

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

METODOLOGIA UNIFICADA PARA
ANÁLISE DE FREQUÊNCIA DE VAZÕES
MÁXIMAS ANUAIS A PARTIR DA
AGREGAÇÃO DA INFORMAÇÃO
HIDROMETEOROLÓGICA REGIONALIZADA

Wilson dos Santos Fernandes

Belo Horizonte

2005

**METODOLOGIA UNIFICADA PARA ANÁLISE DE
FREQUÊNCIA DE VAZÕES MÁXIMAS ANUAIS A
PARTIR DA AGREGAÇÃO DA INFORMAÇÃO
HIDROMETEOROLÓGICA REGIONALIZADA**

Wilson dos Santos Fernandes

Wilson dos Santos Fernandes

**METODOLOGIA UNIFICADA PARA ANÁLISE DE
FREQUÊNCIA DE VAZÕES MÁXIMAS ANUAIS A
PARTIR DA AGREGAÇÃO DA INFORMAÇÃO
HIDROMETEOROLÓGICA REGIONALIZADA**

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

Área de concentração: Recursos Hídricos

Linha de pesquisa: Hidrologia Estatística

Orientador: Mauro da Cunha Naghettini

Belo Horizonte

Escola de Engenharia da UFMG

2005



FOLHA DE APROVAÇÃO

Metodologia Unificada para Análise de Frequência de Vazões Máximas Anuais
a Partir da Agregação da Informação Hidrometeorológica Regionalizada

WILSON DOS SANTOS FERNANDES

Dissertação defendida e aprovada pela banca examinadora constituída pelos Senhores:



Prof. MAURO DA CUNHA NAGHETTINI


Prof. LUIZ RAFAEL PALMIER



Prof. EDUARDO SAVIO PASSOS RODRIGUES MARTINS


PESQ. EBER JOSÉ DE ANDRADE PINTO

Aprovada pelo Colegiado do PG SMARH


Profª. Mônica Maria Diniz Leão
Coordenadora

Versão Final aprovada por


Prof. Mauro da Cunha Naghettini
Orientador

Belo Horizonte, 17 de outubro de 2005.

AGRADECIMENTOS

Agradeço aos meus pais pelo constante incentivo para o meu crescimento acadêmico e pela motivação que sempre despertaram em mim.

À minha esposa Janine, pela paciência e compreensão nas noites que me dediquei à elaboração desse trabalho.

Ao meu orientador e amigo, Professor Mauro Naghettini, pela dedicação e esforço despendido para o aprimoramento dessa dissertação.

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

Agradeço o apoio da FAPEMIG – Fundação de Amparo à Pesquisa do Estado de Minas Gerais – ao projeto CRA576/01, do qual a presente pesquisa é parte integrante.

Por fim, agradeço a todos meus amigos, que foram fonte de grande motivação para a conclusão desta pesquisa.

RESUMO

As variáveis hidrológicas, e as vazões de enchentes em particular, são consideradas variáveis aleatórias e, portanto, sujeitas à análise pela teoria de probabilidade e estatística matemática. As estimativas dos quantis extremos dessas vazões e de suas probabilidade de excedência constituem elementos indispensáveis para o dimensionamento e a avaliação de falha das estruturas hidráulicas de grande porte. Para cheias menos raras, a análise convencional de frequência de vazões apresenta um conjunto satisfatório e rotineiro de métodos de estimação. Para enchentes extremas, entretanto, não existe um conjunto consensual de metodologias a se empregar, dadas as grandes incertezas inerentes à estimação, com base em pequenas amostras.

Nesse sentido o U. S. National Research Council (NRC, 1988) identificou três princípios a serem seguidos para a melhoria das estimativas de quantis extremos de vazão, a saber: “substituição do tempo pelo espaço”, “inclusão de maior estrutura aos modelos utilizados” e “ênfase na cauda superior das distribuições de probabilidade”. A metodologia proposta nesta dissertação fez uso dos três princípios mencionados de uma forma unificada. No contexto dessa metodologia, as vazões de pico excedentes sobre um valor limiar arbitrário e os volumes de cheia, associados a essas excedências, são individualizados e modelados como um processo estocástico pontual marcado usando-se a representação de um processo composto de Poisson. A essência do método proposto consiste em se estimar separadamente a função densidade marginal de probabilidade dos volumes de cheia, para uma duração equivalente ao tempo de base da bacia, e a função densidade das vazões de pico condicionadas aos volumes. Na seqüência, a função de distribuição de probabilidades anuais das vazões de pico pode ser estimada através da integração do produto entre a densidade marginal de probabilidade dos volumes de cheia e a densidade das vazões de pico condicionadas aos volumes. A agregação da informação hidrometeorológica regionalizada serve o propósito de guiar a estimativa da densidade dos volumes de cheia. Essa agregação faz uso de algumas premissas sobre a transformação chuva-vazão, sob condições extremas. Por outro lado, a regionalização das precipitações faz uso da distribuição TCEV (Two Component Extreme Value).

A metodologia proposta combina técnicas estatísticas contemporâneas e busca extrair o máximo de informação dos dados observados em uma bacia. Nesse sentido, espera-se que outras aplicações do método proposto, além da apresentada na presente dissertação, possam melhorar as estimativas de quantis extremos de vazão.

ABSTRACT

Hydrological variables, in general and floodflows, in particular, are considered to be random variables and, consequently, subject to analysis by the theory of probability and mathematical statistics. The estimates of extreme flood quantiles, along with their corresponding exceedance probabilities, are key elements to designing large hydraulic structures and evaluating the associated risks of failure. For less rare floods, the conventional flood frequency analysis provides a number of satisfactory methods of estimation. For extreme floods, however, there is no consensual method to employ, given the large uncertainties inherent to the estimation based on small samples.

In this context, the U.S. National Research Council (NRC, 1988) identified three principles to be followed in order to improve estimation of extreme flood quantiles. These are: “substitution of time for space”, “inclusion of more structure into the employed models”, and “emphasis on the upper tails of probability distributions”. The methodology proposed herein makes use of the three aforementioned principles within a unified context. According to it, the flood peaks that have exceeded an arbitrary threshold and the associated flood volumes are identified and modeled as a marked point stochastic process, using the representation of a compound Poisson process. The essence of the proposed method consists of separately estimating the marginal density function of flood volumes, for a duration equal to the watershed time base, and the density function of flood peaks conditioned to volumes. In the sequence, the annual probability distribution of flood peaks can be estimated by integrating the product of both densities. The aggregation of the regional hydrometeorological information, through a regional TCEV (Two Component Extreme Value) distribution, serves the purpose of guiding the estimation of the flood-volumes density. This aggregation makes use of some premises on the rainfall-runoff transformation, under extreme conditions.

The proposed method combines contemporary statistical techniques and aims to extract the maximum information from the available data. In this sense, it is expected that other applications of the proposed method, in addition to the one described herein, may improve the estimation of extreme flood quantiles.

SUMÁRIO

1	INTRODUÇÃO E OBJETIVOS	1
1.1	INTRODUÇÃO.....	1
1.2	OBJETIVOS.....	3
1.2.1	<i>Objetivo geral.....</i>	<i>3</i>
1.2.2	<i>Objetivos específicos</i>	<i>4</i>
2	MÉTODOS DE ESTIMAÇÃO DE VAZÃO	5
2.1	INTRODUÇÃO.....	5
2.2	CONCEITOS BÁSICOS.....	5
2.3	MÉTODOS PONTUAIS DE ESTIMAÇÃO.....	10
2.3.1	<i>Análise de frequência de séries de máximos anuais.....</i>	<i>11</i>
2.3.2	<i>Séries de duração parcial.....</i>	<i>17</i>
2.4	MÉTODOS REGIONAIS DE ESTIMAÇÃO.....	20
2.4.1	<i>O método do index-flood</i>	<i>22</i>
2.4.2	<i>Etapas para a regionalização por momentos-L</i>	<i>23</i>
2.5	ANÁLISE DE FREQUÊNCIA DE VAZÕES SIMULADAS POR MODELOS CHUVA-VAZÃO.....	27
2.5.1	<i>Modelo estocástico de geração de dados de chuva.....</i>	<i>28</i>
2.5.2	<i>Modelo de transformação chuva-vazão.....</i>	<i>30</i>
2.6	MODELO MISTO	34
3	METODOLOGIA.....	39
3.1	INTRODUÇÃO.....	39
3.2	O MÉTODO PICO-VOLUME.....	41
3.3	O MÉTODO GRADEX	45
3.3.1	<i>Descrição do método GRADEX.....</i>	<i>46</i>
3.3.2	<i>Estimativa do parâmetro gradex.....</i>	<i>52</i>
3.3.3	<i>Extrapolação da distribuição dos volumes de cheia</i>	<i>53</i>
3.4	REGIONALIZAÇÃO HIDROLÓGICA PELA TCEV	54
3.4.1	<i>Definição</i>	<i>54</i>
3.4.2	<i>Primeiro nível de regionalização</i>	<i>55</i>
3.4.3	<i>Segundo nível de regionalização.....</i>	<i>57</i>
3.4.4	<i>Terceiro nível de regionalização.....</i>	<i>58</i>
3.4.5	<i>Estimativa dos parâmetros regionais da TCEV</i>	<i>58</i>
3.5	RELAÇÃO PICO-VOLUME.....	68
3.6	O MÉTODO PVP.....	71
3.6.1	<i>Introdução</i>	<i>71</i>
3.6.2	<i>Método pico-volume com a chuva como variável aleatória auxiliar</i>	<i>72</i>
3.6.3	<i>Determinação da distribuição dos volumes.....</i>	<i>73</i>
3.6.4	<i>Determinação da distribuição dos picos condicionada aos volumes</i>	<i>77</i>
4	RESULTADOS	79
4.1	INTRODUÇÃO.....	79
4.2	A BACIA DO RIO PARÁ	79

4.3	ANÁLISE DAS SÉRIES DE DURAÇÃO PARCIAL DE VAZÃO E PRECIPITAÇÃO.....	80
4.3.1	<i>Vazão</i>	80
4.3.2	<i>Precipitação</i>	84
4.4	RELAÇÕES PICO-VOLUME.....	85
4.4.1	<i>Relação pico-volume para o estado de Minas Gerais</i>	85
4.4.2	<i>Relação pico-volume para a bacia do rio Pará, em Ponte do Vilela</i>	92
4.5	REGIONALIZAÇÃO DOS DADOS PLUVIOMÉTRICOS PELA TCEV	98
4.5.1	<i>Introdução</i>	98
4.5.2	<i>Consistência dos dados e definição da região homogênea</i>	98
4.5.3	<i>Cálculo dos parâmetros regionais e locais da TCEV</i>	101
4.6	A DISTRIBUIÇÃO DOS VOLUMES ESCOADOS PELO MÉTODO GRADEX	104
4.7	A DISTRIBUIÇÃO DOS PICOS DE VAZÃO PELO MÉTODO PVP	107
4.7.1	<i>Análise de frequência convencional</i>	109
4.7.2	<i>Modelo Poisson-Pareto</i>	110
4.7.3	<i>Regionalização por momentos-L</i>	112
4.7.4	<i>Modelo estocástico de simulação de vazões</i>	114
4.7.5	<i>Comparação entre os métodos de estimação de vazão</i>	115
5	CONCLUSÕES E RECOMENDAÇÕES.....	117
6	REFERÊNCIAS BIBLIOGRÁFICAS.....	121
	ANEXO 1 – DISTRIBUIÇÕES DE PROBABILIDADES.....	125
	ANEXO 2 – SÉRIES DE DURAÇÃO PARCIAL PARA AS ESTAÇÕES PLUVIOMÉTRICAS.....	133
	ANEXO 3 – REGRESSÃO LOWESS	136
	ANEXO 4 – PROGRAMA PVP	139

LISTA DE FIGURAS

Figura 2.1 – Fluxograma do módulo de produção do modelo RIO GRANDE (adap. de Naghettini et al., 2002).....	33
Figura 2.2a – Distribuição de freqüência de volumes de cheia (adap. de Naghettini, 1994).....	36
Figura 2.2b – Vazões de pico-volume de cheia simulados e observados (adap. de Naghettini, 1994).....	37
Figura 2.3 – Distribuição de freqüência de vazões máximas (adap. de Naghettini, 1994).....	37
Figura 3.1 – Processo pontual marcado para eventos de vazão.....	41
Figura 3.2 – Fluxograma de aplicação do método PVP.....	45
Figura 3.3 – Relação entre a chuva (P) e o volume escoado (X).....	47
Figura 3.4 – Funções de distribuição de probabilidades acumuladas dos volumes escoados $G(x)$ e dos volumes de chuva $F(p)$, ambos de duração d . Papel de plotagem clássico. A distância r_0 refere-se à distância de translação no domínio das extrapolações D_2	49
Figura 3.5 – Plotagem de $G(x)$ e $F(p)$ em papel de plotagem exponencial ou de Gumbel. A distância r_0 é a distância de translação no domínio das extrapolações D_2 . As durações d da chuva e do volume são as mesmas.....	50
Figura 3.6 – Esquema de um processo pontual marcado tendo a chuva como variável aleatória auxiliar.....	73
Figura 4.1 – Localização da bacia hidrográfica do rio Pará.....	81
Figura 4.2 – Disposição espacial de cada estação pluviométrica.....	84
Figura 4.3 – Localização geográfica das estações utilizadas.....	86
Figura 4.4 – Hidrograma de cheia: definição do tempo de base.....	88
Figura 4.5 – Relação Pico-Volume para cada estação fluviométrica.....	89
Figura 4.6 – Relação entre a área e o intercepto para a região estudada.....	90
Figura 4.7 – Picos observados versus picos simulados.....	91
Figura 4.8 – Hidrogramas unitários para o rio Pomba em Cataguases.....	91
Figura 4.9 – Relação Pico-Volume para a bacia do rio Pará em Ponte do Vilela.....	93
Figura 4.10 – Pares pico-volume simulados.....	97
Figura 4.11 – Modelos pico-volume para a bacia do rio Pará.....	97
Figura 4.12 – Função objetivo do algoritmo genético: DL é o parâmetro Λ^* , TH é o parâmetro θ^* e FO é o valor da função objetivo.....	102
Figura 4.13 – Ajuste da TCEV à chuva média sobre a bacia do rio Pará em Ponte do Vilela.....	104
Figura 4.14 – Distribuição dos volumes de cheia e de precipitação para a bacia do rio Pará.....	107

Figura 4.15 – Distribuição dos picos condicionada aos volumes	108
Figura 4.16 – Vazão máxima anual versus período de retorno em Ponte do Vilela	109
Figura 4.17 – Quantis máximos para Ponte do Vilela segundo várias metodologias	115
Figura A4.1 – Tela de abertura do programa PVP.....	139
Figura A4.2 – Gerenciamento de estações de monitoramento.....	140
Figura A4.3 – Módulo de gerenciamento de relações pico-volume	141
Figura A4.4 – Dados principais do projeto.....	142
Figura A4.5 – Dados para a regionalização pela TCEV	143
Figura A4.6 – Entrada de dados para a relação pico-volume	144
Figura A4.7 – Resultados gerados pelo programa PVP.....	145

LISTA DE TABELAS

Tabela 2.1 –	Peso da cauda superior de algumas distribuições	13
Tabela 2.2 –	Valores críticos para a medida de discordância.....	24
Tabela 2.3 –	Estados para a chuva diária da bacia do rio Pará.....	29
Tabela 2.4 –	Matriz de probabilidade de transição para o mês de janeiro na bacia do rio Pará	29
Tabela 4.1 –	Picos (m^3/s) e volumes ($m^3/s.dia$) para Ponte do Vilela com $\Lambda = 1,5$	83
Tabela 4.2 –	Número de ocorrências (m_k) por ano	83
Tabela 4.3 –	Características das estações pluviométricas da região.....	84
Tabela 4.4 –	Análise da independência serial das estações pluviométricas	85
Tabela 4.5 –	Principais características das estações fluviométricas utilizadas na determinação da relação pico volume para Minas Gerais.....	87
Tabela 4.6 –	Coefficiente de correlação simples entre A, L, S e b	89
Tabela 4.7 –	Quantis de vazão para diversos valores do erro padrão.....	94
Tabela 4.8 –	Erros quadráticos da relação pico-volume.....	95
Tabela 4.9 –	Momentos-L e medida de discordância para a região estudada.....	99
Tabela 4.10 –	Parâmetros das distribuições usadas na estimativa local dos quantis de vazão	110
Tabela 4.11 –	Vazões máximas anuais para Ponte do Vilela calculadas localmente por meio de distribuições de probabilidades.	110
Tabela 4.12 –	Quantis e período de retorno em anos para Ponte do Vilela calculados pelo modelo POT	112
Tabela 4.13 –	Quantis máximos anuais para Ponte do Vilela calculados por Momentos-L.....	113
Tabela 4.14 –	Parâmetros do modelo RIO GRANDE.....	114
Tabela 4.15 –	Quantis máximos anuais simulados para Ponte do Vilela	114

LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS

γ	Coefficiente de assimetria
σ	Desvio padrão
$H_u(x)$	Distribuição acumulada da série de duração parcial
$F_X(x)$	Distribuição acumulada de probabilidades
$f_X(x)$	Função densidade de probabilidade
$g_u(x)$	Função densidade marginal de probabilidade dos volumes de cheia
$G_u(x)$	Distribuição acumulada marginal de probabilidade dos volumes de cheia
$\lambda(t)$	Função intensidade de Poisson
μ_r	Momento populacional central de ordem r em relação à média, <i>index-flood</i>
μ_r'	Momento populacional de ordem r em relação à origem
λ_r	Momento-L de ordem r
α_r	Momento ponderado por probabilidade de ordem r
β_r	Momento ponderado por probabilidade de ordem r
σ^2	Variância populacional
τ_3	Coefficiente de assimetria-L populacional
τ_4	Coefficiente de curtose-L populacional
τ	Coefficiente de variação-L populacional
$f_{Y/X}(y/x)$	Função densidade de probabilidade das vazões de pico condicionadas aos volumes
$F_{Y/X}(y/x)$	Distribuição acumulada de probabilidade das vazões de pico condicionadas aos volumes
Γ	Função gama
$\hat{\mu}$	Média amostral
ℓ_r	Momento-L amostral de ordem r
t_r	Quociente de momentos-L amostral de ordem r

A	Área de drenagem
b	Intercepto da relação pico-volume
CPRM	Companhia de Pesquisa de Recursos Minerais (Serviço geológico do Brasil)
CV	Coefficiente de variação
d	Duração em dias, estatística do teste de independência serial da série de duração parcial
Di	Medida de discordância
DPPV	Distribuição padrão pico-volume
DPVPO	Distribuição pico-volume de primeira ordem
DPVSO	Distribuição pico-volume de segunda ordem
$E(X)$	Valor esperado da variável X
EXP	Distribuição exponencial
GEV	Distribuição generalizada de valores extremos
GLO	Distribuição logística generalizada
GPA	Distribuição generalizada de Pareto
GRADEX	Gradiente de valor extremo
GUM	Distribuição de Gumbel
H	Medida de heterogeneidade
k	Coefficiente de curtose
L	Comprimento do talvegue principal
LNO	Distribuição Log-Normal
LOWESS	<i>Robust locally weighted regression</i>
m	Inclinação da relação pico-volume, número anual de picos acima de u
MPP	Momento ponderado por probabilidade
m_r	Momento amostral de ordem r
$P(A)$	Probabilidade de ocorrer o evento A
PDF	Função densidade de probabilidade
PE3	Distribuição de Pearson tipo III
P_i	Máxima altura de chuva para a duração d
POT	<i>Peaks Over Threshold</i>
Q_p	Vazão de pico

R^2	Coeficiente de determinação
R_i	Déficit de <i>runoff</i>
S	Declividade média do talvegue principal
SCE-UA	<i>Shuffled Complex Evolution</i>
SDP	Série de duração parcial
SEAF	Sistema Especialista para Análise de Freqüência local
SMA	Série dos máximos anuais
T	Período de retorno
T_b	Tempo de base médio da bacia
T_c	Período de retorno a partir do qual a segunda hipótese do método GRADEX é válida
TCEV	<i>Two Component Extreme Value</i>
u	<i>Threshold</i>
V	Volume
V_j	Desvio-padrão do CV-L para a amostra j
$w(X)$	Regressão não paramétrica da variável X
WEI	Distribuição de Weibull
$x(p)$	Quantil para a probabilidade p
X_i	Volumes de cheia de duração d correspondente a Y_i
Y_i	Picos de vazão acima de u , variável adimensionalizada distribuída conforme a TCEV
Z^{DIST}	Medida de aderência da distribuição “ <i>DIST</i> ”
Λ	Taxa anual de ocorrências
σ_ε	Erro padrão da estimativa da regressão de Y e X

1 INTRODUÇÃO E OBJETIVOS

1.1 Introdução

As variáveis hidrológicas, e as vazões de enchentes em particular, são consideradas variáveis aleatórias e, portanto, sujeitas à análise pela teoria de probabilidade e estatística matemática. As estimativas dos quantis extremos dessas vazões e de suas probabilidade de excedência constituem elementos indispensáveis para o dimensionamento e a avaliação de falha das estruturas hidráulicas de grande porte. Para cheias menos raras, a análise convencional de frequência de vazões apresenta um conjunto satisfatório e rotineiro de métodos de estimação. Para enchentes extremas, entretanto, não existe um conjunto consensual de metodologias a se empregar, dadas as grandes incertezas inerentes à estimação, com base em pequenas amostras.

Visando diminuir ou melhor descrever as incertezas envolvidas na determinação de cheias raras, o Comitê de Estudos dos Métodos de Estimação de Probabilidade de Cheias Extremas, formado pelo Conselho Nacional de Pesquisas norte-americano (NRC, 1988), identificou três princípios gerais a serem perseguidos na proposição de novas metodologias para a melhoria das estimativas de probabilidades de enchentes muito raras. São eles:

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

As técnicas convencionais de estimação pontual, tais como a análise de frequência das vazões máximas anuais, não foram concebidas de forma a levar em conta a variabilidade espacial dos processos físicos envolvidos na formação das cheias. Além disso, as séries fluviométricas

disponíveis no Brasil e em outros países são, na grande maioria dos casos, curtas, o que não proporciona uma estimativa confiável dos quantis para grandes períodos de retorno. A escolha da distribuição que representa os máximos locais é motivo de grande debate no meio técnico, não havendo um consenso acerca de como se determinar, com certeza, de qual população a amostra observada foi extraída.

Na análise convencional de frequência dos máximos anuais, também não se leva em conta o terceiro princípio proposto pelo NRC (1988), uma vez que o valor máximo em um determinado ano pode vir a ser inferior ao segundo ou ao terceiro maior valor de um outro ano qualquer, os quais podem constituir elementos amostrais mais representativos da cauda superior. Esse problema pode ser parcialmente contornado pelo uso de séries de duração parcial. A série de duração parcial é formada pelo valores superiores a um determinado nível, o que proporciona a análise sobre os maiores valores históricos. No entanto, não há um critério consensual para se estabelecer o valor limiar desejável, de modo que as observações centrais, ou comuns, sejam excluídas ou não influenciem a estimação da cauda superior.

Atualmente, há uma forte tendência na utilização dos métodos regionais para estimação dos quantis de vazão, sobretudo aqueles baseados nos momentos ponderados por probabilidades ou em suas combinações lineares ou momentos-L. Sabe-se que os métodos regionais têm a vantagem adicional de reduzir os efeitos dos erros amostrais locais e permitir a estimação dos quantis em locais desprovidos de medição. No entanto, a aplicabilidade e a precisão dessas metodologias está sujeita à disponibilidade e à qualidade dos dados observados. Sabe-se que a rede de postos fluviométricos, em especial no Brasil, não é suficientemente densa para permitir uma análise adequada da variabilidade espacial das características estatísticas da região onde se situam. Associados a esse fato, estão todos os problemas referentes à análise de frequência local citados anteriormente. Embora melhores que os métodos pontuais, a análise regional de frequência não utiliza, em sua plenitude, os princípios anteriormente citados. Se por um lado nem sempre é dada atenção exclusiva aos eventos extremos, por outro ressalta-se que nos modelos regionais mais comuns não há uma preocupação em se dar uma maior estrutura à modelagem dos processos físicos envolvidos.

Uma tentativa válida de atender às recomendações do NRC (1988) foi proposta por Naghettini (1994) que desenvolveu um modelo misto que faz uso de três metodologias: método pico-volume, método GRADEX e séries de duração parcial. O modelo reúne as potencialidades e

qualidades dessas três metodologias, de modo a poder utilizar, em princípio, o máximo de informação disponível em uma bacia. Além disso, o método faz uso de técnicas de regionalização e de estatísticas de ordem superior, proporcionando assim, a possibilidade de estimativas mais representativas de quantis extremos. Como será visto no decorrer dessa dissertação, o método proposto por Naghettini (1994) atende aos três princípios do NRC (1988) e, nesse contexto, é uma importante ferramenta na determinação de quantis extremos de vazão. No entanto, sua aplicação às condições brasileiras é difícil devido, principalmente, à disponibilidade e à qualidade de certos dados hidrometeorológicos essenciais a algumas etapas particulares da metodologia.

As limitações do método proposto por Naghettini (1994) constituem uma barreira para a sua aplicação em bacias com pouca disponibilidade de dados, caso comum no Brasil. Nessa dissertação, tal metodologia será descrita em detalhes, permitindo assim, determinar as principais características que impedem sua aplicação no Brasil. Na seqüência, serão introduzidas modificações que possibilitem sua aplicação de forma mais ampla e fácil. A nova metodologia, resultante das modificações propostas, aqui denominada PVP (Precipitação – Volume – Pico de Cheia), constitui o objeto principal da presente dissertação.

Além deste capítulo de introdução e objetivos, esta dissertação é organizada da seguinte forma: no capítulo 2 é apresentada uma revisão dos principais métodos de estimação de quantis máximos de vazão que servirão, posteriormente, para comparação de resultados em relação aos obtidos pela metodologia proposta. No capítulo 3, são descritas todas as partes componentes da metodologia proposta, enquanto no capítulo 4, é realizada uma aplicação completa da metodologia à bacia do Rio Pará, além de comparação de resultados obtidos por outras metodologias existentes. No capítulo 5 conclui-se o trabalho com sugestões e recomendações de uso e futuros avanços e, finalmente, enquanto no capítulo 6 são listadas as referências bibliográficas utilizadas na elaboração do trabalho.

1.2 Objetivos

1.2.1 Objetivo geral

Desenvolver uma seqüência metodológica unificada que permita extrapolar o ramo superior da curva de freqüência de vazões máximas, buscando-se extrair o máximo de informações das estatísticas amostrais de ordem superior das observações pluviométricas disponíveis em uma

dada região, tendo-se em conta a transformação chuva-escoamento superficial sob condições extremas e as características fluviais que governam a relação entre vazão média diária e volume de dada duração.

1.2.2 Objetivos específicos

- Construir e validar um modelo paramétrico regional para a estimação da função densidade condicional $f_{Y/X}(y/x)$ entre vazões médias diárias e volumes de cheia;
- Estruturar e calibrar o modelo regional TCEV para as precipitações de duração especificada;
- Efetuar uma aplicação completa dessa seqüência metodológica em uma bacia, de forma a verificar a adequação do seu emprego, nas condições da rede fluviométrica brasileira; e
- Implementar um pacote computacional, a ser disponibilizado à comunidade técnica, que automatize a seqüência metodológica a ser desenvolvida e que facilite sua utilização em aplicações genéricas de análise de freqüência de vazões médias diárias máximas.

2 MÉTODOS DE ESTIMAÇÃO DE VAZÃO

2.1 Introdução

Em hidrologia, há vários métodos para a estimação dos quantis máximos de vazão. No presente capítulo, são descritos os principais métodos, bem como suas vantagens e desvantagens.

Os métodos de estimação podem ser divididos em dois grupos principais: o primeiro diz respeito às técnicas pontuais de estimação e tem como principais elementos as técnicas de ajuste de distribuições de probabilidade às séries de vazões máximas anuais (SMA), ou às séries de duração parcial (SDP), ou ainda às séries simuladas por modelos chuva-vazão. O segundo grupo refere-se às técnicas regionais de estimação, as quais têm como principal elemento o método da adimensionalização local pela cheia índice ou *index-flood*.

2.2 Conceitos básicos

De modo a uniformizar a linguagem utilizada nos capítulos que se seguem, é feita uma revisão dos conceitos básicos utilizados em hidrologia estatística.

Dada uma variável aleatória X define-se a distribuição acumulada de probabilidades $F_X(x)$ da seguinte forma:

$$F_X(x) = P(X \leq x) \quad (2.1)$$

onde $P(\cdot)$ denota a probabilidade de não superação da variável X em relação ao argumento x .

A função densidade de probabilidade da variável X , $f_X(x)$, é dada pela derivada primeira de $F_X(x)$, enquanto a inversa da equação 2.1 representa a função quantil, $x(F)$, conforme apresentado a seguir:

$$f_X(x) = \frac{dF_X(x)}{dx} \quad (2.2)$$

$$x(F) = F^{-1}(x) \quad (2.3)$$

A equação 2.3 fornece o quantil para o qual a probabilidade de X não exceder o valor $x(F)$ é dada por F .

Define-se o momento populacional de ordem r , em relação à origem da função densidade de probabilidades dada por 2.2, pela seguinte equação:

$$\mu'_r = \int_{-\infty}^{+\infty} x^r f_X(x) dx \quad (2.4)$$

O momento populacional de ordem 1 é conhecido como valor esperado da variável X e representa a média populacional, isto é:

$$\mu = \int_{-\infty}^{+\infty} x f_X(x) dx = E(X) \quad (2.5)$$

Considerando que $p = F(x)$, o valor esperado de X pode tomar a seguinte forma:

$$E(X) = \int_0^1 x(p) dp \quad (2.6)$$

O momento populacional central de ordem r , em relação à média populacional, é:

$$\mu_r = \int_{-\infty}^{+\infty} (x - \mu)^r f_X(x) dx \quad (2.7)$$

Pode-se demonstrar, através das propriedades do valor esperado e da equação 2.7, a seguinte relação:

$$\mu_r = E[(x - \mu)^r]; \quad r \geq 2 \quad (2.8)$$

Os momentos centrais de ordem 2, 3 e 4 são de particular importância para a hidrologia estatística. O momento de ordem 2 representa a variância da distribuição e é uma medida da dispersão desta em relação à média. A partir da variância, dada pela equação 2.9, define-se o desvio padrão (σ) e o coeficiente de variação CV da distribuição:

$$\text{var}(x) = \sigma^2 = \mu_2 \quad (2.9)$$

$$\sigma = \sqrt{\text{var}} = \sqrt{\sigma^2} \quad (2.10)$$

$$CV = \frac{\sigma}{\mu} \quad (2.11)$$

A partir dos momentos centrais de ordem 3 e 4 define-se, respectivamente, o coeficiente de assimetria (γ) e o coeficiente de curtose (k):

$$\gamma = \frac{\mu_3}{(\sigma)^3} \quad (2.12)$$

$$k = \frac{\mu_4}{(\sigma)^4} \quad (2.13)$$

O momentos centrais definidos pela equação 2.7 são ditos momentos convencionais. Alternativamente, Hosking & Wallis (1997) definiram quantidades, de interpretação análoga ao momentos convencionais, denominadas momentos lineares ou, simplesmente, momentos-L.

Os momentos-L são deduzidos a partir de combinações lineares dos momentos ponderados por probabilidades (MPP), os quais foram introduzidos inicialmente por Greenwood et al. (1979). Os MPP de maior interesse na análise estatística são definidos da seguinte forma:

$$\alpha_r = \int_0^1 x(p) \cdot (1-p)^r dp \quad (2.14)$$

$$\beta_r = \int_0^1 x(p) \cdot p^r dp \quad (2.15)$$

Em comparação aos momentos convencionais, os MPP envolvem potências de p enquanto os primeiros envolvem potências de $x(p)$. Esse é um dos pontos favoráveis ao emprego dos MPP, já que o termo $(x - \mu)$ pode ser fortemente afetado pela presença de *outliers* na amostra.

Os MPP das equações 2.14 e 2.15 são de difícil interpretação e não refletem diretamente as características estatísticas das distribuições de probabilidades. Dessa forma, Hosking &

Wallis (1997) definiram os momentos-L, de interpretação mais fácil, a partir dos MPP. O momento-L de ordem $r + 1$ é dado pela seguinte equação:

$$\lambda_{r+1} = (-1)^r \sum_{k=0}^r l_{r,k}^* \cdot \alpha_k = \sum_{k=0}^r l_{r,k}^* \cdot \beta_k \quad (2.16)$$

na qual

$$l_{r,k}^* = (-1)^{r-k} \binom{r}{k} \binom{r+k}{k} = (-1)^{r-k} \frac{(r+k)!}{(k!)^2 (r-k)!} \quad (2.17)$$

Os quatro primeiros momentos-L são:

$$\lambda_1 = \alpha_0 = +\beta_0 \quad (2.18)$$

$$\lambda_2 = \alpha_0 - 2\alpha_1 = -\beta_0 + 2\beta_1 \quad (2.19)$$

$$\lambda_3 = \alpha_0 - 6\alpha_1 + 6\alpha_2 = +\beta_0 - 6\beta_1 + 6\beta_2 \quad (2.20)$$

$$\lambda_4 = \alpha_0 - 12\alpha_1 + 30\alpha_2 - 20\alpha_3 = -\beta_0 + 12\beta_1 - 30\beta_2 + 20\beta_3 \quad (2.21)$$

A quantidade λ_1 é a média ou momento-L de posição e a quantidade λ_2 é o momento-L de escala. Analogamente aos momentos convencionais, podem ser definidos o coeficiente de variação (CV-L), o coeficiente de assimetria (τ_3) e o coeficiente de curtose (τ_4) para os momentos-L:

$$CV-L = \tau = \frac{\lambda_2}{\lambda_1} \quad (2.22)$$

$$\tau_3 = \frac{\lambda_3}{\lambda_2} \quad (2.23)$$

$$\tau_4 = \frac{\lambda_4}{\lambda_2} \quad (2.24)$$

Os momentos populacionais convencionais e lineares para diversas distribuições de probabilidade, conforme estabelecidos anteriormente, encontram-se detalhados no Anexo 1 da presente dissertação.

Os momentos populacionais podem ser estimados por quantidades análogas, calculadas para uma amostra de tamanho n . Para os momentos convencionais, tem-se:

$$\hat{\mu} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (2.25)$$

Os estimadores enviesados dos momentos centrais de ordem r podem ser calculados a partir da seguinte relação:

$$\hat{\mu}_r = m_r = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^r \quad (2.26)$$

Os estimadores não enviesados da variância, coeficiente de variação, coeficiente de assimetria e do coeficiente de curtose são:

$$\hat{\sigma}^2 = s^2 = \frac{n}{n+1} m_2 \quad (2.27)$$

$$CV = \frac{s}{\bar{x}} \quad (2.28)$$

$$\hat{\gamma} = g = \frac{n^2}{(n-1)(n-2)} \cdot \frac{m_3}{s^3} \quad (2.29)$$

$$\hat{k} = k = \frac{n^3}{(n-1)(n-2)(n-3)} \cdot \frac{m_4}{s^4} \quad (2.30)$$

A estimação dos MPP e momentos-L, a partir de uma amostra finita de tamanho n , inicia-se com a ordenação de seus elementos constituintes em ordem crescente, ou seja, $x_{1:n} \leq x_{2:n} \leq \dots \leq x_{n:n}$. Um estimador não-enviesado do MPP β_r pode ser escrito como:

$$b_r = \hat{\beta}_r = \frac{1}{n} \sum_{j=r+1}^n \frac{(j-1)(j-2)\dots(j-r)}{(n-1)(n-2)\dots(n-r)} \cdot x_{j:n} \quad (2.31)$$

Os estimadores para os primeiros momentos-L podem ser calculados da seguinte forma:

$$\ell_1 = \hat{\lambda}_1 = b_0 \quad (2.32)$$

$$\ell_2 = \hat{\lambda}_2 = -b_0 + 2b_1 \quad (2.33)$$

$$\ell_3 = \hat{\lambda}_3 = b_0 - 6b_1 + 6b_2 \quad (2.34)$$

$$\ell_4 = \hat{\lambda}_4 = -b_0 + 12b_1 - 30b_2 + 20b_3 \quad (2.35)$$

Os estimadores para o τ , τ_3 e τ_4 são:

$$\hat{\tau} = t = \frac{\ell_2}{\ell_1} \quad (2.36)$$

$$\hat{\tau}_3 = t_3 = \frac{\ell_3}{\ell_2} \quad (2.37)$$

$$\hat{\tau}_4 = t_4 = \frac{\ell_4}{\ell_2} \quad (2.38)$$

2.3 Métodos pontuais de estimação

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

Segundo Rao & Hamed (2000), a relação entre o evento extremo e o tempo de retorno pode ser obtida localmente através de três modelos: aqueles baseados nas séries dos máximos anuais (SMA), os baseados nas séries de duração parcial (SDP), ou modelos POT (*Peaks over threshold*), como são conhecidos comumente, e os modelos de séries temporais.

Nos modelos SMA somente o maior valor anual é considerado. Neste caso, alguns valores extremos observados não são incluídos na análise, uma vez que o segundo ou o terceiro maior

valor, em determinado ano, pode ser maior que o máximo de um outro ano. Esse fato é contornado nos modelos SDP, nos quais os maiores valores acima de um determinado limiar são considerados na análise.

Os modelos que utilizam séries temporais consideram todos os eventos que são igualmente espaçados no tempo por um intervalo predeterminado. Podem ser utilizados intervalos anuais, mensais ou diários, embora este último seja o mais comum na prática. A análise dos modelos de séries temporais está além dos objetivos do presente trabalho e não será aqui explorada.

2.3.1 Análise de frequência de séries de máximos anuais

Segundo Cândido (2003), a análise de frequência de eventos máximos anuais pode ser dividida em quatro etapas:

- Verificação dos dados amostrais;
- Escolha da distribuição de probabilidades;
- Estimativa dos parâmetros das distribuições; e
- Identificação e tratamento dos pontos atípicos ou *outliers*.

A seguir será descrita cada uma das etapas mencionada.

2.3.1.1 Verificação dos dados amostrais

A análise de frequência convencional parte do pressuposto que a amostra dos dados hidrológicos seja representativa, aleatória, independente, homogênea e estacionária.

A representatividade de uma série é dada pela sua capacidade de caracterizar a variabilidade do processo natural de formação do evento analisado. O fator que mais influencia a representatividade de uma série é o seu tamanho. Séries muito curtas podem enviesar a estimativa dos quantis de vazão. A estimativa de vazões a partir de uma série curta, tomada de um período seco, certamente fornecerá valores diferentes de uma série de mesmo comprimento tomada de um período chuvoso. O tamanho mínimo da série é bastante discutido na literatura, embora não haja um consenso entre os autores.

Segundo Naghettini & Pinto (no prelo), o caráter aleatório de uma série é garantido quando as flutuações da variável hidrológica analisada decorrem de causas naturais. Nesse sentido, as vazões de um curso d'água, fortemente regularizadas pela operação de um reservatório constituiriam um exemplo de uma série não-aleatória.

Uma série é dita independente se qualquer um de seus dados não influencia a probabilidade de ocorrência ou não ocorrência dos demais dados. No caso das séries dos máximos, a independência é garantida quando o máximo de um determinado ano não influencia e não é influenciado pelos máximos dos demais anos. A independência serial de uma amostra pode ser verificada, por exemplo, pelo teste de Wald-Wolfowitz (NAGHETTINI & PINTO, no prelo).

A homogeneidade ocorre quando todos os dados de uma amostra são originados de uma mesma população. Numa série de vazões máximas anuais, por exemplo, pode ocorrer o fato de alguns valores serem originados por eventos locais de precipitação (vazões de baixa magnitude), enquanto outros, mais extremos, serem originados de eventos climáticos regionais ou globais. Neste caso, há uma heterogeneidade e a amostra deve ser separada em duas sub amostras e analisadas sob contextos diferentes. A homogeneidade de uma série pode ser verificada, por exemplo, pelo teste de Mann-Whitney.

A estacionariedade ocorre quando não são verificadas tendências, saltos ou ciclos na série analisada, ou seja, as características estatísticas da série são conservadas ao longo do tempo. A construção de estruturas de controle ao longo dos cursos d'água, alterações climáticas, mudanças no uso do solo, entre outras, podem gerar não estacionariedades nas amostras observadas. O teste de Sperman é usado com frequência para testar a estacionariedade das séries de máximos anuais.

Para detalhes adicionais sobre a aplicação e a recomendação dos vários testes citados, o leitor deve remeter-se a Naghettini & Pinto (no prelo).

2.3.1.2 Escolha da distribuição de probabilidades

Na análise de frequência de máximos anuais, várias técnicas têm sido utilizadas para identificar a distribuição de probabilidades mais adequada à amostra em questão. A despeito do uso dessas técnicas, Hosking & Wallis (1997) fazem as seguintes ponderações quanto à análise das distribuições de probabilidades:

a) Distribuições com limite superior

Impor um limite superior ao modelo probabilístico pode comprometer a obtenção de boas estimativas de quantis para os tempos de retorno que realmente interessam. Embora possa se considerar como fisicamente impossível a ocorrência de determinados valores para casos específicos, o uso de distribuições ilimitadas superiormente pode oferecer melhores estimativas para os quantis do que aquelas limitadas. A menos que se tenha evidências de que a população possui limite superior conhecido, deve-se evitar a imposição de tal limite a priori.

b) Cauda superior da distribuição de probabilidade

O “peso” da cauda superior de uma distribuição de probabilidade determina a taxa de crescimento dos quantis em relação ao aumento do período de retorno na faixa das probabilidades pequenas (período de retorno tendendo ao infinito). A tabela 2.1 mostra a forma da cauda superior de várias distribuições. Uma vez que as curtas amostras de dados hidrológicos não possibilitam a correta definição da forma da cauda superior, a escolha da distribuição a ser utilizada deve ser feita levando-se em consideração um conjunto de distribuições que cubram todas as faixas de caudas superiores possíveis.

Tabela 2.1 – Peso da cauda superior de algumas distribuições

Cauda superior	Forma assintótica ¹	Distribuição ²
Pesada	x^{-a}	GEV, GPA e GLO com parâmetro de forma $k < 0$
↑	$x^{-a \ln x}$	LNO com assimetria positiva
	$\exp(-x^a) \quad 0 < a < 1$	WEI com parâmetro de forma $\lambda < 1$
	$x^a \exp(-bx)$	PE3 com assimetria positiva
	$\exp(-x)$	EXP e GUM
↓	$\exp(-x^a) \quad a > 1$	WEI com parâmetro de forma $\lambda > 1$
Leve	Limitada superiormente	GEV, GPA e GLO com parâmetro de forma $k > 0$. LNO e PE3 com assimetria negativa

1 – a e b são constantes positivas. (adap. de Hosking & Wallis, 1997, p. 75)

2 – Siglas:

- GEV: Generalizada de valores extremos
- GPA: Generalizada Pareto
- GLO: Logística generalizada
- LNO: Log-normal
- WEI: Weibull
- PE3: Pearson III
- EXP: Exponencial
- GUM: Gumbel

Considerações semelhantes às que foram feitas para o limite superior e para a cauda superior das distribuições podem ser feitas para a cauda e para o limite inferior. No entanto, duas ponderações devem ser acrescentadas: no que diz respeito à cauda inferior, NRC (1988) preconiza que a presença de valores baixos na amostra podem comprometer a estimativa da cauda superior da distribuição. Desta forma, NRC (1988) recomenda que para estudos de quantis máximos deve-se dar maior valor às estatísticas de ordem superior; no que diz respeito ao limite inferior da distribuição, pode-se determinar a priori o valor mínimo possível para a amostra. Em alguns casos esse limite pode ser igualado a zero e o parâmetro de posição da distribuição pode ser ajustado de modo a permitir essa imposição.

Cândido (2003) afirma que a inexistência de leis dedutivas para a seleção de uma distribuição de probabilidades ou de uma família de distribuições para a análise de frequência de eventos hidrológicos máximos anuais remete ao analista a decisão de escolher uma determinada distribuição de probabilidade. A decisão é tomada baseando-se em análise multicritério, onde são ponderados fatores subjetivos e testes estatísticos.

Dentre os fatores subjetivos, enquadra-se o exame visual do ajuste entre a distribuição candidata e os dados observados. Pode-se construir um papel de probabilidades no qual os quantis calculados por uma determinada distribuição formam uma linha reta. Assim, a análise da tendência linear dos dados observados plotados nesse papel pode indicar que a amostra se ajusta ou não à distribuição, muito embora, todas as expressões de probabilidade empíricas sejam imprecisas, particularmente aquelas associadas a valores extremos.

A análise visual, apesar de útil, e em alguns casos imprescindível, é pouco informativa e confiável. Desta forma, podem ser propostos vários testes de hipótese de aderência. Os testes de hipótese mais utilizados na prática são o do Qui-Quadrado, o de Kolmogorov-Smirnov, o de Filliben e o de comparação de quocientes de momentos-L. Embora os dois primeiros sejam de uso mais frequente, vários autores recomendam o uso dos dois últimos (STEDINGER *et al.*, 1993; HOSKING & WALLIS, 1997; CÂNDIDO, 2003; NAGHETTINI & PINTO, no prelo). Os testes de hipótese, bem como suas aplicações, são descritos detalhadamente em Naghettini & Pinto (no prelo).

Cabe ressaltar que, de acordo com Cândido (2003), dentre os testes de hipóteses, não há nenhum suficientemente potente para analisar as amostras de pequeno tamanho encontradas em hidrologia.

WMO (1989) afirma que os testes de hipótese não são conclusivos na detecção da distribuição de probabilidade a ser utilizada. Os testes podem rejeitar algumas distribuições, mas não são capazes de distinguir entre as várias distribuições aceitas.

2.3.1.3 Estimativa dos parâmetros da distribuição

Uma distribuição de probabilidades é definida por um conjunto de parâmetros que a descreve integralmente. Uma vez que não se conhece os parâmetros populacionais, os mesmos devem ser estimados a partir das amostras observadas. Considerando-se que as estimativas dos parâmetros são funções de variáveis aleatórias, verifica-se que os parâmetros advindos de uma amostra também serão variáveis aleatórias.

A necessidade da estimação dos parâmetros se faz no cálculo dos quantis e das probabilidades associadas a quantis. Dentre os métodos de estimação dos parâmetros das distribuições, destacam-se três: o método dos momentos (MOM), o método de máxima verossimilhança (MVS) e o método dos momentos-L (MML).

a) Método dos momentos

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

$$\mu_i(\theta_1, \theta_2, \dots, \theta_k) = m_i ; \quad i = 1, 2, \dots, k \quad (2.39)$$

A solução do sistema, definido pelas k equações 2.39, fornece as estimativas dos parâmetros da distribuição.

b) Método da máxima verossimilhança

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

$$L(\theta_1, \theta_2, \dots, \theta_k) = \prod_{i=1}^N f_X(x_i; \theta_1, \theta_2, \dots, \theta_k) \quad (2.40)$$

A equação 2.40 é conhecida como função de verossimilhança. O método da máxima verossimilhança consiste em se maximizar a equação 2.40. Para tal, as derivadas primeiras da função, em relação aos parâmetros, são igualadas a zero, ou seja,

$$\frac{\partial L(\theta_1, \theta_2, \dots, \theta_k)}{\partial \theta_j} = 0; \quad j = 1, \dots, k \quad (2.41)$$

As soluções do sistema, definido pelas k equações 2.41, fornecem as estimativas dos parâmetros da distribuição. Em grande parte dos casos, é mais conveniente trabalhar-se com a função logarítmica de verossimilhança $\ln[L(\theta_1, \theta_2, \dots, \theta_k)]$, por proporcionar maior simplicidade à solução do sistema formado por suas derivadas primeiras.

c) Método dos momentos-L

O método dos momentos-L (MML) para a estimação de parâmetros de distribuições de probabilidades é semelhante ao método dos momentos convencionais. No entanto, para a estimação dos parâmetros são utilizadas as estatísticas-L e seus estimadores amostrais.

De uma forma geral e para grandes amostras, o MVS é mais eficiente que os demais, pois produz estimadores de menor variância. No entanto, para amostras curtas os estimadores do MML são mais acurados. Além disso, o cálculo dos parâmetros pelo MML exige um esforço computacional significativamente menor que o MVS. Em geral, o método MOM fornece as estimativas mais pobres dentre os três métodos apresentados; por outro lado, é o mais simples de se estabelecer na prática.

2.3.1.4 Identificação e tratamento dos pontos atípicos ou outliers

Segundo Hawkins (1980), *outlier* é uma observação que se desvia de tal maneira dos outros pontos da amostra que parece ter sido gerada por fenômenos diferentes daqueles que geraram os pontos típicos. Por outro lado, observações atípicas podem ser frutos de medições erradas, mudança de seção de medição, dentre outros fatores decorrentes da prática de monitoramento de variáveis hidrológicas.

Uma técnica para a identificação de *outliers* de uso corrente em hidrologia, e recomendada pelo United States Water Resources Council (USWRC, 1976), é a estatística modificada de Grubbs & Beck (1972). Por essa técnica, os limites para *outliers* baixos e altos são dados por:

$$X_{baixo} = \exp(\bar{X}_{\ln} - K_{n,\alpha} S_{\ln}) \quad (2.42)$$

$$X_{alto} = \exp(\bar{X}_{\ln} + K_{n,\alpha} S_{\ln}) \quad (2.43)$$

nas quais \bar{X}_{\ln} e S_{\ln} representam, respectivamente, a média aritmética e o desvio-padrão dos logaritmos naturais da amostra, e $K_{n,\alpha}$ denota o valor crítico da estatística de Grubbs e Beck, para um nível de significância α , de uma amostra de tamanho n . Para $\alpha = 10\%$, Pilon (1985) propõe a seguinte aproximação para o valor crítico da estatística:

$$K_{n,\alpha=0,10} = -3,62201 + 6,28446n^{1/4} - 2,49835n^{1/2} + 0,491436n^{3/4} - 0,03791 \ln \quad (2.44)$$

Uma vez identificado um *outlier*, a decisão de mantê-lo na amostra é um tanto complicada. Neste ponto, deve-se valer do grau de certeza de que se trata realmente de um ponto atípico da amostra e qual é a fonte dessa atipicidade.

2.3.2 Séries de duração parcial

Uma série de duração parcial é formada por todos os eventos que excedem um determinado limiar u . Assim, a série de duração parcial, ao contrário da série dos máximos anuais, permite a utilização de um número de dados superior ao número de anos da série hidrológica. O uso de séries de duração parcial é de especial interesse nos casos em que não se dispõe de um conjunto de amostras longas.

Além de permitir a utilização de uma amostra de tamanho maior do que o número de anos de registros, a série de duração parcial contém os maiores eventos observados, independentemente desses serem os máximos anuais. Nesse sentido, e dependendo do valor mínimo da limiar escolhido, o emprego das séries de duração parcial faz com que a inferência estatística leve em consideração somente as maiores observações disponíveis, em detrimento daquelas muito freqüentes e de menor importância, as quais poderiam vir a distorcer ou prejudicar a estimação da distribuição de probabilidades dos eventos raros.

A descrição dos modelos de séries de duração parcial é dada em detalhes no item 3.2. De acordo com a formulação proposta por Todorovic & Zelenhasic (1970), se a série de duração parcial é modelada por uma distribuição $H_u(x)$, para um determinado limiar u , a distribuição dos máximos anuais pode ser obtida da seguinte forma:

$$F_m(x) = \exp\{-\Lambda[1 - H_u(x)]\} \quad (2.45)$$

onde Λ representa a intensidade anual de ocorrências, segundo um processo homogêneo de Poisson.

A distribuição $H_u(x)$ tem sido modelada por várias distribuições de probabilidade. Em especial, são aqui descritos os modelos gerados pela distribuição generalizada de Pareto, generalizada de valores extremos e a logística generalizada, os quais são denominados, respectivamente, modelo Poisson-Pareto, Poisson-GEV e Poisson-Logística.

No item 3.2, será vista uma outra forma de modelar a distribuição $H_u(x)$ que leva em consideração a distribuição dos volumes de duração especificada e a distribuição dos picos de vazão condicionada aos volumes. Essa outra forma de modelar a distribuição $H_u(x)$ é, em última análise, o cerne da metodologia de cálculo de quantis máximos proposta neste trabalho.

- Modelo Poisson-Pareto

Neste caso, $H_u(x)$ é modelada pela distribuição generalizada de Pareto, ou seja,

$$H_u(x) = 1 - \exp(-y) \quad (2.46)$$

na qual,

$$y = -\frac{1}{k} \ln \left[1 - \frac{k}{\alpha} (x - u) \right]; \quad \text{se } k \neq 0$$

ou

$$y = \frac{x - u}{\alpha}; \quad \text{se } k = 0$$

e u , α e k são os parâmetros de posição, forma e escala, respectivamente.

A equação 2.45 pode ser reescrita da seguinte forma:

$$H_u(x) = 1 + \frac{1}{\Lambda} \ln[F_m(x)] \quad (2.47)$$

A função dos quantis para o modelo Poisson-Pareto é obtida, então, igualando-se 2.46 e 2.47:

$$x(F) = u + \frac{\alpha}{k} \left\{ 1 - [-\ln(F)]^{k/\Lambda} \right\}; \quad \text{se } k \neq 0 \quad (2.48)$$

$$x(F) = u + \alpha [-\ln(F)]^{1/\Lambda}; \quad \text{se } k = 0 \quad (2.49)$$

- Modelo Poisson-GEV

Neste caso $H_u(x)$ é modelada pela distribuição generalizada de valores extremos, ou seja:

$$H_u(x) = \exp[-\exp(-y)] \quad (2.50)$$

na qual,

$$y = -\frac{1}{k} \ln \left[1 - \frac{k}{\alpha} (x - u) \right]; \quad \text{se } k \neq 0$$

ou

$$y = \frac{x - u}{\alpha}; \quad \text{se } k = 0$$

e u , α e k são os parâmetros de posição, forma e escala, respectivamente.

A função dos quantis para o modelo Poisson-GEV é obtida, então, igualando-se 2.47 e 2.50:

$$x(F) = u + \frac{\alpha}{k} \left\{ 1 - \left[-\ln \left(\frac{\Lambda + \ln F}{\Lambda} \right) \right]^k \right\}; \quad \text{se } k \neq 0 \quad (2.51)$$

$$x(F) = u + \alpha \left[-\ln \left(\frac{\Lambda + \ln F}{\Lambda} \right) \right]; \quad \text{se } k = 0 \quad (2.52)$$

- Modelo Poisson-Logística

Neste caso, $H_u(x)$ é modelada pela distribuição logística generalizada. Na parametrização proposta por Hosking & Wallis (1997), tem-se:

$$H_u(x) = \frac{1}{1 + \exp(-y)} \quad (2.53)$$

na qual,

$$y = -\frac{1}{k} \ln \left[1 - \frac{k}{\alpha} (x - u) \right]; \quad \text{se } k \neq 0$$

ou

$$y = \frac{x - u}{\alpha}; \quad \text{se } k = 0$$

e u , α e k são os parâmetros de posição, forma e escala, respectivamente.

Igualando-se 2.47 e 2.53, obtêm-se:

$$x(F) = u + \frac{\alpha}{k} \left\{ 1 - \left[-\left(\frac{\ln F}{\Lambda + \ln F} \right) \right]^k \right\}; \quad \text{se } k \neq 0 \quad (2.54)$$

$$x(F) = u - \alpha \ln \left(-\frac{\ln F}{\Lambda + \ln F} \right); \quad \text{se } k = 0 \quad (2.55)$$

2.4 Métodos regionais de estimação

Segundo Tucci (2000), a regionalização pode ser entendida como a técnica de transferência de informações de um local para outro dentro de uma área com comportamento hidrológico semelhante. Além disso, a regionalização hidrológica pode servir também ao propósito de

explicar a variância espacial, por exemplo, de um parâmetro de um modelo distributivo, ou um quantil por ele produzido.

CPRM (2001) afirma que a análise de frequência regional não se restringe apenas à necessidade de transferência espacial de variáveis hidrológicas, mas também à obtenção de estimativas mais confiáveis de parâmetros e quantis de uma distribuição teórica de probabilidades, identificação de regiões com carência de postos de observação e verificação da consistência das séries hidrológicas. Nesse contexto, conforme verificado por Lettenmaier & Potter (1985), o uso de regionalização pode produzir a redução das incertezas associadas às estimativas de quantis, quando comparadas àqueles produzidos pela análise local.

Tucci (1993) classifica os métodos de regionalização em três classes: métodos de regionalização dos parâmetros da distribuição de probabilidades, métodos de regionalização do evento com um determinado risco e os métodos de regionalização da curva adimensional de frequências, denominados métodos de regionalização *index-flood*. A flexibilidade e a possibilidade de se incluir na análise características do processo físico de formação de vazões têm feito do método *index-flood* o de maior aplicação.

De acordo com Pinto & Naghettini (1999), independentemente do método de regionalização a ser utilizado, um dos pontos cruciais é a definição de regiões estatisticamente homogêneas, ou seja, aquelas contendo várias estações cujas séries sejam oriundas de populações regidas pela mesma distribuição de probabilidades, com os parâmetros de posição e escala variando entre as estações.

Hosking & Wallis (1997) desenvolveram um conjunto metodológico para análise regional, baseado no método *index-flood*, que utiliza as estatísticas-L de modo unificado e conciso em todas as etapas de sua aplicação. A regionalização por momentos-L permite uma análise menos subjetiva do comportamento regional das variáveis hidrológicas e, por esse fato, tem sido alvo dos mais recentes estudos no campo da determinação de quantis de vazão. No Brasil, a metodologia foi aplicada com sucesso nos estudos de chuvas intensas da Região Metropolitana de Belo Horizonte (PINHEIRO, 1997) e do estado do Rio de Janeiro (DAVIS & NAGHETTINI, 2000), e na regionalização de vazões máximas e mínimas das sub-bacias do alto São Francisco (CPRM, 2001).

2.4.1 O método do *index-flood*

O método do *index-flood*, desenvolvido por Dalrymple (1960), refere-se a uma forma de adimensionalizar uma determinada variável hidrológica X de vários pontos de observação de modo a utilizá-las como um conjunto comum.

Seja a variável hidrológica X , amostrada em m locais dentro de uma determinada região geográfica com amostras de tamanho n_i ($i = 1, 2, \dots, m$) em cada posto. Seja, também, a função $X_i(F)$, que representa os quantis com probabilidade de não excedência F para o posto i . O método parte do pressuposto que, se os postos formam uma região estatisticamente homogênea, então a função dos quantis é a mesma para todos os postos a menos de um fator de escala. Tal fator de escala é denominado *index-flood*. Assim:

$$X_i(F) = \mu_i \cdot x(F) \quad (2.56)$$

sendo μ_i o fator de escala ou *index-flood* para o posto i e $x(F)$ a curva regional de quantis adimensionais, comum a todos os postos.

O *index-flood* é estimado com base na média amostral da variável X para cada posto. A variável adimensionalizada $x_{i,j} = X_{i,j} / \mu_i$ ($i = 1, 2, \dots, m; j = 1, 2, \dots, n_i$) é usada para estimar os parâmetros regionais da função $x(F)$.

Hosking & Wallis (1997) usaram como estimador dos parâmetros regionais da função $x(F)$ a média ponderada pelo tamanho da amostra (n_i) dos parâmetros locais de cada posto.

A aplicação do método *index-flood* pressupõe o atendimento das seguintes premissas:

- a) as observações adimensionalizadas, em um posto qualquer, são identicamente distribuídas;
- b) as observações em um posto qualquer não apresentam dependência estatística serial;
- c) as observações em diferentes postos são estatisticamente independentes, muito embora, essa seja uma suposição de difícil verificação prática, com conseqüências importantes na precisão de estimativas de quantis (MARTINS & STEDINGER, 2002);

d) as distribuições de frequência em diferentes postos são idênticas, a menos de um fator de escala; e

e) a forma matemática da curva regional de quantis adimensionalizados é corretamente especificada.

2.4.2 Etapas para a regionalização por momentos-L

A seqüência a ser seguida para a regionalização por momentos-L é a seguinte:

- Análise de consistência dos dados

Nessa etapa são verificados erros amostrais das séries e identificados os postos com dados muito discrepantes em relação aos outros postos da região.

- Identificação de regiões homogêneas

Os m postos da região são divididos em grupos de tal forma que as amostras de cada grupo possam ser modeladas por uma única distribuição de probabilidades.

- Escolha da distribuição de frequência regional

Para cada região definida no item anterior é selecionada a distribuição de probabilidades que melhor modela os dados adimensionalizados das amostras da região.

- Estimação dos parâmetros regionais das distribuição de cada região homogênea

Os parâmetros de cada posto são estimados individualmente e em seguida ponderados de forma a estimar os parâmetros regionais.

2.4.2.1 Análise de consistência dos dados

A etapa inicial de qualquer análise de frequência é a verificação da qualidade dos dados a serem trabalhados. Para isso, além dos testes sugeridos no item 2.3.1.1, Hosking & Wallis (1997) propõem uma medida para verificar, dentre as amostras, aquelas que estão grosseiramente discordantes com o grupo. Descreve-se a seguir as etapas necessárias para o cálculo da medida de discordância.

Para uma região homogênea com m postos, seja o vetor $\vec{u}_i = [t^{(i)}, t_3^{(i)}, t_4^{(i)}]^T$ aquele formado pelo Coeficiente de Variação – L, pelo Coeficiente de Assimetria – L e pelo Coeficiente de Curtose – L amostrais, segundo as equações 2.36, 2.37 e 2.38, respectivamente, do posto i .

O vetor formado pelos momentos-L amostrais regionais é obtido a partir da média dos vetores locais:

$$\vec{u}_R = \frac{1}{m} \sum_{i=1}^m \vec{u}_i = [t^R, t_3^R, t_4^R]^T \quad (2.57)$$

Seja, agora, a matriz de covariância \dot{A} , definida da seguinte forma:

$$\dot{A} = \sum_{i=1}^m (\vec{u}_i - \vec{u}_R)(\vec{u}_i - \vec{u}_R)^T \quad (2.58)$$

A medida de discordância é definida por:

$$Di = \frac{1}{3} m (\vec{u}_i - \vec{u}_R)^T \dot{A}^{-1} (\vec{u}_i - \vec{u}_R) \quad (2.59)$$

O posto i é dito discordante caso o valor dado por 2.59 seja maior que o valor crítico para essa medida. Hosking & Wallis (1997) sugerem um valor crítico variando com o número de postos na região de acordo com a tabela 2.2.

Tabela 2.2 – Valores críticos para a medida de discordância

Nº de postos na região	Valor crítico	Nº de postos na região	Valor crítico
5	1,333	11	2,632
6	1,648	12	2,757
7	1,917	13	2,869
8	2,140	14	2,971
9	2,329	≥ 15	3,000
10	2,491		

Fonte: Hosking & Wallis (1997)

2.4.2.2 Identificação das regiões homogêneas

Pela definição de região homogênea, todos os postos de seu interior devem possuir os mesmos quocientes de momentos-L populacionais. No entanto, as flutuações estatísticas fazem com

que os quocientes amostrais difiram dos populacionais. Assim, uma maneira de se estabelecer um critério para verificar a homogeneidade da região é comparar a dispersão dos quocientes de momentos-L amostrais com a dispersão esperada para uma região homogênea. Hosking & Wallis (1997) consideram suficiente comparar somente a variação do CV-L, uma vez que esse exerce um efeito muito mais significativo na estimativa de quantis extremos que a assimetria-L e a curtose-L.

A dispersão esperada para a região homogênea é obtida por simulação através do método de Monte Carlo. Hosking & Wallis (1997) recomendam o uso da distribuição Kapa de 4 parâmetros para simular os dados da região homogênea e justificam essa recomendação pela preocupação de não se comprometer, *a priori*, com as distribuições de 2 e 3 parâmetros. A distribuição Kapa inclui como caso particular a logística generalizada, a generalizada de valores extremos e a generalizada de Pareto e, devido a isso, é capaz de representar a maioria das variáveis hidrológicas.

Suponha uma região com m postos, sendo que cada posto contém uma série de tamanho n_i e quocientes de momentos-L $t^{(i)}, t_3^{(i)}, t_4^{(i)}$. Sejam t^R, t_3^R e t_4^R os quocientes de momentos-L regionais dados pela média ponderada dos quocientes locais.

O desvio-padrão ponderado do CV-L amostral é dados por:

$$V = \sqrt{\frac{\sum_{i=1}^m n_i (t^{(i)} - t^R)^2}{\sum_{i=1}^m n_i}} \quad (2.60)$$

Considere, agora, um conjunto de N_{sim} simulações de regiões com m postos geradas a partir de uma distribuição Kapa ajustada aos quocientes 1, t^R, t_3^R e t_4^R . Para cada simulação tem-se a estimativa V_j ($j = 1, 2, \dots, N_{sim}$) do desvio-padrão do CV-L. A medida de heterogeneidade é calculada a partir da média (μ_V) e do desvio-padrão (σ_V) dos N_{sim} valores de V_j da seguinte forma:

$$H = \frac{V - \mu_V}{\sigma_V} \quad (2.61)$$

Os valores críticos para a medida de heterogeneidade são:

para $H < 1$, a região é homogênea;

para $1 \leq H < 2$, a região é possivelmente heterogênea; e

para $H \geq 2$, a região é definitivamente heterogênea.

2.4.2.3 Escolha da distribuição de frequência regional

Dentre as várias distribuições de probabilidade de uso freqüente em hidrologia estatística, é necessário escolher entre modelos candidatos cujas características sejam consistentes com os padrões de variabilidade da variável em questão. Hosking & Wallis (1997) afirmam que a melhor distribuição não é, necessariamente, aquela que fornece o melhor ajuste e sim aquela que fornece as melhores estimativas para os quantis com período de retorno grande. Nesse sentido, são válidas, também, as ponderações feitas no item 2.3.1.2, principalmente acerca da cauda e do comportamento assintótico das distribuições.

No que se refere à aderência das distribuições, Hosking & Wallis (1997) propuseram uma medida que avalia a capacidade de a distribuição candidata reproduzir um momento-L, de ordem superior ao número de parâmetros que a identifica, próximo ao amostral. Se uma distribuição de 3 parâmetros, tal como a GEV, é ajustada aos dados regionais, ela terá a mesma assimetria-L da média regional. Portanto, a medida de aderência irá verificar a capacidade da GEV em reproduzir a curtose-L regional.

Suponha uma região com m postos, sendo que cada posto contém uma série de tamanho n_i e quocientes de momentos-L $t^{(i)}$, $t_3^{(i)}$ e $t_4^{(i)}$. Seja t^R , t_3^R e t_4^R os quocientes de momentos-L regionais dados pela média ponderada dos quocientes locais.

Dada uma distribuição qualquer de 3 parâmetros, denotada aqui por $DIST$, a medida de aderência por ser estabelecida a partir dos seguintes passos:

- Ajuste da distribuição candidata aos momentos-L 1, t^R e t_3^R e cálculo de sua curtose-L (τ_4^{DIST});
- Ajuste da distribuição Kapa aos quocientes de momentos-L regionais 1, t^R , t_3^R e t_4^R ;

- Simulação de uma grande quantidade (N_{sim}) de regiões com m postos a partir da distribuição Kapa;
- Cálculo do viés (B_4) e do desvio-padrão (σ_4) da curtose-L regional de acordo com as equações 2.62 e 2.63, respectivamente;

$$B_4 = \frac{\sum_{j=1}^{N_{sim}} (t_4^j - t_4^R)}{N_{sim}} \quad (2.62)$$

$$\sigma_4 = \sqrt{\frac{\sum_{j=1}^{N_{sim}} (t_4^j - t_4^R)^2 - N_{sim} B_4^2}{N_{sim} - 1}} \quad (2.63)$$

A medida de aderência é dada pela equação 2.64, qual seja:

$$Z^{DIST} = \frac{\tau_4^{DIST} - t_4^R + B_4}{\sigma_4} \quad (2.64)$$

A distribuição se ajusta adequadamente aos dados se $|Z^{DIST}| \leq 1,64$.

2.4.2.4 Estimação dos parâmetros regionais da distribuição de probabilidades

Escolhida a distribuição de probabilidades conforme feito anteriormente, os parâmetros podem ser obtidos pelo método dos momentos-L de acordo com o exposto no item 2.3.1.3. Nesse caso, as estimativas dos momentos-L amostrais a serem utilizadas são dadas pelas médias ponderadas dos momentos-L de cada posto da região homogênea.

2.5 *Análise de freqüência de vazões simuladas por modelos chuva-vazão*

A idéia central desta abordagem é analisar longas séries de vazões obtidas pela simulação de longas séries de precipitação.

Os modelos chuva-vazão são empregados para a obtenção de séries de vazões a partir de séries de precipitação. Podem ser citados, como exemplos, os modelos SCS, SMAP, ARNO, RIO GRANDE, entre outros.

Os modelos chuva-vazão são de especial interesse na análise hidrológica pois, na maioria dos casos, os dados de chuva são mais abundantes e espacialmente melhor distribuídos que os dados de vazão. Além disso, os dados de chuva são menos sujeitos a erros amostrais e a interferências antrópicas, o que permite a simulação de cenários próximos às condições naturais de uma certa bacia.

Lima (2004) desenvolveu uma metodologia que utiliza um modelo de geração de chuva e um modelo de transformação chuva-vazão de forma integrada, que permite obter séries extensas de vazão para uma determinada bacia. Foi utilizado um modelo estocástico para geração de séries de chuva e posteriormente foram geradas séries de vazão simulada a partir do modelo RIO GRANDE. A seguir são apresentados os principais módulos do modelo integrado desenvolvido por Lima (2004).

2.5.1 Modelo estocástico de geração de dados de chuva

Em um modelo estocástico, assume-se que a variável em análise é fruto de um processo aleatório, não sendo possível determinar seu valor futuro com precisão absoluta. No entanto, pode-se estabelecer uma série sintética que possua as mesmas características estatísticas (eg. variância, média, sazonalidade, dispersão, entre outras) dos valores observados. Essencialmente, essa é a lógica da construção dos modelos de geração de séries sintéticas.

Lima (2004) utilizou a matriz de probabilidade de transição ou cadeia de Markov em multi-estados, como modelo de geração estocástica de precipitação. Foi utilizada uma cadeia de Markov de ordem 1, ou seja, a probabilidade de um determinado valor em um tempo t depende somente do valor no tempo anterior $t - 1$.

A matriz de probabilidade de transição é dada pela matriz de elementos p_{ij} tal que:

$$P(X_t = a_j | X_{t-1} = a_i) = p_{ij} \quad (2.65)$$

Ou seja, p_{ij} é a probabilidade de a variável X estar no estado a_j no tempo t dado que no tempo $t-1$ ela estava no estado a_i .

Além disso, o modelo deve ser capaz de representar a variabilidade sazonal das precipitações. Lima (2004) utilizou 12 estações como forma de representar a distribuição de chuva em cada

mês do ano. Os estados possíveis para a chuva dentro de cada estação (mês) correspondem à discretização dos dados observados em intervalos de classe. O último intervalo de classe de cada estação não deve ser limitado superiormente, possibilitando a ocorrência de eventos raros de precipitação. Para isso, a mesma autora ajustou os dados do último estado por uma distribuição de probabilidades.

Dentro de cada estado, a chuva foi suposta ser distribuída uniformemente e, ao longo dos estados possíveis, a chuva foi suposta ser distribuída exponencialmente. A tabela 2.3 mostra um exemplo de estados possíveis para a chuva média da bacia do rio Pará. A tabela 2.4 mostra a matriz de transição para o mês de janeiro referente aos estados da tabela 2.3.

Tabela 2.3 – Estados para a chuva diária da bacia do rio Pará

Estado	Precipitação – P (mm)
1	Zero
2	$Zero < P \leq 1,05$
3	$1,05 < P \leq 3,50$
4	$3,50 < P \leq 7,50$
5	$7,50 < P \leq 13,3$
6	$13,30 < P \leq 23,0$
7	$23,0 < P$

Fonte: Lima (2004)

Tabela 2.4 – Matriz de probabilidade de transição para o mês de janeiro na bacia do rio Pará

Estado Corrente	Próximo Estado						
	1	2	3	4	5	6	7
1	0,63	0,09	0,08	0,08	0,05	0,05	0,03
2	0,27	0,18	0,18	0,16	0,1	0,1	0,02
3	0,22	0,11	0,16	0,16	0,13	0,12	0,11
4	0,2	0,09	0,17	0,15	0,15	0,13	0,11
5	0,14	0,05	0,12	0,17	0,14	0,19	0,19
6	0,09	0,07	0,09	0,16	0,15	0,19	0,25
7	0,08	0,05	0,06	0,13	0,07	0,28	0,33

Fonte: Lima (2004)

A série sintética pode ser obtida através das seguintes etapas:

- A partir da amostra de precipitação, são feitas as inferências estatísticas de modo a descrever a variabilidade amostral;
- São definidos os estados possíveis para a precipitação;

- Calcula-se a matriz de probabilidade de transição para cada mês do ano;
- Ajusta-se a distribuição de probabilidade que define o último estado através das estatísticas dos dados observados;
- São definidos o estado i e a estação k do valor inicial X_n ;
- Gera-se um número aleatório (Rn) num intervalo de 0 a 1;
- Se $Rn \leq p_{ij}^k$, então $X_{n+1} = 0$;
- Se $\sum_{j=1}^{n-1} p_{ij} > Rn \geq \sum_{j=1}^n p_{ij}$, gera-se uma observação (x) nesse intervalo através do método de Monte Carlo com a distribuição de probabilidades do intervalo e fixa-se o novo valor $X_{n+1} = x$; e
- São repetidos os dois passos anteriores, avançando no tempo, até se atingir o tamanho da série desejado.

Em linhas gerais, o processo permite obter uma série sintética que respeita as características estatísticas amostrais e as matrizes de probabilidade de transição.

2.5.2 Modelo de transformação chuva-vazão

O passo seguinte da metodologia de Lima (2004) é gerar uma série sintética de vazões a partir da série sintética de precipitações. Para isso, Lima (2004) utilizou o modelo de transformação chuva-vazão RIO GRANDE. O modelo RIO GRANDE, assim denominado em função da bacia onde foi inicialmente aplicado, é fruto de pesquisas recentes do Departamento de Engenharia Hidráulica e Recursos Hídricos da Escola de Engenharia da UFMG e foi descrito inicialmente por Naghettini et al. (2002).

O modelo RIO GRANDE é um modelo determinístico, conceitual, semi-distribuído e contínuo. Determinístico no sentido que, dada uma condição inicial (entrada), a resposta do modelo (saída) será sempre a mesma, não envolvendo elementos estocásticos; conceitual no sentido que a modelação dos fenômenos hidrológicos não é necessariamente “fisicamente” fundamentadas; semi-distribuído no sentido que é considerada parcialmente a distribuição

espacial das variáveis envolvidas no processo; e contínuo no sentido que a simulação é feita ao longo do tempo, ao contrário dos modelos que simulam um único evento, sem considerar a situação temporal dos estados de armazenamento da bacia.

O modelo RIO GRANDE é composto de três módulos: o módulo de produção, o de concentração e o de propagação.

2.5.2.1 Módulo de produção

O módulo de produção se refere à formação do escoamento superficial pela modelagem do balanço de umidade do solo. O balanço de umidade do modelo RIO GRANDE segue a estrutura do modelo Xinanjiang, proposto inicialmente por Zhao et al. (1980) e que se mostrou adequado para aplicações em regiões úmidas e semi-úmidas.

Não será descrita aqui a modelagem completa do módulo de produção do modelo. Para detalhes adicionais, o leitor deve remeter-se a Naghettini et al. (2002). No entanto, na figura 2.1 mostra-se o fluxograma que resume a produção do escoamento com base no balanço de umidade. As variáveis que controlam o módulo de produção são:

WM: capacidade de tensão média espacial, a qual é uma medida da aridez da bacia e varia entre 80 mm em áreas úmidas até 180 mm em áreas semi-úmidas.

B: expoente da curva de distribuição em área da tensão capilar, o qual é uma medida da não-uniformidade espacial. O parâmetro *B* pode ser tomado como uma função do tamanho da bacia: 0,1 para áreas de drenagem até 10 km² a 0,4 para áreas de alguns milhares de km².

IMP: fração de área impermeável da bacia; para bacias rurais, em geral, $IMP > 0,02$.

SM: capacidade média espacial de água livre, ou gravitacional, a qual controla a magnitude do escoamento superficial. O valor de *SM* depende da profundidade da camada superficial de maior condutividade hidráulica, podendo ter valores desde 10 mm para solos rasos até 50 mm ou mais para solos espessos e porosos.

EX: expoente da curva de distribuição espacial da curva de capacidade de água livre. Em geral, é tomado entre os valores de 0,5 a 2, com maior frequência entre 1,0 e 1,5.

KSS: coeficiente do fluxo diário do escoamento sub-superficial.

KG: coeficiente do fluxo diário do escoamento subterrâneo.

CI: coeficiente de recessão do fluxo sub-superficial diário, o qual pode ser facilmente obtido por análise de hidrogramas observados. Os valores usuais para *CI* variam entre 0,5 e 0,9.

CG: coeficiente de recessão do fluxo subterrâneo diário, igualmente passível de determinação por análise de hidrogramas observados. Os valores usuais para *CG* variam entre 0,96 e 0,998.

K: coeficiente de ajuste da ETP potencial.

WU: capacidade média espacial da tensão na zona superior, compreendida entre 5 mm para áreas desnudas até 25 mm para bacias com densa vegetação.

WL: capacidade média espacial da tensão na zona inferior, compreendida entre 60 e 90 mm.

WD: capacidade média espacial da tensão na zona profunda.

C: coeficiente de evapotranspiração profunda, com valores entre 0,08 até 0,2 para regiões muito úmidas. Este parâmetro, apesar de inativo durante os períodos de cheia, é importante durante os períodos secos.

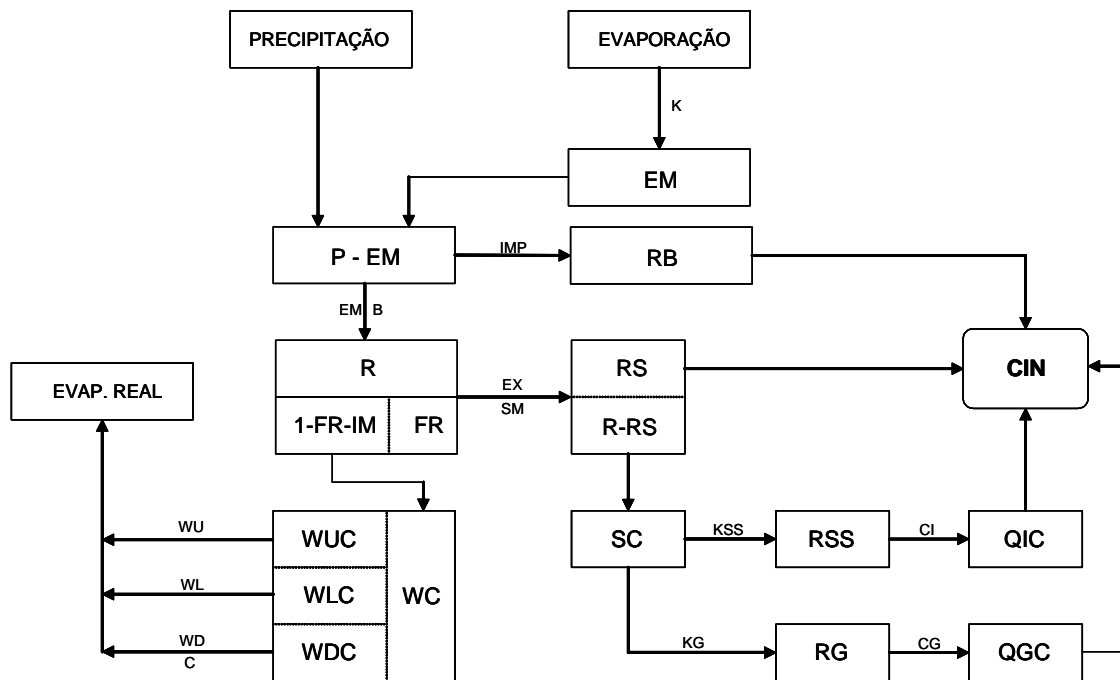


Figura 2.1 – Fluxograma do módulo de produção do modelo RIO GRANDE (adap. de Naghettini et al., 2002)

2.5.2.2 Módulo de concentração

A lâmina de escoamento gerada no módulo de produção (*CIN*) representa um pulso, no intervalo de tempo para o qual foi simulado, distribuído sobre uma área unitária da bacia. No modelo RIO GRANDE presume-se que essa lâmina esteja uniformemente distribuída sobre a bacia. No entanto, diferentes partes da bacia contribuirão em tempos diferentes para a vazão no exutório da bacia. À modelagem desse processo dá-se o nome de concentração.

No modelo RIO GRANDE a concentração é feita por meio de uma função de transferência baseada no método do hidrograma unitário de Clarke, formulado originalmente pelo Hydrologic Engineering Center of U.S. Army Corps of Engineers (HEC, 1981).

2.5.2.3 Módulo de propagação

Segundo Lima (2004), o módulo de propagação do modelo RIO GRANDE é responsável pela distribuição temporal das vazões provenientes do meio superficial (cursos d'água, lagos e reservatórios), do meio sub-superficial (zona aerada) e do meio subterrâneo (zona saturada) concentradas em um determinado trecho de um curso d'água.

No modelo RIO GRANDE a propagação é feita pelo modelo de onda difusiva conhecido por Muskingum-Cunge, o qual é uma abordagem aproximada das equações de Saint Venant.

2.6 Modelo Misto

Na presente dissertação, modelos mistos, são aqueles que fazem uso de componentes de outras metodologias, agregando suas particularidades de uma maneira lógica de forma a constituir um único modelo.

Visando atender às recomendações do NRC (1988), Naghettini (1994) desenvolveu um modelo misto que faz uso de três metodologias: método pico-volume, método GRADEX e séries de duração parcial. As duas primeiras serão descritas em detalhes no capítulo 3 e a última está descrita neste capítulo.

No contexto dessa metodologia, as vazões de pico excedentes sobre um valor limiar arbitrário u , denotadas por Y_i , e os volumes de cheia X_i , associados a essas excedências, são individualizados e modelados como um processo estocástico pontual marcado usando-se a representação de um processo composto de Poisson de intensidade $\lambda(t)$. A essência do método proposto por Naghettini (1994) consiste em se estimar separadamente a função densidade marginal de probabilidade $g_u(x)$ dos volumes de cheia, para uma duração equivalente ao tempo de base da bacia, e a função densidade $f_{Y|X}(y/x)$ das vazões de pico condicionadas aos volumes.

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

$$H_u(y_p) = \int_0^{y_p} \int_0^\infty f_{Y|X}(y|x) g_u(x) dx dy \quad (2.66)$$

e

$$F_M(y_p) = \exp \left\{ - [1 - H_u(y_p)] \int_0^1 \lambda(t) dt \right\} \quad (2.67)$$

onde $H_u(y_p)$ representa a função acumulada das excedências y_p sobre o valor limiar u e $F_M(y_p)$ a função acumulada anual das vazões de pico. A solução do sistema formado pelas

equações 2.66 e 2.67 fornece a probabilidade anual correspondente a um quantil y_p , ou inversamente y_p em função do tempo de retorno.

Os princípios (a), (b) e (c) identificados pelo NRC (1988), mostrados no capítulo 1, são incorporados à metodologia proposta por Naghettini (1994), principalmente na estimativa de $g_u(x)$, da seguinte forma: inicialmente, utiliza-se um modelo regional [princípio (a)], baseado em estatísticas amostrais de ordem superior [princípio (c)] de um conjunto de estações pluviométricas, para se testar a hipótese de exponencialidade da cauda superior da distribuição de probabilidades da altura de precipitação, cuja duração deve ser tomada como igual ao tempo de base de uma bacia dentro da área em estudo. Trata-se de um teste baseado na razão das funções logaritmos de verossimilhança sob a hipótese $H_0 \{k = 0\}$ e sob a alternativa $H_1 \{k \neq 0\}$, onde k representa a estimativa regional do parâmetro de forma de uma distribuição generalizada de Pareto, da qual a exponencial é um caso particular ($k = 0$). Caso não se possa rejeitar a hipótese $H_0 \{k = 0\}$, atende-se então à premissa básica da metodologia GRADEX, desenvolvida por Guillot e Duband (1967), para se transferir a informação hidrometeorológica para as curvas de frequência de volumes de cheia. De acordo com o método GRADEX, a distribuição de probabilidades dos volumes de cheia, cuja função densidade denota-se por $g_u(x)$, pode ser deduzida da distribuição das precipitações de mesma duração efetuando-se a translação das distribuições acumuladas de uma distância fixa ao longo do eixo da variável. A distância de translação é função do gradex, ou parâmetro de escala, da distribuição de precipitações, a qual deve ter necessariamente um comportamento assintótico exponencial. Supõe-se aqui, em condições de saturação e, portanto, para elevados tempos de retorno, qualquer incremento da altura de chuva irá provocar igual incremento do volume de cheia, desde que tomados sob a mesma duração; observa-se uma clara aplicação do princípio (b) do NRC.

A função $f_{Y/X}(y/x)$ é estimada a partir de um modelo de regressão não paramétrica e não linear entre as vazões de pico e os volumes de cheia. Observe que na equação 2.66 a integração deve ser efetuada para elevados valores do quantil y_p ; em consequência, a função densidade $f_{Y/X}(y/x)$ deve ser extrapolada para esses elevados valores y_p . Como não se dispõe de valores observados dessa magnitude, Naghettini (1994) utilizou o modelo de transformação chuva-vazão HSP-F, descrito por Johanson et al. (1984), para simular tormentas extremas transpostas para a bacia de interesse e, em seguida, obter os pares [vazão

de pico, volume de cheia] necessários para a extrapolação da função $f_{Y/X}(y/x)$. A fonte básica para a escolha dessas tormentas críticas foi o catálogo de eventos históricos extremos compilado pelo U.S. Army Corps of Engineers (1945); esse catálogo descreve os eventos chuvosos mais severos que ocorreram nos Estados Unidos nos últimos 100 anos anteriores à sua publicação.

Naghetini (1994) aplicou com sucesso a metodologia acima à bacia do rio Blue, localizada no estado americano de Oklahoma. A bacia do rio Blue possui área de drenagem de 1233 km² e, como era de se esperar, possui registros fluviométricos relativamente mais curtos (cerca de 53 anos de dados) e espacialmente menos abundantes (uma estação) do que os de observações pluviométricas (10 estações com 80 anos de dados comuns) em seu interior e vizinhança. Verifica-se aqui a principal motivação inerente à agregação da informação hidrometeorológica à análise de frequência de cheias, ou seja, a de que os registros pluviométricos são, na maioria dos casos, mais longos, menos sujeitos a erros de medição mais facilmente regionalizáveis do que os registros fluviométricos.

A figura 2.2 mostra os resultados parciais da aplicação da metodologia à bacia do rio Blue. A figura 2.2a mostra a aplicação do método GRADEX para a obtenção da distribuição dos volumes, considerando que a bacia em questão tem tempo de base de 5 dias. Na figura 2.2b são mostrados os pares [vazão de pico, volume de cheia] observados e simulados utilizados para a obtenção da função $f_{Y/X}(y/x)$.

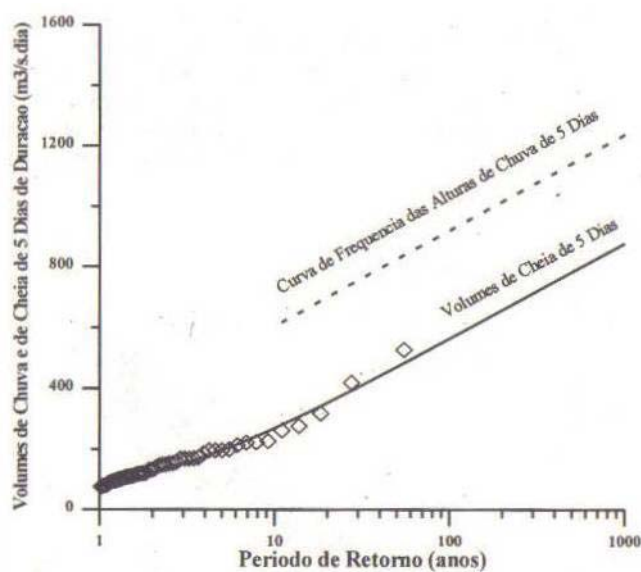


Figura 2.2a – Distribuição de frequência de volumes de cheia (adap. de Naghetini, 1994)

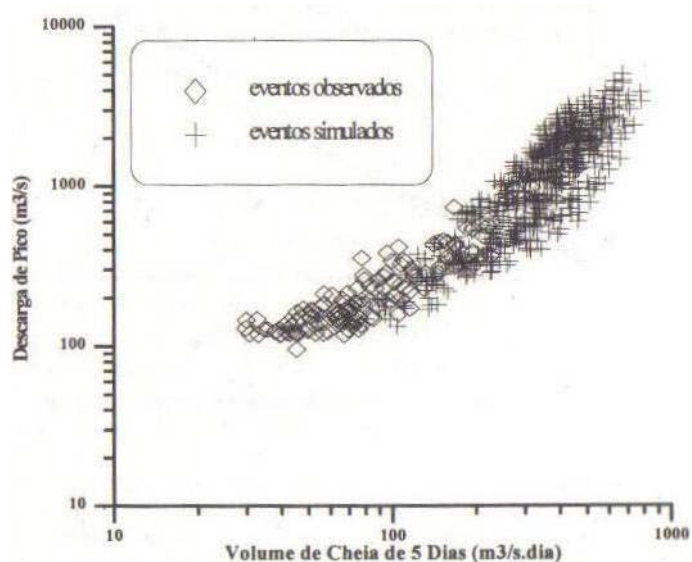


Figura 2.2b – Vazões de pico-volume de cheia simulados e observados (adap. de Naghettini, 1994)

Com o auxílio das funções obtidas na figura 2.2, Naghettini (1994) obteve a distribuição de frequência de vazões máximas anuais para o rio Blue pela solução do sistema formado pelas equações 2.66 e 2.67. A figura 2.3, a seguir, resume os resultados da aplicação à bacia do rio Blue.

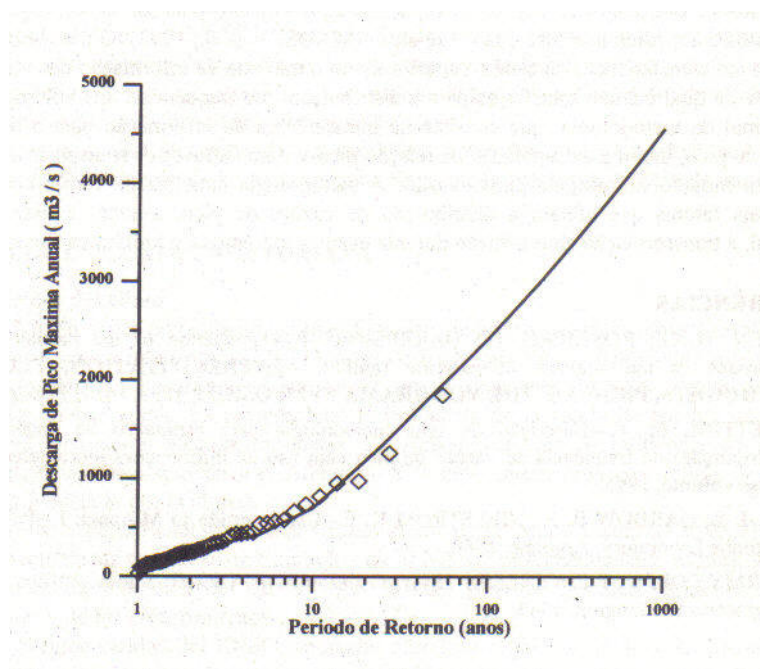


Figura 2.3 – Distribuição de frequência de vazões máximas (adap. de Naghettini, 1994)

Tal como originalmente formulada, a metodologia proposta por Naghettini (1994) apresenta limitações quanto à sua aplicação, podendo ser destacadas as seguintes:

- Ela perde em generalidade por exigir que a distribuição regional de precipitações tenha uma cauda superior assintoticamente exponencial. Como será mostrado no capítulo 3, trata-se de um requisito sem o qual não se é possível aplicar as premissas GRADEX na transferência da informação hidrometeorológica para a curva de frequência dos volumes de cheia;
- Por construção, o modelo regional de precipitações utiliza somente as estatísticas amostrais superiores na estimativa de parâmetros e quantis. Apesar de serem esses os valores considerados mais representativos da cauda superior e um dos atributos da metodologia, verifica-se aqui mais um fator que limita sua aplicação em nosso país. Isso deve-se às séries pluviométricas relativamente curtas e espacialmente pouco disseminadas de nossa rede hidrometeorológica;
- A extrapolação da densidade condicionada $f_{Y/X}(y/x)$ exige a simulação de tormentas extremas transpostas para a bacia de interesse e, por conseguinte, pressupõe a existência de um catálogo dessas tormentas. Com exceção de alguns poucos casos, para os quais existem estimativas de eventos hipotéticos como os de precipitação máxima provável (PMP), não se tem notícia da existência desse tipo de catálogo no Brasil.

As limitações apresentadas acima constituem uma barreira para a aplicação do método em bacias com pouca disponibilidade de dados, caso comum no Brasil. No capítulo 3, a metodologia proposta por Naghettini (1994) será descrita em detalhes, permitindo assim, determinar as principais características que impedem sua aplicação no Brasil. Identificadas as limitações do método, serão introduzidas modificações que possibilitem sua aplicação de forma mais ampla e fácil.

3 METODOLOGIA

3.1 Introdução

Conforme visto no item 2.6 a seqüência metodológica introduzida por Naghettini (1994) traz consigo um série de barreiras que impedem sua aplicação direta no Brasil. De forma mais crítica, a disponibilidade e a qualidade da informação hidrometeorológica no Brasil constituem a principal limitação para a aplicação do método. Por outro lado, a metodologia proposta pelo autor introduz a possibilidade de se utilizar o máximo de informação hidrometeorológica de uma bacia e, por esse motivo, torna-se uma importante ferramenta para a determinação de quantis extremos de vazão em local com escassez de dados.

Neste capítulo a metodologia proposta por Naghettini (1994) será retomada de forma a possibilitar sua aplicação no Brasil. O método aqui proposto, doravante denominado método PVP, preserva a estrutura geral da seqüência metodológica proposta por Naghettini (1994), principalmente no que diz respeito à transferência de informação hidrometeorológica por meio do método GRADEX. No entanto, são introduzidas duas modificações importantes: a primeira refere-se ao emprego do modelo regional TCEV para a estimação do parâmetro de escala das precipitações de duração igual ao tempo de base e a segunda refere-se à modelação, em escala regional, da relação entre vazões médias diárias e volumes de cheia, de duração igual ao tempo de base, ou seja, da relação pico-volume.

Naghettini (1994) adaptou o modelo regional de freqüência desenvolvido por Smith (1989) para estimar a cauda superior das precipitações de duração igual ao tempo de base. O modelo adaptado utiliza somente as estatísticas de ordem superior dos dados de chuva sendo que o parâmetro de escala da distribuição é determinado a partir de informações hidrogeológicas locais.

No método PVP, o modelo regional de precipitações acomoda o conjunto integral das amostras locais de alturas de precipitação de duração especificada. Ao invés de se restringir apenas às caudas exponenciais unipopulacionais, o modelo regional contempla também as situações em que os eventos chuvosos provenham de populações distintas. São conhecidos os efeitos do fenômeno *El Niño*, em algumas regiões, ou de atividades ciclônica, em outras, sobre a formação de eventos hidrometeorológicos extraordinários. Esses eventos certamente possuem características populacionais distintas daquelas dos eventos mais freqüentes. Neste

sentido e buscando uma maior generalidade, o modelo regional de chuva é estimado com base no modelo distributivo descrito por Rossi et al. (1984), denominado TCEV – “Two-Component Extreme Value”, o qual compõem-se de uma mistura de duas distribuições exponenciais.

A aplicação do método PVP, bem como a do método proposto por Naghettini (1994), envolve a extrapolação da função densidade de probabilidade das vazões de pico condicionadas aos volumes $f_{Y/X}(y/x)$ para valores extremos de y_p . Naghettini (1994) obteve essa extrapolação a partir da simulação de chuvas críticas, selecionadas em catálogos de eventos históricos extremos, no modelo chuva-vazão HSP-F. Assim, foi possível obter pares [vazão de pico, volume de cheia] que na seqüência foram modelados por um modelo de regressão não paramétrico. No entanto, esse método de obtenção de $f_{Y/X}(y/x)$, embora robusto, não pode ser aplicado a bacias brasileiras devido a inexistência de tais catálogos. Por outro lado, mesmo que se tenha catalogado tais eventos ou que se utilize tormentas hipotéticas, a simulação em modelos de transformação chuva-vazão é complexa e envolve, além da necessidade de mais dados, um levantamento criterioso das condições em que serão feitas as simulações. Os dados utilizados na calibração dos parâmetros do modelo chuva-vazão, o algoritmo de otimização do erro, as condições de umidade do solo e a própria escolha do modelo de transformação são algumas das questões a serem analisadas antes da utilização desse método para obter os pares [pico de vazão, volume de cheia].

No método PVP, é introduzida uma nova perspectiva para a determinação dos pares [pico de vazão, volume de cheia]. Os pares são determinados por uma relação regional adimensional obtida a partir de picos e volumes de hidrogramas observados em várias estações fluviométricas. Dessa maneira, o método PVP prescinde da utilização de modelos de transformação chuva-vazão e, por conseqüência, de catálogos de tormentas extremas. O método pode ser utilizado em locais desprovidos de dados fluviométricos e/ou pluviométricos uma vez que os picos de vazão podem ser determinados a partir de características físicas da bacia em análise para qualquer volume de cheia.

A essência do método proposto por Naghettini (1994), e mantida no método PVP, tem origem na modelagem dos processos estocásticos pontuais marcados. Esses encontram-se descritos no item que se segue.

3.2 O método pico-volume

A figura 3.1 mostra uma representação esquemática de um processo estocástico pontual marcado. O eixo das ordenadas representa a vazão escoada e o eixo das abscissas representa o tempo. O processo mostrado na figura 3.1 é caracterizado por: T_i , o tempo de ocorrência do i -ésimo pico que excede o valor limiar (*threshold*) u ; Y_i , o pico associado ao tempo T_i ; e X_i , o volume escoado, com duração d , associado ao pico Y_i .

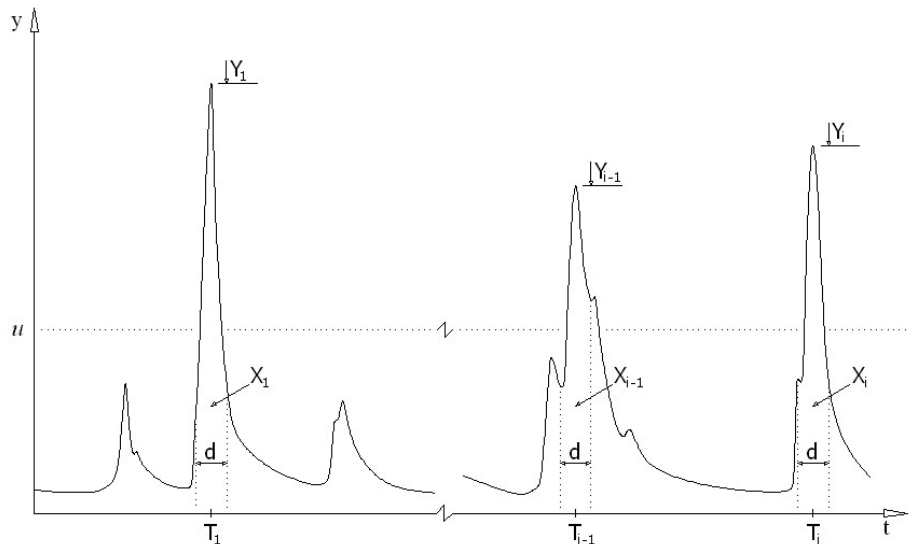


Figura 3.1 – Processo pontual marcado para eventos de vazão

Os processos pontuais marcados, tal como o processo $\{T_i, X_i, Y_i; i = 1, 2, \dots\}$, foram modelados, primeiramente, por Todorovic & Zelenhasic (1970) usando uma representação de um processo de Poisson composto. Para essa representação, duas condições são necessárias:

- O número de ocorrência de cheias $N_{\alpha\beta}$ em um determinado intervalo de tempo $[\alpha, \beta]$ é representado por um processo de Poisson não homogêneo com uma taxa de ocorrência dependente do tempo $\lambda(t)$, ou seja:

$$P(N_{\alpha\beta} = n) = \frac{\left[\int_{\alpha}^{\beta} \lambda(t) dt \right]^n \exp \left[- \int_{\alpha}^{\beta} \lambda(t) dt \right]}{n!} \quad (3.1)$$

- Y_i é uma seqüência de variáveis aleatórias mutuamente independentes, com distribuição de freqüência dependente de T_i .

Dividindo-se o intervalo $[\alpha, \beta]$ em k_0 subintervalos e assumindo-se que k_0 é suficientemente grande para que a distribuição de Y_i em cada subintervalo não dependa do tempo, tem-se:

$$P(M_j \leq y) = P(N_j = 0) + \sum_{n=1}^{\infty} P\left[\bigcap_{i=1}^n (Y_{ij} \leq y) \cap N_j = n\right] \quad (3.2)$$

sendo,

N_j o número de ocorrências (picos) no j -ésimo subintervalo;

M_j o maior pico no j -ésimo subintervalo; e

Y_{ij} o i -ésimo pico acima de u no j -ésimo subintervalo.

Pela condição de independência mútua de Y_{ij} , a equação 3.2 torna-se:

$$P(M_j \leq y) = P(N_j = 0) + \sum_{n=1}^{\infty} P(N_j = n) [H_{uj}(y)]^n \quad (3.3)$$

sendo H_{uj} a função acumulada de probabilidade dos picos de vazão que excedem o limiar u no subintervalo j .

Substituindo-se a equação 3.1 na equação 3.3, tem-se:

$$P(M_j \leq y) = \exp\left\{-[1 - H_{uj}(y)] \int_j \lambda(t) dt\right\} \quad (3.4)$$

De acordo com North (1980), a distribuição do máximo valor em $[\alpha, \beta]$, denotada por $M(\alpha, \beta)$, pode ser deduzida da seguinte forma:

$$P[M(\alpha, \beta) \leq y] = P\left(\bigcap_{j=1}^{k_0} M_j \leq y\right) \quad (3.5)$$

Sob a hipótese de independência mútua, tem-se:

$$P[M(\alpha, \beta) \leq y] = \prod_{j=1}^{k_0} P(M_j \leq y) \quad (3.6)$$

A partir da equação 3.4, pode-se escrever 3.6 da seguinte forma:

$$P[M(\alpha, \beta) \leq y] = \exp \left\{ - \sum_{j=1}^{k_0} [1 - H_{u_j}(y)] \int_j \lambda(t) dt \right\} \quad (3.7)$$

Quando $k_0 \rightarrow \infty$ a equação 3.7 torna-se:

$$P[M(\alpha, \beta) \leq y_p] = \exp \left\{ - \int_{\alpha}^{\beta} [1 - H_u(y_p | t)] \lambda(t) dt \right\} \quad (3.8)$$

A partir da equação 3.8 é possível calcular a probabilidade de ocorrência de picos em intervalos anuais. Se $\alpha = 0$ representa o início do ano e $\beta = 1$ representa o final, a distribuição dos máximos anuais dos picos de vazão é dada por:

$$P[M(0,1) \leq y_p] = F_M(y_p) = \exp \left\{ - \int_0^1 [1 - H_u(y_p | t)] \lambda(t) dt \right\} \quad (3.9)$$

onde y_p representa o quantil de interesse.

Segundo Bradley & Potter (1992), a distribuição dos picos condicionada a u , denotada por $H_u(y|t)$, pode ser deduzida da distribuição dos volumes escoados e da distribuição dos picos condicionada aos volumes.

A distribuição dos volumes de cheia X_i , condicionada ao nível u e ao tempo de ocorrência t , é dada por:

$$G_u(x | t) = P(X_i \leq x | Y_i > u, T_i = t) \quad (3.10)$$

As variáveis Y e X são relacionadas pela distribuição dos picos condicionada aos volumes e ao tempo de ocorrência. Formalmente:

$$F_{Y|X}(y | x, t) = P(Y_i \leq y | X_i = x, T_i = t) \quad (3.11)$$

Se $g_u(x|t)$ e $f_{y|x}(y|x,t)$ representam as correspondentes funções densidade de probabilidade, a função distribuição de probabilidade conjunta de x e y é dada pelo produto $f_{y|x}(y|x,t) \cdot g_u(x|t)$. Assim, por definição, a distribuição marginal $H_u(y|t)$ é dada por:

$$H_u(y_p|t) = \int_0^{y_p} \int_0^{\infty} f_{y|x}(y|x,t) \cdot g_u(x|t) dx dy \quad (3.12)$$

Se $g_u(x|t)$ e $f_{y|x}(y|x,t)$ não dependem do tempo, as equações 3.9 e 3.12 se reduzem, respectivamente, a:

$$H_u(y_p) = \int_0^{y_p} \int_0^{\infty} f_{y|x}(y|x) \cdot g_u(x) dx dy \quad (3.13)$$

e

$$F_M(y_p) = \exp \left\{ - \left[1 - H_u(y_p) \right] \cdot \int_0^1 \lambda(t) dt \right\} \quad (3.14)$$

No método PVP a distribuição dos volumes $g_u(x)$ é obtida por meio do método GRADEX. Inicialmente, estima-se a distribuição da chuva pela regionalização dos dados pluviométricos utilizando-se a TCEV. Na seqüência, estima-se a distribuição dos volumes pela translação da distribuição da chuva por uma distância proporcional ao parâmetro de escala (gradex) da cauda superior da TCEV. A distribuição dos picos condicionada aos volumes é estimada por meio de um modelo regional paramétrico não linear que correlaciona as vazões de pico com os volumes escoados.

Uma vez determinadas as funções densidade (f e g) e a intensidade $\Lambda = \int_0^1 \lambda(t) dt$, a solução do sistema formado pelas equações 3.13 e 3.14 para valores extremos de y_p fornece a distribuição anual de probabilidades para os picos de vazão. A seguir é mostrado um fluxograma que resume a aplicação do método PVP.

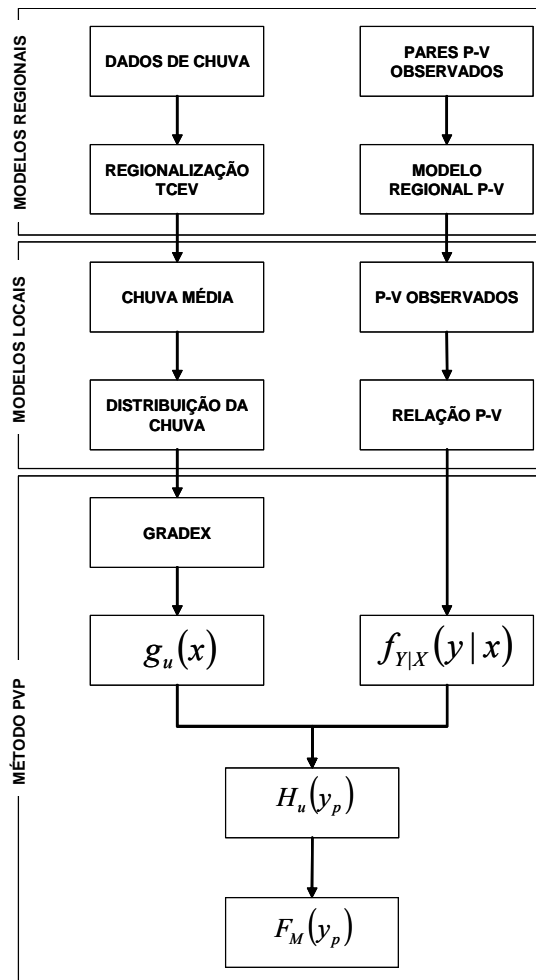


Figura 3.2 – Fluxograma de aplicação do método PVP

Na seqüência, serão descritos, isoladamente, os componentes da seqüência metodológica proposta, sem a preocupação de unificá-los em um conjunto lógico. Ao final do capítulo, no item 3.6, apresenta-se o contexto lógico, dentro do qual são incorporadas as características essenciais de cada componente, de modo a estabelecer uma metodologia unificada para a determinação de quantis extremos que leve em consideração os princípios estabelecidos pelo U.S. National Research Council (NRC, 1988).

3.3 O método GRADEX

O método GRADEX – gradiente de valores extremos – foi desenvolvido pela companhia elétrica francesa EDF (*Electricité de France*), tendo sido descrito primeiramente por Guillot e Duband (1967). Esse método tem sido extensivamente usado na França e em outros países para o desenvolvimento de projetos de hidrelétricas e de projetos de dimensionamento de vertedouros (DUBAND et al., 1988). O método GRADEX tem por objetivo a extrapolação da curva de freqüência dos volumes de cheia usando séries de chuvas, baseando-se em duas

suposições. A primeira refere-se à relação entre a chuva e o volume escoado (*runoff*) quando a umidade do solo aproxima-se da condição de saturação. Supõe-se que, em condições de saturação, qualquer incremento de volume da chuva, para uma duração suficientemente longa, tende a produzir um incremento equivalente no volume de *runoff* de igual duração. A segunda refere-se à cauda superior da distribuição dos volumes de chuva, a qual é suposta ser uma função de decaimento exponencial, ou seja:

$$1 - F(p) = \exp\left(-\frac{p - K}{a}\right) \quad (3.15)$$

na qual, a constante positiva K e a são os parâmetros de posição e escala, respectivamente.

Em geral, esses parâmetros são estimados pelo ajuste de uma distribuição que tenha um comportamento assintótico exponencial na cauda superior (e.g. Gumbel, Gamma, Log-Normal, TCEV) aos dados de chuva. Combinando essas duas suposições, verifica-se que a cauda superior da distribuição dos volumes de *runoff* tem comportamento assintótico exponencial com o mesmo parâmetro de escala a (o parâmetro gradex) estimado para a cauda superior da distribuição dos volumes de chuva.

3.3.1 Descrição do método GRADEX

Suponha que P_i denote a máxima altura de chuva (para uma duração d) sobre uma bacia em um dado mês, estação ou ano. A duração d é usualmente especificada como o tempo de base médio da bacia, obtido de hidrogramas observados. Suponha, também, que X_i represente o volume de *runoff* (para a mesma duração d) associado à P_i , e seja R_i o déficit de *runoff*, definido como $R_i = P_i - X_i$. A figura 3.3 mostra o gráfico esquemático da relação entre X e P .

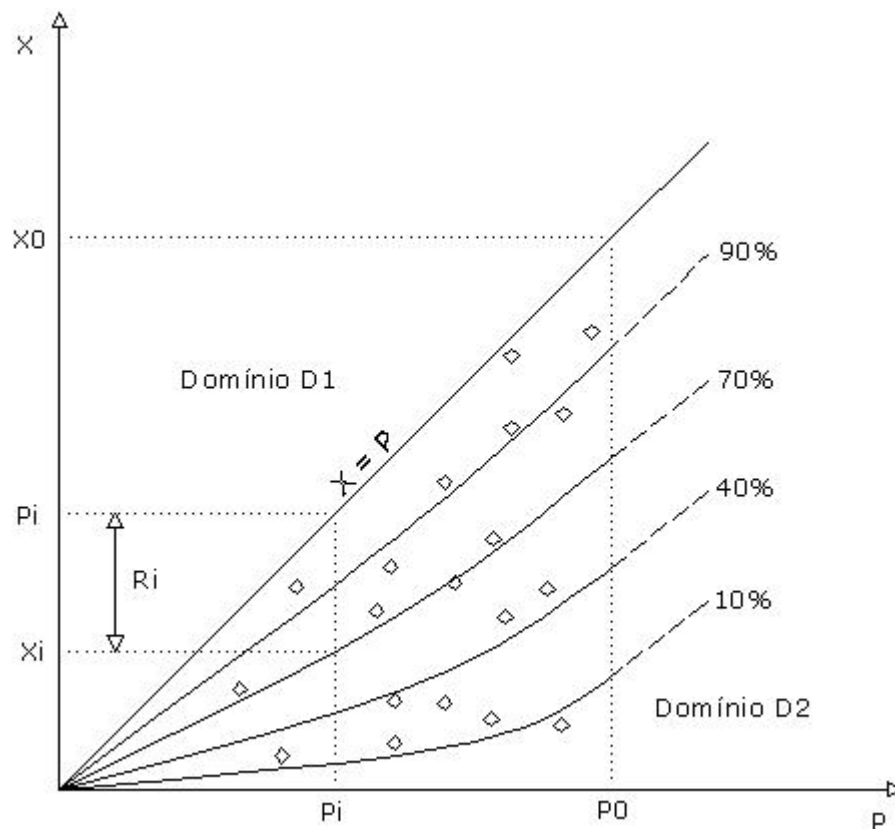


Figura 3.3 – Relação entre a chuva (P) e o volume escoado (X)

Na figura 3.3, os pontos (P_i, X_i) estão todos abaixo da linha $X = P$. O valor de R depende de muitos fatores complexos e interdependentes tais como as condições de umidade anterior do solo, o armazenamento de água subterrânea e a distribuição espaço-temporal da chuva sobre a bacia. No método GRADEX, R é tratada como uma variável aleatória com distribuição de probabilidades, condicionada à P , caracterizado na figura 3.3 por uma curva de quantis hipotética. A primeira suposição do método GRADEX impõe que essa curva de quantis tenda assintoticamente para a linha $X = P$, à medida que a bacia se aproxima da saturação. A posição em que cada curva começa a tender assintoticamente depende das condições iniciais de umidade do solo, sendo que a curva aproxima-se da reta $X = P$ mais rapidamente em terrenos úmidos do que em terrenos secos. Em outras palavras, a distribuição acumulada de R , condicionada a P tende a ter parâmetro de forma estável e variância constante para $P > P_0$.

O plano definido pelos pontos (P_i, X_i) na figura 3.3 pode ser dividido em dois domínios:

- Domínio D1: definido por todos os pontos $P < P_0$ e $X < X_0$, para os quais a distribuição de probabilidades pode ser estimada através dos dados observados; e

- Domínio D2: o domínio das extrapolações.

Seja $f(p)$, $g(x)$ e $h(r)$ as funções densidade de probabilidade de P , X e R , respectivamente, e seja $j(p, r)$ a função densidade de probabilidade conjunta de P e R . Dado que $R_i = P_i - X_i$, a função densidade marginal de $g(x)$ é dada por:

$$g(x) = \int_0^{\infty} j(x+r, r) dr \quad (3.16)$$

Se $h(r | X+R)$ é a função densidade de probabilidade de R , condicionada à variável $(X+R)$, então a equação 3.16 pode ser reescrita como:

$$g(x) = \int_0^{\infty} f(x+r) h_{X+R}(r) dr \quad (3.17)$$

No domínio D2, supõe-se que a distribuição de R não depende de P . Em outros termos, a função $h_{X+R}(r)$ torna-se $h(r)$. Assim,

$$g(x) = \int_0^{\infty} f(x+r) h(r) dr \quad (3.18)$$

A segunda suposição do método GRADEX diz respeito ao comportamento da cauda superior da função de distribuição acumulada $F(p)$, a qual presume-se tender assintoticamente para uma cauda do tipo exponencial, segundo a seguinte equação:

$$1 - F(x+r) = \exp\left(-\frac{x+r-K}{a}\right) \quad (3.19)$$

na qual o parâmetro de posição K é uma constante positiva e o parâmetro de escala a refere-se ao parâmetro gradex da chuva. Nesse caso, a função densidade $f(p)$ torna-se:

$$f(x+r) = \frac{1}{a} \exp\left(-\frac{x+r-K}{a}\right) \quad (3.20)$$

ou

$$f(x+r) = f(x) \exp\left(-\frac{r}{a}\right) \quad (3.21)$$

Substituindo-se a equação 3.21 na equação 3.18, segue-se que:

$$g(x) = f(x) \int_0^{\infty} \exp\left(-\frac{r}{a}\right) h(r) dr \quad (3.22)$$

Nessa expressão, a integral é definida e igual a uma constante positiva menor ou igual a 1. Assumindo-se que essa constante seja igual a $\exp(-r_0/a)$, a equação 3.22 torna-se:

$$g(x) = f(x+r_0) \quad (3.23)$$

Portanto, a função densidade de probabilidade g no domínio D2 pode ser deduzida de f por uma simples translação por uma quantidade r_0 , o que é válido, também, para as funções de distribuição acumuladas G e F . As figuras 3.4 e 3.5 mostram esquematicamente esses resultados em papel de plotagem clássico (sem transformação) e em papel de plotagem exponencial, respectivamente.

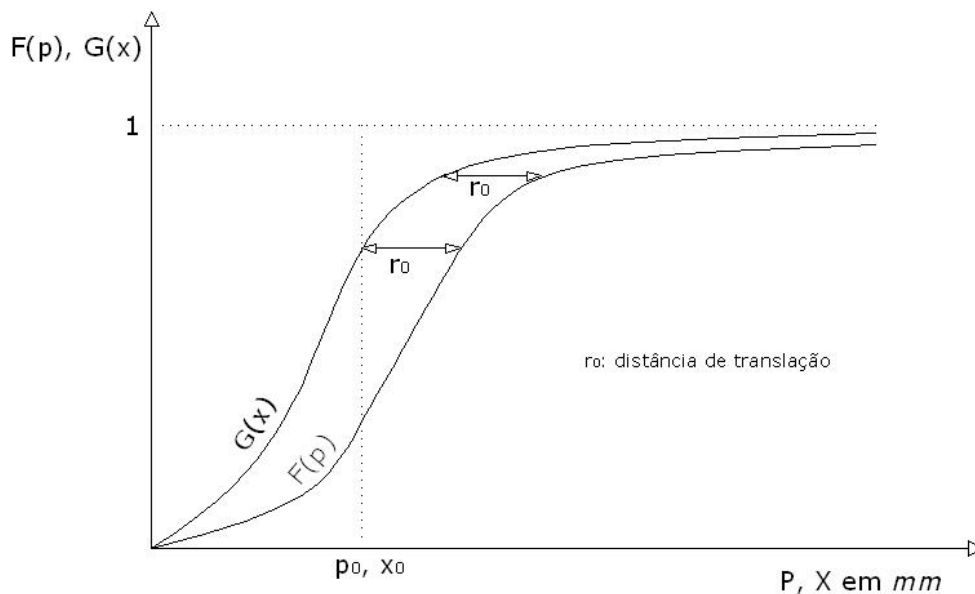


Figura 3.4 – Funções de distribuição de probabilidades acumuladas dos volumes escoados $G(x)$ e dos volumes de chuva $F(p)$, ambos de duração d . Papel de plotagem clássico. A distância r_0 refere-se à distância de translação no domínio das extrapolações D2.

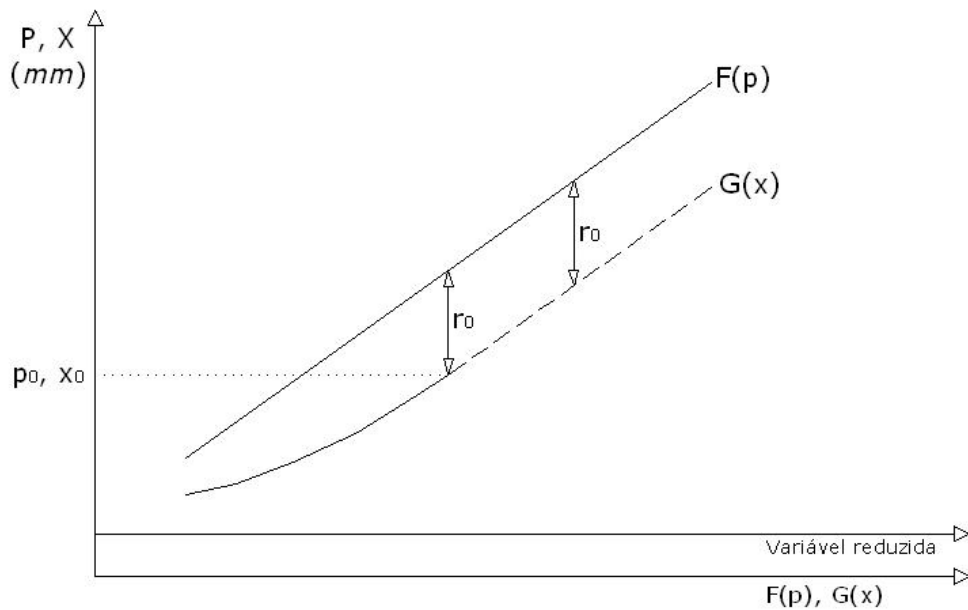


Figura 3.5 – Plotagem de $G(x)$ e $F(p)$ em papel de plotagem exponencial ou de Gumbel. A distância r_0 é a distância de translação no domínio das extrapolações D2. As durações d da chuva e do volume são as mesmas.

Para a distância de translação r_0 a integral na equação 3.22 representa o valor esperado de $\exp(-r/a)$. Como resultado tem-se:

$$r_0 = -a \ln \left\{ E \left[\exp \left(-\frac{r}{a} \right) \right] \right\} \quad (3.24)$$

A distância de translação r_0 é referenciada como um “limite prático” para R (DUBAND et al., 1988). Na prática, Guillot & Duband (1967) recomendam o uso de uma distribuição empírica para os dados observados dos volumes máximos anuais (ou sazonais) até um período de retorno de 10 a 20 anos para bacias relativamente impermeáveis e até 50 anos para bacias com alta capacidade de infiltração. A partir desse ponto, as funções de distribuição acumuladas dos volumes de chuva e vazão são separadas por uma distância r_0 . Equivalentemente, no domínio D2, as duas distribuições podem ser plotadas em papel exponencial como uma linha reta, ambas com uma inclinação igual ao parâmetro gradex da chuva a . Portanto, para a aplicação do método GRADEX em uma bacia é necessário somente a estimativa do parâmetro gradex da chuva a .

A validade da equação 3.23 está condicionada à suposição de que a cauda superior da distribuição dos volumes de chuva é do tipo exponencial, e não somente a suposição de que R

e P são independentes para $P > P_0$. A justificativa para essa afirmação é dada em C.T.G.R.E.F. (1972). Inicialmente, as equações 3.18 e 3.23 são igualadas, resultando em:

$$f(x + r_0) = \int_0^{\infty} f(x + r)h(r)dr \quad (3.25)$$

Denotando $(x + r_0)$ por ν e substituindo-o na equação 3.25, segue-se que:

$$f(\nu) = \int_0^{\infty} f(\nu + r - r_0)h(r)dr \quad (3.26)$$

ou

$$\int_0^{\infty} \frac{f(\nu + r - r_0)}{f(\nu)}h(r)dr = 1 \quad (3.27)$$

Diferenciando 3.27 em relação a ν , obtém-se:

$$\frac{d}{d\nu} \int_0^{\infty} \frac{f(\nu + r - r_0)}{f(\nu)}h(r)dr = 0 \quad (3.28)$$

ou

$$\int_0^{\infty} \frac{d}{d\nu} \left[\frac{f(\nu + r - r_0)}{f(\nu)} \right] h(r)dr = 0 \quad (3.29)$$

Para que a equação 3.29 seja satisfeita, a taxa $f(\nu + r - r_0)/f(\nu)$ deve ser constante em relação a ν . As únicas funções que possuem tal propriedade são as funções exponenciais da forma $A \exp(B\nu)$, sendo A e B constantes. Como resultado, f e g devem ser funções exponenciais dessa forma. Entre as distribuições de probabilidade que exibem uma cauda superior do tipo exponencial estão a Normal, a Log-Normal, a Gamma, a Gumbel e a TCEV. Por exemplo, se m denota a variável reduzida $(p - K)/a$, então a distribuição de probabilidades de Gumbel é dada por:

$$F'(p) = \exp[-\exp(-m)] \quad (3.30)$$

Expandindo $F'(p)$ em uma série de Taylor segue-se que:

$$F'(p) = 1 - \exp(-m) + \frac{1}{2!} [\exp(-m)]^2 - \dots + \frac{(-1)^n}{n!} [\exp(-m)]^n \quad (3.31)$$

Como p tende para o infinito, a equação 3.31 tende para:

$$F'(p) \rightarrow 1 - \exp(-m) \quad (3.32)$$

Logo, a distribuição exponencial e a de Gumbel tendem assintoticamente para uma cauda exponencial à medida que p tende para o infinito. Desenvolvimento similar à equação 3.31, para outras distribuições de probabilidade, pode ser encontrado em Gumbel (1958).

3.3.2 Estimativa do parâmetro gradex

Em aplicações convencionais do método GRADEX, o parâmetro a pode ser estimado pelo ajuste de uma distribuição com cauda exponencial aos dados de chuva (máximos anuais ou sazonais). O procedimento mais simples de estimação do parâmetro gradex é ajustar a distribuição de Gumbel à série de máximos anuais de chuva, com duração d igual ao tempo de base da bacia. No entanto, o procedimento mais usado consiste em se estimar o parâmetro gradex através do ajuste de uma distribuição de cauda exponencial à série de máximos sazonais (mensais) de chuva e então combinar cada distribuição para obter a distribuição anual. Esse procedimento tem a vantagem de levar em consideração as condições sazonais de produção de escoamento, reduzir o erro amostral pelo aumento da amostra e atenuar a influência de *outliers* na estimativa do gradex. A seguir será descrito esse procedimento com aplicação mensal.

Sejam F_1, F_2, \dots, F_{12} as distribuições acumuladas dos volumes máximos mensais de chuva para uma duração d ocorridos nos meses 1, 2, ..., 11 e 12 respectivamente. Sejam P_1, P_2, \dots, P_{11} e P_{12} as variáveis que representam esses máximos. Assim:

$$P(P \leq p) = P\left(\bigcap_{i=1}^{12} P_i \leq p\right) \quad (3.33)$$

Os máximos mensais de duração d podem ser considerados, na maioria dos casos, como variáveis aleatórias mutuamente independentes. Como resultado, tem-se:

$$P(P \leq p) = \prod_{i=1}^{12} P(P_i \leq p) \quad (3.34)$$

ou

$$F(p) = P(P \leq p) = \prod_{i=1}^{12} F_i(p) \quad (3.35)$$

Supondo que a distribuição de probabilidades de cada mês possui uma cauda superior exponencial, com parâmetros estimados a partir da séries de máximos mensais, tem-se:

$$F_i(p) = 1 - \exp\left(-\frac{p - K_i}{a_i}\right), \quad i = 1, 2, \dots, 12 \quad (3.36)$$

Substituindo-se 3.36 em 3.35, tem-se

$$F(p) = \prod_{i=1}^{12} \left[1 - \exp\left(-\frac{p - K_i}{a_i}\right) \right] \quad (3.37)$$

Rigorosamente, a cauda superior da distribuição anual $F(p)$, calculada através da equação 3.37, não pode ser considerada exponencial. No entanto, o produto consecutivo da expressão $\{1 - \exp[-(p - K_i)/a_i]\}$ gera uma distribuição anual que é dominada pela distribuição mensal com o maior gradex. Para valores crescentes de p , o parâmetro gradex anual a tende para o maior parâmetro gradex mensal (NAGHETTINI, 1994).

3.3.3 Extrapolação da distribuição dos volumes de cheia

O parâmetro de escala da cauda superior da distribuição dos volumes de cheia é supostamente igual ao parâmetro gradex da chuva. No entanto, a estimativa do gradex do volume de cheia deve levar em consideração as unidades em que são expressas as chuvas e as vazões. Dessa maneira, se AD é a área de drenagem da bacia em km^2 , d é a duração em horas e os parâmetros gradex são expressos em unidades do SI, tem-se:

$$\begin{array}{l} \text{gradex das vazões} = \frac{AD}{3,6d} \cdot \text{gradex da chuva} \\ \text{(m}^3\text{/s)} \qquad \qquad \text{(km}^2\text{/hora)} \quad \text{(mm)} \end{array}$$

Se AD e d são expressos em km^2 e dias, respectivamente, tem-se:

$$\begin{array}{ccc} \text{gradex das vazões} & = & \frac{AD}{86,4d} \cdot \text{gradex da chuva} \\ (m^3/s) & & (km^2/dia) \quad (mm) \end{array}$$

Usualmente, a extrapolação da distribuição dos volumes de cheia, de acordo com o parâmetro gradex, começa a partir da cheia de 10 anos de período de retorno para bacia pequenas e relativamente impermeáveis, ou a partir da cheia de 20 anos para bacias grandes ou, ainda, a partir da cheia de 50 anos para bacias com relevo suave e/ou grande capacidade de infiltração. No entanto, são raros os casos em que se tem uma definição exata do ponto de saturação da bacia. Conseqüentemente, a definição do quantil necessário para saturar a bacia é feita de maneira bastante arbitrária. Nesses casos, recomenda-se que a definição do ponto de extrapolação da distribuição dos volumes escoados seja feita a partir dos dados observados, dando-se maior ênfase às estatísticas de ordem superior.

3.4 Regionalização hidrológica pela TCEV

3.4.1 Definição

A função distribuição acumulada de probabilidade (FAP) do modelo TCEV (*Two Component Extreme Value*) é obtida pelo produto de duas distribuições Gumbel através da seguinte expressão:

$$F_X(x) = P(X \leq x) = \exp(-\Lambda_1 e^{-x/\theta_1} - \Lambda_2 e^{-x/\theta_2}) \quad (3.38)$$

com,

$$x > 0$$

$$\theta_2 > \theta_1 > 0$$

$$\Lambda_1 > \Lambda_2 \geq 0$$

A distribuição é definida para valores de x não negativos, apresentando uma descontinuidade em $x = 0$.

A função densidade de probabilidade (FDP) para o modelo TCEV é dada pela derivada primeira da equação 3.38 em relação a x . Ou seja:

$$f_X(x) = \frac{\partial}{\partial x} F_X(x) = \left(\frac{\Lambda_1}{\theta_1} e^{-x/\theta_1} + \frac{\Lambda_2}{\theta_2} e^{-x/\theta_2} \right) \exp(-\Lambda_1 e^{-x/\theta_1} - \Lambda_2 e^{-x/\theta_2}) \quad (3.39)$$

A regionalização de dados meteorológicos, com base no modelo TCEV, tem sido utilizada largamente na Itália através da aplicação da metodologia desenvolvida por Fiorentino *et al.* (1987) e Gabriele & Arnell (1991). Nessa metodologia, a regionalização é feita em três níveis hierárquicos de modo a otimizar a estimativa dos parâmetros da TCEV.

Fiorentino *et al.* (1987) mostraram que, ao invés de se usar os dois níveis comuns à análise regional convencional, o primeiro objetivando identificar a variação de escala da variável hidrológica com o período de retorno e o segundo objetivando identificar a variação do parâmetro de posição, a análise regional pode ser separada, hierarquicamente, em três níveis distintos, um para cada parâmetro estatístico. No primeiro, identifica-se uma região homogênea com respeito ao coeficiente de assimetria. Essa região é subdividida, então, em sub-regiões homogêneas com respeito ao coeficiente de variação, assim constituindo o segundo nível hierárquico. No terceiro nível, aplica-se o método *index flood* e são determinadas as estimativas dos quantis para os vários períodos de retorno.

3.4.2 Primeiro nível de regionalização

No primeiro nível de regionalização, identifica-se uma região homogênea segundo o coeficiente de assimetria. Formalmente, considerando a existência de k estações dentro da região homogênea, define-se a variável aleatória Y' como:

$$Y' = Y_{ij} = \frac{X_{ij}}{\theta_{li}} - \ln \Lambda_{li} \quad i = 1, \dots, k \quad e \quad j = 1, \dots, n_i \quad (3.40)$$

sendo n_i o tamanho da amostra da i -ésima estação; e θ_{li} , Λ_{li} os parâmetros da componente básica da i -ésima estação, sob a hipótese de Y ser homogênea.

Y' é uma variável aleatória independente e identicamente distribuída (i.i.d.) e se distribui segundo a TCEV da seguinte forma:

$$F_{Y'}(y') = P(Y' \leq y') = \exp(-e^{-y'} - \Lambda_* e^{-y'/\theta_*}) \quad (3.41)$$

sendo,

$$\theta_* = \frac{\theta_{2i}}{\theta_{1i}}; e \quad (3.42)$$

$$\Lambda_* = \frac{\Lambda_{2i}}{\Lambda_{1i}^{1/\theta_*}} \quad (3.43)$$

Se a variável Y' se distribui de acordo com a TCEV, Gabriele & Iiritano (1994) mostram serem válidas as seguintes propriedades:

- O coeficiente de assimetria teórico (γ_1) é único em toda a região. Esse coeficiente é obtido pela seguinte equação:

$$\gamma_1 = \frac{3,7553 - 3S_2 - 3,4632S_1 + 6S_1S_0 + 3,9350S_0 + 3,4632S_0^2 - S_0^3}{(1,6448 + 2S_1 + 1,1544S_0 - S_0^2)^{3/2}} \quad (3.44)$$

na qual,

$$S_i = \sum_{j=1}^{\infty} \frac{(-1)^j \Lambda_*^j \Gamma^{(i)}(j/\theta_*)}{j!}$$

e $\Gamma^{(i)}$ é a i -ésima derivada da função gama; e

- As relações entre os parâmetros θ_{1i} e θ_{2i} são constantes em toda a região.

Pela hipótese de homogeneidade do primeiro nível de regionalização, θ_* e Λ_* são constantes, o que equivale a admitir que o coeficiente de assimetria γ_1 é o mesmo para a região estudada.

Uma vez que o coeficiente de assimetria, dado pela equação 3.44, depende dos parâmetros da TCEV e esses não são conhecidos *a priori*, pode-se definir a região homogênea com base no coeficiente de assimetria amostral g , ou seja:

$$\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}}$$

e

$$\hat{\gamma}_1 = g = \frac{n^2}{(n-1)(n-2)} \cdot \frac{m_3}{s^3}$$

sendo,

μ_r = momento central de ordem r ; e

m_r = momento central amostral de ordem r dado por $\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^r$, $m_1 = 0$

3.4.3 Segundo nível de regionalização

Nesse nível, calcula-se Λ_{1s} ou Λ_1 regional em cada sub-região homogênea, sob o aspecto do coeficiente de variação. Segundo Gabriele & Iiritano (1994), a estimativa de Λ_{1s} pode ser realizada a partir dos parâmetros Λ_{1i} , calculados para cada estação i no nível 1. Considerando o valor médio de Λ_{1s} e os valores transformados de Λ_{1i} , têm-se:

$$CV_{1i} = \frac{0,557}{\log \Lambda_{1i} + 0,251} \quad (3.45)$$

$$C\bar{V}_{1s} = \frac{\sum_{j=1}^{n_i} CV_{1j}}{n_i} \quad (3.46)$$

$$\hat{\Lambda}_{1s} = 10^{\frac{0,557}{C\bar{V}_{1s}} - 0,251} \quad (3.47)$$

A estimativa de Λ_{1s} pela equação 3.47 baseia-se num procedimento misto de aplicação do método da máxima verossimilhança e do método dos momentos. Alternativamente, Gabriele & Iiritano (1994) propuseram um método de máxima verossimilhança modificado, baseado na curva adimensional de quantis (*growth curve*). Além desses métodos, pode-se utilizar as

propriedades dos momentos-L, descritos por Hosking & Wallis (1997), para a determinação dos parâmetros em cada nível. Os métodos de estimativa dos parâmetros regionais da TCEV serão vistos no item 3.4.5.

3.4.4 Terceiro nível de regionalização

O terceiro nível de regionalização refere-se à aplicação do método do *index-flood* com o objetivo de se determinar o quantil para vários períodos de retorno. Nessa etapa, define-se um modelo de regressão que relaciona as características físicas e climáticas de cada região homogênea a um fator de escala local.

O objetivo da aplicação da regionalização pela TCEV, dentro do contexto da metodologia proposta no presente trabalho, é definir o *gradex*, ou seja, o parâmetro de escala regional da chuva. Portanto, aqui não é necessária a estimativa dos quantis de chuva para diferentes períodos de retorno. Busca-se a identificação do parâmetro de escala regional da cauda superior da TCEV, ou seja, $\theta_2^{regional}$. Assim, na aplicação da metodologia, a ser apresentada no capítulo 4, não foi realizada essa etapa do método de regionalização originalmente proposto por Gabriele e Iiritano (1994).

3.4.5 Estimativa dos parâmetros regionais da TCEV

3.4.5.1 Primeiro nível de regionalização – Estimativa de θ_* e Λ_*

A estimativa dos parâmetros θ_* e Λ_* pelo método da máxima verossimilhança consiste em procurar os valores de θ_* e Λ_* que maximizam a seguinte função:

$$L_{\max} = \sum_{i=1}^k \sum_{j=1}^{n_i} \ln f(y_{ij}, \theta_*, \Lambda_*) \quad (3.48)$$

na qual o valor de y_{ij} é dado pela equação 3.40, k é o número de estações da região homogênea e n_i é o número de dados da i -ésima estação.

Os valores de θ_{1i} e Λ_{1i} , usados no cálculo de y_{ij} , são aqueles que maximizam a função de verossimilhança de cada estação, ou seja:

$$L_{\max}^{(i)} = f(x_{ij}, \Lambda_{li}, \theta_{li} | \theta_*, \Lambda_*) \quad (3.49)$$

A partir das equações 3.48 e 3.49, têm-se:

$$LR = \sum_{i=1}^k \sum_{j=1}^{n_i} \ln \left[-e^{-y_{ij}} - \Lambda_* e^{-y_{ij}/\theta_*} \right] \exp \left[-e^{-y_{ij}} - \Lambda_* e^{-y_{ij}/\theta_*} \right] \quad (3.50)$$

$$LS^{(i)} = \sum_{j=1}^{n_i} \ln \left\{ \left[\frac{\Lambda_{li}}{\theta_{li}} \exp \left(-\frac{x_{ij}}{\theta_{li}} \right) - \frac{\Lambda_* \Lambda_{li}^{1/\theta_*}}{\theta_* \theta_{li}} \exp \left(-\frac{x_{ij}}{\theta_* \theta_{li}} \right) \right] \times \right. \\ \left. \times \exp \left[-\Lambda_{li} \exp \left(-\frac{x_{ij}}{\theta_{li}} \right) - \Lambda_* \Lambda_{li}^{1/\theta_*} \exp \left(-\frac{x_{ij}}{\theta_* \theta_{li}} \right) \right] \right\} \quad (3.51)$$

Os parâmetros que maximizam as equações 3.50 e 3.51 são obtidos igualando-se as zero as derivadas de LR em relação a θ_* e Λ_* , e de $LS^{(i)}$ em relação a θ_{li} e Λ_{li} . Dessa forma, obtêm-se um sistema com $2 + 2k$ equações o qual pode ser resolvido por vários esquemas.

Fiorentino & Gabriele (1985) propuseram um método para estimar θ_* , Λ_* , θ_{li} e Λ_{li} , com base em um esquema que resolve, em separado, as equações 3.50 e 3.51. O método proposto é constituído das seguintes etapas:

- a) Para cada série são calculados os valores iniciais de θ_{li} e Λ_{li} , com base nos parâmetros α_i e ε_i da distribuição de Gumbel, estimados pelo método da máxima verossimilhança. Formalmente,

$$\theta_{li} = \frac{1}{\alpha_i} \quad (3.52)$$

$$\Lambda_{li} = e^{\alpha_i \varepsilon_i}$$

- b) A partir da equação 3.40, calcula-se o valor de Y' ;

- c) Com o valor de Y' , são determinados os valores de θ_* e Λ_* que maximizam a equação 3.50;

- d) Com os valores de θ_* e Λ_* estimados em (c), são calculados, para cada uma das k séries que compõem a região, os valores de θ_{li} e Λ_{li} que maximizam a equação 3.51; e
- e) Repete-se o algoritmo a partir de (b) até que a estimativa atual dos parâmetros não seja diferente da estimativa anterior, dada uma tolerância pré definida.

O método descrito permite estimar os $2 + 2k$ parâmetros, sob a hipótese de se ter Λ_{li} diferente em cada local. Fiorentino & Gabriele (1985) propuseram um método misto que combina a solução iterativa com o método de Newton-Raphson.

O método proposto por Fiorentino & Gabriele (1985) utiliza as propriedades e qualidades dos métodos iterativo e de Newton-Raphson. A estimativa dos parâmetros θ_* e Λ_* é feita pelo método iterativo. Nesse caso, iguala-se a zero a derivada parcial de 3.50 em relação a θ_* e Λ_* , ou seja:

$$\theta_* = \frac{\partial}{\partial \theta_*} LR(\theta_*, \Lambda_*, y') \quad (3.53)$$

$$\Lambda_* = \frac{\partial}{\partial \Lambda_*} LR(\theta_*, \Lambda_*, y') \quad (3.54)$$

A partir das equações 3.53 e 3.54, obtêm-se:

$$\theta_* = \left[\sum_{i=1}^k \sum_{j=1}^{n_i} \frac{y_{ij} e^{-y_{ij}/\theta_*}}{e^{-y_{ij}} + \frac{\Lambda_*}{\theta_*} e^{-y_{ij}/\theta_*}} \right] \left[\sum_{i=1}^k \sum_{j=1}^{n_i} y_{ij} e^{-y_{ij}/\theta_*} + \sum_{i=1}^k \sum_{j=1}^{n_i} \frac{y_{ij} e^{-y_{ij}/\theta_*}}{e^{-y_{ij}} + \frac{\Lambda_*}{\theta_*} e^{-y_{ij}/\theta_*}} \right]^{-1} \quad (3.55)$$

$$\Lambda_* = \Lambda_* \left[\sum_{i=1}^k \sum_{j=1}^{n_i} \frac{e^{-y_{ij}/\theta_*}}{e^{-y_{ij}} + \frac{\Lambda_*}{\theta_*} e^{-y_{ij}/\theta_*}} \right] \left[\theta_* \sum_{i=1}^k \sum_{j=1}^{n_i} e^{-y_{ij}/\theta_*} e^{-y_{ij}/\theta_*} \right]^{-1} \quad (3.56)$$

Analogamente, as $2k$ equações para o cálculo de Λ_{li} e θ_{li} são:

$$\theta_{1i} = \left[\sum_{j=1}^{n_i} \frac{x_{ij} e^{-x_{ij}/\theta_{1i}}}{\Psi(x_{ij})} + \frac{\Lambda_{2i}}{\theta_*^2 \Lambda_{1i}} \sum_{j=1}^{n_i} \frac{x_{ij} e^{-x_{ij}/\theta_{2i}}}{\Psi(x_{ij})} \right] \times \left[\sum_{j=1}^{n_i} \frac{e^{-x_{ij}/\theta_{1i}}}{\Psi(x_{ij})} + \frac{\Lambda_{2i}}{\theta_* \Lambda_{1i}} \sum_{j=1}^{n_i} \frac{e^{-x_{ij}/\theta_{2i}}}{\Psi(x_{ij})} + \sum_{j=1}^{n_i} x_{ij} e^{-x_{ij}/\theta_{1i}} \frac{\Lambda_{2i}}{\theta_* \Lambda_{1i}} \sum_{j=1}^{n_i} x_{ij} e^{-x_{ij}/\theta_{1i}} \right]^{-1} \quad (3.57)$$

$$\Lambda_{1i} = \frac{\Lambda_{1i}}{\theta_{1i}} \left[\sum_{j=1}^{n_i} \frac{x_{ij} e^{-x_{ij}/\theta_{1i}}}{\Psi(x_{ij})} + \frac{\Lambda_{2i}}{\theta_*^2 \Lambda_{1i}} \sum_{j=1}^{n_i} \frac{x_{ij} e^{-x_{ij}/\theta_{2i}}}{\Psi(x_{ij})} \right] \left[\sum_{j=1}^{n_i} e^{-x_{ij}/\theta_{1i}} + \frac{\Lambda_{2i}}{\theta_* \Lambda_{1i}} \sum_{j=1}^{n_i} e^{-x_{ij}/\theta_{2i}} \right]^{-1} \quad (3.58)$$

com

$$\Psi(x_{ij}) = \frac{\Lambda_{1i}}{\theta_{1i}} e^{-x_{ij}/\theta_{1i}} + \frac{\Lambda_{2i}}{\theta_{2i}} e^{-x_{ij}/\theta_{2i}}$$

O método misto de estimativa dos parâmetros pode não convergir caso haja, entre as séries da região homogênea, uma série com características marcadamente aderentes a uma exponencial simples ou a uma distribuição de Gumbel. Nesse caso, o valor de θ_* se aproxima de 1 e as estimativas dos demais parâmetros podem tender ao infinito.

Gabriele & Villani (1992) propuseram um método que, antes de estimar os parâmetros, verifica a assimetria da amostra e, caso essa seja inferior a um determinado valor, utiliza o modelo de Gumbel para a determinação dos parâmetros Λ_{1i} e θ_{1i} . Esse método permite eliminar as eventuais falhas dos demais métodos descritos e não altera significativamente as propriedades estatísticas dos parâmetros. No entanto, esse método carece de uma definição clara sobre o valor de assimetria amostral abaixo do qual os parâmetros são estimados pela distribuição de Gumbel.

3.4.5.2 Segundo nível de regionalização – Estimativa de Λ_{1s}

Um método de estimativa de Λ_{1s} alternativo àquele apresentado no item 3.4.3 pode ser desenvolvido considerando-se a transformação a seguir:

$$X' = \frac{X}{E[X]}$$

A variável X' é aleatória, independente e igualmente distribuída segundo a distribuição de probabilidades acumulada denominada curva adimensional de frequência, que no caso da TCEV assume a seguinte forma:

$$F(x') = \exp \left[-\Lambda_1 \exp(-\eta x') - \Lambda_* \Lambda_1^{1/\theta_*} \exp\left(-\frac{\eta x'}{\theta_*}\right) \right] \quad (3.59)$$

sendo,

$$x' > 0; \Lambda_1 > 0; \Lambda_* > 0; \theta_* > 0;$$

$$\eta = \ln \Lambda_1 + \gamma_e - \sum_{j=1}^{\infty} \frac{(-1)^j \Lambda_*^j}{j!} \Gamma\left(\frac{j}{\theta_*}\right) \quad (3.60)$$

$$\gamma_e = 0,5772 \quad (\text{constante de Euler})$$

A função de máxima verossimilhança pode ser obtida a partir da função densidade de probabilidade, a qual é dada por:

$$f(x') = \frac{\partial}{\partial x'} F(x') = F(x') \left[\eta \Lambda_1 \exp(-\eta x') + \eta \frac{\Lambda_*}{\theta_*} \Lambda_1^{1/\theta_*} \exp\left(-\frac{\eta x'}{\theta_*}\right) \right] \quad (3.61)$$

A equação 3.60 pode ser reescrita da seguinte forma:

$$\eta = \ln \Lambda_1 + k \quad (3.62)$$

com

$$k = \gamma_e - \sum_{j=1}^{\infty} \frac{(-1)^j \Lambda_*^j}{j!} \Gamma\left(\frac{j}{\theta_*}\right) \quad (3.63)$$

Substituindo-se 3.62 em 3.59 e 3.61, obtêm-se:

$$F(x') = \exp \left[-\Lambda_1^{1-x'} \exp(-kx') - \Lambda_* \Lambda_1^{\frac{1-x'}{\theta_*}} \exp\left(-\frac{kx'}{\theta_*}\right) \right] \quad (3.64)$$

e

$$f(x') = F(x') \left(\ln \Lambda_1 + k \right) \left[\Lambda_1^{1-x'} \exp(-kx') - \frac{\Lambda_*}{\theta_*} \Lambda_1^{\frac{1-x'}{\theta_*}} \exp\left(-\frac{kx'}{\theta_*}\right) \right] \quad (3.65)$$

A equação 3.65 permite obter a função logaritmo de verossimilhança (LV), a saber:

$$LV = n \ln(\ln \Lambda_1 + k) + \sum_{i=1}^n \left\{ -\Lambda_1^{1-x_i'} \exp(-kx_i') - \frac{\Lambda_*}{\theta_*} \Lambda_1^{\frac{1-x_i'}{\theta_*}} \exp\left(-\frac{kx_i'}{\theta_*}\right) \right\} + \ln \left[\Lambda_1^{1-x_i'} \exp(-kx_i') + \frac{\Lambda_*}{\theta_*} \Lambda_1^{\frac{1-x_i'}{\theta_*}} \exp\left(-\frac{kx_i'}{\theta_*}\right) \right] \quad (3.66)$$

Maximizar a função LV em relação ao parâmetro Λ_1 implica em:

$$\frac{\partial}{\partial \Lambda_1} LV = 0$$

Assim, tem-se que:

$$\frac{n}{\Lambda_1 (\ln \Lambda_1 + k)} + \sum_{i=1}^n \left\{ -(1-x_i') \Lambda_1^{-x_i'} \exp(-kx_i') - \frac{\Lambda_*}{\theta_*} (1-x_i') \Lambda_1^{\frac{1-x_i'}{\theta_*}-1} \exp\left(-\frac{kx_i'}{\theta_*}\right) + \frac{(1-x_i') \Lambda_1^{-x_i'} \exp(-kx_i') + (1-x_i') \frac{\Lambda_*}{\theta_*} \Lambda_1^{\frac{1-x_i'}{\theta_*}-1} \exp\left(-\frac{kx_i'}{\theta_*}\right)}{\Lambda_1^{1-x_i'} \exp(-kx_i') + \frac{\Lambda_*}{\theta_*} \Lambda_1^{\frac{1-x_i'}{\theta_*}} \exp\left(-\frac{kx_i'}{\theta_*}\right)} \right\} = 0 \quad (3.67)$$

Fazendo

$$\phi_i = \Lambda_1^{-x_i'} e^{-kx_i'} \quad (3.68)$$

e

$$\psi_i = \frac{\Lambda_*}{\theta_*} \Lambda_1^{\frac{1-x_i'}{\theta_*}-1} e^{-\frac{kx_i'}{\theta_*}} \quad (3.69)$$

a função de máximo verossimilhança torna-se:

$$\frac{1}{\Lambda_1} = \frac{\ln \Lambda_1 + k}{n} \sum_{i=1}^n (1 - x_i') \left(\phi_i + \psi_i - \frac{\phi_i + \frac{\psi_i}{\theta_i}}{\Lambda_1 (\phi_i + \psi_i)} \right) \quad (3.70)$$

A equação 3.70 fornece a estimativa de Λ_1 para cada sub-região, considerando os valores de θ_* e Λ_* calculados no primeiro nível. A série de dados x_i' é obtida a partir de todas as séries da sub-região adimensionalizadas pelas suas respectivas médias amostrais.

3.4.5.3 Estimativa vinculada de Λ_{1s} , θ_* e Λ_*

Baseado nos métodos descritos anteriormente, Gabriele & Iiritano (1994) propuseram um método que estimar Λ_{1s} , θ_* e Λ_* de uma forma vinculada. O método apresenta a vantagem de reduzir o número de equações necessárias para a obtenção dos parâmetros de cada região. Além disso, o método permite superar a possibilidade de erro na estimativa separada dos parâmetros.

Segundo os métodos anteriormente descritos, a regionalização no primeiro e segundo nível de uma região composta de k séries e M sub-regiões homogêneas necessita da estimativa de $2 + 2k + M$ parâmetros. O método de estimativa vinculada estima os M parâmetros Λ_{1s} sem considerar a estimativa dos k parâmetros Λ_{1i} . Dessa forma, o número total de parâmetros estimados é reduzido para $2 + k + M$.

O procedimento para a aplicação do método é o seguinte:

- a) Atribui-se o valor inicial para os parâmetros Λ_{1s} , θ_* e Λ_* para cada sub região;
- b) Para cada estação estima-se o parâmetro $\theta_{1j|s}$ dado pela equação 3.57;
- c) Calcula-se o valor de y' pela equação

$$Y' = Y_{ij|s} = \frac{X_{ij}}{\theta_{1j|s}} - \ln \Lambda_{1s}$$

com,

$$i = 1, \dots, k$$

$$j = 1, \dots, n_i$$

$$s = 1, \dots, M$$

d) Com y' , estimam-se os valores de Λ_{1s} , θ_* e Λ_* que resolvem as equações 3.55, 3.56 e 3.70;

e) Com os valores de Λ_{1s} , θ_* e Λ_* calcula-se o novo valor de $\theta_{1j|s}$ pela equação 3.57; e

f) Repete-se o algoritmo a partir do ponto (c) até que a diferença dos parâmetros em duas iterações sucessivas seja menor que uma tolerância pré determinada.

3.4.5.4 Estimativa dos parâmetros por momentos-L

Os métodos para a determinação dos parâmetros regionais mostrados anteriormente apresentam o inconveniente de não convergir para o valor ótimo, caso pelo menos uma das condições a seguir enunciadas seja verificada:

- Presença de séries com baixos coeficientes de assimetria;
- Presença de séries com características marcadamente de uma exponencial simples ou da distribuição Gumbel; ou
- Valores iniciais dos parâmetros muito distantes do valor ótimo.

A determinação dos parâmetros a partir dos momentos-L é menos influenciada pelas condições acima. Com efeito, a determinação dos parâmetros não se baseia nos coeficientes de variação e assimetria convencionais. Além disso, as expressões definidas a seguir são bem mais simples que aquelas de máximo de verossimilhança, o que torna a solução numérica mais robusta e menos dependente dos valores iniciais.

Para o cálculo dos parâmetros da TCEV, é necessário estabelecer os quatro primeiros momentos-L. Beran et al. (1986) deduziram os momentos ponderados por probabilidade para a TCEV da seguinte forma:

$$\beta_r = E[xF(x)^r] = MPP_r^{(1)} + \frac{\theta_1}{r+1} \cdot T_r \quad (3.71)$$

na qual,

$$MPP_r^{(1)} = \frac{\theta_1}{r+1} [\gamma_e + \ln(\Lambda_1) + \ln(r+1)]; \quad (3.72)$$

$$T_r = \sum_{j=1}^{\infty} \frac{(-1)^{j-1} \cdot \Lambda_*^j \cdot (r+1)^{j(1-1/\theta_*)} \cdot \Gamma\left(\frac{1}{\theta_*}\right)}{j!} e \quad (3.73)$$

γ_e é a constante de Euler ($\gamma_e = 0,5772\dots$)

Os momentos-L são obtidos a partir da combinação linear dos momentos ponderados pela probabilidade β_r , ou seja:

$$\lambda_{r+1} = \sum_{k=0}^r l_{r,k}^* \cdot \beta_k \quad (3.74)$$

sendo,

$$l_{r,k}^* = (-1)^{r-k} \binom{r}{k} \binom{r+k}{k} = \frac{(-1)^{r-k} (r+k)!}{(k!)^2 (r-k)!} \quad (3.75)$$

Os quatro primeiros momentos-L são, portanto:

$$\lambda_1 = \beta_0 = \theta_1 (\gamma + \ln \Lambda_1 + T_0) \quad (3.76)$$

$$\lambda_2 = 2\beta_1 - \beta_0 = \theta_1 (\ln 2 + D_1) \quad (3.77)$$

$$\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0 = \theta_1 (\ln 9/8 + 2D_2 - D_1) \quad (3.78)$$

$$\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 = \theta_1 (\ln 2^{16}/3^{10} + 5D_3 - 5D_2 + D_1) \quad (3.79)$$

com $D_r = T_r - T_{r-1}$ e T_r dados pela equação 3.73.

Para a estimativa dos parâmetros da TCEV, para uma única série, basta igualar os momentos das equações 3.76 a 3.79 aos momentos-L amostrais. Para a estimativa regional dos parâmetros, é necessário definir as seguintes quantidades:

$$CV - L \quad \equiv \tau = \frac{\lambda_2}{\lambda_1} \quad (3.80)$$

$$Assimetria - L \quad \equiv \tau_3 = \frac{\lambda_3}{\lambda_2} \quad (3.81)$$

$$Curtose - L \quad \equiv \tau_4 = \frac{\lambda_4}{\lambda_2} \quad (3.82)$$

Os parâmetros regionais θ_* e Λ_* podem ser obtidos igualando-se a assimetria-L regional (t_3^R) e a curtose-L regional (t_4^R) com seus valores esperados τ_3 e τ_4 , respectivamente. Hosking & Wallis (1997) estimaram os valores regionais para a assimetria-L e a curtose-L a partir da média ponderada de seus valores locais. Formalmente, tem-se:

$$t_k^R = \frac{\sum_{i=1}^m n_i \cdot \tau_k^i}{\sum_{i=1}^m n_i}; \quad k = 3 e 4 \quad (3.83)$$

sendo m o número de estações da região considerada, n_i o número de dados da série i e τ_k^i o quociente de momento-L de ordem k , da série i .

A partir das equações 3.81, 3.82 e 3.83, têm-se:

$$\frac{\ln 9/8 + 2D_2 - D_1}{\ln 2 + D_1} - t_3^R = 0 \quad (3.84)$$

$$\frac{\ln 2^{16}/3^{10} + 5D_3 - 5D_2 + D_1}{\ln 2 + D_1} - t_4^R = 0 \quad (3.85)$$

As soluções do sistema formado pelas equações 3.84 e 3.85 fornecem os valores regionais θ_* e Λ_* .

3.5 Relação pico-volume

A relação entre volume e pico do escoamento direto é de importância em uma larga variedade de análises hidrológicas, especialmente onde os dados são escassos (SINGH & AMINIAN, 1986). Tais relações prestam-se não só às tarefas afeitas à engenharia de recursos hídricos, como, por exemplo, as estimativas de cheias de projeto, como também a argumentações importantes sobre os conceitos de linearidade e não-linearidade de bacias hidrográficas, tal como introduzidos por Rogers (1980). Como será visto no item 3.6, a relação pico-volume é de fundamental importância na determinação da distribuição dos picos condicionada aos volumes $f_{y|x}(y|x)$.

Segundo Sivapalan et al. (2002), duas definições de “linearidade” ou “não linearidade” têm sido usadas em hidrologia. A primeira definição diz respeito ao que aqueles autores chamam de ‘propriedade dinâmica’ da bacia, ou seja, a transformação chuva-vazão. Nesse sentido, linearidade refere-se à dependência linear entre alturas de chuva (entrada) e volumes de cheia (resposta). A segunda definição de linearidade diz respeito à dependência estatística entre uma determinada variável (e.g: vazão média anual) e a área de drenagem da bacia. As duas definições são intuitivas e bem estabelecidas do ponto de vista matemático, mas são diferentes e de algum modo controversas do ponto de vista hidrológico. De fato, pode-se demonstrar que as duas definições de linearidade existem independentemente uma da outra.

No presente trabalho, o termo “linearidade” se refere à primeira definição, ou seja, linearidade ou bacia de drenagem linear é definida como a bacia de drenagem na qual o volume de escoamento superficial total é diretamente proporcional ao volume precipitado (ROGERS, 1982). Como se verá a seguir, essa definição traz consigo algumas conseqüências importantes para a relação entre vazões de pico e volumes de cheia.

As condições físicas para a ocorrência da linearidade hidrológica envolvem outras variáveis hidrológicas, notadamente a infiltração e a evaporação. Para uma linearidade perfeita entre

precipitação e volume escoado, a bacia de drenagem deve ter sua superfície uniforme e impermeável. Além disso, deve-se ter uma declividade constante, com uma precipitação uniforme sobre toda a área. Dessa maneira, não há perdas por infiltração e evaporação, e a velocidade de escoamento em qualquer parte é a mesma. Sob tais condições, a função que relaciona o volume precipitado e o escoado é dada por uma reta com inclinação +1. De modo mais geral, não há necessidade de se ter infiltração e evaporação nulas; basta que estas sejam constantes em toda bacia para se ter a condição de linearidade. Essas são premissas também utilizadas na formulação da teoria do hidrograma unitário, as quais implicam em linearidade mútua entre alturas de chuva efetiva, volumes de escoamento superficial e, devido à fixação da forma da função de transferência, vazões de pico. Ressalte-se, entretanto, que a metodologia do hidrograma unitário foi desenvolvida para escoamento superficial direto. Assim, é necessário proceder à separação do escoamento para a aplicação do método; as hipóteses formuladas por Rogers (1980), a serem descritas a seguir, prescindem da separação do escoamento.

Várias relações entre vazão de pico e volume escoado vêm sendo propostas nos últimos anos. Rogers (1980) estudou dados de 42 bacias de drenagem dos EUA com áreas entre 5 e 700 km^2 . Posteriormente, o estudo foi estendido para bacias com áreas de até 23.000 km^2 (ROGERS, 1982; ROGERS & ZIA, 1982). Assim, foram estabelecidas relações entre a vazão de pico e o volume de escoamento superficial dos seguintes tipos:

$$\ln Q_p = b + m \cdot \ln(V) \quad (3.86)$$

$$\ln(Q_p/V) = b + (m - 1) \cdot \ln(V) \quad (3.87)$$

$$\ln(Q_p/V^2) = b + (m - 2) \cdot \ln(V) \quad (3.88)$$

nas quais Q_p é o pico do hidrograma em m^3/s , V é o volume total escoado em cm , b é o intercepto da relação considerada, correspondente a um *runoff* unitário e m é a inclinação da reta que melhor se ajusta aos pares P-V.

O modelo dado pela equação 3.86 é denominado “Distribuição Padrão Pico-Volume” (DPPV). O modelo dado pela equação 3.87 é denominado “Distribuição Pico-Volume de Primeira Ordem” (DPVPO) e o modelo da equação 3.88 é denominado “Distribuição Pico-Volume de Segunda Ordem” (DPVSO).

Nota-se, claramente, que os modelos DPVPO e DPVSO são deduzidos do modelo DPPV. No primeiro caso, divide-se o pico pelo volume e, no segundo, divide-se pelo volume ao quadrado. Rogers (1980) justifica o emprego do modelo DPVSO a partir de considerações físicas, considerando o tempo de base e os tempos de ascensão e recessão do hidrograma.

Os picos serão diretamente proporcionais aos volumes se a inclinação m no modelo DPPV for +1, enquanto no modelo DPVPO a inclinação deve ser zero e no modelo DPVSO a inclinação deve ser -1. Pequenos coeficientes de inclinação indicam que não há linearidade hidrológica (MIMIKOU, 1983).

Mimikou (1983) aplicou os modelos referidos anteriormente a oito bacias da Grécia com áreas de drenagem entre 200 e 6000 km^2 . Os modelos foram ajustados aos dados pelo método dos mínimos quadrados. A inclinação para o modelo DPPV variou entre 0,279 e 0,960, enquanto para o modelo DPVSO a inclinação variou entre -0,721 e -0,040 e para o modelo DPVSO a inclinação variou entre -1,721 e -1,040. Tais coeficientes são compatíveis com aqueles encontrados por Rogers (1980, 1982) para algumas bacias nos EUA.

Segundo Mimikou (1983), sucessivas subtrações de $LogV$ no modelo original (DPPV) causam distorções na construção da relação entre pico e volume. No caso do modelo DPVPO, há um completo desaparecimento da independência entre as variáveis. Além disso, há grandes variações no coeficiente de determinação com a alteração da inclinação do modelo (coeficiente m). No caso do modelo DPVSO, a distorção aparece com o aumento da inclinação. Para inclinações próximas a +1 para o modelo original, no caso de bacias lineares, o coeficiente de determinação tende a se aproximar de 1. No entanto, quando o modelo DPVSO foi aplicado a essas bacias o coeficiente de determinação decresceu com o aumento da inclinação ($m \rightarrow 1$).

O modelo DPPV possui toda a informação necessária para determinar a resposta da bacia e avaliar o grau de proporcionalidade entre os volumes e os picos. Isso evidencia a não obrigatoriedade de padronização das variáveis, tal como nos modelos DPVPO e DPVSO. Além disso, a eventual possibilidade de correlação espúria é eliminada ao se utilizar o modelo DPPV (MOLFINO & CRUISE, 1990). Assim, para a determinação da distribuição $f_{Y|X}(y|x)$, de acordo com a metodologia a ser proposta no item 3.6, utilizou-se o modelo DPPV, representado pela equação 3.86.

3.6 O método PVP

3.6.1 Introdução

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

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

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

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

Uma vantagem adicional do método PVP é que a análise probabilística envolvendo chuva e vazão evita o absurdo de se ter um volume escoado superior ao precipitado, para uma mesma duração e um mesmo período de retorno. Entretanto, ressalta-se que, para uma aplicação bem sucedida do método PVP, estão implícitas as seguintes condicionantes:

- Admite-se que a relação entre os volumes escoados e o pico de vazão é bem definida pelo modelo pico-volume;
- A densidade de postos pluviométricos é alta o suficiente para refletir a variabilidade espaço-temporal da chuva média sobre a bacia;
- A introdução de fontes extras de incerteza (entrada do modelo, estrutura do modelo e parâmetros) não afeta, significativamente, as estimativas dos picos de vazão.

3.6.2 Método pico-volume com a chuva como variável aleatória auxiliar

A figura 3.6 mostra um processo pontual marcado para eventos de vazão tendo a chuva como variável aleatória auxiliar. Conforme visto anteriormente, se $g_u(x)$ representa a distribuição de probabilidades dos volumes e $f_{Y|X}(y|x)$ representa a distribuição dos picos dado o volume, a distribuição anual de probabilidades dos picos de vazão pode ser determinada pela solução conjunta das duas equações a seguir:

$$H_u(y_p) = \int_0^{y_p} \int_0^{\infty} f_{Y|X}(y|x) \cdot g_u(x) dx dy \quad (3.89)$$

e

$$F_M(y_p) = \exp\{-\Lambda[1 - H_u(y_p)]\} \quad (3.90)$$

Na formulação original do método pico-volume, por Bradley & Potter (1992), a distribuição dos volumes é estimada a partir da análise de frequência convencional. Naghettini (1994) utilizou os totais precipitados de duração d para extrapolar a distribuição dos volumes acima dos valores observados, ou seja, para valores extremos de volumes. Nessa adaptação da metodologia do pico-volume, as duas condições do método GRADEX devem ser utilizadas em conjunto com as equações 3.89 e 3.90.

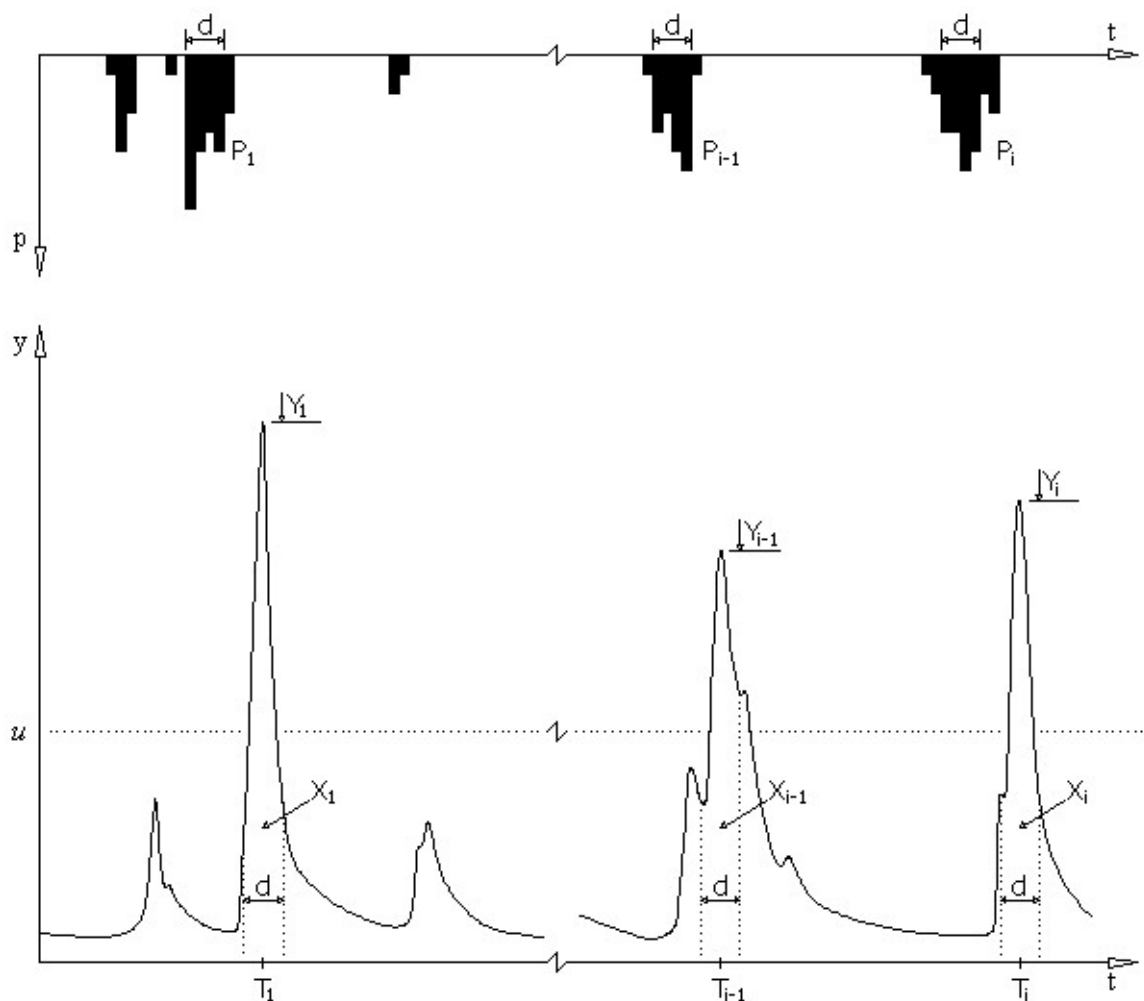


Figura 3.6 – Esquema de um processo pontual marcado tendo a chuva como variável aleatória auxiliar

3.6.3 Determinação da distribuição dos volumes

A distribuição dos volumes de cheia de duração d , a qual é tomada como igual ao tempo de base da bacia, pode ser obtida a partir da distribuição da chuva de mesma duração d com a metodologia convencional do GRADEX. O modelo TCEV, descrito anteriormente, pode ser usado para tal fim. A TCEV tem a vantagem de ser flexível o suficiente para abranger as condições impostas pelo método GRADEX, particularmente a hipótese de exponencialidade da cauda superior.

A TCEV é dada pelo produto de duas distribuições Gumbel, uma modelando os dados mais frequentes e a outra modelando os dados raros. Uma vez que a distribuição Gumbel tem cauda superior exponencial e a TCEV tende para uma Gumbel para valores extremos, respeita-se a condição de exponencialidade do método GRADEX.

Por outro lado, pode-se ajustar a TCEV de modo que sua cauda superior esteja situada no domínio onde os incrementos de volumes da chuva produzem incrementos equivalentes de volume de *runoff*, ou seja, no domínio onde a bacia se encontra saturada e a segunda hipótese do método GRADEX pode ser verificada. Para isso, considere a distribuição TCEV reescrita na forma do produto de duas distribuições Gumbel, isto é:

$$F(x) = \exp\left[-\exp\left(-\frac{x-\beta_1}{\alpha_1}\right)\right] \cdot \exp\left[-\exp\left(-\frac{x-\beta_2}{\alpha_2}\right)\right] \quad (3.91)$$

Em comparação com a equação 3.38, têm-se:

$$\Lambda_i = \exp\left(\frac{\beta_i}{\alpha_i}\right), \quad i = 1, 2 \quad (3.92)$$

e

$$\theta_i = \alpha_i, \quad i = 1, 2 \quad (3.93)$$

O método consiste em se ajustar separadamente a cauda inferior e a superior da distribuição dos volumes escoados, de modo a considerar os dados locais e regionais da bacia em questão. Inicialmente ajusta-se a TCEV sobre os dados de chuva média obtendo-se os parâmetros locais Λ_1^{PLU} , Λ_2^{PLU} , θ_1^{PLU} e θ_2^{PLU} .

Os parâmetros Λ_1^{PLU} e θ_1^{PLU} são obtidos por momentos-L, por meio da solução do sistema formado pelas equações 3.76 e 3.77. Os parâmetros da cauda superior são obtidos pela relação entre os parâmetros da cauda inferior e os parâmetros regionais deduzidos da regionalização das alturas pluviométricas, ou seja:

$$\Lambda_2^{PLU} = \Lambda_* \cdot (\Lambda_1^{PLU})^{1/\theta_*} \quad (3.94)$$

e

$$\theta_2^{PLU} = \theta_* \cdot \theta_1^{PLU} \quad (3.95)$$

A aplicação do método GRADEX exige que a cauda superior da distribuição dos volumes escoados seja paralela à cauda superior da distribuição da chuva. Isso implica que o parâmetro θ_2 da distribuição da chuva seja proporcional ao da distribuição dos volumes. Tal imposição não permite a aplicação da análise de frequência convencional aos dados de volume de forma a obter o ajuste da TCEV. Uma forma de superar esse obstáculo é considerar que as caudas inferior e superior da TCEV sejam independentes entre si. Assim, pode-se modelar cada parte da distribuição em momentos separados.

A cauda inferior da distribuição dos volumes é ajustada, por análise de frequência convencional, até o período de retorno T_c . A partir desse ponto a distribuição é extrapolada pelo método GRADEX tendo como auxílio a distribuição da chuva.

O período de retorno T_c é de difícil definição. Vários autores (DUBAND et al., 1988; GUILLOT & DUBAND, 1967) propuseram valores arbitrários variando entre 10 e 50 anos dependendo das condições geomorfológicas da bacia. Recomenda-se que, a menos que se tenha boas razões físicas, T_c seja definido com base nos dados observados de volumes de cheia utilizando alguma medida de tendência central das estatísticas de ordem superior da série. Uma justificativa para essa recomendação é que, para séries observadas longas, pode-se admitir que os dados com baixas probabilidades empíricas de excedência (grande períodos de retorno) contemplem a condição de saturação da bacia. Adverte-se, no entanto, que a aplicabilidade dessa justificativa é inversamente proporcional ao tamanho e à homogeneidade da bacia. Assim, para grandes bacias, com geomorfologia bastante heterogênea, as séries observadas de volumes de cheia podem não fornecer uma boa estimativa de T_c .

Dado o valor de T_c , os parâmetros da distribuição dos volumes podem ser determinados por meio das seguintes relações:

$$\alpha_1^{FLU} = \sqrt{\frac{\sigma_{FLU}^2}{1,645}} \quad (3.96)$$

$$\beta_1^{FLU} = \mu_{FLU} - 0,577 \cdot \alpha_1^{FLU} \quad (3.97)$$

$$\alpha_2^{FLU} = K \cdot \alpha_2^{PLU} \quad (3.98)$$

$$\beta_2^{FLU} = X(T_c) + \alpha_2^{FLU} \cdot \ln \left[-\ln \left(1 - \frac{1}{T_c} \right) \right] \quad (3.99)$$

nas quais σ_{FLU}^2 é a variância da série parcial dos volumes escoados de mesma duração d e intensidade de Poisson da série parcial igual à da chuva média; μ_{FLU} é a média da série parcial dos volumes escoados de mesma duração d e intensidade de Poisson da série parcial igual à da chuva média; K é uma constante que converte o gradex da chuva para a mesma unidade de medida da série dos volumes, conforme mostrado no tópico 3.3.3; e, $X(T_c)$ é o quantil estimado pela distribuição da cauda inferior dos volumes a partir do qual a segunda hipótese do método GRADEX é válida.

$X(T_c)$ é obtido pela função inversa da distribuição de Gumbel no ponto T_c . Formalmente, tem-se:

$$X(T_c) = \beta_1^{FLU} - \alpha_1^{FLU} \cdot \ln \left[-\ln \left(1 - \frac{1}{T_c} \right) \right] \quad (3.100)$$

As equações 3.96 a 3.99 foram obtidas pela análise de frequência convencional utilizando-se o método dos momentos. No entanto, não há restrição quanto ao uso de outra metodologia para a determinação dos parâmetros, desde que a estrutura do método GRADEX seja mantida.

Retomando a equação 3.38, a distribuição acumulada dos volumes de cheia é dada por:

$$G(x) = \exp \left[-\Lambda_1^{FLU} \cdot \exp \left(-\frac{x}{\theta_1^{FLU}} \right) - \Lambda_2^{FLU} \cdot \exp \left(-\frac{x}{\theta_2^{FLU}} \right) \right] \quad (3.101)$$

A distribuição de probabilidades $g(x)$ é obtida derivando-se $G(x)$ em relação a x .

A partir das equações 3.96 a 3.99, e com as equações 3.92 e 3.93, define-se a distribuição dos volumes de cheia. É interessante notar que a aplicação do método GRADEX, conforme aqui proposta, agrega de forma simples e sistemática as características regionais da bacia em relação ao comportamento da cauda superior da distribuição da chuva através do parâmetro θ_* , bem como a transformação chuva-vazão em condições extremas através da passagem

$\alpha_2^{PLU} \rightarrow \alpha_2^{FLU}$. Isso evidencia o atributo do método PVP em buscar a agregação de mais informação disponível para a determinação dos quantis extremos de vazão.

3.6.4 Determinação da distribuição dos picos condicionada aos volumes

A distribuição condicionada $f_{Y|X}(y|x)$ é aqui estimada usando um modelo regional não linear que relaciona as vazões máximas diárias, ou de pico, aos respectivos volumes de cheia de duração igual ao tempo de base. Tal modelo regional, cujas particularidades serão objeto do item 4.4.1 do capítulo 4, tem a seguinte estrutura geral:

$$Y = w(X) + \varepsilon \quad (3.102)$$

na qual Y representa as vazões máximas diárias (ou de pico, para um certo hidrograma), X denota os volumes e ε os erros da regressão.

A função w é a equação que relaciona Y e X . Para efeito de construção de um modelo regional, tais variáveis podem ser expressas, como se verá no item 4.4.1, sob forma de deflúvios.

Admitindo-se que os erros da regressão 3.102 são homocedásticos, a quantidade $Y - w(X)$ é normalmente distribuída com média zero e desvio padrão σ_ε . Uma vez que Y representa os picos acima de um limiar u , a distribuição dos picos condicionada aos volumes é uma normal truncada abaixo de u . Formalmente, tem-se:

$$F_{Y|X}(y|x) = \frac{\Phi\left[\frac{y-w(x)}{\sigma_\varepsilon}\right] - \Phi\left[\frac{u-w(x)}{\sigma_\varepsilon}\right]}{1 - \Phi\left[\frac{u-w(x)}{\sigma_\varepsilon}\right]} \quad (3.103)$$

sendo $\Phi[\cdot]$ a distribuição acumulada de probabilidade para a variável normal padrão.

O desvio padrão σ_ε é estimado usando os dados observados de pico e volume. Diferenciando a equação 3.103, obtém-se a distribuição $f_{Y|X}(y|x)$.

Na verdade, não é necessário diferenciar a equação 3.103 para utilizá-la na solução da equação 3.89. Como $f_{y|x}(y|x)$ representa a densidade condicionada de y para um valor fixo de x , pode-se separar a integral da equação 3.89 e resolvê-la separadamente para o domínio de y , ou seja:

$$H_u(y_p) = \int_0^\infty \left\{ \int_0^{y_p} f_{y|x}(y|x) \cdot g(x) dy \right\} dx \quad (3.104)$$

Como $g(x)$ não depende de y tem-se:

$$H_u(y_p) = \int_0^\infty F_{y|x}(y_p|x) \cdot g(x) dx \quad (3.105)$$

Na equação anterior, $F_{y|x}$ é dada pela equação 3.103 e g pela derivada da equação 3.101.

No capítulo seguinte o método PVP será aplicado a uma bacia de forma a ilustrar, por meio de um estudo de caso, os vários passos metodológicos apresentados neste capítulo.

4 RESULTADOS

4.1 Introdução

A metodologia apresentada no capítulo 3 foi objeto de aplicação à bacia do rio Pará, em Ponte do Vilela. Essa bacia foi escolhida por ter sido alvo de recentes estudos de regionalização de vazão e precipitação por diferentes metodologias, como, por exemplo, em CPRM (2001), Quintão (2003) e Lima (2004). Esse fato permitiu a verificação dos resultados gerados e a análise comparativa entre os diversos métodos de estimação de quantis máximos de vazão.

Além disso, a bacia escolhida possui uma extensa série de dados de vazão e uma boa distribuição de estações pluviométricas, o que permitiu uma análise completa de cada parte da metodologia desenvolvida.

4.2 A bacia do rio Pará

O rio Pará se localiza entre as longitudes 44° e 45° e latitudes 20° e 21° na bacia do alto rio São Francisco, no estado de Minas Gerais. A nascente é na Serra da Cebola, a uma altitude de 1160 m, no município de Resende Costa, no estado de Minas Gerais. Seus principais afluentes são os rios Itapecerica, Lambari e Picão, pela margem esquerda, e o rio São João, pela margem direita. O rio Pará percorre uma extensão de cerca de 310 km até desaguar no rio São Francisco.

Segundo CPRM (1995), as características principais da região estudada são:

- Quanto ao relevo, a bacia é caracterizada por planaltos, depressões e áreas dissecadas resultantes da alternância de atuação de processos morfoclimáticos associados ao condicionamento geológico. A unidade geomorfológica predominante é “Planalto Dissecado do Centro-Sul e Leste de Minas”;
- Quanto à geologia, a região situa-se basicamente sobre rochas de idade pré-cambriana, desde arqueanas até proterozóicas inferiores;
- A vegetação nativa da região é constituída predominantemente pelo cerrado, que se estendia originalmente por quase toda a sub-bacia. Esse tipo de vegetação é característico de regiões de clima semi-úmido, constituindo-se principalmente de gramíneas, arbustos e árvores de médio porte, tendo como principais características troncos e galhos retorcidos e

folhas espessas e coriáceas. Atualmente, principalmente nas regiões mais planas, grandes áreas de cerrado vêm sendo substituídas por pastagens e áreas de cultivo mecanizado, ou mesmo devastadas para a exploração de lenha e carvão. Nas regiões de relevo mais acidentado, entretanto, o cerrado se encontra mais preservado;

- Quanto ao clima, a região é classificada pela tipologia Cwb, de acordo com a classificação de Köppen. Há uma predominância de temperaturas medianas a elevadas na região durante quase todo o ano, principalmente na primavera e no verão. O mês mais quente, janeiro, apresenta média das máximas variando entre 28 e 30°C. Já no período outono/inverno ocorre um significativo decréscimo na temperatura, sendo que a média das mínimas do mês mais frio, julho, varia entre 8 e 10°C. A amplitude térmica anual é da ordem de 12 a 14°C. Os totais pluviométricos anuais variam, de forma geral, entre 1700 *mm* nas cabeceiras da bacia a 1150 *mm* na região próxima à barragem de Três Marias. O regime pluviométrico é típico das regiões de clima tropical, com a ocorrência dos valores mensais máximos no período de verão e dos mínimos no inverno;
- O sistema de aquíferos é caracterizado por rochas ígneas e metamórficas de alto grau. Por englobar tipos rochosos bastante distintos, essa unidade apresenta grande variabilidade nos parâmetros hidráulicos como função direta da densidade de estruturas de descontinuidade e do grau de intemperismo. De modo geral, as rochas mais antigas, submetidas a sucessivos eventos tectono-metamórficos, exibem rede mais densa de estruturas e, portanto, maior potencial hidrogeológico. Rochas intrusivas, de idade cambriana ou superior, são as que exibem menor capacidade de produção de água e podem funcionar, localmente, como barreiras hidráulicas, em virtude da baixa permeabilidade.

O ponto de referência adotado para a aplicação da metodologia foi a estação fluviométrica Ponte do Vilela (Código 40130001), localizada a 73 *km* da nascente do rio Pará. A figura 4.1 mostra a localização da bacia em relação ao estado de Minas Gerais.

4.3 Análise das séries de duração parcial de vazão e precipitação

4.3.1 Vazão

A estação fluviométrica utilizada foi a de Ponte do Vilela, operada pela Companhia Energética de Minas Gerais (CEMIG). Essa estação possui uma área de drenagem de 1620

km^2 , abrangendo quase a totalidade da área a montante da represa de Cajuru. O rio Pará, até esse ponto, tem um comprimento de 73 km e uma declividade média de 0,30%.

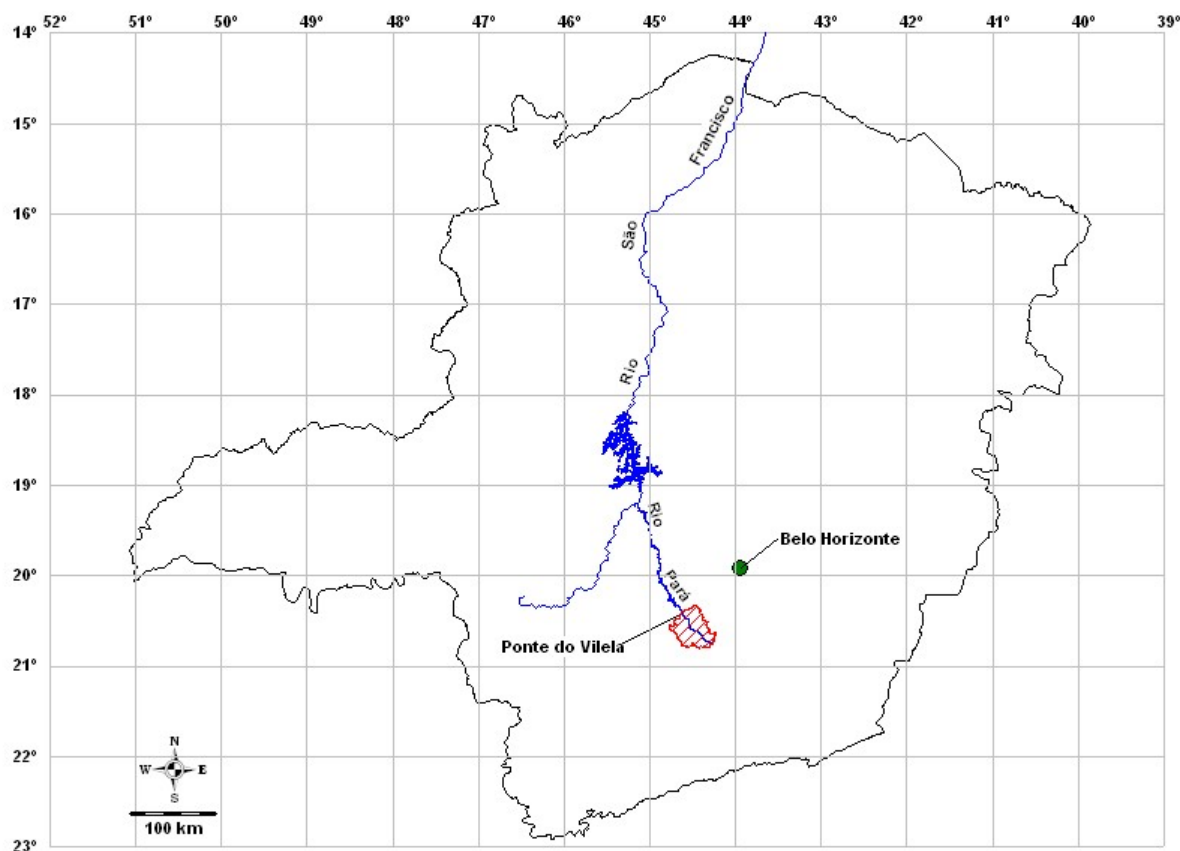


Figura 4.1 – Localização da bacia hidrográfica do rio Pará

A estação Ponte do Vilela possui dados diários de vazão de julho de 1938 a dezembro de 2001, dispondo, portanto, de 63 anos de observações. O tempo de base médio, obtido a partir de 26 hidrogramas selecionados, é de 8 dias.

A escolha da taxa anual de ocorrências Λ deve ser feita de modo a extrair o máximo de informação da série de dados mantendo, entretanto, a independência serial entre as vazões médias diárias máximas e a hipótese distributiva de Poisson quanto ao número médio de excedências anuais. Dessa forma, para uma faixa de valores de Λ entre 1,0 e 3,0, utilizou-se o teste de Cunnane (1979) para verificar a hipótese distributiva de Poisson para valores superiores ao limiar u constituintes da série de duração parcial.

Cunnane (1979) propôs um teste para verificar a hipótese de Poisson com base na estatística de dispersão de Fisher. Em seu desenvolvimento, Cunnane (1979) considerou que o número

anual de picos acima de u , denotado por m_i ($i = 1, 2, \dots, n$), pode ser modelado por uma distribuição normal com média e variância igual a Λ . Formalmente, tem-se:

$$m \sim N\left(\Lambda, \Lambda^{1/2}\right) \quad (4.1)$$

Desse modo, é possível escrever:

$$\frac{m - \Lambda}{\sqrt{\Lambda}} \sim N(0, 1) \quad (4.2)$$

Assim, pela definição da estatística do χ^2 , obtém-se:

$$\sum_{i=1}^N \left(\frac{m_i - \Lambda}{\sqrt{\Lambda}} \right)^2 \sim \chi_N^2 \quad (4.3)$$

Se Λ é estimado através da média do número anual de picos acima de u , ou seja, \bar{m} , perde-se um grau de liberdade e a equação 4.3 torna-se:

$$d = \sum_{i=1}^N \left(\frac{m_i - \hat{\Lambda}}{\sqrt{\hat{\Lambda}}} \right)^2 \sim \chi_{N-1}^2 \quad (4.4)$$

O maior valor obtido foi $\Lambda = 1,5$ para um $u = 96,6 \text{ m}^3/\text{s}$. Acima desse valor, as séries geradas apresentaram muitas ocorrências para determinados anos e nenhuma ocorrência para outros, indicando a possibilidade de dependência serial. Vale salientar que o teste do χ^2 serve ao propósito de verificação da hipótese de Poisson, não sendo válida para a hipótese de independência serial das amostras. É necessário, em paralelo, uma análise gráfica das séries. A tabela 4.1 mostra a série de picos e de volumes de cheias, de 8 dias de duração, para $\Lambda = 1,5$. A tabela 4.2 mostra o número de ocorrências para cada ano da série.

Para os 63 anos de dados de Ponte do Vilela e com um $\Lambda = 1,5$, a estatística do teste do χ^2 é $d = 78,5$. Como $d < \chi_{63;0,95}^2 = 82,5$, conclui-se não haver elementos para rejeitar a hipótese de que o número de excedências tenha uma distribuição de Poisson.

Tabela 4.1 – Picos (m^3/s) e volumes ($m^3/s.dia$) para Ponte do Vilela com $\Lambda = 1,5$

Volume	Pico	Volume	Pico	Volume	Pico	Volume	Pico
38,05	98,00	74,64	137,00	89,26	113,00	116,75	135,00
40,77	115,00	75,14	102,00	89,50	98,00	119,94	213,00
50,87	98,00	76,33	188,00	90,78	166,00	119,94	101,00
53,75	104,23	76,50	179,00	90,78	166,00	123,15	168,00
53,88	113,00	77,97	184,00	91,26	110,00	123,29	214,00
56,35	121,00	78,70	150,00	92,06	120,00	126,40	184,00
57,25	112,00	79,79	102,00	92,06	120,00	126,40	184,00
58,13	121,00	79,79	180,00	92,61	113,00	127,25	153,00
60,59	125,00	80,75	98,60	93,35	135,00	129,25	164,00
61,32	127,00	81,29	106,00	93,49	107,00	133,00	177,00
62,04	126,00	81,29	106,00	94,84	213,00	135,60	260,00
63,40	108,00	81,30	185,00	95,73	141,00	136,80	285,00
64,11	106,00	81,91	107,00	96,00	126,00	139,20	244,00
65,89	126,00	82,94	148,00	99,13	111,00	140,38	169,99
67,14	116,00	82,94	148,00	102,70	198,00	144,05	222,00
68,56	145,00	83,03	128,00	103,00	188,00	144,06	380,00
69,21	121,00	83,04	111,00	103,24	104,00	145,38	380,00
69,36	99,30	83,16	127,00	106,39	130,00	148,76	298,00
69,58	117,00	83,89	115,00	106,70	168,00	150,63	224,00
69,75	99,20	84,01	129,00	108,59	112,00	151,88	213,00
69,94	99,30	84,17	129,00	110,48	153,00	177,38	230,00
70,35	102,00	85,00	118,53	111,56	193,00	207,88	313,00
70,36	171,00	86,15	115,00	112,16	181,00	215,75	477,00
72,81	124,00	86,15	115,00	114,00	141,00		

Tabela 4.2 – Número de ocorrências (m_k) por ano

Ano	m_k	Ano	m_k	Ano	m_k	Ano	m_k
1938/1939	4	1954/1955	2	1970/1971	0	1986/1987	1
1939/1940	3	1955/1956	1	1971/1972	1	1987/1988	1
1940/1941	4	1956/1957	2	1972/1973	3	1988/1989	0
1941/1942	1	1957/1958	0	1973/1974	0	1989/1990	1
1942/1943	4	1958/1959	0	1974/1975	2	1990/1991	1
1943/1944	2	1959/1960	2	1975/1976	0	1991/1992	1
1944/1945	1	1960/1961	3	1976/1977	0	1992/1993	1
1945/1946	0	1961/1962	0	1977/1978	0	1993/1994	2
1946/1947	1	1962/1963	1	1978/1979	1	1994/1995	3
1947/1948	2	1963/1964	1	1979/1980	2	1995/1996	1
1948/1949	4	1964/1965	3	1980/1981	2	1996/1997	4
1949/1950	2	1965/1966	1	1981/1982	3	1997/1998	0
1950/1951	3	1966/1967	0	1982/1983	6	1998/1999	2
1951/1952	2	1967/1968	0	1983/1984	2	1999/2000	2
1952/1953	0	1968/1969	1	1984/1985	3	2000/2001	0
1953/1954	0	1969/1970	0	1985/1986	0		

4.3.2 Precipitação

As estações pluviométricas utilizadas no estudo foram selecionadas de acordo com sua proximidade à bacia do rio Pará em Ponte do Vilela e com a disponibilidade de dados de cada série. Assim, foram selecionadas 11 estações pluviométricas, cuja características são apresentadas na tabela 4.3. A figura 4.2 mostra a disposição espacial de cada estação. O Anexo 2 traz as séries de duração parcial das estações pluviométricas.

Tabela 4.3 – Características das estações pluviométricas da região

Código	Nome	Latitude	Longitude	Altitude	Período de dados
02044000	Ponte do Rio do Peixe	20° 55' 00"	44° 27' 00"	878	01/1941 - 12/1971
02044002	Itaúna - Montante	20° 04' 17"	44° 34' 13"	859	01/1941 - 11/2004
02044003	Carmo do Cajuru	20° 11' 32"	44° 47' 37"	746	01/1941 - 09/2004
02044005	Carmo da Mata	20° 33' 28"	44° 52' 03"	846	01/1942 - 12/2002
02044006	Divinópolis	20° 08' 13"	44° 53' 31"	672	01/1941 - 11/2004
02044009	Fazenda Campo Grande	20° 37' 31"	44° 26' 00"	915	01/1942 - 12/2002
02044011	Carmópolis de Minas	20° 33' 00"	44° 38' 00"	(1)	01/1942 - 12/1949
02044016	Fazenda Benedito Chaves	20° 10' 09"	44° 30' 54"	944	05/1970 - 07/2004
02044027	Ponte Fernão Dias	20° 44' 27"	44° 47' 00"	882	08/1966 - 12/2003
02044036	Itaguara	20° 24' 00"	44° 28' 00"	(1)	01/1942 - 12/2002
02044042	Carmo da Mata (ETA - COPASA)	20° 33' 45"	44° 52' 03"	749	01/1977 - 11/2004

(1) Dado não disponível

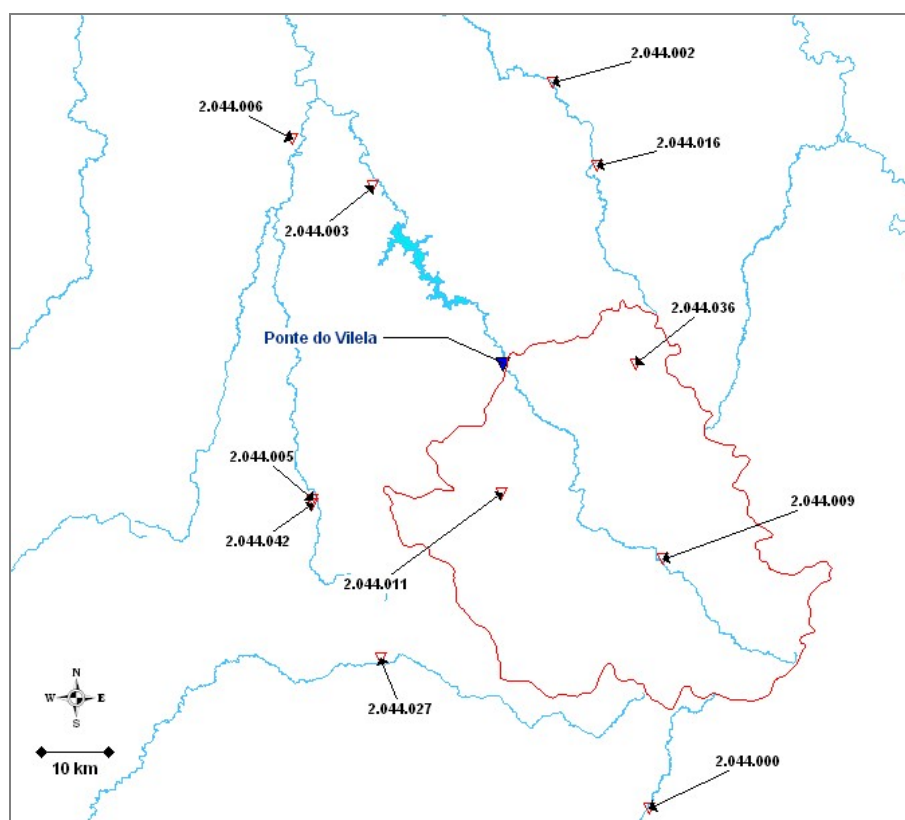


Figura 4.2 – Disposição espacial de cada estação pluviométrica

A chuva média sobre a bacia foi calculada a partir das estações 02044005, 02044009 e 02044036, por meio dos polígonos de Thiessen. Os coeficientes de ponderação para as estações são, respectivamente, 0,09, 0,64 e 0,27.

As séries parciais de precipitação foram obtidas considerando a mesma taxa de ocorrência anual Λ da série fluviométrica. A duração da chuva foi tomada como o tempo de base médio da bacia, ou seja, 8 dias. Para cada série, foi feito o teste do χ^2 ou de Cunnane, de modo a verificar a hipótese distributiva de Poisson. A tabela 4.4 traz as principais informações para a realização desse teste de hipótese.

Tabela 4.4 – Análise da independência serial das estações pluviométricas

Estação	N _{anos}	N _{série}	<i>u</i>	Λ	γ	$\chi^2_{N_{anos};0,95}$
02044000	28	43	143,0	1,54	9,8	41,3
02044002	53	79	150,2	1,49	30,0	71,0
02044003	64	96	142,4	1,50	18,5	83,7
02044005	61	93	139,9	1,52	28,7	80,2
02044006	57	85	136,6	1,49	25,3	75,6
02044009	61	97	132,5	1,59	27,3	80,2
02044011	8	12	141,1	1,50	3,8	15,5
02044016	34	51	149,6	1,50	13,5	48,6
02044027	36	54	129,4	1,50	20,5	51,0
02044036	61	94	134,0	1,54	27,3	80,2
02044042	28	42	142,3	1,50	12,5	41,3
Chuva média	61	91	127,2	1,49	19,3	80,2

N_{anos} Número de anos da estação pluviométrica

N_{série} Número de dados da série parcial

u Precipitação limiar (*mm*)

Λ Taxa anual de ocorrência (frequência de Poisson)

γ Estatística do teste de Cunnane

$\chi^2_{N_{anos};0,95}$ Distribuição χ^2 com N_{anos} graus de liberdade a um nível de significância de 5%

4.4 Relações Pico-Volume

4.4.1 Relação pico-volume para o estado de Minas Gerais

Para a determinação da relação pico-volume para o estado de Minas Gerais foi ajustado o modelo dado pela equação 3.86 aos dados obtidos de 24 estações fluviométricas, selecionadas ao longo de várias bacias hidrográficas nos estados de Minas Gerais e Goiás. As estações têm área de drenagem entre 500 e 32000 *km*².

Para a seleção das estações, levou-se em consideração as diversidades climática e geomorfológica das diversas bacias do estado de Minas Gerais. Além disso, as estações

utilizadas estão todas a montante dos principais reservatórios de regularização de vazão das bacias do rio Grande, rio São Francisco e rio Paranaíba. Assim, os hidrogramas de cada ponto não são influenciados pelo controle do reservatório, tendo-se hidrogramas que refletem a condição natural de escoamento na bacia. A figura 4.3 mostra a localização de cada estação e a hidrografia principal da área estudada.

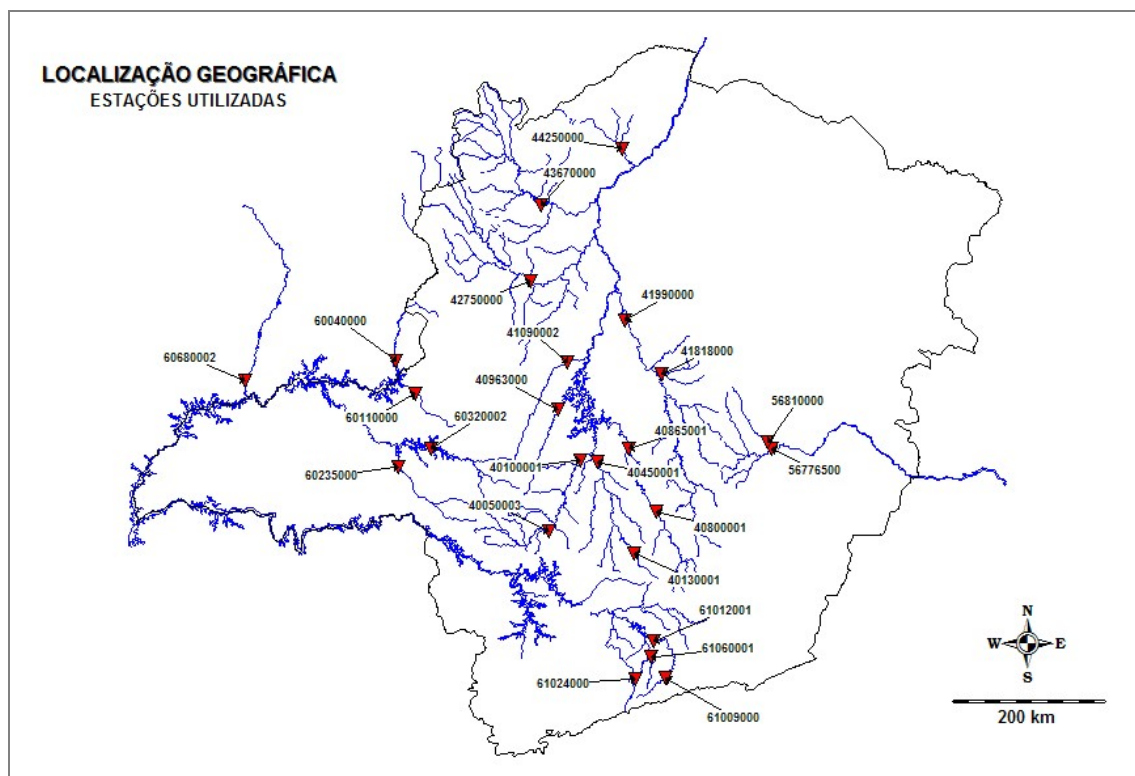


Figura 4.3 – Localização geográfica das estações utilizadas

A tabela 4.5 traz as principais características das estações consideradas. Os dados utilizados abrangem o período entre 1931 e 2002, sendo que a estação com a maior série possui 71 anos de dados, enquanto a estação com a menor série tem 25 anos de dados. As características mostradas são: T_b , o tempo de base médio da bacia; L , o comprimento do talvegue principal; S , a declividade média do talvegue principal; m , a inclinação da reta que melhor se ajusta aos pares P-V; b , o intercepto da reta P-V; e R^2 , o coeficiente de determinação da regressão P-V.

Tabela 4.5 – Principais características das estações fluviométricas utilizadas na determinação da relação pico volume para Minas Gerais

Estação	Código	T _b [dia]	Área [km ²]	L [km]	S [%]	m ¹	b ¹	R ²
40050003	IGUATAMA-MONTANTE	7	4.846	184,99	0,22	0,99 (0,2058)	6,77 (0,1223)	0,97
40100001	PORTO DAS ANDORINHAS	12	13.300	399,39	0,11	0,81 (0,1092)	7,36 (0,0543)	0,93
40130001	PONTE DO VILELA	8	1.620	73,07	0,30	1,08 (0,0726)	6,07 (0,0525)	0,95
40450001	PORTO PARA	12	11.300	238,82	0,15	1,05 (0,0535)	7,74 (0,0328)	0,98
40800001	PTE.NOVA DO PARAOPEBA	9	5.860	238,00	0,08	0,93 (0,0567)	7,00 (0,0351)	0,96
40865001	PORTO MESQUITA	10	10.300	421,05	0,06	0,89 (0,0726)	7,42 (0,0473)	0,94
40963000	PORTO INDAIA	8	2.260	239,40	0,12	1,00 (0,0577)	6,44 (0,0465)	0,95
41090002	PONTE BR040	8	5.320	260,80	0,11	0,91 (0,0805)	7,23 (0,0542)	0,93
41818000	SANTO HIPOLITO	11	16.900	587,65	0,07	1,00 (0,1809)	8,19 (0,0859)	0,91
41990000	VARZEA DA PALMA	13	26.600	742,35	0,06	0,88 (0,1028)	8,44 (0,0634)	0,90
42750000	CAATINGA	11	32.600	422,73	0,05	0,88 (0,0927)	8,26 (0,0432)	0,95
43670000	VILA URUCUIA	12	18.700	219,90	0,11	0,69 (0,0856)	7,51 (0,0376)	0,93
44250000	USINA PANDEIROS	10	3.812	111,20	0,14	1,04 (0,1916)	6,62 (0,0801)	0,88
56776500	FAZENDA OURO FINO	9	6.260	172,96	0,19	1,04 (0,0813)	7,43 (0,0542)	0,93
56810000	FAZENDA MELOSO	8	2.060	155,62	0,29	1,02 (0,0662)	6,06 (0,0357)	0,95
60040000	FAZENDA SAO DOMINGOS	10	10.659	521,13	0,03	0,94 (0,0756)	7,28 (0,0532)	0,97
60110000	ABADIA DOS DOURADOS	7	1.906	124,50	0,16	0,98 (0,1601)	6,26 (0,0892)	0,84
60235000	SANTA JULIANA	8	3.970	221,55	0,14	1,07 (0,0819)	6,89 (0,0532)	0,94
60320002	PONTE JOAO CANDIDO	8	8.490	253,05	0,08	1,00 (0,0873)	7,33 (0,054)	0,93
60680002	MEIA PONTE	10	11.400	360,20	0,17	0,93 (0,1694)	7,43 (0,085)	0,87
61009000	BOM JARDIM DE MINAS	10	509	78,24	0,52	1,06 (0,0702)	4,51 (0,045)	0,96
61012001	MADRE DE DEUS DE MINAS	8	2.230	168,59	0,05	0,90 (0,0651)	5,84 (0,0494)	0,91
61024000	AIURUOCA	7	536	55,91	1,13	0,83 (0,0501)	4,62 (0,0539)	0,89
61060001	FAZENDA LARANJEIRAS	12	2.050	150,05	0,43	0,55 (0,0607)	5,64 (0,0702)	0,85
Regressão Total						0,51	6,11	0,12
Média						0,94	6,85	0,93
Máxima						1,08	8,44	0,98
Mínima						0,55	4,51	0,84

1 - Os valores entre parêntesis refere-se ao erro padrão do parâmetro calculado

Com os dados diários de cada estação, foram selecionados hidrogramas que melhor representassem as cheias ocorridas naquele ponto. A separação do escoamento de base desses hidrogramas foi feita por meio do procedimento descrito por Linsley *et al.* (1975), o qual identifica o escoamento superficial entre o ponto de subida do hidrograma e o ponto localizado N dias após a ocorrência do pico, conforme ilustrado na figura 4.4. O valor de N é dado pela seguinte equação:

$$N = 0,8 \times A^{0,2} \quad (4.5)$$

sendo A a área de drenagem da bacia em km^2 .

Essa técnica foi utilizada para a definição do início e fim do hidrograma. Uma vez que a metodologia não necessita de separação do escoamento de base, foram utilizados os valores totais escoados.

Procurou-se estabelecer os hidrogramas decorrentes de eventos de chuvas isolados, de modo que o mesmo tivesse somente um pico. Dessa maneira, foram isolados 575 hidrogramas. O tempo de base médio das bacias (T_b) na figura 4.4 variou entre 7 e 13 dias.

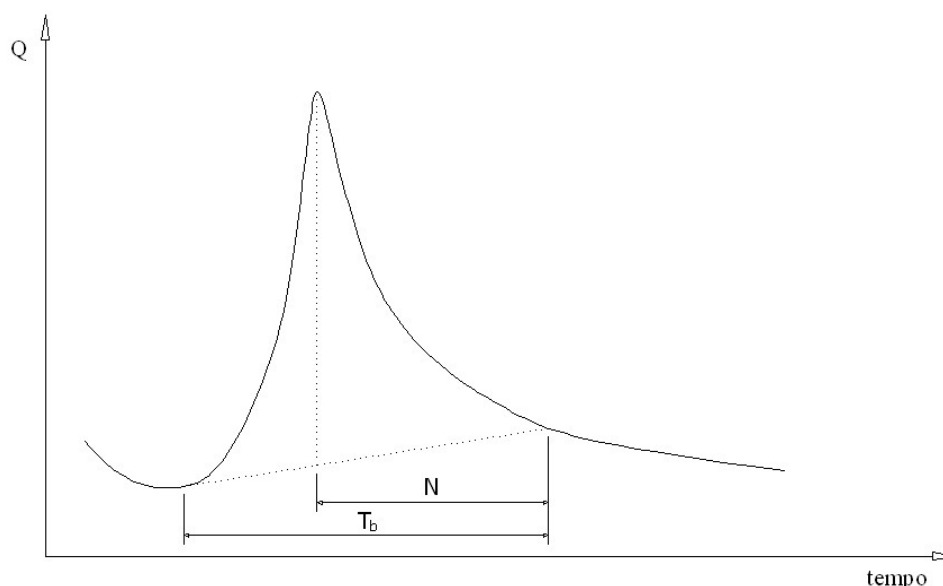


Figura 4.4 – Hidrograma de cheia: definição do tempo de base

Para cada estação, foi feita a regressão linear pelo método dos mínimos quadrados. Os resultados dessas regressões constam da tabela 4.5, precisamente nas colunas indicadas por “*m*” (inclinação) e por “*b*” (intercepto).

A figura 4.5 mostra as retas individuais encontradas para cada estação. Nota-se que, com exceção das estações 43670000 e 61060001, todas as retas seguem o mesmo comportamento, ou seja, as inclinações são aproximadamente iguais. Dessa forma, é possível definir uma inclinação única para a região ou um parâmetro regional para o modelo. O valor médio da inclinação é 0,9369, indicando que as bacias estudadas apresentam-se com características próximas àquelas de bacias lineares ($m \approx 1,0$).

Se fosse considerado apenas o volume de escoamento superficial, a exponencial do coeficiente *b*, ou seja e^b , representaria o pico do hidrograma unitário, ou seja, representaria o pico quando o volume escoado é igual a 1 *cm*. Uma vez que, de maneira geral, o volume de escoamento subterrâneo é pequeno em relação ao volume total precipitado, pode-se tomar o valor e^b como uma estimativa do pico do hidrograma unitário, mesmo sabendo-se que para a relação aqui encontrada, não foi feita a separação do escoamento. É sabido que o pico do hidrograma depende de vários fatores e varia de local para local. Sendo assim, é imediata a idéia de se encontrar uma relação regional que explique o comportamento do intercepto, de modo a particularizar a equação de regressão entre picos e volumes para o local onde se pretende calcular o pico, a partir da precipitação efetiva.

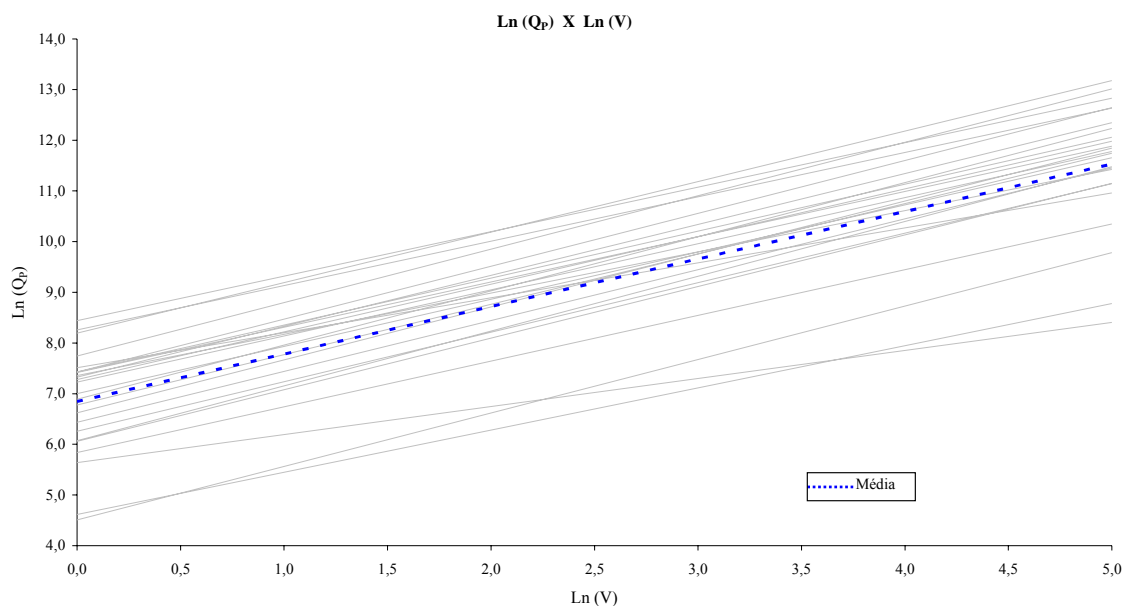


Figura 4.5 – Relação Pico-Volume para cada estação fluviométrica

Foi verificado que o intercepto b possui correlações importantes com certas características morfológicas da bacia, notadamente a área, a declividade média longitudinal do talvegue e o comprimento do curso d'água. Procedeu-se, então, à análise de regressão múltipla entre os logaritmos das variáveis b , A , S e L . A tabela 4.6 mostra a matriz dos coeficientes de correlação simples entre essas variáveis.

Tabela 4.6 – Coeficiente de correlação simples entre A , L , S e b

	A	L	S	b
A	1,00	0,88	-0,75	0,96
L	0,88	1,00	-0,81	0,84
S	-0,75	-0,81	1,00	-0,75
b	0,96	0,84	-0,75	1,00

Pela tabela 4.6, verifica-se que as variáveis A e L são altamente correlacionadas, não havendo sentido estatístico em manter ambas no modelo de regressão. Como A possui maior correlação com b , essa foi mantida na relação final. A regressão $b = f(A, S)$ apresentou um coeficiente de determinação de 0,92 enquanto a regressão $b = f(A)$ apresentou o valor de 0,91. Ou seja, houve um ganho muito pouco significativo em se manter a variável S na relação final. Portanto, pelo princípio da parcimônia estatística, o modelo usado para explicar b relacionou somente a variável A . A relação entre A e b que apresentou o melhor coeficiente de correlação ($R^2 = 0,93$) é dada pela equação 4.6 e ilustrada pela figura 4.6.

$$b = -0,7376 + 0,8877 \ln(A) \quad (4.6)$$

O modelo Pico-Volume para a região estudada do estado de Minas Gerais é obtido através da combinação das equações 3.86 e 4.6, conforme mostrado a seguir:

$$\ln(Q_p) = -0,7376 + 0,8877 \ln(A) + 0,9369 \times \ln(V) \quad (4.7)$$

ou

$$Q_p = 0,4783 \times A^{0,8877} \times V^{0,9369} \quad (4.8)$$

sendo A a área de drenagem em km^2 , V o volume escoado (deflúvio) em cm e Q_p a vazão de pico em m^3/s .

Esse modelo possui características próximas àquelas de bacias lineares, no sentido proposto por Rogers (1980).

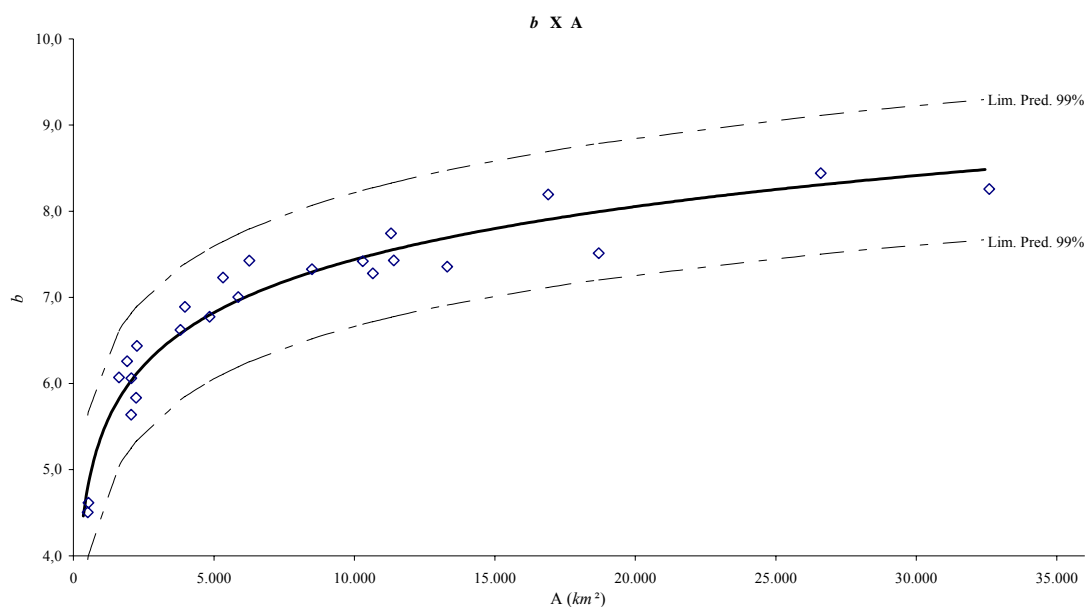


Figura 4.6 – Relação entre a área e o intercepto para a região estudada

A figura 4.7 mostra a concordância entre os valores estimados pelo modelo e os observados. Nota-se que para vazões inferiores a $30 m^3/s$ o modelo subestima as vazões observadas. No entanto, esse fato não traz maiores problemas uma vez que o foco de interesse está na cauda superior do modelo.

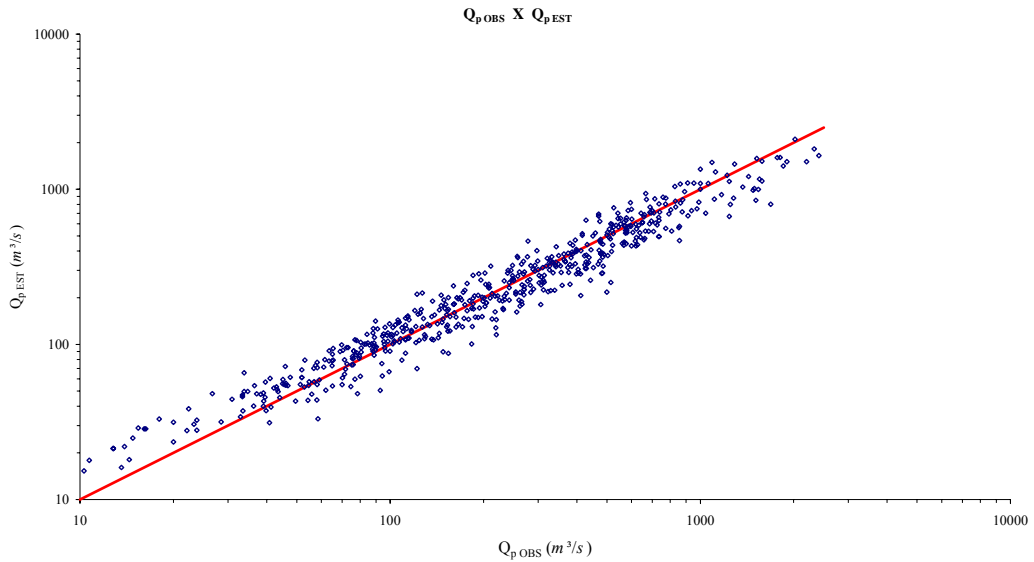


Figura 4.7 – Picos observados versus picos simulados

No intuito de verificar o modelo aqui proposto, buscou-se a reprodução das vazões de pico de hidrogramas unitários estimados para locais diferentes daqueles listados na tabela 4.5, ressaltando, entretanto, que a relação pico-volume expressa pela equação 4.8 não é perfeitamente linear. Serebrenick & Carvalho (1966) estabeleceram os hidrogramas unitários naturais para o rio Pomba em Cataguases, Minas Gerais, segundo a proposta clássica de Sherman (1932). A figura 4.8 mostra os 9 hidrogramas unitários estimados pelos autores a partir de hidrogramas de cheia observados no período de 1930 a 1940.

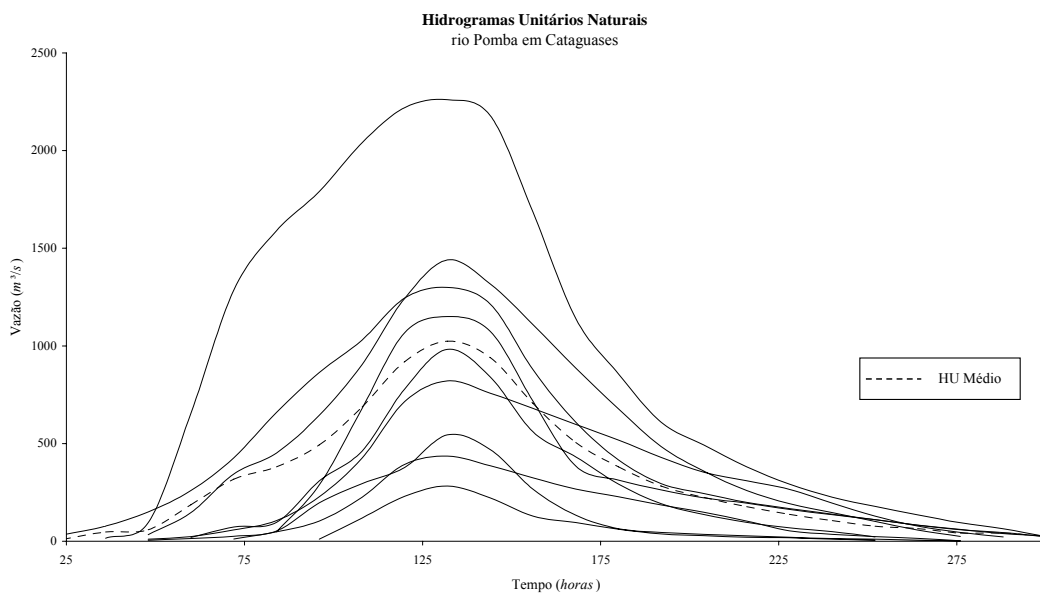


Figura 4.8 – Hidrogramas unitários para o rio Pomba em Cataguases

O hidrograma médio mostrado na figura 4.8 foi obtido a partir da média dos valores de cada hidrograma. O pico do hidrograma médio é $1024 \text{ m}^3/\text{s}$. A área de drenagem da bacia do rio Pomba em Cataguases é de 5910 km^2 . Portanto, para um volume de escoamento unitário ($1,0 \text{ cm}$) a equação 4.8 fornece $Q_P = 1066 \text{ m}^3/\text{s}$.

O pico encontrado utilizando o modelo aqui descrito difere em 4% do pico encontrado a partir das médias dos hidrogramas unitários estimados por Serebrenick e Carvalho (1966), demonstrando, ainda que de forma apenas parcial, a adequação dos resultados encontrados. Ressalte-se ainda a oportunidade de aplicação do modelo P-V em locais desprovidos de dados fluviométricos, bem como o relaxamento da condição de estrita linearidade.

Os resultados encontrados para a inclinação m regional são compatíveis com aqueles encontrados por Mimikou (1983) para bacias da Grécia e por Rogers (1980) para bacias dos Estados Unidos.

A vazão de pico pode ser determinada pela chuva efetiva e por uma única característica morfológica da bacia, qual seja, a área de drenagem A . Uma característica importante da relação pico-volume encontrada é seu caráter regional. Pelo fato de a equação ser originada a partir de dados de diferentes locais com diferentes características, sua aplicação é possível em outros locais com ausência de dados. Além disso, a incorporação de um grande número de eventos pico-volume, em caráter regional, certamente produz melhores estimativas do que aqueles obtidos somente com dados locais.

A característica regional é resguardada pelo parâmetro b . Tal parâmetro é um indicativo da vazão de pico do hidrograma unitário. Mimikou (1983) encontrou uma boa correlação ao relacionar b com o índice AS/L . No entanto, a inclusão de novas variáveis sem uma definição clara das diversas interdependências entre as mesmas, além da possível presença de correlação espúria, pode resultar eventualmente em maior erro padrão da estimativa.

4.4.2 Relação pico-volume para a bacia do rio Pará, em Ponte do Vilela

Em bacias para as quais se dispõe de registros fluviométricos suficientemente longos, pode-se utilizar os pares pico-volume observados para estimar o modelo que explica a variância das vazões máximas diárias a partir dos volumes de cheia.

A relação entre o pico e o volume de cheia é complexa e depende de vários fatores, incluindo a atenuação do pico devido à inundação do leito maior da bacia e características geomorfológicas particulares de cada local. Nesse caso, a relação pico-volume pode ser estimada através de modelos não paramétricos. Bradley & Potter (1992) usaram uma técnica de regressão chamada LOWESS (*robust locally weighted regression*) para estimar a relação entre os picos e os volumes de três dias de duração para uma bacia norte-americana, obtendo bons resultados. Em muitos casos, a técnica LOWESS, desenvolvida por Cleveland (1979), parece ser adequada para a modelagem pico-volume. A técnica LOWESS é descrita com detalhes no Anexo 3.

A relação pico-volume para a bacia do rio Pará foi estimada em duas etapas: a primeira, relacionando os dados observados, e a seguinte, extrapolando a relação através do modelo regional descrito no item 4.4.1.

A partir dos dados observados, mostrados na tabela 4.1, ajustou-se um modelo não paramétrico utilizando a técnica LOWESS conforme descrito no Anexo 3. Para valores superiores aos observados, foi utilizada a equação regional dada pela equação 4.8 para extrapolar a curva pico-volume do rio Pará em Ponte do Vilela. A figura 4.9 mostra a relação encontrada, incluindo os valores observados e os extrapolados.

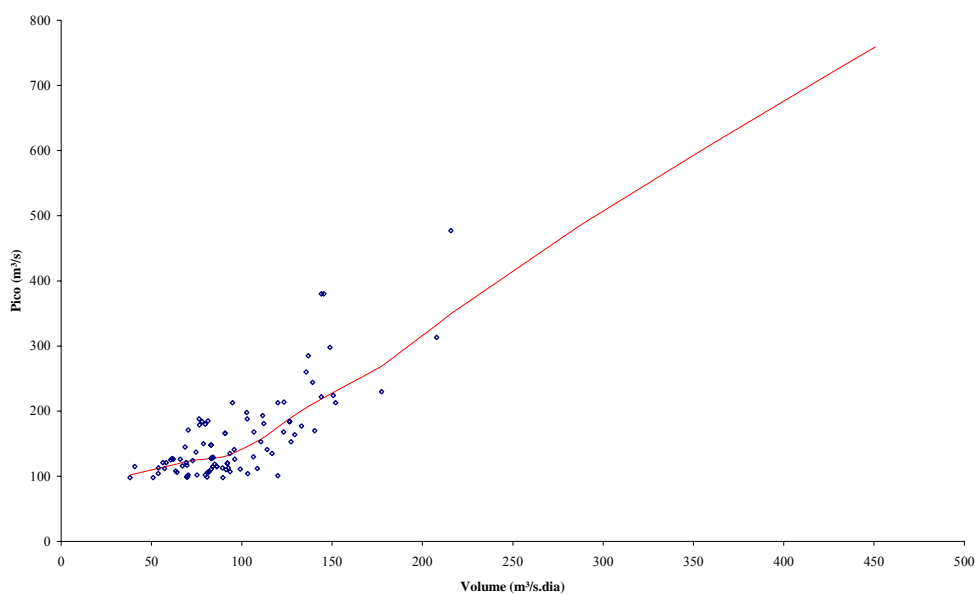


Figura 4.9 – Relação Pico-Volume para a bacia do rio Pará em Ponte do Vilela

O erro padrão da regressão (σ_ϵ) mostrada acima é fortemente afetado pela presença de *outliers*. Assim, o uso de todos os dados para o cálculo do erro pode superestimar essa

estatística. A tabela 4.7 apresenta os quantis calculados pelo método PVP considerando vários valores de σ_ε . Verifica-se que o método PVP é bastante sensível à estimativa do erro padrão. Para o período de retorno de 2 anos, há uma variação de 63% nos quantis calculados, considerando o erro padrão variando de 5 a 50. Nessa mesma faixa de variação do erro padrão, há um desvio de 12% para os quantis de 10.000 anos de período de retorno. Portanto, o erro padrão deve ser avaliado de forma cuidadosa, para que não produza estimativas tendenciosas para os quantis de vazão.

Tabela 4.7 – Quantis de vazão para diversos valores do erro padrão

T (anos)	Quantil (m³/s)										Média	Desvio ¹
	$\sigma_\varepsilon = 5$	$\sigma_\varepsilon = 10$	$\sigma_\varepsilon = 15$	$\sigma_\varepsilon = 20$	$\sigma_\varepsilon = 25$	$\sigma_\varepsilon = 30$	$\sigma_\varepsilon = 35$	$\sigma_\varepsilon = 40$	$\sigma_\varepsilon = 45$	$\sigma_\varepsilon = 50$		
2	196,46	214,9	229,53	232,04	242,95	263,13	279,89	300,12	310,54	320,77	259,03	63%
5	263,71	284,75	302,75	319,96	325,59	336,14	351,61	371,59	392,04	413,75	336,19	57%
10	309,86	328,55	348,92	368,35	380,21	388,55	400,86	417,11	435,6	454,19	383,22	47%
50	465,99	485,51	505,42	524,46	544,09	563,86	579,33	585,42	593,13	602,76	545,00	29%
75	533,66	552,26	572,62	591,99	611,07	631,11	650,05	666,05	671,58	676,83	615,72	27%
100	587,59	606,19	625,9	644,96	664,61	682,95	703,38	722,44	735,71	741,31	671,50	26%
200	719,04	737,63	756,49	776,47	795	815,38	834,92	854,38	873,43	893,18	805,59	24%
300	795,96	815,17	834,06	853,52	872,65	892,92	912,89	931,24	951,12	971,77	883,13	22%
400	849,58	868,72	888,14	907,09	926,76	947,52	965,73	986,47	1.005,25	1.025,72	937,10	21%
500	893,05	911,57	931,06	949,68	969,15	989,17	1.008,30	1.028,12	1.047,52	1.067,90	979,55	20%
1.000	1.021,54	1.040,91	1.059,63	1.078,86	1.098,92	1.118,61	1.138,67	1.158,43	1.176,82	1.196,63	1108,90	17%
2.000	1.150,10	1.169,03	1.187,98	1.208,61	1.226,54	1.247,57	1.267,07	1.286,52	1.305,66	1.325,11	1237,42	15%
3.000	1.225,27	1.243,48	1.262,36	1.282,31	1.302,01	1.321,70	1.341,26	1.361,23	1.380,27	1.399,92	1311,98	14%
4.000	1.277,35	1.296,58	1.315,75	1.336,06	1.354,26	1.374,31	1.394,25	1.414,32	1.433,88	1.453,66	1365,04	14%
5.000	1.318,77	1.337,48	1.357,14	1.376,40	1.396,39	1.414,71	1.434,81	1.455,04	1.474,75	1.493,73	1405,92	13%
6.000	1.351,55	1.370,28	1.389,78	1.409,01	1.428,85	1.448,84	1.468,25	1.487,32	1.507,42	1.526,59	1438,79	13%
7.000	1.379,30	1.399,20	1.417,52	1.437,80	1.457,25	1.475,98	1.496,21	1.516,25	1.534,95	1.555,22	1466,97	13%
8.000	1.403,17	1.423,75	1.441,96	1.461,89	1.480,75	1.500,30	1.520,16	1.540,16	1.559,90	1.579,82	1491,19	13%
9.000	1.425,32	1.444,43	1.464,06	1.483,64	1.503,16	1.522,07	1.542,18	1.560,65	1.580,54	1.600,71	1512,68	12%
10.000	1.443,60	1.463,14	1.483,92	1.502,61	1.521,06	1.543,83	1.560,14	1.580,75	1.600,74	1.621,01	1532,08	12%

1 - Desvio calculado em relação aos quantis para os erros padrão iguais a 5 e 50

A tabela 4.8 apresenta os erros médios quadráticos da regressão mostrada na figura 4.9. Uma técnica freqüente para identificar valores atípicos de uma amostra faz uso do critério da amplitude inter-quartis (AIQ). Segundo tal técnica, é considerado um ponto atípico superior todo elemento da amostra superior a:

$$LS = Q_3 + 1,5AIQ \quad (4.9)$$

e, analogamente, um ponto atípico inferior é todo e qualquer elemento menor do que:

$$LI = Q_1 - 1,5AIQ \quad (4.10)$$

sendo que Q_1 e Q_3 representam o primeiro e o terceiro quartil, respectivamente, e AIQ é a amplitude inter-quartis dada pela diferença entre Q_1 e Q_3 .

A partir dos dados da tabela 4.8 tem-se:

$$Q_1 = 134,49$$

$$Q_3 = 1412,02$$

$$AIQ = 1277,53$$

Os limites superior e inferior, dados pelas equações 4.9 e 4.10, respectivamente, são:

$$LI = -1781,80$$

$$LS = 3328,31$$

Tabela 4.8 – Erros quadráticos da relação pico-volume

Volume	Erro	Volume	Erro	Volume	Erro	Volume	Erro
53,75	0,01	93,35	137,68	81,29	398,04	140,38	1495,93
84,01	0,06	85,00	145,76	82,94	426,53	102,70	1661,56
83,03	0,15	40,77	149,18	82,94	439,08	93,49	1714,80
84,17	0,55	58,13	185,64	89,26	461,65	127,25	1883,49
69,58	1,60	83,89	189,73	92,06	475,24	99,13	1960,81
83,16	1,68	65,89	196,92	79,79	509,72	116,75	2075,45
106,70	3,60	56,35	229,80	92,06	547,23	70,36	2725,95
63,40	5,97	69,36	246,03	133,00	566,00	79,79	3025,46
67,14	10,72	70,35	256,35	123,29	590,03	76,50	3166,34
38,05	16,23	62,04	266,46	78,70	674,48	108,59	3189,63
72,81	19,37	74,64	274,76	80,75	711,25	135,60	3270,31
112,16	26,59	86,15	277,80	90,78	740,36	103,24	3303,61
64,11	27,10	60,59	284,18	96,00	751,30	81,30	3454,90
57,25	29,07	83,04	287,57	90,78	820,16	77,97	3671,22
50,87	30,72	69,75	298,88	103,00	823,27	94,84	3958,12
150,63	37,68	110,48	321,57	119,94	907,01	76,33	4360,61
151,88	45,47	69,94	321,98	91,26	913,21	136,80	6426,26
69,21	45,78	86,15	325,84	68,56	993,49	148,76	6717,59
53,88	62,82	61,32	327,27	92,61	1023,11	119,94	7083,52
126,40	63,38	75,14	370,47	106,39	1137,73	207,88	8031,66
177,38	70,63	123,15	378,06	129,25	1197,07	145,38	27484,14
126,40	104,04	81,29	381,06	114,00	1383,86	144,06	28093,23
95,73	115,42	81,91	383,18	139,20	1385,18	215,75	63420,92
144,05	131,30	111,56	384,25	89,50	1438,86		

O erro padrão da relação pico-volume pode ser estimado por:

$$\sigma_{\varepsilon} = \sqrt{\frac{\sum_{i=1}^n (Y_i^{obs} - Y_i^{est})^2}{n}} \quad (4.11)$$

Considerando os limites do critério da AIQ tem-se $\sigma_{\varepsilon} = 25,9$.

Uma forma de avaliar a relação encontrada sob o ponto de vista da adequação do modelo às características da estação Ponte do Vilela, sobretudo na região das extrapolações, é a

utilização de séries sintéticas longas. Nesse sentido, a relação aqui encontrada foi comparada aos dados pico-volume de séries sintéticas encontradas na região.

Lima (2004) estabeleceu um modelo estocástico de chuva que permite a obtenção de longas séries de precipitação simuladas, com as mesmas características estatísticas da série observada. Em seu trabalho, Lima (2004) gerou uma série pluviométrica para a bacia do rio Pará com 10.000 anos de dados diários. Posteriormente, Lima (2004) utilizou o modelo de transformação chuva-vazão RIO GRANDE para obter a série de vazões médias diárias para um período de 10.000 anos.

A partir da série de vazões simulada por Lima (2004), foram obtidos 10.000 pares pico-volume. Foram considerados somente os picos máximos anuais e o volume correspondente de 8 dias de duração. Obteve-se, então, um modelo de regressão para esses pares na forma da equação 3.86.

A figura 4.10 mostra os pares pico-volume simulados e a figura 4.11 mostra a regressão obtida a partir desses dados em comparação à relação regional obtida no item 4.4.1.

O modelos mostrados na figura 4.11 são:

$$Q_p^{SIM} = 315,29 \cdot \left(\frac{8,64 \cdot Vol}{1620} \right)^{0,9946} = 1,7297 \cdot Vol^{0,9946} \quad (4.12)$$

$$Q_p^{REG} = 337,90 \cdot \left(\frac{8,64 \cdot Vol}{1620} \right)^{0,9396} = 2,4722 \cdot Vol^{0,9396} \quad (4.13)$$

sendo Vol o volume em $m^3/s.dia$ e Q_p a vazão de pico em m^3/s .

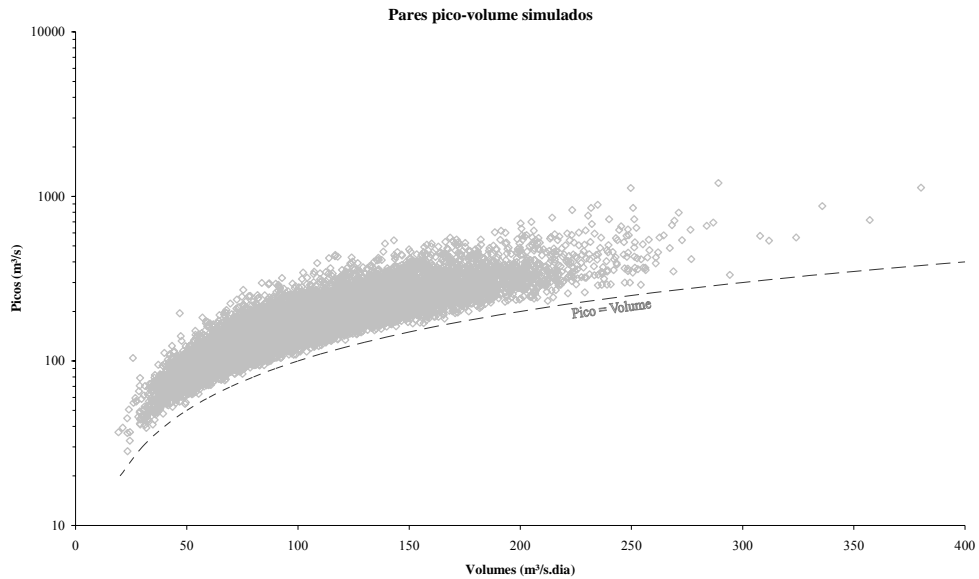


Figura 4.10 – Pares pico-volume simulados

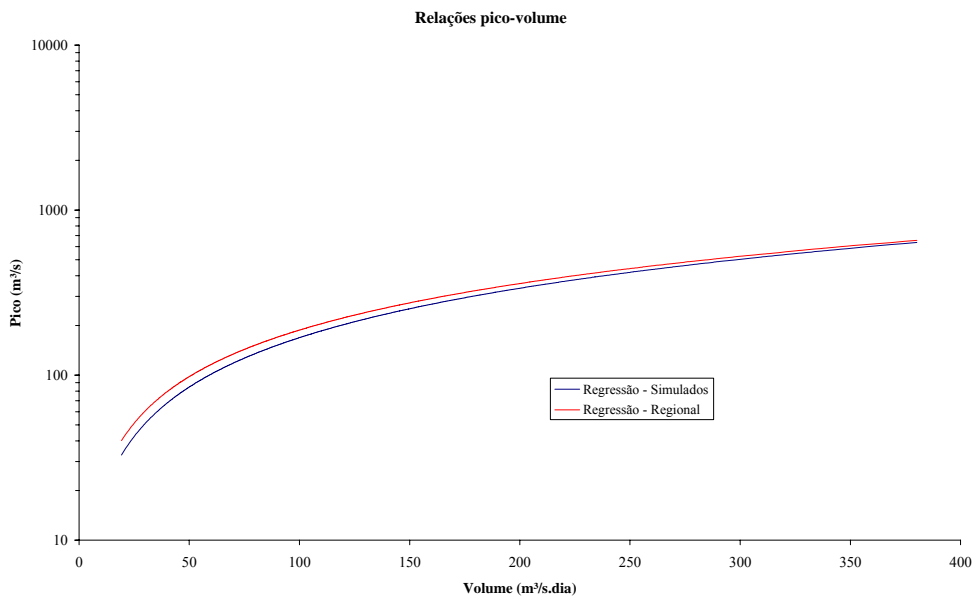


Figura 4.11 – Modelos pico-volume para a bacia do rio Pará

A grande semelhança entre o modelo regional e o modelo obtido por simulação, notadamente para volumes maiores, evidencia a aplicabilidade do modelo regional para extrapolar os pares pico-volume para a bacia do rio Pará.

Outro fato a ser observado diz respeito à linearidade da bacia. O coeficiente que relaciona os volumes aos picos é bastante próximo de 1, demonstrando um comportamento hidrológico aproximadamente linear.

4.5 Regionalização dos dados pluviométricos pela TCEV

4.5.1 Introdução

Conforme visto no capítulo 3, a aplicação do método GRADEX depende da definição do parâmetro de escala da distribuição de probabilidades da chuva. Com o objetivo de dar uma maior estrutura ao método e agregar mais informação ao resultado final da aplicação da metodologia PVP, o parâmetro de escala da distribuição da chuva foi definido por meio de um estudo regional.

Dentre os métodos de estimação dos parâmetros regionais da TCEV, discutidos no tópico 4.3, aquele realizado por momentos-L foi o que se mostrou mais viável para a aplicação pretendida. Os demais métodos foram testados e rejeitados por não convergirem para um resultado satisfatório. Possivelmente, isso ocorreu devido à estrutura geral de cada método e das características das séries utilizadas.

Em geral, as séries de duração parcial apresentam uma assimetria relativamente baixa, o que faz os métodos de estimação baseados nas equações de máximo de verossimilhança falharem na busca dos parâmetros ótimos. Não é necessário que todas as séries tenham baixa assimetria para a não convergência dos métodos; de fato, a presença de apenas uma, com essas características, já é suficiente.

Possivelmente, dentre as 11 séries utilizadas, pode ter havido alguma com características fortes de uma única componente ($\theta_2/\theta_1 \approx 1$), o que fez os métodos de busca falharem. No entanto, a identificação dessa característica, *a priori*, é difícil e não há métodos robustos para tal.

Assim, a regionalização por momentos-L se mostrou a melhor alternativa para a estimação dos parâmetros regionais. A regionalização baseada em momentos-L é pouco influenciada pela assimetria das séries e não encontra problemas quando, dentre todas as estações, houver alguma com características marcadamente de uma única componente.

4.5.2 Consistência dos dados e definição da região homogênea

A regionalização dos dados pluviométricos pela TCEV utilizando momentos-L deve ser precedida por uma análise da consistência dos dados e da homogeneidade da região. Dessa

maneira procedeu-se ao cálculo da medida de discordância, de acordo com a equação 2.59, e da medida de heterogeneidade, de acordo com a equação 2.61.

A tabela 4.9 traz os quocientes de momentos-L para cada estação, bem como a medida de discordância (Di).

Tabela 4.9 – Momentos-L e medida de discordância para a região estudada

Código	N _{dados}	CV-L	Assimetria-L	Curtose-L	Di
02044000	43	0,173	0,274	0,152	1,71
02044002	79	0,138	0,412	0,306	1,58
02044003	96	0,145	0,338	0,223	0,17
02044005	93	0,136	0,281	0,150	0,93
02044006	85	0,151	0,259	0,157	0,25
02044009	97	0,144	0,225	0,177	0,96
02044011	12	0,141	0,329	0,318	2,21
02044016	51	0,147	0,346	0,207	0,45
02044027	54	0,161	0,324	0,214	0,52
02044036	94	0,128	0,212	0,140	1,65
02044042	42	0,155	0,325	0,173	0,57

O valor crítico para Di, numa região onde há 11 postos pluviométricos, é, segundo Hosking & Wallis (1997), 2,632. Como todos Di foram inferiores ao valor crítico, não há nenhuma estação discordante do comportamento médio esperado.

O cálculo da medida de heterogeneidade, conforme estabelecido no tópico 2.4.2.2, é feito pela comparação das variâncias entre os CV-L observado e simulado. Hosking & Wallis (1997) recomendam o uso da distribuição Kapa para a simulação da região. Os parâmetros da distribuição Kapa são calculados pelo método dos momentos-L utilizando os quocientes de momentos-L regionais ou as médias ponderadas dos quocientes locais. A distribuição de probabilidades acumulada da distribuição Kapa é dada por:

$$F(x) = \left\{ 1 - h \left[1 - \frac{k(x - \xi)}{\alpha} \right]^{1/k} \right\}^{1/h} \quad (4.13)$$

Os valores da tabela 4.9 permitem o cálculo dos quocientes-L regionais por meio da média ponderada tomando o tamanho da série como peso, conforme originalmente feito por Hosking & Wallis (1997). Assim, têm-se:

$$\text{CV-L} \quad \frac{\lambda_2}{\lambda_1} = \tau_2^R = 0,1449$$

$$\text{Assimetria-L} \quad \frac{\lambda_3}{\lambda_2} = \tau_3^R = 0,2942$$

$$\text{Curtose-L} \quad \frac{\lambda_4}{\lambda_2} = \tau_4^R = 0,1917$$

Os parâmetros da distribuição Kapa podem ser calculados pela comparação desses momentos-L regionais amostrais com os populacionais, quais sejam:

$$\lambda_1 = \xi + \alpha(1 - g_1)/k \quad (4.14)$$

$$\lambda_2 = \alpha(g_1 - g_2)/k \quad (4.15)$$

$$\tau_3 = (-g_1 + 3g_2 - 2g_3)/(g_1 - g_2) \quad (4.16)$$

$$\tau_4 = (-g_1 + 6g_2 - 10g_3 + 5g_4)/(g_1 - g_2) \quad (4.17)$$

sendo,

$$g_r = \begin{cases} \frac{r\Gamma(1+k)\Gamma(r/k)}{h^{1+k}\Gamma(1+k+r/h)}, & h > 0 \\ \frac{r\Gamma(1+k)\Gamma(-k-r/h)}{(-h)^{1+k}\Gamma(1-r/h)}, & h < 0 \end{cases} \quad (4.18)$$

A solução do sistema formado pelas equações 4.14 a 4.17 forneceu os seguintes valores para os parâmetros regionais da distribuição Kapa:

$$\xi = 0,8301$$

$$\alpha = 0,2000$$

$$k = -0,1215$$

$$h = 0,2955$$

Foram simuladas 500 populações com esses parâmetros, o que forneceu uma medida de heterogeneidade de -1,28. Conseqüentemente:

$$H = -1,28 < 1 \rightarrow \text{região "aceitavelmente homogênea"}$$

O valor negativo de H indica que há menos dispersão entre os valores amostrais de CV-L do que se esperaria de uma região homogênea com *distribuições individuais de frequência independentes*. A causa mais provável para esse valor negativo é a presença de correlação significativa entre os dados dos diferentes postos. Hosking & Wallis (1997) preconizam que se valores muito negativos, como $H < -2$, são observados durante a regionalização, isso pode ser uma indicação de que há muita correlação cruzada entre as distribuições individuais de frequência ou de que há uma regularidade excessiva dos valores amostrais de CV-L. Como o H calculado é maior que -2, aceita-se a hipótese de que se trata de uma região homogênea.

4.5.3 Cálculo dos parâmetros regionais e locais da TCEV

O cálculo dos parâmetros regionais da TCEV é feito por meio da solução do sistema formado pelas equações 3.84 e 3.85. A solução do sistema é complexa, uma vez que envolve o somatório de várias funções gama, com o parâmetro a ser determinado como argumento. Os métodos convencionais de solução de sistemas não lineares por meio de derivadas (e.g. método de Newton) falham por exigir boa precisão para o valor inicial dos parâmetros e por serem excessivamente complexos do ponto de vista computacional. O fato da função-objetivo, estabelecida por meio de derivadas, não ter variações muito acentuadas pode levar os métodos convencionais a se prenderem em máximos locais, ou não convergirem para uma solução adequada. Assim, foi aplicado um algoritmo do tipo genético para a solução do sistema utilizando como funções-objetivo as equações 3.84 e 3.85, em detrimento de suas derivadas.

O método SCE-UA (*Shuffled Complex Evolution*) foi inicialmente proposto por Duan (1991), sendo uma variante de otimização por algoritmos genéticos no contexto da busca dos parâmetros ótimos de modelos hidrológicos. Trata-se de um método de calibração global que incorpora, ao processo de busca, uma reprodução do processo de evolução natural (seleção natural), de forma a aumentar sua efetividade. Neste sentido, o método é considerado um algoritmo genético. Assim, a partir de uma população inicial de parâmetros (indivíduos) e das probabilidades de sobrevivência, eliminação, mutação, cruzamento e duplicação de indivíduos, é feito um mapeamento da função-objetivo dentro deste espaço paramétrico

multidimensional. A geração de indivíduos, bem como as mutações, cruzamentos e duplicações, ocorrem de forma aleatória. Desta forma, é avaliado um maior número de parâmetros, tornando o resultado mais confiável. Embora mais lento que os métodos convencionais, o algoritmo genético fornece melhores resultados uma vez que não se prende a máximos locais. Além disso, a otimização por algoritmo genético é bem menos sensível aos valores iniciais dos parâmetros, tornando-o mais robusto que os demais métodos.

A figura 4.12 apresenta o gráfico da variação da função objetivo, tal como utilizada no algoritmo genético.

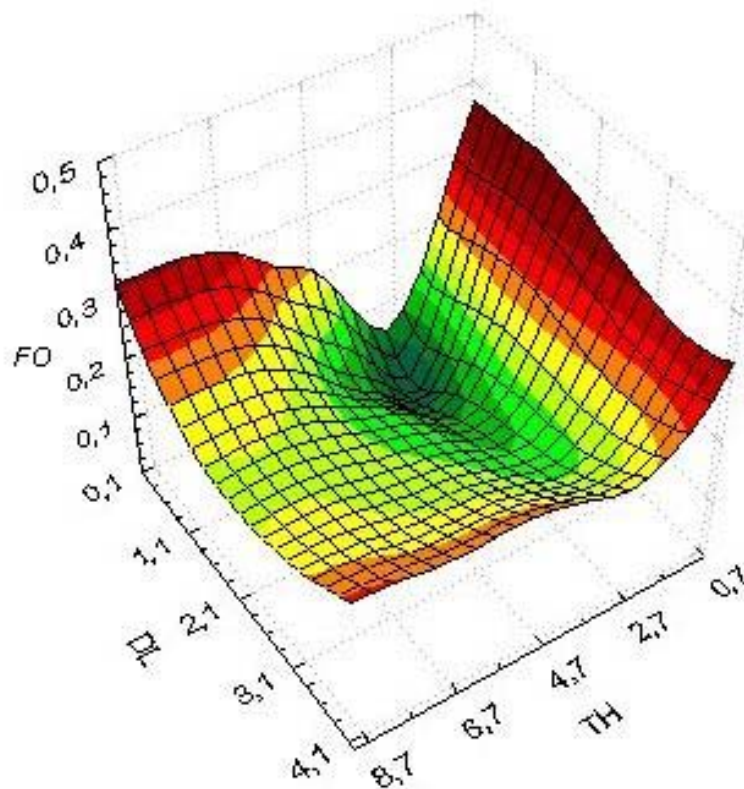


Figura 4.12 – Função objetivo do algoritmo genético: *DL* é o parâmetro Λ^* , *TH* é o parâmetro θ^* e *FO* é o valor da função objetivo

Como resultados da aplicação do método de otimização anteriormente descrito, os valores obtidos para os parâmetros foram:

$$\theta_* = 2,810$$

$$\Lambda_* = 0,801$$

Os parâmetros locais da TCEV são obtidos por meio dos momentos-L da série de duração parcial da chuva média sobre a bacia e dos parâmetros regionais determinados anteriormente. As equações 3.76 e 3.77 podem ser reescritas de modo a fornecer os parâmetros locais da TCEV, quais sejam:

$$\theta_1 = \frac{\lambda_2}{\ln 2 + T_1 - T_0} \quad (4.19)$$

$$\Lambda_1 = \exp\left(\frac{\lambda_1}{\theta_1} - 0,577 - T_0\right) \quad (4.20)$$

sendo que λ_i é o momento-L de ordem i da série parcial da chuva média e T_i é dado pela equação 3.73.

Uma vez que T_i depende somente dos parâmetros regionais da TCEV, os parâmetros locais podem ser determinados diretamente. Assim, para a bacia do rio Pará em Ponte do Vilela tem-se:

$$\theta_1 = 17,985$$

$$\theta_2 = \theta_* \cdot \theta_1 = 50,547$$

$$\Lambda_1 = 4.674,127$$

$$\Lambda_2 = \Lambda_* \cdot \Lambda_1^{1/\theta_*} = 16,203$$

A figura 4.13 demonstra a boa aderência da TCEV às alturas de chuva média sobre a bacia, para a duração de 8 dias.

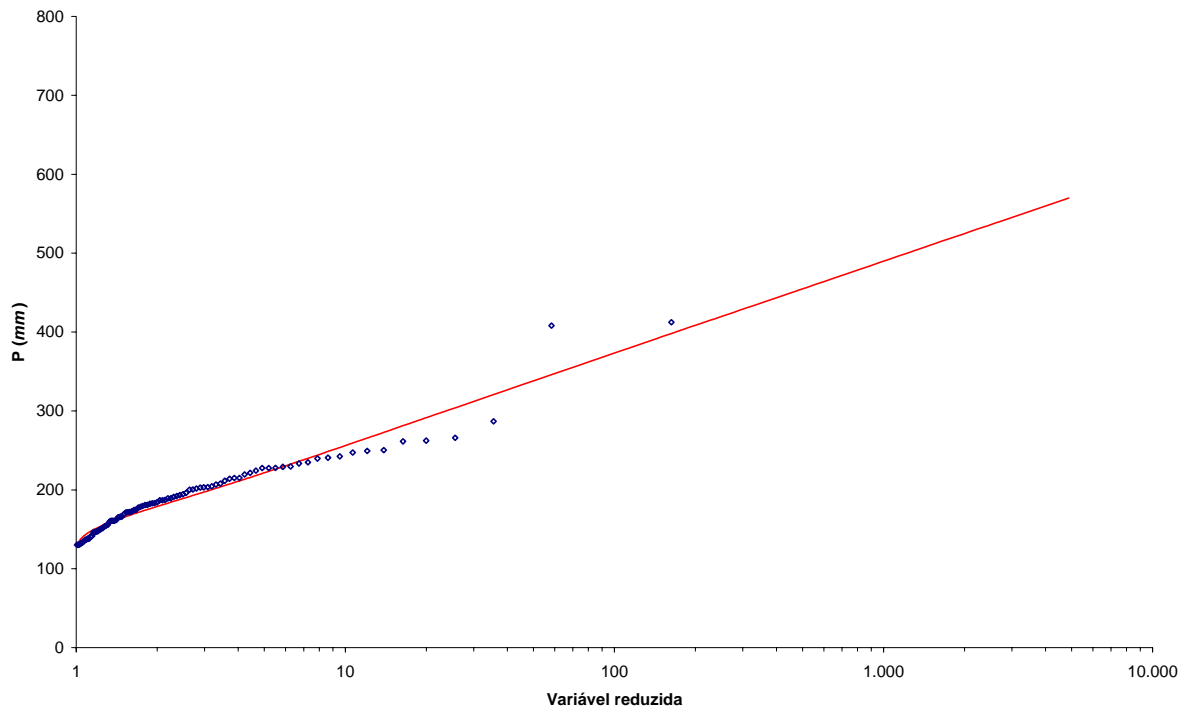


Figura 4.13 – Ajuste da TCEV à chuva média sobre a bacia do rio Pará em Ponte do Vilela

4.6 A distribuição dos volumes escoados pelo método GRADEX

A distribuição dos volumes foi estabelecida de acordo com a metodologia apresentada no item 3.6.3. Inicialmente, é necessário converter o gradex da chuva (θ_2) para as mesmas unidades em que estão expressos os volumes de cheia. Uma vez que o gradex da chuva está expresso em *mm* e os volumes de cheia estão expressos em $m^3/s.dia$, a seguinte transformação é necessária:

$$gradex (volume) = \frac{Ad}{86,4 \cdot d} \cdot gradex (chuva) \quad (4.21)$$

na qual,

Ad é a área de drenagem em km^2 ; e

d é a duração em *dias*.

Assim,

$$\text{gradex}(\text{volume}) = \frac{1620}{86,4 \cdot 8} \cdot 50,547 = 118,470 \text{ m}^3 / \text{s.dia}$$

A distribuição dos volumes foi obtida pelo ajuste da TCEV aos dados observados levando-se em consideração o gradex calculado acima. Fez-se o ajuste de cada parte da TCEV em momentos separados. Para a cauda inferior da distribuição foi utilizado o modelo de Gumbel com os parâmetros calculados pelo método dos momentos convencionais. As estimativas são:

$$\alpha_1^{FLU} = \sqrt{\frac{\sigma_{FLU}^2}{1,645}} = \sqrt{\frac{1086,786}{1,645}} = 25,703$$

$$\beta_1^{FLU} = \mu_{FLU} - 0,577 \cdot \alpha_1^{FLU} = 95,784 - 0,577 \cdot 25,703 = 80,913$$

O desvio padrão e a média foram calculados com base nos dados apresentados na tabela 4.1.

Para o ajuste da cauda superior é necessário conhecer T_c , ou seja, o tempo de retorno a partir do qual a bacia é considerada saturada e as suposições do método GRADEX são tidas como válidas. Conforme visto no capítulo 3, para bacias onde não há uma evidência forte do ponto de saturação e da chuva que provoca essa saturação, T_c deve ser definido com base nos dados observados de volumes de cheia utilizando alguma medida de tendência central das estatísticas de ordem superior da série. Esse é o caso da bacia do rio Pará, onde não foi encontrado nenhum estudo que quantifique, de forma precisa, o ponto de saturação e a chuva necessária para alcançar tal ponto.

Para a aplicação em questão, fez-se a suposição de que os 5 maiores valores da série de duração parcial dos volumes contemplam a condição de saturação da bacia. Assim, espera-se que, durante os 63 anos de monitoramento de vazão, a condição de saturação tenha sido alcançada pelo menos 5 vezes e que tal fato esteja refletido nos 5 maiores volumes escoados, de 8 dias de duração.

O quantil referente ao ponto de saturação foi estimado, de modo supostamente robusto, como a mediana dos 5 maiores valores da série dos volumes. Assim,

$$X(T_c) = 177,38 \text{ m}^3 / \text{s.dia}$$

O período de retorno T_c pode ser encontrado a partir dos parâmetros da cauda inferior já que a mesma será prolongada até este valor. Com a distribuição de Gumbel, têm-se:

$$F = \exp\left[-\exp\left(-\frac{X - \beta_1}{\alpha_1}\right)\right] = 0,977$$

e

$$T_c = \frac{1}{1 - F} \approx 43 \text{ anos}$$

Com os valores assim estimados, pode-se calcular o parâmetro β_2 , por meio da equação 3.99, ou seja:

$$\beta_2 = -266,91$$

Os parâmetros da TCEV podem, então, ser determinados pelas relações 3.92 e 3.93. As estimativas são:

$$\Lambda_1 = \exp\left(\frac{\beta_1}{\alpha_1}\right) = 23,289$$

$$\Lambda_2 = \exp\left(\frac{\beta_2}{\alpha_2}\right) = 0,105$$

$$\theta_1 = \alpha_1 = 25,703$$

$$\theta_2 = \alpha_2 = 118,470$$

A figura 4.14 mostra a distribuição dos volumes de cheia de 8 dias de duração. Também é mostrada, nessa figura, a cauda superior da distribuição da chuva.

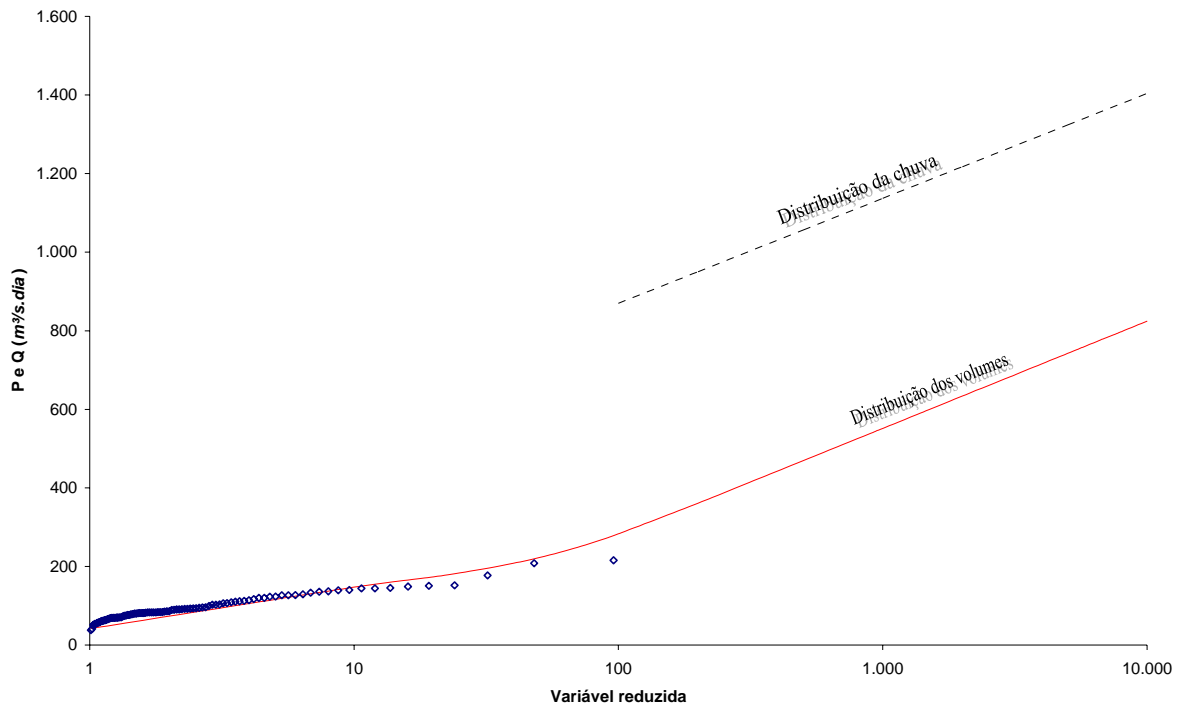


Figura 4.14 – Distribuição dos volumes de cheia e de precipitação para a bacia do rio Pará

4.7 A distribuição dos picos de vazão pelo método PVP

A distribuição dos picos condicionados aos volumes e a distribuição dos volumes foram combinadas, por meio das equações 3.89 e 3.90, para se obter a distribuição anual das vazões médias diárias máximas ou de pico. A integral da equação 3.89 não pode ser resolvida analiticamente. Por isso, utilizou-se o método de integração de Romberg para obter seu valor numérico. O método de Romberg pode ser visto como uma generalização da regra de Simpson para cálculos numéricos de integrais.

A distribuição dos picos condicionados aos volumes $F_{Y|X}(y|x)$ é definida somente para valores superiores ao *threshold* da série parcial de vazão ($u = 96,6 \text{ m}^3/\text{s}$) e isso faz com que a função não seja suave nas proximidades desse valor. A figura 4.15 mostra a distribuição $F_{Y|X}(y|x)$ para vários valores de Y_p . Observa-se que a forma da distribuição é a mesma para valores de Y_p superiores a u ($Y_p > 150 \text{ m}^3/\text{s}$), diferenciando apenas de uma translação a partir do eixo da distribuição. A distância de translação entre cada distribuição é proporcional a Y_p .

Para valores de Y_p superiores a u e inferiores a $150 \text{ m}^3/\text{s}$ a distribuição dos picos condicionada aos volumes apresenta dificuldades, de ordem numérica, para ser avaliada ao longo de seu

domínio entre 0 e 1, com comportamento bastante diferente do esperado. Na figura 4.15, o comportamento esperado da função é mostrado por meio de linhas tracejadas.

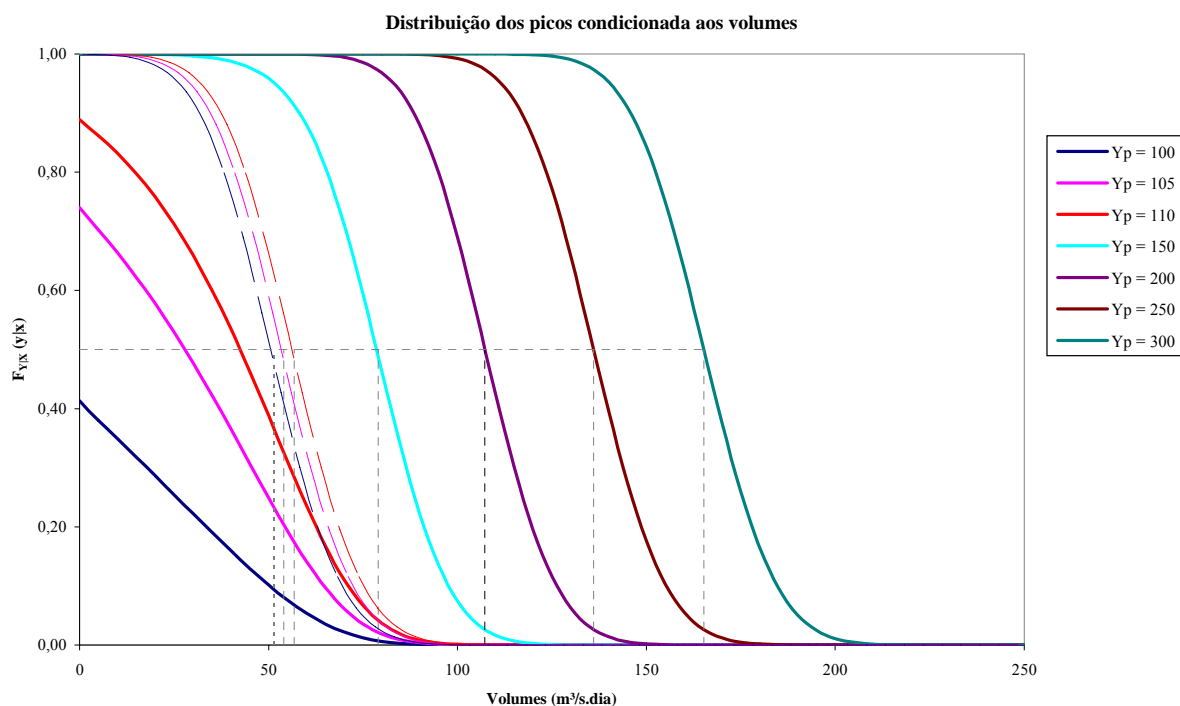


Figura 4.15 – Distribuição dos picos condicionada aos volumes

Assim, na integração da equação 3.89, utilizou-se a distribuição esperada dos picos condicionada aos volumes para valores de Y_p superiores a u e inferiores a $150 \text{ m}^3/\text{s}$. Por outro lado, para valores superiores a $150 \text{ m}^3/\text{s}$, a distribuição foi calculada pela equação 3.103.

A figura 4.16 mostra os dados observados de máximas anuais de vazão para Ponte do Vilela, com posição de plotagem de Gringorten, e a distribuição anual das vazões médias diárias máximas dada pelas equações 3.89 e 3.90.

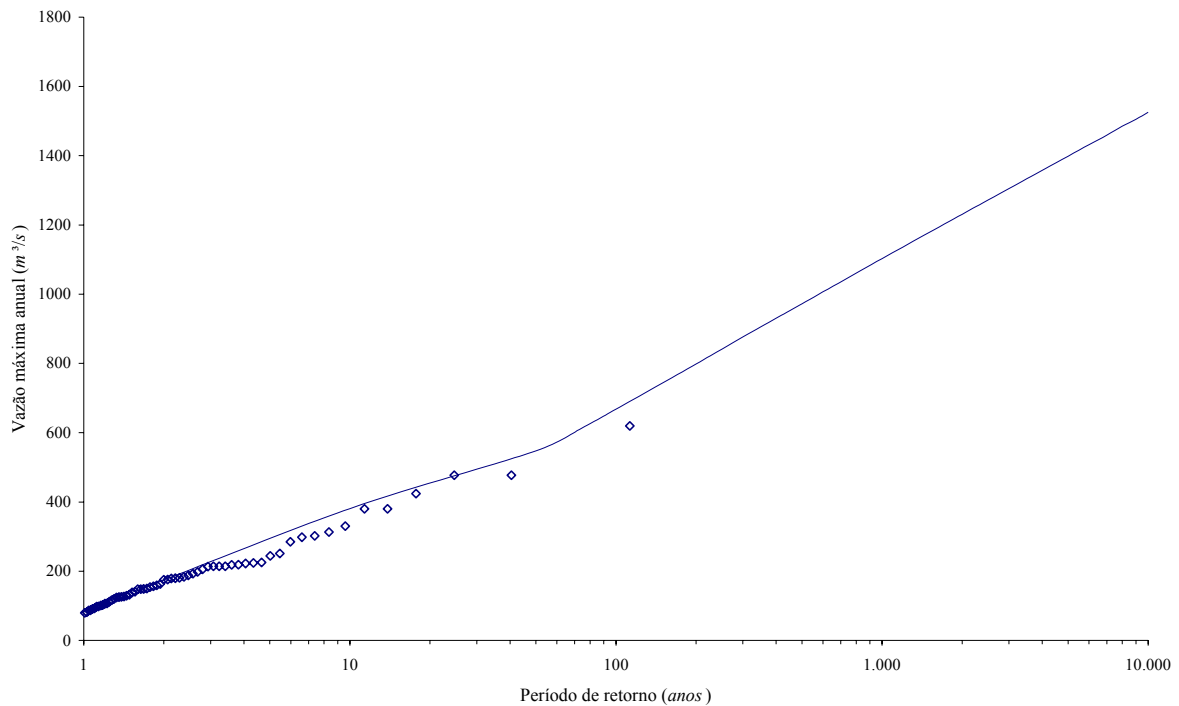


Figura 4.16 – Vazão máxima anual *versus* período de retorno em Ponte do Vilela

4.7.1 Análise de frequência convencional

A curva de quantis apresentada na figura 4.16 deve ser analisada sob o ponto de vista de outras metodologias de cálculo de vazão máxima. Inicialmente, fez-se a análise local dos dados de vazão por meio do ajuste de 11 distribuições de probabilidades. Os parâmetros de cada distribuição foram estimados pelo método dos momentos-L. A tabela 4.10 mostra os parâmetros de cada distribuição e a tabela 4.11 traz os quantis de vazão máxima anual.

Dentre as distribuições utilizadas fez-se uma análise das distribuições de probabilidades de 2 e 3 parâmetros por meio do *software SEAF* – Sistema Especialista para Análise de Frequência local – disponível no endereço eletrônico www.ehr.ufmg.br. O software seleciona as melhores distribuições de probabilidades a serem utilizadas na análise de frequência local por meio da metodologia proposta por Cândido (2003). Para os dados de Ponte do Vilela foi selecionada a GEV com um nível de confiança de 0,98, dentro do contexto de lógica difusa proposto por Cândido (2003).

Tabela 4.10 – Parâmetros das distribuições usadas na estimativa local dos quantis de vazão

Distribuição	Nº Parâmetros	Símbolo	Posição	Escala	Forma		
Exponencial	2	EXP	86,0787	109,9874			
Gama	2	GAM	3,7884	51,7548			
Gumbel	2	GUM	150,2703	79,3392			
Normal	2	NOR	196,0661	97,4738			
Normal generalizada	3	GNO	165,2894	80,2931	-0,6811		
Pareto generalizada	3	GPA	84,9625	113,3587	0,0203		
Generalizada de valores extremos	3	GEV	143,1810	61,2582	-0,2270		
Logística generalizada	3	GLO	168,2095	45,9578	-0,3244		
Pearson tipo III	3	PE3	196,0661	109,3233	1,9462		
Kappa	4	KAP	127,6033	74,0787	-0,1534	0,3720	
Wakeby	5	WAK	74,7524	129,9792	3,7069	81,3486	0,1318

Tabela 4.11 – Vazões máximas anuais para Ponte do Vilela calculadas localmente por meio de distribuições de probabilidades.

T (anos)	DISTRIBUIÇÃO										
	EXP	GAM	GUM	NOR	GNO	GPA	GEV	GLO	PE3	KAP	WAK
2	162,3	179,1	179,3	196,1	165,3	163,0	166,6	168,2	163,2	165,5	166,2
5	263,1	272,0	269,3	278,1	256,5	264,5	252,6	248,7	263,7	256,4	255,6
25	440,1	402,5	404,0	366,7	435,8	438,2	431,1	423,7	437,9	434,4	436,0
50	516,4	453,3	459,8	396,3	524,9	511,3	527,6	527,2	512,6	523,9	526,2
75	560,9	482,2	492,3	412,1	580,8	553,6	591,3	598,9	556,2	580,6	583,0
100	592,6	502,3	515,2	422,8	622,3	583,3	640,0	655,5	587,1	623,0	625,1
200	668,8	549,9	570,4	447,1	728,8	654,4	771,2	815,4	661,6	733,0	733,4
300	713,4	577,2	602,7	460,5	795,6	695,5	857,9	926,8	705,1	802,9	801,6
500	769,6	611,1	643,3	476,6	884,6	746,8	979,1	1089,5	759,9	897,4	892,7
1.000	845,8	656,5	698,3	497,3	1014,7	815,6	1167,7	1357,9	834,2	1038,0	1026,7
2.000	922,1	701,1	753,3	516,8	1156,1	883,4	1388,3	1693,8	908,4	1194,3	1173,4
3.000	966,7	726,9	785,5	527,8	1244,3	922,6	1534,4	1928,3	951,8	1293,8	1265,7
4.000	998,3	745,1	808,3	535,3	1309,5	950,3	1646,5	2114,3	982,6	1368,2	1334,2
5.000	1022,9	759,1	826,0	541,1	1361,6	971,6	1738,6	2271,1	1006,5	1428,2	1389,2
6.000	1042,9	770,6	840,5	545,8	1405,1	989,0	1817,4	2407,9	1026,0	1478,8	1435,3
7.000	1059,9	780,2	852,7	549,7	1442,6	1003,6	1886,6	2530,0	1042,5	1522,7	1475,2
8.000	1074,6	788,6	863,3	553,0	1475,6	1016,2	1948,6	2640,8	1056,8	1561,6	1510,4
9.000	1087,5	795,9	872,6	556,0	1505,1	1027,3	2004,8	2742,6	1069,4	1596,5	1542,0
10.000	1099,1	802,5	881,0	558,6	1531,9	1037,3	2056,4	2837,1	1080,6	1628,3	1570,6

Verifica-se que a análise de frequência convencional possibilita uma faixa muito ampla de respostas. Para o quantil de 10.000 anos, por exemplo, a variação de possibilidades cobre uma faixa de 559 m^3/s até 2837 m^3/s . Mesmo com a análise realizada pelo *SEAF*, não há como prever qual das distribuições fornece o quantil mais próximo da realidade.

4.7.2 Modelo Poisson-Pareto

Uma forma de se obter um melhor resultado da análise de frequência convencional é, na maioria das vezes, aumentar a quantidade de dados da amostra. A utilização das séries de duração parcial proporciona esse aumento. Ajustou-se o modelo descrito no item 2.3.2 à série

parcial dos picos de vazão mostrada na tabela 4.1. A seguir são apresentadas as etapas de cálculo dos parâmetros da distribuição de Pareto.

Nos cálculos que se seguem, x_i indica a série $Q_i - Q_1$, sendo Q_1 o *threshold* igual a 96,6 m^3/s .

$$\bar{x} = 59,17$$

$$S_x = 66,98$$

A determinação dos parâmetros da distribuição generalizada de Pareto pelo método dos momentos é feita pelas seguintes equações:

$$\hat{\alpha} = \frac{\bar{x}}{2} \left(\frac{\bar{x}^2}{S_x^2} + 1 \right) \quad (4.22)$$

$$\hat{k} = \frac{1}{2} \left(\frac{\bar{x}^2}{S_x^2} - 1 \right) \quad (4.23)$$

As estimativas obtidas são $\hat{\alpha} = 52,68$ e $\hat{k} = -0,1098$.

Se a distribuição da série parcial é modelada pela distribuição generalizada de Pareto, pode-se mostrar, a partir das equações 2.45 e 2.46, que a distribuição anual é uma GEV com os parâmetros dados a seguir:

$$\hat{\alpha}_* = \hat{\alpha} \cdot [\hat{\Lambda}(1)]^{\hat{k}} = 52,68 \cdot (1,5)^{0,1098} = 55,07$$

$$\hat{u} = Q_1 + \frac{\hat{\alpha} - \hat{\alpha}_*}{\hat{k}} = 96,6 + \frac{52,68 - 55,07}{-0,1098} = 118,44$$

Os quantis podem, então, ser determinados pela função inversa da GEV, ou seja:

$$Q_T = u + \frac{\alpha_*}{k} \left\{ 1 - [-\ln F(T)]^k \right\} \quad (4.24)$$

onde T representa o período de retorno em anos da série de máximos anuais.

A tabela 4.12 mostra os quantis calculados para Ponte do Vilela considerando o modelo Poisson-Pareto para séries parciais.

Tabela 4.12 – Quantis e período de retorno em anos para Ponte do Vilela calculados pelo modelo POT

Quantis POT - m^3/s			
T	Q	T	Q
2	139,0	1.000	687,6
5	208,2	2.000	772,3
10	259,0	3.000	824,9
50	386,7	4.000	863,7
75	422,0	5.000	894,6
100	448,0	6.000	920,4
200	514,0	7.000	942,7
300	554,9	8.000	962,2
400	585,0	9.000	979,7
500	609,1	10.000	995,6

4.7.3 Regionalização por momentos-L

CPRM (2001) fez uma estudo de regionalização de vazões máximas anuais para as sub-bacias 40 e 41, que compreende uma área desde as nascentes do rio São Francisco até a barragem de Três Marias inclusive. O estudo abrange, portanto, a bacia do rio Pará. O estudo de regionalização das vazões máximas anuais foi realizado com a aplicação do método *index-flood*, utilizando momentos-L e empregando as estatísticas-L. O cálculo dos quantis máximos de vazão foi feito pela metodologia proposta por Hosking & Wallis (1997).

A bacia do rio Pará se localiza na região homogênea denominada MAX-VI, segundo CPRM (2001). Para essa região, a equação que define o *index-flood* é:

$$Q_{MAX-MÉD} = 0,2451 \cdot A^{0,8751} \quad (4.25)$$

na qual A é a área de drenagem em km^2 .

Para Ponte do Vilela, o *index-flood* calculado pela equação 4.25 é $157,76 m^3/s$. Embora esse valor possa ser usado sem restrições, em locais onde há medição de vazão, obtêm-se melhores resultados utilizando o *index-flood* calculado pelos dados amostrais. Assim, evita-se a propagação do erro inerente aos modelos de regressão, como no caso da equação 4.25.

O *index-flood* local é dado pela média dos máximos anuais da estação considerada. No caso de Ponte do Vilela tem-se $Q_{MAX-MÉD} = 196,07 \text{ m}^3 / \text{s}$.

A distribuição de probabilidades regional foi definida pelo teste de aderência Z proposto por Hosking & Wallis (1997). A distribuição que mostrou os melhores resultados foi a Logística Generalizada, com os seguintes parâmetros:

$$\xi = 0,918$$

$$\alpha = 0,223$$

$$k = -0,213$$

Os quantis de vazão máxima anual adimensionais são dados pela inversa da distribuição logística de acordo com a equação a seguir:

$$q(T) = \xi + \frac{\alpha}{k} \left\{ 1 - \left[\frac{1 - F(T)}{F(T)} \right]^k \right\} \quad (4.26)$$

O quantis anuais são dados pelo método *index-flood* de acordo com a equação 4.27, ou seja:

$$Q(T) = q(T) \cdot Q_{MAX-MÉD} \quad (4.27)$$

A tabela 4.13 mostra os quantis calculados para Ponte do Vilela considerando a regionalização por momentos-L.

Tabela 4.13 – Quantis máximos anuais para Ponte do Vilela calculados por Momentos-L

Quantis ML - m^3/s			
T	Q	T	Q
2	180,0	1.000	868,5
5	250,5	2.000	1010,8
10	302,5	3.000	1104,3
50	445,0	4.000	1175,7
75	488,1	5.000	1234,2
100	521,0	6.000	1284,1
200	608,6	7.000	1327,8
300	666,0	8.000	1366,9
400	709,8	9.000	1402,2
500	745,7	10.000	1434,6

4.7.4 Modelo estocástico de simulação de vazões

O modelo de geração de séries sintéticas desenvolvido por Lima (2004) tem como característica a manutenção das propriedades estatísticas das séries observadas. Lima (2004) gerou uma série de chuva de 10.000 anos com base na chuva média dada pelas estações 02044005, 02044036 e 02044042. Posteriormente calibrou o modelo chuva-vazão RIO GRANDE, com base na estação fluviométrica 40130002, na série pluviométrica média e na estação evaporimétrica 1945017. A calibração foi feita sobre o período de dados comuns de 10/1974 a 09/1983 e forneceu, como parâmetros de calibração, os valores apresentados na tabela 4.14.

Tabela 4.14 – Parâmetros do modelo RIO GRANDE

Parâmetro	Intervalo de Variação	Valores Calibrados
K	0,40 a 0,85	0,84
IMP (%)	0 a 10	0
WU (mm)	3 a 25	22
WL (mm)	50 a 100	86,08
WD (mm)	15 a 70	66,79
SM (mm)	5 a 90	50
B	0,1 a 2,0	1
EX	0,3 a 2,0	0,35
C	0 a 0,25	0
KSS	0,1 a 0,5	0,32
KG	0,3 a 0,8	0,63
CI	0,5 a 0,9	0,79
CG	0,960 a 0,999	0,995

FONTE: Lima (2004)

A tabela 4.15 mostra os quantis simulados para Ponte do Vilela obtidos por transformação chuva-vazão.

Tabela 4.15 – Quantis máximos anuais simulados para Ponte do Vilela

Quantis Simulados - m³/s			
T	Q	T	Q
1,02	67	400	630
2	173	500	669
5	246	1.000	764
10	299	2.000	873
50	431	3.000	1031
100	492	4.000	1127
200	562	5.000	1132
300	600	10.000	1207

FONTE: Lima (2004)

4.7.5 Comparação entre os métodos de estimação de vazão

O gráfico da figura 4.17 mostra os quantis obtidos pelas diversas metodologias apresentadas. Foram plotados os quantis para períodos de retorno superiores a 10 anos. Para quantis inferiores o gráfico se torna de difícil leitura não favorecendo a análise comparativa entre as metodologias.

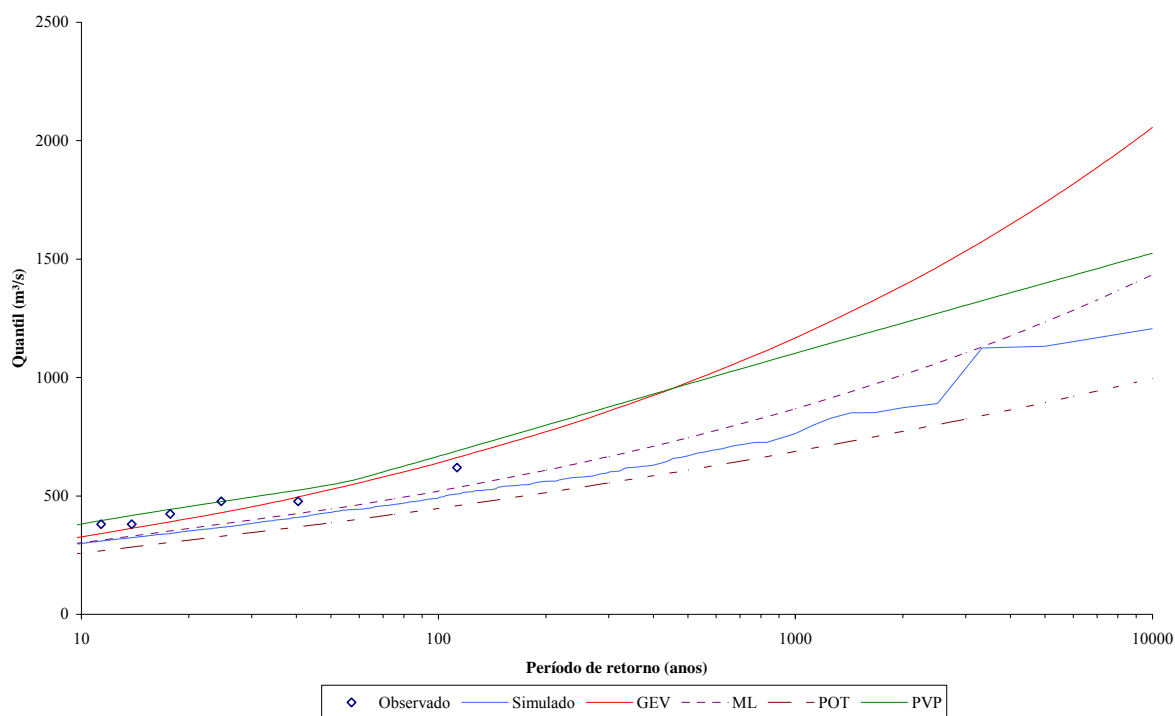


Figura 4.17 – Quantis máximos para Ponte do Vilela segundo várias metodologias: PVP – metodologia desenvolvida neste trabalho; Simulado – valores obtidos por Lima (2004); GEV – ajuste pela análise de frequência convencional; ML – regionalização por momentos-L feita por CPRM (2001); POT – modelo Poisson-Pareto para séries parciais

Observa-se uma grande variação entre os valores. Para o quantil de 10.000 anos de período de retorno, a metodologia do PVP se aproxima bastante dos resultados obtidos pela regionalização por momentos-L. Esse fato pode ser visto como uma vantagem a favor da metodologia proposta neste trabalho, uma vez que diversos autores preconizam o uso das estatísticas L para a determinação de quantis extremos de vazão.

O trabalho elaborado por Wallis (1989) sobre técnicas de regionalização *index-flood*, utilizando momentos-L, representou um avanço nos estudos de regionalização, permitindo reduzir a subjetividade das etapas necessárias ao seu desenvolvimento. Bobée & Rasmussen

(1995) consideram a contribuição de Hosking & Wallis como a mais relevante dos últimos anos para a obtenção de melhores estimativas das probabilidades de eventos raros.

Desta forma, pode-se considerar, mesmo que de maneira subjetiva, que a metodologia do PVP proporciona uma boa estimativa de quantis extremos. Além disso, a metodologia do PVP foi estruturada de forma a utilizar os três princípios indicados pelo NRC (1988) para incrementar a análise de frequência. Assim, espera-se que a utilização de mais informações produza uma melhor precisão dos quantis extremos, muito embora, não tenha sido analisado, neste trabalho, o efeito da inclusão de novas fontes de incertezas na estimativa dos quantis.

Por outro lado, a característica aleatória das vazões de um rio não permite saber qual o valor correto do quantil para um determinado período de retorno. Isso seria viável caso se conhecesse a população dos valores possíveis de vazão. Como o monitoramento de vazões é feito por meio de uma rede com poucas estações fluviométricas, e os dados raramente ultrapassam um período de 50 anos, verifica-se a grande subjetividade na determinação de vazões com período de retorno de 10.000 anos, por exemplo.

As várias respostas possíveis mostradas na figura 4.17 são um reflexo das incertezas quanto à aleatoriedade das vazões de um rio e de como cada metodologia trabalha essa aleatoriedade. Embora não se possa saber ao certo qual é o quantil correto, pode-se verificar que a metodologia do PVP oferece respostas condizentes com as demais metodologias utilizadas atualmente, dentro de um contexto de busca da maior suficiência dos estimadores de quantis extremos.

5 CONCLUSÕES E RECOMENDAÇÕES

Neste trabalho foi desenvolvida e aplicada uma metodologia que faz uso dos três princípios identificados pelo U. S. National Research Council (NRC, 1988) para a melhoria das estimativas de quantis extremos de vazão, a saber:

- a) substituição do tempo pelo espaço;
- b) introdução de maior estrutura aos modelos; e
- c) enfoque na cauda superior das distribuições de probabilidade.

Inicialmente, fez-se uma regionalização (princípio *a*) dos dados de chuva considerando as séries de duração parcial (princípio *c*), de modo a incluir, na análise, a variabilidade espacial dos elementos formadores das vazões de pico.

O uso de séries de duração parcial, além de possibilitar um aumento dos dados em análise, quando comparadas às séries dos máximos anuais, permite orientá-la para os maiores valores observados, a despeito desse ser ou não um máximo anual. Desta forma, evita-se enviesar a determinação da cauda superior da distribuição de probabilidade em função de dados centrais mais freqüentes.

Por outro lado, o uso da TCEV permite o ajuste da cauda inferior e superior em momentos separados. O fato de a TCEV ser uma produto de duas distribuições Gumbel, uma modelando a cauda inferior e a outra modelando a cauda superior, dá uma grande flexibilidade à análise. Embora a TCEV tenha problemas quanto à determinação de seus parâmetros, sua inclusão na metodologia desenvolvida possibilitou atingir os objetivos procurados. As propriedades da TCEV poder modelar separadamente as estatísticas de ordem superior e de ter uma cauda superior do tipo exponencial permitiu seu uso na estrutura desenvolvida por Naghettini (1994).

A etapa seguinte foi a determinação da distribuição dos volumes de cheia. Conforme visto, quando a bacia está saturada, incrementos de volume de chuva tendem a produzir incrementos proporcionais de volume de cheia. Assim, a distribuição de probabilidades dos volumes de cheia pode ser deduzida da distribuição das precipitações de mesma duração efetuando-se a translação das distribuições acumuladas de uma distância fixa ao longo do eixo da variável.

Essa é a motivação principal do emprego das premissas do método GRADEX. Da forma como é proposto, o método GRADEX vai ao encontro do princípio (b) do NRC. Com efeito, tal princípio indica a possibilidade, por exemplo, de se equacionar a transformação chuva-vazão em condições extremas, o que, em última análise, é a própria aplicação do método GRADEX.

Neste trabalho, as propriedades da TCEV foram empregadas para a aplicação do método GRADEX. Embora se pudesse utilizar a distribuição empírica dos valores observados até o quantil referente à saturação da bacia e, a partir desse ponto, extrapolar a distribuição dos volumes pelo método GRADEX, optou-se por ajustar a TCEV de modo a, concomitantemente, possibilitar uma boa aderência da cauda baixa e respeitar a aplicação do GRADEX para a cauda superior. A vantagem dessa aplicação é que se obtém uma distribuição única e contínua para os volumes de cheia, o que torna sua análise e utilização mais simples.

Em seguida determinou-se a distribuição dos picos condicionada aos volumes. Essa determinação foi possibilitada pela relação regional entre os picos e os volumes encontrada no capítulo 4. Na aplicação original do método PVP, Naghettini (1994) utilizou dados históricos de chuvas extremas catalogadas pelo U. S. Army Corps of Engineers e posteriormente transformou esses dados em vazão por meio do modelo de simulação chuva-vazão HSP-F. Assim, obteve os pares pico-volume necessários para a construção da distribuição dos picos condicionada aos volumes. No Brasil, não há disponibilidade de cadastros semelhantes, com informações sobre todas as bacias. Por outro lado, os registros pluviométricos não são longos o suficiente para incorporar todas as condições de chuvas extremas possíveis. Dessa forma, não há como se aplicar o método PVP da forma como foi proposto originalmente de maneira segura. A relação pico-volume regional, encontrada no capítulo 4, possibilitou contornar esse problema e adequar o método PVP às condições de disponibilidade de dados no Brasil.

A relação pico-volume regional possibilita não somente a utilização do método PVP mas, também, a análise da hidráulica fluvial das bacias dentro da sua região de aplicação. A possibilidade de obtenção do pico do hidrograma unitário em locais desprovidos de monitoramento de vazão é um produto de especial interesse para dimensionamentos hidráulicos e que não era esperado inicialmente.

Finalmente, as várias metodologias foram agrupadas numa estrutura que permitiu o cálculo dos quantis máximos anuais de vazão. A estrutura metodológica proposta por Naghettini (1994) foi preservada e as várias partes do processo foram modificadas de modo a atender as condições de disponibilidade e qualidade de dados no Brasil. Conforme visto nos capítulos que formaram esta dissertação, o método preserva os princípios identificados pelo NRC. Nesse contexto, a metodologia proposta apresenta interessantes atributos, quando comparada às demais metodologias analisadas.

Um ponto desfavorável do método PVP é a grande quantidade de cálculos e etapas para a determinação dos quantis. Visando transpor essa dificuldade foi elaborado um pacote computacional que automatiza todas as etapas de cálculo. O software, denominado PVP, foi construído em ambiente Microsoft Windows[®] utilizando a linguagem Visual Basic .NET 2002[®]. A descrição do pacote é dada no Anexo 4. O software poderá ser obtido brevemente a partir da URL www.ehr.ufmg.br.

Os resultados obtidos pela aplicação do método foram satisfatórios quando comparados a outras metodologias de cálculo de vazões máximas.

A comparação entre metodologias, para quantis com grande período de retorno, é um tanto controvertida. Como a população dos valores possíveis para a vazão não é conhecida, não se pode dizer qual metodologia fornece os melhores resultados. Nesse sentido, como recomendação para futuros estudos, deveria se estabelecer a variância para os quantis calculados pelo método PVP. Somente dessa maneira poderia se verificar se a metodologia proposta fornece quantis com erros menores que as demais metodologias. A determinação da variância dos quantis gerados pelo PVP de forma analítica é por demais complicada. Assim sugere-se que a variância seja estimada por meio de simulações com o método de Monte Carlo.

Da forma como foi proposta, a metodologia não possui uma grande capacidade de generalização, sobretudo em locais desprovidos de monitoramento de vazão. Isso se deve ao fato de que, para a aplicação da metodologia, deve-se ter, *a priori*, uma estimativa do erro padrão de relação pico-volume (σ_ε) e do quantil (Q_{Tc}), a partir do qual a bacia pode ser considerada saturada. Nesse sentido, recomenda-se as seguintes direções para a obtenção de avanços, em relação àqueles aqui apresentados:

- A relação pico-volume regional, aqui obtida, pode ser ampliada a um maior número de estações fluviométricas, notadamente aquelas com pequena área de drenagem. A inclusão de novas estações pode vir a fornecer uma melhor estimativa dos picos de vazão;
- No que diz respeito ao erro padrão da relação pico-volume podem ser tomadas duas decisões: em locais desprovidos de dados, pode-se utilizar o erro total gerado pela regressão regional, o que de certa forma não permite incluir a variância local na análise dos quantis. Pode-se, também, construir um modelo regional, a partir das características físicas e/ou climáticas da bacia, que agregue a variabilidade espacial do erro. Essa última opção é a que parece, numa primeira análise, a mais adequada para os propósitos da generalização requerida; e
- Quanto ao quantil Q_{Tc} , sua regionalização é mais complicada, pois envolve a sua medição em vários pontos da região de interesse. No caso do uso da TCEV, o quantil Q_{Tc} pode ser visto como um parâmetro de posição. Mesmo visto dessa forma, o quantil não pode ser definido apenas pelas características das séries fluviométricas observadas. Assim, recomenda-se que sejam realizados estudos hidrogeológicos de forma a determinar o valor do ponto de saturação da bacia e da chuva necessária para atingir tal ponto. Com o propósito de generalizar o método PVP para locais sem medição de vazão, os estudos hidrogeológicos deveriam apontar para uma relação entre Q_{Tc} e as características físicas e climáticas da bacia.

Assim, espera-se que futuros desenvolvimentos melhorem a metodologia proposta. Por fim, espera-se que os desenvolvimentos realizados nesta pesquisa tenham colaborado com o avanço da área de conhecimento.

6 REFERÊNCIAS BIBLIOGRÁFICAS

- BERAN, M., HOSKING, J. R. M. & ARNELL, N. Comment on – Two component extreme value distribution for flood frequency analysis – by F. Rossi, M. Fiorentino and P. Versace, *Water Resource Research*, 22(2), p. 263-266, 1986.
- BOBÉE, B. & RASMUSSEN, P. Recent advances in flood frequency analysis. U.S. National Report to IUGG, 1991-1994, *Geophysics*, vol. 33 Suppl., 1995.
- BRADLEY, A. A. & POTTER, K. W. Flood frequency analysis of simulated flows, *Water Resources Research*, 28(9), p. 2375-2385, 1992.
- C.T.G.R.E.F. L'application de la méthode du gradex à l'estimation des crues de faible fréquence, Centre Technique de Génie Rural, des Eaux et des Forêts, Ministère de l'Agriculture, Paris, França, 1972.
- CÂNDIDO, M. O. SEAF – Um protótipo de um sistema especialista para análise de frequência local de eventos hidrológicos máximos anuais, Dissertação de Mestrado, Programa de Pós Graduação em Saneamento, Meio Ambiente e Recursos Hídricos – UFMG, Belo Horizonte, 2003.
- CLEVELAND, W. S. Robust locally weighted regression and smoothing scatterplots, *Journal of American Statistical Association*, 74(368), p. 829-836, 1979.
- COMPANHIA DE PESQUISA DE RECURSOS MINERAIS. Projeto Análise de Consistência de Dados Pluviométricos – Bacia do Rio São Francisco. Sub-Bacias 40 a 44: Relatório Técnico Parcial. Sub-Bacia 40, DNAEE/CPRM, Belo Horizonte, 1995.
- COMPANHIA DE PESQUISA DE RECURSOS MINERAIS. Regionalização de vazões. Sub-Bacias 40 e 41: Relatório Final. Vazões Máximas. Vol. 4, CPRM/ANEEL. Belo Horizonte, 2001.
- CUNNANE, C. A note on the Poisson assumption in partial duration series models, *Water Resources Research*, 15(2), p. 489-494, 1979.
- DALRYMPLE, T. Flood-frequency analyses, *Manual of Hydrology: Part. 3. Flood-flow Techniques*, Geological Survey Water Supply Paper 1543-A, U. S. Government Printing Office, Washington, D.C., 1960.
- DAVIS, E. G. & NAGHETTINI, M. C. Estudo de Chuvas Intensas no Estado do Rio de Janeiro. CPRM, Brasília, 2000.
- DUAN, Q. A Global Optimization Strategy for Efficient and Effective Calibration of Hydrologic Models, PhD thesis, University of Arizona, 1991.
- DUBAND, D., MICHEL, C., GARROS, H. & ASTIER, J. Estimating extreme value floods and the design flood by the gradex method. In: *Proceedings of the 16th International Congress on Large Dams*, p. 1009-1047, San Francisco, California, 1988.
- FIORENTINO, M. & GABRIELE, S. Distribuzione TCEV: metodi di stima dei parametri e proprietà statistiche degli stimatori, *Geodata* 25, Cosenza, Itália, 1985.
- FIORENTINO, M., GABRIELE, S., ROSSI, F. & VERSACE, P. Hierarchical approach for regional flood frequency analysis. In: *Regional Flood Frequency Analysis*, V. P. Singh (ed.), Reidel, Dordrecht, p. 35-49, 1987.

- GABRIELE, S. & ARNELL, N. W. A hierarchical approach to regional flood frequency analysis, *Water Resources Research*, 26(6), p. 1281-1289, 1991.
- GABRIELE, S. & IIRITANO, G. *Análise regional das chuvas na Região Sudoeste do Estado do Rio de Janeiro*, Relatório interno nº 414, CNR-IRPI, Rende, Itália, 1994.
- GABRIELE, S. & VILLANI, P. An investigation of the feasibility space of parameter estimation using POME and ML to the TCEV distribution. In: *Entropy and energy dissipation in water resources*, V. P. Singh and M. Fiorentino (eds.), p. 193-203, Kluwer Academic Publishers, 1992.
- GREENWOOD, J. A., LANDWEHR, J. M., MATALAS, N. C. & WALLIS, J. R. Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. In: *Water Resources Research*, American Geophysical Union. V.15, n.5, p.1049-1054, 1979.
- GRUBBS, F. E. & BECK, G. Extension of sample sizes and percentage points for significance tests of outlying observations, *Technometrics*, 14(4), p. 847-854, 1972.
- GUILLOT, P. & DUBAND, D. La méthode du gradex pour le calcul de la probabilité des crues à partir des pluies. In: *Floods and their computation – Proceedings of the Leningrad Symposium*, p. 560-569, IASH Publication no. 84, 1967.
- GUMBEL, E. J. *Statistics of extremes*, Columbia University Press, New York, 1958.
- HAWKINS, D. M. *Identification of outliers*. Monographs on applied probability and statistics, Chapman and Hall, New York, 1980.
- HOSKING, J. R. M. & WALLIS, J. R. *Regional Frequency Analysis - an approach based on L moments*. Cambridge University Press, Cambridge, 1997.
- HEC - HYDROLOGIC ENGINEERING CENTER. *HEC-1 Flood Hydrograph Package - Reference Manual*, U. S. Army Corps of Engineers, Davis, 1981.
- JOHANSON, R. C., IMHOFF J. C., KITTLE J. L., DONIGIAN A. S. *Hydrological Simulation Program – FORTRAN (HSPF): Users manual for release 8.0*, Rep. EPA-600/3-84-066, U.S. Environmental Protection Agency, Washington, DC, 1984.
- KITE, G. W. *Frequency and risk analysis in hydrology*, Water Resources Publications, Fort Collins, 1977.
- LETTENMAIER, D. P. & POTTER, K. W. Testing flood frequency estimation methods using a regional model. *Water Resources Research*, 21(2), p. 1903-1914, 1985.
- LIMA, A. A. *Metodologia integrada para determinação da enchente de projeto de estruturas hidráulicas por meio de séries sintéticas de precipitação e modelos chuva-vazão*, Dissertação de Mestrado, Programa de Pós Graduação em Saneamento, Meio Ambiente e Recursos Hídricos – UFMG, Belo Horizonte, 2004.
- LINSLEY, R. K., KOHLER, M. A. & PAULHUS, J. L. H. *Hydrology for Engineers*, Ed. 2, McGraw-Hill, New York, 1975.
- MARTINS, E. S. & STEDINGER, J. R. Cross-correlation among estimators of shape, *Water Resources Research*, 38(11), doi:10.1029/2002WR001589, 2002.
- MIMIKOU, M. A study of drainage basin linearity and non-linearity, *Journal of Hydrology*, vol. 64, p. 113-134, 1983.

- MOLFINO, M. E. & CRUISE, J. F. An additional analysis of peak-volume relations and standardization procedures, *Water Resources Bulletin*, 26(4), p. 687-692, 1990.
- NAGHETTINI, M. C. Methodology for estimating the upper tail of flood-peak frequency distribution using hydrometeorological information, PHD thesis, Colorado University, Colorado, 1994.
- NAGHETTINI, M. C. & PINTO, E. J. A. Hidrologia Estatística – Boletim Técnico, CPRM, Belo Horizonte, no prelo.
- NAGHETTINI, M. C., NASCIMENTO, N. O., THIMOTTI, T., LIMA, A. A. & SILVA, F. E. O. Modelo Rio Grande de Simulação Hidrológica para Previsão de Vazões de Curto Prazo: Formulação Teórica, Departamento de Engenharia Hidráulica e Recursos Hídricos da UFMG, Belo Horizonte, 2002.
- NORTH, M. Time-dependent stochastic model of floods, *ASCE, Journal of the Hydraulics Division*, 106(HY5), p. 717-731, 1980.
- NRC. Estimating Probabilities of Extreme Floods, National Academy Press, Washington, DC, 1988.
- PILON, P. J. Consolidated frequency analysis – CFA. User Manual for Version I, Environment Canada, Ottawa, 1985.
- PINHEIRO, M. M. G. Estudo de chuvas intensas na região metropolitana de Belo Horizonte – RMBH. Dissertação de Mestrado, Programa de Pós Graduação em Saneamento, Meio Ambiente e Recursos Hídricos – UFMG, Belo Horizonte, 1997.
- PINTO, E. J. A. & NAGHETTINI, M. C. Definition of homogeneous regions and frequency analysis of annual maximum daily precipitation over the upper São Francisco river basin, in southeastern Brazil. In: *ASCE's 1999 International Water Resources Engineering Conference*, ASCE-American Society of Civil Engineer, Seattle, 1999.
- PRESCOTT, P. & WALDEN, A. T. Maximum likelihood estimation of the three-parameter generalized extreme value distribution from censored samples, *J. Statist. Comput. Simul.*, vol. 16, p. 241-250, 1983.
- QUINTÃO, C. M. F. Regionalização hidrológica utilizando o método TCEV (Two Component Extreme Value) com apoio de um sistema de informações geográficas, Dissertação de Mestrado, Programa de Pós Graduação em Saneamento, Meio Ambiente e Recursos Hídricos – UFMG, Belo Horizonte, 2003.
- RAO, A. R. & HAMED, K. H. Flood frequency analysis, CRC Pres, Boca Raton, Flórida, 2000.
- ROGERS, W. F. & ZIA, H. A. Linear and nonlinear runoff from large drainage basins, *Journal of Hydrology*, vol. 55, p. 267-278, 1982.
- ROGERS, W. F. Some characteristics and implications of drainage basin linearity and non-linearity, *Journal of Hydrology*, vol. 55, p. 247-265, 1982.
- ROGERS, W. F., A practical model for linear and nonlinear runoff, *Journal of Hydrology*, vol. 46, p. 51-78, 1980.
- ROSSI, F., FIORENTINO, M. & VERSACE, P. Two-Component Extreme Value Distribution for flood frequency analysis, *Water Resources Research*, 20 (7), p. 847-856, 1984.

- SEREBRENICK, R. & CARVALHO, N. O. Definições, finalidade, estabelecimento e aplicações da técnica do hidrograma unitário para uma bacia hidrográfica, Ministério das Minas e Energia, Rio de Janeiro, 1966.
- SHERMAN, L. K. Stream flow from rainfall by the unitgraph method, Eng. News-Rec., vol.108, p. 501, 1932.
- SINGH, V. P. & AMINIAN, H. An empirical relation between volume and peak of direct runoff, Water Resources Bulletin, 22 (5), p. 725-730, 1986.
- SIVAPALAN, M., JOTHITYANGKOON, C. & MENABDE, M. Linearity and nonlinearity of basin response as a function of scale: discussion of definitions, Water Resources Research, 38 (2), 2002.
- SMITH, J. A. Regional flood frequency analysis using extreme order statistics of the annual peak record, Water Resources Research, 25 (2), 311-317, 1989.
- STEDINGER, J. R., VOGEL, R. M. & FOUFOULA-GEORGIU, E. Frequency analysis of extreme events, Chapter 18. In: Handbook of hydrology, ed. D. R. Maidment, McGraw-Hill, New York, 18.1-18.66, 1993.
- TODOROVIC, P. & ZELENHASIC, E. A stochastic model for flood analysis. Water Resources Research, 6(6), p. 411-424, 1970.
- TUCCI, C. E. M. Regionalização de vazões In: Hidrologia: ciência e aplicação. Porto Alegre, Ed. UFRGS/ABRH/EDUSP, p.573-611, 1993.
- TUCCI, C. E. M. Regionalização de vazões. Porto Alegre, ANEEL/UFRGS/IPH , 2000.
- U. S. Army Corps of Engineers. Storm Rainfall in the United States, Office of the Chief of Engineers, Washington, 1945.
- U. S. WATER RESOURCES COUNCIL. Guidelines for determining flood flow frequency – bulletin 17, USWRC Hydrology Committee, Washington, 1976.
- WALLIS, J. R. Regional frequency studies using L-moments. In: IBM Research Report, New York, IBM Research Division, RC 14597, p. 469-476, 1989.
- WMO. Statistical distributions for flood frequency analysis, Operational Hydrology Report 33, World Meteorological Organization, Geneva, Switzerland, 1989.
- ZHAO, R. J., ZHUANG, Y. L., FANG, L. R., LIU, X. R. & ZHANG, Q. S. The Xinanjiang Model, Hydrological Forecasting Proceedings, Oxford Symposium, IAHS 129, pp. 351-356, 1980.

ANEXO 1 – DISTRIBUIÇÕES DE PROBABILIDADES

A seguir são dadas os estimadores dos parâmetros das principais distribuições utilizadas em hidrologia. Essa descrição foi retirada de Naghettini & Pinto (no prelo). No texto que se segue, MOM indica o método dos momentos, MVS indica o método da máxima verossimilhança e MML indica o método dos momentos-L.

Distribuição Normal (NOR)

Notação: $X \sim N(\mu, \sigma)$

Parâmetros: μ e σ

$$\text{FDP: } f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \text{ para } -\infty < x < \infty$$

Média: $E[X] = \mu$

Variância: $\text{Var}[X] = \sigma^2$

Coefficiente de Assimetria: $\gamma=0$

Curtose: $\kappa=3$

$$\text{Função Geratriz de Momentos: } \phi(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$$

Estimação de parâmetros:

Método MOM:

$$\hat{\mu}_X = \bar{x}$$

$$\hat{\sigma}_X = s_X$$

Método MVS:

$$\hat{\mu}_X = \bar{x}$$

$$\hat{\sigma}_X = s_X$$

Método MML:

$$\hat{\mu}_X = l_1$$

$$\hat{\sigma}_X = \sqrt{\pi} l_2$$

Distribuição Exponencial (EXP)

Notação: $X \sim E(\theta)$

Parâmetro: θ

$$\text{FDP: } f_X(x) = \frac{1}{\theta} \exp\left(-\frac{x}{\theta}\right), x \geq 0$$

$$\text{Função de quantis: } x(F) = -\theta \ln(1 - F)$$

$$\text{Média: } E[X] = \theta$$

$$\text{Variância: } \text{Var}[X] = \theta^2$$

$$\text{Coeficiente de Assimetria: } \gamma = 2$$

$$\text{Curtose: } \kappa = 9$$

$$\text{Função Geratriz de Momentos: } \phi(t) = \frac{1}{1 - \theta t} \text{ para } t < \frac{1}{\theta}$$

Estimação de parâmetros

$$\text{Método MOM: } \hat{\theta} = \bar{x}$$

$$\text{Método MVS: } \hat{\theta} = \bar{x}$$

$$\text{Método MML: } \hat{\theta} = l_1$$

Distribuição Generalizada de Valores Extremos (GEV)

Notação: $Y \sim \text{GEV}(\alpha, \beta, \kappa)$

Parâmetros: α , β e κ

$$\text{FDP: } f_Y(y) = \frac{1}{\alpha} \left[1 - \kappa \left(\frac{y - \beta}{\alpha} \right) \right]^{1/\kappa - 1} \exp \left\{ - \left[1 - \kappa \left(\frac{y - \beta}{\alpha} \right) \right]^{1/\kappa} \right\} \text{ se } \kappa \neq 0 \text{ e}$$

$$f_Y(y) = \frac{1}{\alpha} \exp \left[- \frac{y - \beta}{\alpha} - \exp \left(- \frac{y - \beta}{\alpha} \right) \right] \text{ se } \kappa = 0$$

$$\text{Função de Quantis: } x(F) = \beta + \frac{\alpha}{\kappa} \left[1 - (-\ln F)^\kappa \right]$$

$$\text{Média: } E[Y] = \beta + \frac{\alpha}{\kappa} \left[1 - \Gamma(1 + \kappa) \right]$$

$$\text{Variância: } \text{Var}[Y] = \left(\frac{\alpha}{\kappa} \right)^2 \left[\Gamma(1 + 2\kappa) - \Gamma^2(1 + \kappa) \right]$$

$$\text{Coeficiente de Assimetria: } \gamma = \langle \text{sinal de } \kappa \rangle \frac{-\Gamma(1+3\kappa) + 3\Gamma(1+\kappa)\Gamma(1+2\kappa) - 2\Gamma^3(1+\kappa)}{[\Gamma(1+2\kappa) - \Gamma^2(1+\kappa)]^{3/2}}$$

Estimação de parâmetros

Método MOM:

Alternativa 1: resolver para κ , utilizando método de Newton, a equação para a variância substituindo $Var[Y]$ pelo seu valor amostral.

Alternativa 2:

para coeficientes de assimetria amostrais $1,1396 < g_X < 10$ ($g = g_X$):

$$\hat{\kappa} = 0,2858221 - 0,357983g + 0,116659g^2 - 0,022725g^3 + 0,002604g^4 - 0,000161g^5 + 0,000004g^6$$

para coeficientes de assimetria amostrais $-2 < g_X < 1,1396$ ($g_X = g$):

$$\hat{\kappa} = 0,277648 - 0,322016g + 0,060278g^2 + 0,016759g^3 - 0,005873g^4 - 0,00244g^5 - 0,00005g^6$$

para coeficientes de assimetria amostrais $-10 < g_X < 0$ ($g = g_X$):

$$\hat{\kappa} = -0,50405 - 0,00861g + 0,015497g^2 + 0,005613g^3 + 0,00087g^4 + 0,000065g^5$$

$$\text{Em seguida, } \hat{\alpha} = \frac{s_X \hat{\kappa}}{\sqrt{\Gamma(1+2\hat{\kappa}) - \Gamma^2(1+\hat{\kappa})}} \text{ e } \hat{\beta} = \bar{x} - \frac{\hat{\alpha}}{\hat{\kappa}} [1 - \Gamma(1+\hat{\kappa})]$$

Método MVS:

$\hat{\alpha}, \hat{\beta}, \hat{\kappa}$ são as soluções simultâneas (método de Newton) do seguinte sistema:

$$\frac{1}{\alpha} \left[\sum_{i=1}^N \exp(-y_i - \kappa y_i) - (1 - \kappa) \sum_{i=1}^N \exp(\kappa y_i) \right] = 0 \quad (C)$$

$$\frac{1}{\kappa \alpha} \left[\sum_{i=1}^N \exp(-y_i - \kappa y_i) - (1 - \kappa) \sum_{i=1}^N \exp(\kappa y_i) + N - \sum_{i=1}^N \exp(-y_i) \right] = 0 \quad (D)$$

$$\frac{1}{\kappa^2} \left[\sum_{i=1}^N \exp(-y_i - \kappa y_i) - (1 - \kappa) \sum_{i=1}^N \exp(\kappa y_i) + N - \sum_{i=1}^N \exp(-y_i) \right] +$$

$$+ \frac{1}{\kappa} \left[- \sum_{i=1}^N y_i + \sum_{i=1}^N y_i \exp(y_i) + N \right] = 0 \quad (E)$$

$$\text{onde } y_i = \frac{1}{\kappa} \ln \left[1 - \kappa \left(\frac{x_i - \beta}{\alpha} \right) \right].$$

A resolução desse sistema é complexa; sugere-se a referência Prescott e Walden (1983) para algoritmo de resolução.

Método MML:

$$\hat{\kappa} = 7,8590C + 2,9554C^2, \text{ onde } C = 2/(3 + t_3) - \ln 2/\ln 3$$

$$\hat{\alpha} = \frac{l_2 \hat{\kappa}}{\Gamma(1 + \hat{\kappa})(1 - 2^{-\hat{\kappa}})}$$

$$\hat{\beta} = l_1 - \frac{\hat{\alpha}}{\hat{\kappa}}[1 - \Gamma(1 + \hat{\kappa})]$$

Distribuição Gumbel (GUM)

Notação: $Y \sim \mathbf{Gum}_{\max}(\alpha, \beta)$

Parâmetros: α e β

$$\text{FDP: } f_Y(y) = \frac{1}{\alpha} \exp\left[-\frac{y - \beta}{\alpha} - \exp\left(-\frac{y - \beta}{\alpha}\right)\right]$$

$$\text{Função de Quantis: } y(F) = \beta - \alpha \ln[-\ln(F)]$$

$$\text{Média: } E[Y] = \beta + 0,5772\alpha$$

$$\text{Variância: } \text{Var}[Y] = \sigma_Y^2 = \frac{\pi^2 \alpha^2}{6}$$

$$\text{Coeficiente de Assimetria: } \gamma = 1,1396$$

$$\text{Curtose: } \kappa = 5,4$$

Estimação de parâmetros

Método MOM:

$$\hat{\alpha} = 0,7797 s_x$$

$$\hat{\beta} = \bar{x} - 0,45 s_x$$

Método MVS:

$\hat{\alpha}$ e $\hat{\beta}$ são as soluções do seguinte sistema de equações:

$$\frac{\partial}{\partial \alpha} \ln[L(\alpha, \beta)] = -\frac{N}{\alpha} + \frac{1}{\alpha^2} \sum_{i=1}^N (x_i - \beta) - \frac{1}{\alpha^2} \sum_{i=1}^N (x_i - \beta) \exp\left(-\frac{x_i - \beta}{\alpha}\right) = 0 \quad (\text{F})$$

$$\frac{\partial}{\partial \beta} \ln[L(\alpha, \beta)] = \frac{N}{\alpha} - \frac{1}{\alpha} \sum_{i=1}^N \exp\left(-\frac{x_i - \beta}{\alpha}\right) = 0 \quad (\text{G})$$

Manipulando-se ambas equações, chega-se a

$$F(\alpha) = \sum_{i=1}^N x_i \exp\left(-\frac{x_i}{\alpha}\right) - \left(\frac{1}{N} \sum_{i=1}^N x_i - \alpha\right) \sum_{i=1}^N \exp\left(-\frac{x_i}{\alpha}\right) = 0 \quad (\text{H})$$

A solução de (H), pelo método de Newton, fornece $\hat{\alpha}$.

Em seguida, $\hat{\beta} = \hat{\alpha} \ln \left[\frac{N}{\sum_{i=1}^N \exp(-x_i/\alpha)} \right]$.

Método MML:

$$\hat{\alpha} = \frac{l_2}{\ln 2}$$

$$\hat{\beta} = l_1 - 0,5772\hat{\alpha}$$

Distribuição Log-Normal de 2 parâmetros (LNO)

Notação: $X \sim LN(\mu_Y, \sigma_Y)$

Parâmetros: μ_Y e σ_Y , com $Y = \ln(X)$

$$\text{FDP: } f_X(x) = \frac{1}{x \sigma_{\ln(X)} \sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left[\frac{\ln(X) - \mu_{\ln(X)}}{\sigma_{\ln(X)}} \right]^2 \right\} \text{ para } x > 0$$

$$\text{Média: } E[X] = \mu_X = \exp \left[\mu_{\ln(X)} + \frac{\sigma_{\ln(X)}^2}{2} \right]$$

$$\text{Variância: } \text{Var}[X] = \sigma_X^2 = \mu_X^2 \left[\exp(\sigma_{\ln(X)}^2) - 1 \right]$$

$$\text{Coeficiente de Variação: } CV_X = \sqrt{\exp[\sigma_{\ln(X)}^2] - 1}$$

$$\text{Coeficiente de Assimetria: } \gamma = 3 CV_X + (CV_X)^3$$

$$\text{Curtose: } \kappa = 3 + \left(e^{\sigma_{\ln(X)}^2} - 1 \right) \left(e^{3\sigma_{\ln(X)}^2} + 3e^{2\sigma_{\ln(X)}^2} + 6e^{\sigma_{\ln(X)}^2} + 6 \right)$$

Estimação de parâmetros

Método MOM:

$$\hat{\sigma}_Y = \sqrt{\ln(CV_X^2 + 1)}$$

$$\hat{\mu}_Y = \ln \bar{x} - \frac{\hat{\sigma}_Y^2}{2} \quad \text{com } Y = \ln X$$

Método MVS:

$$\hat{\mu}_Y = \bar{y}$$

$$\hat{\sigma}_Y = s_Y$$

Método MML:

$$\hat{\sigma}_Y = 2 \text{erf}^{-1}(t)$$

$\hat{\mu}_Y = \ln l_1 - \frac{\hat{\sigma}_Y^2}{2}$ onde $erf(w) = \frac{2}{\sqrt{\pi}} \int_0^w e^{-u^2} du$. A inversa $erf^{-1}(t)$ é igual a $u/\sqrt{2}$, com u representando a variável Normal padrão correspondente $\Phi[(t+1)/2]$.

Distribuição Log-Pearson Tipo III (LP3)

Aplica-se aqui, as mesmas definições da distribuição de Pearson Tipo III lembrando que se $X \sim PE3$ então, $Z = \ln(X) \sim LP3$.

Estimação de parâmetros

Método MOM:

Lembrando que $\mu'_r = \frac{\exp(\gamma r)}{(1-r\alpha)^\beta}$ são estimados por m'_r , $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$ são as soluções de:

$$\ln m'_1 = \gamma - \beta \ln(1 - \alpha)$$

$$\ln m'_2 = 2\gamma - \beta \ln(1 - 2\alpha)$$

$$\ln m'_3 = 3\gamma - \beta \ln(1 - 3\alpha)$$

Para a solução desse sistema, Kite (1977) sugere:

$$\text{defina } B = \frac{\ln m'_3 - 3 \ln m'_1}{\ln m'_2 - 2 \ln m'_1}, \quad A = \frac{1}{\alpha} - 3 \quad \text{e} \quad C = \frac{1}{B - 3}$$

$$\text{para } 3,5 < B < 6, \quad A = -0,23019 + 1,65262C + 0,20911C^2 - 0,04557C^3$$

$$\text{para } 3,0 < B \leq 3,5, \quad A = -0,47157 + 1,99955C$$

$$\hat{\alpha} = \frac{1}{A + 3}$$

$$\hat{\beta} = \frac{\ln m'_2 - 2 \ln m'_1}{\ln(1 - \hat{\alpha})^2 - \ln(1 - 2\hat{\alpha})}$$

$$\hat{\gamma} = \ln m'_1 + \hat{\beta} \ln(1 - \hat{\alpha})$$

Método MVS:

$\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$ são as soluções (método de Newton) do seguinte sistema:

$$\sum_{i=1}^N (\ln x_i - \gamma) = N\alpha\beta$$

$$N\Psi(\beta) = \sum_{i=1}^N \ln \left[(\ln x_i - \gamma) / \alpha \right]$$

$$N = \alpha (\beta - 1) \sum_{i=1}^N \frac{1}{\ln x_i - \gamma}$$

$$\text{onde } \Psi(\beta) = \frac{\Gamma'(\beta)}{\Gamma(\beta)}$$

Método MML:

As estimativas pelo método MML podem ser obtidas por procedimento idêntico ao ilustrado para a distribuição Pearson Tipo III, com a transformação $z_i = \ln(x_i)$.

Distribuição Pearson Tipo III (PE3)

Notação: $X \sim \text{PIII}(\alpha, \beta, \gamma)$

Parâmetros: α , β e γ

$$\text{FDP: } f_X(x) = \frac{1}{\alpha \Gamma(\beta)} \left(\frac{x - \gamma}{\alpha} \right)^{\beta-1} \exp\left(-\frac{x - \gamma}{\alpha} \right)$$

$$\text{Média: } E[X] = \alpha\beta + \gamma$$

$$\text{Variância: } \text{Var}[X] = \alpha^2 \beta$$

$$\text{Coeficiente de Assimetria: } \gamma = \frac{2}{\sqrt{\beta}}$$

$$\text{Curtose: } \kappa = 3 + \frac{6}{\sqrt{\beta}}$$

Estimação de parâmetros

Método MOM:

$$\hat{\beta} = \left(\frac{2}{g_X} \right)^2$$

$$\hat{\alpha} = \sqrt{\frac{s_X^2}{\hat{\beta}}}$$

$$\hat{\gamma} = \bar{x} - \sqrt{s_X^2 \hat{\beta}}$$

Método MVS:

$\hat{\alpha}, \hat{\beta}, \hat{\gamma}$ são as soluções (método de Newton) do seguinte sistema:

$$\sum_{i=1}^N (x_i - \gamma) = N\alpha\beta$$

$$N\Psi(\beta) = \sum_{i=1}^N \ln[(\ln x_i - \gamma) / \alpha]$$

$$N = \alpha (\beta - 1) \sum_{i=1}^N \frac{1}{x_i - \gamma}$$

$$\text{onde } \Psi(\beta) = \frac{\Gamma'(\beta)}{\Gamma(\beta)}$$

Método MML:

$$\text{Para } t_3 \geq 1/3 \text{ e com } t_m = 1 - t_3, \hat{\beta} = \frac{0,36067t_m - 0,5967t_m^2 + 0,25361t_m^3}{1 - 2,78861t_m + 2,56096t_m^2 - 0,77045t_m^3}.$$

$$\text{Para } t_3 < 1/3 \text{ e com } t_m = 3\pi t_3^2, \hat{\beta} = \frac{1 + 0,2906t_m}{t_m + 0,1882t_m^2 + 0,0442t_m^3}.$$

$$\hat{\alpha} = \sqrt{\pi} l_2 \frac{\Gamma(\hat{\beta})}{\Gamma(\hat{\beta} + 0,5)}$$

$$\hat{\gamma} = l_1 - \hat{\alpha}\hat{\beta}$$

Além das distribuições aqui relacionadas, são de especial interesse as distribuições Kappa e Wakeby. A descrição dessas distribuições podem ser encontradas em Rao & Hamed (2000) e Hosking & Wallis (1997).

ANEXO 2 – SÉRIES DE DURAÇÃO PARCIAL PARA AS ESTAÇÕES PLUVIOMÉTRICAS

02044000 - Ponte do Rio do Peixe (mm)									
182,50	190,50	286,60	198,60	209,20	150,20	286,60	160,50	239,00	297,60
168,90	273,60	418,00	246,80	373,40	397,60	239,20	230,00	206,60	163,80
242,40	185,00	183,20	221,00	143,60	259,00	167,40	173,20	191,80	232,40
340,20	183,40	334,60	354,80	168,00	152,60	243,40	148,40	215,60	207,40
162,00	179,20	175,60							

02044002 - Itaúna - Montante (mm)									
196,90	189,00	206,90	222,20	197,20	159,20	222,10	152,70	183,60	218,20
163,20	165,30	168,00	159,00	151,00	269,10	624,90	163,80	162,40	180,60
186,00	232,00	152,40	252,50	175,70	215,00	210,40	210,00	167,00	198,90
151,20	156,50	157,60	192,70	389,30	196,60	178,60	152,70	195,80	183,50
166,10	238,50	162,00	220,90	193,80	202,90	161,30	153,50	263,30	191,50
156,00	192,40	261,50	269,70	169,70	182,70	161,90	237,90	155,80	318,30
222,90	199,20	356,70	154,80	195,60	152,90	153,70	157,60	197,10	216,20
182,00	156,00	194,00	190,80	162,90	193,70	169,30	233,00	169,90	

02044003 - Carmo do Cajuru (mm)									
182,20	186,30	154,10	174,60	175,40	168,20	293,20	162,80	185,50	189,00
189,00	170,80	194,80	267,00	214,70	279,90	153,90	190,40	166,40	208,30
155,20	154,40	145,60	192,40	247,50	203,20	186,70	158,10	152,50	166,90
203,90	253,60	148,10	249,60	303,10	216,00	213,70	255,20	172,00	226,20
154,90	165,80	155,20	201,50	148,00	200,90	164,20	142,70	225,30	161,70
217,90	181,80	163,60	151,00	187,70	150,60	341,00	175,00	143,80	150,20
220,60	415,40	241,00	201,40	234,80	190,40	232,60	213,80	156,40	218,60
190,60	166,00	148,00	307,40	163,30	147,20	169,40	552,80	304,40	223,20
187,80	314,20	215,80	147,60	173,40	177,50	153,80	237,50	155,40	183,00
206,20	184,00	253,20	188,20	232,50	256,90				

02044005 - Carmo da Mata (mm)									
163,00	191,00	167,80	166,80	189,50	195,80	201,40	206,80	149,20	178,90
147,80	219,00	247,80	236,20	243,90	227,60	145,20	233,50	150,80	148,60
174,10	143,90	216,70	207,10	148,70	173,80	189,30	175,60	289,80	204,70
283,00	194,80	195,70	256,60	181,80	272,80	144,70	237,60	168,70	151,70
172,70	301,60	253,00	201,50	142,00	196,00	171,50	150,90	143,00	163,20
198,40	163,00	191,00	204,10	167,90	254,70	272,90	204,80	155,80	154,40
163,50	142,60	324,20	179,40	226,70	305,80	154,00	392,40	155,30	166,10
145,00	250,40	144,20	291,50	224,70	185,90	198,00	150,00	180,00	160,80
174,20	158,70	148,40	301,10	161,90	181,90	192,10	233,20	182,20	142,30
165,70	185,90	165,20							

02044006 - Divinópolis (mm)									
165,10	207,00	192,20	277,80	182,50	183,00	160,50	157,40	208,80	184,30
295,40	240,70	168,80	240,40	161,90	141,80	200,10	334,70	223,30	333,40
171,80	234,30	165,30	274,30	146,20	154,10	184,70	191,80	139,20	156,20
164,20	164,00	187,10	182,60	139,00	160,70	142,30	203,00	351,50	138,80
142,60	166,40	166,10	144,70	410,70	233,10	219,80	204,60	171,80	190,80
197,40	177,30	154,80	163,00	203,10	145,40	212,40	147,20	293,90	192,50
208,40	190,60	268,40	246,70	136,70	270,90	242,30	190,10	336,00	171,20
157,00	238,20	297,00	200,20	146,00	221,00	264,20	220,80	149,70	304,10
180,90	221,90	233,50	201,80	220,60					

02044009 - Fazenda Campo Grande (mm)									
203,10	152,20	191,00	136,50	241,40	226,60	181,80	137,40	144,00	140,00
187,00	208,40	196,40	272,80	241,60	213,60	194,80	253,10	222,80	135,90
201,20	209,10	197,90	142,80	259,20	234,60	157,20	187,40	145,30	247,40
141,00	170,00	245,60	202,70	239,80	146,40	232,80	200,90	240,10	234,30
173,00	241,50	168,20	199,00	225,00	209,20	170,00	206,90	193,30	142,40
210,40	205,70	151,30	264,00	201,00	169,30	179,20	246,80	146,80	156,10
286,10	216,20	237,80	136,30	221,30	155,20	139,00	167,40	199,10	149,30
166,00	289,30	145,50	205,60	473,00	179,20	220,60	195,00	179,20	276,40
157,20	289,60	154,00	154,00	177,80	139,80	201,20	301,80	228,10	462,40
156,50	145,40	228,80	167,60	180,90	169,50	163,30			

02044011 - Carmópolis de Minas (mm)									
260,00	179,90	217,90	156,60	163,50	204,00	160,60	180,20	174,30	149,90
174,40	301,90								

02044016 - Fazenda Benedito Chaves (mm)									
169,10	150,20	197,90	172,30	169,10	189,40	163,00	169,20	188,40	309,40
238,60	161,40	164,00	168,50	313,00	150,40	152,00	223,20	222,80	185,10
235,20	186,40	323,20	164,00	181,00	206,60	257,70	327,00	157,40	277,00
162,40	240,00	234,50	265,90	161,40	185,80	182,70	439,20	151,20	187,60
238,10	186,60	189,00	152,20	188,80	209,80	176,00	196,20	175,20	274,30
149,80									

02044027 - Ponte Fernão Dias (mm)									
179,40	220,00	212,40	192,20	140,00	182,00	141,00	141,20	129,50	147,10
190,00	168,40	230,40	146,60	162,60	272,70	148,50	178,30	160,40	232,50
157,20	177,60	144,00	168,00	250,00	130,70	135,30	185,00	298,60	179,20
200,80	283,10	143,60	227,80	140,40	237,20	340,00	257,20	187,00	203,00
195,40	303,60	188,70	224,00	450,60	173,40	156,20	149,20	133,00	139,30
172,20	178,10	161,40	150,90						

02044036 - Itaguara (mm)									
184,90	154,30	151,40	161,30	173,80	198,40	260,30	171,60	174,80	194,70
185,20	199,30	213,60	282,90	263,80	199,10	240,10	163,60	193,60	139,00
174,20	151,70	173,10	187,90	202,70	183,10	147,10	159,70	200,00	190,90
272,60	250,20	257,10	202,80	214,50	260,30	137,90	211,70	223,40	176,40
264,40	186,40	220,60	134,20	172,50	196,00	169,60	147,80	146,60	145,20
148,20	178,40	227,00	158,30	318,90	149,00	226,70	190,60	142,00	181,40
144,10	212,30	172,20	159,00	271,00	215,40	148,40	142,60	282,00	157,90
140,60	168,50	217,30	146,80	204,40	246,80	201,00	199,00	140,00	209,10
134,20	194,80	184,20	141,40	253,10	345,60	182,90	155,40	139,00	177,20
173,10	189,90	144,20	166,60						

02044042 - Carmo da Mata [ETA - Copasa] (mm)									
204,10	167,90	254,70	272,90	204,80	155,80	154,40	163,50	142,60	324,20
179,40	226,70	305,80	154,00	392,40	155,30	166,10	145,00	250,40	144,20
291,50	224,70	185,90	198,00	150,00	180,00	160,80	174,20	158,70	148,40
301,10	161,90	181,90	192,10	233,20	182,20	165,70	185,90	165,20	206,90
203,60	285,30								

Chuva média - Ponte do Vilela (mm)									
184,17	147,02	182,31	146,48	219,51	233,43	174,66	150,59	160,69	183,11
202,90	203,08	262,27	227,84	200,38	247,30	196,19	154,84	138,08	171,87
194,23	183,28	224,24	206,57	147,80	193,30	130,92	136,55	140,88	214,05
227,71	189,21	239,73	186,72	234,85	200,23	203,36	240,75	189,26	229,13
149,80	187,38	242,53	214,90	141,61	191,12	166,20	204,50	186,86	136,74
229,67	192,13	160,82	171,46	221,43	151,80	286,76	215,11	208,03	134,60
201,71	137,86	155,83	171,61	165,33	261,25	132,02	165,82	407,97	161,92
180,76	172,73	177,42	249,17	146,69	265,85	169,38	180,93	132,81	178,36
250,25	227,68	412,35	159,48	130,34	130,21	211,57	174,40	179,42	153,80
160,86									

ANEXO 3 – REGRESSÃO LOWESS

A descrição a seguir foi adaptada de Naghettini (1994).

Seja uma amostra com n dados (x_i, y_i) $i = 1, 2, \dots, n$ de um modelo da forma:

$$y_i = g(x_i) + \varepsilon_i \quad (\text{A 3.1})$$

onde, $g(\)$ é a função de regressão e ε_i são os erros, assumidos serem independentes e identicamente distribuídos com média zero e variância constante σ_ε^2 .

Seja \hat{y}_i a estimativa de y_i dada por $g(x_i)$. O valor de \hat{y}_i pela técnica Lowess pode ser obtido seguindo os seguintes passos:

Passo 1 – Seleção do número de pontos a ser usado na regressão local

Seja f um parâmetro da regressão entre 0 e 1 tal que seu valor represente a porcentagem de dados a serem utilizados na regressão. Se f é igual a 1 todos os dados são usados na regressão. Seja r o número de dados a serem utilizados na regressão igual a $f.n$, arredondado para o número inteiro mais próximo. A regressão é feita com os r pontos mais próximos x_i . Defina h_i a partir do r -ésimo ponto próximo de x_i de tal forma que h_i seja o r -ésimo menor valor de $|x_j - x_i|$ para $j = 1, 2, \dots, n$.

Passo 2 – Determinação da função de ponderação

A função de ponderação $W(\)$ é dada pela função tricúbica a seguir.

$$W(z) = (1 - z^3)^3, \quad |z| < 1 \quad (\text{A 3.2})$$

$$W(z) = 0, \quad |z| \geq 1 \quad (\text{A 3.3})$$

A função de ponderação para cada ponto da regressão local em x_i é:

$$w_k(x_i) = W\left(\frac{x_k - x_i}{h_i}\right) \quad (\text{A 3.4})$$

Passo 3 – Regressão linear ponderada

A partir dos pesos calculados na passo anterior é feita a regressão linear ponderada para o i -ésimo ponto a partir do modelo:

$$y_i = \beta_0 - \beta_1 x_i + \varepsilon_i \quad (\text{A 3.4})$$

Os parâmetros β_0 e β_1 para o i -ésimo ponto são encontrados minimizando-se a seguinte expressão:

$$\sum_{k=1}^n w_k(x_i)(y_k - \beta_0 - \beta_1 x_k)^2 \quad (\text{A 3.5})$$

O valor ajustado para y no ponto x_i é:

$$\hat{y}_i = \beta_0 + \beta_1 x_i = \sum_{k=1}^n r_k(x_i) y_k \quad (\text{A 3.6})$$

Os passos 2 e 3 são repetidos para todos os n pontos.

Passo 4 – Cálculo dos pesos robustos

Os pontos \hat{y}_i do passo 3 são estimadores de máxima verossimilhança para $g(x_i)$ assumindo que os erros ε_i são normalmente distribuídos. Se a distribuição de ε_i é desconhecida, uma forma robusta de estimar $g(x_i)$ é dar menos peso aos pontos com maior resíduo. Denotando os resíduos da primeira regressão como

$$e_i = y_i - \hat{y}_i \quad (\text{A 3.7})$$

e definindo s como a mediana de $|e_i|$ os pesos robustos para a regressão podem ser definidos como

$$\delta_k = W\left(\frac{e_k}{6s}\right) \quad (\text{A 3.8})$$

Passo 5 – Regressão linear ponderada com os pesos robustos

A regressão linear é repetida, mas dessa vez as estimativas dos parâmetros β_0 e β_1 é feita pela minimização da seguinte expressão:

$$\sum_{k=1}^n \delta_k w_k(x_i) (y_k - \beta_0 - \beta_1 x_k)^2 \quad (\text{A 3.9})$$

Os passos 4 e 5 são repetidos para todos os pontos n .

Passo 6 – Refinamento da regressão

Usando as novas estimativas de \hat{y}_i do passo 5 os pesos robustos δ_i são recalculados (passo 4) e uma terceira regressão é realizada (passo 5) usando os valores ajustados \hat{y}_i como estimador de $g(x_i)$.

O passo 6 é repetido para todos os n valores da amostra.

ANEXO 4 – PROGRAMA PVP

A metodologia apresentada no capítulo 3 é de difícil aplicação e envolve cálculos complexos e, por vezes, demorados. Além disso, a grande quantidade de dados a serem analisados torna impossível sua aplicação de forma manual. Visando transpor essas dificuldade foi elaborado um pacote computacional que automatiza todas as etapas de cálculo.

O software, denominado PVP, foi construído em ambiente Microsoft Windows® utilizando a linguagem Visual Basic .NET 2002®. Em breve, o software poderá ser obtido através do endereço eletrônico www.ehr.ufmg.br. A figura a seguir mostra a tela de abertura do programa.



Figura A4.1 – Tela de abertura do programa PVP

O programa PVP é formado por três módulos:

- Gerenciamento de estações de monitoramento
- Relações Pico-Volume
- Quantis

Gerenciamento de estações de monitoramento

Neste módulo as estações fluviométricas e pluviométricas a serem utilizadas no cálculo dos quantis são adicionadas ou excluídas do banco de dados do programa. A figura a seguir mostra a tela com o módulo de gerenciamento de estações.

Gerenciador de estações fluviométricas e pluviométricas

Escolha a estação: 40130001 - Ponte do Vilela Tipo de Dado: Fluviométricos

Data	Dia 1	Dia 2	Dia 3	Dia 4	Dia 5	Dia 6	Dia 7	Dia 8
jul/1938	21,20	21,20	21,20	21,20	20,90	20,90	20,90	20,90
ago/1938	17,20	17,20	16,60	16,60	17,70	18,30	21,20	27,40
set/1938	15,70	15,40	15,40	15,40	15,40	16,60	24,20	34,30
out/1938	20,30	17,20	16,00	15,70	16,00	32,80	29,70	22,70
nov/1938	21,80	20,60	23,30	22,40	20,60	20,10	20,10	17,70
dez/1938	34,30	49,20	184,00	123,00	52,20	43,80	42,30	51,30
jan/1939	169,00	121,00	103,00	95,60	71,20	43,40	39,30	38,60
fev/1939	132,79	112,81	88,32	76,28	71,56	66,30	62,88	61,57
mar/1939	52,07	50,75	47,31	45,72	44,93	43,87	42,54	41,75
abr/1939	39,35	40,42	38,82	38,29	38,29	38,29	38,03	36,96
mai/1939	28,96	30,56	30,83	30,56	28,96	29,22	28,96	28,96
jun/1939	22,57	22,39	22,22	22,57	22,04	21,87	21,87	21,69
jul/1939	21,87	25,07	22,92	21,01	19,83	19,67	19,50	19,18
ago/1939	18,34	18,18	17,53	17,23	16,91	16,58	16,42	16,29
set/1939	13,98	13,68	14,14	14,28	15,04	15,34	15,04	14,44
out/1939	14,00	21,20	37,90	28,00	33,60	33,60	25,40	21,20
nov/1939	11,90	11,70	10,70	10,70	10,70	10,70	11,40	12,40
dez/1939	20,60	28,70	25,40	23,60	20,30	16,00	25,40	44,20
jan/1940	31,80	31,40	28,70	34,60	34,30	24,50	23,60	31,40
fev/1940	55,70	45,70	37,90	43,40	50,40	68,70	95,00	90,60
mar/1940	29,40	29,70	30,00	50,90	35,00	30,40	30,40	33,60
abr/1940	36,80	35,70	34,60	32,10	37,20	37,50	32,80	32,10
mai/1940	24,50	23,60	24,80	46,10	36,80	26,40	24,50	24,80
jun/1940	23,30	21,50	20,60	19,20	18,90	18,90	18,30	18,60

Figura A4.2 – Gerenciamento de estações de monitoramento

Relações Pico-Volume

Neste módulo o usuário pode adicionar, visualizar ou editar as relações pico-volume. Ao adicionar uma novo modelo pico-volume o usuário deve especificar os vários parâmetros que o definem. Todos os modelos têm a forma da equação 3.86 e o parâmetro b pode ter a forma logarítmica, como na equação 3.90, linear, exponencial ou potencial. A seguir é mostrada a tela onde foi definido a relação pico-volume dada pela equação 3.92.

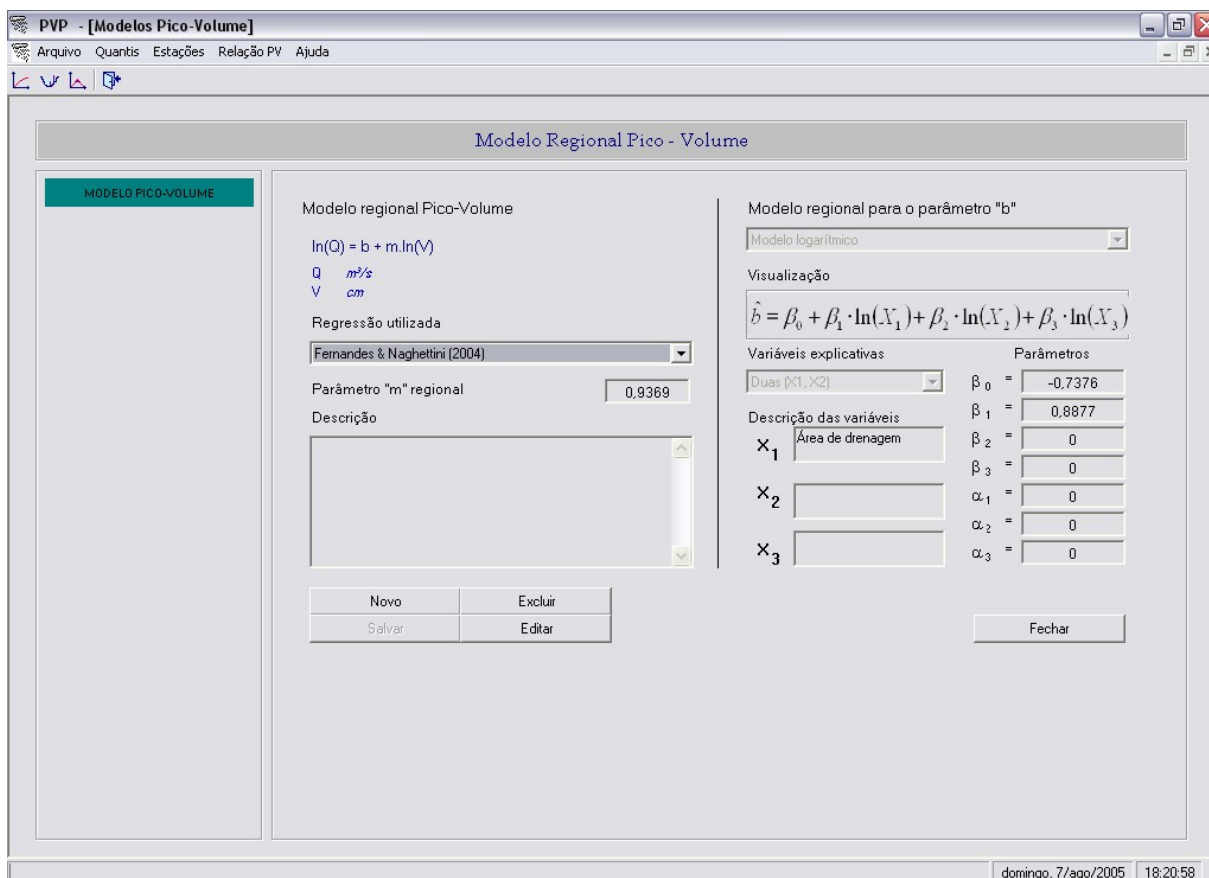


Figura A4.3 – Módulo de gerenciamento de relações pico-volume

Quantis

Neste módulo a metodologia desenvolvida no capítulo 3 é aplicada conforme as especificações do usuário. O módulo conta com quatro partes:

- Dados principais

Nesta parte são especificadas as características gerais do projeto, tais como: Nome do projeto, Bacia analisada, Tempo de base da bacia, Capacidade de infiltração e/ou T_c (tempo de retorno a partir do qual são válidas as suposições do método GRADEX).

A figura A4.4 mostra a tela referente a esta parte. Na tela são mostradas as entradas utilizadas para a geração dos resultados mostrados no Capítulo 4.

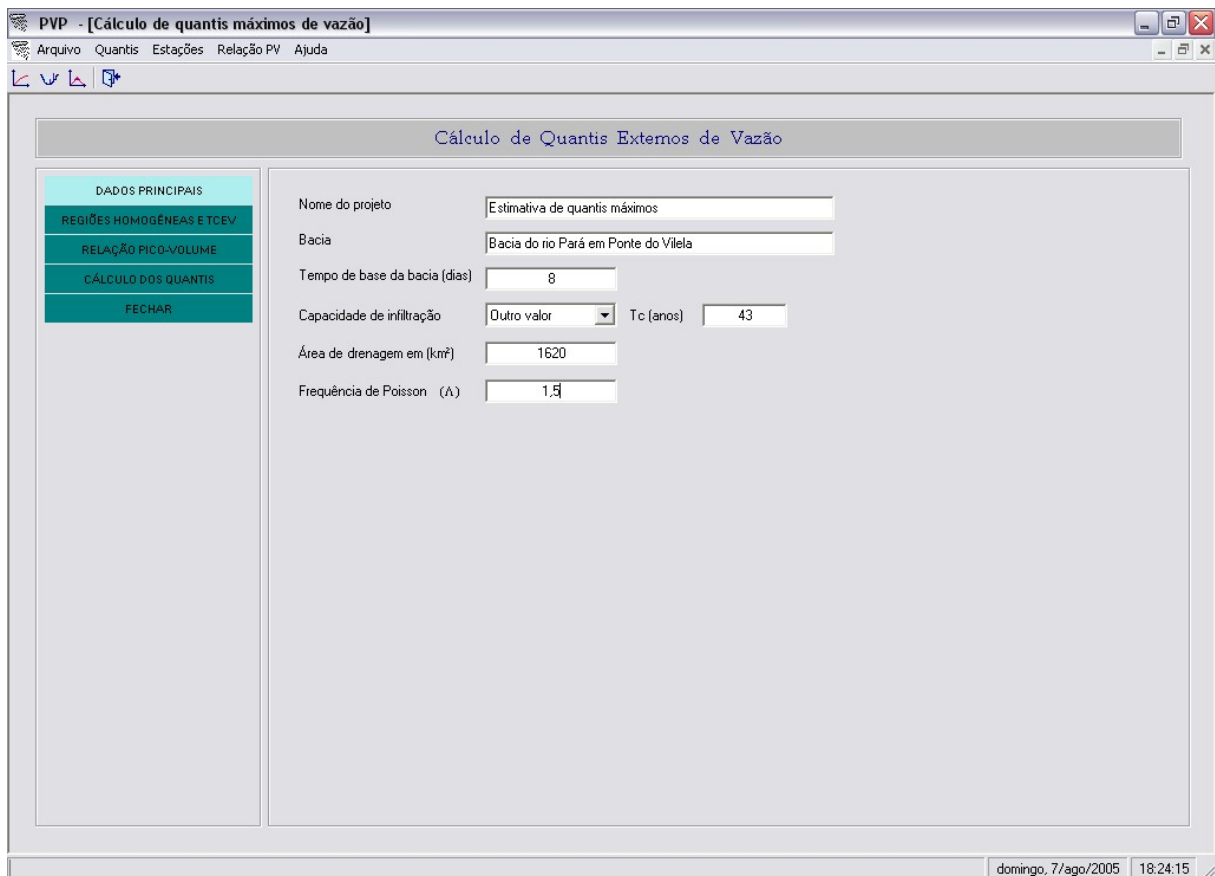


Figura A4.4 – Dados principais do projeto

- Regiões homogêneas e TCEV

Nesta parte são especificadas as estações pluviométricas da região homogênea utilizadas para a regionalização pela TCEV e como consequência para o cálculo do gradex da chuva. O usuário tem a opção de especificar o gradex da chuva e não realizar a regionalização. Isso é especialmente útil quando se deseja fazer análise sobre várias estações pluviométricas dentro de uma mesma região homogênea. Neste caso é necessário proceder uma única regionalização pela TCEV, o que economiza muito tempo de processamento.

Outro dado a ser fornecido neste ponto é a série diária de chuva média sobre a bacia de interesse. Recomenda-se o uso de polígonos de Thiessen para tal. O usuário tem a opção ainda de eliminar os outliers da série por meio da estatística de Grubbs & Beck.

A seguir é mostrada a entrada de dados que gerou os resultados da regionalização de chuva mostrada no Capítulo 4.

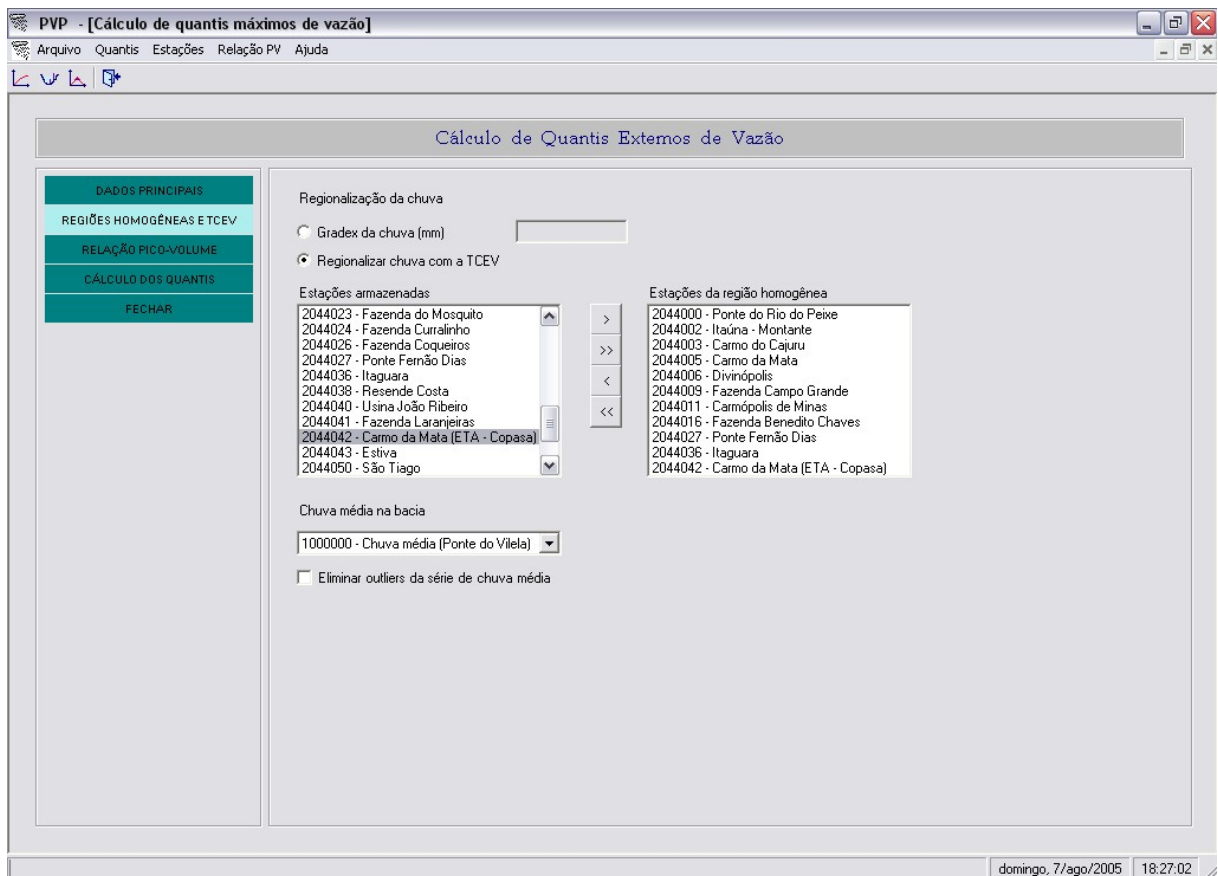


Figura A4.5 – Dados para a regionalização pela TCEV

- Relação pico-volume

Nesta parte deve ser especificada qual relação pico-volume deverá ser utilizada para o cálculo dos quantis. Deve ser especificada, também, as variáveis explicativas da relação pico-volume selecionada.

Nesta parte é especificada a estação fluviométrica que servirá de referência para o cálculo dos quantis. É a partir desta estação que são obtidos os pares pico-volume observados. Com esses dados o programa ajusta a distribuição dos volumes e constrói a relação pico-volume na faixa dos valores observados por meio da regressão Lowess.

O erro padrão da estimativa da relação pico-volume pode ser calculada a partir dos dados observados ou especificado pelo usuário. A seguir é mostrada a tela de entrada de dados referente a esta parte do programa. Os dados mostrados são os mesmos utilizados no capítulo 4.

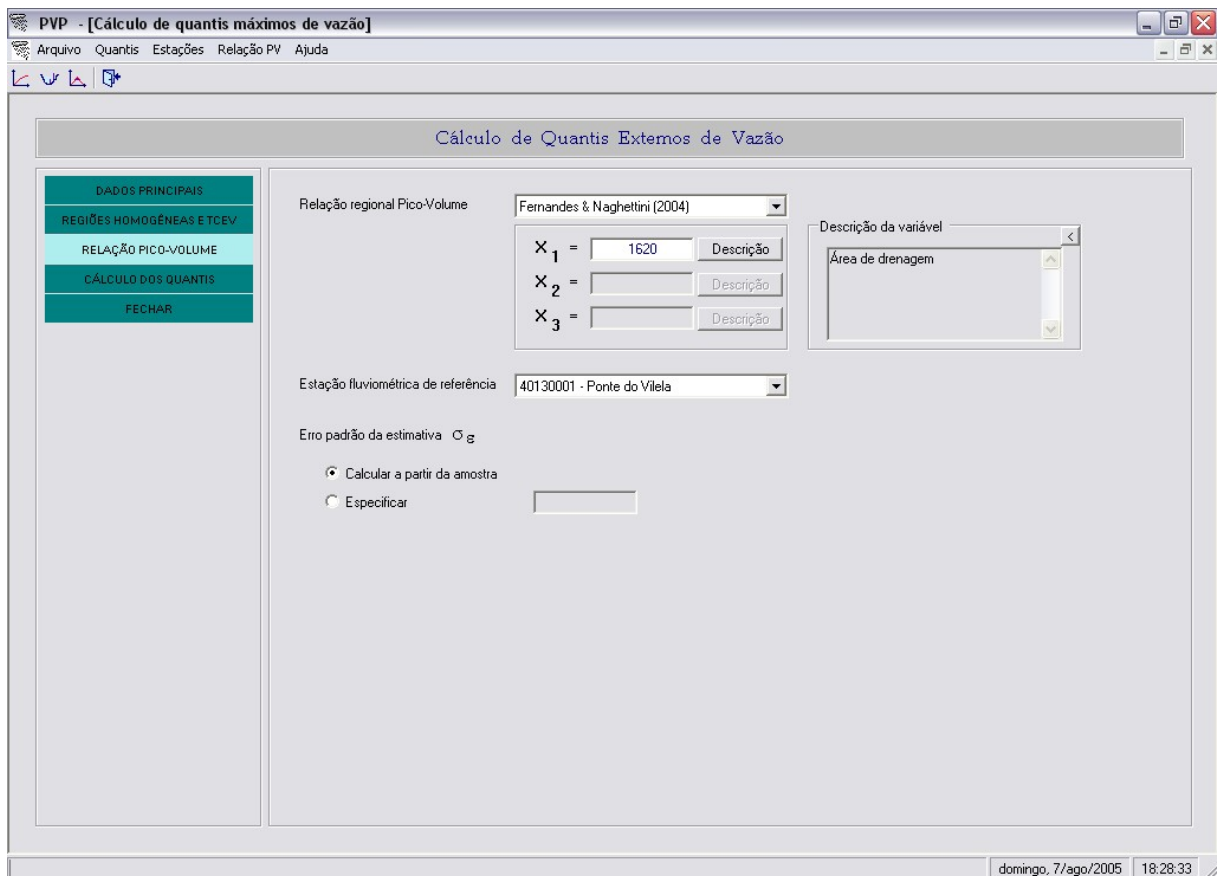


Figura A4.6 – Entrada de dados para a relação pico-volume

- Cálculo de quantis

A única variável a ser fornecida neste ponto é o período de retorno máximo para o qual se deseja calcular a curva de quantis. A seguir o programa faz uma aplicação completa da metodologia desenvolvida no capítulo 3. A figura A4.7 mostra a tela com os resultados gerados.

O tempo de processamento depende de vários fatores, sendo os principais aqueles relacionados com a quantidade e qualidade dos dados de entrada, período de retorno máximo de cálculo e capacidade de processamento do computador utilizado.

Com os dados de entrada referenciados no capítulo 4 o tempo de processamento foi de aproximadamente, 9,5 minutos para um período de retorno máximo de 10.000 anos. O computador utilizado no processamento tem a seguinte configuração básica:

Processador: AMD Athlon XP 2400+

Clock: 2,00 GHz

Memória: 512 MB (RAM)

Sistema: Windows XP® 2002 – SP2

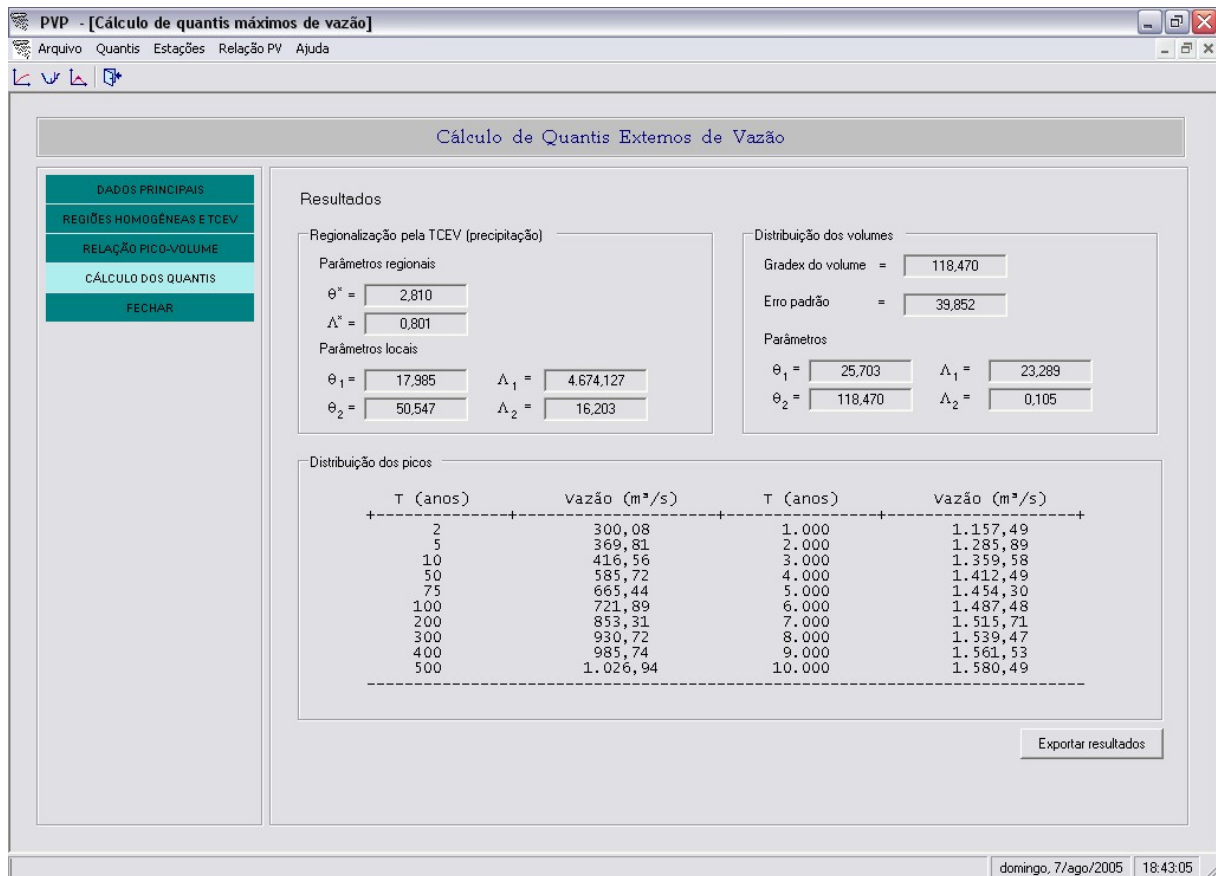


Figura A4.7 – Resultados gerados pelo programa PVP