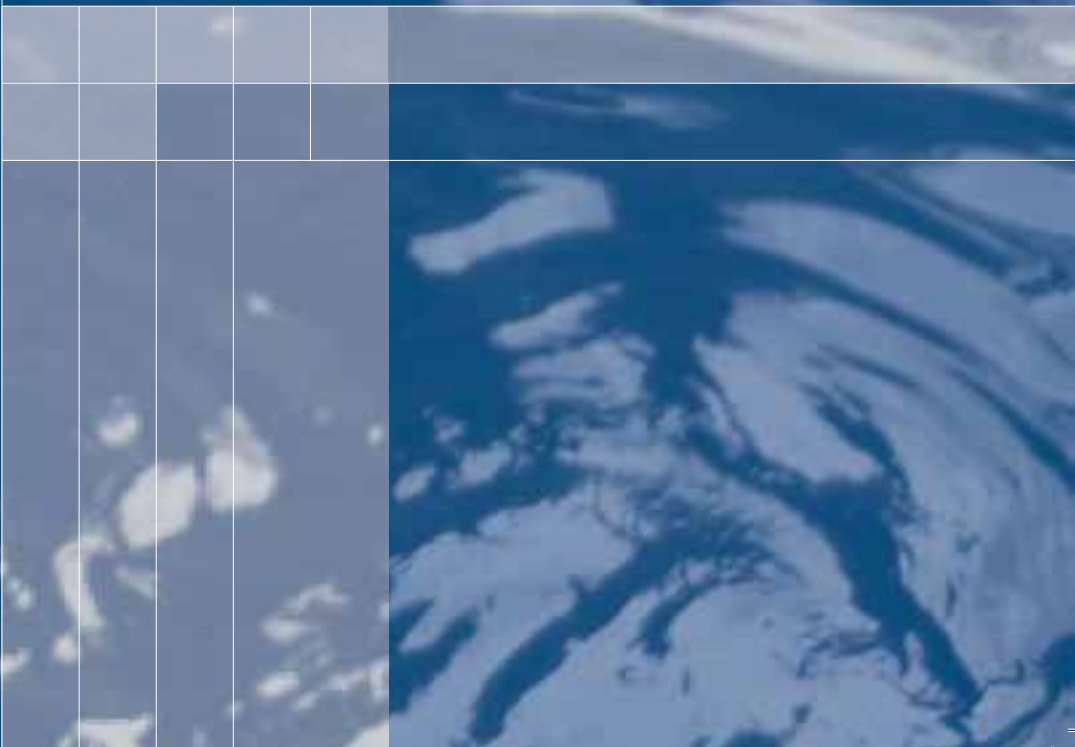




CAPÍTULO 8



**ANÁLISE LOCAL DE FREQUÊNCIA
DE VARIÁVEIS HIDROLÓGICAS**





ANÁLISE LOCAL DE FREQUÊNCIA DE VARIÁVEIS HIDROLÓGICAS

Os sistemas hidrológicos podem ser impactados por eventos extremos, tais como tempestades, grandes cheias e secas. A magnitude de um evento extremo é inversamente relacionada à sua frequência de ocorrência, uma vez que os eventos muito severos ocorrem com menor frequência do que os mais moderados. O objetivo da análise de frequência das variáveis hidrológicas é relacionar a magnitude dos eventos com sua frequência de ocorrência por meio do uso de uma distribuição de probabilidade. Os resultados da análise de frequência são necessários para a solução de vários problemas de engenharia, tais como, por exemplo, os projetos de vertedores de barragens, pontes, bueiros e estruturas de controle de cheias, além de problemas que envolvem a estimativa de algum valor característico, tal como a vazão mínima com 7 dias de duração e 10 anos de tempo de retorno.

Do ponto de vista da extensão espacial das informações envolvidas, a análise de frequência pode ser classificada em local ou regional. Na análise de frequência local, a definição dos quantis de interesse, ou seja, dos valores da variável hidrológica associados a certas probabilidades de excedência, é realizada utilizando uma única série de registros hidrométricos ou hidrometeorológicos, observados em certa estação fluviométrica, ou pluviométrica, ou climatológica. No caso da análise regional, são utilizados os dados de várias estações de uma região geográfica. Na análise regional, as informações podem ser agrupadas em subconjuntos que apresentam semelhanças fisiográficas, climáticas e/ou estatísticas, as quais justificam a transferência de informações de um local para outro, dentro daquele contexto geográfico. Atualmente, a utilização da análise de frequência regional não se restringe apenas à estimativa de variáveis hidrológicas em locais que não possuem uma coleta sistemática de informações, mas também para aumentar a confiabilidade das estimativas dos parâmetros de uma distribuição de probabilidades, para identificar a ausência de postos de observação em partes de uma região, bem como para verificar a consistência das séries hidrológicas. A análise de frequência regional será abordada, em maior profundidade, no capítulo 10.

A análise de frequência, tanto local como regional, pode ser realizada a partir das chamadas séries de duração anual ou de duração parcial. As séries de duração anual, ou séries anuais, são formadas por um único valor para cada ano de observações, tendo como referência temporal o ano hidrológico ou o ano civil, a depender da variável hidrológica sob análise. As séries de duração parcial consistem

das observações independentes de magnitude superior (ou inferior) a certo valor limiar de referência. Por exemplo, em uma determinada estação fluviométrica, as vazões de pico, extraídas de hidrogramas de cheia convenientemente isolados e independentes entre si, e que superaram um valor limiar especificado, irão formar uma série de duração parcial, a qual pode ter um número de elementos superior ou inferior ao da série anual, a depender da especificação do valor limiar. Existem relações importantes entre a distribuição de probabilidade para máximos anuais e a frequência de eventos em uma série de duração parcial, as quais serão examinadas ao final do presente capítulo.

As séries constituídas para a análise de frequência devem ser representativas da variável em questão, não apresentando erros de observação ocasionais e/ou sistemáticos, além de possuir um número suficiente de elementos que permita realizar extrapolações confiáveis. Além disso, na análise de frequência, é necessário que os dados sejam homogêneos e independentes. A condição de homogeneidade pretende assegurar que todas as observações tenham sido extraídas de uma única população. Para o caso de análise de vazões, por exemplo, pretende-se assegurar que o uso e a ocupação da bacia não tenham sido modificados ou, ainda, que não tenham sido implantadas estruturas hidráulicas que hajam modificado o escoamento natural nos cursos d'água. Por outro lado, a condição de independência procura assegurar que não exista dependência serial entre os elementos que constituem a série. Para efeito de ilustração, considere o caso de uma bacia hipotética, situada na região sudeste do Brasil, sobre a qual abateu-se uma precipitação duradoura, que resultou da formação de uma zona de convergência do Atlântico Sul (ZCAS), dando origem a dois ou mais eventos de cheia. Nesse caso, para garantir a hipótese de independência, apenas a vazão de pico da enchente de maior magnitude deve ser representada na série.

A análise de frequência pode ser realizada de modo analítico, caso se admita que uma função paramétrica descreva o comportamento probabilístico da variável hidrológica. A análise de frequência também pode ser efetuada de modo empírico. Nesse último caso, o analista grafá as observações ordenadas contra uma escala de probabilidades e utiliza seu melhor julgamento para determinar a associação entre as magnitudes de ocorrências passadas, ou eventos hipotéticos, e os respectivos tempos de retorno. Na análise de frequência analítica de variáveis hidrológicas, além dos problemas afetos à inferência estatística, surge ainda a questão de identificação do modelo distributivo a ser empregado. Em algumas aplicações da estatística, nas quais é possível conhecer *a priori* a distribuição populacional da variável aleatória sob análise, o problema se restringe à estimação dos parâmetros populacionais a partir dos dados amostrais. Porém, em se tratando de variáveis hidrológicas, para as quais não se conhece *a priori* a distribuição

populacional, é ainda necessário especificar um certo modelo distributivo, o qual deve ser capaz de descrever o comportamento probabilístico da variável analisada. De fato, várias distribuições têm sido propostas para a modelagem estatística das variáveis hidrológicas, não havendo, todavia, uma distribuição específica consensual que seja capaz de, sob quaisquer condições, descrever o comportamento da variável em foco. Em suma, em uma análise de frequência típica, o analista procura selecionar, dentre as diversas distribuições candidatas, aquela que parece ser a mais capaz, por um lado, de sintetizar as principais características estatísticas amostrais, e, por outro, de predizer quantis hipotéticos com confiabilidade razoável. De modo resumido, as etapas para análise de frequência local são as seguintes:

- Optar pela utilização de séries anuais ou séries de duração parcial.
- Avaliar os dados das séries, quanto aos atributos de homogeneidade, independência e representatividade.
- Propor uma ou algumas distribuições teóricas de probabilidade, com a estimativa de seus respectivos parâmetros, quantis e intervalos de confiança, seguida da verificação de aderência à distribuição empírica.
- Realizar a identificação e tratamento de eventuais pontos atípicos, com possível repetição de algumas etapas precedentes.
- Selecionar o modelo distributivo mais apropriado.

Os procedimentos de realização da análise local de frequência de variáveis hidrológicas serão analisados em detalhes nos itens que se seguem. Inicialmente serão descritos os métodos para a construção de papeis de probabilidade, bem como algumas técnicas utilizadas na estimação das probabilidades de eventos observados, as quais são etapas importantes da análise de frequência empírica.

8.1 – Análise de Frequência com Gráficos de Probabilidade

A análise de frequência hidrológica local pode ser realizada com ou sem a hipótese de que os dados amostrados sejam oriundos da população de uma determinada distribuição de probabilidades. Em não se admitindo tal hipótese, a análise de frequência se restringe a grafar, ou plotar, os pares constituídos pelas frequências empíricas e pelas observações devidamente ordenadas.

Nessa análise gráfica, a associação das observações ordenadas às respectivas probabilidades empíricas de excedência, ou aos respectivos tempos de retorno, apresenta consideráveis incertezas que dependem, principalmente, do tamanho e representatividade da amostra. Há ainda a incerteza posta pela questão de extrapolação para tempos de retorno muito superiores ao número de anos de

observações amostrais. Essas incertezas podem ser parcialmente reduzidas (a) pela construção dos chamados papéis “de probabilidades” e/ou (b) pela definição de critérios para estimar as probabilidades empíricas associadas às observações amostrais. Esses tópicos são os objetos dos itens que se seguem.

8.1.1 – Construção de Papéis de Probabilidade

Os gráficos podem ser traçados em escalas aritméticas ou em escalas transformadas. De modo geral, quando se plota, em escala aritmética uma certa função acumulada de probabilidades, $F_X(x)$, versus o valor da variável aleatória X , têm-se um gráfico não linear, tal como exemplificado na Figura 8.1.

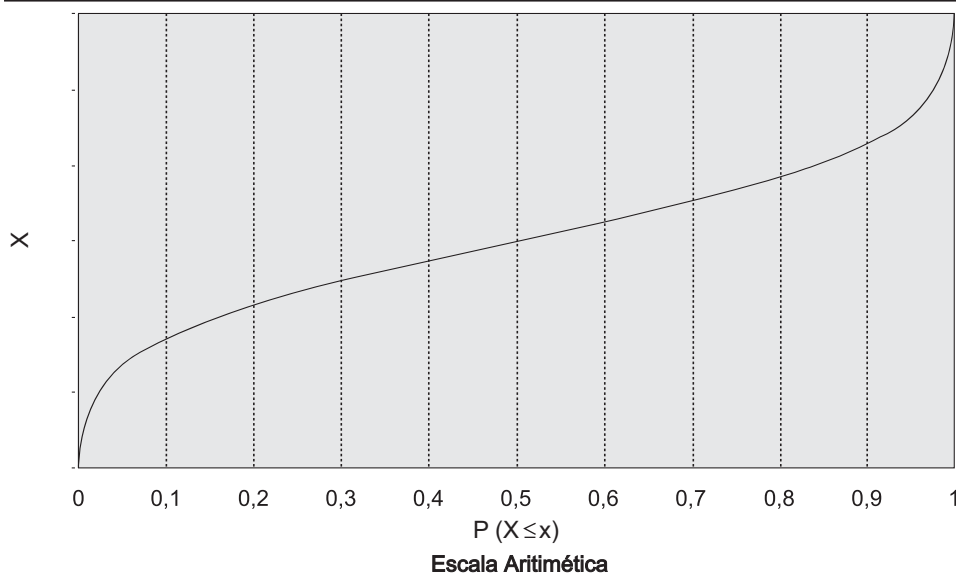


Figura 8.1 – Distribuição Normal em escala aritmética

Os papéis de probabilidade são gráficos para plotagem de observações amostrais e suas respectivas probabilidades empíricas, cujas escalas são previamente transformadas de modo a linearizar a relação entre $F_X(x)$ (ou $[1 - F_X(x)]$) e X , tal como ilustrado na Figura 8.2. A escala apropriada para a linearização de uma certa função acumulada de probabilidades $F_X(x)$, descrita por não mais de dois parâmetros, é geralmente construída por meio da *variável padrão* ou *variável reduzida da distribuição*. A verificação visual de linearidade de um conjunto de dados amostrais, plotados em um papel de probabilidades, pode ser empregada para aceitar ou rejeitar, ainda que empiricamente, a hipótese de aderência a um certo modelo de distribuição de probabilidades.

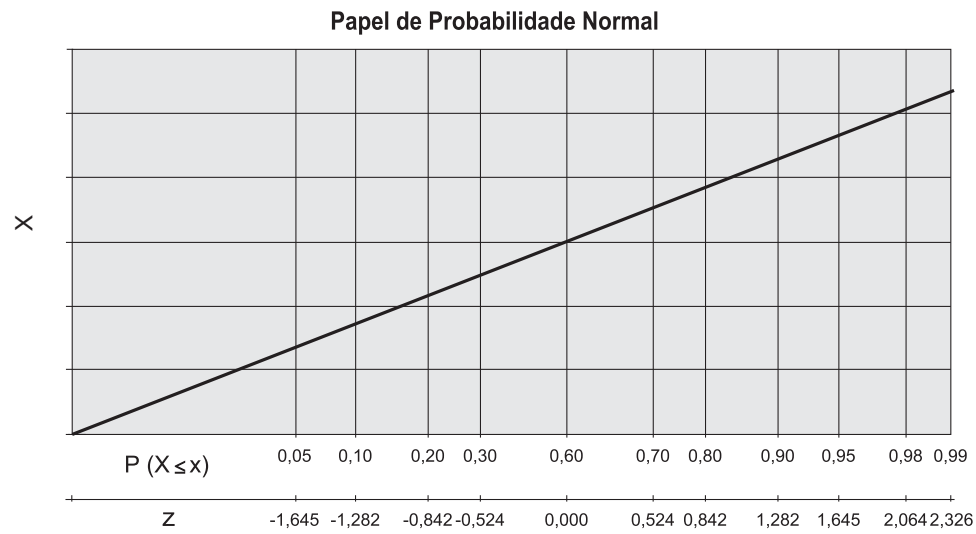


Figura 8.2 – Distribuição Normal no papel de probabilidade Normal

Exemplo 8.1 - Construir o papel de probabilidades da distribuição Normal. Solução: O papel de probabilidade Normal, ou Gaussiano, é construído com base na *distribuição normal padrão* (ver capítulo 5; equações 5.12 e 5.13, e a Tabela 5.1). O eixo das ordenadas, em escala aritmética, representa o valor da variável X , como está ilustrado na Figura 8.2. O eixo das abscissas é composto por duas escalas paralelas, uma em escala aritmética que representa os valores da *variável normal central reduzida*, Z , enquanto que a outra escala mostra os valores da *distribuição normal padrão*, $\Phi(Z)$, correspondentes aos valores de Z , tal como está apresentado na Figura 8.2. Como foi visto no capítulo 5, quando uma variável é normalmente distribuída, o quantil é calculado pela relação $x = \mu_x + Z \cdot \sigma_x$, a qual é a equação de uma reta, onde σ_x é o coeficiente angular e μ_x é o coeficiente linear. A Tabela 8.1 apresenta alguns valores de Z e $\Phi(Z)$ da Tabela 5.1, para a construção do eixo das abscissas. Portanto, para construir um gráfico de probabilidades Normal basta plotar a *variável normal central reduzida*, $Z = [(x - \mu) / \sigma]$, a qual está associada a uma probabilidade de não excedência da *distribuição normal padrão* [$\Phi(Z)$], versus x . Em geral, omite-se a escala da *variável normal central reduzida*, Z .

Tabela 8.1 - Valores de Z e $\Phi(Z)$ para construção do papel normal

Z	-1,645	-1,282	-0,842	-0,524	0,000	0,524	0,842	1,282	1,645	2,054	2,326
$P(X \leq x) = \Phi(Z)$	0,05	0,10	0,20	0,30	0,50	0,70	0,80	0,90	0,95	0,98	0,99

Exemplo 8.2 - Construir o papel de probabilidades da distribuição exponencial

Solução: A construção de um papel de probabilidade passa pela definição de uma *variável padrão* apropriada para uma função acumulada de probabilidades, que permita a linearização do gráfico $F_X(x)$ versus x . No caso da distribuição exponencial temos que a FAP é dada por:

$$F_X(x) = 1 - \exp(-\lambda \cdot x) \quad (8.1)$$

A equação anterior pode ser linearizada por anamorfose logarítmica, ou seja,

$$\exp(-\lambda \cdot x) = 1 - F_X(x) \quad (8.2)$$

$$\ln[\exp(-\lambda \cdot x)] = \ln[1 - F_X(x)] \quad (8.3)$$

$$-\lambda \cdot x = \ln[1 - F_X(x)] \quad (8.4)$$

$$x = \frac{1}{\lambda} \{-\ln[1 - F_X(x)]\} \quad (8.5)$$

Assim, para uma variável aleatória exponencialmente distribuída, plotando $\{-\ln[1 - F_X(x)]\}$, a variável padrão, versus x , obtém-se uma reta com coeficiente angular igual a $1/\lambda$. De forma que, no papel da distribuição exponencial, $\{-\ln[1 - F_X(x)]\}$ é grafado nas abscissas e x nas ordenadas, como ilustrado na Figura 8.3.

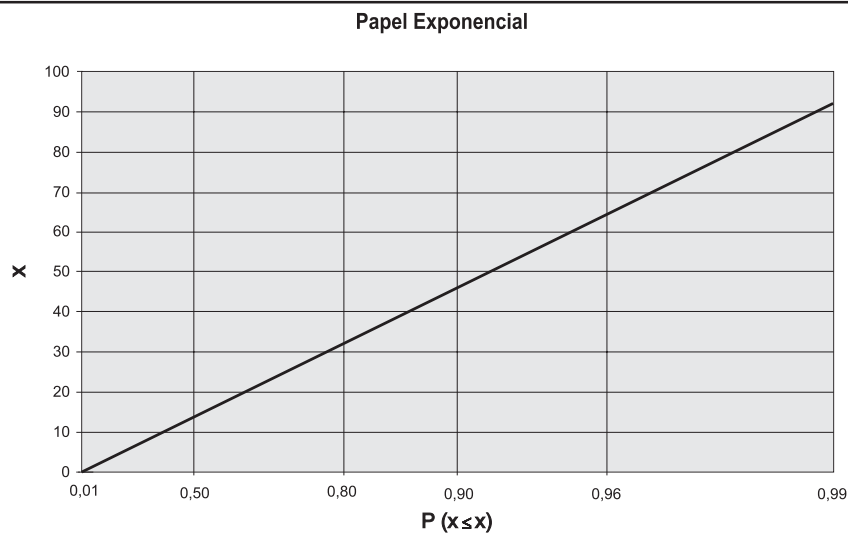


Figura 8.3 – Papel de probabilidade Exponencial

8.1.2 – Posição de Plotagem

Conforme definição anterior, um gráfico de probabilidade associa as magnitudes das observações ordenadas às suas respectivas probabilidades empíricas. No caso de eventos máximos, a estimação da probabilidade empírica de excedência, associada a um certo ponto, é geralmente referida como a determinação da posição de plotagem, a qual pode ser expressa como uma fração entre 0 e 1, ou como uma porcentagem entre 0 e 100. No caso da população, a determinação da posição de plotagem é meramente um problema de determinação da “fração” da população cujos valores são maiores ou iguais ao valor em questão. Assim, para uma variável aleatória de máximos, o menor valor da população terá uma posição de plotagem igual a 1 (um) e o maior valor terá uma posição de plotagem igual a 0 (zero). A definição das posições de plotagem para dados amostrais não é tão direta como no caso populacional, uma vez que nunca haverá certeza de que a amostra contém o maior e/ou o menor valor da população. Assim, para dados amostrais, as posições de plotagem 0 e/ou 1 devem ser evitadas, à exceção dos casos em que existirem informações definitivas acerca dos limites populacionais. Em síntese, para o caso de máximos, uma fórmula para a estimação da posição de plotagem deve especificar a frequência com que um, entre n valores ordenados de modo decrescente, será igualado ou superado.

A estimação da posição de plotagem de dados hidrológicos requer observações individuais independentes entre si e representativas da população. Gumbel (1958) estabeleceu os seguintes critérios para definição das fórmulas para estimativa das posições de plotagem:

- A posição de plotagem deve ser tal que todas as observações possam ser plotadas;
- A posição de plotagem deve estar compreendida entre $(i-1)/n$ e i/n , onde i denota a ordem de classificação de uma amostra ordenada de tamanho n ;
- No caso de séries anuais, o tempo de retorno de um valor maior ou igual à maior observação (ou menor ou igual à menor observação) deve convergir para n , para valores elevados de n .
- As observações devem ser igualmente espaçadas na escala de frequências;
- A posição de plotagem deve ser intuitiva, analiticamente simples e fácil de usar.

Várias fórmulas de posição de plotagem têm sido apresentadas na literatura, as quais, geralmente, produzem valores similares no centro da distribuição, mas variam consideravelmente nas caudas. Algumas dessas fórmulas e seus atributos de aplicação encontram-se apresentados na Tabela 8.2.

As expressões apresentadas na Tabela 8.2 são casos particulares da fórmula mais geral introduzida por Cunnane (1978):

$$q_i = \frac{i - a}{n + 1 - 2a} \tag{8.6}$$

onde a é uma constante que pode ter diferentes valores em conformidade com as hipóteses distributivas. Se $a = 0$, obtém-se a fórmula de Weibull; se $a = 0,44$, a fórmula de Gringorten; se $a = 0,375$, a fórmula de Blom; se $a = 0,50$, a fórmula de Hazen e se $a = 0,40$, a fórmula de Cunnane.

De uma forma geral, as séries hidrológicas de valores máximos ou médios anuais são ordenadas de forma decrescente, o que faz com que a posição de plotagem represente a probabilidade da variável X ser maior ou igual a um certo quantil x , ou seja, $P(X \geq x)$.

Entretanto, quando os valores de uma série são ordenados de forma crescente, como na análise de valores mínimos anuais, a posição de plotagem denota a probabilidade de não-excedência, ou seja, a probabilidade da variável X ser menor ou igual a x , ou seja, $P(X \leq x)$.

Tabela 8.2 – Fórmulas para estimativa das posições de plotagem

Fórmula	Autor	Atributos de aplicação
$q_i = \frac{i}{n+1}$	Weibull	Probabilidades de excedência não enviesadas para todas as distribuições
$q_i = \frac{i - 0,44}{n + 0,12}$	Gringorten	Usada para quantis das distribuições de Gumbel e GEV
$q_i = \frac{i - 0,375}{n + 0,25}$	Blom	Quantis não enviesados para as distribuições Normal e Log-Normal
$q_i = \frac{i - 0,5}{n}$	Hazen	Usada para quantis da distribuição Gama de 3 parâmetros
$q_i = \frac{i - 0,40}{n + 0,20}$	Cunnane	Quantis aproximadamente não enviesados para todas as distribuições

i é posição na amostra ordenada e n é o tamanho da amostra

A estimativa do conjunto das posições de plotagem dos eventos observados, chamada de distribuição empírica, permite a elaboração de um gráfico de probabilidades em conformidade com as seguintes etapas:

- a) classificação dos dados em ordem decrescente (análise de máximos) ou crescente (análise de mínimos);
- b) cálculo da posição de plotagem por uma das fórmulas apresentadas na Tabela 8.2;

- c) seleção do tipo de gráfico, em escala aritmética ou papel de probabilidades apropriado; e
- d) plotagem dos pares $[q_p, x_i]$, formando o gráfico da distribuição empírica.

Quando são plotadas as distribuições empíricas dos dados hidrológicos, freqüentemente, um ou dois eventos extremos da amostra parecem ter comportamento atípico em relação aos outros pontos amostrais, como pode ser visto na Figura 8.4. Nessa figura, estão plotadas as alturas diárias de chuva máximas anuais, por ano hidrológico, da estação pluviográfica de Caeté (MG), código 01943010. Foram utilizados 47 máximos anuais na montagem da série (41/42 a 99/2000), sendo que o maior valor é de 210,2mm, registrado em 15/02/1978, e a segunda maior precipitação diária é de 147,1mm. Por meio da fórmula de Gringorten, o tempo de retorno empírico para a precipitação de 210,2mm é de 84 anos, o qual foi estimado pelo inverso da probabilidade de excedência de excedência com $i = 1$ e $n = 47$. Entretanto, observa-se no gráfico da Figura 8.4 que este evento deveria estar associado a um tempo de retorno maior, caso fosse mantida a tendência do alinhamento dos dados amostrais. Trata-se, nesse exemplo específico, de uma observação atípica em relação àquele conjunto particular de observações amostrais. Esse comportamento atípico pode decorrer de diversas razões, entre as quais, podem ser citadas a eventual existência de erros grosseiros de medição ou, ainda, a associação de uma probabilidade empírica incorreta àquela observação específica, como resultado do pequeno tamanho da amostra. Nesse último caso, supondo que a série de Caeté tivesse, digamos, 150 anos de observações e que, ainda assim, a altura de chuva de 210,2 mm continuasse sendo o maior valor amostral, o seu tempo de retorno seria de 268 anos, pela fórmula de Gringorten. Esse exemplo hipotético demonstra a incerteza intrínseca à associação de tempos de retorno empíricos às observações amostrais. O tratamento desses *outliers* é uma questão não resolvida e controvertida, sendo freqüente a ocorrência de tais observações em amostras de dados hidrológicos. De fato, como foi visto anteriormente, a probabilidade de um evento de T anos de tempo de retorno, ocorrer pelo menos uma vez em um período de n anos, é calculada pela relação $1 - (1 - 1/T)^n$. Assim, a probabilidade de ocorrer pelo menos um evento de 100 anos de tempo de retorno T , durante um período de observação de 30 anos, é de 0,26 ou 26%.

Ao se grafar a distribuição empírica, em certo um papel de probabilidades, os pares $[q_p, x_i]$ podem apresentar uma tendência a se alinharem ao longo de uma reta, a qual pode ser parcimoniosamente extrapolada para tempos de retorno

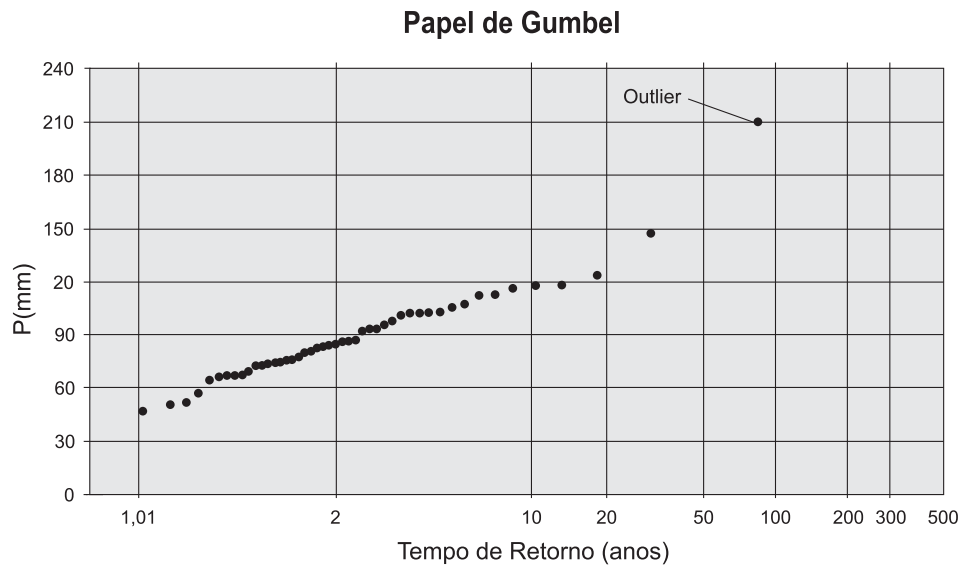


Figura 8.4 – Série com presença de pontos atípicos

superiores àquele associado à maior observação. Todavia, tal situação é pouco freqüente na análise de freqüência de séries hidrológicas. Em geral, os pares $[q_r, x_i]$ apresentam curvaturas e irregularidades que se devem a fatores diversos, entre os quais, os mais importantes são (a) a inadequação do modelo distributivo implicitamente definido pela seleção do papel de probabilidades e (b) problemas de amostragem decorrentes das séries hidrológicas de curta extensão. Tais dificuldades impõem um limite ao uso da análise de freqüência com gráficos de probabilidade, principalmente, quando os quantis de tempos de retorno muito elevados são o principal interesse.

8.1.3 – Posição de Plotagem de Eventos Históricos

As observações sistemáticas de níveis d'água e as medições de vazões nos rios brasileiros tiveram início nos princípios do século XX, com as décadas de 40, 50, 60 e 70 sendo os períodos de maior expansão da rede fluviométrica. Atualmente, em algumas regiões do país, existe um número razoável de séries fluviométricas cujas extensões variam de 30 a 60 anos de observações. Todavia, em alguns locais, é possível obter informações sobre eventos históricos de cheias que ocorreram anteriormente ao início do programa de coleta sistemática de dados hidrológicos. Essas informações podem ser incorporadas à análise de freqüência e obtidas por meio de pesquisas em arquivos públicos e particulares, os quais

guardam documentos de instituições ligadas aos recursos hídricos ou que sofreram as conseqüências das cheias, tais como, institutos históricos e geográficos, museus, empresas relacionadas ao projeto, construção e operação de sistemas de transporte ferroviário, rodoviário e fluvial; arquivos particulares e públicos com fotos e filmes de enchentes; arquivos de jornais e revistas locais, regionais e nacionais; registros paroquiais, entre outras fontes. Informações sobre grandes cheias ocorridas no passado longínquo também podem ser obtidas por meio do uso dos chamados métodos paleohidrológicos. Em síntese, esses métodos fazem uso de técnicas de datação para reconstituir a cronologia de grandes cheias, ocorridas em passado longínquo, ao longo de certo trecho fluvial, a partir das evidências de sua passagem, tais como depósitos de sedimentos e outras marcas deixadas nas seções transversais próximas.

A incorporação de dados históricos nas estimativas de frequência de vazões de enchentes tem sido objeto de considerável debate na literatura especializada (Hirsch, 1987; Hosking e Wallis, 1986; Sutcliffe, 1987). Do mesmo modo, a utilização dos métodos paleohidrológicos também recebe grande atenção, principalmente nos Estados Unidos (Baker, 1987; Stedinger *et al.*, 1993).

Uma das questões ligadas à utilização de informações sobre eventos históricos está relacionada à estimativa de suas respectivas posições de plotagem. Essa questão pode ser ilustrada pelo diagrama da Figura 8.5. Nessa figura, h representa o número de anos do período histórico e s denota o período de coleta sistemática de dados, enquanto que e indica o número de vazões extremas observadas no período sistemático e e' refere-se ao número de eventos extremos do período histórico. O limite Q_o , indicado por uma linha tracejada na Figura 8.5, refere-se à vazão acima da qual as cheias foram consideradas extremas. De acordo com Bayliss e Reed (2001), o limiar Q_o pode ser definido pelos registros históricos e corresponde a um nível de referencia acima do qual as vazões extremas foram percebidas. Ainda segundo Bayliss e Reed (2001), em algumas situações, o limite é determinado por uma cheia extrema recente, sendo razoável supor que o limite seja pouco inferior a esse evento de grandes proporções. Para Hirsch (1987), o limite Q_o pode ser estabelecido pela vazão que produz destruição e sérios prejuízos econômicos.

Quando as informações históricas são incorporadas como no formato da Figura 8.5, Hirsch (1987), Hirsch e Stedinger (1987) e Salas *et al.* (1994) sugerem a utilização das seguintes fórmulas para cálculo da posição de plotagem:

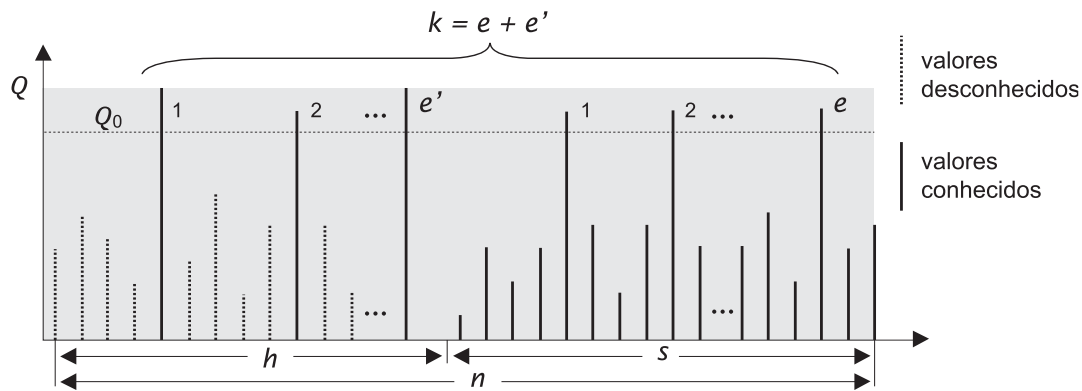


Figura 8.5 – Registros sistemáticos e informações históricas - Modificado de Bayliss e Reed (2001)

$$\left\{ \begin{array}{l} q_i = \frac{i-a}{k+1-2a} \cdot \frac{k}{n}, \quad i = 1, \dots, k \quad (a) \\ q_i = \frac{k}{n} + \frac{n-k}{n} \cdot \frac{i-k-a}{s-e+1-2a}, \quad i = k+1, \dots, k+s-e \quad (b) \end{array} \right. \quad (8.7)$$

onde a é a constante de posição de plotagem de Cunnane; n é o número de anos resultante da união das séries de dados sistemáticos e informações históricas, ou seja, $n = h + s$; k representa o número total de vazões extremas, ou seja, o número de vazões que superam o valor limite Q_0 no período combinado, $k = e + e'$.

As equações, contidas no sistema 8.7, permitem a plotagem do gráfico de probabilidade, ou seja, a probabilidade anual de excedência versus a magnitude das vazões. A equação 8.7a é aplicada a todas as vazões que estão acima da vazão limite. Em outras palavras, ela é utilizada para todas as vazões das séries histórica e sistemática que estão acima do limite Q_0 . A equação 8.7b é aplicada às vazões da série sistemática abaixo do limite.

A publicação britânica *Flood Estimation Handbook*, mencionada por Bayliss e Reed (2001), sugere que as maiores vazões da série combinada de informações históricas e dados sistemáticos sejam plotadas, por meio da utilização da fórmula de Gringorten. Isso difere da recomendação da utilização da equação 8.7a, com $a = 0,44$, apesar das diferenças serem pequenas. As diferenças são significativas somente quando k é muito pequeno em comparação a n , como por exemplo, quando o limite Q_0 é tão alto que poucas vazões, por século, são plotadas.

A vantagem do uso das equações do sistema 8.7 é que elas permitem que os dados sistemáticos abaixo do limite sejam plotados de um modo consistente e compatível aos outros dados. A seguir, apresenta-se um exemplo para ilustrar a utilização de informações históricas na análise de frequência.

Exemplo 8.3 (Modificado de Bayliss e Reed, 2001) - O rio Avon, em Evesham Worcestershire, na Inglaterra, com uma área de drenagem de 2200 km², é monitorado sistematicamente desde 1937. Por meio de pesquisas em jornais, publicações técnicas e arquivos do Severn River Authority, Bayliss e Reed (2001) selecionaram, a partir de 1822, 15 eventos históricos que foram superiores a 265 m³/s. Plotar em um mesmo papel de probabilidades as distribuições empíricas das séries sistemática (1937-1998) e combinada (1822-1998).

Solução: No caso do período sistemático, os eventos máximos por ano hidrológico foram ordenados de forma decrescente e associados às suas respectivas posições de plotagem, por meio da fórmula de Gringorten. A série sistemática ordenada, as posições de plotagem e os respectivos períodos de retorno calculados estão na Tabela 8.3.

A série combinada, formada pela série sistemática acrescida dos 15 eventos históricos, também foi ordenada de forma decrescente. As posições de plotagem foram calculadas com as equações do sistema 8.7. A equação 8.7a foi aplicada a todos os eventos superiores a 265 m³/s, que é o limite definido por Bayliss e Reed (2001) para incorporação de vazões históricas extremas. A equação 8.7b definiu a posição de plotagem das vazões da série sistemática inferiores ao limite de 265 m³/s. Os parâmetros das equações são: $n = 177$ anos (1998 - 1822 + 1); $h = 115$ anos; $s = 62$ anos; $k = 19$ (eventos superiores a 265 m³/s); $e = 4$ (eventos do período sistemático superiores a 265 m³/s); $e' = 15$ (eventos do período histórico superiores a 265 m³/s) e $a = 0,44$.

Os resultados com as posições de plotagem e tempos de retorno calculados também estão apresentados na Tabela 8.3. As distribuições empíricas das séries sistemática e combinada foram grafadas em um papel de probabilidades de Gumbel, conforme ilustrado na Figura 8.6. Por essa figura observa-se o efeito da inclusão da informação histórica. Além das mudanças óbvias nos pontos acima da vazão limite, outro efeito perceptível ocorre sobre os pontos da série combinada logo abaixo do limite. Os tempos de retorno desses pontos são levemente menores do que os calculados sem a utilização da informação histórica.

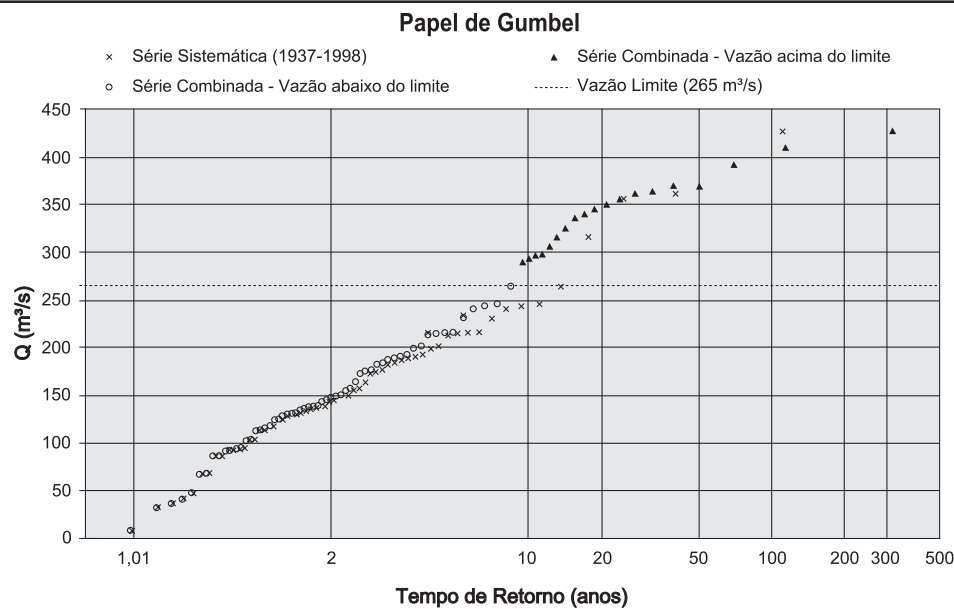


Figura 8.6 – Distribuições empíricas sistemática e combinada

Tabela 8.3 – Cálculo das posições de plotagem das séries sistemática e combinada

Série Sistemática (1937-1998)					Série Combinada (1822-1998)				
Ordem	AH	Q(m³/s)	Gring.	T(anos)	Ordem	AH	Q(m³/s)	Eq. (8.7)	T(anos)
1	1997	427,000	0,0090148	110,93	1	1997	427	0,003144	318,07
2	1967	361,909	0,0251127	39,82	2	1900*	410	0,008758	114,18
3	1946	356,187	0,0412106	24,27	3	1848*	392	0,014373	69,58
4	1939	316,213	0,0573084	17,45	4	1852*	370	0,019987	50,03
5	1981	264,091	0,0734063	13,62	5	1829*	370	0,025601	39,06
6	1959	245,633	0,0895042	11,17	6	1882*	364	0,031215	32,04
7	1958	243,687	0,1056021	9,47	7	1967	362	0,03683	27,15
8	1938	240,382	0,1216999	8,22	8	1946	356	0,042444	23,56
9	1979	230,596	0,1377978	7,26	9	1923*	350	0,048058	20,81
10	1980	215,716	0,1538957	6,50	10	1875*	345	0,053672	18,63
11	1960	215,279	0,1699936	5,88	11	1931*	340	0,059287	16,87
12	1978	214,387	0,1860914	5,37	12	1888*	336	0,064901	15,41
13	1992	212,600	0,2021893	4,95	13	1874*	325	0,070515	14,18
14	1942	201,259	0,2182872	4,58	14	1939	316	0,076129	13,14
15	1968	198,944	0,2343851	4,27	15	1935*	306	0,081744	12,23
16	1987	192,414	0,2504829	3,99	16	1932*	298	0,087358	11,45
17	1954	190,617	0,2665808	3,75	17	1878*	296	0,092972	10,76
18	1971	188,904	0,2826787	3,54	18	1885*	293	0,098586	10,14
19	1940	187,123	0,2987766	3,35	19	1895*	290	0,104201	9,60
20	1941	183,657	0,3148744	3,18	20	1981	264,091	0,115946	8,62
21	1950	181,934	0,3309723	3,02	21	1959	245,633	0,131304	7,62
22	1976	176,653	0,3470702	2,88	22	1958	243,687	0,146663	6,82
23	1984	174,533	0,3631681	2,75	23	1938	240,382	0,162022	6,17
24	1974	172,612	0,3792659	2,64	24	1979	230,596	0,177381	5,64
25	1989	163,307	0,3953638	2,53	25	1980	215,716	0,19274	5,19
26	1970	157,400	0,4114617	2,43	26	1960	215,279	0,208099	4,81
27	1982	155,035	0,4275596	2,34	27	1978	214,387	0,223457	4,48
28	1998	149,700	0,4436574	2,25	28	1992	212,600	0,238816	4,19
29	1949	148,908	0,4597553	2,18	29	1942	201,259	0,254175	3,93

* Eventos históricos

Tabela 8.3 – Continuação

Série Sistemática (1937-1998)					Série Combinada (1822-1998)				
Ordem	AH	Q(m³/s)	Gring.	T(anos)	Ordem	AH	Q(m³/s)	Eq.(8.7)	T(anos)
30	1965	148,443	0,4758532	2,10	30	1968	198,944	0,269534	3,71
31	1985	145,447	0,4919511	2,03	31	1987	192,414	0,284893	3,51
32	1993	143,400	0,5080489	1,97	32	1954	190,617	0,300252	3,33
33	1991	138,800	0,5241468	1,91	33	1971	188,904	0,31561	3,17
34	1956	138,782	0,5402447	1,85	34	1940	187,123	0,330969	3,02
35	1957	137,556	0,5563426	1,80	35	1941	183,657	0,346328	2,89
36	1973	135,722	0,5724404	1,75	36	1950	181,934	0,361687	2,76
37	1990	134,179	0,5885383	1,70	37	1976	176,653	0,377046	2,65
38	1966	131,490	0,6046362	1,65	38	1984	174,533	0,392405	2,55
39	1952	130,432	0,6207341	1,61	39	1974	172,612	0,407763	2,45
40	1951	130,432	0,6368319	1,57	40	1989	163,307	0,423122	2,36
41	1986	128,578	0,6529298	1,53	41	1970	157,400	0,438481	2,28
42	1994	124,300	0,6690277	1,49	42	1982	155,035	0,45384	2,20
43	1977	123,646	0,6851256	1,46	43	1998	149,700	0,469199	2,13
44	1963	117,402	0,7012234	1,43	44	1949	148,908	0,484558	2,06
45	1988	115,592	0,7173213	1,39	45	1965	148,443	0,499916	2,00
46	1995	113,900	0,7334192	1,36	46	1985	145,447	0,515275	1,94
47	1972	112,565	0,7495171	1,33	47	1993	143,400	0,530634	1,88
48	1944	103,298	0,7656149	1,31	48	1991	138,800	0,545993	1,83
49	1983	102,542	0,7817128	1,28	49	1956	138,782	0,561352	1,78
50	1969	94,897	0,7978107	1,25	50	1957	137,556	0,576711	1,73
51	1955	93,851	0,8139086	1,23	51	1973	135,722	0,592069	1,69
52	1961	92,290	0,8300064	1,20	52	1990	134,179	0,607428	1,65
53	1948	91,377	0,8461043	1,18	53	1966	131,490	0,622787	1,61
54	1953	86,275	0,8622022	1,16	54	1952	130,432	0,638146	1,57
55	1945	86,275	0,8783001	1,14	55	1951	130,432	0,653505	1,53
56	1962	67,913	0,8943979	1,12	56	1986	128,578	0,668864	1,50
57	1947	67,110	0,9104958	1,10	57	1994	124,300	0,684222	1,46
58	1937	47,021	0,9265937	1,08	58	1977	123,646	0,699581	1,43
59	1964	41,032	0,9426916	1,06	59	1963	117,402	0,71494	1,40
60	1975	35,937	0,9587894	1,04	60	1988	115,592	0,730299	1,37
61	1996	31,880	0,9748873	1,03	61	1995	113,900	0,745658	1,34
62	1943			1,01	62	1972	112,565	0,761017	1,31
					63	1944	103,298	0,776375	1,29
					64	1983	102,542	0,791734	1,26
					65	1969	94,897	0,807093	1,24
					66	1955	93,851	0,822452	1,22
					67	1961	92,290	0,837811	1,19
					68	1948	91,377	0,85317	1,17
					69	1953	86,275	0,868528	1,15
					70	1945	86,275	0,883887	1,13
					71	1962	67,913	0,899246	1,11
					72	1947	67,110	0,914605	1,09
					73	1937	47,021	0,929964	1,08
					74	1964	41,032	0,945323	1,06
					75	1975	35,937	0,960681	1,04
					76	1996	31,880	0,97604	1,02
					77	1943	7,574	0,991399	1,01

* Eventos históricos

8.2 – Análise de Frequência Analítica

A análise convencional de frequência de realizações de uma variável aleatória, da qual se conhece uma amostra e a distribuição de probabilidades da população de onde a amostra foi retirada, consiste em estimar os parâmetros populacionais a partir dos dados observados e, em seguida, estimar os quantis para a probabilidade desejada. No caso de eventos máximos (e/ou mínimos) de variáveis hidrológicas, a distribuição de probabilidades da população não é conhecida e tem-se somente uma amostra de dados observados. Esse fato complicador leva à proposição de modelos probabilísticos, ou sejam funções paramétricas de probabilidade, as quais, em função de suas características de assimetria e da eventual existência de limites superiores (e/ou inferiores) no domínio de definição da variável aleatória, se atribuem propriedades de modelarem os fenômenos hidrológicos. Muitas distribuições têm sido propostas para a modelação estatística dos valores máximos anuais de variáveis hidrológicas ou hidrometeorológicas, mas não há uma distribuição específica consensual que seja capaz de, sob quaisquer condições, descrever o comportamento da variável em foco. Portanto, em uma análise típica, cabe ao analista selecionar, entre as diversas distribuições candidatas, aquela que parece mais apropriada à modelação dos dados amostrais. Os procedimentos típicos de uma análise de frequência local de séries anuais estão descritos nos próximos subitens.

a) Avaliação dos dados amostrais

A qualidade e a aplicabilidade da análise de frequência dependem diretamente dos dados utilizados para estimação de seus parâmetros. Desse modo, é um fato reconhecido que, por mais sofisticado que seja, a qualidade de um modelo estocástico jamais superará a dos dados disponíveis para a estimação de seus parâmetros. Nesse sentido, cabe ao hidrólogo julgar a qualidade dos registros hidrológicos disponíveis para dar prosseguimento à análise de frequência.

É um pressuposto da análise de frequência convencional que a amostra de dados disponível seja uma entre um número infinito de outras amostras possíveis, as quais representariam realizações, com igual chance de sorteio, de uma única população. Também são pressupostos da análise de frequência convencional que os dados hidrológicos devem satisfazer as condições de independência, estacionariedade e representatividade. De modo sintético, pode-se dizer que os eventos são considerados independentes quando não há correlação entre os valores da série. Sendo assim, a independência significa a inexistência de correlação entre um registro de um dado ano e o registro posterior (ou anterior), considerados todos os anos disponíveis. Por outro lado, uma série de dados hidrológicos é dita

estacionária quando não ocorrem modificações nas características estatísticas de sua série ao longo do tempo. A análise de frequência de séries hidrológicas não estacionárias e, por conseguinte, a estimação de parâmetros e quantis com tendências ou variações temporais são objetos de investigações recentes [e.g: Cox et al. (2002) e Clarke (2002)] e não serão aqui consideradas. Em termos da análise de frequência convencional, dados não estacionários devem ser analisados em sub-séries homogêneas ou ajustados de modo a corrigir as heterogeneidades encontradas. As causas principais de possíveis não-estacionariedades em uma série hidrológica ou hidrometeorológica são: a relocação das estações de observação, a construção de barragens a montante, a urbanização ou o desmatamento das bacias, as eventuais modificações do leito fluvial, a ocorrência de cheias catastróficas, além, evidentemente, de mudanças climáticas.

A confiabilidade das estimativas dos parâmetros de uma dada distribuição de probabilidade está intrinsecamente ligada ao tamanho da amostra e à sua representatividade. Os dados da amostra devem ser representativos da variabilidade inerente ao processo natural ou experimento em foco. Em se tratando de variáveis hidrológicas ou hidrometeorológicas, uma amostra, obtida ao longo de um período predominantemente seco (ou úmido), irá certamente distorcer os resultados da análise, produzindo, em consequência, estimativas tendenciosas dos parâmetros populacionais. Por outro lado, uma amostra de dados possui propriedades estatísticas apenas similares às da população; elas serão idênticas se e somente se toda a população tiver sido amostrada. Yevjevich (1972) resume a questão afirmando que tanto a presença de erros sistemáticos em uma amostra, os quais podem ser provenientes de problemas de processamento e medição, de heterogeneidades e falta de representatividade, quanto os erros aleatórios, esses inerentes às naturais flutuações amostrais em torno de valores populacionais, podem produzir grandes incertezas quanto às estimativas de parâmetros estatísticos, realizadas a partir de amostras de tamanho relativamente pequeno. De qualquer modo, é um pressuposto básico dos métodos de inferência estatística a inexistência de erros sistemáticos, atribuindo somente às flutuações amostrais as diferenças entre estimativas e valores populacionais.

Benson (1960), utilizou uma série sintética de 1000 anos de vazões máximas e demonstrou que para se estimar uma cheia de 50 anos são necessárias amostras de pelo menos 39 anos, para que as estimativas ficassem na faixa de 24% do valor correto, em 95% dos casos. Caso a confiança de acerto decresça para 80%, o período mínimo de dados necessário seria de 15 anos. É frequente encontrar na literatura referências à consideração que de uma série de máximos anuais de n valores pode-se estimar, com alguma confiabilidade, quantis com tempos de retorno de até $2n$. Watt et al. (1988), editor do guia “*Hydrology of*

Floods in Canada - A Guide to Planning and Design”, preparado para o Conselho Nacional de Pesquisas do Canadá, relacionam o tamanho da amostra ao tipo de abordagem a ser tomada pela análise de frequência de vazões máximas. Nesse guia, a análise de frequência local de vazões máximas anuais é recomendada apenas para as amostras com mais de 10 anos de dados e para estimativas de quantis com tempos de retorno no máximo menores do que quatro vezes o tamanho da série. Apesar de existirem outras formas de avaliar qualitativamente a aplicabilidade da análise de frequência, não se pode negar a importância do tamanho da amostra como uma forma de avaliação qualitativa dos estimadores amostrais e quantis, uma vez que a variância de todos eles é inversamente proporcional a alguma potência do tamanho da amostra.

Testes estatísticos paramétricos e não paramétricos podem ser usados como ferramentas auxiliares na identificação da presença de dependência e heterogeneidade serial. Os testes paramétricos são fundamentados em suposições distributivas mais severas do que as exigidas por testes não paramétricos similares. Geralmente, em sua formulação, os testes paramétricos são baseados na suposição de uma distribuição de probabilidades específica para os dados amostrais. Conforme visto no capítulo 7, os testes não paramétricos, também chamados de “testes livres de distribuição”, não exigem a premissa de uma distribuição de probabilidade específica e têm suas estatísticas de decisão construídas com base em características indiretas dos dados originais. Portanto, tendo como motivação não assumir *a priori* compromissos com as características distributivas populacionais durante a etapa de verificação de dados amostrais, é claramente recomendável o uso dos testes não paramétricos, discutidos no capítulo 7, para a identificação da eventual presença de heterogeneidade e dependência serial na amostra. Cabe esclarecer, entretanto, que, embora os testes estatísticos sejam válidos para pequenas amostras e sob situações diversas, eles devem ser vistos apenas como indicadores, pois não são constituem, por si, argumentos suficientemente fortes para se abandonar uma amostra caso indiquem, por exemplo, a presença de dependência serial entre seus dados. Nesses casos, deve-se procurar uma evidência física que justifique o resultado do teste.

Ainda na etapa de verificação inicial de dados, deve-se lembrar que alguns cuidados devem ser tomados durante a seleção dos eventos de modo a assegurar a independência serial da amostra. Em regiões com sazonalidade muito acentuada, a seleção de eventos para compor uma dada amostra deve ser feita de forma diferenciada para vazões máximas e mínimas anuais; por exemplo, no estado de Minas Gerais, como de resto em grande parte da região sudeste do Brasil, a estação chuvosa vai de Outubro a Março, com grande possibilidade de ocorrência de eventos máximos em Dezembro. Neste caso, as vazões máximas anuais devem

ser individualizadas por ano hidrológico, o qual corresponde a um período fixo de 12 meses, a começar no início do período chuvoso (Outubro) e terminar no final da estação seca (Setembro). Mesmo em regiões com sazonalidade não tão evidente como o sudeste brasileiro, tais como o sul de Santa Catarina e grande parte do Rio Grande do Sul, o ano hidrológico de Maio a Abril deve ser empregado para a seleção de eventos. Por outro lado, no caso da seleção da amostra de vazões mínimas anuais, a abordagem anterior merece restrições, já que uma estiagem prolongada pode fazer com que valores dependentes sejam escolhidos. Neste caso, os períodos anuais devem ser limitados pelos meses mais chuvosos.

b) Definição da distribuição de probabilidade, estimação de seus parâmetros e a verificação de aderência à distribuição empírica.

Existe um conjunto não muito extenso de funções de distribuição de probabilidades que podem ser empregadas para a modelação de eventos máximos anuais de variáveis hidrológicas e hidrometeorológicas. Dentro desse conjunto, pode-se distinguir as distribuições oriundas da teoria clássica de valores, quais sejam as distribuições Gumbel, Fréchet, Weibull e a Generalizada de Valores Extremos (GEV), e aquelas ditas não-extremais, entre as quais as de maior uso são: as distribuições Exponencial e sua forma mais geral que é a Generalizada de Pareto, Pearson III, Log-Pearson III e Log-Normal de 2 parâmetros. Embora a adequação destas distribuições candidatas dependa de critérios variados, incluindo alguns de caráter subjetivo, talvez o atributo mais desejável seja a capacidade dessas distribuições de reproduzir algumas características amostrais relevantes. Apresenta-se, a seguir, as principais considerações a levar em conta quando da seleção de um modelo probabilístico local.

No que concerne às distribuições limitadas à direita, é um fato que algumas quantidades físicas possuem limites superiores inerentemente definidos; é o caso, por exemplo, da concentração de oxigênio dissolvido em um corpo d'água, limitado fisicamente em um valor entre 9 a 10 mg/l, a depender da temperatura ambiente. Outras quantidades podem igualmente possuir um limite superior, muito embora, tal limite possa não ser conhecido *a priori*, fato decorrente da insuficiente compreensão e/ou quantificação de todos os processos físicos causais envolvidos. A esse respeito, é conhecida a controvérsia quanto à existência da Precipitação Máxima Provável (PMP), originalmente formulada como um limite superior de produção de precipitação pelo ar atmosférico. Admitindo-se que esse limite exista de fato, é consensual que sua determinação fica comprometida pela insuficiente quantificação da variabilidade espaço-temporal das variáveis que lhe dão origem. Entretanto, pode-se conjecturar que seria fisicamente impossível a ocorrência de uma vazão, digamos de 100.000 m³/s, em uma pequena bacia hidrográfica, por

exemplo, da ordem de 100 km² de área de drenagem. Por essa razão, alguns pesquisadores, como Boughton (1980) e Laursen (1983), recomendam que somente distribuições limitadas superiormente devem ser usadas para modelar variáveis com essas características. Hosking e Wallis (1997) consideram errônea essa recomendação e sustentam que, se o objetivo da análise de frequência é o de estimar o quantil de tempo de retorno de 100 anos, é irrelevante considerar como “fisicamente impossível” a ocorrência do quantil de 100.000 anos. Acrescentam que impor um limite superior ao modelo probabilístico pode comprometer a obtenção de boas estimativas de quantis para os tempos de retorno que realmente interessam. Hosking e Wallis (1997) concluem afirmando que, ao se empregar uma distribuição ilimitada superiormente, as premissas implícitas são (i) que o limite superior não é conhecido e nem pode ser estimado com a precisão necessária e (ii) que no intervalo de tempos de retorno de interesse do estudo, a distribuição de probabilidades da população pode ser melhor aproximada por uma função ilimitada do que por uma que possua um limite superior. Evidentemente, quando existem evidências empíricas que a distribuição populacional possui um limite superior, ela deve ser aproximada por uma distribuição limitada superiormente. Seria o caso, por exemplo, do ajuste da distribuição Generalizada de Valores Extremos a uma certa amostra, cuja tendência de possuir um limite superior estaria refletida na estimativa de um valor positivo para o parâmetro de forma κ .

O chamado “peso” da cauda superior de uma função distribuição de probabilidades determina a intensidade com que os quantis aumentam, à medida que os tempos de retorno tendem para valores muito elevados. Em outras palavras, o peso da cauda superior é proporcional às probabilidades de excedência associadas a quantis elevados e é reflexo da intensidade com que a função densidade $f_x(x)$ decresce quando x tende para valores muito elevados. Os pesos das caudas superiores de algumas das principais funções de distribuição de probabilidades encontram-se relativizados na Tabela 8.4.

Para a maioria das aplicações envolvendo variáveis hidrológicas/hidrometeorológicas, a correta prescrição da cauda superior de uma distribuição de probabilidades é de importância fundamental e, em muitos casos, representa a motivação primeira da análise de frequência. Entretanto, os tamanhos das amostras disponíveis para essas aplicações são invariavelmente insuficientes para se determinar, com alguma precisão, a forma da cauda superior do modelo probabilístico. Segundo Hosking e Wallis (1997), não havendo razões suficientes para se recomendar o emprego exclusivo de somente um tipo de cauda superior, é aconselhável utilizar um grande conjunto de distribuições candidatas cujos pesos de suas caudas superiores se estendam por um amplo espectro.

Tabela 8.4 – Pesos das caudas superiores de algumas distribuições de probabilidade		
Cauda Superior	Forma de $f(x)$ para valores elevados de x	Distribuição
Pesada	x^{-A}	Generalizada de valores extremos, generalizada de Pareto e Logística generalizada com parâmetro de forma $k < 0$.
↑	$x^{-A \ln x}$	Lognormal
	$\exp(-x^A)$ $0 < A < 1$	Weibull com parâmetro de forma $\lambda < 1$.
	$x^A \exp(-Bx)$	Pearson tipo III com assimetria positiva.
	$\exp(-x)$	Exponencial, Gumbel.
↓	$\exp(-x^A)$ $A > 1$	Weibull com parâmetro de forma $\lambda < 1$.
Leve	Existe um Limite superior	Generalizada de valores extremos, generalizada de Pareto e Logística generalizada com parâmetro de forma $k > 0$; e Pearson tipo III com assimetria negativa.

A e B representam constantes positivas. (adap. de Hosking e Wallis, 1997, p. 75)

Considerações semelhantes às anteriores se aplicam à cauda inferior, ou seja, é necessário utilizar um conjunto razoável de distribuições candidatas cujos pesos de suas caudas inferiores se estendam por um amplo espectro. Entretanto, se o interesse do estudo encontra-se centrado em se prescrever a melhor aproximação da cauda superior, a forma da cauda inferior é irrelevante. Em alguns casos, conforme enfatizado no relatório “*Estimating Probabilities of Extreme Floods, Methods and Recommended Research*” do *National Research Council (NRC, 1987)*, a presença de “*outliers*” baixos em uma dada amostra pode inclusive vir a comprometer a correta estimação das características da cauda superior.

As reflexões sobre o limite superior também são aplicáveis ao limite inferior. Contudo, diferentemente do limite superior, o inferior é, em geral, conhecido ou pode ser igualado a zero; algumas distribuições, como a Generalizada de Pareto, permitem, com facilidade, o ajuste do parâmetro de posição, quando se conhece ou se prescreve o limite inferior. Hosking e Wallis (1997) ressaltam, entretanto, que, em diversos casos, a prescrição de limite inferior nulo é inútil e que melhores resultados podem ser obtidos sem nenhuma prescrição *a priori*. Exemplificam afirmando que os totais anuais de precipitação em regiões úmidas, apesar de números positivos, são muito superiores a zero; para esse exemplo, uma distribuição de probabilidades realista deve ter um limite inferior muito maior do que zero.

As distribuições oriundas da teoria clássica de valores extremos (Gumbel, 1958), quais sejam os modelos Gumbel, Fréchet e Weibull, são as únicas para as quais existem justificativas teóricas para seu emprego na modelação de valores máximos (ou mínimos) de dados empíricos. Por exemplo, o modelo de valores extremos do tipo I para máximos (EV1 ou Gumbel) é a distribuição assintótica do maior valor de uma seqüência ilimitada de variáveis aleatórias independentes e igualmente distribuídas (*iid*), a distribuição das quais possui uma cauda superior do tipo exponencial. Analogamente, a distribuição do tipo II para máximos (EV2 ou Fréchet) relaciona-se a variáveis *iid* com cauda superior do tipo polinomial, enquanto a distribuição do tipo III (EV3 ou Weibull) refere-se a variáveis *iid* que possuem um limite superior finito. Sob as premissas da teoria de valores extremos, por exemplo, a distribuição de probabilidades das vazões médias diárias máximas anuais de uma certa bacia hidrográfica, depende da distribuição inicial única dos valores diários considerados independentes. A maior objeção ao uso das distribuições oriundas da teoria de valores extremos em hidrologia refere-se à premissa de variáveis iniciais *iid*, a qual muito dificilmente é satisfeita por variáveis hidrológicas ou hidrometeorológicas. A esse respeito, transcreve-se o seguinte comentário escrito por Perichi e Rodríguez-Iturbe (1985, p. 515) :

“Presumir que duas vazões médias diárias, observadas digamos no dia 15 de maio e em 20 de Dezembro, são variáveis aleatórias identicamente distribuídas, é uma clara violação da realidade hidrológica. Essa premissa ‘regulariza’ as distribuições históricas iniciais afirmando não só que elas são do mesmo tipo, mas também que elas possuem os mesmos parâmetros (e.g. média e variância) para qualquer dia do ano. Sob essa premissa, não se pode admitir o fato que se uma mesma vazão média diária foi observada em dois dias diferentes, é mais provável que aquele que possui a maior variância produzirá cheias maiores do que aquele de menor variância. A realidade hidrológica é que a combinação da média e da variância de um dado mês faz com que alguns meses do ano sejam mais suscetíveis à ocorrência de cheias do que outros.”

Além dessas considerações, a seqüência de variáveis hidrológicas/hidrometeorológicas, amostradas em intervalos horários (ou diários) ao longo de um ano, pode apresentar correlação serial significativa e o número anual de intervalos de tempo não é uma garantia de convergência para alguma das três formas assintóticas extremas.

As características probabilísticas de um fenômeno aleatório não são facilmente definidas, tanto que a dedução teórica do modelo probabilístico necessário para definir tais características não é uma tarefa simples. Sob certas circunstâncias, as

bases ou propriedades do fenômeno físico em análise podem sugerir a forma da distribuição a ser utilizada. Por exemplo, se o processo é composto pelo somatório de muitos efeitos individuais, como no caso da precipitação anual que é a soma das precipitações diárias, a distribuição Normal pode ser utilizada com base no *teorema do limite central*. Além disso, o fato que variáveis hidrológicas e hidrometeorológicas dificilmente satisfazem as premissas da teoria clássica de valores extremos, pode vir a justificar o uso de distribuições não-extremais, tais como a Log-Normal, na análise local de frequência de eventos máximos anuais. Chow (1954) apresenta a seguinte justificativa para o emprego da distribuição Log-Normal: os fatores causais de várias variáveis hidrológicas agem de forma multiplicativa, ao invés de aditiva, e a soma dos logaritmos desses fatores, em consequência do *teorema central limite* da teoria de probabilidades, tende a ser normalmente distribuída. Stedinger et al. (1993) afirmam que algumas variáveis como a diluição, por exemplo, podem resultar do produto de fatores causais. Entretanto, para o caso de enchentes ou precipitações máximas, a interpretação dessa ação multiplicativa não é evidente.

As objeções anteriores referem-se às justificativas teóricas inerentes à distribuição Log-Normal, bem como às distribuições oriundas da teoria clássica de valores extremos. Entretanto, conforme menção anterior, tais objeções não têm a finalidade de excluir os modelos citados do elenco de distribuições candidatas à modelação de variáveis hidrológicas e hidrometeorológicas. No contexto da análise de frequência local de variáveis hidrológicas, elas devem ser consideradas candidatas como quaisquer outras distribuições e, portanto, devem ser discriminadas de acordo com outros critérios, tais como suas *medidas de aderência* aos dados amostrais.

Com relação ao número de parâmetros desconhecidos de uma distribuição de probabilidades, Hosking e Wallis (1997) afirmam que as distribuições de dois parâmetros produzem estimativas precisas de quantis quando as características distributivas populacionais a elas se assemelham. Entretanto, quando isso não ocorre, pode-se produzir estimativas tendenciosas dos quantis. A busca de um modelo probabilístico mais geral e flexível levou as agências do governo norte-americano a preconizarem o uso da distribuição Log-Pearson do tipo III para a análise local de frequência de cheias máximas anuais em projetos com participação federal. O modelo Log-Pearson III é uma distribuição de três parâmetros, resultante da transformação logarítmica de variáveis aleatórias distribuídas de acordo com Gama ou Pearson do tipo III. Embora os seus três parâmetros confirmem flexibilidade de forma a essa distribuição, a sua estimação, com base exclusiva em dados locais, é uma fonte de controvérsias. Bobée (1975) reporta situações em que a simples alteração do método de inferência estatística faz com que o parâmetro de forma dessa distribuição passe de negativo a positivo, o que a torna limitada

superiormente ou inferiormente de acordo com o sinal do parâmetro. São essas características indesejáveis da distribuição Log-Pearson do tipo III que levaram, por exemplo, Reich (1977) a argumentar contra a sua utilização na análise de frequência local de vazões máximas anuais. No contexto da análise regional, Hosking e Wallis (1997) observam que, obedecido o preceito da *parcimônia estatística*, recomenda-se o uso de distribuições de mais de dois parâmetros por produzirem estimativas menos tendenciosas dos quantis nas caudas superior e inferior. No contexto da análise local, entretanto, resta apenas o preceito da parcimônia de parâmetros na especificação da função de distribuição de probabilidades.

As considerações anteriores, revelando 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, remetem o analista a critérios variados e de algum modo subjetivos, entre os quais aqueles relacionados à capacidade descritiva dos modelos propostos. Alguns especialistas utilizam, como um possível critério de escolha, a comparação entre o coeficiente de assimetria amostral e o valor de assimetria teórico esperado para uma determinada distribuição de probabilidade. Por exemplo, enquanto estimativas amostrais do coeficiente de assimetria amostral próximas de zero podem sugerir a distribuição Normal como candidata à modelação estatística, amostras com assimetrias próximas a 1,14 ou -1,14 indicam a prescrição de uma distribuição de Gumbel. A utilização deste critério está sujeita à precisão da estimativa do coeficiente de assimetria, a qual cresce com o aumento do tamanho da amostra, e serve apenas como um indicador de ajuste, tornando necessário o emprego de outros critérios, tais como indicadores de aderência, para selecionar uma distribuição probabilidades apropriada.

Apesar de ser um procedimento subjetivo, o exame visual do ajuste entre as distribuições de probabilidades candidatas e os dados observados pode ser útil na seleção da distribuição de probabilidades apropriada. Para isto, os dados observados são ordenados de forma decrescente, para análise de máximos, e plotados em um papel de probabilidade específico para cada distribuição. A tendência linear dos pontos plotados em papel de probabilidade apropriado é um indício de que a amostra pode ter sido extraída daquela população. Por exemplo, uma tendência linear em um papel de probabilidade normal é uma evidência que os dados amostrais possam ter sido sorteados de uma distribuição normal. No caso de distribuições de 3 parâmetros, o exame visual ainda pode ser realizado nos papéis de probabilidade mais comuns, tais como exponencial ou normal. Entretanto, neste caso, serão observadas tendências curvilíneas e não mais lineares. Embora útil, o exame visual dos dados é adequado para amostras de grandes

tamanhos, uma vez que, amostras pequenas são muito mais sensíveis à presença de erros de amostragem ou de imprecisão na estimação da posição de plotagem, os quais podem tornar a análise visual pouco informativa ou, mesmo, pouco confiável.

A definição de um modelo distributivo que descreva as características probabilísticas de um fenômeno hidrológico é um problema complexo e passa também pela estimação dos seus parâmetros. Conforme exposição anterior, as distribuições freqüentemente utilizadas em hidrologia apresentam um número de parâmetros bastante variado. Apesar dos modelos de 3 parâmetros apresentarem maior flexibilidade de forma, de modo geral, quando se dispõe de amostras curtas (50 valores ou menos), é aconselhável que se investigue primeiramente apenas as funções que estão definidas por um ou dois parâmetros, pois a qualidade da estimativa é proporcional ao tamanho e à representatividade da amostra. O cálculo dos parâmetros pode ser realizado por vários métodos, mas os mais empregados são o método dos momentos, o método da máxima verossimilhança e o método dos momentos-L, que foram descritos e discutidos no capítulo 6.

Além das considerações anteriores sobre a definição do modelo que descreve o comportamento probabilístico de uma variável hidrológica, outro aspecto importante é a verificação do ajuste ou aderência da distribuição teórica à curva da distribuição empírica. Essa verificação de ajuste ou aderência pode ser realizada aplicando alguns testes, como por exemplo, os testes do Qui-Quadrado, de Anderson-Darling, de Kolmogorov-Smirnov, de Filliben e do teste visual, os quais foram descritos no item 7.4 do capítulo 7.

c) Identificação e tratamento de pontos atípicos

O ajuste entre a distribuição empírica e teórica pode ser comprometido pela presença de *outliers* como foi examinado no item 7.5 do capítulo 7. Esses eventos atípicos podem ser identificados por diferentes métodos. Nesta publicação destacamos os procedimentos dos quartis amostrais e da amplitude inter-quartis, descritos no item 2.1.4 do capítulo 2, e o teste de Grubbs e Beck (1972) exposto no item 7.5 do capítulo 7. Entretanto, caso ocorra a identificação de pontos atípicos, a exclusão desses *outliers* da análise de frequência é uma decisão que exige cuidados, conforme discutido no item 7.5 do capítulo 7.

8.3 – Análise de Frequência Utilizando o Fator de Frequência

De acordo com Chow (1964), um quantil de uma variável hidrológica pode ser representado pela média μ_x , somada a um desvio ΔX , da seguinte forma:

$$X = \mu_x + \Delta X \quad (8.8)$$

O termo ΔX depende da dispersão característica da distribuição de X , do tempo de retorno e de outros parâmetros do modelo probabilístico. Ainda segundo Chow (1964), o termo ΔX pode ser assumido igual ao produto do desvio padrão σ por um fator de frequência k_T , ou seja, $\Delta X = \sigma k_T$. O fator de frequência é uma função do tempo de retorno e da distribuição de probabilidades empregada na análise. Desse modo, a equação 8.8 pode ser reescrita como

$$X = \mu_x + \sigma k_T \quad (8.9)$$

Substituindo pelas estimativas amostrais tem-se

$$x_T = \bar{x} + s k_T \quad (8.10)$$

onde, x_T é estimativa do quantil associado ao tempo de retorno T ; \bar{x} é a média amostral; s é o desvio-padrão amostral e k_T é o fator de frequência associado ao modelo probabilístico e ao tempo de retorno T . A aplicação do método dos fatores de frequência aos modelos distributivos mais usuais é objeto dos itens que se seguem.

8.3.1 – Distribuição Normal

Quando uma variável é normalmente distribuída, o quantil é calculado pela relação $x = \mu_x + Z \sigma_x$, onde σ_x e μ_x são, respectivamente, o desvio padrão e a média da variável e Z é a *variável normal central reduzida*. Por analogia, conclui-se que o fator de frequência da distribuição normal, k_T , é igual à *variável normal central reduzida* Z . Os valores de Z podem ser obtidos nas Tabela 5.1 e 8.1 ou aproximados pela seguinte equação apresentada por Kite (1977):

Para $0 \leq P(X \leq x) \leq 0,5$:

$$Z \approx - \left(W - \frac{C_0 + C_1 W + C_2 W^2}{1 + d_1 W + d_2 W^2 + d_3 W^3} \right), \text{ com } W = \sqrt{\ln \left(\frac{1}{P(X \leq x)^2} \right)} \quad (8.11)$$

para $P(X \leq x) > 0,5$

$$Z \approx \left(W - \frac{C_0 + C_1 W + C_2 W^2}{1 + d_1 W + d_2 W^2 + d_3 W^3} \right) \text{ com } W = \sqrt{\ln \left(\frac{1}{(1 - P(X \leq x))^2} \right)} \quad (8.12)$$

Nos dois casos, os valores das constantes são: $C_0 = 2,515517$; $C_1 = 0,802853$; $C_2 = 0,010328$; $d_1 = 1,432788$; $d_2 = 0,189269$ e $d_3 = 0,001308$.

8.3.2 – Distribuição Log-Normal

Como para a distribuição Log-Normal, os logaritmos neperianos dos elementos da amostra devem ser ajustados a uma distribuição normal, tem-se que o fator de frequência k_T também deve ser igual à *variável normal central reduzida* Z e que, portanto, a equação 8.10 pode ser reescrita da seguinte forma:

$$x_T = \exp(\bar{x}_{\ln x} + s_{\ln x} k_T) \quad (8.13)$$

onde x_T é a estimativa do quantil associado ao tempo de retorno T ; $\bar{x}_{\ln x}$ é a média dos logaritmos de X ; $s_{\ln x}$ é o desvio-padrão dos logaritmos de X e k_T é o fator de frequência, igual à *variável normal central reduzida* Z .

8.3.3 – Distribuição Log-Pearson Tipo III

Kite (1977) apresenta a seguinte equação para a estimação dos quantis da distribuição Log-Pearson tipo III por meio do fator de frequência:

$$Y_T = \ln x_T = \bar{y} + k_T s_Y \quad (8.14)$$

na qual, x_T é a estimativa do quantil associado ao tempo de retorno T , \bar{y} é a média dos logaritmos neperianos de X ; s_Y é o desvio padrão dos logaritmos de X ; e k_T é o fator de frequência, o qual pode ser obtido por meio de tabelas apropriadas (United States Water Resources Council, Guidelines for Determining Flood Frequency – Bulletin 17-B, U. S. Government Printing Office, Washington, 1982) ou aproximado pela transformação de Wilson-Hilferty, dada por

$$k_T \approx Z + (Z^2 - 1) \frac{\gamma_Y}{6} + \frac{1}{3} (Z^3 - 6Z) \left(\frac{\gamma_Y}{6} \right)^2 - (Z^2 - 1) \left(\frac{\gamma_Y}{6} \right)^3 + Z \left(\frac{\gamma_Y}{6} \right)^4 + \frac{1}{3} \left(\frac{\gamma_Y}{6} \right)^5 \quad (8.15)$$

$$\text{para } \begin{cases} 0,01 \leq \frac{1}{T} \leq 0,99 \\ |\gamma_Y| < 2 \end{cases}$$

onde o coeficiente de assimetria γ_Y pode ser estimado por

$$g_Y = \frac{n \sum_1^n (\ln x - \bar{Y})^3}{(n-2) \left[\sum_1^n (\ln x - \bar{Y})^2 \right]^{3/2}} \quad (8.16)$$

Z é a variável normal central reduzida.

8.3.4 – Distribuição de Gumbel

A FAP da distribuição de Gumbel para máximos é representada pela equação 5.56 e a sua inversa pode ser escrita da seguinte forma:

$$x(T) = \beta - \alpha \ln \left(-\ln \left(1 - \frac{1}{T} \right) \right) \quad (8.17)$$

onde β é parâmetro de posição; α é o parâmetro de escala e T é o tempo de retorno em anos. Estimando os parâmetros da distribuição pelo método dos momentos obtêm-se:

$$\hat{\beta} = \bar{X} - 0,45s_x \quad (8.18)$$

$$\hat{\alpha} = \frac{s_x}{1,283} \quad (8.19)$$

nas quais \bar{X} e s_x denotam a média e o desvio padrão amostrais.

Substituindo os parâmetros estimados pelas equações 8.18 e 8.19 em 8.17, e fazendo algumas simplificações encontra-se:

$$x(T) = \bar{X} + \left[-0,45 - \frac{1}{1,283} \ln \left(-\ln \left(1 - \frac{1}{T} \right) \right) \right] s_x \quad (8.20)$$

Comparando as equações 8.20 e 8.10, conclui-se que o fator de frequência da distribuição de Gumbel pode ser expresso pela seguinte equação:

$$k_T = - \left[0,45 + \frac{1}{1,283} \ln \left(- \ln \left(1 - \frac{1}{T} \right) \right) \right] \quad (8.21)$$

De acordo com Kite (1977), o fator de forma da distribuição de Gumbel também pode ser calculado considerando o tamanho das amostras disponíveis com a estimativa dos quantis através da equação:

$$x_T = \bar{x} + sk_T(n) \quad (8.22)$$

na qual x_T é a estimativa do quantil associado ao tempo de retorno T ; \bar{x} é a média amostral; s é o desvio-padrão amostral e $k_T(n)$ é fator de frequência em função do tamanho da amostra.

O fator de frequência, $k_T(n)$, pode ser obtido em tabelas (Haan, 1979 e Kite, 1977) ou calculado pela seguinte equação:

$$k_T(n) = \frac{Y_T - \mu_{Y_i}}{\sigma_{Y_i}} \quad (8.23)$$

na qual Y_T é a variável reduzida de Gumbel, associada a tempo de retorno T , calculada por

$$Y_T = - \ln \left\{ - \ln \left[1 - \frac{1}{T} \right] \right\} \quad (8.24)$$

Denota-se por $\hat{\mu}_{Y_i}$ a média dos $Y_i(n)$, enquanto o desvio padrão é representado por $\hat{\sigma}_{Y_i}$, o qual pode ser estimado por

$$\hat{\sigma}_{Y_i} = \sqrt{\frac{n \sum Y_i^2(n) - (\sum Y_i(n))^2}{n^2}} \quad (8.25)$$

Nessa equação, $Y_i(n) = - \ln \{ - \ln [F(x)] \}$ é a variável reduzida de Gumbel calculada para cada posição i de uma amostra ordenada de tamanho n . Admitindo que a posição de plotagem é calculada pela fórmula de Weibull, obtém-se, então, a seguinte equação:

$$Y_i(n) = - \ln \left(- \ln \left(1 - \frac{i}{n+1} \right) \right) \quad (8.26)$$

na qual i é a ordem de classificação do elemento amostral e n é o tamanho da amostra.

Os fatores de frequência calculados com a equação 8.21 correspondem aos resultados assintóticos, resultantes da utilização da equação 8.23, quando o tamanho da amostra tende para infinito ($n \rightarrow \infty$).

8.3.5 – Distribuição Weibull (mínimos)

A estimativa dos quantis da distribuição de Weibull para análise de mínimos também pode ser realizada por meio da equação 8.10. Nesse caso, o fator de frequência, k_T , segundo Kite (1977), é dado por:

$$k_T = A(\lambda) + B(\lambda) \left\{ \left[-\ln \left(1 - \frac{1}{T} \right) \right]^{\frac{1}{\lambda}} - 1 \right\} \quad (8.27)$$

na qual,

$$A(\lambda) = \left[1 - \Gamma \left(1 + \frac{1}{\lambda} \right) \right] B(\lambda) \quad (8.28)$$

$$B(\lambda) = \left[\Gamma \left(1 + \frac{2}{\lambda} \right) - \Gamma^2 \left(1 + \frac{1}{\lambda} \right) \right]^{-\frac{1}{2}} \quad (8.29)$$

$$\lambda = \frac{1}{H_0 + H_1\gamma + H_2\gamma^2 + H_3\gamma^3 + H_4\gamma^4} \quad \text{para } -1,0 \leq \gamma \leq 2 \quad (8.30)$$

Nas equações acima, $\Gamma(\cdot)$ denota a função Gama; $H_0 = 0,2777757913$; $H_1 = 0,3132617714$; $H_2 = 0,0575670910$; $H_3 = -0,0013038566$; $H_4 = -0,0081523408$ e γ é o coeficiente de assimetria estimado pela equação:

$$\hat{\gamma} = \frac{n \sum_1^n (x - \bar{x})^3}{(n-2) \left[\sum_1^n (x - \bar{x})^2 \right]^{\frac{3}{2}}} \quad (8.31)$$

Exemplo 8.4 – Calcular o fator de frequência da distribuição Gumbel, $k_T(n)$, referente ao tempo de retorno de 50 anos para uma amostra de 10 elementos.
Solução: A primeira etapa consiste em calcular a variável reduzida de Gumbel, $Y_i(n)$, para cada posição i através da equação 8.26. A Tabela 8.5 apresenta os resultados. Em seguida é estimada a média dos valores de $Y_i(n)$ e o desvio padrão pela equação 8.25. Os resultados estão na Tabela 8.5.

A variável reduzida de Gumbel para o tempo de retorno de 50 anos é calculada pela equação 8.24:

$$Y_T = -\ln\left\{-\ln\left[1 - \frac{1}{T}\right]\right\} = -\ln\left\{-\ln\left[1 - \frac{1}{50}\right]\right\} = 3,9019 \quad (8.32)$$

Tabela 8.5 – Cálculo dos $Y_m(n)$

i	$\frac{i}{n+1}$	$Y_i(n)$	i	$\frac{i}{n+1}$	$Y_i(n)$
1	0,090909	2,350619	8	0,727273	-0,26181
2	0,181818	1,60609	9	0,818182	-0,53342
3	0,272727	1,144278	10	0,909091	-0,87459
4	0,363636	0,794106		$\hat{\mu}_{Y_i}$	0,4952
5	0,454545	0,500651		$\hat{\sigma}_{Y_i}$	0,9496
6	0,545455	0,237677			
7	0,636364	-0,01153			

Como, $\hat{\mu}_Y = 0,4952$ e $\hat{\sigma}_Y = 0,9496$, o fator de frequência pode ser calculado pela equação 8.23, de forma que:

$$k_{50}(10) = \frac{Y_{50} - \hat{\mu}_Y}{\hat{\sigma}_Y} = \frac{3,9019 - 0,4952}{0,9496} = 3,5874 \quad (8.33)$$

Assim, o fator de frequência da distribuição de Gumbel para o tempo de retorno de 50 anos e uma amostra de 10 elementos é igual a 3,5874.

Exemplo 8.5 – Admitindo que uma série de vazões mínimas com 7 dias de duração apresenta um coeficiente de assimetria de -0,10, calcular o fator de frequência da distribuição de Weibull para os tempos de retorno de 2, 5, 10, 20, 50 e 100 anos.

Solução: Utilizando o valor do coeficiente de assimetria $\gamma = -0,10$, calcula-se o parâmetro λ pela equação 8.30. O valor de λ é igual a 4,048160583. O parâmetro permite que se calcule $B(\lambda)$ pela equação 8.29 e em seguida

$A(\lambda)$ através da equação 8.28, a saber $B(\lambda) = 3,972674215$ e $A(\lambda) = 0,369376575$.

Com esses valores é possível estimar o fator de frequência pela equação 8.27,

$$k_T = 0,369376575 + 3,972674215 \cdot \left\{ \left[-\ln\left(1 - \frac{1}{T}\right) \right]^{4,048160583} - 1 \right\} \quad (8.34)$$

Substituindo os tempos de retorno na equação 8.34, calcula-se os valores dos fatores de frequência. Neste exemplo, tem-se:

T (anos)	2	5	10	20	50	100
k_T	0,0255	-0,8607	-1,3247	-1,6959	-2,0881	-2,3281

8.4 – Intervalo de Confiança para os Quantis

Os intervalos de confiança para os quantis estimados podem ser definidos a partir da equação 6.23, como foi detalhado no item 6.6 do capítulo 6. Naquele item foi visto que, assintoticamente (para grandes valores de n), os estimadores de quantis \hat{x}_T são normalmente distribuídos. Sendo assim, com base na equação 6.23, o intervalo de confiança aproximado para um quantil \hat{x}_T a um nível de confiança $100(1-\alpha)\%$ é definido por:

$$\hat{x}_T - Z_{1-\frac{\alpha}{2}} s_T \leq \hat{x}_T \leq \hat{x}_T + Z_{1-\frac{\alpha}{2}} s_T \quad (8.35)$$

onde $Z_{1-\frac{\alpha}{2}}$ é a *variável normal central reduzida* associada à probabilidade $(1-\alpha/2)$ e s_T é o erro-padrão da estimativa de \hat{x}_T , o qual varia com o modelo distributivo em análise.

No capítulo 6, analisou-se a definição dos intervalos de confiança quando os parâmetros das distribuições foram estimados pelos métodos dos momentos, da máxima verossimilhança e dos momentos-L. Para ilustrar a definição de intervalos de confiança aproximados de quantis, apresenta-se a seguir as expressões para os erros-padrão para algumas distribuições, cujos parâmetros foram estimados pelo método dos momentos.

• Normal

$$s_T = \sqrt{\frac{s_X^2}{n} \left(1 + \frac{Z^2}{2} \right)} \quad (8.36)$$

na qual T é o tempo de retorno; s_X é o desvio padrão amostral e Z é a *variável normal central reduzida*.

• Log-Normal

$$s_T = \sqrt{\exp \left[\left(\bar{Y} + Z_{1-\frac{1}{T}} s_Y \right)^2 \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{s_Y^2}{n} \left(1 + \frac{1}{2} Z_{1-\frac{1}{T}}^2 \right)} \right]} \quad (8.37)$$

na qual $Y = \ln(X)$; s_Y é o desvio padrão dos logaritmos dos dados observados; T é o tempo de retorno e Z é a *variável normal central reduzida*.

• Log-Pearson Tipo III

Segundo Kite (1977), o erro-padrão para a distribuição Log-Pearson Tipo III pode ser estimado, no espaço logarítmico, a partir da seguinte equação:

$$s_{T,Y} = \delta \sqrt{\frac{s_Y^2}{n}} \quad (8.38)$$

na qual $s_{T,Y}$ é o erro-padrão dos logaritmos dos eventos observados; n é o tamanho da amostra; s_Y é o desvio padrão dos logaritmos dos dados observados e δ pode ser obtido a partir da Tabela 8.6, em dependência do tempo de retorno e do coeficiente de assimetria dos logaritmos dos dados amostrais.

O erro padrão pode ser convertido para o espaço aritmético por meio da relação:

$$s_T = \frac{\hat{x}_T (e^{s_{T,Y}} - e^{-s_{T,Y}})}{2,0} \quad (8.39)$$

Tabela 8.6 – Parâmetro δ para estimativa do erro padrão da Log-Pearson Tipo III

Assimetria	Tempo de retorno (anos)					
	2	5	10	20	50	100
0,00	1,0801	1,1698	1,3748	1,6845	2,1988	2,6363
0,10	1,0808	1,2006	1,4367	1,7810	2,3425	2,8168
0,20	1,0830	1,2309	1,4989	1,8815	2,4986	3,0175
0,30	1,0866	1,2609	1,5610	1,9852	2,6656	3,2365
0,40	1,0918	1,2905	1,6227	2,0915	2,8423	3,4724
0,50	1,0987	1,3199	1,6838	2,1998	3,0277	3,7238
0,60	1,1073	1,3492	1,7441	2,3094	3,2209	3,9895
0,70	1,1179	1,3785	1,8032	2,4198	3,4208	4,2684
0,80	1,1304	1,4082	1,8609	2,5303	3,6266	4,5595
0,90	1,1449	1,4385	1,9170	2,6403	3,8374	4,8618
1,00	1,1614	1,4699	1,9714	2,7492	4,0522	5,1741
1,10	1,1799	1,5030	2,0240	2,8564	4,2699	5,4952
1,20	1,2003	1,5382	2,0747	2,9613	4,4896	5,8240
1,30	1,2223	1,5764	2,1237	3,0631	4,7100	6,1592
1,40	1,2457	1,6181	2,1711	3,1615	4,9301	6,4992
1,50	1,2701	1,6643	2,2173	3,2557	5,1486	6,8427
1,60	1,2952	1,7157	2,2627	3,3455	5,3644	7,1881
1,70	1,3204	1,7732	2,3081	3,4303	5,5761	7,5339
1,80	1,3452	1,8374	2,3541	3,5100	5,7827	7,8783
1,90	1,3690	1,9091	2,4018	3,5844	5,9829	8,2196
2,00	1,3913	1,9888	2,4525	3,6536	6,1755	8,5562

• Weibull

Segundo Kite (1977), o erro-padrão para a distribuição de Weibull, para mínimos, pode ser estimado por:

$$s_T = \delta_w \sqrt{\frac{s_x^2}{n}} \tag{8.40}$$

na qual, s_x é o desvio padrão amostral; n é o tamanho da amostra e δ_w pode ser obtido a partir da Tabela 8.7, na dependência do tempo de retorno e do coeficiente de assimetria amostral.

Tabela 8.7 – Parâmetro δ_w para estimativa do erro padrão da distribuição de Weibull (mínimos)

Assimetria	Tempo de retorno (anos)					
	2	5	10	20	50	100
-0,80	0,9265	1,3665	1,8116	2,2267	2,6325	2,7650
-0,70	0,9743	1,3556	1,7517	2,1869	2,7877	3,2475
-0,60	1,0242	1,3492	1,6940	2,1413	2,8843	3,5450
-0,50	1,0710	1,3434	1,6356	2,0820	2,9084	3,6757
-0,40	1,0954	1,3259	1,5738	1,9846	2,7731	3,5067
-0,30	1,0886	1,2934	1,5063	1,8351	2,4456	3,0047
-0,20	1,0952	1,2624	1,4374	1,7320	2,2300	2,7011
-0,10	1,1065	1,2282	1,3709	1,6181	2,0938	2,5248
0,00	1,1157	1,1916	1,3042	1,5255	1,9631	2,3559
0,10	1,1244	1,1532	1,2374	1,4371	1,8437	2,2043
0,20	1,1318	1,1130	1,1711	1,3529	1,7336	2,0658
0,30	1,1394	1,0712	1,1078	1,2814	1,6496	1,9627
0,40	1,1460	1,0281	1,0467	1,2172	1,5775	1,8740
0,50	1,1517	0,9839	0,9905	1,1653	1,5236	1,8065
0,60	1,1567	0,9392	0,9414	1,1287	1,4905	1,7623
0,70	1,1605	0,8943	0,8981	1,1014	1,4661	1,7262
0,80	1,1636	0,8500	0,8646	1,0895	1,4583	1,7074
0,90	1,1657	0,8072	0,8422	1,0914	1,4630	1,7006
1,00	1,1671	0,7669	0,8319	1,1064	1,4788	1,7047
1,10	1,1678	0,7303	0,8348	1,1338	1,5049	1,7189
1,20	1,1681	0,6988	0,8507	1,1719	1,5394	1,7413
1,30	1,1680	0,6739	0,8792	1,2196	1,5815	1,7715
1,40	1,1676	0,6569	0,9196	1,2745	1,6291	1,8075
1,50	1,1669	0,6488	0,9673	1,3354	1,6816	1,8488
1,60	1,1658	0,6494	1,0218	1,3987	1,7355	1,8921
1,70	1,1643	0,6585	1,0807	1,4638	1,7908	1,9376
1,80	1,1622	0,6742	1,1406	1,5274	1,8446	1,9823
1,90	1,1596	0,6940	1,1987	1,5877	1,8952	2,0247
2,00	1,1544	0,7148	1,2523	1,6421	1,9405	2,0628

Exemplo 8.6 – Realizar uma análise de frequência com os dados de vazões diárias máximas anuais do rio Paraopeba em Ponte Nova do Paraopeba apresentados no Anexo 2. Considerar como candidatas as distribuições Log-Normal, Gumbel, Exponencial, Pearson III, Log-Pearson III e Generalizada de eventos extremos (GEV).

Solução: A primeira etapa consiste no cálculo das estatísticas e os momentos-L da série. Os resultados estão apresentados na Tabela 8.8.

Tabela 8.8 – Estatísticas de série de vazões diárias máximas anuais de Ponte Nova do Paraopeba

	Estatísticas descritivas	Estatísticas dos Logaritmos	Momentos-L	
Tamanho da Amostra	57	57	/1	534,2
Valor Máximo	1017	6,9250	/2	99,63
Valor Mínimo	246	5,5050	t3	0,1288
Média	534,2	6,2270	t4	0,1070
Desvio-Padrão	176,0	0,3320		
Coefficiente de Assimetria	0,6040	-0,0972		

O segundo passo consiste em aplicar alguns testes para verificar as hipóteses de independência e homogeneidade da série. Nesse exemplo a independência foi verificada com o teste não-paramétrico proposto por Wald e Wolfowitz (1943), descrito no item 7.3.2, e a homogeneidade pelo teste de Mann e Whitney (1947), descrito no item 7.3.3. A série pode ser considerada independente e homogênea a um nível de significância de 5%.

A terceira etapa é a verificação da presença de eventos atípicos na amostra. Nesse caso foi aplicado teste de Grubbs e Beck (1972), descrito no item 7.5. Na série analisada, a um nível de significância de 10%, não foi observada a presença de *outliers*.

Após a análise inicial dos dados, são calculados os parâmetros das distribuições candidatas. Nesse exemplo os parâmetros foram calculados pelo método dos momentos-L apresentado no capítulo 6. Os resultados estão na Tabela 8.9.

Tabela 8.9 – Parâmetros das distribuições candidatas

Distribuição	Posição (ξ)	Escala (α)	Forma (κ)
Log-Normal	6,2274	0,3382	
Gumbel	451,2123	143,7298	
Exponencial	334,9236	199,2519	
Pearson-III	534,1754	180,0157	0,7854
Log-Pearson-III	6,2274	0,3383	-0,1226
GEV	455,6143	152,0965	0,0650

Definidos os parâmetros das distribuições, é possível calcular os quantis associados a diferentes tempos de retorno a partir das inversas das FAP's dos modelos candidatas. As distribuições candidatas foram detalhadas nos capítulos 5 e 6. Na Tabela 8.10 são apresentadas algumas funções inversas das distribuições candidatas.

Tabela 8.10 – Funções Inversas da FAP de algumas distribuições

Distribuição	Inversa $x(T)$	Observações
Log-Normal	$x(T) = \exp(\xi + \alpha \cdot Z_T)$	Z_T é a variável normal central reduzida associada à probabilidade $(1-1/T)$
Gumbel	$x(T) = \xi - \alpha \cdot \ln(-\ln(-1/T))$	
Exponencial	$x(T) = \xi - \alpha \cdot \ln(1/T)$	
GEV	$x(T) = \xi + \alpha \cdot \left\{ \frac{1 - [-\ln(1 - 1/T)]^k}{k} \right\}$	para $k \neq 0$

Os parâmetros de posição μ , escala σ e forma γ da distribuição Pearson Tipo III podem ser calculados com as equações

$$\mu = \lambda_1, \quad \sigma = \frac{\lambda_2 \pi^{1/2} c^{1/2} \Gamma(c)}{\Gamma\left(c + \frac{1}{2}\right)} \quad \text{e} \quad \gamma = 2c^{-1/2} \text{sin}(\tau_3) .$$

A variável c é estimada considerando duas situações. A primeira, se $0 < |\tau_3| < 1/3$, nesse caso adotar $z = 3\pi\tau_3^2$ e aplicar a

equação $c \approx \frac{1 + 0,2906z}{z + 0,1882z^2 + 0,0442z^3}$. A segunda, se $1/3 \leq |\tau_3| < 1$, nessa

situação adota-se $z = 1 - |\tau_3|$ e emprega-se a equação

$$c \approx \frac{0,36067z - 0,59567z^2 + 0,25361z^3}{1 - 0,78861z + 2,56096z^2 - 0,77045z^3} .$$

A distribuição Pearson Tipo III com parâmetros de posição μ , escala σ e forma γ , apresenta algumas relações importantes com as distribuições Gama e Normal, as quais facilitam a estimação dos quantis. Quando o parâmetro de forma γ é positivo, a Pearson-III está associada à distribuição Gama. Se o parâmetro de forma γ é negativo, a Pearson-III está associada à distribuição Gama refletida. E, quando o parâmetro de forma γ é igual a zero, a Pearson-III está relacionada à distribuição Normal. Considerando que uma variável X segue uma distribuição Pearson tipo III, com parâmetros de posição μ , escala σ e forma γ , a relação entre esses parâmetros e os das distribuições Gama e Normal são as seguintes:

- Se $\gamma > 0$, então $X - \mu + \frac{2\sigma}{\gamma}$ segue uma distribuição Gama com

parâmetros $\alpha = \frac{4}{\gamma^2}$ e $\beta = \frac{\sigma\gamma}{2}$. Desse modo, os quantis da Pearson-III com parâmetro de forma positivo podem ser calculados pela equação:

$$x(T) = \mu - \frac{2\sigma}{\gamma} + G^{-1}\left(1 - \frac{1}{T}, \alpha, \beta\right) \quad (8.41)$$

onde T é o tempo de retorno e $G^{-1}()$ é a inversa da distribuição Gama com parâmetros α e β .

- Se $\gamma < 0$, então $-X + \mu - \frac{2\sigma}{\gamma}$ segue uma distribuição Gama com parâmetros $\alpha = \frac{4}{\gamma^2}$ e $\beta = \left|\frac{\sigma\gamma}{2}\right|$. Desse modo, os quantis da Pearson-III com parâmetro de forma negativo podem ser calculados pela equação:

$$x(T) = \mu - \frac{2\sigma}{\gamma} - G^{-1}\left(\frac{1}{T}, \alpha, \beta\right) \quad (8.42)$$

onde T é o tempo de retorno e $G^{-1}()$ é a inversa da distribuição Gama com parâmetros α e β .

- Se $\gamma = 0$, então X segue uma distribuição Normal com parâmetros μ e σ . Assim, os quantis da Pearson-III com parâmetro de forma nulo podem ser calculados pela equação:

$$x(T) = \mu + \sigma Z_T \quad (8.43)$$

onde T é o tempo de retorno e Z_T é a variável normal central reduzida associada uma probabilidade $(1 - 1/T)$. Recorde que, no programa Microsoft EXCEL, a inversa da distribuição Gama com parâmetros α e β pode ser calculada com a função INVGAMA() e a variável normal central reduzida com a função INV.NORMP().

Quando uma variável X segue a distribuição Log-Pearson tipo III, é um fato matemático que a variável transformada $Y = \ln(X)$ distribui-se de acordo com a Pearson tipo III. Assim, os parâmetros podem ser calculados por meio dos logaritmos dos valores observados e os quantis são estimados por meio das seguintes equações:

- Para $\gamma_{\ln X} > 0$

$$x(T) = \exp\left\{\mu_{\ln X} - \frac{2\sigma_{\ln X}}{\gamma_{\ln X}} + G^{-1}\left(1 - \frac{1}{T}, \alpha, \beta\right)\right\} \quad (8.44)$$

• Para $\gamma_{\ln X} < 0$

$$x(T) = \exp\left\{\mu_{\ln X} - \frac{2\sigma_{\ln X}}{\gamma_{\ln X}} - G^{-1}\left(\frac{1}{T}, \alpha, \beta\right)\right\} \quad (8.45)$$

• Para $\gamma_{\ln X} = 0$

$$x(T) = \exp(\mu_{\ln X} + \sigma_{\ln X} Z_T) \quad (8.46)$$

Os quantis das distribuições candidatas foram estimados por meio das funções inversas anteriormente apresentadas, pela substituição das estatísticas populacionais pelas amostrais. Os resultados estão apresentados na Tabela 8.11.

Tabela 8.11 – Quantis calculados para o exemplo 8.1 (m³/s)

Distribuição	T (anos)						
	2	5	10	50	100	200	1000
Log-Normal	506,4	673,2	781,2	1014,3	1112,2	1210,1	1440,0
Gumbel	503,9	666,8	774,7	1012,0	1112,4	1212,4	1444,0
Exponential	473,0	655,6	793,7	1114,4	1252,5	1390,6	1711,3
Pearson-III	510,8	674,8	774,7	974,5	1052,8	1128,1	1294,4
Log-Pearson-III	510,0	674,5	777,7	992,1	1079,1	1164,4	1358,5
GEV	510,7	673,0	774,0	979,8	1060,4	1137,1	1302,1

Antes de se iniciar a verificação do ajuste entre as distribuições teóricas e a empírica, a escolha do modelo probabilístico mais adequado pode ser feita por meio da análise dos parâmetros das distribuições candidatas e das estatísticas amostrais. Nesse exemplo, observa-se que o parâmetro de forma da GEV é positivo (ver Tabela 8.9), indicando uma distribuição com limite superior, o que, para alguns especialistas, conforme discussão no presente capítulo, não é adequado para a análise de máximos. Outra distribuição candidata que pode ser excluída da análise é a Log-Pearson tipo III, pois o coeficiente de assimetria no espaço logaritmo é negativo, indicando que esta distribuição também apresenta um limite superior. Considerando esses critérios, restam como candidatas as distribuições Log-Normal, Gumbel, Exponencial e Pearson-III.

Após essa seleção inicial, o próximo passo na escolha da distribuição teórica que melhor se ajustou à distribuição empírica é a verificação do ajuste por meio de testes de aderência e análise visual dos gráficos de probabilidades. Os testes de aderência foram descritos no item 7.4. Neste exemplo, foi aplicado o teste de Filliben, no qual as probabilidades empíricas para a verificação das distribuições Log-Normal e Pearson foram calculadas por meio da fórmula de posição de plotagem de Blom; para as distribuições de Gumbel e Exponencial, foi utilizada a fórmula de Gringorten. Os resultados obtidos do teste de Filliben estão na Tabela 8.12.

Tabela 8.12 – Resultados do teste de Filliben

Distribuição	$r_{crit,\alpha}$	r	Situação
Log-Normal ($\alpha = 10\%$)	0,9835	0,9952	Aceita
Gumbel ($\alpha = 10\%$)	0,9760	0,9919	Aceita
Exponencial ($\alpha = 10\%$)	0,9716	0,9616	Rejeitada
Pearson-III ($\alpha = 5\%$)	0,9860	0,9958	Aceita

As Figuras 8.7 e 8.8 permitem a verificação visual do ajuste entre as distribuições empíricas e teóricas. As probabilidades empíricas foram calculadas com ordenamento decrescente da amostra e a utilização das fórmulas de Blom e Gringorten, com os resultados na Tabela 8.13.

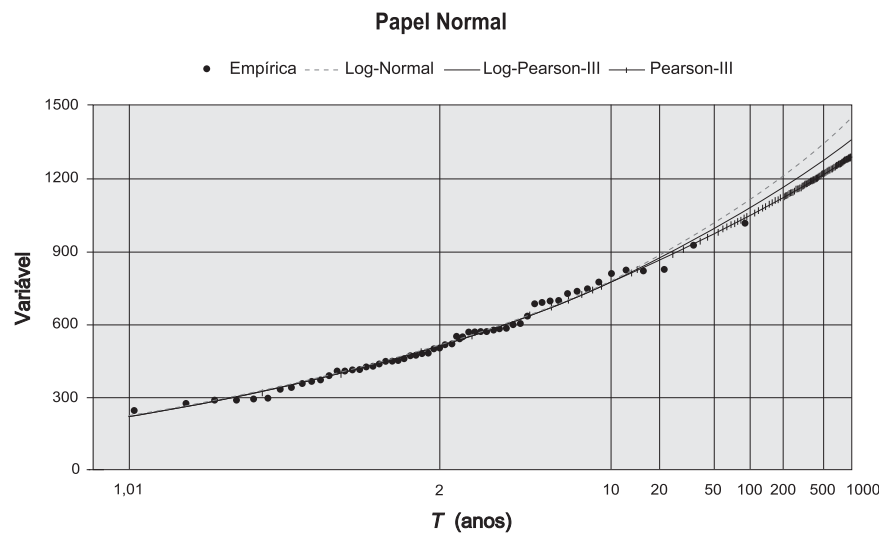


Figura 8.7 – Ajuste das distribuições Log-Normal, Pearson-III e Log-Pearson III

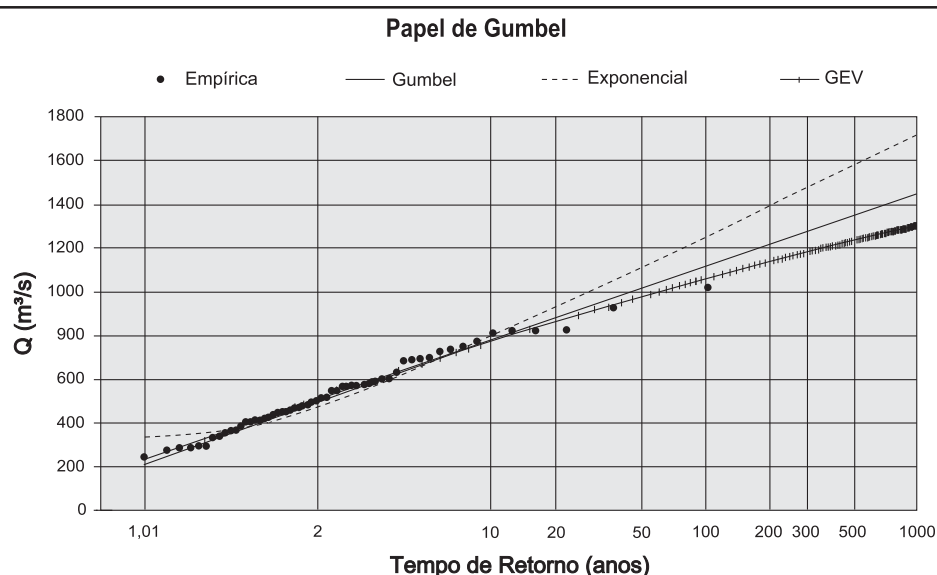


Figura 8.8 – Ajuste das distribuições de Gumbel, Exponencial e GEV

Após a aplicação do teste de Filliben e a verificação do ajuste visual restaram três distribuições candidatas, Log-Normal, Gumbel e Pearson-III. Qualquer um desses modelos pode ser adotado como a distribuição das vazões máximas anuais no rio Paraopeba em Ponte Nova do Paraopeba, ou seja, a partir desse ponto a escolha do modelo incorpora critérios muito subjetivos.

Nesse caso, por se tratar de uma análise de máximos, se o interesse é por tempos de retorno inferiores a 1000 anos, a escolha da distribuição de Gumbel ou da Log-Normal seria praticamente indiferente.

Tabela 8.13 – Probabilidades empíricas

<i>i</i>	AH	Q(m³/s)	Blom	<i>T</i>	Gring	<i>T</i>	<i>i</i>	AH	Q(m³/s)	Blom	<i>T</i>	Gring	<i>T</i>
1	84/85	1017	0,0109	91,6	0,0098	102,0	30	54/55	498	0,5175	1,93	0,5175	1,93
2	90/91	927	0,0284	35,2	0,0273	36,6	31	89/90	481	0,5349	1,87	0,5350	1,87
3	91/92	827	0,0459	21,8	0,0448	22,3	32	68/69	478	0,5524	1,81	0,5525	1,81
4	60/61	822	0,0633	15,8	0,0623	16,0	33	40/41	472	0,5699	1,75	0,5700	1,75
5	78/79	822	0,0808	12,4	0,0798	12,5	34	55/56	470	0,5873	1,70	0,5875	1,70
6	48/49	810	0,0983	10,2	0,0973	10,3	35	41/42	458	0,6048	1,65	0,6050	1,65
7	56/57	774	0,1157	8,6	0,1148	8,7	36	67/68	450	0,6223	1,61	0,6225	1,61
8	63/64	748	0,1332	7,5	0,1324	7,6	37	73/74	449	0,6397	1,56	0,6401	1,56
9	77/78	736	0,1507	6,6	0,1499	6,7	38	59/60	448	0,6572	1,52	0,6576	1,52
10	65/66	726	0,1681	5,9	0,1674	6,0	39	85/86	437	0,6747	1,48	0,6751	1,48
11	82/83	698	0,1856	5,4	0,1849	5,4	40	98/99	427	0,6921	1,44	0,6926	1,44
12	95/96	695	0,2031	4,9	0,2024	4,9	41	92/93	424	0,7096	1,41	0,7101	1,41
13	50/51	690	0,2205	4,5	0,2199	4,5	42	39/40	414	0,7271	1,38	0,7276	1,37
14	42/43	684	0,2380	4,2	0,2374	4,2	43	61/62	414	0,7445	1,34	0,7451	1,34
15	94/95	633	0,2555	3,9	0,2549	3,9	44	43/44	408	0,7620	1,31	0,7626	1,31

Tabela 8.13 – Continuação

<i>i</i>	AH	Q(m ³ /s)	Blom	<i>T</i>	Gring	<i>T</i>	<i>i</i>	AH	Q(m ³ /s)	Blom	<i>T</i>	Gring	<i>T</i>
16	93/94	603	0,2729	3,7	0,2724	3,7	45	58/59	408	0,7795	1,28	0,7801	1,28
17	87/88	601	0,2904	3,4	0,2899	3,4	46	57/58	388	0,7969	1,25	0,7976	1,25
18	83/84	585	0,3079	3,2	0,3074	3,3	47	44/45	371	0,8144	1,23	0,8151	1,23
19	66/67	580	0,3253	3,1	0,3249	3,1	48	49/50	366	0,8319	1,20	0,8326	1,20
20	38/39	576	0,3428	2,9	0,3424	2,9	49	74/75	357	0,8493	1,18	0,8501	1,18
21	46/47	570	0,3603	2,8	0,3599	2,8	50	69/70	340	0,8668	1,15	0,8676	1,15
22	51/52	570	0,3777	2,6	0,3775	2,6	51	45/46	333	0,8843	1,13	0,8852	1,13
23	64/65	570	0,3952	2,5	0,3950	2,5	52	97/98	296	0,9017	1,11	0,9027	1,11
24	71/72	568	0,4127	2,4	0,4125	2,4	53	53/54	295	0,9192	1,09	0,9202	1,09
25	79/80	550	0,4301	2,3	0,4300	2,3	54	52/53	288	0,9367	1,07	0,9377	1,07
26	86/87	549	0,4476	2,2	0,4475	2,2	55	88/89	288	0,9541	1,05	0,9552	1,05
27	72/73	520	0,4651	2,2	0,4650	2,2	56	75/76	276	0,9716	1,03	0,9727	1,03
28	62/63	515	0,4825	2,1	0,4825	2,1	57	70/71	246	0,9891	1,01	0,9902	1,01
29	47/48	502	0,5000	2,0	0,5000	2,0							

Como visto no exemplo 8.6, a seleção do modelo probabilístico que melhor se ajusta aos dados amostrais não é uma tarefa fácil, o que obriga o analista a fazer uso de uma combinação de critérios objetivos e subjetivos. A subjetividade presente no processo de escolha do modelo pode gerar soluções diferenciadas para uma mesma série hidrológica dependendo dos critérios aplicados pelo analista.

De qualquer forma é importante ressaltar que devido ao pequeno tamanho das amostras disponíveis é impossível comprovar que o modelo selecionado representa a verdadeira distribuição populacional.

Dentre as ferramentas disponíveis para a análise de frequência local, os sistemas especialistas computacionais, que emulam os princípios de raciocínio de um especialista humano ao selecionar uma distribuição de probabilidades, têm-se mostrado muito úteis. Um exemplo desse tipo de sistema é o SEAF (Sistema Especialista para Análise de Frequência local de eventos máximos anuais), disponível a partir da URL <http://www.ehr.ufmg.br/downloads.php>.

Exemplo 8.7 – No Anexo 2, encontram-se os dados de vazões mínimas, para diversas durações, da estação fluviométrica de Ponte Nova do Paraopeba, código 40800001. Ajustar as distribuições de Gumbel (de mínimos) e Weibull (de mínimos) às vazões mínimas com duração de 3 dias. Qual distribuição apresenta o melhor ajuste ?

Solução: Inicialmente são calculadas as estatísticas da série: $n = 59$; $\bar{X} = 27,778 \text{ m}^3/\text{s}$; $s = 7,683 \text{ m}^3/\text{s}$; $g = 0,04706$, esse calculado pela equação 8.31. Para ajustar a distribuição de Weibull pelo método do fator de

freqüência, é necessário estimar os parâmetros da equação 8.27 por meio das equações 8.28 a 8.30. Os resultados obtidos foram: $\lambda = 3,417092$; $B(\lambda) = 3,441457$ e $A(\lambda) = 0,34891$. Substituindo esses valores na equação 8.27, referente à estimativa do fator de freqüência e, em seguida, aplicando-a na equação geral de freqüência (equação 8.10), foram calculadas as vazões mínimas associadas a diferentes tempos de retorno, conforme apresentado na Tabela 8.14. Estimando os parâmetros da distribuição de Gumbel para mínimos (ver item 6.7.9), obtém-se $\hat{\alpha} = 5,990612$ e $\hat{\beta} = 31,23533$. Os quantis da distribuição de Gumbel para mínimos são estimados pela equação:

$$x(T) = \beta + \alpha \ln\left(-\ln\left(1 - \frac{1}{T}\right)\right) \quad (8.47)$$

onde β e α são os parâmetros de posição e escala respectivamente, e T é tempo de retorno. A Tabela 8.14 apresenta os quantis calculados pela equação 8.47.

	T (anos)	2	5	10	15	25	50
Weibull	k_t	-0,00111	-0,8738	-1,31127	-1,51884	-1,74288	-1,99398
	Q_T (m³/s)	27,77	21,06	17,70	16,11	14,39	12,46
Gumbel	Q_T (m³/s)	29,04	22,25	17,75	15,22	12,07	7,86

A Figura 8.9 apresenta as distribuições empírica e teóricas, grafadas em um papel de probabilidade de Gumbel. A série foi ordenada de forma crescente e a posição de plotagem da distribuição empírica foi calculada por meio da fórmula de Gringorten, conforme Tabela 8.15.

Analisando a Figura 8.9, percebe-se visualmente que a distribuição de Weibull se ajustou melhor à distribuição empírica. Em algumas análises de vazões mínimas, a primeira tentativa de ajuste entre a distribuição empírica e a teórica não apresenta resultados satisfatórios. Uma das causas pode ser a presença de valores altos na amostra que não permitem o ajuste adequado. Como esses valores estão, em geral, fora da parte de maior interesse da análise, novas tentativas podem ser realizadas para tentar melhorar o ajuste, retirando da série alguns valores elevados e refazendo os cálculos.

Tabela 8.15 – Distribuição empírica das vazões mínimas de Ponte Nova de Paraopeba com 3 dias de duração

<i>m</i>	Ano	Q-3 dias	PP	T (anos)	<i>m</i>	Ano	Q-3 dias	PP	T (anos)
1	1999	11,97	0,009472	105,6	31	1961	27,50	0,516915	1,93
2	1971	12,80	0,026387	37,9	32	1968	27,50	0,533829	1,87
3	1955	15,20	0,043302	23,1	33	1980	28,53	0,550744	1,82
4	1998	15,80	0,060217	16,6	34	1984	28,77	0,567659	1,76
5	1954	17,90	0,077131	13,0	35	1958	28,97	0,584574	1,71
6	1963	17,90	0,094046	10,6	36	1993	29,03	0,601488	1,66
7	1964	18,13	0,110961	9,0	37	1997	29,73	0,618403	1,62
8	1959	19,33	0,127876	7,8	38	1940	29,90	0,635318	1,57
9	1990	20,50	0,14479	6,9	39	1944	30,30	0,652233	1,53
10	1960	20,80	0,161705	6,2	40	1951	30,70	0,669147	1,49
11	1969	21,20	0,17862	5,6	41	1973	30,70	0,686062	1,46
12	1956	21,47	0,195535	5,1	42	1991	31,90	0,702977	1,42
13	1975	21,70	0,212449	4,7	43	1966	32,00	0,719892	1,39
14	1995	21,87	0,229364	4,4	44	1952	32,20	0,736806	1,36
15	1988	22,70	0,246279	4,1	45	1946	32,70	0,753721	1,33
16	1976	23,93	0,263194	3,8	46	1950	34,00	0,770636	1,30
17	1972	24,00	0,280108	3,6	47	1982	34,57	0,787551	1,27
18	1953	24,03	0,297023	3,4	48	1939	34,83	0,804465	1,24
19	1978	24,17	0,313938	3,2	49	1965	34,83	0,82138	1,22
20	1974	24,40	0,330853	3,0	50	1949	35,37	0,838295	1,19
21	1948	24,67	0,347767	2,9	51	1992	35,37	0,85521	1,17
22	1957	24,67	0,364682	2,7	52	1942	37,00	0,872124	1,15
23	1989	24,90	0,381597	2,6	53	1941	37,33	0,889039	1,12
24	1987	24,93	0,398512	2,5	54	1979	37,60	0,905954	1,10
25	1994	25,07	0,415426	2,4	55	1945	38,27	0,922869	1,08
26	1986	25,17	0,432341	2,3	56	1947	38,60	0,939783	1,06
27	1970	25,40	0,449256	2,2	57	1938	42,13	0,956698	1,05
28	1962	25,77	0,466171	2,1	58	1985	44,00	0,973613	1,03
29	1967	26,97	0,483085	2,1	59	1943	50,00	0,990528	1,01
30	1996	27,23	0,5	2,0					

Papel de Gumbel

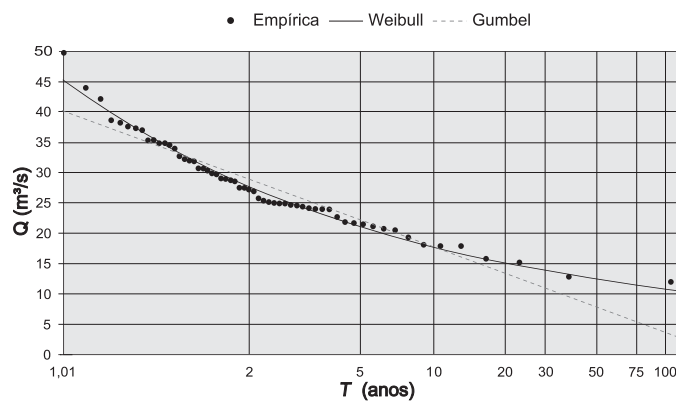


Figura 8.9 – Distribuições ajustadas às vazões mínimas de Ponte Nova de Paraopeba com 3 dias de duração

8.5 – Análise de Frequência de Séries de Duração Parcial

Conforme menção anterior, a modelação probabilística das variáveis hidrológicas pode ser realizada por meio de duas abordagens gerais. A primeira, utilizando as séries de máximos anuais, as quais consideram apenas o maior evento em cada ano hidrológico, e a segunda, empregando as séries de duração parcial (SDP), também denominadas de séries de picos acima de um limiar (POT, da expressão inglesa *Peaks over Threshold*), as quais incluem todos os picos que superaram certo valor de referência ou limiar. A principal objeção à utilização de séries de máximos anuais está relacionada ao fato de se empregar somente o maior evento de cada ano hidrológico, não considerando que o segundo maior evento de um ano pode ser superior aos picos de outros anos, circunstância que é comum em regiões mais secas. A análise com séries de duração parcial evita este tipo de problema, pois considera todos os picos independentes que superam um limite especificado. Entretanto, o uso das séries de duração parcial apresenta a dificuldade adicional de se definir os critérios para identificar somente aqueles eventos superiores ao limite estabelecido que sejam independentes, com a garantia de que não sejam utilizados duas ou mais ocorrências que tenham, como origem, o mesmo mecanismo ou evento causal, conforme comentário no início deste capítulo. Essa dificuldade adicional torna a análise com séries de duração parcial um pouco mais trabalhosa.

Como a série de duração parcial é formada por eventos que superaram um valor limiar, a modelação probabilística para esse tipo de série requer respostas para duas questões importantes. A primeira refere-se à decisão de qual deve ser o modelo que melhor representa a frequência ou a taxa anual de excedências dos eventos maiores que o limiar estipulado, ou seja, qual é a distribuição que descreve o número médio anual de eventos que superaram o valor de referência. A segunda refere-se à decisão de qual deve ser o modelo distributivo das magnitudes das excedências acima do valor limiar. Em geral, a distribuição de Poisson é frequentemente usada para modelar a taxa de excedências dos eventos, enquanto a distribuição exponencial é empregada para descrever a magnitude dos picos excedentes sobre o limiar estabelecido (Stedinger et al. 1993).

Uma vez que, em geral, o interesse volta-se para o intervalo de tempo anual, é necessário calcular as estimativas das probabilidades anuais de excedência a partir das séries de duração parcial. Supondo que o número médio anual de eventos maiores que um limiar estabelecido u , seja um estimador da taxa de excedências v , da distribuição de Poisson, é possível demonstrar, conforme dedução apresentada no Anexo 9, que a relação entre a função acumulada de probabilidades para máximos anuais $F_a(x)$, a razão de ocorrência dos eventos acima do limite

estipulado v , e a distribuição acumulada da série de duração parcial $H_u(x)$ é dada pela seguinte equação:

$$F_a(x) = \exp\{-v[1 - H_u(x)]\} \quad (8.48)$$

Uma vez que a probabilidade de excedência anual é dada por $[1 - F_a(x)]$, a equação 8.48 pode ser alterada para:

$$1 - F_a(x) = 1 - \exp\{-v[1 - H_u(x)]\} \quad (8.49)$$

Como a probabilidade de excedência anual, $[1 - F_a(x)]$, é igual a $1/T_a$, onde T_a é o período de retorno anual, e a correspondente probabilidade de excedência para um valor x em uma série de duração parcial, $[1 - H_u(x)]$, pode ser representada por q_p , verifica-se que a equação 8.49 pode ser transformada em:

$$\frac{1}{T_a} = 1 - \exp\{-vq_i\} \quad (8.50)$$

Segundo Stedinger et al. (1993), o tempo de retorno da série parcial T_p é expresso pela relação

$$T_p = \frac{1}{vq_i} \quad (8.51)$$

Substituindo esse resultado na equação 8.50, obtém-se

$$\frac{1}{T_a} = 1 - \exp\left\{-\frac{1}{T_p}\right\} \quad (8.52)$$

Após algumas transformações da equação 8.52 obtêm-se as seguintes relações:

$$T_a = \frac{1}{1 - \exp\left(-\frac{1}{T_p}\right)} \quad (8.53)$$

ou

$$T_p = \frac{1}{\ln(T_a) - \ln(T_a - 1)} \quad (8.54)$$

A relação entre as funções acumuladas de probabilidades de séries anual e parcial, representada pela equação 8.48, está intrinsecamente relacionada à taxa média dos eventos excedentes v , ou seja ao número médio anual de eventos a ser especificado. Como mencionado no Anexo 9, a experiência de alguns estudos anteriores indica que especificar o valor de \hat{v} entre 2 e 3, parece trazer vantagens para o uso das séries de duração parcial, facilitando, desse modo, a garantia de

independência serial dos eventos selecionados. Além disso, outro aspecto importante na aplicação do modelo expresso pela equação 8.48, é a verificação da adequação da distribuição de Poisson às taxas de excedência dos eventos v . Uma das maneiras de se verificar esta condição é por meio de um teste proposto por Cunnane (1979), o qual se fundamenta na aproximação da distribuição de Poisson pela distribuição Normal. Esse teste encontra-se descrito em detalhes no Anexo 9.

Exemplo 8.8 – Partindo da equação 8.48, deduzir o modelo Poisson-Pareto. Nessa situação, a taxa de excedências é poissoniana e as magnitudes dos eventos que superam o limite estabelecido seguem a distribuição de Generalizada de Pareto. (Ver exemplos 5.5 e 5.10)

Solução: A FAP da distribuição Generalizada de Pareto é dada por:

$$H(x) = 1 - \exp(-y) \quad (8.55)$$

$$\text{com } y = \begin{cases} -\frac{\text{Ln}\left[1 - \frac{k(x-\xi)}{\alpha}\right]}{k} & k \neq 0 \\ \frac{(x-\xi)}{\alpha} & k = 0 \end{cases}$$

onde ξ é o parâmetro de posição, α é o parâmetro de escala e κ é o parâmetro de forma. Os limites de variação de x são: para $k > 0$ $\xi \leq x \leq \xi + \frac{\alpha}{k}$; e para $k \leq 0$ $\xi \leq x < \infty$

Para facilitar a dedução do modelo Poisson-Pareto, as representações de $F_a(x)$ e $H_u(x)$ foram trocadas por $F(x)$ e $H(x)$, respectivamente. Assim, a equação 8.48 foi reescrita como:

$$F(x) = \exp\{-v[1 - H(x)]\} \quad (8.56)$$

Desenvolvendo a equação 8.56 obtém-se:

$$\ln(F(x)) = -v[1 - H(x)]$$

$$\frac{\ln(F(x))}{v} = H(x) - 1$$

$$H(x) = 1 + \frac{1}{v} \ln[F(x)] \quad (8.57)$$

Igualando as equações 8.55 e 8.57, tem-se o desenvolvimento

$$\begin{aligned}
 1 - \exp(-y) &= 1 + \frac{1}{v} \ln[F(x)] \\
 -\exp(-y) &= \frac{1}{v} \ln[F(x)] \\
 \exp(-y) &= -\frac{1}{v} \ln[F(x)] \\
 \exp(-y) &= -\ln[F(x)]^{1/v} \\
 -y &= \ln\left\{-\ln[F(x)]^{1/v}\right\} \\
 y &= -\ln\left\{-\ln[F(x)]^{1/v}\right\} \tag{8.58}
 \end{aligned}$$

• Para $\kappa = 0$, na distribuição Generalizada de Pareto $y = \frac{(x - \xi)}{\alpha}$.

Substituindo y na equação 8.58, segue-se que

$$\frac{(x - \xi)}{\alpha} = -\ln\left\{-\ln[F(x)]^{1/v}\right\}$$

e

$$x = \xi - \alpha \ln\left\{-\ln[F(x)]^{1/v}\right\} \tag{8.59}$$

na qual $F(x) = 1 - \frac{1}{T(\text{anos})}$

Na equação 8.59, tem-se $-\ln[F(x)]^{1/v} = -\frac{1}{v} \ln[F(x)]$ e desenvolvendo

a equação 8.59, os quantis também são dados por

$$\begin{aligned}
 x &= \xi - \alpha \ln\left\{-\frac{1}{v} \ln[F(x)]\right\}, \quad \ln(ab) = \ln(a) + \ln(b) \\
 x &= \xi - \alpha \left\{ \ln\left(\frac{1}{v}\right) + \ln(-\ln[F(x)]) \right\}, \quad \ln\left(\frac{a}{b}\right) = \ln(a) - \ln(b) \\
 x &= \xi - \alpha \{ \ln(1) - \ln(v) + \ln(-\ln[F(x)]) \}, \quad \ln(1) = 0 \\
 x &= \xi - \alpha \{ -\ln(v) + \ln(-\ln[F(x)]) \} \quad \text{ou} \\
 x &= \xi + \alpha \{ \ln(v) - \ln(-\ln[F(x)]) \} \tag{8.60}
 \end{aligned}$$

- Para $\kappa \neq 0$, na distribuição Generalizada de Pareto

$$y = -\frac{\text{Ln}\left[1 - \frac{k(x-\xi)}{\alpha}\right]}{k}. \text{ Substituindo } y \text{ na equação 8.58, segue-se que}$$

$$-\frac{\text{ln}\left[1 - \frac{k(x-\xi)}{\alpha}\right]}{k} = -\text{ln}\left\{-\text{ln}[F(x)]^{\frac{1}{v}}\right\}$$

$$\text{ln}\left[1 - \frac{k(x-\xi)}{\alpha}\right] = k \text{ ln}\left\{-\text{ln}[F(x)]^{\frac{1}{v}}\right\}$$

$$\text{ln}\left[1 - \frac{k(x-\xi)}{\alpha}\right] = \text{ln}\left\{-\text{ln}[F(x)]^{\frac{1}{v}}\right\}^k$$

$$1 - \frac{k(x-\xi)}{\alpha} = \left\{-\text{ln}[F(x)]^{\frac{1}{v}}\right\}^k$$

$$\frac{k(x-\xi)}{\alpha} = 1 - \left\{-\text{ln}[F(x)]^{\frac{1}{v}}\right\}^k$$

$$x = \xi + \frac{\alpha}{k} \left\{1 - \left[-\text{ln}[F(x)]^{\frac{1}{v}}\right]^k\right\} \text{ ou}$$

$$x = \xi + \frac{\alpha}{k} \left\{1 - \left[-\frac{\text{ln}[F(x)]}{v}\right]^k\right\} \quad (8.61)$$

$$\text{na qual } F(x) = 1 - \frac{1}{T(\text{anos})}$$

Em resumo, pode-se dizer que, conhecendo-se a taxa de excedência v e os parâmetros da distribuição Generalizada de Pareto, esses estimados a partir das excedências sobre o limiar estabelecido u , os quantis anuais podem ser calculados por meio das equações 8.60 ou 8.61, conforme o caso.

Exemplo 8.9 – Ajustar o modelo Poisson-Pareto aos dados de uma série de duração parcial de precipitação com duas horas de duração da estação pluviográfica de Entre Rios de Minas, código 02044007. O período de dados disponíveis é de 13 anos hidrológicos (73/74 a 85/86); o valor limiar estabelecido para definição da série é 39 mm e a taxa de excedência v , é igual a 2.

Solução: A primeira etapa consiste em verificar se as taxas de excedências anuais seguem um modelo poissoniano. Esta verificação é realizada com o teste de Cunnane (1979), que está descrito no Anexo 9. Inicialmente é feita a contagem do número de eventos por ano que superam o limite estabelecido. Esses valores permitem a estimativa da estatística do teste de Cunnane, equação A9.11 do Anexo 9. O número de excedências e a estatística de Cunnane estão na Tabela 8.16.

Tabela 8.16 – Contagem das excedências anuais

AH	73/74	74/75	75/76	76/77	77/78	78/79	79/80	80/81	81/82	82/83	83/84	84/85	85/86	
<i>m</i>	1	3	1	2	2	3	2	4	2	2	0	3	1	Soma
<i>γ</i>	0,5	0,5	0,5	0	0	0,5	0	2	0	0	2	0,5	0,5	7

O valor da estatística do teste deve ser comparado ao quantil $\chi^2_{1-\alpha, n}$ da distribuição do Qui-Quadrado, com 12 graus de liberdade ($n-1$), e nível de significância α de 5%. Analisando o Anexo 6, verifica-se que $\chi^2_{0,95;12}$ é igual a 21. Como a estatística de Cunnane, $\gamma = 7$, é menor que o quantil $\chi^2_{0,95;12}$ da distribuição Qui-Quadrado, aceita-se, a um nível de significância de 5%, a hipótese de que as excedências anuais ocorrem segundo um modelo poissoniano.

Em seguida, pode-se calcular a distribuição empírica por meio da estimativa da posição de plotagem e dos tempos de retorno da série parcial e o seu correspondente anual. O cálculo da posição de plotagem foi realizado com a fórmula de Gringorten, $q_i = (i - 0,44)/(n + 0,12)$; o tempo de retorno parcial foi estimado com equação 8.51, $T_p = 1/\nu q_i$, e o seu correspondente anual com a equação 8.53, $T_a = 1/\{1 - \exp(-1/T_p)\}$. A série parcial e os resultados de cálculo estão na Tabela 8.17.

A FAP da distribuição Generalizada de Pareto está apresentada no exemplo 8.8, equação 8.55. A estimativa dos parâmetros pelo método dos momentos é realizada do modo descrito a seguir.

ξ é o parâmetro de posição e nesse caso é igual ao limite estabelecido, ou seja, 39 mm, α é o parâmetro de escala, estimado por

$$\hat{\alpha} = \frac{\bar{X}}{2} \left(\frac{\bar{X}^2}{S_X^2} + 1 \right) \quad \text{com } X = x_i - \xi \quad (8.62)$$

e k é o parâmetro de forma, cuja estimativa é dada por

$$\hat{k} = \frac{1}{2} \left(\frac{\bar{X}^2}{S_x^2} - 1 \right) \text{ com } X = x_i - \xi \quad (8.63)$$

onde \bar{X} e S_x são a média e o desvio padrão amostrais da variável $X = x - \xi$

A média e o desvio padrão da variável $X = x - \xi$ são, respectivamente, 10,57692mm e 11,06318mm. Os parâmetros estimados são $\hat{\alpha} = 10,12226$ e $\hat{k} = -0,04299$. Como o parâmetro de forma, k , é negativo, o cálculo dos quantis anuais é realizado pela equação 8.61, ou seja,

$$x(F) = 39 + \frac{10,12226}{-0,04299} \left\{ 1 - \left[-\frac{\text{Ln}(F(x))}{2} \right]^{-0,04299} \right\} \quad (8.64)$$

na qual $F(x) = 1 - \frac{1}{T(\text{anos})}$

Tabela 8.17 – Cálculo da distribuição empírica do exemplo 8.9

<i>i</i>	AH	P (mm)	X - ξ	<i>q_e</i>	<i>T_p</i>	<i>T_a</i>
1	77/78	80	41	0,0214	23,3214	23,8250
2	84/85	73,4	34,4	0,0597	8,3718	8,8817
3	80/81	64,1	25,1	0,0980	5,1016	5,6179
4	84/85	63,1	24,1	0,1363	3,6685	4,1912
5	78/79	61,1	22,1	0,1746	2,8640	3,3931
6	81/82	57,2	18,2	0,2129	2,3489	2,8843
7	77/78	55,6	16,6	0,2511	1,9909	2,5325
8	82/83	53,1	14,1	0,2894	1,7275	2,2755
9	73/74	51,1	12,1	0,3277	1,5257	2,0799
10	84/85	48,6	9,6	0,3660	1,3661	1,9266
11	79/80	48,4	9,4	0,4043	1,2367	1,8034
12	81/82	48,3	9,3	0,4426	1,1298	1,7026
13	75/76	47,4	8,4	0,4809	1,0398	1,6187
14	76/77	44,5	5,5	0,5191	0,9631	1,5481
15	80/81	44,3	5,3	0,5574	0,8970	1,4880
16	80/81	43,6	4,6	0,5957	0,8393	1,4363
17	82/83	43,3	4,3	0,6340	0,7886	1,3916
18	85/86	41,2	2,2	0,6723	0,7437	1,3525
19	78/79	41	2	0,7106	0,7037	1,3183
20	80/81	40,8	1,8	0,7489	0,6677	1,2881
21	74/75	40,5	1,5	0,7871	0,6352	1,2613
22	78/79	40,2	1,2	0,8254	0,6058	1,2375
23	74/75	40	1	0,8637	0,5789	1,2162
24	74/75	39,6	0,6	0,9020	0,5543	1,1971
25	76/77	39,4	0,4	0,9403	0,5318	1,1799
26	79/80	39,2	0,2	0,9786	0,5110	1,1645

$n = 26; v = 2; u = \xi = 39 \text{ mm}$ e o número de anos igual a 13

A equação 8.64 permite que se calcule os quantis anuais associados a diferentes tempos de retorno. Alguns quantis estão apresentados na Tabela 8.18. A Figura 8.10 apresenta os quantis anuais do modelo Poisson-Pareto, calculados pela equação 8.64, e os quantis empíricos, apresentados da Tabela 8.17, ambos grafados em um papel de probabilidade de Gumbel.

Tabela 8.18 – Quantis anuais - Modelo Poisson-Pareto

T (anos)	2	5	10	20	30	50	75	100
Quantis (mm)	50,0	62,3	70,8	79,2	84,1	90,4	95,5	99,2

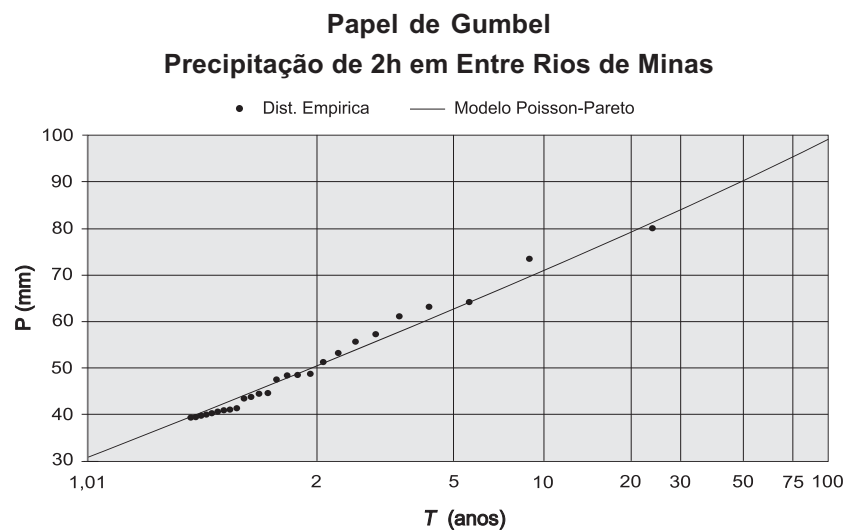


Figura 8.10 – Ajuste do modelo Poisson-Pareto à distribuição empírica

Exercícios

- 1) Construir os papeis de probabilidade das seguintes distribuições:
 - a) Log-Normal de 2 parâmetros
 - b) Gumbel

- 2) Grafar os dados de vazões médias anuais do rio Paraopeba em Ponte Nova do Paraopeba, apresentados na Tabela 7.1, nos papeis de probabilidade Normal e Log-Normal. Qual dessas distribuições parece se ajustar melhor aos dados?

- 3) Grafar os dados de vazões médias diárias máximas anuais do rio Paraopeba em Ponte Nova do Paraopeba, apresentados no Anexo 2, utilizando todas as fórmulas de posição de plotagem da Tabela 7.19 no mesmo papel de probabilidade. Avaliar as diferenças entre os resultados.

4) Os dados da Tabela 8.19 referem-se às descargas médias diárias máximas anuais (m^3/s) observadas no rio Hercílio em Ibirama, Sta. Catarina. A área de drenagem é de 3314 km^2 . Utilize o papel de probabilidades de Gumbel, construído para o exercício 1, e plote as vazões versus posição de plotagem, utilizando a fórmula de Gringorten $[(m-0,44)/(n+0,12)]$.

Tabela 8.19 – Dados do exercício 4

Ano	Q (m^3/s)	Ano	Q (m^3/s)	Ano	Q (m^3/s)	Ano	Q (m^3/s)	Ano	Q (m^3/s)
1935	1342	1945	474	1955	969	1965	708	1975	1406
1936	625	1946	763	1956	566	1966	998	1976	801
1937	619	1947	592	1957	1300	1967	477	1977	741
1938	797	1948	981	1958	526	1968	298	1978	1002
1939	1250	1949	438	1959	520	1969	872	1979	1090
1940	271	1950	281	1960	487	1970	483	1980	faltoso
1941	263	1951	556	1961	897	1971	1040	1981	589
1942	566	1952	393	1962	582	1972	1010	1982	490
1943	649	1953	726	1963	510	1973	1240	1983	2475
1944	236	1954	897	1964	faltoso	1974	697	1984	2125

5) Ajuste a distribuição de Gumbel à amostra do rio Hercílio em Ibirama, calcule os quantis para $T = 5, 25, 50, 100$ e 500 anos pelo método dos fatores de frequência e plote a reta obtida no gráfico do exercício 1. Calcule e plote também os intervalos de confiança a um nível $100(1-\alpha) = 95\%$ para os quantis estimados.

6) Ajuste uma distribuição log-Pearson III à amostra do rio Hercílio em Ibirama, calcule os quantis e os intervalos de confiança a 95% , correspondentes a $T = 5, 25, 50, 100$ e 500 anos, usando o método dos fatores de frequência.

7) O programa ALEA, disponível para download a partir da URL <http://www.ehr.ufmg.br>, contém rotinas para ao ajuste da distribuição GEV, pelos métodos dos momentos e da máxima verossimilhança. Utilize o programa ALEA para ajustar a distribuição Generalizada Valores Extremos à amostra do rio Hercílio em Ibirama, pelo método da máxima verossimilhança. Calcule os quantis para $T = 5, 25, 50, 100$ e 500 anos pela expressão da função inversa da GEV. Plote os quantis da GEV no gráfico do exercício 5 e comente sobre a influência do sinal do parâmetro de forma nos resultados obtidos.

8) A Tabela 8.20 apresenta os dados de precipitação máxima diária de Caeté, código 01943010. Realizar uma análise analítica de frequência de máximos, calculando os parâmetros das distribuições candidatas pelo método dos momentos-L.

Tabela 8.20 – Dados do exercício 8

AH	P (mm)	AH	P (mm)	AH	P (mm)	AH	P (mm)
41/42	72,8	53/54	87	77/78	210,2	89/90	97,7
42/43	69,4	54/55	112,8	78/79	92,1	90/91	116,2
43/44	77,8	56/57	80,1	79/80	86,5	91/92	100,9
44/45	74,2	58/59	95,7	80/81	86,3	92/93	66,2
45/46	102,2	59/60	102,3	81/82	123,6	93/94	84,2
46/47	93,4	60/61	105,5	82/83	84,6	94/95	93,4
47/48	75	64/65	75,9	83/84	64,6	95/96	147,1
48/49	117,4	66/67	112,7	84/85	80,7	96/97	118,2
49/50	47,2	67/68	50,7	85/86	73	97/98	67,5
50/51	67,4	69/70	82,8	86/87	83,4	98/99	107,3
51/52	76	70/71	52	87/88	73,6	99/00	102,8
52/53	102,6	76/77	66,9	88/89	57,2		

9) Repetir o exercício 8 realizando a análise de frequência com métodos do fator de frequência utilizando as mesmas distribuições candidatas. Comparar com os resultados do exercício 8.

10) A série utilizada nos exercício 8 apresenta um *outlier* de 210,2mm. Qual é a probabilidade desse evento atípico ocorrer em período de 50 anos, admitindo que as precipitações máximas diárias de Caeté seguem a distribuição ajustada no exercício 8.

11) O Anexo 2 apresenta os dados de vazões mínimas com duração de 7 dias da estação fluviométrica de Ponte Nova do Paraopeba, código 40800001. Utilizando esses dados, ajustar as seguintes distribuições:

a) Gumbel para mínimos com os parâmetros calculados pelo método dos momentos-L

b) Weibull (2P) com os parâmetros calculados pelo método dos momentos-L

Para a solução deste exercício, destaca-se que Stedinger et al. (1993) indicam que, caso uma variável aleatória X se ajuste à distribuição de Weibull, então a variável $Y = -\ln[X]$ se distribui segundo a distribuição de Gumbel. Assim, os procedimentos de estimativa dos parâmetros e os testes de ajuste disponíveis para a distribuição de Gumbel podem ser utilizados para a distribuição de Weibull.

Desse modo, $+\ln[X]$ possui uma média $\lambda_{1,(\ln X)}$ e o momento-L $\lambda_{2,(\ln X)}$, os parâmetros de ajuste da distribuição de Weibull (2P) para a variável X são os seguintes:

$$k = \frac{\ln(2)}{\lambda_{2,(\ln X)}} \quad (8.65)$$

$$\alpha = \exp\left(\lambda_{1,(\ln X)} + \frac{0,5772}{k}\right) \quad (8.66)$$

Para efetuar o ajuste da distribuição de Weibull, é necessário calcular os logaritmos naturais dos valores das séries. Em seguida, são calculados os momentos-L. Os valores dos momentos-L permitem o cálculo dos parâmetros da distribuição de Weibull através das equações 8.65 e 8.66.

12) Grafar as distribuições empírica e teóricas ajustadas no exercício 11 em um papel de probabilidades de Gumbel utilizando a fórmula de posição de plotagem de Gringorten.

13) Considerando o exercício 11, calcular a probabilidade de ocorrência de vazões mínimas com duração de 7 dias inferiores a $Q_{7,10}$ durante um período de 5 anos.

14) Montar uma tabela que contenha os valores do fator de frequência da distribuição de Weibull, para mínimos, em função da assimetria amostral e do tempo de retorno.

15) Montar uma tabela que contenha os valores do fator de frequência da distribuição de Gumbel (máximos), em função do tempo de retorno e do tamanho da amostra.

16) Ajustar o modelo Poisson-Pareto à série de duração parcial de precipitações, com duração de 3h, na estação de Pium-í, código 02045012, apresentada na Tabela 8.21. Essa série se refere ao período de anos (75/76 a 85/86), com taxa média de excedências de 2 eventos por ano para um limite de 44,5mm.

Tabela 8.21 – Dados do exercício 16

AH	P (mm)	AH	P (mm)	AH	P (mm)	AH	P (mm)
75/76	70,2	78/79	47,6	81/82	53	83/84	46,6
75/76	50	79/80	49,8	82/83	47,9	84/85	72,2
76/77	47,2	79/80	46	82/83	59,4	84/85	46,4
77/78	52	79/80	46,8	82/83	50,2	85/86	48,4
77/78	47,6	80/81	50,6	82/83	53,4		
77/78	47,4	81/82	44,1	82/83	59,4		

17) O Rio Alva em Ponte de Mucela, em Portugal, apresenta um número médio de 3 excedências por ano sobre a descarga de referência de 65 m³/s. Testes estatísticos comprovaram serem plausíveis as hipóteses nulas do número Poissoniano de excedências, independência serial e exponencialidade da cauda superior, a um nível de significância de 5%. Se a média das excedências é de 72,9 m³/s, estime a descarga máxima anual de tempo de retorno 500 anos.

18) A Tabela 8.22 apresenta as 205 maiores enchentes ao longo dos 72 anos contínuos (1896-1967) de registros fluviométricos do Rio Greenbrier em Alderson (West Virginia, EUA) e que excederam 17000 cfs.

a) Escolha o maior valor possível para o número médio anual de cheias ($\hat{\Lambda}$), tal que as excedências possam ser modeladas por um processo de Poisson. Verifique a conveniência de sua escolha através do teste da hipótese Poissoniana pela estatística

$$\gamma = \sum_{k=1}^N \left(\frac{m_k - \hat{\Lambda}}{\hat{\Lambda}} \right)^2 \text{ lembrando que essa segue uma distribuição do Qui-Quadrado}$$

com $(N-1)$ graus de liberdade, onde N indica o número de anos de registros, e que o número de excedências que ocorrem no ano k é representado por m_k .

b) Depois de escolhido o maior valor possível para o número médio anual de enchentes, modele-as através da distribuição generalizada de Pareto dada por

$$H(x) = 1 - \exp(-y) \text{ onde } y = -\frac{\ln \left[1 - \frac{k}{\alpha} (x - \xi) \right]}{k} \text{ para } k \neq 0, \quad y = \frac{x - \xi}{\alpha} \text{ para}$$

$k = 0$ e ξ , α e k são, respectivamente, os parâmetros de posição, escala e forma. Lembre-se que a distribuição generalizada de Pareto é ilimitada superiormente para $k \leq 0$ e possui limite superior para $k > 0$. Observe que, quando $k = 0$, ela se reduz à distribuição exponencial com parâmetros ξ e α .

c) Calcule os quantis de cheias anuais para diversos tempos de retorno (2 a 1000 anos) invertendo a expressão da função de distribuição acumulada de probabilidades anuais do modelo Poisson-Pareto dada por $F(x) = \exp\{-\Lambda[1 - H(x)]\}$. Plote os quantis com o tempo de retorno (T em coordenadas logarítmicas).

**Tabela 8.22 – Vazões do rio Greenbrier em Alderson
(West Virginia, EUA) superiores a 17.000 cfs**

Ano	Q (cfs)	Ano	Q (cfs)	Ano	Q (cfs)	Ano	Q (cfs)
1896	28800	1915	34000	1935	20800	1954	29700
1897	27600		40800	1936	19400		18800
	54000	1916	27200		20800	1955	32000
	40900		24400		27100		28000
1898	17100	1917	17300		58600		44400
	18600		43000		28300		26200
	52500		28000	1937	21200	1956	18200
1899	25300	1918	17900		22300	1957	23900
	20000		77500		36600		28900
	23800		24000		26400		22000
	48900	1919	28600	1938	21200	1958	21800
1900	17100		24800		32800		23900
1901	56800		49000		22300		22200
	21100	1920	38000	1939	40200		17500
	20400		20700		41600		26700
	19300		33500		21200	1959	17200
	20000	1922	21500		17200		23900
1902	36700		20100		19400	1960	17800
	43800		22200	1940	29900		35500
1903	25300	1923	19500		21500		32500
	29600	1924	26500		19400	1961	25000
	33500		20400		18700		21800
	34400		36200	1942	35300		31400
	48900		17900	1943	33600		17200
1904	25700	1926	20700		17200	1962	34700
	25700		17600		36200		20100
1905	29600	1927	17900		21200		21500
	37600		24000	1944	25200		17800
1906	18200		40200		17200		23200
	26000		18800	1945	17900		35500
1907	17500		19500		19000	1963	22700
	52500	1928	18000	1946	43600		34800
1908	17800	1929	22800	1947	20000		47200
	23000		32700		24400		26100
	31500		23800	1948	35200		30400
	52500		20000		23500	1964	19100
	26800	1930	36600		40300		39600
	27600	1932	50100	1949	18500		22800
	31500		17600		37100	1965	22000
1909	20000		31500		26300		28400
1910	45900		27500		23200		19800
1911	43800		21900	1950	31500		18600
	20000	1933	26400	1951	25600	1966	26400
	23800	1934	32300		27800	1967	54500
	18900		20500		26700		39900
	18900		27900		18500		20900
	35500	1935	19400		19800		
	27200		49600		29300		
	20000		22300	1952	17800		
	21100		17900		19100		
1913	21800		24800		27600		
	64000		20100	1953	47100		
	20000		24800		20100		

