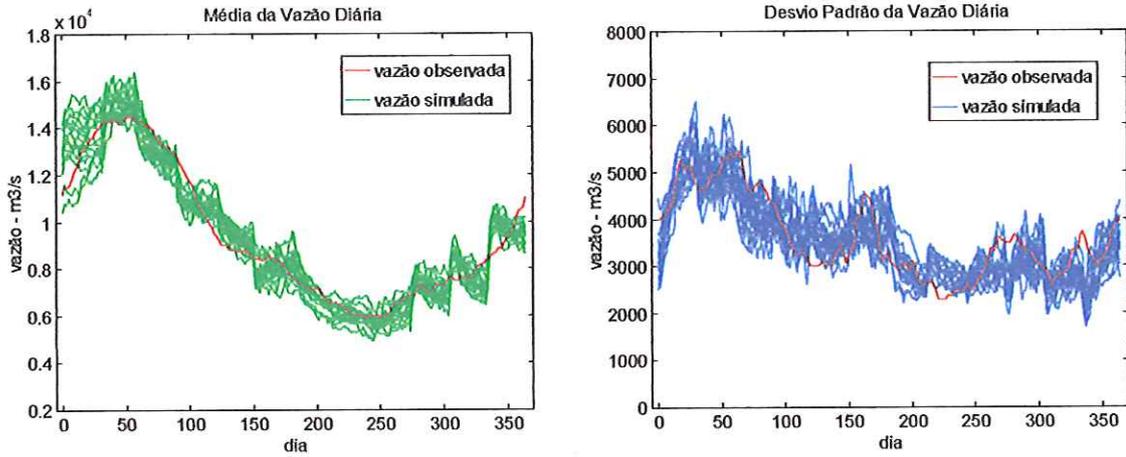


b) Vazões geradas utilizando o modelo de distribuição exponencial de pulsos

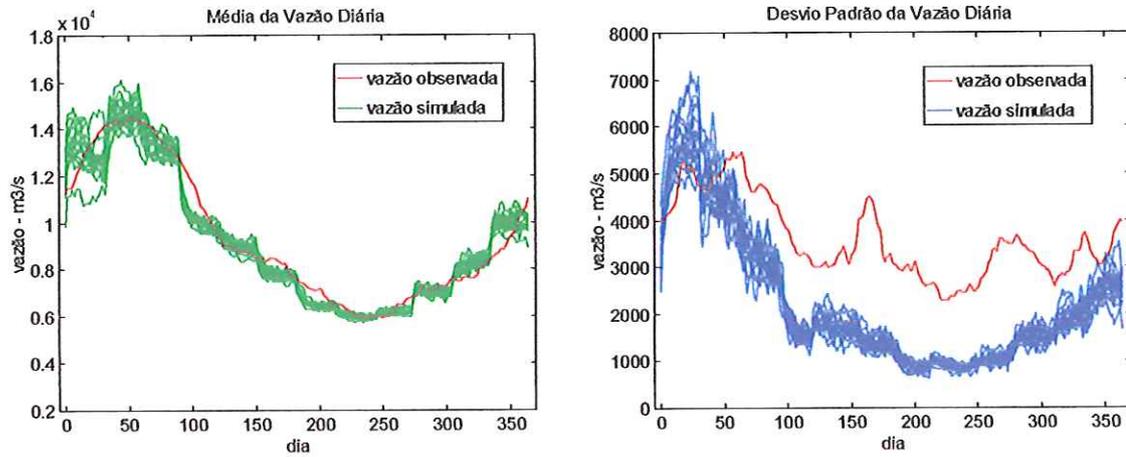


c) Vazões geradas utilizando o modelo de distribuição exponencial modificada de pulsos

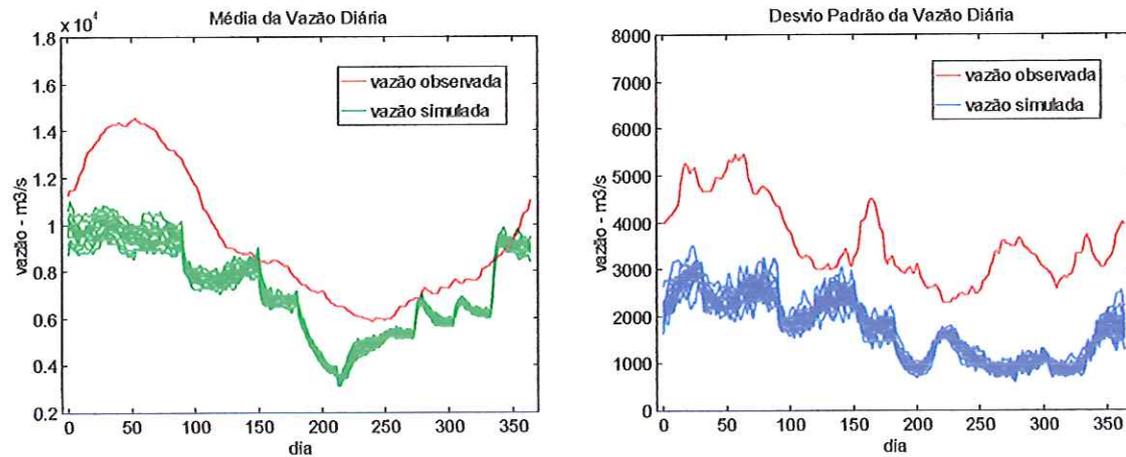
FIGURA 47 – Hidrógrafas representativas das vazões diárias históricas e simuladas geradas através dos três modelos de geração de pulsos de vazões.



a) vazões geradas utilizando o modelo de distribuição limitada de pulsos

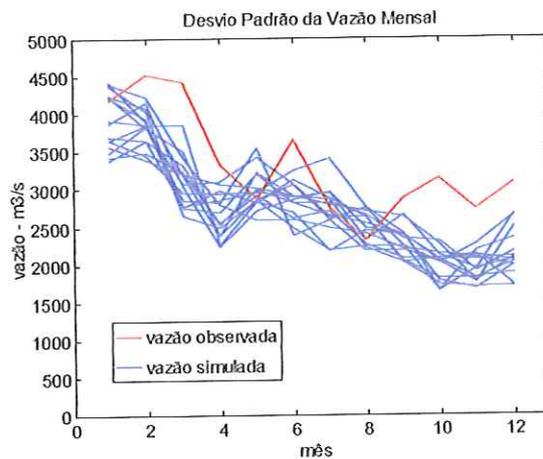
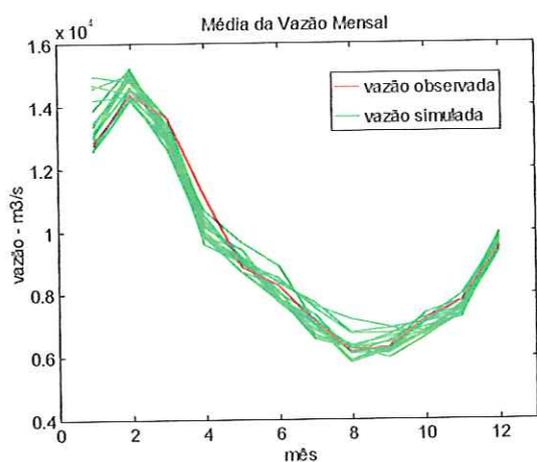


b) vazões geradas utilizando o modelo de distribuição exponencial de pulsos

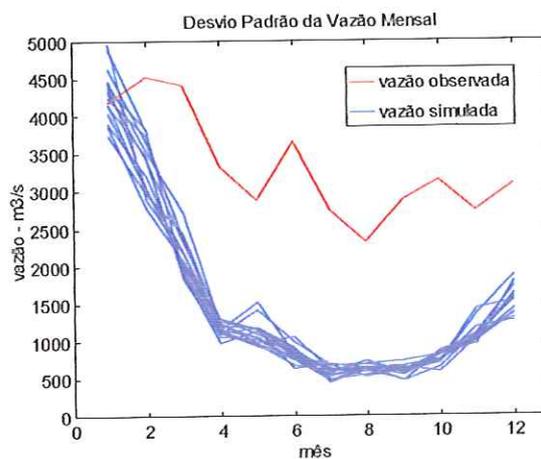
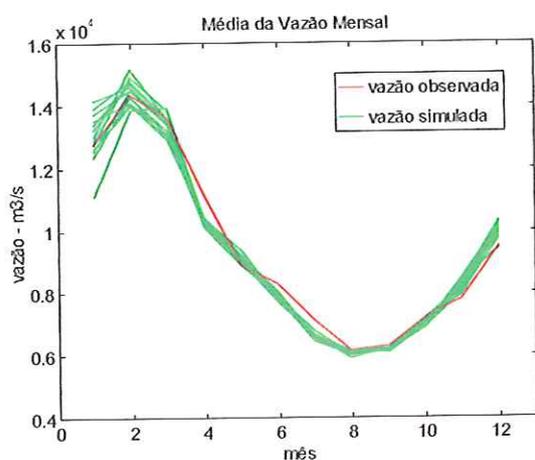


c) vazões geradas utilizando o modelo de distribuição exponencial modificada de pulsos

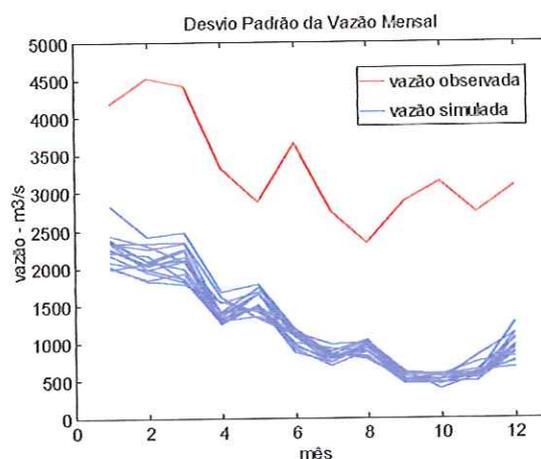
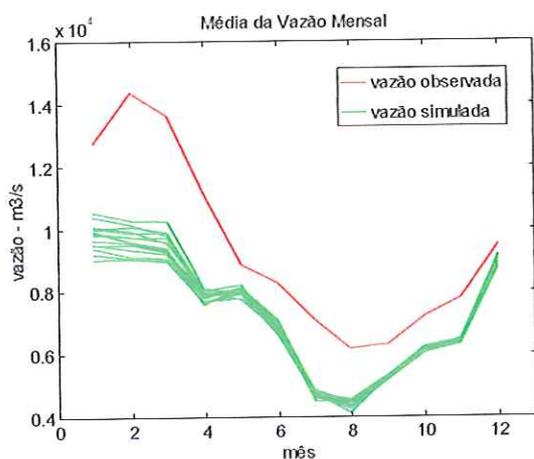
FIGURA 48 - Média e Desvio Padrão das vazões diárias, históricas e simuladas, referentes às 15 séries geradas por cada um dos três modelos de pulsos.



a) vazões geradas utilizando o modelo de distribuição limitada de pulsos

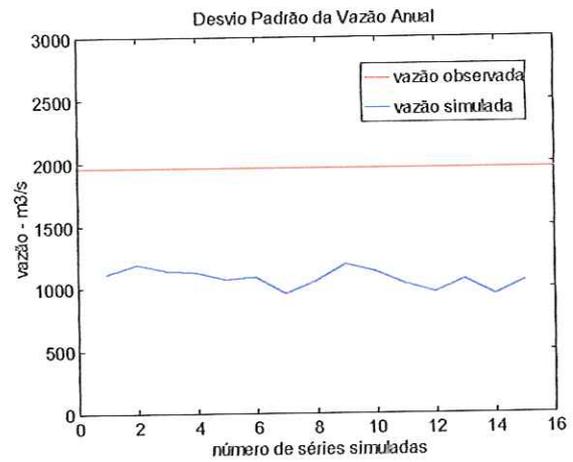
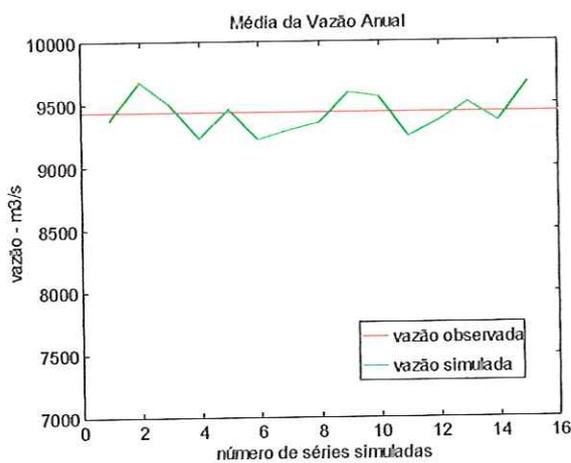


b) vazões geradas utilizando o modelo de distribuição exponencial de pulsos

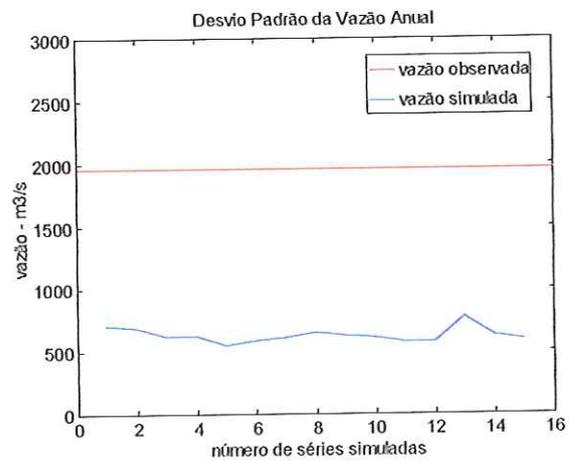
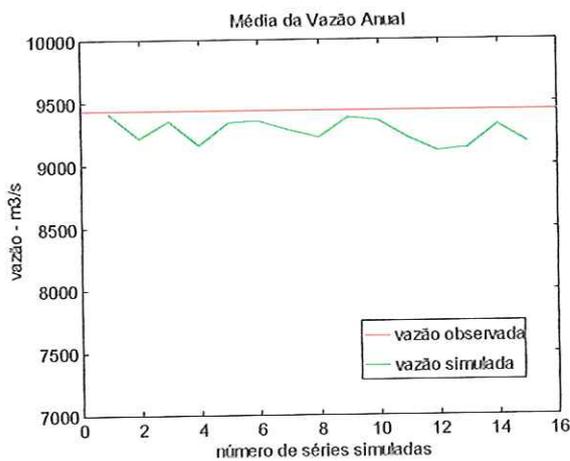


c) vazões geradas utilizando o modelo de distribuição exponencial modificada de pulsos

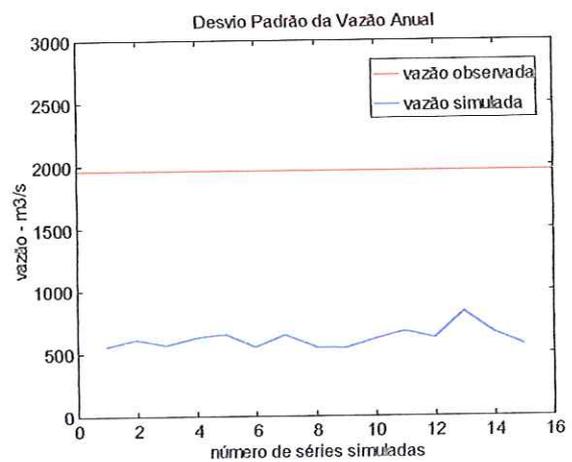
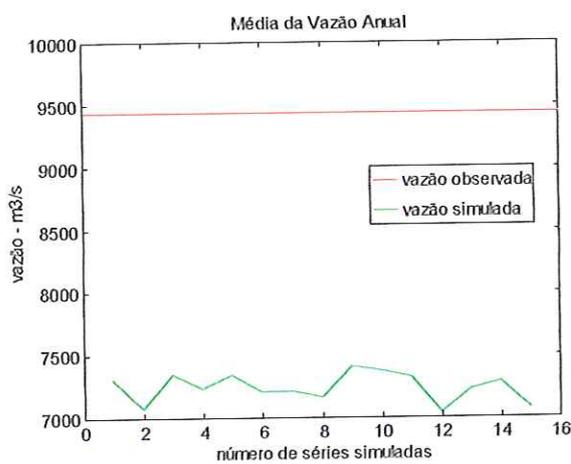
FIGURA 49 - Média e Desvio Padrão das vazões mensais, históricas e simuladas, referentes às 15 séries geradas por cada um dos três modelos de pulsos.



a) vazões geradas utilizando o modelo de distribuição limitada de pulsos

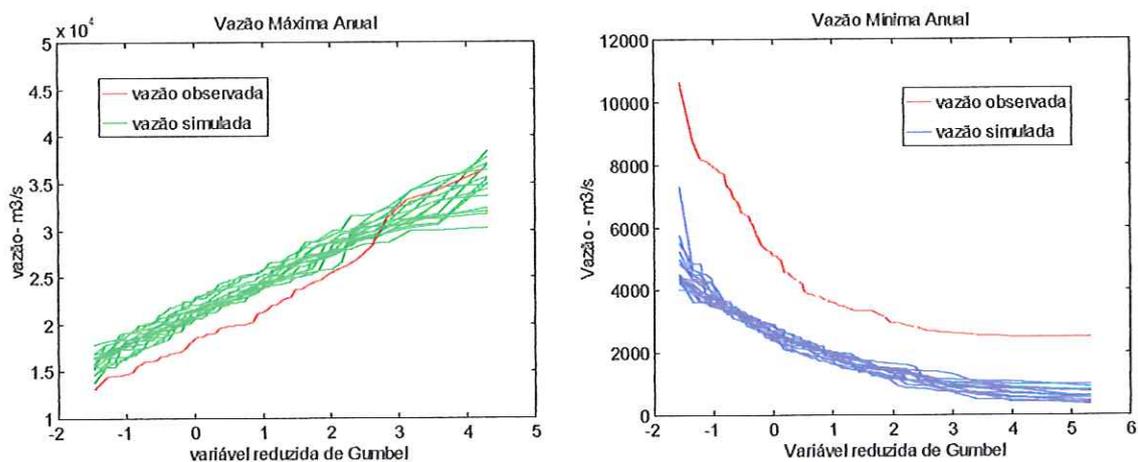


b) vazões geradas utilizando o modelo de distribuição exponencial de pulsos

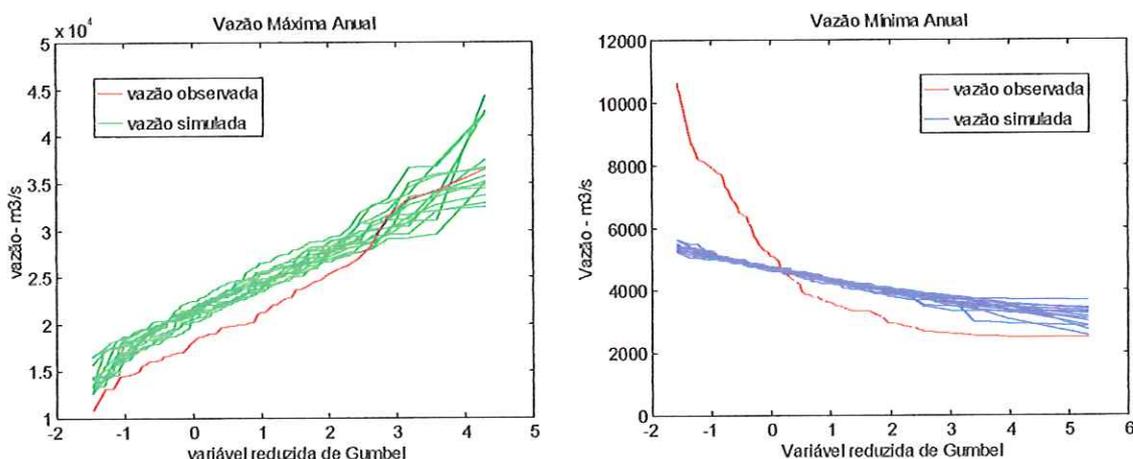


c) vazões geradas utilizando o modelo de distribuição exponencial modificada de pulsos

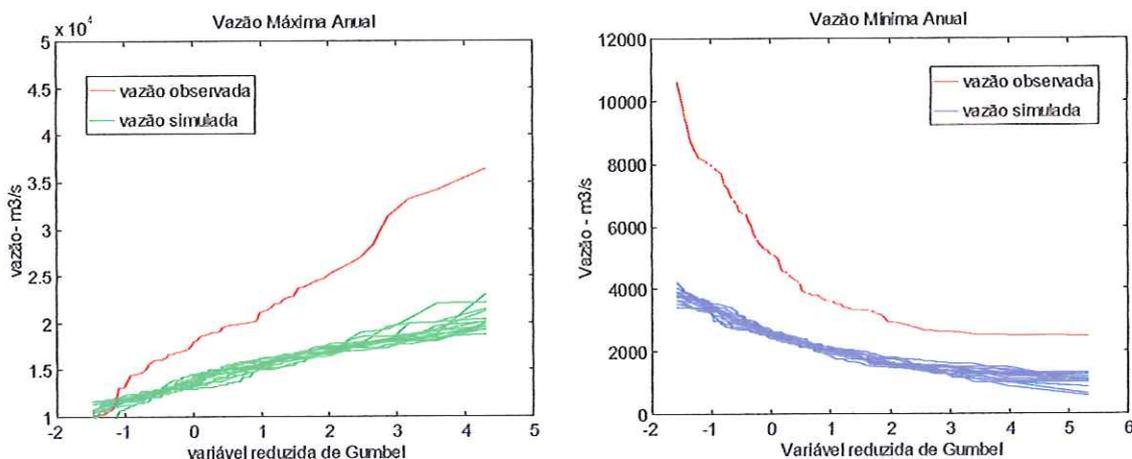
FIGURA 50 - Média e Desvio Padrão das vazões anuais, históricas e simuladas, referentes às 15 séries geradas por cada um dos três modelos de pulsos.



a) vazões geradas utilizando o modelo de distribuição limitada de pulsos



b) vazões geradas utilizando o modelo de distribuição exponencial de pulsos



c) vazões geradas utilizando o modelo de distribuição exponencial modificada de pulsos

FIGURA 51 - Vazões extremas, históricas e simuladas, referentes às 15 séries de vazões diárias geradas por cada um dos três modelos de pulsos.

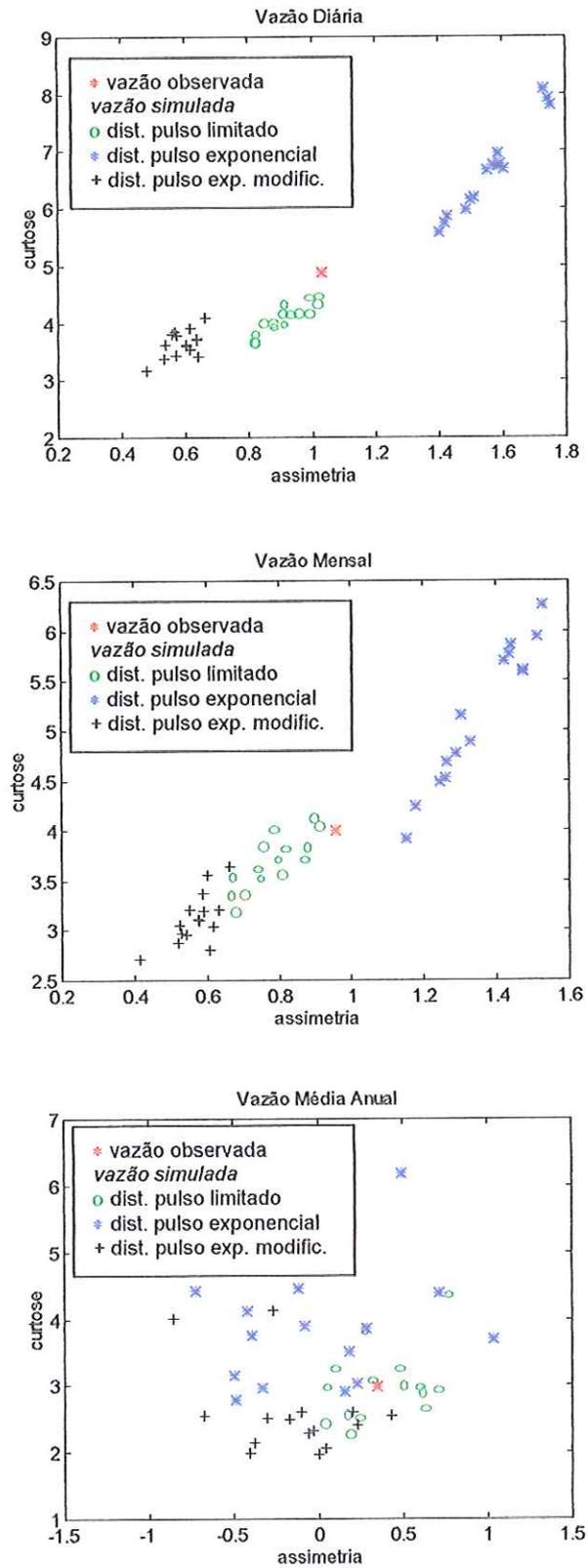


FIGURA 52 - Diagrama de momentos amostrais das vazões diárias, mensais e anuais referentes às 15 séries simuladas e a série histórica.

Embora não exista valores críticos para ser comparados aos resultados dos dois testes, pois não há valores crítico para dados correlacionados como é o caso de vazões diárias, os mesmos são utilizados para comparação dos modelos, onde o melhor desempenho está associado ao menor valor do resultado de cada teste.

Os resultados dos testes  $\chi^2$  e Kolmogorov-Smirnov mostrados na Tabela 13, confirmam o bom desempenho do modelo de vazões diárias quando simula pulsos de vazões através da distribuição limitada e sua superioridade em relação ao uso dos outros dois modelos.

TABELA 13 - Testes da qualidade de ajustes dos modelos utilizados.

Ordem das Séries de Vazões	Modelo de Pulso de Vazão com Distribuição de Probabilidade					
	<i>Limitada</i>		<i>Exponencial</i>		<i>Exp. modificada</i>	
	$\chi^2$	K-S	$\chi^2$	K-S	$\chi^2$	K-S
1	306,6	0,0483	1778,9	0,1363	4741,2	0,2259
2	147,3	0,0342	2649,2	0,1312	5662,5	0,2520
3	248,7	0,0431	3150,3	0,1354	4517,7	0,2122
4	248,1	0,0410	3343,9	0,1345	4965,7	0,2377
5	146,5	0,0340	2989,8	0,1311	4493,8	0,2109
6	282,4	0,0530	3036,8	0,1318	5112,6	0,2332
7	289,0	0,0525	3155,6	0,1325	5038,0	0,2285
8	351,9	0,0432	2526,6	0,1390	5459,7	0,2415
9	187,9	0,0345	2768,6	0,1335	4188,9	0,1872
10	163,7	0,0294	3321,0	0,1358	4348,2	0,2053
11	217,3	0,0436	2852,7	0,1362	4595,1	0,2118
12	373,5	0,0480	3450,6	0,1338	5992,0	0,2691
13	280,4	0,0477	2835,4	0,1298	4685,8	0,2267
14	247,9	0,0360	3337,7	0,1362	4785,0	0,2184
15	212,7	0,0493	2853,4	0,1328	5668,8	0,2540

$\chi^2$  - teste de qui-quadrado e K-S - teste de Kolmogorov-Smirnov; Graus de liberdade = 26642

Os coeficientes de correlação com defasagem 1 (um) das 15 séries sintéticas de vazões diárias, mensais e anuais, encontram-se em torno de 0,97; 0,68 e 0,06, para o modelo de distribuição de pulsos limitada; 0,95; 0,63 e 0,12, para o modelo exponencial; 0,93; 0,66 e 0,33, para o modelo exponencial modificada e 0,99; 0,77 e 0,41 para a série histórica, respectivamente. Os resultados dos coeficientes de correlação das séries sintéticas de vazões diárias, mensais e anuais geradas através do uso do modelo de distribuição de pulsos limitada, são os que melhor se aproximam dos respectivos valores da série histórica. Isto indica melhor adequação do modelo de geração de vazões quando adota a distribuição de pulsos limitada.

#### *b) Índice de Comparação das Séries Geradas Com a Série Histórica*

Os valores médios, máximos e mínimos de índice referente à média, desvio padrão, assimetria e curtose das 15 séries geradas, com 73 anos, nos níveis diário, mensal e anual, além dos valores extremos, bem como, índice histórico, correspondente a cada nível, encontram-se descritos nas Tabelas 14a e 14b. Apresentam-se, ainda, nas referidas tabelas, os valores da proporção das vezes que o índice sintético é maior que o índice histórico (distribuição empírica dos dados) para as 4 referidas estatísticas em todos os níveis citados. Quanto mais próximo o valor da proporção de índice estiver de 0,5 melhor será a adequação do modelo.

Os valores relacionados nas duas tabelas revelam o bom desempenho do modelo proposto, quando simula vazões a partir da geração de pulsos com distribuição limitada e sua superioridade em relação as outras distribuições, notadamente, em relação aos índices da média e desvio padrão todos os níveis considerados. Houve uma discordância significativa da assimetria das vazões anuais em relação à histórica, referente aos modelos exponencial e exponencial modificada, decorrente da inabilidade dos modelos em descrever as estatísticas históricas, como já foi demonstrado através dos resultados apresentado anteriormente.

Os resultados do índice de avaliação global encontram-se descritos na Tabela 15. Para este teste, quando menor for o valor absoluto da nota -  $\Psi$ , referente à distância padronizada, e quanto mais próximo de 0.5 estiver o valor  $d_j$ , que corresponde a distância ao centro da distribuição dos índices, mais adequado será o modelo utilizado.

TABELA 14a - Avaliação de desempenho dos três modelos de distribuição de pulsos, considerando os momentos das vazões diárias e mensais.

Modelo de Distribuição	VAZÕES DIÁRIAS				VAZÕES MENSAIS			
	$\mu_d$	$\sigma_d$	$\gamma_d$	$\kappa_d$	$\mu_m$	$\sigma_m$	$\gamma_m$	$\kappa_m$
<i>Histórica:</i>								
Índice <sub>(hist.)</sub>	9410,3	4615,4	1,03	4,89	9436,8	4050,3	0,86	4,00
<i>Limitada:</i>								
Índice <sub>(méd.)</sub>	9430,5	4661,4	0,92	4,08	9460,4	4053,8	0,79	3,65
Índice <sub>(max.)</sub>	9687,6	4980,3	1,02	4,44	9719,2	4365,5	0,91	4,13
Índice <sub>(min.)</sub>	9226,0	4455,2	0,82	3,67	9259,4	3827,2	0,68	3,19
$P_{(sint.>hist.)}$	0,47	0,53	0,00	0,00	0,47	0,47	0,00	0,13
<i>Exponencial:</i>								
Índice <sub>(méd.)</sub>	9321,6	3867,6	1,67	6,79	9350,2	3390,5	1,39	5,27
Índice <sub>(max.)</sub>	9461,3	4101,1	1,86	8,28	9489,2	3600,0	1,58	6,55
Índice <sub>(min.)</sub>	9162,4	3649,7	1,50	5,66	9189,9	3155,1	1,20	4,13
$P_{(sint.>hist.)}$	0,00	0,00	1,00	0,80	0,00	0,00	1,00	0,93
<i>Exp. modif.:</i>								
Índice <sub>(méd.)</sub>	7284,1	2664,5	0,69	3,58	7295,8	2325,6	0,61	3,18
Índice <sub>(max.)</sub>	7456,5	2812,8	0,78	4,03	7470,2	2490,6	0,71	3,81
Índice <sub>(min.)</sub>	7078,4	2479,4	0,56	3,11	7088,4	2130,4	0,46	2,73
$P_{(sint.>hist.)}$	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

$\mu$  - média;  $\sigma$  - desvio padrão;  $\gamma$  - coeficiente de assimetria;  $\kappa$  - coeficiente de curtose,  
 $P_{(sint.>hist.)}$  - proporção das vezes que o índice sintético é superior ao histórico (distribuição empírica)



TABELA 14b - Avaliação de desempenho dos três modelos de distribuição de pulsos, considerando os momentos das vazões anuais e máximas anuais e a vazão máxima.

Modelo de Distribuição	VAZÕES MÉDIAS ANUAIS				VAZÕES MÁXIMAS ANUAIS					
	$\mu_a$	$\sigma_a$	$\gamma_a$	$\kappa_a$	$\mu_{mx}$	$\sigma_{mx}$	$\gamma_{mx}$	$\kappa_{mx}$	$Q_{mx}$	
<i>Histórica:</i>										
Índice <sub>(hist.)</sub>	9436,8	1958,6	0,36	2,99	19807,1	5055,2	1,01	4,60	36374,0	
<i>Limitada:</i>										
Índice <sub>(méd.)</sub>	9430,5	1078,5	0,39	2,99	23153,5	4015,1	0,69	3,36	35357,8	
Índice <sub>(max.)</sub>	9687,6	1200,5	0,78	4,35	24487,8	4469,9	1,32	5,12	42595,0	
Índice <sub>(min.)</sub>	9226,0	951,30	0,04	2,26	22251,2	3541,2	0,21	2,21	30327,6	
$P_{(sint.>hist.)}$	0,47	0,00	0,47	0,40	1,00	0,00	0,13	0,07	0,33	
<i>Exponencial:</i>										
Índice <sub>(méd.)</sub>	9270,8	633,9	0,01	3,82	22988,6	4321,2	0,75	4,27	37166,8	
Índice <sub>(max.)</sub>	9409,8	778,6	0,70	2,38	23920,7	5313,6	1,36	8,83	44315,4	
Índice <sub>(min.)</sub>	9112,5	551,8	-0,72	2,79	22058,9	3479,3	0,19	3,11	32555,4	
$P_{(sint.>hist.)}$	0,00	0,00	0,02	0,80	1,00	0,13	0,20	0,20	0,47	
<i>Exp. modif.:</i>										
Índice <sub>(méd.)</sub>	7244,4	621,8	-0,15	2,57	14658,2	2067,0	0,39	3,02	20311,3	
Índice <sub>(max.)</sub>	7415,9	835,1	0,43	4,13	15048,2	2352,4	1,24	5,55	22989,5	
Índice <sub>(min.)</sub>	7039,8	542,0	-0,85	1,97	13908,0	1761,7	-0,55	2,30	18778,8	
$P_{(sint.>hist.)}$	0,00	0,00	0,07	0,13	0,00	0,00	0,07	0,07	0,00	

$\mu$  - média;  $\sigma$  - desvio padrão;  $\gamma$  - coeficiente de assimetria;  $\kappa$  - coeficiente de curtose;  $Q_{mx}$  - vazão máxima.  
 $P_{(sint.>hist.)}$  - proporção das vezes que o índice sintético é superior ao histórico (distribuição empírica)

TABELA 15 - Índice de avaliação global.

Modelo de Pulso com Distribuição:	Distância Padronizada (Nota - $\psi$ )	Distância ao Centro da Distribuição - $d_j$
Limitada	-0,2315	0,533
Exponencial	-0,7102	0,733
Exponencial Modificada	155,84	0,000

Os resultados do teste mostrado na Tabela 15 confirma a superioridade da modelo proposto quando simula pulsos de vazões através da distribuição limitada. O modelo consegue reproduzir as estatísticas dos dados históricos nos níveis diário, mensal e anual, bem como em relação aos valores extremos. Isto é mais um indicativo do bom desempenho e adequabilidade do novo modelo na geração de vazões diárias.

#### 6.2.4 Análise dos Valores Extremos Máximos

Os 3 blocos com 15 séries de vazões diárias geradas com os 3 modelos de pulsos, renderam 15 séries de vazões máximas anuais, com 73 anos cada uma, para cada modelo de pulsos. As referidas séries de cheias anuais foram modeladas através da distribuição GEV, usando momentos LH. Os resultados dos ajustes e dos teste de sua qualidade (WANG, 1998) e da forma da distribuição ajustada (HOSKING et al., 1985a) estão apresentados na Tabela 16.

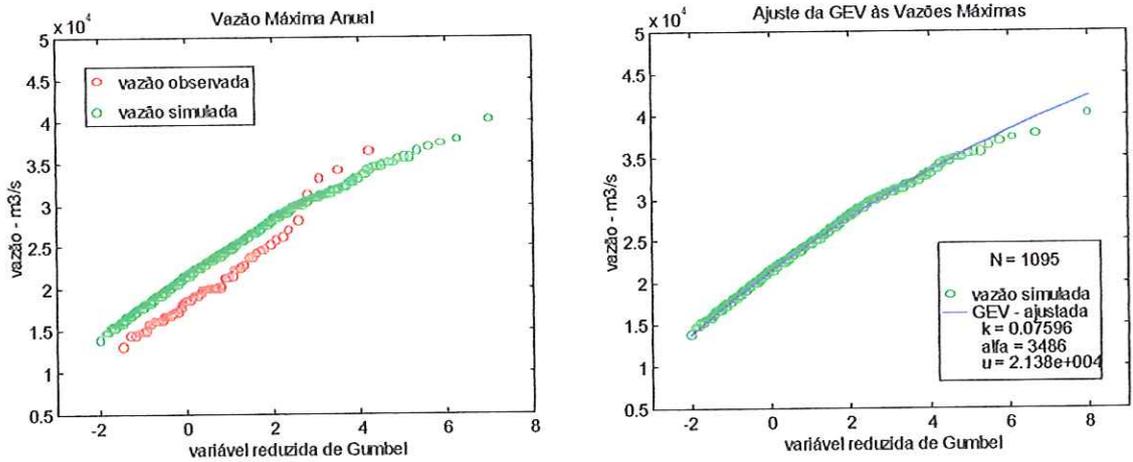
A geração das 15 séries de vazões diárias, através de processo contínuo uma após outro, permite transforma-las em uma única série com 1095 valores. Na Figura 53, ilustram-se as distribuições de frequência das séries de vazões máximas com 1095 anos, referentes aos 3 modelos de pulsos, da série histórica e da distribuição GEV ajustada às referidas séries. O ajuste, realizado através de momentos LH, resultaram nos valores de  $k$ ,  $\alpha$  e  $u$  mostrados em cada gráfico da citada figura.

As estimativas do parâmetro de forma  $k$  das 15 séries de cheias anuais, referentes aos 3 modelos de pulsos, indicam a ocorrência das três formas de

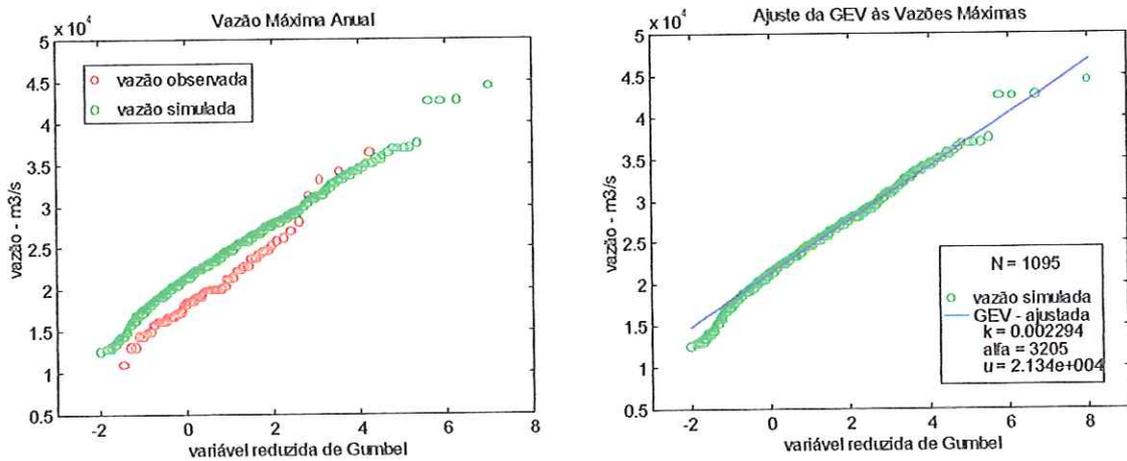
TABELA 16 – Resultados do ajuste da distribuição GEV às 15 séries de vazões máximas com 73 anos, referentes aos três modelos de pulsos de vazões.

Nº das amostras	Distribuição limitada de pulsos					Distribuição exponencial de pulsos					Distribuição exponencial modificada de pulsos				
	k	$\alpha$	u	Z <sub>W</sub>	Z <sub>H</sub>	k	$\alpha$	u	Z <sub>W</sub>	Z <sub>H</sub>	k	$\alpha$	u	Z <sub>W</sub>	Z <sub>H</sub>
1	0,101	3638	22541	0,32*	1,15*	0,005	4041	21742	0,55*	0,05*	0,215	1903	13963	0,73*	2,48
2	0,057	3652	22559	0,16*	0,64*	-0,097	3553	20427	0,31*	1,10*	-0,042	1570	13402	0,02*	0,48*
3	0,080	3674	22040	0,57*	0,91*	0,053	3128	21637	0,12*	0,60*	0,326	2228	14218	0,05*	3,71
4	0,002	3255	20522	0,28*	0,03*	-0,014	3225	20390	0,27*	0,16*	0,092	1892	13623	0,06*	1,04*
5	0,000	3516	20542	0,93*	0,01*	0,035	2869	21502	0,07*	0,40*	0,279	2292	14235	0,55*	3,18
6	0,132	4121	20703	0,51*	1,50*	0,016	3506	21469	0,15*	0,18*	0,303	2073	13714	0,40*	3,42
7	0,049	3431	21436	0,51*	0,56*	-0,128	2162	22059	1,04*	1,46*	0,183	1646	14162	0,07*	2,08
8	0,103	3567	20867	0,72*	1,17*	0,126	4205	21153	0,27*	1,44*	0,174	2021	13247	1,18*	2,00
9	-0,004	2751	21520	0,27*	0,04*	-0,058	3000	21803	0,08*	0,66*	0,254	1747	14343	0,15*	2,89
10	0,177	3378	21588	1,46*	2,02	0,189	2736	20225	0,38*	2,15	-0,040	1702	14038	0,06*	0,45*
11	0,120	3654	20686	0,04*	1,37*	0,075	3214	20962	0,81*	0,86*	0,339	2361	13742	0,78*	3,86
12	-0,113	2752	20701	0,05*	1,28*	0,102	3365	20429	0,97*	1,17*	0,003	1667	12999	0,95*	0,37*
13	0,1141	3604	21430	0,50*	1,30*	0,071	3988	20258	0,08*	0,80*	0,485	2541	14081	0,17*	5,52
14	0,191	4180	20596	0,76*	2,17	-0,216	1928	22262	1,57*	2,45	0,261	2107	14082	0,08*	2,97
15	-0,050	2507	22526	0,10*	0,54*	-0,011	3199	20992	0,28*	0,13*	0,140	1969	13490	0,10*	1,59*

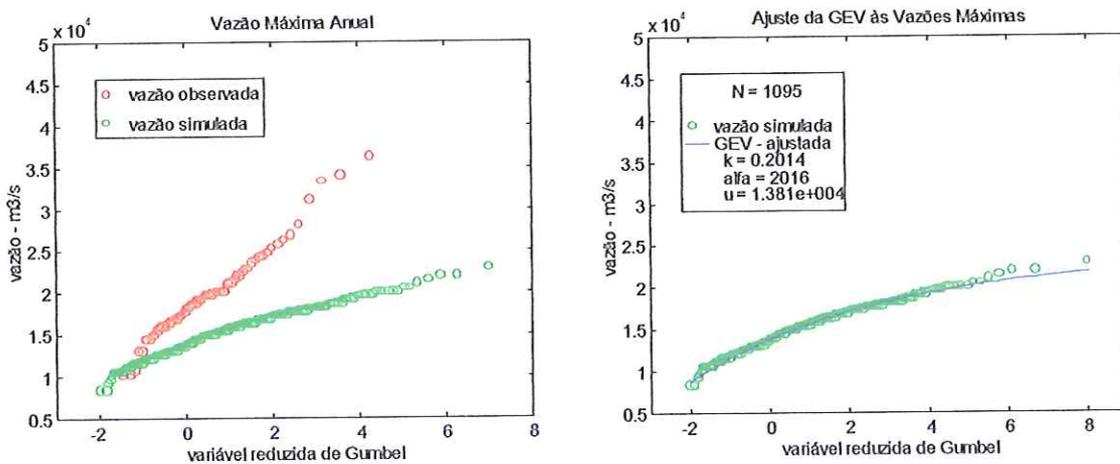
Z<sub>W</sub> – Teste WANG, Z<sub>H</sub> – Teste de HOSKING, \* significativo ao nível de 5% (Z<sub>0,025</sub> = 1,96)



a) vazões geradas utilizando o modelo de distribuição limitada de pulsos



b) vazões geradas utilizando o modelo de distribuição exponencial de pulsos



c) vazões geradas utilizando o modelo de distribuição exponencial modificada de pulsos

FIGURA 53 - Distribuição das vazões máximas observadas, simuladas e dos respectivos ajustes da GEV, referentes aos três modelos de pulsos de vazões.

distribuição de valores extremos com o processo de modelação. A qualidade dos ajustes é altamente significativa, como revela os resultado do teste de WANG.

Os resultado do teste sobre a forma da distribuição ajustada mostram que, das 15 séries de cheias anuais correspondentes a geração de pulsos com a distribuição limitada, apenas duas são estatisticamente diferentes da distribuição Gumbel. O mesmo número se repete para o modelo de pulsos com distribuição exponencial. No caso da distribuição exponencial modificada, 10 séries se apresentam diferentes da Gumbel.

É importante observar que em amostras com 73 anos, o parâmetro de forma só é estatisticamente diferente de zero para valores de  $k$  superiores a 0,1722 ou inferiores a  $-0,1722$ . Isto significa que em análise de frequência descritiva, utilizando amostras com 73 anos, a distribuição GEV com parâmetro de forma entre  $-0,1722$  e  $0,1722$  não difere da distribuição Gumbel ( $k = 0$ ) com nível de significância estatística de 5 %, segundo o teste de HOSKING. Do ponto de vista da análise de frequência preditiva, esses resultados não se aplicam, pois o teste se baseia na assintoticidade dos dados.

Os gráfico da Figura 53 demostram o efeito do tamanho da amostra na forma da distribuição dos valores extremos, evidenciando a convergência assintótica das 15 séries de cheias anuais à uma forma de valores extremos, específica ao tipo de distribuição da variável inicial, quando o tamanho da amostra atinge 1095 valores.

Observa-se que a séries de valores extremos com 1095 anos relacionada ao modelo de geração de pulsos com a distribuição limitada, resultou na forma assintótica da distribuição GEV limitada. Enquanto a série de cheias com 1095 anos referente ao modelo de geração de pulsos através da distribuição exponencial, resultou na forma assintótica equivalente a distribuição Gumbel. Tal distribuição tem como subjacente, a distribuição exponencial.

Na Figura 54, comparam-se as distribuições ajustadas às séries de valores máximos, correspondentes aos modelos de pulsos e aos valores máximos observados, plotadas juntamente com os valores da série histórica.

Os resultados dos testes aplicados para avaliar a qualidade de ajuste e a forma da distribuição ajustada; as estimativas dos parâmetros da GEV e os valores de quantis de cheias de 10000 anos e de período de retorno infinito, estimados pelas referidas distribuições, encontram-se relacionados na Tabela 17.

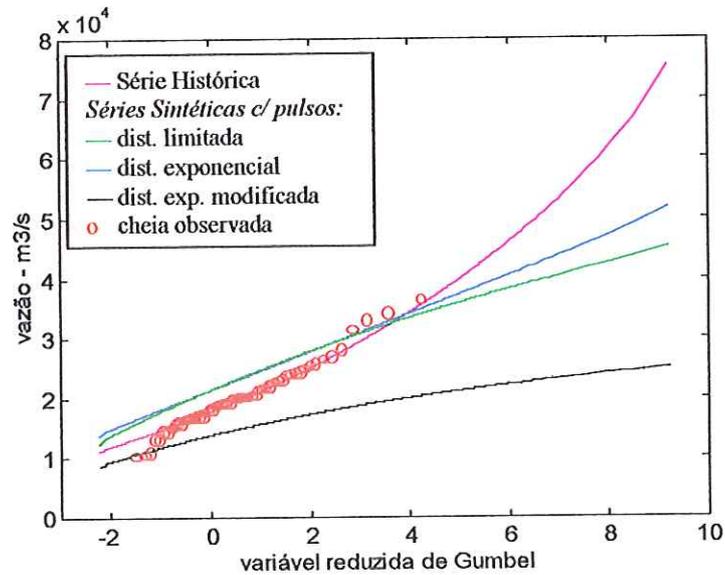


FIGURA 54 - Ajustes da GEV às séries de vazões máximas históricas e simuladas com distribuição limitada, exponencial e exponencial modificada, para os pulsos de vazões.

TABELA 17- Resultados do ajuste da distribuição GEV às séries de vazões máximas observadas e simuladas através dos três modelos de pulsos de vazões.

Série de Vazão	Parâmetros da GEV			Testes de ajuste		Quantis de cheia	
	$k$	$\alpha$	$U$	$Z_{Wang}$	$Z_{Hosking}$	$Q_{10000}$	$Q_{\infty}$
Histórica	-0,118	3218	17731	0,079	-4,970	71314	$\infty$
Sintética: <i>Limitada</i>	0,076	3486	21380	0,747	3,349	44470	67248
<i>Exponencial</i>	0,002	3201	21340	1,450	0,101*	50553	$\infty$
<i>Exp. modific.</i>	0,201	2016	13810	0,510	8,878	22926	25429

\* significativo ao nível de 5 % ( $Z_{0,025} = 1,96$ )

Os valores do teste de qualidade de ajuste mostram que as séries de máximos foram ajustada adequadamente através da GEV, com nível de significância de 5% de probabilidade. Quanto à forma das distribuições ajustadas, o teste de HOSKING aceita a hipótese de  $k = 0$  com nível de significância de 5 % de probabilidade, apenas, para o modelo de distribuição de pulsos exponencial, indicando convergência para a forma assintótica VEI, que corresponde a distribuição Gumbel.

Assim, o modelo proposto produziu a série com 1095 valores máximos de vazões, que quando ajustada à distribuição GEV, resultou num parâmetro de forma  $k=0,076$  estatisticamente diferente de zero, convergindo assintótica para a forma VEIII, que refere-se a distribuição GEV limitada.

A influência da forma da distribuição na estimação dos altos quantis de cheia é verificada através da Figura 54 e dos valores de cheia decamilenares descritos na Tabela 17. O ajuste da distribuição amostral, com  $k < 0$ , e o ajuste referente a distribuição de pulsos exponencial, com  $k \cong 0$ , fornecem valores de  $Q_{10000}$  bem mais elevados de que o ajuste referente ao modelo de distribuição de pulsos limitada, que apresentou  $k > 0$ . Embora, as estimativas de  $Q_{10000}$  fornecidas pelos ajustes correspondentes aos modelos de distribuição de pulsos limitada e exponencial apresentem valores bem próximos, a grande diferença refere-se ao significado físico dos dois resultados. Onde, um provem da distribuição que converge para um valor superior finito, sendo condizente com a realidade física dos processos hidrológicos naturais, enquanto o outro, vem de uma distribuição sem limite superior que não condiz com a realidade.

Os gráficos da Figura 55 ilustram o comportamento dos ajustes da GEV às séries com  $N= 20, 35, 50, 100, 500$  e 1095 primeiros valores da referida série de valores máximos. Os gráficos desta figura mostram que, para as séries com  $N \leq 100$  ocorrem as 3 formas da distribuição GEV, diferindo-se da forma verdadeira. Isto mostra que o efeito do tamanho da amostra sobre a qualidade dos ajustes e as incertezas da estimação da distribuição ocorrem em amostra pequenas, com menos de 100 anos, independente da distribuição subjacente. Além disso, a convergência assintótica só acontece quando a mostra torna-se superior a 500 anos.

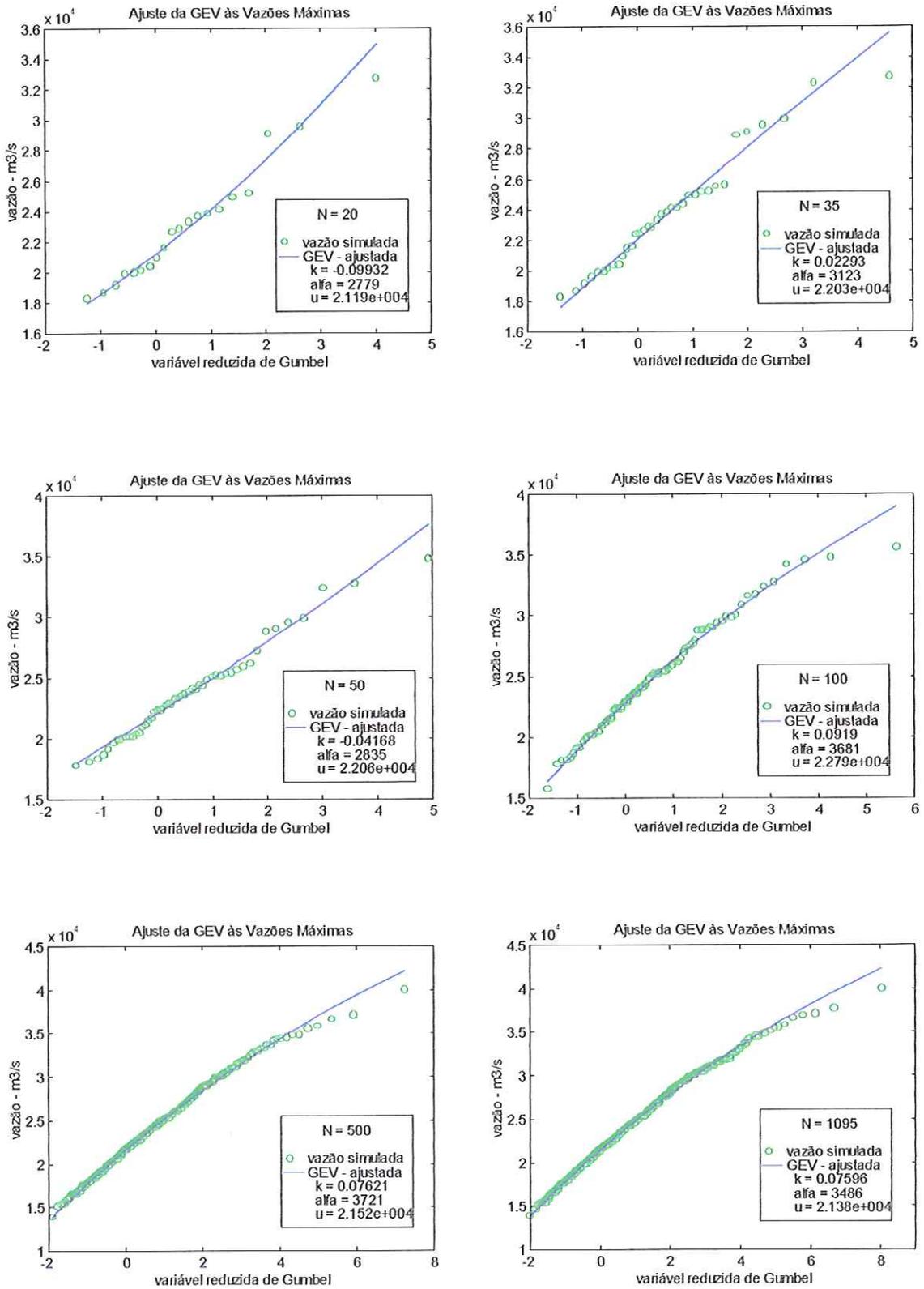


FIGURA 55 - Ajuste da distribuição GEV às séries de vazões máximas formadas pelos primeiros 20, 35, 50, 100, 500 e 1095 valores da série total de vazões geradas através do modelo proposto.

## 7 CONCLUSÕES

O modelo de distribuição Generalizada de Valores Extremos - GEV tem encontrado muitas aplicações em hidrologia, com crescente aceitação para descrever eventos máximos naturais, principalmente, dados de cheias máximas. Eventos máximos de variáveis hidrológicas, como cheias anuais, ocorrem na natureza numa faixa de valores fisicamente limitada. A distribuição dos valores máximos de vazões anuais é, apropriadamente, descrita através da distribuição GEV limitada, para seqüências com longos períodos de observações. Neste estudo, propôs-se investigar o comportamento do ajuste da distribuição GEV, às incertezas da estimação de seus parâmetros e altos quantis de cheias, sob o efeito de diferentes tamanhos e características das séries de cheias anuais; verificar a forma assintótica dos valores de vazões máximas provenientes da geração de vazões diárias através de alguns modelos estocásticos diários usados na hidrologia operacional e propor um modelo de simulação de vazões diárias que gere valores máximos com distribuição assintótica limitada.

O estudo sobre a estimabilidade da distribuição GEV procedeu-se com base na relação teórica existente entre a distribuição GEV limitada e sua subjacente, a distribuição inicial limitada, definida por meio de seus parâmetros. Para realização do estudo, utilizou-se séries sintéticas de cheias anuais com tamanhos variando de 20 a 1000 anos, obtidas através de simulação Monte Carlo de vazões diárias, usando o modelo de distribuição limitada como gerador de vazões diárias, para as diferentes combinações de valores dos parâmetros  $k$ ,  $\alpha$  e  $u$  comuns às duas distribuições. Todas as séries de cheias anuais foram submetidas ao ajuste da distribuição GEV através de momentos  $L$ .

Para verificar a forma da distribuição dos valores de vazões máximas provenientes da simulação estocástica diária com modelos usados na hidrologia operacional, ajustaram-se os modelos Diana, Shot-Noise, Modelo Simplificado de

Desagregação e Modelo de Sargent às séries de vazões diárias observadas no rio Paraná e no rio dos Patos e geraram-se séries de vazões máximas com 1000 valores.

Propõe-se um modelo de simulação de vazões diárias com base no modelo de SARGENT (1979), o qual utiliza a distribuição de probabilidade limitada na modelação dos pulsos de vazões de cada mês, onde a seqüência de dias que devem ocorrer ou não pulsos é simulada através de uma cadeia de Markov de dois estados, com base nas probabilidades de transição dia seco – dia chuvoso.

Os resultados da estimação dos parâmetros  $k$ ,  $\alpha$  e  $u$  sobre as diferentes amostras ocorrem simetricamente em torno do valor real do respectivo parâmetro, independente do efeito do tamanho da amostra e da variação de cada um deles. As estimativas dos três parâmetros mostram que a dispersão de seus valores, em relação ao valor real, aumenta consideravelmente quando o tamanho da amostra diminui. Este feito é mais marcante em amostras retiradas de uma população com distribuição Gumbel ou GEV limitada com parâmetro de forma inferior a 0,05.

A comparação dos resultados da estimação de  $k$  em função dos diferentes valores reais de  $\alpha$  e  $u$  indica que os mesmos não afetam a sua estimação. Mostra, ainda, que para amostras com 20 valores, provenientes de uma distribuição Gumbel ou GEV limitada, tanto ocorrem valores de  $k$  positivos como negativos, além dos nulos. Isto acontece, também, com amostras com  $20 < N < 100$  anos retiradas da Gumbel ou GEV limitada com  $k \leq 0,20$ .

Com base nas estimativas de  $\alpha$  pode-se concluir que, além do efeito do tamanho da amostra, a variabilidade dos dados amostrais afeta sua estimação. A magnitude dos dados expressa pelo valor de  $u$  (tamanho do rio) e a forma da distribuição de valores extremos, indicado pelo sinal de  $k$ , não intervém nos resultados da estimação de  $\alpha$ .

Os resultados da estimação de  $u$  revelam que a variabilidade amostral afeta diretamente sua estimativa. A magnitude das vazões, assim como o tamanho da amostra, influenciam na estimativa de  $u$ , à medida que seus valores reais diminuem (rios menores). A forma da distribuição não afeta os resultados da estimação de  $u$ .

Os resultados da estimação de  $k$ ,  $\alpha$  e  $u$  permitem concluir que para amostras com menos de 100 anos, retiradas de uma população com distribuição Gumbel ou GEV limitada com  $k$  inferior a 0,2, pode ocorrer qualquer tipo de distribuição de

valores extremos, com a modelação da distribuição GEV. A precisão da estimação dos mesmos é mais fortemente afetada com o número reduzido de dados amostrais do que com a variação de um ou outro parâmetro.

As estimativas da cheia decamilenar retratam o efeito marcante dos valores negativos e nulo do parâmetro de forma na sua estimação, sendo, nestas circunstâncias, também, fortemente afetada com a variação de  $\alpha$  e  $u$ . Os resultados revelam que em amostras com menos de 100 anos de dados, de uma distribuição Gumbel ou GEV limitada com  $k \leq 0,2$ , as cheias decamilenares são superestimadas com maior frequência do que subestimadas. A superestimação da cheia decamilenar pode ser extremamente elevada em modelação de séries de cheias anuais composta com baixo número de dados,  $N \leq 50$  anos, advindas de uma distribuição Gumbel ou GEV limitada com  $k \leq 0,1$ .

Os resultados da raiz do erro médio quadrático percentual referentes aos parâmetros  $k$ ,  $\alpha$  e  $u$  e à correspondente cheia decamilenar mostram que a incerteza de estimação do parâmetro de forma, extremamente elevada para valores de  $k$  próximos de zero, é 10 vezes maior do que a do parâmetro de escala e 100 vezes superior ao parâmetro de posicionamento. Tais valores resultam na incerteza de estimação da cheia decamilenar similar à do parâmetro  $\alpha$ . Onde pode-se concluir, que as altas incertezas da estimação da cheia decamilenar são decorrentes das incertezas presentes no parâmetro de forma e daquelas referentes aos parâmetros  $\alpha$  e  $u$ , quando  $k$  é negativo.

A comparação dos diagramas de momentos L das cheias simuladas com aquelas observadas em várias partes do mundo, mostra que amostras curtas,  $N < 100$  anos, retiradas de uma população com distribuição GEV limitada, assintótica, apresentam assimetria e curtosis similares as das vazões observadas nos vários rios. Isto indica que a distribuição assintótica de cheias anuais observadas em diferentes rios, pode ser adequadamente descrita pela distribuição GEV limitada. Esta comparação, retrata, também, o efeito do tamanho da amostra na qualidade do ajuste.

Os resultados da estimação dos parâmetros  $k$ ,  $\alpha$  e  $u$  através de momentos L e LH renderam bons ajustes da GEV às séries de cheias utilizadas. Os resultados do teste de qualidade de ajuste proposto por WANG (1998) comparados aos resultados

gráficos, mostram que o teste foi eficiente na determinação da melhor qualidade de ajuste entre as diferentes ordens de momentos LH.

Os resultados dos ajustes da distribuição GEV às séries de cheias anuais com diferentes tamanhos, mostram que a presença de outliers ocorre, apenas, em amostras curtas, provenientes de uma distribuição Gumbel ou GEV limitada com  $k \leq 0,10$ . Com base nos ajustes realizados, conclui-se que a ocorrência de valores atípicos é resultante de um problema de posição de plotagem do referido dado, comum em amostras pequenas. Nestes casos, a fórmula de posição de plotagem APL aproxima a distribuição ajustada aos dados amostrais mais adequadamente.

Os ajustes realizados através de momentos LH sobre as séries sintéticas de cheias anuais, com diferentes tamanhos, referentes aos diversos valores de  $k$ , revelam que, em amostras com menos de 100 anos, retiradas de uma Gumbel ou GEV limitada com  $k \leq 0,20$ , podem ocorrer qualquer uma das três formas assintóticas de valores extremos com o ajuste da distribuição GEV, onde, em qualquer caso, a distribuição ajustada difere da verdadeira distribuição. Tais ajustes mostram ainda, que a convergência à forma assintótica da distribuição teórica ocorre, de modo geral, com o uso de amostras com 1000 valores. Em alguns casos, tal convergência acontece em seqüência de cheias anuais, a partir de 500 anos.

Os valores extremos máximos de longas seqüências estocásticas de vazões diárias, geradas através dos modelos Diana, Shot-Noise, modelo simplificado de desagregação e modelo de Sargent, ajustados aos dados de vazões diárias dos rios dos Patos e Paraná, apresentam a forma assintótica da distribuição Gumbel.

Os resultados da modelação de picos de vazões mostram que a distribuição inicial limitada descreve eficientemente as seqüências de picos de cheias de todos os meses do ano. Em que os momentos de pesos probabilísticos da distribuição limitada fornecem estimativas dos parâmetros da referida distribuição, adequadamente.

A análise comparativa dos resultados da simulação de vazões diárias com o modelo proposto, indica que o novo método reproduz as estatísticas da série histórica de vazões do rio Paraná nos níveis diário, mensal e anual, bem como os valores extremos de vazões, satisfatoriamente. Os resultados dos testes aplicados confirmam a superioridade do modelo em simular vazões diárias, quando utiliza a distribuição limitada para modelar os picos de cheias em cada mês. Os valores máximos das

---

vazões diárias geradas com o novo modelo, apresentam convergência à forma assintótica da distribuição GEV limitada.

Os resultados obtidos neste estudo satisfazem aos objetivos propostos, de forma plena. No entanto, a realização de uma análise sobre as incertezas que podem ocorrer com a geração de séries sintéticas de vazões de longas duração, decorrente da influência do tamanho da série de vazões diária, utilizada na calibração do modelo estocástico diário, complementaria o presente estudo.

As incertezas da estimação paramétrica dos modelos utilizados na simulação de vazões estão sempre presentes no processo de ajuste, independente do método de estimação adotado, decorrentes dos erros presentes nos dados amostrais e da disponibilidade de dados. Quando, num processo de ajuste, a calibração de um modelo resulta num gerador, que difere do gerador natural, as séries sintéticas obtidas poderão não conter toda a gama de seqüência de vazões possíveis de ocorrer na secção do rio estudado, e mais do que isso, não conter as verdadeiras probabilidades de ocorrência. A idéia é investigar o efeito das diferentes extensões de dados sobre a capacidade do modelo ajustado em reproduzir as possibilidades de ocorrências das seqüências de vazões diárias naturais. Além disso, desenvolver um procedimento que incorpore nos modelos de geração de vazões o efeito da incerteza paramétrica, que possibilite a modelação da vazões com variância apropriada.

## 8 REFERÊNCIAS BIBLIOGRÁFICAS

- ANG, A. H-S.; TANG, W. H. (1984). *Probability Concepts in Engineering Planning and Design*. Vol. II, Decision, Risk, and Reliability, John Wiley & Sons, New York, 562p.
- BALAKRISHNAN, N.; COHEN, A. C., (1991). *Order Statistics and Inference Estimation Methods*, Academic Press, INC. Boston.
- BARNETT, V.; LEWIS, T. (1978). *Outliers in Statistical Data*, Wiley, New York, N.Y. 378pp.
- BARRICELLI, N. A. (1943). Les plus grands et les petits maxima ou minima annuels d'une variable climatique. *Arch. Math. Naturv.*, Oslo, 46, n.6.
- BARTOLONI, P.; SALAS, D. (1993). Modeling of streamflow process at different time scales. *Water Resour. Res.*, v.29, n.8, p.2576-2587.
- BEARD, L. M. (1967). Simulation of daily streamflow proceedings. In: *The Inter. Hydrol. Symp.*, Sept. 6-8, 1967, Fort Collins, Colorado, v.1, paper n.78, June, 1967, Fort Collins, p.624-632.
- BEARD, L. R., (1974). *Flood Flow Frequency Techniques*, Technical Repot 119, Center of Research in Water Resources, The Univ. of Texas at Austin.
- BENSON, L. R., (1968). Uniform flood - frequency estimating methods for federal agencies, *Water Resour. Res.*, v.4, n.5, p.891-908.

- BOBÉE, B.; RASMUSSEN, P. F., (1995) Recent advances in flood frequency analysis, U.S. Natl. Rep. Int. Union Geol. Geophys. 1991-1994, Ver. *Geophys*, v.33, p.111-1116.
- BOBÉE, B. et al. (1993). Towards a systematic approach to comparing distributions used in flood frequency analysis. *J. of Hydrology*, v.142, p.121-136.
- BROOKS, C. E. P.; CARRUTHERS, N. (1953). *Handbook of Statistical Methods in Meteorology*, London (H.M.S.O.), p.131-134.
- CHAN, L.K. (1967). On characterization of distributions by expected values of extreme order statistics. *Am. Math. Mthly*, v.74, p. 950-951.
- CHOW, V. T. (1954). The log-probability law and its engineering applications. *Proc. ASCE*, v.80, n.5, p.1-25.
- CHOW, V. T.; MAIDMENT, D. R.; MAYS, L. W. (1988). *Applied Hydrology*. McGraw-Hill, Water Resources, 583p.
- CHOWDHURY, J. U.; STEDINGER, J. R.; LU, L. (1991). Goodness-of-fit test for regional generalized extreme value flood distributions. *Water Resour. Res.*, v.27, n.7, p.1765-1776.
- CICIONI, G.; GUILIANO, G.; SPAZIANI, F.M., (1973). Best fitting of probability functions to a set of data for flood studies. Floods and Droughts, In: *Proc. 2nd Int. Symp. in Hydrol.* 11- 13 September 1972, Fort Collins, CO. Water Resour. Publ., Fort Collins, CO, pp. 304 –314.
- CHOWDHURY, J. U., STEDINGER, J. R.; LU, L.-H. (1991). Goodness of fit test for regional generalized extreme value flood distributions. *Water Resour. Res.*, v.27, n.7, p.1765-1776.
- CLARKE, R.T. (1994). *Statistical modeling in hydrology*. John Wiley & Sons, 412p.

- CONWAY, K. M. (1971). *Flood frequency analysis of some New South Wales coastal rivers*. University of New South Wales, Sydney, N.S.W.
- CRAMER, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press, Princeton, N. J.
- CROWDER, M. J. et al. (1991). *Statistical Analysis of Reliability Data*. Chapman and Hall, London.
- CUNNANE, C. (1988). Methods and merits of regional flood frequency analysis. *J. of Hydrology*, v.100, n.1/3, p 269-290.
- DAVID, H. A., (1981). *Order Statistics*. 2nd ed. New York: Wiley.
- DOWNTON, F. (1966). Linear estimates with polynomial coefficients. *Biometrika*, v.53, p.129-141.
- DUPUIS, D. J. (1997). Extreme value theory based on the r largest annual events: a robust approach. *J. Hydrology*, v.200, p.295-306.
- ELETROBRÁS, CENTRAIS ELETRICAS BRASILEIRAS SA (1987). *Guia para cálculo de cheia de projeto de vertedores*. Departamento de Planejamento e Engenharia, Rio de Janeiro.
- EPSTEIN, B. (1960). Elements of the theory of extreme values. *Technometrics*, v.2, n.1, p.27- 41.
- FIERING, M. B. (1961). Queuing theory and simulation in reservoir design. Proc. Am. Soc. Civ. Eng., *J. Hydraulic Div.*, 87(HY6), p.39-69.
- FISHER, R. A.; TIPPETT, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, v.24, p.180-190.

- FOSTER, H. A. (1924). Theoretical frequency curves and their application to engineering problems. *American Society of Civil Engineers*, vol. 87, p.142-173.
- GARCIA, L.A.V. de; BRAGA JR., B. P. F. (1997). Análise de frequência conjugada. *Revista Brasileira de Recursos Hídricos*, ABRH, v.2, n.2, p.21-44.
- GNEDENKO, B.V. (1943). Sur la distribution limite du terme maximum d'une serie Aleatories. *Ann. Math.*, v.44, p.423-453.
- GREEN, N. M. D. (1973), A synhetic model for daily streamflow. *J. of Hydrology*, v.20, p.351-364.
- GREWOOD, J.A. et al. (1979). Probability weighted moments: definition and relation to parameters of distributions expressible in inverse form. *Water Resour. Res.*, v.15, n.5, p.1049-1054.
- GREIS, N. P.; WOOD, E. F. (1981). Regional flood frequency estimation and network design. *Water resour. Res.*, v.17, n.4, p.1167-1177.
- GRYGIER, J.C.; STEDINGER, J.R. (1988). Condensed procedures and conservation correctins for stochastic hydrology. *Water Resour. Res.*,v.24, n.10, p.1547-1584.
- GUMBEL, E. J. (1935). *Les valeurs extrêmes des distributions statistiques*. Ann. l'Inst. Henri Poincaré, p.4-115.
- GUMBEL, E. J. (1941). The return period of flood flows. *Ann Math. Stat.*, v.12, p.163-190.
- GUMBEL, E. J. (1958). *Statistics of Extreme*. Columbia University Press, New York, 396p.
- GUNASEKARA, T. A .G.; CUNNANE, C. (1992). Splait sampling technique for selecting a flood frequency analysis procedure. *J. Hydrology*, v.130, p.189-200.

- GUTTMAN, N.B. (1996). Statistical characteristics of US historical climatology network temperature distribution. *Clim. Res.*, v.6, n.1, p.33-43.
- HAAN, C. T., (1977). *Statistical Methods in Hydrology*. The Iowa State University Publication, 378p.
- HAKTANIR, T.; BOZDUMAN, A. (1995). A study on sensitivity of the probability Weighted moments method on the choice of the plotting position formula. *J. Hydrology.*, v.168, p.265-281.
- HAKTANIR, T. (1992). Comparison of various flood frequency distributions using flood peaks data of rivers in Anatolia. *J. of Hydrology*, v.136, p.1-31.
- HALTER, A. N.; MILLER, S. F. (1967). *River basin planning: a simulation approach*. Oregon State Agricultural Experiment Station, Corvallis, Oregon.
- HAZEN, A. (1914). Storage to be provided in impounding reservoirs for municipal water supply. *Trans. ASCE*, v.1308, n.77, p.1547-1550.
- HARMS, A.A.; CAMPBELL, T.H. (1967). Na extension to the Thomas - Fiering model for the sequential generation of streamflow. *Water Resour. Res.*, v. 3, n.3, p.653-661.
- HORTON, R. E. (1913). Frequency of recurrence of Hudson River floods. U.S. *Weather Bur. Bull.*, v.2, p.109-112.
- HOSKING, J. R. M. (1985). Algorithm AS215: maximum-likelihood estimation of the parameter of the generalized extrem - value distribution. *Appl. Stat.*, v.34, p.301-310.
- HOSKING, J. R. M. (1990). L - moments: analysis and estimation of distributions using linear combinations of order statistics. *J. R. Statist. Soc, Ser. B*, v.52, n.2, p.105-124.

- HOSKING, J.R.M.; WALLIS, J.R. (1986). The value of historical data in flood frequency analysis. *Water Resour. Res.*, v.22, n.11, p.1606-1612.
- HOSKING, J.R.M.; WALLIS, J.R. (1988). The effect of intersite dependence on regional frequency analysis. *Water Resour. Res.*, v.24, n.1, p.588-600.
- HOSKING, J.R.M.; WALLIS, J. R. (1993). Some statistic useful in regional frequency analysis. *Water Resour. Res.*, v.29, n.2, p.271-281.
- HOSKING, J.R.M.; WALLIS, J.R. (1995). A comparison of unbiased and plotting-position estimators of L moments. *Water Resour. Res.*, v.31, n.8, p.2019-2025.
- HOSKING, J.R.M.; WALLIS, J.R.; WOOD, E.F. (1985a). Estimation of the generalized extreme value distribution by the method of probability weighted moments. *Technometrics*, v.27, n.3, p.251-261.
- HOSKING, J. R. M.; WALLIS, J. R.; WOOD, E. F. (1985b). An appraisal of the regional flood frequency procedure in the UK Flood Studies Report, *Hydrol. Sci. J.*, v.30, n.1 p.85-109.
- HOYBYE, J.; IRITZ, L. (1997). Analysis of extreme hydrological events in a monsoon climate catchment: the Hongru River, China, *Hidrol. Scien. Jour Des Sciences Hydrologiques*, v.42, n.3, p.343-356.
- HOUGHTON, J. C. (1978). Birth of a parent: the Wakeby distribution for modeling flood flows. *Water Resour. Res.*, v.14, n.6, p.1105 – 1110.
- INSTITUTION OF ENGINEERS AUSTRALIAN – I.E.A. (1977). *Australian rainfall and runoff: Flood analysis and design*. Inst. Eng., Australia, 149p.
- JENKINSON, A. F. (1955). The frequency distribution of the annual maximum (or minimum) of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, v.81, p.158-171.

- KELMAN, J. (1980). A stochastic model for daily streamflow. *J. of Hydrology*, v.47, p.235-249.
- KELMAN, J.; DAMÁZIO, J.M.; COSTA, J.P. (1983). Geração de séries sintéticas de vazões diárias – modelo Diana. *Revista Bras. de Eng.*, v.1 n.2, p.5-22.
- KELMAN, J. (1987a). Modelos estocásticos para gerenciamento de recursos hídricos. In: *Modelos para gerenciamento de recursos hídricos*. São Paulo: Nobel ABRH, cap.4, p.327-420.
- KELMAN, J. (1987b). *Controle de Cheias e Aproveitamentos Hidrelétricos*. Rio de Janeiro: ABRH, 175p.
- KELMAN, J. (1983). *Cheias e Aproveitamentos Hidrelétricos*. Rio de Janeiro, 1983. 210p. Tese (Prof. Titular) – Universidade Federal do Rio de Janeiro – Escola de Engenharia.
- KIRBY, W. (1974). Algebraic boundedness of sample statistic. *Water Resour. Res.*, v.10, n.2, p.220-222.
- KONHEIM, A. G. (1971). A note on order statistics. *Am. Math. Mthly*, v.78, p.524.
- KOPITTKER, R.A.; STEWART, B.J.; TICKLE, K.S., (1976). Frequency analysis of flood data. In: *Queensland. Hydrol. Symp.*, 1976, Inst. Eng, Aust., Natl. Conf. Publ.n.76/2, p.20-24.
- KOTTEGODA, N. T. (1972). Stochastic five daily stream flow model. *J. of Hydraulics Div.*, HY9, p.1469-1485.
- KOUTSOYIANNIS, D. (1992). A nonlinear disaggregation method with reduced Parameter set for simulation of hydrologic series. *Water Resour. Res.*, v.28, n.12, p.3175-3191.

- KOUTSOYIANNIS, D.; MANETAS, A. (1996). Simple disaggregation by accurate adjusting procedures. *Water Resour. Res.*, v.32, n.7, p.2105-2117.
- KUCZERA, G. (1992). Uncorrelated measurement error in flood frequency inference, *Water Resour. Res.*, v.28, n.1, p.183-188.
- KUCZERA, G. (1996). Correlated rating curve error in flood frequency inference, *Water Resour. Res.*, v.32, n.7, p.2119-2127.
- LANDWEHR, J.M.; MATALAS, N.C.; WALLIS, J.R., (1979). Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. *Water Resour. Res.*, v.15, n.5, p.1055-1064.
- LAWLESS, J. F. (1982). *Statistical Models and Methods for Lifetime Data*. Univ. of Waterloo, John Wiley & Sons, New York. 592p.
- LETTENMAIER, D. P.; WALLIS, J. R.; WOOD, E. F. (1987). Effect of regional heterogeneity on flood frequency estimation. *Water Resour. Res.*, v.23, n.2, p.313-323.
- MADSEN, H.; RASMUSSEN, P.F.; ROSBJERG, D.(1997a). Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events. 1, at-site modeling. *Water Resour. Res.*, v.33, n.4, p.747-758.
- MADSEN, H.; PEARSON, C. P.; ROSBJERG, D. (1997b). Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events. 2, Regional modeling. *Water Resour. Res.*, v.33, n.4, p.759-770.
- MALLOWS, C. L. (1973). Bounds on distribution functions in terms of expectations of order-statistics. *Ann Probab.*, v.1, p.297-303.

- MARTINS, S.E.; STEDINGER, J.R. (2000). Generalized maximum-likelihood generalized extreme - value quantile estimators for hydrologic data. *Water Resour. Res.*, v.36, n.3, p.737-744.
- McMAHON, T. A.; SRIKANTHAN, R. (1981). Log Pearson III distribution - is it applicable to flood frequency analysis of Australian streams? *J. Hydrology*, v.52, p.139-148.
- MEIGH, J.R.; FARQUHARSON, F.A.K.; SUTCLIFFE, J.V. (1997). A worldwide comparison of regional flood estimation methods and climate. *Hidrol. Scien. Jour. Des Sciences Hydrologiques*, v.42, n.2, p.225-244.
- MEJÍA, J. M.; ROUSSELLE, J. (1976). Disaggregation models in hydrology revisited. *Water Resour. Res.*, v.12, n.2, p.185-186.
- MINE, M.R.M. (1990). Geração de vazões médias diárias pelo SOSN. *Revista Bras. de Eng.*, v.8, n.1, p. 51-64.
- MITSIPOULOS, J.; HAIMES, Y.Y.; LI, D. (1991). Approximating catastrophic risk through statistics of extreme. *Water Resour. Res.*, v.27, n.6, p.1223-1230.
- MKHANDI S. H.; KACHROO, R. K.; GUO, S. L. (1996). Uncertainty analysis of flood quantile estimates with reference to Tanzania. *J. Hydrology*, v.185, p.317-333.
- MÜLLER, F.; FILL, H.D. (1997). Quantis de cheias GEV regionais –uma aplicação Prática. *Revista Brasileira de Recursos Hídricos*, ABRH, v.2, n.2, p.107-122.
- MÜLLER, F. (1998). Dois modelos de desagregação de vazões mensais em diárias. *Revista Brasileira de Recursos Hídricos*, ABRH, v.3, n.2, p.23-45.

- NATHAN, R. J.; WEINMANN, P. E. (1991). Application of at-site and regional flood frequency analysis. In: *Proc. Int. Hydrol. Water Resour. Symp.*, Perth, 2-4 October, p.769-774.
- NATIONAL ENVIRONMENT RESEARCH COUNCIL – NERC (1975). *Flood Studies Report*, v.1. London.
- O'CONNELL, P. E.; JONES, D. A. (1978). Some experience with the development of models for the stochastic simulation of daily flows. In: *Risk and Reliability Symposium*, 1978, Waterloo, Canada.
- ÖNÖZ, B.; BAYAZIT, M., (1995). Best - fit distributions of largest available flood samples. *J. Hydrology*, v.167, p.195-208.
- OTTEN, A.; VAN MONFORT, M. A. J. (1980). Maximum-likelihood estimation of the general extreme-value distribution parameters. *J. of Hydrology*, v.47, p.187-192.
- PAYNE, K. et al., (1969). Daily streamflow simulation. *J. of Hydraulics Div.*, HY4, p.1163-1179.
- PEARSON, C. P.; McKERCHAR, A. I.; WOODS, R. A. (1991). Regional flood frequency analysis of Western Australian data using L – moments. In: *Proc. Int. Hydrol. & Water Resour. Symp.*, Perth, 2-4 October 1991, p.631-632.
- PINHEIRO, M. C., DAMAZIO, J. M. KELMAN, J., 1981, Análise de frequência de cheias para as regiões Sul e Sudeste do Brasil, *IV Simpósio Brasileiro de Hidrologia e Recursos Hídricos*, 15 a 19 de Novembro, Fortaleza, Ceará.
- PORTELA, M. M.; SANTOS, E. G. (1989). Geração de séries sintéticas de caudais diários. In: *IV Simpósio Luso-Brasileiro*. Lisboa.

- POTTER, K. W.; WALKER, J. F. (1981). A model of discontinuous measurement error and its effects on the probability distribution of flood discharge measurements. *Water Resour. Res.*, v.17, n.5, p.1505-1509.
- POTTER, K. W.; WALKER, J. F. (1985). An empirical study of flood measurement Error. *Water Resour. Res.*, v.21, n.3, p.403-406.
- PRESCOTT, P.; WALDEN, A. T. (1980). Maximum - likelihood estimation of the parameters of the generalized extreme - value distribution. *Biometrika*, v.67, p.723-724.
- QUIMPO, R. H. (1968). Stochastic analysis of daily river flows. *J. of Hydraulics Div.*, HY1, p.43-57.
- REIGIER, M.H.; HAMDAM, M. A. (1971). Correlation in bivariate normal distribution with truncation in both variables. *Austral. J. of Statist.*, v.13, n.2, p.77-82.
- ROSSI, F.; FIORENTINO, M.; VERSACE, P. (1984). Two component extreme value distribution for flood frequency analysis. *Water Resour. Res.*, v.20, n.7, p.847-856.
- SALAS et al. (1980). *Applied modeling of hydrologic time series*. Colorado: Water Resourcer Publ., 484p.
- SANTOS, E. G.; SALAS, J. D. (1992). Stepwise disaggregation scheme for synthetic hydrology. *J. of Hydraulic Eng.*, v.118, n.5, p.765-784.
- SARGENT, D.M.A. (1979). A simplified model for the generation of daily stream-flows. *Hydrological Sciences-Bulletin – des Sciences Hydrologiques*, v.24, n.4, p.509-527.

- SILLITTO, G. P. (1969). Derivation of approximates to the inverse distribution function of a continuous univariate population from the order statistics of a sample. *Biometrika*, v.56, p.641-650.
- SRIKANTHAN, R. (1978). Sequential generation of monthly streamflows. *J. of Hydrology*, v.38, p.71-80.
- STEDINGER, J.R.; VOGEL, R.M.; FOUFOULA-GEORGIU, E. (1993). Frequency analysis of extreme events. In: *Handbook of Hydrology*, p.18.1-18.66, McGraw-Hill, New York.
- STEDINGER, J. R.; PEI, D.; COHN, T. A. (1985). A condensed disaggregation model for incorporating parameter uncertainty into monthly reservoir simulation. *Water Resour. Res.*, v.21, n.5, p.665-675.
- SVANIDZE, G. G. (1980). *Mathematical modeling of hydrologic series: for hydroelectric and water resources computations*. Fort Collins, Colorado: Water Resources Publ., 314p.
- THOMAS, H.A.; FIERING, M. B. (1962). Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation. In: *Design of Water Resource Systems*, Chapter 12. MAASS, et al., Harvard University Press, Cambridge, Massachusetts.
- TAO, P.C.; DELLEUR, J.W.(1976). Multistation, multiyear synthesis of hydrologic time series by disaggregation. *Water Resour. Res.*, v.12, n.6, p.1303-1312.
- TODOROVIC, P.; WOOLHISER, D. A. (1975). A stochastic model of on-day Precipitation. *Journal of Applied Meteorology*, v.14.

- TREIBER, B.; PLATE, E. S. (1975). A stochastic model for the simulation of daily flows. Paper presented at the *Workshop on the Application of Mathematical Models in Hydrology and Water Resources Systems*, Bratislava, Czechoslovakia, September 1975.
- TUCCI, C. E. M. (1997), Vazão máxima e hidrograma de projeto. In: *Hidrologia – Ciência e Aplicação*. Porto Alegre: Editora da Universidade: ABRH. Cap. 14, p.527 – 572.
- UNIVERSIDADE DE SÃO PAULO. Escola de Engenharia de São Carlos. Serviço de Biblioteca (1996). *Diretrizes para elaboração de dissertações e teses na EESC-USP*. São Carlos.
- VALENCIA, D. R.; SCHAAKE JR, J.L. (1972). Disaggregation process in stochastic hydrology. *Water Resour. Res.*, v.9, n.3, p.580-585.
- VOGEL, R. M.; FENNESSY, N. M. (1993). L moment diagrams should replace product moment diagrams, *Water Resour. Res.*, v.29, n.6, p.1745-1752.
- VOGEL, R.M.; McMAHON, T.A.; CHIEW, F.H.S. (1993b). Floodflow frequency model selection in Australia. *J. Hydrology*, v.146, p.421-449.
- VOGEL, R.M.; THOMAS, W.; McMAHON, T.A. (1993a). Floodflow frequency model selection in Southwestern U.S.A. *J. Water Resour. Plann. Manage., Div. Am. Soc. Civ. Eng.*, v.119, n.3, p.353-366.
- VOGEL, R.M.; WILSON, I. (1996). Probability distribution of annual maximum, mean, and minimum streamflows in the United State. *J. Hydrol. Eng. ASCE*, v.2, n.1, p.69-76.
- WALLIS, J.R.; WOOD, E.F.(1985). Relative accuracy of log Pearson III procedures. *J. Hydraul. Eng.*, v.111, n.7, p.1043-1056.

- WANG, Q.J. (1990a). Estimation of the GEV distribution from censored samples by method of partial probability weighted moments. *J. Hydrology*, v.120, p.103-114.
- WANG, Q. J. (1990b). Unbiased estimation of probability weighted moments and partial probability weighted moments from systematic and historical flood information and their application to estimating the GEV distribution. *J. Hydrology*, v.120, p.115-124.
- WANG, Q. J. (1996). Using partial probability-weighted moments to fit the extreme value distribution to censored samples. *Water Resour. Res.*, v.32, n.6, p.1767-1772.
- WANG, Q. J. (1997a). Using higher probability weighted moments for flood frequency analysis. *J. Hydrology*, v.194, p.95-106.
- WANG, Q. J. (1997b). LH moments for statistical analysis of extreme events. *Water Resour. Res.*, v.33, n.12, p.2841-2848.
- WANG, Q. J. (1998). Approximate goodness-of-fit test of fitted generalized extreme value distribution using LH moments. *Water Resour. Res.*, v.34, n.12, p.3497-3502.
- WEISS, G. (1977). Shot noise models for the generation of synthetic streamflow data. *Water Resour. Res.*, v.13, n.1, p.101-108.
- WORLD METEOROLOGICAL ORGANIZATION – WMO (1973). *Estimation of Probable Maximum Precipitation*. n.332, Secretariat of the WMO, Geneva.
- U.S. WATER RESOURCES COUNCIL - USWRC (1981). *Guidelines for determining flood flow frequency*. U.S. Water Resour. Counc. Bull. 17B, 222pp.
- U.S. WATER RESOURCES COUNCIL - USWRC (1967). *A uniform technique for determining flood flow frequencies*. Bull. 15 US WRC, Washington, DC, 13pp.

YAKOWITZ, S. J. (1979). A nonparametric Markov model for daily river flow.  
*Water Resour. Res.*, v.15, n.5, p.1035-1043.

YEVJEVICH, V. (1968). Misconceptions in hydrology and their consequences.  
*Water Resour. Res.*, v.4, n.2, p.225-234.

YEVJEVICH, V. (1972). *Probability and Statistics in Hydrology*. Water Resources  
Publication, Fort Collins, 300p.

## 9 APÊNDICES

### 9.1 APÊNDICE I

Para mostrar que  $b_s$ , definido através da equação (41), é um estimador não viesado de  $\beta_s$ , equação (39), considere-se a equação (35) que define o valor esperado da  $i$ -ésima estatística de ordem. Assim, o valor esperado de  $b_s$  pode ser expresso como:

$$E[\beta_s] = \frac{1}{n} \sum_{i=1}^{n-k} \left[ \frac{\binom{n-i}{k}}{\binom{n-1}{k}} \right] E[X_{(i)}]$$
$$E[\beta_s] = \frac{1}{n} \sum_{i=1}^{n-k} \left[ \frac{\binom{n-i}{k}}{\binom{n-1}{k}} \right] \binom{n}{i} \int_0^1 x F^{i-1} (1-F)^{n-1} dF$$

fazendo  $i-1 = j$  note que

$$\frac{1}{n} \left[ \frac{\binom{n-i}{k}}{\binom{n-1}{k}} \right] \binom{n}{i} = \binom{n-1-k}{j}$$
$$E[\beta_s] = \int_0^1 x \left[ \sum_{j=0}^{n-1-k} \binom{n-1-k}{j} F^j (1-F)^{n-j-k-1} \right] (1-F)^k dF$$

observa-se ainda que

$$\sum_{j=0}^{n-1-k} \binom{n-1-k}{j} F^j (1-F)^{n-j-k-1} = [F + (1-F)]^{n-k-1} = 1$$

logo

$$E[\beta_s] = \int_0^1 x (1-F)^k dF = \beta_s$$

De acordo com (39)  $E[\beta_s] = \beta_s$ . Portanto,  $b_s$  é um estimador não viesado de  $\beta_s$ .

9.2 APÊNDICE II

TABELA 18: Valores de coeficiente de correlação  $\rho'$  em uma distribuição normal bivariada com ambas variáveis truncadas  $\beta^*$ .

$\beta$	$\rho$																		
	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50	0,55	0,60	0,65	0,70	0,75	0,80	0,85	0,90	0,95
-2,50	48	96	144	192	240	289	337	386	436	485	535	585	636	687	738	790	842	890	947
-2,00	44	89	134	179	225	271	318	366	414	463	513	564	615	668	721	775	830	886	943
-1,50	39	78	118	159	201	244	287	332	379	427	476	527	580	634	690	749	809	871	935
-1,30	36	73	111	149	189	230	272	316	361	408	457	508	561	616	674	734	797	862	930
-1,10	33	68	103	139	177	215	256	298	342	388	437	487	541	597	656	718	783	852	924
-1,00	32	65	99	134	170	208	248	289	332	378	426	476	530	586	646	709	776	846	921
-0,90	30	62	94	128	164	200	239	280	322	367	415	465	518	575	635	700	768	840	918
-0,80	29	59	90	123	157	193	230	270	312	356	403	453	506	563	624	689	759	834	914
-0,70	28	56	86	118	151	185	222	260	302	345	392	441	494	552	613	679	750	827	910
-0,60	26	54	82	112	144	178	213	251	291	334	380	429	482	539	601	668	741	820	906
-0,50	25	51	78	107	138	170	205	242	281	323	368	417	470	527	589	657	731	812	902
-0,40	24	48	75	102	132	163	196	232	271	312	356	405	457	514	577	645	721	804	897
-0,30	22	46	71	97	126	156	188	223	260	301	345	392	444	501	564	633	710	796	892
-0,20	21	43	67	93	120	149	180	214	250	290	333	380	432	489	551	621	699	787	886
-0,10	20	41	64	88	114	142	172	205	241	279	322	368	419	476	538	609	688	778	881
0,00	19	39	60	84	109	136	165	196	231	269	310	356	406	463	526	596	677	769	875
0,10	18	37	57	79	103	129	157	188	222	258	299	344	396	450	512	584	665	759	869
0,20	17	35	54	75	98	123	150	180	213	248	288	332	381	437	500	571	653	749	868
0,30	16	33	51	72	93	117	143	172	204	239	277	321	369	424	486	558	641	739	858
0,40	15	31	49	68	89	112	137	165	195	229	267	309	357	412	474	545	629	729	849
0,50	14	29	46	64	84	106	131	157	187	220	257	298	345	399	461	532	617	718	842
0,60	13	28	44	61	81	101	125	150	179	211	247	288	334	387	448	520	605	707	835
0,70	12	26	41	58	76	96	119	144	172	203	238	277	323	375	435	507	592	697	827
0,80	12	25	39	55	72	92	113	137	164	194	228	267	312	363	423	494	580	686	820
0,90	11	23	37	52	69	88	108	131	157	187	220	258	301	352	411	482	568	674	812
1,00	10	22	35	49	65	83	103	125	150	179	211	248	291	340	399	469	555	663	804
1,10	9	21	33	47	62	79	98	120	144	172	203	239	280	329	387	457	543	652	795
1,30	8	18	30	42	56	71	90	110	132	158	187	222	261	308	365	433	519	629	778
1,50	8	16	28	38	50	66	81	100	121	145	173	205	244	289	343	410	495	607	761
2,00	0	0	18	33	43	45	71	82	97	118	141	171	202	246	294	356	439	551	716
2,50	0	0	0	0	0	0	0	0	0	61	134	151	176	218	259	311	386	495	673

Os valores de  $\rho'$  são dados com três casas após a vírgula, que está omitida na tabela. \*(REIGIER & HAMDAM, 1971).

### 9.3 APÊNDICE III

9.3.1 Rotina em Matlab que determina os momentos de pesos probabilísticos de uma amostra de dados.

```
function [b,k,alfa,u,]=fmwp(x)
%-----
% esta função calcula os seis primeiros momentos de pesos probabilísticos
% os parâmetros da distribuição GEV
% os parâmetros da distribuição limitada
%-----
N=length(x);

for t=1:N
y=N+1-t;
xw(t)=x(y);
end

for j=1:6
r=j-1;
if r==0
den=1;
elseif r==1
den=N-1;
elseif r==2
den=(N-1)*(N-2)/2;
elseif r==3
den=(N-1)*(N-2)*(N-3)/6;
elseif r==4
den=(N-1)*(N-2)*(N-3)*(N-4)/24;
elseif r==5
den=(N-1)*(N-2)*(N-3)*(N-4)*(N-5)/gamma(5+1);
elseif r==6
den=(N-1)*(N-2)*(N-3)*(N-4)*(N-5)*(N-6)/gamma(6+1);
end
sigma=0;
for i=1:N
U=N-i-r;
if U>=0
if r==0
coeff=1;
elseif r==1
coeff=N-i;
elseif r==2
coeff=(N-i)*(N-i-1)/2;
elseif r==3
coeff=(N-i)*(N-i-1)*(N-i-2)/6;
```

```

elseif r==4
    coeff=(N-i)*(N-i-1)*(N-i-2)*(N-i-3)/24;
elseif r==5
    coeff=(N-i)*(N-i-1)*(N-i-2)*(N-i-3)*(N-i-4)/gamma(5+1);
elseif r==6
    coeff=(N-i)*(N-i-1)*(N-i-2)*(N-i-3)*(N-i-4)*(N-i-5)/gamma(6+1);
end
conf=coeff*xw(i);
sigma=sigma+conf;
end
end
b(j)=(sigma/N)*(1/den);
end

%'Elementos de b'
b';
%'Valores dos 6 primeiros momentos mwp são'
bo=b(1);
b1=b(2);
b2=b(3);
b3=b(4);
b4=b(5);
b5=b(6);
b=[bo b1 b2 b3 b4 b5];

%cálculo dos parâmetros da GEV
c=(2*b1-bo)/(3*b2-bo)-log(2)/log(3);
k=7.8590*c+2.9554*c^2;
alfa=k*(2*b1-bo)/(gamma(1+k)*(1-2^(-k)));
u=bo+(alfa*(gamma(1+k)-1))/k;

n=N;
% cálculo dos parâmetros da distribuição limitada
k=(3*b2+10*b1-8*bo)/(2*(b2-b1));
alfa=(2*b1-bo)*(k+1)*(k+2)/n^k;
u=bo+a*n^k/(k*(k+1))-a/k;

```

### 9.3.2 Rotina em Matlab que determina os momentos L, os parâmetros da distribuição GEV e da Gumbel e as taxas de momentos L amostral

```

function [k,alfa,u,alfag,ug,cv,cas,cut]=momenl(x)
%-----
% esta função calcula os 4 primeiros momentos L
% determina os parâmetros da GEV
% determina os parâmetros da Gumbel
%-----
N=length(x);

```

```

for j=1:5
    r=j-1;
    if r==0
        den=1;
    elseif r==1
        den=N-1;
    elseif r==2
        den=(N-1)*(N-2)/2;
    elseif r==3
        den=(N-1)*(N-2)*(N-3)/6;
    else
        den=(N-1)*(N-2)*(N-3)*(N-4)/24;
    end
    sigma=0;
    for i=1:N
        U=N-i-r;
        if U>=0
            if r==0
                coeff=1;
            elseif r==1
                coeff=N-i;
            elseif r==2
                coeff=(N-i)*(N-i-1)/2;
            elseif r==3
                coeff=(N-i)*(N-i-1)*(N-i-2)/6;
            else
                coeff=(N-i)*(N-i-1)*(N-i-2)*(N-i-3)/24;
            end
            end
            conf=coeff*x(i);
            sigma=sigma+conf;
        end
    end
    a(j)=(sigma/N)*(1/den);
end
%'Elementos de a'
a';
%'Valores dos 4 primeiros momentos-L são'
l1=a(1);
l2=a(1)-2*a(2);
l3=a(1)-6*a(2)+6*a(3);
l4=a(1)-12*a(2)+30*a(3)-20*a(4);
%[l1 l2 l3 l4]
cv=l2/l1; % coef. de variação
cas=l3/l2; % coef. de assimetria
cut=l4/l2; % curtose
%'Coef. de variação, Assimetria e Curtose Amostral'
[cv cas cut];

```

```
%cálculo dos parâmetros da GEV
```

```
c=2*12/(13+3*12)-log(2)/log(3);
k=7.8590*c+2.9554*c^2;
alfa=k*12/(gamma(1+k)*(1-2^(-k)));
u=11+(alfa*(gamma(1+k)-1))/k;
[k alfa u];
```

```
%parâmetros da Gumbel
```

```
alfag=12/log(2);
ug=11-0.5772*alfag;
[alfag ug];
```

9.3.3 Rotina em Matlab que determina os momentos LH, os parâmetros da GEV, da Gumbel e as taxas de momentos LH.

```
function [k,alfa,u,CV,CS,CK,Alfag,Ug]=fmomlh(x)
```

```
%-----
%Calcula os 4 primeiros momentos-LH
% determina os parâmetros da GEV para ni=0, 1, 2, 3 e 4
% determina os parâmetros da Gumbel para ni=0, 1, 2, 3 e 4
%-----
```

```
N=length(x);
for j=1:5
    n=j-1;
    if n==0
        k1=N;
        k2=k1*(N-1)/2;
        k3=k2*(N-2)/3;
        k4=k3*(N-3)/4;
    elseif n==1
        k1=N*(N-1)/2;
        k2=k1*(N-2)/3;
        k3=k2*(N-3)/4;
        k4=k3*(N-4)/5;
    elseif n==2
        k1=N*(N-1)*(N-2)/6;
        k2=k1*(N-3)/4;
        k3=k2*(N-4)/5;
        k4=k3*(N-5)/6;
    elseif n==3
        k1=N*(N-1)*(N-2)*(N-3)/24;
        k2=k1*(N-4)/5;
        k3=k2*(N-5)/6;
        k4=k3*(N-6)/7;
```

```

else
    k1=N*(N-1)*(N-2)*(N-3)*(N-4)/120;
    k2=k1*(N-5)/6;
    k3=k2*(N-6)/7;
    k4=k3*(N-7)/8;
end
sigma1=0; sigma2=0; sigma3=0; sigma4=0;

for i=1:N

if n==0
    F1=i-1;
    F2=1;
    F4=F1*(i-2)/2;
    F6=F4*(i-3)/3;
elseif n==1
    F2=i-1;
    F1=F2*(i-2)/2;
    F4=F1*(i-3)/3;
    F6=F4*(i-4)/4;
elseif n==2
    F2=(i-1)*(i-2)/2;
    F1=F2*(i-3)/3;
    F4=F1*(i-4)/4;
    F6=F4*(i-5)/5;
elseif n==3
    F2=(i-1)*(i-2)*(i-3)/6;
    F1=F2*(i-4)/4;
    F4=F1*(i-5)/5;
    F6=F4*(i-6)/6;
else
    F2=(i-1)*(i-2)*(i-3)*(i-4)/24;
    F1=F2*(i-5)/5;
    F4=F1*(i-6)/6;
    F6=F4*(i-7)/7;
end

    F3=N-i;
    F5=F3*(N-i-1)/2;
    F7=F5*(N-i-2)/3;

    coef1=F2;
    coef2=F1-F2.*F3;
    coef3=F4-2.*F1.*F3+F2.*F5;
    coef4=F6-3.*F4.*F3+3.*F1.*F5-F2.*F7;
    sigma1=sigma1+coef1*x(i);
    sigma2=sigma2+coef2*x(i);
    sigma3=sigma3+coef3*x(i);
    sigma4=sigma4+coef4*x(i);

```

```

end
LH1(j)=sigma1./k1;
LH2(j)=sigma2./(2*k2);
LH3(j)=sigma3./(3*k3);
LH4(j)=sigma4./(4*k4);
CV(j)=LH2(j)./LH1(j); % coeficiente de variação
CS(j)=LH3(j)./LH2(j); % Coeficiente de assimetria
CK(j)=LH4(j)./LH2(j); % Coeficiente de curtose

end

%Cálculo dos parâmetros da GEV

%coeficientes para o cálculo de k (Wang,1997)
Ao=[0.2849 0.4823 0.5914 0.6618 0.7113];
A1=[-1.8213 -2.1494 -2.3351 -2.4548 -2.5383];
A2=[0.8140 0.7269 0.6442 0.5733 0.5142];
A3=[-0.2835 -0.2103 -0.1616 -0.1273 -0.1027];

for i=1:5
ni=i-1;
k(i)=Ao(i)+A1(i)*CS(i)+A2(i)*CS(i)^2+A3(i)*CS(i)^3;
alfa(i)=(LH2(i)*2*k(i))/((ni+2)*(-(ni+2)^(-k(i)))+(ni+1)^(-k(i)))*gamma(1+k(i));
u(i)=LH1(i)-alfa(i)*(1-gamma(1+k(i))^(ni+1)^(-k(i)))/k(i);
Alfag(i)=LH2(i)*2/((ni+2)*(log(ni+2)-log(ni+1))),% Parâmetros da GEV para k=0
Ug(i)=LH1(i)-Alfag(i)*(0.5772+log(ni+1));
end

no=[0 1 2 3 4];
%'Valores de ni LH1 LH2 LH3 LH4 CV CS CK k alfa u'
[no' LH1' LH2' LH3' LH4' CV' CS' CK' k' alfa' u'];
%'valores de k alfa u da GEV'
[k' alfa' u'];
%'valores de alfa u da Gumbel'
[Alfag' Ug'];

```