

UNIVERSIDADE FEDERAL DO PARANÁ

Matheus Henrique Tavares

Avaliação do método dos momentos-L no
ajuste de distribuições a máximos
hidrológicos

Curitiba
2015

Matheus Henrique Tavares

Avaliação do método dos momentos-L no
ajuste de distribuições a máximos
hidrológicos

Trabalho apresentado como requisito parcial à obtenção do grau de bacharel em Engenharia Ambiental, Setor de Tecnologia da Universidade Federal do Paraná.

Orientador: Paulo Justiniano Ribeiro Jr.

Co-orientador: Maurício Felga Gobbi

Curitiba
2015

Agradecimentos

Agradeço primeiramente ao meu orientador, Paulo Justiniano, por ter aceitado a ideia de me orientar, mesmo sem haver um tema bem definido inicialmente, e por ser tão dedicado e entusiasmado em ensinar.

Agradeço ao professor Maurício, que propôs o tema, e que ajudou com toda a parte de hidrologia presente neste trabalho.

Agradeço à minha namorada Marianna, pelo carinho e suporte em todas as horas.

Agradeço a meus familiares e amigos, por me aturar.

Agradeço a todos que contribuíram, direta ou indiretamente, com este trabalho.

Muito obrigado!

Lista de Figuras

- 5.1 Densidade dos dados de vazão máxima anual do Rio dos Patos, para o período de 1938 a 2005, e as distribuições teóricas de Gumbel para os parâmetros estimados pelo Método dos Momentos (MOM), pelo Método da Máxima Verossimilhança (MMV) e pelo Método dos Momentos-L (MML) 26
- 5.2 Densidade dos dados de vazão máxima anual do Rio dos Patos e as distribuições teóricas de Gumbel para os parâmetros estimados pelo Método dos Momentos (MOM), pelo Método da Máxima Verossimilhança (MMV) e pelo Método dos Momentos-L (MML), na região da cauda à direita 27
- 5.3 Densidade dos dados de vazão máxima anual do Rio São Salvador, para o período de 1977 a 2008, e as distribuições teóricas de Pearson tipo III para os parâmetros estimados pelo Método dos Momentos (MOM), pelo Método da Máxima Verossimilhança (MMV) e pelo Método dos Momentos-L (MML) 29
- 5.4 Densidade dos dados de vazão máxima anual do Rio São Sebastião, para o período de 1977 a 2008, desconsiderando o ano de 1981, e as distribuições teóricas de Pearson tipo III para os parâmetros estimados pelo Método dos Momentos (MOM), pelo Método da Máxima Verossimilhança (MMV) e pelo Método dos Momentos-L (MML) 30
- 5.5 Densidade dos dados de vazão máxima anual do Rio Paraguai, para o período de 1968 a 2007, desconsiderando os anos de 1989, 1990, 1991 e 1992, e as distribuições teóricas de Weibull para os parâmetros estimados pelo Método dos Momentos (MOM), pelo Método da Máxima Verossimilhança (MMV) e pelo Método dos Momentos-L (MML) 32

5.6	Densidade dos dados de vazão máxima anual do Rio Paraguai e as distribuições teóricas de Weibull para os parâmetros estimados pelo Método dos Momentos (MOM), pelo Método da Máxima Verossimilhança (MMV) e pelo Método dos Momentos-L (MML), na região da cauda à direita	34
6.1	Densidade das vazões para os tempos de retorno de 50, 100 e 1000 anos estimadas pelo Método dos Momentos (MOM), pelo Método da Máxima Verossimilhança (MMV) e pelo Método dos Momentos-L (MML), para a simulação Pearson tipo III – Pearson tipo III, $n = 30$	37
6.2	Densidade das vazões para tempo de retorno de 50, 100 e 1000 anos estimadas pelo Método dos Momentos (MOM), pelo Método da Máxima Verossimilhança (MMV) e pelo Método dos Momentos-L (MML), para a simulação Gumbel – Gumbel, $n = 30$, com contaminação	40
6.3	Densidade das vazões para tempo de retorno de 50, 100 e 1000 anos estimadas pelo Método dos Momentos (MOM), pelo Método da Máxima Verossimilhança (MMV) e pelo Método dos Momentos-L (MML), para a simulação Gumbel – Pearson tipo III, $n = 50$, com contaminação	43
6.4	Densidade das vazões para tempo de retorno de 50, 100 e 1000 anos estimadas pelo Método dos Momentos (MOM), pelo Método da Máxima Verossimilhança (MMV) e pelo Método dos Momentos-L (MML), para a simulação Gumbel – Weibull, $n = 50$, com contaminação	45
6.5	Densidade das vazões para tempo de retorno de 50, 100 e 1000 anos estimadas pelo Método dos Momentos (MOM), pelo Método da Máxima Verossimilhança (MMV) e pelo Método dos Momentos-L (MML), para a simulação Pearson tipo III – Weibull, $n = 30$	47
6.6	Densidade das vazões para tempo de retorno de 50, 100 e 1000 anos estimadas pelo Método dos Momentos (MOM), pelo Método da Máxima Verossimilhança (MMV) e pelo Método dos Momentos-L (MML), para a simulação Weibull – Gumbel, $n = 30$, com contaminação	49

Lista de Tabelas

5.1	Vazões máximas anuais do rio dos Patos, para o período de 1938 a 2005 . . .	25
5.2	Vazões para os tempos de retorno de 50, 100 e 1000 anos estimadas pelos Métodos dos Momentos (MOM), da Máxima Verossimilhança (MMV) e dos Momentos-L (MML) para o rio dos Patos	25
5.3	Vazões máximas anuais do rio São Salvador, para o período de 1977 a 2008	28
5.4	Vazões para os tempos de retorno de 50, 100 e 1000 anos estimadas pelos Métodos dos Momentos (MOM), da Máxima Verossimilhança (MMV) e dos Momentos-L (MML) para o rio São Salvador, considerando e desconsiderando o valor extremo	31
5.5	Vazões máximas anuais do rio Paraguai, para o período de 1968 a 2007 . .	31
5.6	Vazões para os tempos de retorno de 50, 100 e 1000 anos estimadas pelos Métodos dos Momentos (MOM), da Máxima Verossimilhança (MMV) e dos Momentos-L (MML) para o rio Paraguai	33
6.1	Viés, variância e erro quadrático médio (EQM) das vazões para os tempos de retorno de 50, 100 e 1000 anos estimadas pelos Métodos dos Momentos (MOM), da Máxima Verossimilhança (MMV) e dos Momentos-L (MML), para a simulação Pearson tipo III – Pearson tipo III, $n = 30$	36
6.2	Viés, variância e erro quadrático médio (EQM) das vazões para os tempos de retorno de 50, 100 e 1000 anos estimadas pelos Métodos dos Momentos (MOM), da Máxima Verossimilhança (MMV) e dos Momentos-L (MML), para a simulação Pearson tipo III – Pearson tipo III, $n = 50$	38

6.3	Viés, variância e erro quadrático médio (EQM) das vazões para os tempos de retorno de 50, 100 e 1000 anos estimadas pelos Métodos dos Momentos (MOM), da Máxima Verossimilhança (MMV) e dos Momentos-L (MML), para a simulação Gumbel – Gumbel, $n = 30$, com contaminação	39
6.4	Viés, variância e erro quadrático médio (EQM) das vazões para os tempos de retorno de 50, 100 e 1000 anos estimadas pelos Métodos dos Momentos (MOM), da Máxima Verossimilhança (MMV) e dos Momentos-L (MML), para a simulação Weibull – Weibull, $n = 30$	39
6.5	Viés, variância e erro quadrático médio (EQM) das vazões para os tempos de retorno de 50, 100 e 1000 anos estimadas pelos Métodos dos Momentos (MOM), da Máxima Verossimilhança (MMV) e dos Momentos-L (MML), para a simulação Gumbel – Pearson tipo III, $n = 50$, com contaminação . .	42
6.6	Viés, variância e erro quadrático médio (EQM) das vazões para os tempos de retorno de 50, 100 e 1000 anos estimadas pelos Métodos dos Momentos (MOM), da Máxima Verossimilhança (MMV) e dos Momentos-L (MML), para a simulação Gumbel – Weibull, $n = 50$, com contaminação	44
6.7	Viés, variância e erro quadrático médio (EQM) das vazões para os tempos de retorno de 50, 100 e 1000 anos estimadas pelos Métodos dos Momentos (MOM), da Máxima Verossimilhança (MMV) e dos Momentos-L (MML), para a simulação Pearson tipo III – Weibull, $n = 30$	46
6.8	Viés, variância e erro quadrático médio (EQM) das vazões para os tempos de retorno de 50, 100 e 1000 anos estimadas pelos Métodos dos Momentos (MOM), da Máxima Verossimilhança (MMV) e dos Momentos-L (MML), para a simulação Weibull – Gumbel, $n = 30$, com contaminação	48

Sumário

Resumo	ix
Abstract	xi
1 Introdução	1
2 Objetivos	4
2.1 Objetivo Geral	4
2.2 Objetivos específicos	4
3 Referencial Teórico	5
3.1 Métodos	5
3.1.1 Método dos Momentos	5
3.1.2 Método da Máxima Verossimilhança	7
3.1.3 Método dos Momentos-L	9
3.2 Distribuições	11
3.2.1 Distribuição de Gumbel	11
3.2.2 Distribuição de Pearson tipo III	14
3.2.3 Distribuição de Weibull	18
4 Metodologia	21
5 Aplicação a dados reais	24
5.1 Rio dos Patos	24
5.2 Rio São Salvador	26
5.3 Rio Paraguai	30

6	Estudo de simulação	35
6.1	Simulações com apenas uma distribuição	35
6.1.1	Pearson tipo III – Pearson tipo III, $n = 30$	35
6.1.2	Pearson tipo III – Pearson tipo III, $n = 50$	37
6.1.3	Gumbel – Gumbel, $n = 30$, com contaminação	38
6.1.4	Weibull – Weibull, $n = 30$	39
6.1.5	Discussão	40
6.2	Simulações com duas distribuições	41
6.2.1	Gumbel – Pearson tipo III, $n = 50$, com contaminação	42
6.2.2	Gumbel – Weibull, $n = 50$, com contaminação	43
6.2.3	Pearson tipo III – Weibull, $n = 30$	44
6.2.4	Weibull – Gumbel, $n = 30$, com contaminação	45
6.2.5	Discussão	46
7	Conclusão	51
	Referências Bibliográficas	53
	Anexos	57

Resumo

Em hidrologia, estudos de previsão de eventos extremos de vazão ou de precipitação contribuem para evitar danos a estruturas suscetíveis a tais eventos. A análise estatística por meio de distribuições de probabilidades é uma das formas de se estudar estes eventos. Partindo do pressuposto de que os dados de vazão e precipitação anuais, sejam eles de máximo ou mínimo, são variáveis aleatórias independentes e representam o comportamento hidrológico da região estudada, uma distribuição de probabilidades é ajustada a estes dados e utilizada para obter estimativas de quantis e probabilidade de não-excedência. Para o ajuste destas curvas, os parâmetros da distribuição devem ser estimados a partir dos dados. Na literatura, dois métodos se destacam nesta área: o método dos momentos e o método da máxima verossimilhança. Contudo, estudos mostram que os parâmetros estimados por estes métodos nem sempre são satisfatórios sob determinadas condições que são comuns a dados de vazão máxima, como amostras pequenas de dados e presença de valor atípicos. Hosking (1990) introduziu o método dos momentos-L e estudos mostram que seu desempenho pode ser melhor que os referidos métodos no ajuste sob tais condições. Este trabalho se propôs a comparar o método dos momentos-L com os métodos já estabelecidos da máxima verossimilhança e dos momentos no ajuste de distribuições de probabilidade a dados hidrológicos de vazão máxima anual. Para isto, foram feitos ajustes a dados reais e estudos de simulação, utilizando dois tamanhos de amostra, 30 e 50. O estudo incluiu situações com amostras contaminadas por valores extremos. Considerou-se a três distribuições de probabilidade: Gumbel, Pearson tipo III e Weibull de 3 parâmetros. Verificou-se que, para amostras grandes, os três métodos fizeram boas estimativas, porém o da máxima verossimilhança foi considerado o mais indicado. Para amostras pequenas, o método dos momentos se mostrou o melhor estimador. Nas simulações, o método dos momentos-L nunca apresentou o menor erro quadrático médio comparado com os demais. Também foi observado que os valores extremos tiveram impacto bastante negativo nas estimações, e no caso de dados reais, o método dos momentos se mostrou o mais sensível. Além disso, os resultados reiteram a importância da escolha criteriosa de uma distribuição para o ajuste em estudos hidrológicos deste ramo.

Palavras-chave: Hidrologia estatística, ajuste de curvas, métodos de estimação, eventos

extremos, tempos de retorno

Abstract

In hydrology, studies to predict extreme flood or rainfall events contribute to avoid damage to structures that are susceptible to such events. Statistical analysis is one of the ways to study these events. Starting from the assumption that the flood and rainfall annual data, whether they are of maxima or minima, are random variables that represent the hydrological behaviour of the region studied, a probability distribution is fitted to these data and used to estimate its quantiles and non-exceedance probabilities. To fit these curves, the distribution parameters must be estimated from the data. In the literature, two methods stand out in this area: the Method of Moments and the Method of Maximum Likelihood. However, studies show that the parameters estimated by these methods are not always satisfying under some conditions common to annual maximum flood data, such as small samples and outliers. Hosking (1990) introduced the Method of L-Moments recently, and studies show that its performance may be better than the referred methods in the fitting under these conditions. This work proposes to compare the Method of L-Moments with the well established Methods of Moments and Maximum Likelihood in the fitting of probability distributions to annual flood maxima hydrological data. To do so, curves were fitted to real data and simulations were made, using two sample sizes, 30 and 50. The study included cases where the samples were contaminated with extreme values. Three probability distributions were accounted for: Gumbel, Pearson type III and three-parameter Weibull distributions. It was found that, to large samples, all three methods estimated the parameters well, but the Maximum Likelihood was considered the most recommendable. To small samples, the Method of Moments was the best estimator. In the simulations, the Method of L-Moments was never the estimator with the least mean squared error, when compared to the other methods. It was also observed that the extreme values had a negative impact on the estimations, and in the case of real data, the Method of Moments was the most sensitive to it. Furthermore, this study reiterates the importance of choosing judiciously a distribution for the fitting in hydrological studies in this field.

Keywords: Hydrologic statistics, curve fitting, estimation methods, extreme events, return periods

1. Introdução

Em hidrologia, um dos principais problemas está na previsão de eventos extremos. No projeto das ruas de uma cidade, é importante saber, por exemplo, um certo volume alto de chuva que seja razoavelmente frequente naquela cidade, para embasar o dimensionamento das galerias e dos bueiros que captam a água da chuva, de forma a evitar alagamentos (Serrano, 1997). Porém, fazer esta previsão baseada em todos os processos físicos que influenciam esta precipitação extrema é uma tarefa bastante complexa. Por isto, em hidrologia, um modo bastante comum de se fazer este tipo de previsão é através de análise estatística. Esta ferramenta utiliza os dados disponíveis, de vazão de um rio ou de precipitação em uma bacia, para fazer previsões a respeito da expectativa de ocorrência temporal de eventos de determinada magnitude (Chow et al., 1988). Neste ramo denominado hidrologia estatística, usualmente os dados são tratados como realizações independentes de uma variável aleatória, e é ajustada uma distribuição de probabilidade ao conjunto de dados, de forma a se obter uma função de probabilidades que resume este conjunto e caracteriza os eventos. Desta forma, tendo uma série histórica de dados de uma região, seja ela, por exemplo, de mínimas de vazão anuais ou de máximas de precipitação diária mensais, é possível se obter uma curva que represente estes dados. Com isto, previsões podem ser feitas para a região em questão, utilizando quantis e probabilidades de não-excedência para valores que a função pode assumir (Serrano, 1997). Isto é muito empregado, por exemplo, na previsão de eventos extremos, como secas e enchentes, e de vazões de projeto de estruturas vulneráveis a enchentes, como pontes e portos fluviais. A desvantagem desta metodologia, porém, se deve à limitação quanto aos dados utilizados: dados de vazão e precipitação diárias e mensais podem ser correlacionados, e, portanto, métodos mais simples, que supõe independência entre observações não podem ser empregados. Por outro lado, dados de maior suporte ou espaçamento temporal apresentam comportamento independente e desta forma, a análise

estatística fica, na prática, muitas vezes restrita a dados de máximos e mínimos anuais.

Para que a distribuição represente bem o comportamento hidrológico na região estudada, dois itens são necessários: dados com uma qualidade mínima e um método de estimação consistente com a análise. Os dados devem ser representativos do comportamento hidrológico e meteorológico da área de estudo como um todo. Logo, a série histórica a ser utilizada deve conter um intervalo mínimo de dados contínuo, de forma a conter os padrões de variação inter-sazonal e interanual da região. Chow et al. (1988) sugere que o número mínimo de anos de dados seja de 20 anos, enquanto, em estudos hidrológicos brasileiros, o número mínimo de dados para que uma série histórica seja considerada confiável é de 30 dados segundo Eletrobrás (2000). Com relação ao método ou métodos empregados no ajuste, estes devem ter uma eficácia mínima que a curva ajustada seja consistente com o conjunto de dados disponível, para assegurar que a distribuição possa representar bem o comportamento da região estudada.

Embora, geralmente, a consistência dos dados hidrológicos seja estudada e, de certa forma, contornada com outras ferramentas, como a regionalização (Hosking e Wallis, 1997), muitas vezes um método é empregado no ajuste sem nenhum critério ou sem alguma forma de avaliação de sua performance na sumarização do conjunto de dados disponível. Na literatura, há dois métodos já consolidados e que são aplicados na grande maioria das vezes na estimação dos parâmetros de uma distribuição de probabilidade no ajuste de curvas: o Método dos Momentos, devido à sua fácil implementação, e o Método da Máxima Verossimilhança, por ser, assintoticamente (quando o tamanho da amostra tende ao infinito), o melhor estimador possível (Fisher, 1912). Estes métodos, porém, nem sempre rendem resultados satisfatórios. Cousineau (2009), por exemplo, empregou alguns métodos, entre eles o dos Momentos e da Máxima Verossimilhança, no ajuste de distribuições para amostras pequenas usando simulações de Monte Carlo, e concluiu que o Método da Máxima Verossimilhança foi bastante ineficiente e que deveria ser até evitado, estimando com erros até duas vezes maiores do que os demais métodos.

O Método dos Momentos-L foi proposto por Hosking (1990) como uma alternativa a estes métodos já bem estabelecidos na análise de dados hidrológicos, visando uma melhor estimação dos parâmetros nos casos em que os dados exibem características que podem limitar o desempenho destes métodos clássicos: amostras pequenas e presença de *outliers*

(ou valores atípicos). O Método dos Momentos-L foi desenvolvido a partir do trabalho de Greenwood et al. (1979), e tem se tornado popular na hidrologia devido a diversos trabalhos recentes na área, como os de Delicado e Gorla (2008) e Goda et al. (2010), por exemplo, que concluíram que ele é bastante eficiente nesse contexto.

Dada a importância da qualidade dos ajustes das curvas em estudos hidrológicos como o de projetos de usinas hidrelétricas (que envolvem bastante risco), faz-se necessários estudos na área que avaliem o desempenho dos métodos mais frequentemente aplicados nestes casos. Sendo assim, o que este trabalho se propõe a fazer é um estudo comparativo entre estes métodos de ajustes de distribuição a dados hidrológicos. São apresentados ajustes a dados de vazão de três rios com características diferentes. Além disto, foi conduzido um estudo de simulação utilizando diferentes distribuições de probabilidades usualmente adotadas em hidrologia e considerando-se dois tamanhos de amostra e a presença de *outliers*.

2. Objetivos

2.1 Objetivo Geral

Comparar o Método dos Momentos-L, mais recentemente proposto na literatura, com os métodos já estabelecidos da Máxima Verossimilhança e dos Momentos no ajuste de distribuições de probabilidade a dados hidrológicos de vazão máxima anual.

2.2 Objetivos específicos

1. Revisar os fundamentos dos diferentes métodos empregados neste trabalho e sua implementação computacional.
2. Avaliar o desempenho do Método dos Momentos-L em comparação com a dos Métodos dos Momentos e da Máxima Verossimilhança no ajuste de curvas a dados hidrológicos reais, provenientes de séries históricas de estações fluviométricas.
3. Comparar o desempenho dos três métodos na obtenção de parâmetros de distribuições de probabilidade frequentemente utilizadas na hidrologia em simulações com diferentes cenários.

3. Referencial Teórico

3.1 Métodos

Os métodos empregados foram, além do Método dos Momentos-L, que é o foco deste trabalho, dois métodos bem estabelecidos na literatura na estimação de parâmetros de distribuições: o Método dos Momentos e o Método da Máxima Verossimilhança. Eles são populares por características distintas: o Método dos Momentos é, geralmente, facilmente aplicado, e seu estimador, apesar de consistente (isto é, a medida que o tamanho amostra aumenta, o estimador $\hat{\theta}$ converge em probabilidade ao valor real θ), não é tão eficiente. Já o Método da Máxima Verossimilhança é, geralmente, o estimador mais eficiente, porém é muitas vezes muito difícil de ser aplicado.

A seguir é feita uma breve revisão dos três métodos.

3.1.1 Método dos Momentos

O Método dos Momentos foi criado em 1894, quando Pearson “importou” os momentos da física à estatística para estimar os coeficientes de duas distribuições normais ajustadas a dados de razão entre a largura da frente e o comprimento do corpo de 1000 caranguejos (Pearson, 1894). Desde então, tem sido um dos métodos mais utilizados para ajuste de curvas a dados em várias áreas do conhecimento.

Os momentos de uma população são os valores esperados de potências da variável aleatória em questão, e são definidos como:

$$\mu_r = E[X^r] = \int_{-\infty}^{\infty} x^r f(x) dx \quad (3.1)$$

em que μ_r é o r-ésimo momento da população.

Os momentos centrados são momentos em torno da média, e dados por:

$$\mu'_r = E[X - E(X)]^r = \int_{-\infty}^{\infty} (x - \mu)^r f(x) dx \quad (3.2)$$

Muito da forma de distribuições de probabilidade é explicado por medidas que são definidas pelos momentos ou são dependentes deles. O primeiro momento é a média da população. A variância, σ^2 , é definida como o segundo momento centrado. O coeficiente de variação CV, uma medida de dispersão da distribuição em torno da média, é dado por:

$$CV = \frac{\sigma}{\mu} \quad (3.3)$$

A obliquidade γ , uma medida de assimetria, e a curtose κ , uma medida de dispersão da distribuição, são definidas como razões entre momentos:

$$\gamma = \frac{\mu_3}{\mu_2^{3/2}} \quad (3.4)$$

$$\kappa = \frac{\mu_4}{\mu_2^2} \quad (3.5)$$

Os momentos de uma amostra são dados de forma análoga por:

$$m_r = \frac{1}{n} \sum_{i=1}^n (x_i)^r \quad (3.6)$$

em que m_r é o r-ésimo momento da amostra, e n é o tamanho da amostra. Os momentos amostrais centrados, de forma similar:

$$m'_r = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^r \quad (3.7)$$

O Método dos Momentos consiste em se estimar os momentos da população pelos momentos amostrais, isto é, igualar os momentos amostrais dos dados aos momentos da distribuição escolhida para o ajuste, chegando-se, assim, a equações envolvendo os momentos amostrais e os parâmetros da distribuição. Em geral, a ordem dos momentos usada é igual ao número de parâmetros da distribuição escolhida.

Muitas vezes, é possível isolar os parâmetros para obter equações analíticas para os mesmos; quando isto não ocorre, geralmente, se obtém uma equação não-linear que é dependente de apenas um dos parâmetros, que é estimado a partir de métodos numéricos.

O Método dos Momentos é largamente utilizado pela facilidade de implementação, em especial para distribuições com poucos parâmetros, já que em geral a estimação dos parâmetros se dá pela solução de equações analíticas, e o estimador é consistente. Porém, os estimadores são por vezes inviesados (em situações como quando a amostra é pequena, por exemplo), e além disso, os momentos, devido à sua definição, dão mais peso a grandes diferenças (exponenciação dos x_i), e por isso o método tende a não ter bom desempenho em amostras com muitos *outliers* (Hosking e Wallis, 1997).

3.1.2 Método da Máxima Verossimilhança

O Método da Máxima Verossimilhança foi formalmente estruturado e posteriormente popularizado por R. A. Fisher entre 1912 e 1922, apesar de já ter sido aplicado anteriormente sob outros nomes (Hald, 1999; Fisher, 1912, 1921). Este método foi proposto depois de Fisher questionar o Método dos Momentos de Pearson, que o considerava um estimador ineficaz (Fisher, 1937).

O método consiste em, dada uma amostra, se obter os parâmetros da distribuição escolhida para o ajuste que têm a maior probabilidade de ser a distribuição que gerou a amostra. Em termos matemáticos, os parâmetros estimados são os que maximizam a função de verossimilhança, definida como:

$$\mathcal{L}(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i, \theta) \quad (3.8)$$

em que θ é o parâmetro, escalar ou vetor de parâmetros, a ser estimado.

Como a função deve ser maximizada, o logaritmo da função de verossimilhança, chamada de função de log-verossimilhança, é normalmente utilizada, já que o máximo de ambas as funções ocorre no mesmo ponto. A escala logarítmica facilita a manipulação algébrica, pois o logaritmo transforma o produto em um somatório. A função de log-verossimilhança é dada, então, por:

$$\ell(x_1, x_2, \dots, x_n; \theta) = \sum_{i=1}^n \log f(x_i, \theta) \quad (3.9)$$

Para a maximização da função, pode ser empregada análise matemática, calculando-se as derivadas parciais da função no ponto de máximo, isto é, igualando-as a zero e resolvendo o sistema de equações resultante, ou análise computacional, utilizando métodos numéricos de otimização.

Como o Método da Máxima Verossimilhança estima os valores do conjunto de parâmetros da distribuição que tem a maior probabilidade de ter gerado a amostra de dados, ele é, em geral, pouco enviesado (de fato é assintoticamente não enviesado) e mais eficaz na estimação dos parâmetros. Ele apresenta algumas propriedades assintóticas desejadas que o tornam eficiente:

1. Eficiência: quando a amostra tende ao infinito, o erro quadrático médio do estimador do Método da Máxima Verossimilhança tende a ser o menor do que qualquer outro estimador consistente.
2. Invariância: o Método da Máxima Verossimilhança não varia com transformações; então, por exemplo, se $\hat{\theta}$ é o estimador de θ , então $\hat{\theta}^2$ é o estimador de θ^2 .
3. Normalidade assintótica: sob certas condições regulares, quando a amostra tende ao infinito, a distribuição do estimador do método tende a uma distribuição Normal, com média θ .

Contudo, para amostras pequenas, outros estimadores podem ser mais eficientes, e o método nem sempre estima os parâmetros satisfatoriamente (Delicado e Gorla, 2008). Oztekin (2005), por exemplo, analisou dados de vazão máxima de 50 rios na Turquia, estimando os parâmetros da distribuição generalizada de Pareto por cinco métodos diferentes, e concluiu que o método dos momentos, na maioria dos casos, foi o melhor estimador. O método pode ainda apresentar outros problemas, como nos casos em que a função de verossimilhança apresenta vários máximos, nenhum ponto de máximo, ou máximos na fronteira do espaço paramétrico.

3.1.3 Método dos Momentos-L

O Método dos Momentos-L foi proposto por J. R. M. Hosking em 1990 como uma alternativa a estes dois métodos descritos (Hosking, 1990). Ele foi desenvolvido a partir da ideia de “*momentos ponderados por probabilidade*” (tradução livre de “*probability weighted moments*”) de Greenwood et al. (1979). Esses momentos são definidos como:

$$M_{p,r,s} = E\{X^p F(X)^r [1 - F(X)]^s\} \quad (3.10)$$

em que p , r e s são números inteiros, e $F(X)$ é a função distribuição acumulada.

Chamando de $\alpha_r = M_{1,0,r}$, de $\beta_r = M_{1,r,0}$, e fazendo $u = F(x)$, temos que:

$$\alpha_r = \int_0^1 x(u)(1-u)^r du \quad (3.11)$$

$$\beta_r = \int_0^1 x(u)u^r du \quad (3.12)$$

Os momentos-L são, então, definidos como combinações lineares destas integrais de $x(u)$, ponderados pelos “*polinômios modificados de Legendre*” (tradução livre de “*shifted Legendre polynomials*”), que são definidos como:

$$P_r^*(u) = \sum_{k=1}^r p_{r,k}^* u^k \quad (3.13)$$

em que

$$p_{r,k}^* = (-1)^{r-k} \binom{r}{k} \binom{r+k}{k} = \frac{(-1)^{r-k} (r+k)!}{(k!^2)(r-k)!} \quad (3.14)$$

Sendo assim, temos que os quatro primeiros momentos-L populacionais são dados por:

$$\begin{aligned} \lambda_1 &= \alpha_0 = \beta_0 \\ \lambda_2 &= \alpha_0 - 2\alpha_1 = 2\beta_1 - \beta_0 \\ \lambda_3 &= \alpha_0 - 6\alpha_1 + 6\alpha_2 = 2\beta_2 - 6\beta_1 + \beta_0 \\ \lambda_4 &= \alpha_0 - 12\alpha_1 + 30\alpha_2 - 20\alpha_3 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \end{aligned} \quad (3.15)$$

e, em geral:

$$\lambda_{r+1} = (-1)^r \sum_{k=0}^r p_{r,k}^* \alpha_k = \sum_{k=0}^r P_{r,k}^* \beta_k \quad (3.16)$$

em que λ_{r+1} é o $(r+1)$ -ésimo momento-L.

Assim, o r -ésimo momento-L da população é definido em termos de $x(u)$ como:

$$\lambda_r = \int_0^1 x(u) P_{r-1}^*(u) du. \quad (3.17)$$

Os momentos-L amostrais são calculados de forma análoga. Sendo n o tamanho da amostra, e $x_{1:n} \leq x_{2:n} \leq \dots \leq x_{n:n}$ a amostra ordenada de forma crescente, onde $x_{r:n}$ é o r -ésimo menor valor da amostra, os momentos-L amostrais são definidos como:

$$\begin{aligned} l_1 &= b_0 \\ l_2 &= 2b_1 - b_0 \\ l_3 &= 2b_2 - 6b_1 + b_0 \\ l_4 &= 20b_3 - 30b_2 + 12b_1 - b_0 \end{aligned} \quad (3.18)$$

em que (de Landwehr et al., 1979)

$$b_r = \frac{1}{n} \sum_{j=r+1}^n \frac{(j-1)(j-2)\dots(j-r)}{(n-1)(n-2)\dots(n-r)} x_{j:n} \quad (3.19)$$

e, em geral

$$l_{r+1} = \sum_{k=0}^r p_{r,k}^* b_k \quad (3.20)$$

O Método dos Momentos-L funciona de forma similar ao Método do Momentos: tomando o momento amostral como o estimador do momento da população, para se obter os parâmetros da função ajustada aos dados. E, assim como os momentos, os momentos-L definem quantidades que sumarizam a distribuição de probabilidade: λ_1 é a média da distribuição, λ_2 é uma medida de escala, τ é o CV-L (coeficiente de variação-L), τ_3 é a obliquidade-L, e τ_4 é a curtose-L (Hosking, 1989), em que:

$$\tau = \frac{\lambda_2}{\lambda_1} \quad (3.21)$$

e

$$\tau_r = \frac{\lambda_r}{\lambda_2} \quad (3.22)$$

De modo similar aos momentos convencionais, o Método dos Momentos-L muitas vezes se resume à resolução de equações analíticas para se estimar os parâmetros, e por isso é mais simples e mais facilmente implementado do que o Método da Máxima Verossimilhança. Sua desvantagem, novamente, é que os estimadores podem ser viesados, dependendo das circunstâncias, e na maior parte das vezes o estimador é menos eficiente, se comparado com a Máxima Verossimilhança. Contudo, estudos mostram que, quando a amostra de dados é pequena e contém *outliers*, o Método dos Momentos-L pode ser mais eficiente (Hosking, 1990), particularmente devido à exponenciação de $1 - u$, em vez de $x(u)$ no caso do Método dos Momentos, dando menor peso a estes valores extremos.

3.2 Distribuições

As distribuições escolhidas para este trabalho são distribuições comumente aplicadas em hidrologia estatística, e em especial, na previsão de eventos extremos: Gumbel, Pearson tipo III e Weibull de 3 parâmetros. A distribuição de Gumbel é provavelmente a distribuição mais utilizada no ajuste de dados de vazão máxima, sendo, por exemplo, recomendada no Brasil em estudos e projetos de Pequenas Centrais Hidrelétricas, quando a obliquidade é menor do que 1,5 (Eletrobrás, 2000). A distribuição de Pearson tipo III é também bastante popular, possivelmente por ser bastante versátil, já que ela pode ter obliquidade negativa, positiva ou nula; ela é recomendada, por exemplo, pela Interagency Advisory Committee on Water Data (1982) em sua forma logarítmica para máximas anuais. Já a distribuição de Weibull de 3 parâmetros, que também pode ter obliquidade negativa e positiva, é mais usada para ajuste de dados de mínimas, mas também é frequentemente utilizada na previsão de cheias, entre outras aplicação na hidrologia.

3.2.1 Distribuição de Gumbel

A distribuição de Gumbel é uma distribuição da família da “*Distribuição de Valores Extremos Generalizada*” (tradução livre de “*generalized extreme value distribution*”, abreviada por GEV), sendo chamada de Valor Extremo tipo I. Ela tem este nome em

homenagem a Emil Julius Gumbel, que a decreteu entre 1935 e 1941 (Gumbel, 1935, 1941). É muito utilizada, juntamente com as outras distribuições da família GEV, no campo de Teoria de Valores Extremos, principalmente na hidrologia, para modelar máximas de vazão e de precipitação e secas.

A distribuição de Gumbel é uma distribuição de dois parâmetros, um parâmetro de localidade ξ , que pode variar de $-\infty$ a $+\infty$, e um parâmetro de escala β , que é sempre positivo; assim, a distribuição também pode assumir valores entre $-\infty$ e $+\infty$. Sua função densidade de probabilidade é dada por:

$$f(x) = \frac{1}{\beta} e^{-(z+e^{-z})} \quad (3.23)$$

em que $z = \frac{x-\xi}{\beta}$, e sua função distribuição cumulativa, por:

$$F(x) = e^{-e^{-z}} \quad (3.24)$$

Método dos Momentos

Para a distribuição de Gumbel, o primeiro momento é dado por (Mahdi e Cenac, 2005):

$$\mu_1 = \beta\phi + \xi \quad (3.25)$$

em que $\phi \simeq 0,577216$ é a constante de Euler-Mascheroni. O segundo momento centrado é dado por por:

$$\mu_2' = \frac{(\beta\pi)^2}{6} \quad (3.26)$$

Como se pode ver, a equação 3.26 depende apenas do parâmetro β . Segue-se, então, que:

$$\hat{\beta} = \frac{s\sqrt{6}}{\pi} \quad (3.27)$$

em que s é o desvio-padrão da amostra. A partir de $\hat{\beta}$, ξ é facilmente estimado por:

$$\hat{\xi} = \bar{x} - \phi\hat{\beta} \quad (3.28)$$

em que \bar{x} é a média da amostra.

Método da Máxima Verossimilhança

A função de log-verossimilhança da distribuição de Gumbel é:

$$\ell(\beta, \xi) = -\sum_{i=1}^n \frac{x_i - \xi}{\beta} - n \ln \beta - \sum_{i=1}^n \exp\left(-\frac{\xi - x_i}{\beta}\right) \quad (3.29)$$

As derivadas parciais em relação a β e a ξ são dadas por:

$$\frac{\partial \ell(\beta, \xi)}{\partial \xi} = \frac{1}{\beta} \left[n - \sum_{i=1}^n \exp\left(-\frac{x_i - \xi}{\beta}\right) \right] \quad (3.30)$$

e

$$\frac{\partial \ell(\beta, \xi)}{\partial \beta} = \sum_{i=1}^n \left(\frac{x_i - \xi}{\beta^2} \right) - \frac{n}{\beta} - \sum_{i=1}^n \left[\left(\frac{x_i - \xi}{\beta^2} \right) \exp\left(-\frac{x_i - \xi}{\beta}\right) \right] \quad (3.31)$$

Igualando as derivadas parciais a zero e resolvendo o sistema $\frac{\partial \ln \ell(\beta, \xi)}{\partial \beta} = \frac{\partial \ln \ell(\beta, \xi)}{\partial \xi} = 0$, de Mahdi e Cenac (2005), temos que:

$$\xi = \beta \left[\ln n - \ln \sum_{i=1}^n \exp\left(-\frac{x_i}{\beta}\right) \right] \quad (3.32)$$

e

$$\beta = \bar{x} - \frac{\sum_{i=1}^n x_i \exp\left(-\frac{x_i}{\beta}\right)}{\sum_{i=1}^n \exp\left(-\frac{x_i}{\beta}\right)} \quad (3.33)$$

Assim, β é estimado pela equação 3.33. Neste trabalho, optou-se por fazer a estimação através do Método de Newton-Raphson, utilizando um *loop* em que a condição a ser satisfeita é $\delta = 0,0001$, e dada uma estimativa inicial de β , aqui chamada de β_0 , calcula-se β_1 pela equação 3.33, em seguida a diferença $\delta = |\beta_1 - \beta_0|$, e então é analisado se esta diferença satisfaz a condição; se sim, $\hat{\beta} = \beta_1$; se não, o *loop* é iniciado novamente, com $\beta_0 = \beta_1$. Então, partir de $\hat{\beta}$, $\hat{\xi}$ é calculado diretamente pela equação 3.32.

Quando ocorreu do Método de Newton-Raphson não convergir para um valor de β após 100 iterações, os dois parâmetros da distribuição foram estimados diretamente pela otimização da equação 3.29, utilizando a função *optim* do pacote *stats4* do software *R*, com o Método de Nelder-Mead (R Core Team, 2015).

Método dos Momentos-L

O primeiro e o segundo momentos-L da distribuição de Gumbel são dados por (de Hosking 1990):

$$\begin{aligned}\lambda_1 &= \beta\phi + \xi \\ \lambda_2 &= \beta \log 2\end{aligned}\tag{3.34}$$

onde ϕ é a constante de Euler-Mascheroni.

Rearranjando as equações, assim como no Método dos Momentos, é possível se obter uma equação analítica para se estimar β :

$$\hat{\beta} = \frac{l_2}{\log 2}\tag{3.35}$$

E a partir de $\hat{\beta}$, $\hat{\xi}$ é calculado por:

$$\hat{\xi} = l_1 - \phi\hat{\beta}.\tag{3.36}$$

3.2.2 Distribuição de Pearson tipo III

A distribuição de Pearson tipo III, também chamada de distribuição Gama de 3 parâmetros, já que ela é a distribuição Gama, com parâmetros de forma α e de escala β , acrescida de um parâmetro de localidade, ξ . O parâmetro α é sempre positivo, enquanto que β e ξ podem assumir qualquer valor entre $-\infty$ e $+\infty$. Ela é, como mencionado, uma distribuição da família das distribuições de Pearson, que têm este nome por terem sido descritas por Karl Pearson em seu artigo de 1895. Seu uso é bem difundido no campo da hidrologia estatística, tanto na forma convencional quanto na forma logarítmica (distribuição de Log-Pearson tipo III). Sua função densidade de probabilidade é dada por:

$$f(x) = \frac{1}{|\beta|^\alpha \Gamma(\alpha)} |x - \xi|^{\alpha-1} e^{-\left(\frac{x-\xi}{\beta}\right)}\tag{3.37}$$

em que Γ é a função Gama.

A função assume valores que dependem do valor de β . Se $\beta > 0$, $\xi \leq x < \infty$, a

função distribuição cumulativa é dada por:

$$F(x) = \frac{G(\alpha, \frac{x-\xi}{\beta})}{\Gamma(\alpha)} \quad (3.38)$$

em que G é a função gama incompleta inferior (Hosking e Wallis, 1997). Se $\beta < 0$, $-\infty < x \leq \xi$, a função distribuição cumulativa é dada por:

$$F(x) = 1 - \frac{G(\alpha, \frac{x-\xi}{\beta})}{\Gamma(\alpha)} \quad (3.39)$$

Se $\beta = 0$, a distribuição de Pearson tipo III torna-se uma distribuição Normal, em que $-\infty < x < +\infty$, e função densidade de probabilidade:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \frac{(x - \mu)^2}{2\sigma^2} \quad (3.40)$$

em que μ é a média e σ^2 é a variância.

Método dos Momentos

O primeiro momento da distribuição de Pearson é dada por:

$$\mu = \xi + \alpha\beta \quad (3.41)$$

o segundo momento centrado, por:

$$\sigma^2 = \alpha\beta^2 \quad (3.42)$$

e a obliquidade, por:

$$\gamma = \frac{2\beta}{|\beta|\sqrt{\alpha}} \quad (3.43)$$

O valor da obliquidade depende apenas de α , e o sinal, apenas de β . Assim, se a obliquidade da amostra for positiva, $\hat{\beta}$ é positivo; e se a obliquidade da amostra for negativa, $\hat{\beta}$ é negativo. De qualquer forma, α é estimado por:

$$\hat{\alpha} = \left(\frac{2}{g_2}\right)^2 \quad (3.44)$$

em que g_2 é a obliquidade sem viés da amostra, que neste trabalho foi calculada como

recomendado por Bobée e Robitaille (1977):

$$g_2 = \left(1 + \frac{0,85}{n}\right)g_1 \quad (3.45)$$

em que g é a obliquidade da amostra usual, dada por:

$$g_1 = \frac{n^2}{(n-1)(n-2)} \frac{m_3}{s^3} \quad (3.46)$$

em que n é o tamanho da amostra, m_3 é o terceiro momento amostral, e s é o desvio-padrão da amostra. A partir de $\hat{\alpha}$, os outros parâmetros são calculados então por:

$$\hat{\beta} = \left(\frac{s}{\hat{\alpha}}\right)^{\frac{1}{2}} \quad (3.47)$$

e

$$\hat{\xi} = \bar{x} - \hat{\alpha}\hat{\beta}. \quad (3.48)$$

Método da Máxima Verossimilhança

A função de log-verossimilhança da distribuição de Pearson tipo III é dada por:

$$\ell(\alpha, \beta, \xi) = -n\alpha \log \beta - n \log \Gamma(\alpha) + (\alpha + 1) \sum_{i=1}^n \log(x_i - \xi) - \sum_{i=1}^n \left(\frac{x_i - \xi}{\beta}\right) \quad (3.49)$$

As derivadas parciais em relação a α , β e ξ são dadas por:

$$\frac{\partial \ell(\alpha, \beta, \xi)}{\partial \alpha} = -n \log \beta - n\Psi(\alpha) + \sum_{i=1}^n \log(x_i - \xi) \quad (3.50)$$

$$\frac{\partial \ell(\alpha, \beta, \xi)}{\partial \beta} = -\frac{n\alpha}{\beta} + \sum_{i=1}^n \left(\frac{x_i - \xi}{\beta^2}\right) \quad (3.51)$$

$$\frac{\partial \ell(\alpha, \beta, \xi)}{\partial \xi} = -(\alpha - 1) \sum_{i=1}^n \left(\frac{1}{x_i - \xi}\right) + \frac{n}{\beta} \quad (3.52)$$

em que Ψ é a função digama (Bai et al., 1990). Rearranjando as equações 3.51 e 3.52, de

Bai et al. (1990), se obtém uma equação analítica para β em função apenas de ξ :

$$\hat{\beta} = \sum_{i=1}^n \left(\frac{x_i - \hat{\xi}}{n} \right) - \frac{n}{\sum_{i=1}^n (x_i - \hat{\xi})^{-1}} \quad (3.53)$$

e uma equação analítica para α em função de β e ξ :

$$\hat{\alpha} = \sum_{i=1}^n \left(\frac{x_i - \hat{\xi}}{n\hat{\beta}} \right) \quad (3.54)$$

Assim, dada uma estimativa de ξ , é possível se calcular β e α . Neste trabalho, esta metodologia foi empregada com a função de log-verossimilhança: foram usados 1000 valores ξ , a partir do menor x_i , ou seja, do menor valor da amostra, e decrescendo de 1 a cada cálculo, foram calculados os estimadores e o valor da função de log-verossimilhança; e então, os valores que maximizam a função de log-verossimilhança foram escolhidos como os estimadores dos parâmetros da distribuição.

Método dos Momentos-L

O primeiro momento-L da distribuição de Pearson tipo III, como o primeiro momento convencional, é dado por:

$$\lambda_1 = \xi + \alpha\beta \quad (3.55)$$

O segundo momento-L é dado por:

$$\lambda_2 = \frac{\beta\Gamma(\alpha + \frac{1}{2})}{\sqrt{\pi}\Gamma(\alpha)} \quad (3.56)$$

e a obliquidade-L, por:

$$\tau_3 = 6I_{\frac{1}{3}}(\alpha, 2\alpha) - 3 \quad (3.57)$$

em que $I_x(a, b)$ é a função razão beta incompleta (Hosking e Wallis, 1997). Como a equação 3.57 depende apenas de α , $\hat{\alpha}$ pode ser calculado a partir de t_3 , que é a obliquidade-L da amostra; esta equação é, contudo, difícil de ser manipulada, então usa-se uma aproximação

de τ_3 , conforme Hosking e Wallis (1997): se $0 < |t_3| < \frac{1}{3}$, então $z = 3\pi t_3^2$, e:

$$\hat{\alpha} \approx \frac{1 + 0,2906z}{z + 0,1882z^2 + 0,0442z^3} \quad (3.58)$$

e se $\frac{1}{3} \leq |t_3| < 1$, então $z = 1 - |t_3|$, e:

$$\hat{\alpha} \approx \frac{0,36067z - 0,59567z^2 + 0,25361z^3}{1 - 2,78861z + 2,56096z^2 - 0,77045z^3} \quad (3.59)$$

A partir de $\hat{\alpha}$, $\hat{\beta}$ é calculado pela equação 3.56, e então $\hat{\xi}$ é calculado pela equação 3.55.

3.2.3 Distribuição de Weibull

A distribuição de Weibull leva este nome por ter sido derivada por Weibull (1951) como uma generalização da distribuição exponencial, e é bastante utilizada em estudos de falhas de um certo mecanismo, por exemplo. Ela é uma distribuição de 3 parâmetros: ξ , que é o parâmetro de localidade, que pode assumir qualquer valor, α , que é o parâmetro de escala, e β , que é o parâmetro de forma; estes dois últimos são sempre positivos. Para qualquer valor de ξ , $\xi < x < \infty$, e quando ξ é zero, a distribuição é a de Weibull convencional (Weibull de 2 parâmetros). Sua função densidade de probabilidade é dada por:

$$f(x) = \frac{\beta}{\alpha} \left(\frac{x - \xi}{\alpha}\right)^{\beta-1} e^{-\left(\frac{x-\xi}{\alpha}\right)^\beta} \quad (3.60)$$

Sua função distribuição cumulativa é dada por:

$$F(x) = 1 - \exp\left[-\left(\frac{x - \xi}{\alpha}\right)^\beta\right] \quad (3.61)$$

Método dos Momentos

A esperança de distribuição de Weibull é dada por:

$$\mu_1 = \xi + \alpha\Gamma\left(1 + \frac{1}{\beta}\right) \quad (3.62)$$

O segundo momento centrado é dado por:

$$\mu'_2 = \alpha^2 \left[\Gamma\left(1 + \frac{2}{\beta}\right) - \Gamma^2\left(1 + \frac{1}{\beta}\right) \right] \quad (3.63)$$

E a obliquidade, por:

$$\gamma = \frac{2[\Gamma(1 + \frac{1}{\beta})]^3 - 3\Gamma(1 + \frac{1}{\beta})\Gamma(1 + \frac{2}{\beta}) + \Gamma(1 + \frac{3}{\beta})}{\{\Gamma(1 + \frac{2}{\beta}) - [\Gamma(1 + \frac{1}{\beta})]^2\}^{\frac{3}{2}}} \quad (3.64)$$

Como a obliquidade só depende de um parâmetro, β , este pode ser estimado a partir da obliquidade amostral, por um método de otimização. A partir de $\hat{\beta}$, $\hat{\alpha}$ é calculado pela equação 3.63:

$$\hat{\alpha} = \frac{s}{[\Gamma(1 + \frac{2}{\hat{\beta}}) - \Gamma^2(1 + \frac{1}{\hat{\beta}})]^{1/2}} \quad (3.65)$$

e em seguida, $\hat{\xi}$ é calculado pela equação 3.62:

$$\hat{\xi} = \bar{x} - \hat{\alpha}\Gamma(1 + \frac{1}{\hat{\beta}}). \quad (3.66)$$

Método da Máxima Verossimilhança

A função de log-verossimilhança da distribuição de Weibull é dada por:

$$\ell(\alpha, \beta, \xi) = n[\log(\beta) - \beta \log(\alpha)] + (\beta - 1) \sum_{i=1}^n [\log(x_i - \xi)] - \frac{1}{\alpha^\beta} \sum_{i=1}^n (x_i - \xi)^\beta \quad (3.67)$$

As derivadas parciais da função de log-verossimilhança são:

$$\frac{\partial \ell(\alpha, \beta, \xi)}{\partial \alpha} = \frac{n\beta}{\alpha} - \frac{\beta}{\alpha^{\beta+1}} \sum_{i=1}^n (x_i - \xi)^\beta \quad (3.68)$$

$$\frac{\partial \ell(\alpha, \beta, \xi)}{\partial \beta} = -\frac{n}{\beta} - \sum_{i=1}^n \log\left(\frac{x_i - \xi}{\alpha}\right) + \sum_{i=1}^n \log\left(\frac{x_i - \xi}{\alpha}\right)^\beta \log\left(\frac{x_i - \xi}{\alpha}\right) \quad (3.69)$$

$$\frac{\partial \ell(\alpha, \beta, \xi)}{\partial \xi} = \frac{\beta}{\alpha^\beta} \left[\sum_{i=1}^n (x_i - \xi)^{\beta-1} \right] - (\beta - 1) \sum_{i=1}^n \left(\frac{1}{x_i - \xi} \right) \quad (3.70)$$

Igualando cada equação a zero, o sistema resultante não apresenta nenhuma solução fechada para os parâmetros (Rockette, Antle and Klimko, 1974 *apud* Akram e Hayat 2014).

O que foi feito neste trabalho, então, foi otimizar diretamente a equação 3.67 para os três parâmetros da distribuição, utilizando a função *optim* do pacote *stats4* do software *R*, pelo Método de Nelder-Mead (R Core Team, 2015).

Método dos Momentos-L

O primeiro e o segundo momentos-L da distribuição de Weibull são dados por (de Goda et al. 2010):

$$\lambda_1 = \xi + \alpha\Gamma\left(1 + \frac{1}{\beta}\right) \quad (3.71)$$

$$\lambda_2 = \alpha\Gamma\left(1 + \frac{1}{\beta}\right)\left(1 - \frac{1}{2^{1/\beta}}\right) \quad (3.72)$$

e a obliquidade-L, por:

$$\tau_3 = \frac{1 - \frac{3}{2^{1/\beta}} + \frac{2}{3^{1/\beta}}}{1 - \frac{1}{2^{1/\beta}}} \quad (3.73)$$

Novamente, como no Método dos Momentos, a obliquidade-L depende apenas do parâmetro β , que é obtido por otimização pela equação 3.73, utilizando a função *uniroot.all*, do pacote *rootSolve* (Soetaert, 2009). Então, a partir de $\hat{\beta}$, $\hat{\alpha}$ é calculado pela equação 3.72:

$$\hat{\alpha} = \frac{l_2}{\Gamma\left(1 + \frac{1}{\hat{\beta}}\right)\left(1 - \frac{1}{2^{1/\hat{\beta}}}\right)} \quad (3.74)$$

e $\hat{\xi}$, pela equação 3.71:

$$\hat{\xi} = l_1 - \hat{\alpha}\Gamma\left(1 + \frac{1}{\hat{\beta}}\right) \quad (3.75)$$

Nos casos em que a otimização não obteve nenhuma valor, β foi calculado através da aproximação proposta por Goda et al. (2010):

$$\hat{\beta} = 285,3t_3^6 - 658,6t_3^5 + 622,8t_3^4 - 317,2t_3^3 + 98,52t_3^2 - 21,256t_3 + 3,516 \quad (3.76)$$

em que t_3 é a obliquidade-L da amostra. A partir de $\hat{\beta}$, os demais parâmetros são calculados do mesmo modo pelas equações 3.72 e 3.71.

4. Metodologia

A simulação de dados aleatórios das distribuições de probabilidade, a implementação dos métodos, o cálculo dos quantis e a elaboração dos gráficos foram realizados no ambiente do software *R* 3.2.2 (R Core Team, 2015). O pacote *FAdist* (Aucoin, 2015) foi utilizado para gerar a distribuição teórica, os quantis e valores aleatórios das distribuições de Gumbel e de Weibull de 3 parâmetros, e o pacote *PearsonDS* (Becker e Klößner, 2013) para a distribuição de Pearson tipo III, distribuições estas que não estão contidas no *Rbase*. A implementação do Método dos Momentos foi feita com o auxílio dos pacotes *moments* (Komsta e Novomestky, 2015) e *rootSolve* (Soetaert, 2009); a implementação do Método da Máxima Verossimilhança foi feita com o auxílio do pacote *MASS* (Venables e Ripley, 2002); e a implementação do Método dos Momentos-L foi feita diretamente pelo pacote *lmom* (Hosking, 2015).

Inicialmente, os métodos foram aplicados para ajustar curvas a dados reais, utilizando as vazões máximas anuais extraídas das séries históricas de estações fluviométricas disponíveis na plataforma *Hidroweb* (Agência Nacional de Águas – ANA, 2005), para três rios: rio dos Patos, rio São Salvador e rio Paraguai. Na obtenção destes dados, foram descartados os meses com mais de 7 dias de dados faltantes, e os anos com mais de 3 meses de dados descartados ou faltantes, seguidos ou não. Estes rios foram escolhidos como uma forma de representar três cenários reais com que um estudo hidrológicos poderia se deparar, e de forma que as três distribuições escolhidas neste trabalho fosse empregadas: um rio com vazão pequena, com muitos anos de dados, e com poucos valores atípicos, que é o caso do rio dos Patos; um rio com vazão média, com poucos anos de dados, sendo um extremo à direita, cuja densidade apresenta, então, uma cauda pesada à direita, que é o caso do rio São Salvador; e um rio com vazão grande e também poucos anos de dados, e que apresenta cauda tanto à direita como à esquerda, que é o caso do rio Paraguai.

O ajuste das curvas foi feito pelos três métodos, utilizando, como mencionado, uma distribuição para cada rio. A escolha da distribuição para cada caso foi um pouco arbitrária, e baseada parcialmente na obliquidade dos dados. Para cada ajuste, além da análise visual de aderência das distribuições teóricas dadas pelos parâmetros estimados por cada método, foi feito um comparativo entre as vazões, dadas pelos quantis de cada distribuição, para os tempos de retorno T de 50, 100 e 1000 anos (chamadas de Q50, Q100 e Q1000). Estes tempos de retorno foram escolhidos por serem comumente utilizados como vazão de projeto em estudos de viabilização e dimensionamento de obras hidráulicas: por exemplo, 50 anos, para pontes ferroviárias; 100 anos, para portos fluviais; e 1000 anos, para grandes estruturas hidráulicas, como usinas hidrelétricas (Serrano, 1997).

Num segundo momento, foi feito um estudo de simulação, para se analisar o desempenho dos métodos em diferentes cenários. Nestas simulações, foi suposto que os dados reais de vazão máxima anual seguem uma dada distribuição de probabilidade, e então, em cada caso, foram gerados 5000 dados aleatórios, representando 5000 anos de vazões máximas anuais, e calculadas as vazões para os tempo de retorno de 50, 100 e 1000 anos, que foram as vazões "*corretas*" da população em cada caso. Os parâmetros das distribuições foram escolhidos de forma que elas ficassem semelhantes aos dados de vazão máxima anual do rio Paraguai. Em seguida, foram retiradas amostras aleatórias destes dados, e feitas as estimativas dos parâmetros das distribuições pelos três métodos a partir destas amostras, e calculadas as vazões para os tempos de retorno definidos para comparar os resultados.

Foram realizadas simulações com dois tamanhos de amostra, 30 e 50 dados (consideradas aqui como de tamanho pequeno e grande, respectivamente), e com e sem contaminação à direita, para comparar o comportamento dos métodos quando as amostras são pequenas, e quando há dados extremos no conjunto. Para a contaminação das amostras, foram adicionados 5 dados extremos, extraídos aleatoriamente do grupo dos 500 maiores valores do conjunto original de 5000 dados, mas mantendo os tamanhos de amostra em 30 e 50 dados. Este processo foi repetido 1000 vezes em cada simulação, de forma a gerar 9 conjuntos de dados em cada caso: Q50, Q100 e Q1000 estimadas para cada método. A densidade destas vazões estimadas foram então representadas em gráficos para comparação, juntamente com as vazões respectivas da população de onde as amostras foram extraídas, e calculados o viés, a variância e o erro quadrático médio (EQM), em relação ao valor da

população. O viés é dado por:

$$\text{viés}(\hat{\theta}) = \frac{1}{N} \sum_{i=1}^N (\theta - \hat{\theta}) \quad (4.1)$$

em que θ é o valor "*correto*" da vazão, $\hat{\theta}$ são os valores estimados, e N é o número de estimações. O erro quadrático médio é calculado, então, por:

$$EQM(\hat{\theta}) = \text{viés}(\hat{\theta})^2 + \text{Var}(\hat{\theta}) \quad (4.2)$$

em que Var é a variância do conjunto de valores estimados por cada método.

5. Aplicação a dados reais

5.1 Rio dos Patos

O rio dos Patos é um pequeno rio situado na região centro-sul do Estado do Paraná. Os dados de vazão máxima anual deste rio foram extraídos da série histórica de 68 anos de dados de vazão, de 1938 a 2005, da estação fluviométrica Rio dos Patos (código ANA 64620000), situada no município de Prudentópolis (Agência Nacional de Águas – ANA, 2005). Outros estudos verificaram uma possível regularização da vazão deste rio a partir do ano 2000, porém estes dados foram incluídos neste estudo, e esta regularização não foi investigada, dado que não é intuito deste trabalho fazer um estudo hidrológico definitivo da região. Assim, as vazões máximas anuais observadas para este rio podem ser vistas na Tabela 5.1.

Para este conjunto de dados, optou-se por ajustar uma distribuição de Gumbel. A densidade destes dados de vazão máxima, juntamente com as distribuições teóricas com os parâmetros estimados pelos três métodos, podem ser vistas na Figura 5.1.

Como esperado, devido ao grande volume de dados, os três métodos ajustaram distribuições parecidas e com boa aderência aos dados. A distribuição ajustada pelo Método da Máxima Verossimilhança aderiu melhor aos dados na região do entorno da média, mas superestima levemente os dados à esquerda e subestima os dados da cauda à direita, que contém os dados extremos deste conjunto. O Método dos Momentos ajustou uma distribuição que aderiu bem aos dados à esquerda, mas não tanto na região do entorno da média, e na região da cauda os valores são um pouco maiores do que os estimados pela máxima verossimilhança. Já a distribuição ajustada pelo Método dos Momentos-L exibiu comportamento de meio-termo entre esses dois. O comportamento das distribuições ajustadas pelos três métodos na região da cauda pode ser visto na Figura 5.2. Observa-se

Tabela 5.1: Vazões máximas anuais do rio dos Patos, para o período de 1938 a 2005

Ano	Q (m ³ /s)	Ano	Q (m ³ /s)	Ano	Q (m ³ /s)
1938	233,36	1961	224,54	1984	187,03
1939	221,00	1962	133,62	1985	240,02
1940	56,75	1963	219,24	1986	147,13
1941	163,96	1964	192,74	1987	318,99
1942	162,30	1965	190,98	1988	248,33
1943	135,08	1966	168,95	1989	223,39
1944	144,43	1967	107,44	1990	266,62
1945	251,86	1968	141,31	1991	129,88
1946	189,21	1969	134,76	1992	595,37
1947	215,71	1970	98,87	1993	318,99
1948	201,57	1971	185,16	1994	125,20
1949	69,68	1972	194,92	1995	436,61
1950	160,64	1973	246,25	1996	192,95
1951	229,84	1974	117,41	1997	242,10
1952	178,92	1975	170,20	1998	416,55
1953	132,17	1976	172,07	1999	445,33
1954	240,64	1977	70,82	2000	278,36
1955	259,35	1978	147,13	2001	134,24
1956	152,32	1979	221,32	2002	116,79
1957	306,52	1980	179,55	2003	111,70
1958	149,10	1981	133,00	2004	105,15
1959	127,80	1982	266,62	2005	134,24
1960	167,29	1983	484,19		

que as vazões para os tempos de retorno estimadas, relacionadas na Tabela 5.2, foram bastante semelhantes; a maior diferença relativa entre elas é de apenas 9%.

Tabela 5.2: Vazões para os tempos de retorno de 50, 100 e 1000 anos estimadas pelos Métodos dos Momentos (MOM), da Máxima Verossimilhança (MMV) e dos Momentos-L (MLM) para o rio dos Patos

Método	Vazão (m ³ /s)		
	Q50	Q100	Q1000
MOM	458	511,67	689,01
MMV	422,97	469,74	624,31
MLM	443,44	494,05	661,29

Obviamente, os dados extremos na cauda da distribuição não podem ser totalmente suavizados pela distribuição, por serem consideravelmente maiores do que os demais; isto fica mais evidente ao se comparar os valores estimados de vazão para o tempo de retorno do 50 anos com o valor real, que é de 533,87 m³/s (por interpolação), ou seja, a vazão foi bastante subestimada. Ainda assim, pode-se dizer que a distribuição de Gumbel se ajustou

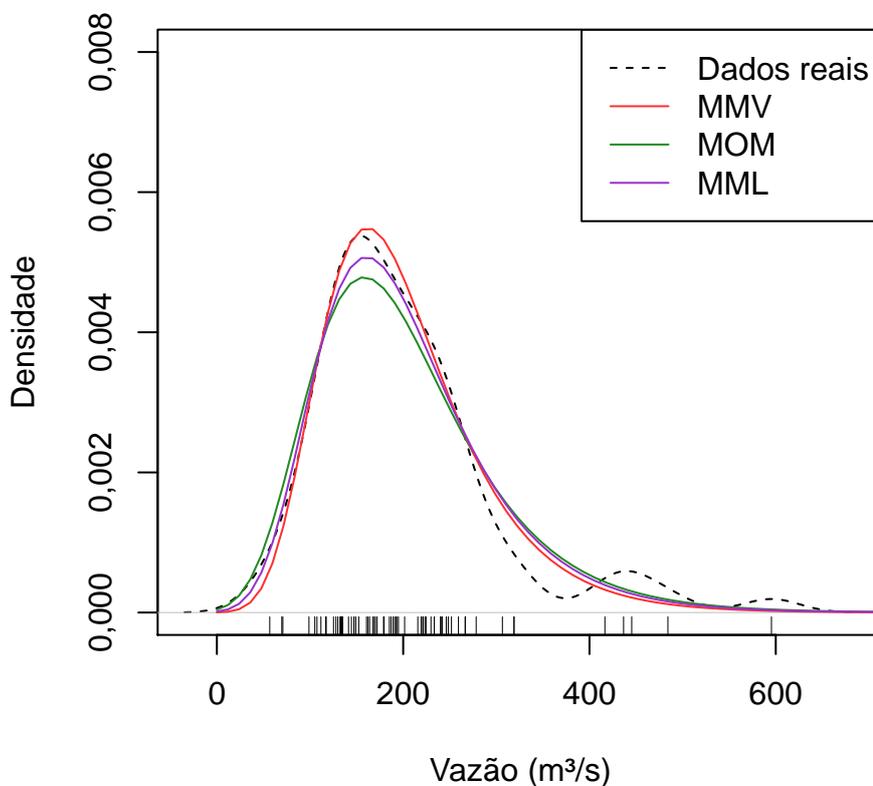


Figura 5.1: Densidade dos dados de vazão máxima anual do Rio dos Patos, para o período de 1938 a 2005, e as distribuições teóricas de Gumbel para os parâmetros estimados pelo Método dos Momentos (MOM), pelo Método da Máxima Verossimilhança (MMV) e pelo Método dos Momentos-L (MML).

bem aos dados. Além disso, os três métodos ajustaram as curvas satisfatoriamente, e é possível se afirmar que qualquer um deles poderia ser usado para o ajuste neste caso. Em uma estimativa mais conservadora, o Método dos Momentos seria indicado, ainda que as diferenças relativas entre as estimações tenham sido pequenas, como dito anteriormente.

5.2 Rio São Salvador

O rio São Salvador, também chamado de rio Andrada, é um rio de vazão média situado na região oeste do Estado do Paraná. Os dados de vazão máxima anual deste rio foram extraídos da série histórica de 32 anos de dados de vazão, do período de 1977 a 2008, da estação fluviométrica São Sebastião (código ANA 65979000), localizada no município

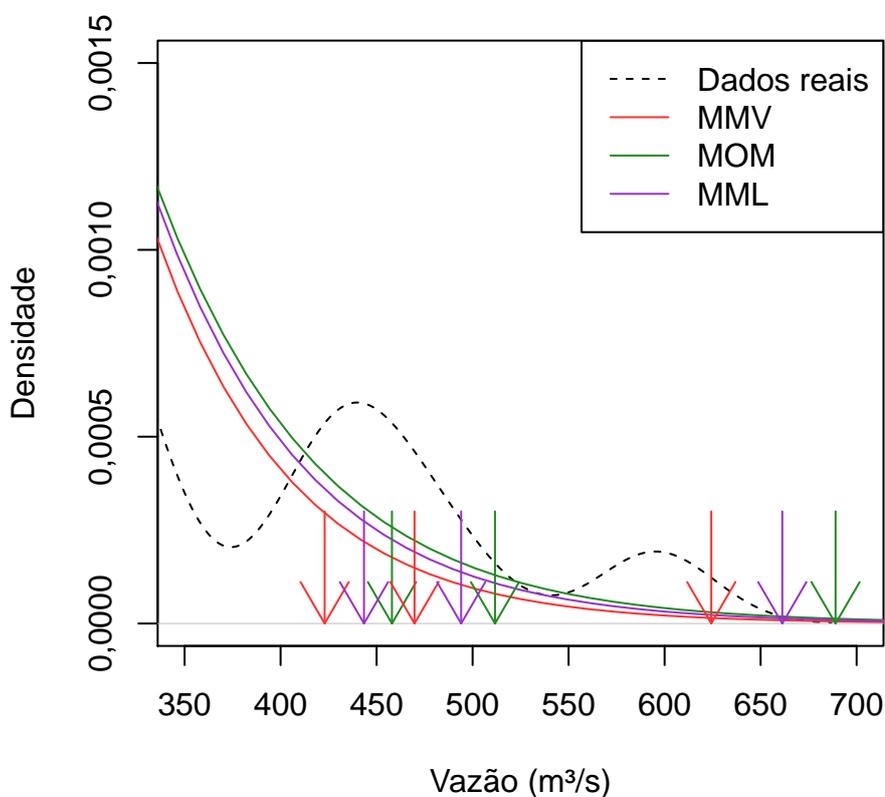


Figura 5.2: Densidade dos dados de vazão máxima anual do Rio dos Patos e as distribuições teóricas de Gumbel para os parâmetros estimados pelo Método dos Momentos (MOM), pelo Método da Máxima Verossimilhança (MMV) e pelo Método dos Momentos-L (MML), na região da cauda à direita. As setas indicam as vazões para os tempos de retorno de 50, 100 e 1000 anos.

de Capitão Leônidas Marques (Agência Nacional de Águas – ANA, 2005). Estes dados de vazão máxima anual podem ser vistos na Tabela 5.3.

A distribuição de Pearson tipo III foi a escolhida para o ajuste. Este conjunto de dados apresenta um valor, no ano de 1981, bastante destoante dos demais valores de máxima. Este valor não foi decorrido de erro, mas sim de um evento extremo, sendo o mesmo verificado em estações e bacias vizinhas à deste rio. Por isso, para estes de dados, o ajuste foi considerando e desconsiderando este *outlier*, para se avaliar seu efeito no desempenho dos métodos.

A densidade dos dados de vazão máxima anual, com e sem o *outlier*, juntamente com as distribuições teóricas com os parâmetros estimados pelos três métodos, podem ser

Tabela 5.3: Vazões máximas anuais do rio São Salvador, para o período de 1977 a 2008

Ano	Q (m ³ /s)	Ano	Q (m ³ /s)	Ano	Q (m ³ /s)
1977	188	1988	658	1999	798
1978	119	1989	298	2000	413
1979	500	1990	428	2001	216
1980	205	1991	360	2002	590
1981	1968	1992	411	2003	609
1982	466	1993	628	2004	485
1983	524	1994	492	2005	603
1984	438	1995	330	2006	112
1985	185	1996	561	2007	630
1986	278	1997	572	2008	383
1987	519	1998	596		

vistas nas Figuras 5.3 e 5.4.

Está claro que esse dado causou uma diferença muito grande nas estimativas, que gerou uma cauda bastante pesada nos três métodos, e que o Método dos Momentos foi especialmente sensível. As estimações feitas pelos métodos da Máxima Verossimilhança e dos Momentos-L foram parecidas, mas não aderiram muito bem aos dados, principalmente devido à obliquidade: no conjunto de dados com o *outlier* ela é igual a 3,07, e sem o *outlier* é igual a -0,25; por isso, as estimações consideram a obliquidade positiva, mesmo que na região de concentração dos dados ela seja negativa. Já na estimação feita sem o dado extremo, o Método dos Momentos-L adere melhor aos dados na região da média, mas subestima os valores na cauda à direita. A estimação do Métodos da Máxima Verossimilhança aderiu melhor aos dados na região da cauda à direita, mas nem tanto no entorno da média; já o Método dos Momentos estimou parâmetros bastante semelhantes ao Método dos Momentos-L. Isto fica mais claro com as vazões Q50, Q100 e Q1000 estimadas nos dois casos, vistas na Tabela 5.4. As vazões estimadas para o tempo de retorno de 1000 anos pelos três métodos foi semelhante, no caso sem o dado extremo, e mostra como a vazão máxima do rio São Salvador no ano de 1981 foi bastante anormal: ela foi quase o dobro da vazão que se espera que ocorra apenas uma vez a cada 1000 anos. Mostra, também, como os *outliers* influenciam na estimação dos métodos, em particular o Método dos Momentos, o que já foi discutido na literatura (Hosking e Wallis, 1997).

A distribuição de Pearson tipo III se mostrou uma boa escolha para estes dados. Devido à sua característica de pode ter obliquidade tanto negativa quanto positiva, mesmo

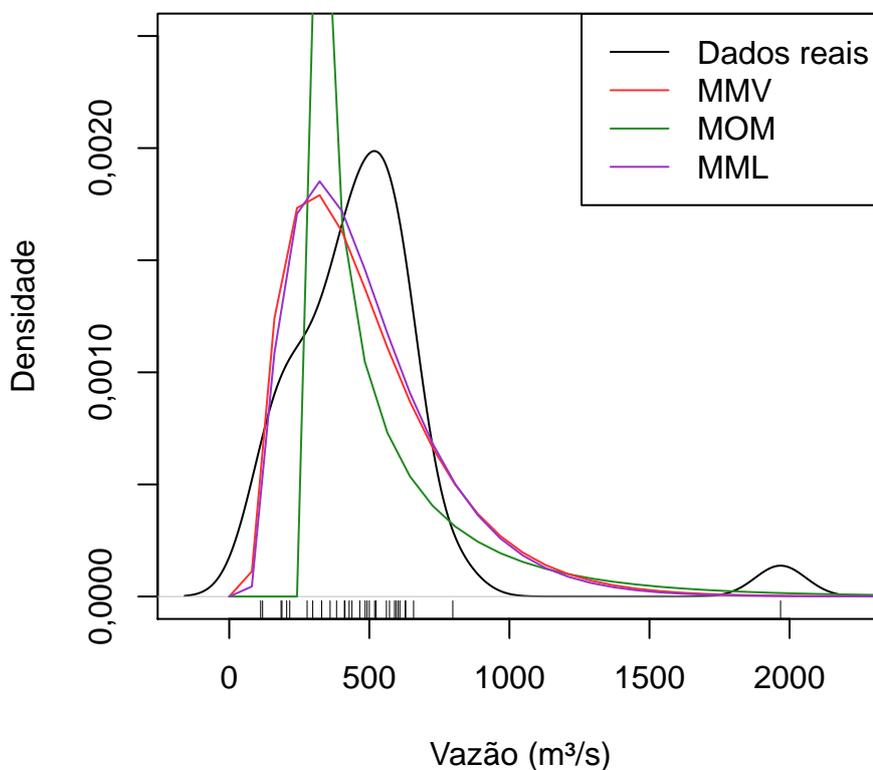


Figura 5.3: Densidade dos dados de vazão máxima anual do Rio São Salvador, para o período de 1977 a 2008, e as distribuições teóricas de Pearson tipo III para os parâmetros estimados pelo Método dos Momentos (MOM), pelo Método da Máxima Verossimilhança (MMV) e pelo Método dos Momentos-L (MML).

a remoção do dado extremo, que causou a mudança da obliquidade dos dados, a distribuição mostrou boa aderência, independente do método de ajuste. Quando o *outlier* foi considerado, dificilmente alguma distribuição conseguiria se ajustar muito bem aos dados amostrais, dada a característica deste extremo. Com relação ao Método dos Momentos-L, neste caso em que há poucos dados e um deles é extremo, observa-se que sua estimação foi bastante parecida ao da Máxima Verossimilhança, e no caso sem o dado extremo, a aderência foi um pouco melhor aos dados, mas a distribuição ajustada subestima os valores grandes. Considerando que este fosse um estudo hidrológico real, haveria duas possibilidades: primeiro, considerando o dado extremo para se fazer a estimação normalmente, e então tanto a Máxima Verossimilhança quanto os Momentos-L poderiam ser empregados

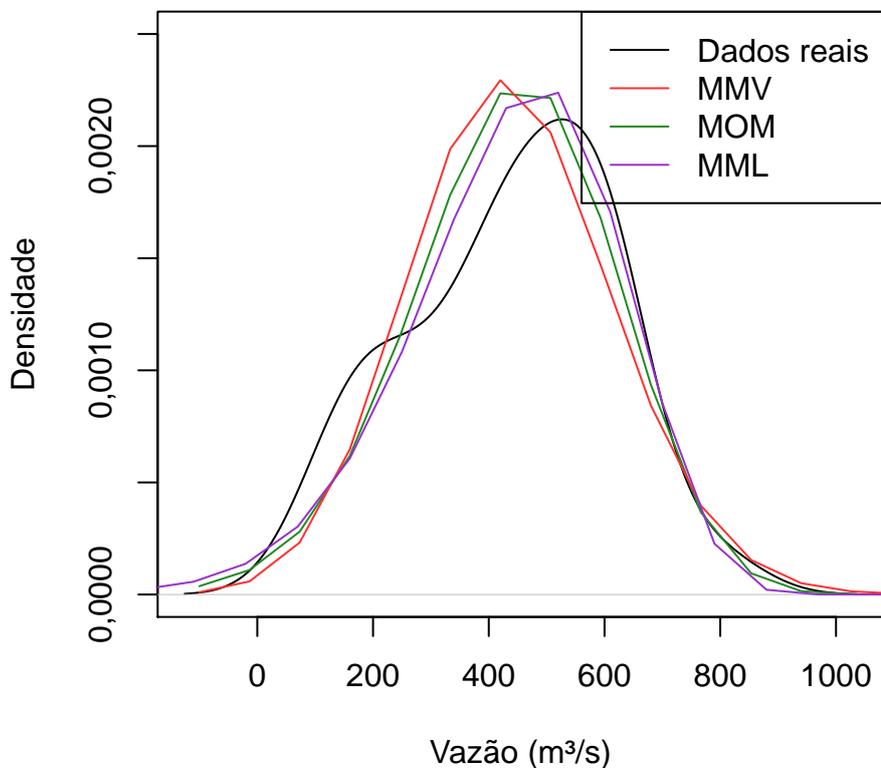


Figura 5.4: Densidade dos dados de vazão máxima anual do Rio São Sebastião, para o período de 1977 a 2008, desconsiderando o ano de 1981, e as distribuições teóricas de Pearson tipo III para os parâmetros estimados pelo Método dos Momentos (MOM), pelo Método da Máxima Verossimilhança (MMV) e pelo Método dos Momentos-L (MML).

(o Método dos Momentos deveria ser preterido); ou então, um peso (menor do que 1) seria adotado a este dado extremo, de forma a diminuir seu impacto nas estimativas. Em uma estimativa mais conservadora, o Método da Máxima Verossimilhança seria o preferido, por estimar as vazões de projeto, independente de qual seja, maiores do que os outros métodos, apesar da diferença não ser muito grande. Neste sentido, o Método dos Momentos-L seria, então, o menos indicado.

5.3 Rio Paraguai

O rio Paraguai é um grande rio situado na região centro-oeste do Brasil, banhando também Argentina, Bolívia e Paraguai. Para os dados de vazão máxima anual, foram

Tabela 5.4: Vazões para os tempos de retorno de 50, 100 e 1000 anos estimadas pelos Métodos dos Momentos (MOM), da Máxima Verossimilhança (MMV) e dos Momentos-L (MML) para o rio São Salvador, considerando e desconsiderando o valor extremo

Método	Vazão (m ³ /s)					
	Incluindo o <i>outlier</i>			Excluindo o <i>outlier</i>		
	Q50	Q100	Q1000	Q50	Q100	Q1000
MOM	1504,45	1802,15	2833,32	773,77	812,88	917,14
MMV	1241,84	1398,28	1901,32	810,58	863,44	1015,23
MLM	1193,45	1337,43	1798,46	755,5	786,43	862,94

usadas as vazões históricas de 40 anos de dados, de 1968 a 2007, da estação fluviométrica São Francisco (código ANA 66810000), localizada no município de Corumbá, Mato Grosso do Sul. As vazões máximas para os anos de 1989 a 1992 foram desconsideradas, devido à falta de dados. As vazões máximas anuais observadas para este rio podem ser vistas na Tabela 5.5.

Tabela 5.5: Vazões máximas anuais do rio Paraguai, para o período de 1968 a 2007

Ano	Q (m ³ /s)	Ano	Q (m ³ /s)	Ano	Q (m ³ /s)
1968	1181,868	1980	4238,595	1996	3242,146
1969	1063,758	1981	3276,473	1997	3731,764
1970	1208,015	1982	5043,494	1998	2531,341
1971	869,6035	1983	3208,119	1999	2517,469
1972	1086,582	1984	2837,718	2000	2702,836
1973	1150,711	1985	4153,854	2001	1771,384
1974	3091,36	1986	2355,884	2002	3026,245
1975	2058,541	1987	2807,27	2003	3225,095
1976	2559,276	1988	4897,128	2004	2369,01
1977	3191,218	1993	3058,658	2005	1843,275
1978	2822,46	1994	2035,245	2006	3562,151
1979	4132,891	1995	5295,533	2007	3140,958

Foi feito um ajuste da distribuição de Weibull a estes dados de vazão máxima. A densidade dos dados, juntamente com as distribuições teóricas com os parâmetros estimados pelos três métodos, pode ser vista na Figura 5.5.

Observa-se que os Métodos da Máxima Verossimilhança e dos Momentos ajustam curvas bem semelhantes, com boa aderência as dados amostrais como um todo. De fato, analisando as vazões estimadas pelos dois métodos para os tempos de retorno do 50, 100 e 1000 anos (Tabela 5.6), observa-se que a maior diferença relativa encontrada foi de apenas 5%. Já a curva estimada pelo Método dos Momentos-L não apresenta boa

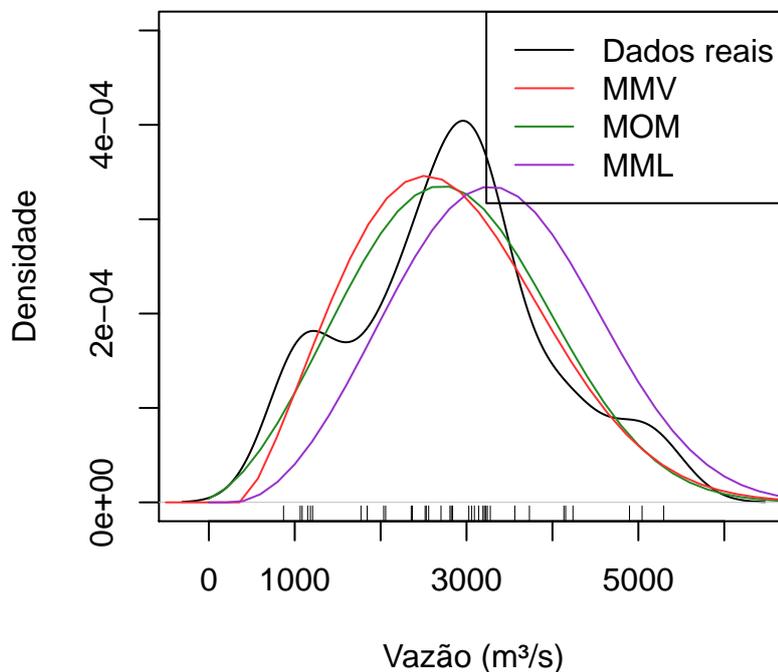


Figura 5.5: Densidade dos dados de vazão máxima anual do Rio Paraguai, para o período de 1968 a 2007, desconsiderando os anos de 1989, 1990, 1991 e 1992, e as distribuições teóricas de Weibull para os parâmetros estimados pelo Método dos Momentos (MOM), pelo Método da Máxima Verossimilhança (MMV) e pelo Método dos Momentos-L (MML).

aderência aos dados na parte da cauda à esquerda e no entorno da média, apresentando-se consideravelmente diferente das outras duas curvas, mas na região da cauda à direita tem comportamento semelhante, e que tende a ficar mais parecida com as outras curvas à medida que a probabilidade de não-excedência aumenta, como pode ser observado na Figura 5.6.

É interessante ressaltar, neste exemplo, a variação da máxima verossimilhança com os valores iniciais dos parâmetros da estimação. Foi observado neste exemplo que, para valores iniciais pequenos para os parâmetros, e também de ξ positivos, a estimativa do método se assemelhava mais a dos Momentos-L, e enquanto que para os demais valores iniciais a estimativa se assemelhava a do Método dos Momentos. Neste caso, foram utilizados como valores iniciais tanto os parâmetros estimados pelo Método dos Momentos

Tabela 5.6: Vazões para os tempos de retorno de 50, 100 e 1000 anos estimadas pelos Métodos dos Momentos (MOM), da Máxima Verossimilhança (MMV) e dos Momentos-L (MLM) para o rio Paraguai

Método	Vazão (m ³ /s)		
	Q50	Q100	Q1000
MOM	5236,37	5554,7	6429,31
MMV	5337,95	5701,44	6724,74
MLM	5769,84	6082,59	6939,25

quanto os do Método dos Momentos-L, e o resultado foi o mesmo e o adotado para este conjunto de dados. As razões por isso, no entanto, não foram investigadas neste trabalho.

Para este conjunto de dados, caso fosse necessário extrair as vazões de projeto para algum estudo hidrológico, o Método dos Momentos-L deveria ser preterido, pois não caracterizou tão bem os dados como um todo. O Método da Máxima Verossimilhança seria o mais indicado neste caso, para um estimativa mais conservadora, apesar de as diferenças relativas entre as estimativas deste e as do Método do Momentos serem pequenas.

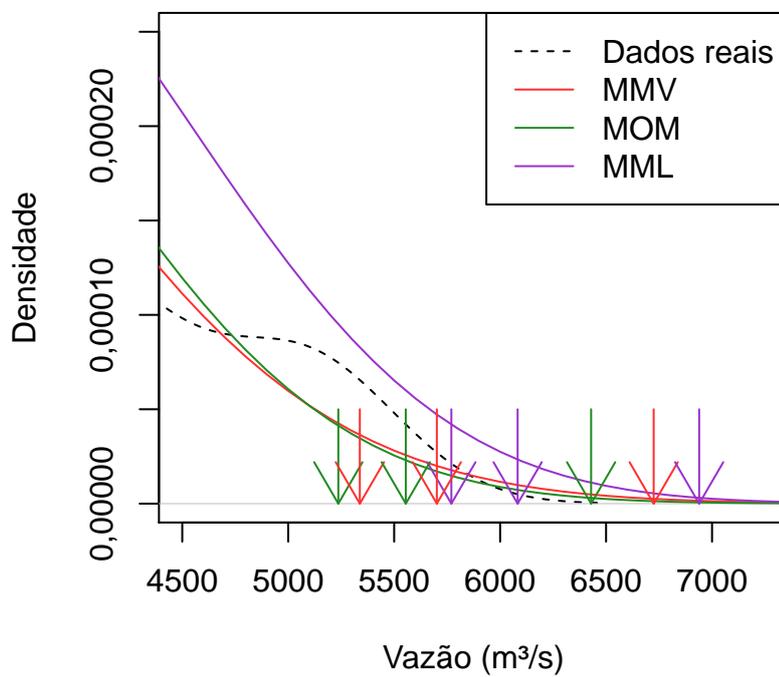


Figura 5.6: Densidade dos dados de vazão máxima anual do Rio Paraguai e as distribuições teóricas de Weibull para os parâmetros estimados pelo Método dos Momentos (MOM), pelo Método da Máxima Verossimilhança (MMV) e pelo Método dos Momentos-L (MML), na região da cauda à direita. As setas indicam as vazões para os tempos de retorno de 50, 100 e 1000 anos.

6. Estudo de simulação

Neste estudo foram feitas, ao todo, 36 simulações. Devido ao grande número de dados, foram incluídos nesta seção apenas alguns exemplos que representam o que foi observado neste estudo, e o restante foi incluído no Anexo. A nomenclatura adotada neste capítulo, para cada simulação, foi definida como: *Geradora – Estimada*, $n = x$, em que *Geradora* é a distribuição que gerou os dados da população simulada, *Estimada* é a distribuição escolhida para o ajuste aos dados, e $n = x$ é o tamanho de amostra utilizada nas estimações; além disso, foi especificado se houve contaminação de dados extremos à direita.

As simulações foram divididas em dois grupos: no primeiro, estão as simulações em que a distribuição geradora é a mesma distribuição ajustada; no segundo, são duas distribuições diferentes. Estes dois grupos de simulação serão discutidos em dois momentos distintos, a seguir.

6.1 Simulações com apenas uma distribuição

Foram realizadas 12 simulações com apenas uma distribuição, 4 para cada distribuição, e variando o tamanho da amostra e adicionando ou não contaminação. Em geral, os resultados foram bastante similares, e por isso, foram selecionados nesta seção quatro exemplos, de modo a ilustrar o padrão de comportamento de cada método, e também de cada distribuição nas simulações.

6.1.1 Pearson tipo III – Pearson tipo III, $n = 30$

O primeiro exemplo abordado dos estudos de simulação foi o caso em que a distribuição de Pearson tipo III é a geradora e a ajustada, e a amostra de dados utilizada

pelos métodos na estimação continha 30 dados. Neste caso, como pode-se ver na Figura 6.1, o Método da Máxima Verossimilhança, apesar de estimar os valores reais de Q50, Q100 e Q1000 na maior parte das vezes, em várias falha ao estimar bem os parâmetros, como pode ser observado pelas caudas à direita, o que aumenta o viés e a variância e, portanto, o erro quadrático médio do estimador, conforme a Tabela 6.1. Isto se deve a problemas de convergência do método: em uma amostra em que há muitos valores grandes de vazão, por exemplo, a função de máxima verossimilhança pode convergir para valores bastante distintos dos parâmetros da população, ou até mesmo não convergir a nenhum ponto, gerando, por exemplo, vários pontos em que a função é maximizada. O Método dos Momentos-L apresentou o menor viés entre os métodos, de apenas 3,5 para Q50, por exemplo, mas assim como a máxima verossimilhança, também em várias vezes falhou em estimar bem os parâmetros, o que gerou alta variância dos quantis e também alto EQM. Já o Método dos Momentos teve um viés razoavelmente pequeno, mas estimou bem os parâmetros na maioria das vezes, gerando pouca variância, e, por isso, o menor EQM entre os três métodos.

Tabela 6.1: Viés, variância e erro quadrático médio (EQM) das vazões para os tempos de retorno de 50, 100 e 1000 anos estimadas pelos Métodos dos Momentos (MOM), da Máxima Verossimilhança (MMV) e dos Momentos-L (MLM), para a simulação Pearson tipo III – Pearson tipo III, $n = 30$

Tempo de retorno	Medida	Método		
		MOM	MMV	MLM
Q50	Viés	-53,77	128,66	3,5
	Variância	$2,97 \times 10^5$	$3,43 \times 10^5$	$3,43 \times 10^5$
	EQM	3×10^5	$3,6 \times 10^5$	$3,43 \times 10^5$
Q100	Viés	-60,71	192,86	11,16
	Variância	$4,19 \times 10^5$	$4,98 \times 10^5$	$5,04 \times 10^5$
	EQM	$4,23 \times 10^5$	$5,36 \times 10^5$	$5,04 \times 10^5$
Q1000	Viés	-78,64	424,57	43,55
	Variância	$1,03 \times 10^6$	$1,3 \times 10^6$	$1,34 \times 10^6$
	EQM	$1,04 \times 10^6$	$1,48 \times 10^6$	$1,34 \times 10^6$

Nesta simulação, assim como nas simulações em há contaminação com dados à direita, o Método dos Momentos foi o que teve o menor EQM dos três métodos, e a diferença este método e os demais foi bastante alta; ele foi, por isso, considerado o que teve melhor performance, seguido do Método dos Momentos-L, e por último o Método da Máxima Verossimilhança.

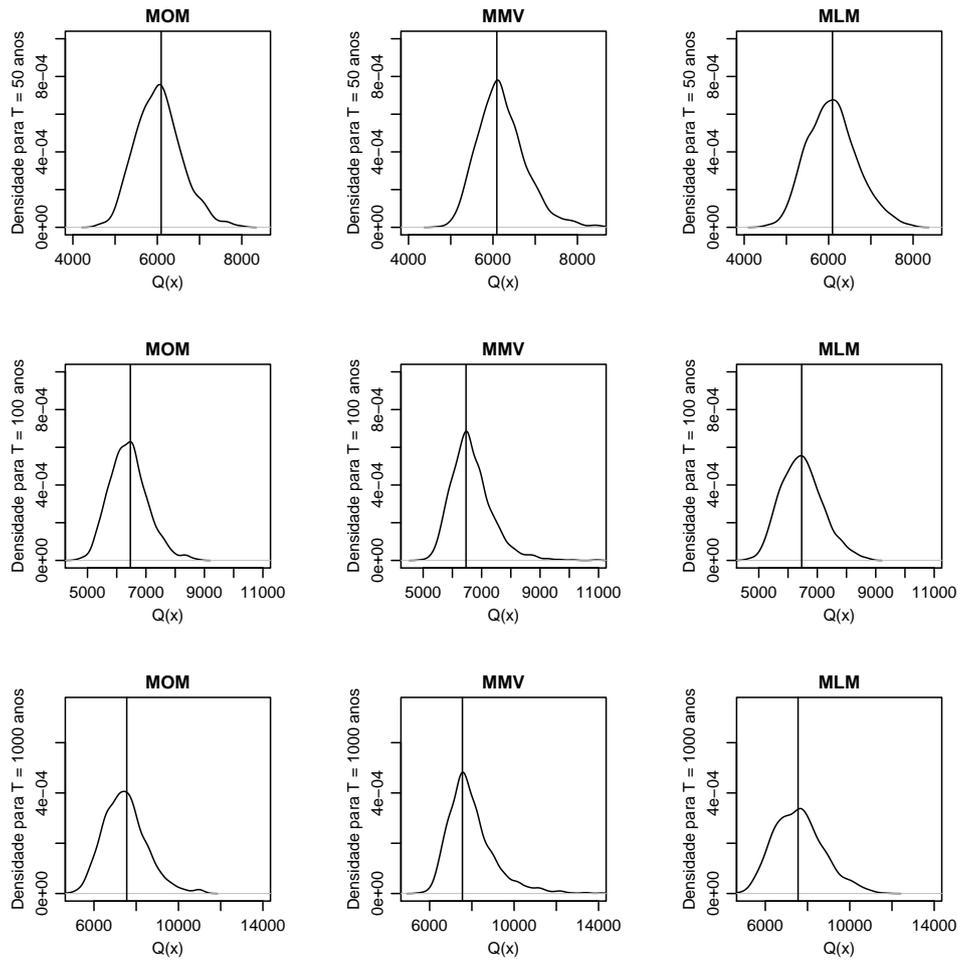


Figura 6.1: Densidade das vazões para os tempos de retorno de 50, 100 e 1000 anos estimadas pelo Método dos Momentos (MOM), pelo Método da Máxima Verossimilhança (MMV) e pelo Método dos Momentos-L (MLM), para a simulação Pearson tipo III – Pearson tipo III, $n = 30$. A linha vertical representa o valor da vazão para a população em cada caso.

6.1.2 Pearson tipo III – Pearson tipo III, $n = 50$

Nesta simulação a distribuição de Pearson tipo III é a geradora e a ajustada, e a amostra de dados utilizada pelos métodos na estimação era de 50 dados. Neste caso, como pode-se ver na Tabela 6.2, o Método das Máxima Verossimilhança teve desempenho bastante parecido ao do Método dos Momentos, se comparados os erros quadráticos médios para cada método, apesar de o viés ter sido novamente o maior entre os três métodos. Este caso ilustra a principal característica da máxima verossimilhança: conforme a amostra aumenta, a precisão e acurácia do método também aumentam, e quando a amostra tende

ao infinito, o método é o melhor entre todos possíveis (Hald, 1999; Hosking, 1990). Este comportamento foi observado para todas as distribuições utilizadas neste trabalho.

Tabela 6.2: Viés, variância e erro quadrático médio (EQM) das vazões para os tempos de retorno de 50, 100 e 1000 anos estimadas pelos Métodos dos Momentos (MOM), da Máxima Verossimilhança (MMV) e dos Momentos-L (MML), para a simulação Pearson tipo III – Pearson tipo III, $n = 50$

Tempo de retorno	Medida	Método		
		MOM	MMV	MLM
Q50	Viés	-29,52	89,02	5,65
	Variância	$1,93 \times 10^5$	$1,83 \times 10^5$	$2,16 \times 10^5$
	EQM	$1,94 \times 10^5$	$1,91 \times 10^5$	$2,16 \times 10^5$
Q100	Viés	-26,85	137,77	17,28
	Variância	$2,77 \times 10^5$	$2,58 \times 10^5$	$3,18 \times 10^5$
	EQM	$2,78 \times 10^5$	$2,77 \times 10^5$	$3,18 \times 10^5$
Q1000	Viés	-13,34	312,78	61,31
	Variância	$7,06 \times 10^5$	$6,35 \times 10^5$	$8,45 \times 10^5$
	EQM	$7,06 \times 10^5$	$7,33 \times 10^5$	$8,48 \times 10^5$

6.1.3 Gumbel – Gumbel, $n = 30$, com contaminação

Nesta simulação a distribuição de Gumbel gerou os dados e foi utilizada para o ajuste, e a amostra de dados utilizada pelos métodos na estimação conteve 30 dados, sendo que 5 destes eram extremos. Neste caso, o Método da Máxima Verossimilhança foi superior em todas as medidas (viés, variância e EQM), como pode ser verificado na Tabela 6.3. O Método dos Momentos apresentou desempenho semelhante, enquanto que o Método dos Momentos-L, por sua vez, apresentou desempenho bastante inferior do que os demais. Este resultado também foi verificado nas outras três simulações em que a Gumbel é geradora e ajustada aos dados.

Além disso, neste caso, é visível o impacto da contaminação com dados extremos na estimação dos três métodos. Como se pode ver na Figura 6.2, as estimações dos métodos está bastante deslocada dos valores de vazão da população. O menor viés obtido foi em torno de $1000 \text{ m}^3/\text{s}$, um valor bastante alto, considerando que as vazões da população se assemelham à vazão do rio Paraguai (Figura 5.5, cuja média é em torno de $3500 \text{ m}^3/\text{s}$).

Tabela 6.3: Viés, variância e erro quadrático médio (EQM) das vazões para os tempos de retorno de 50, 100 e 1000 anos estimadas pelos Métodos dos Momentos (MOM), da Máxima Verossimilhança (MMV) e dos Momentos-L (MLM), para a simulação Gumbel – Gumbel, $n = 30$, com contaminação

Tempo de retorno	Medida	Método		
		MOM	MMV	MLM
Q50	Viés	1095,05	1077,5	1248,98
	Variância	$2,86 \times 10^5$	$2,89 \times 10^5$	$2,86 \times 10^5$
	EQM	$1,48 \times 10^6$	$1,45 \times 10^6$	$1,85 \times 10^6$
Q100	Viés	1242,61	1223,55	1428,87
	Variância	$3,85 \times 10^5$	$3,84 \times 10^5$	$3,84 \times 10^5$
	EQM	$1,93 \times 10^6$	$1,88 \times 10^6$	$2,43 \times 10^6$
Q1000	Viés	1730,2	1706,13	2023,27
	Variância	$8,32 \times 10^5$	$8,03 \times 10^5$	$8,24 \times 10^5$
	EQM	$3,83 \times 10^6$	$3,71 \times 10^6$	$4,92 \times 10^6$

6.1.4 Weibull – Weibull, $n = 30$

Nesta simulação a distribuição de Weibull foi geradora e usada no ajuste, e a amostra de dados utilizada pelos métodos na estimação era de 30 dados. Neste caso, o Método da Máxima Verossimilhança apresentou viés alto, mas convergiu bem na maioria das vezes, gerando variância e EQM pequenos, se comparado aos demais, conforme a Tabela 6.4. Além disso, foi verificado que o método não convergiu em 35 da 1000 simulações, porém as razões não foram investigadas. O Método dos Momentos-L foi o que apresentou maior viés, variância e EQM entre os métodos, e foi considerado o pior estimador nos ajustes da distribuição de Weibull.

Tabela 6.4: Viés, variância e erro quadrático médio (EQM) das vazões para os tempos de retorno de 50, 100 e 1000 anos estimadas pelos Métodos dos Momentos (MOM), da Máxima Verossimilhança (MMV) e dos Momentos-L (MLM), para a simulação Weibull – Weibull, $n = 30$

Tempo de retorno	Medida	Método		
		MOM	MMV	MLM
Q50	Viés	20,75	-49,83	69,97
	Variância	$1,82 \times 10^5$	$1,73 \times 10^5$	$2,13 \times 10^5$
	EQM	$1,82 \times 10^5$	$1,75 \times 10^5$	$2,18 \times 10^5$
Q100	Viés	28,48	-57,73	89,85
	Variância	$2,46 \times 10^5$	$2,3 \times 10^5$	$2,98 \times 10^5$
	EQM	$2,47 \times 10^5$	$2,34 \times 10^5$	$3,06 \times 10^5$
Q1000	Viés	58,55	-75,33	159,36
	Variância	$5,34 \times 10^5$	$4,79 \times 10^5$	$6,94 \times 10^5$
	EQM	$5,37 \times 10^5$	$4,85 \times 10^5$	$7,19 \times 10^5$

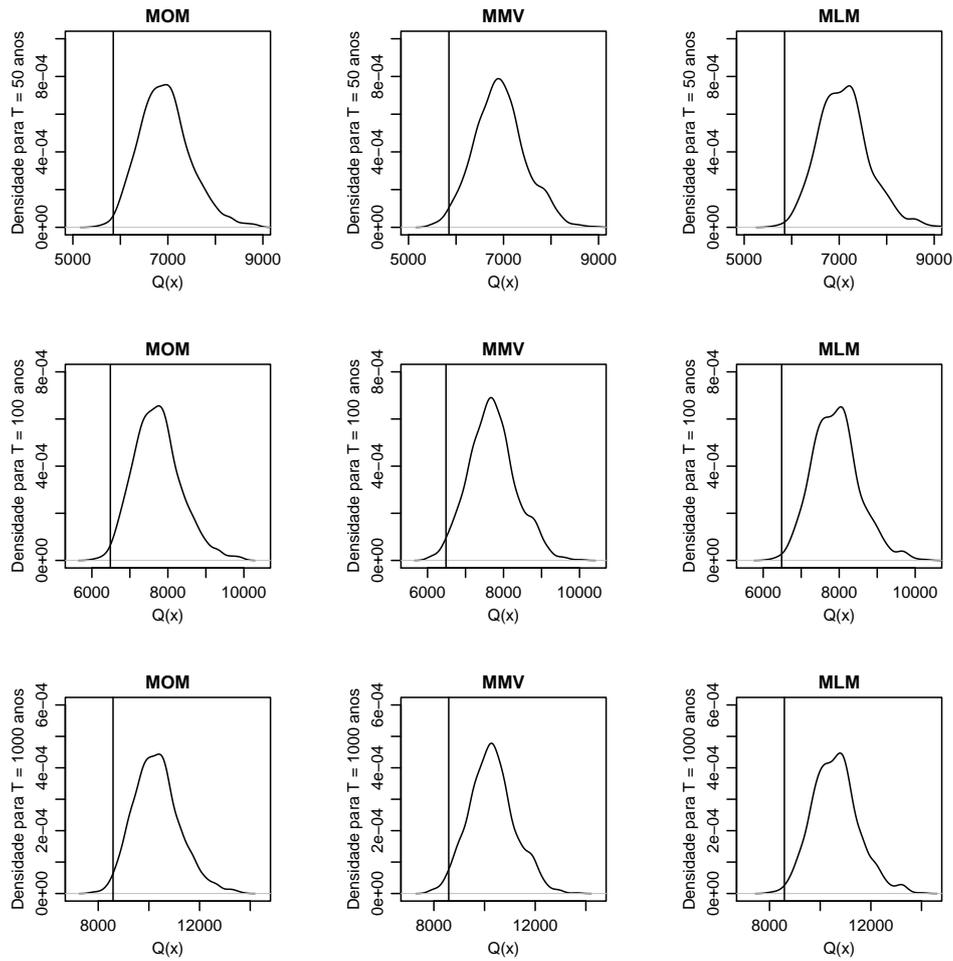


Figura 6.2: Densidade das vazões para tempo de retorno de 50, 100 e 1000 anos estimadas pelo Método dos Momentos (MOM), pelo Método da Máxima Verossimilhança (MMV) e pelo Método dos Momentos-L (MLM), para a simulação Gumbel – Gumbel, $n = 30$, com contaminação. A linha vertical representa o valor da vazão para a população em cada caso.

Para esta distribuição de probabilidade, o Método da Máxima Verossimilhança mostrou-se mais eficiente para amostras maiores, com ou sem contaminação, ainda que as estimações fossem bastante enviesadas quando houve contaminação nas amostras.

6.1.5 Discussão

Para as simulações onde a distribuição geradora é a mesma ajustada às amostras, observou-se um padrão: para a distribuição de Pearson tipo III, para amostras pequenas, o Método dos Momentos foi estimador com menor erro, e à medida que a amostra aumenta, o Método da Máxima Verossimilhança tende a ser superior, o que já era esperado do

método. Além disso, observou-se que a máxima verossimilhança, em alguns casos, estima valores de vazão com bastante desvio do valor real, como pode-se ver na Figura 6.1, o que contrasta com os picos das distribuições dos quantis, que foram bem parecidos com o valor real. Isto mostra que, para determinadas amostras, consideradas "*ruins*" para a máxima verossimilhança, como contendo vários valores extremos, ou que representem mal a população de dados, por exemplo, o método não converge bem, mas para os demais casos o método converge bastante bem. O Método dos Momentos-L foi o que estimou com o menor viés entre os métodos, porém variância alta, e o Método dos Momentos foi o que estimou com menor erro quadrático médio, e por isso foi considerado, para esta distribuição, o melhor estimador.

Com relação às outras duas distribuições, a de Gumbel e a de Weibull, o Método da Máxima Verossimilhança foi o melhor; em geral, a variância das vazões estimadas foi a menor entre os três métodos, e mesmo para amostras pequenas, com ou sem contaminação, o método foi estimador com menor EQM. O Método dos Momentos-L, em algumas das simulações, estimou com o menor viés entre os métodos, mas, em geral, foi o que teve o pior desempenho, e em alguns casos com diferenças bastante significativas, o que vai de encontro com alguns trabalhos encontrados na literatura (Delicado e Gorla, 2008; Hosking, 1989).

Outro resultado interessante foi a influência dos *outliers* na estimação. Assim como observado para o caso do rio São Salvador, em que o dado extremo causou um aumento drástico nas vazões para os tempos de retorno de 50, 100 e 1000 anos estimadas pelos métodos, nas simulações, todas as estimações foram bastante enviesadas à direita, como pode ser observado na Figura 6.2. Em estudos hidrológicos, isto deve ser estudado com cautela, mas de modo a considerar este valor, para não estimar a vazão de projeto com viés elevado, como ocorreu aqui.

6.2 Simulações com duas distribuições

Foram realizadas 24 simulações no caso de duas distribuições diferentes. Para ilustrar o comportamento dos métodos nestas simulações, foram selecionados quatro exemplos, que serão discutidos a seguir.

6.2.1 Gumbel – Pearson tipo III, $n = 50$, com contaminação

Nesta simulação a distribuição de Gumbel gerou a população, e a distribuição de Pearson tipo III foi usada no ajuste, e a amostra de dados utilizada pelos métodos na estimação continha 50 dados, sendo que 5 deles era extremos. Neste caso, o Método dos Momentos foi o que estimou com menor viés e variância, como pode ser visto na Tabela 6.5, sendo o melhor dos três estimadores. Já o Método da Máxima Verossimilhança teve o pior desempenho dos três, tanto em relação ao viés quanto à variância. Como pode ser visto na Figura 6.3, este método teve problemas em convergência em muitos casos, o que gerou dois picos na distribuição dos quantis estimados e variância bastante alta, e consequentemente erro bastante elevado.

Tabela 6.5: Viés, variância e erro quadrático médio (EQM) das vazões para os tempos de retorno de 50, 100 e 1000 anos estimadas pelos Métodos dos Momentos (MOM), da Máxima Verossimilhança (MMV) e dos Momentos-L (MLM), para a simulação Gumbel – Pearson tipo III, $n = 50$, com contaminação

Tempo de retorno	Medida	Método		
		MOM	MMV	MLM
Q50	Viés	509,12	766,34	736,36
	Variância	$2,65 \times 10^5$	$3,48 \times 10^5$	$3,01 \times 10^5$
	EQM	$5,24 \times 10^5$	$9,35 \times 10^5$	$8,43 \times 10^5$
Q100	Viés	471,98	833,87	787,05
	Variância	$4,01 \times 10^5$	$5,4 \times 10^5$	$4,72 \times 10^5$
	EQM	$6,24 \times 10^5$	$1,23 \times 10^6$	$1,09 \times 10^6$
Q1000	Viés	229,23	967,68	858,51
	Variância	$1,13 \times 10^6$	$1,57 \times 10^6$	$1,42 \times 10^6$
	EQM	$1,19 \times 10^6$	$2,51 \times 10^6$	$2,16 \times 10^6$

Este comportamento da máxima verossimilhança foi observado para todos os casos em que houve contaminação. Para as simulações sem contaminação, para $n = 30$ seu desempenho foi parecido com o do Método dos Momentos, e para $n = 50$, foi superior aos demais. O Método dos Momentos-L teve desempenho pior que o Método dos Momentos na maior parte das simulações, e foi melhor do que o Método da Máxima Verossimilhança apenas nas simulações com contaminação. Nota-se, também, que a contaminação não teve um impacto tão grande nas estimações como ocorreu nos casos em que usou-se a mesma distribuição para gerar os dados e para fazer o ajuste.

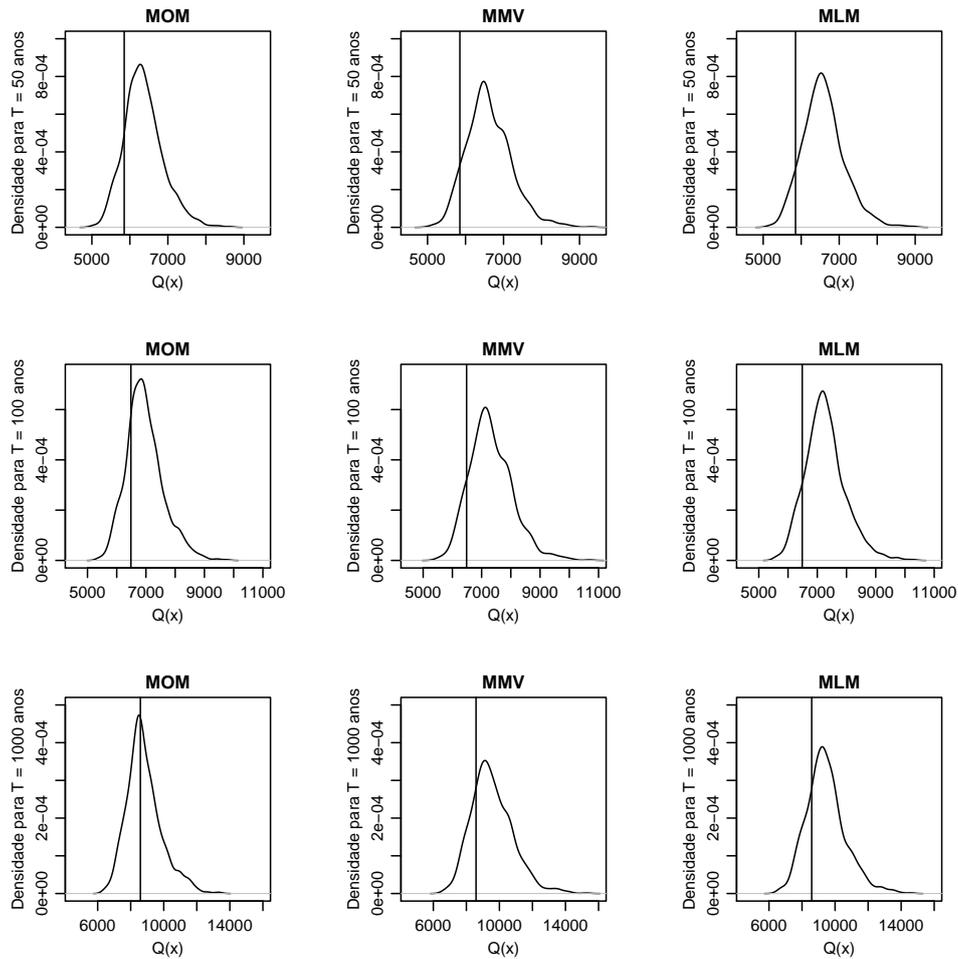


Figura 6.3: Densidade das vazões para tempo de retorno de 50, 100 e 1000 anos estimadas pelo Método dos Momentos (MOM), pelo Método da Máxima Verossimilhança (MMV) e pelo Método dos Momentos-L (MLM), para a simulação Gumbel – Pearson tipo III, $n = 50$, com contaminação. A linha vertical representa o valor da vazão para a população em cada caso.

6.2.2 Gumbel – Weibull, $n = 50$, com contaminação

Nesta simulação, a distribuição de Gumbel gerou os dados de vazão máxima da população, e a distribuição de Weibull foi ajustada às amostras, que continham 50 dados, sendo 5 deles extremos à direita. Verificou-se que o Método da Máxima Verossimilhança teve um desempenho bastante insatisfatório, não convergindo em 258 de 1000 amostras, apesar de que, nas vezes em que convergiu, a estimação foi boa (baixo viés e variância, se comparada aos outros métodos), como pode-se ver na Tabela 6.6. Essa baixa taxa de convergência também foi verificada nas demais simulações Gumbel – Weibull. A performance do Método

dos Momentos-L também foi bastante ruim, tendo o maior viés e maior variância entre os três métodos. O Método dos Momentos, neste caso, foi considerado o melhor estimador.

Tabela 6.6: Viés, variância e erro quadrático médio (EQM) das vazões para os tempos de retorno de 50, 100 e 1000 anos estimadas pelos Métodos dos Momentos (MOM), da Máxima Verossimilhança (MMV) e dos Momentos-L (MLM), para a simulação Gumbel – Weibull, $n = 50$, com contaminação

Tempo de retorno	Medida	Método		
		MOM	MMV	MLM
Q50	Viés	505,25	391,18	660,7
	Variância	$2,78 \times 10^5$	$2,71 \times 10^5$	$3,06 \times 10^5$
	EQM	$5,33 \times 10^5$	$4,25 \times 10^5$	$7,43 \times 10^5$
Q100	Viés	426,2	263,87	641,69
	Variância	$4,23 \times 10^5$	$3,97 \times 10^5$	$4,82 \times 10^5$
	EQM	$6,05 \times 10^5$	$4,67 \times 10^5$	$8,93 \times 10^5$
Q1000	Viés	-25,98	-354,07	400,46
	Variância	$1,2 \times 10^6$	$1,02 \times 10^6$	$1,46 \times 10^6$
	EQM	$1,2 \times 10^6$	$1,14 \times 10^6$	$1,62 \times 10^6$

Novamente, nota-se que a contaminação com dados extremos não aumentou tanto o viés quanto nos casos em que usou-se apenas uma distribuição. Como pode-se ver na Figura 6.4, os picos das distribuições dos quantis estimados pelos três métodos está pouco deslocado do valor real das vazões, em especial no caso da vazão para o tempo de retorno de 1000 anos, em que a vazão estimada pelos Métodos dos Momentos e da Máxima Verossimilhança foi, muitas vezes, subestimada.

6.2.3 Pearson tipo III – Weibull, $n = 30$

Nesta simulação, a distribuição de Pearson tipo III gerou os dados da população, e a distribuição de Weibull foi ajustada às amostras, que continham 30 dados. De acordo com a Tabela 6.7, nota-se que o Método dos Momentos-L foi o que estimou com menor viés, porém com a maior variância entre os três métodos, gerando, por isso, o maior erro quadrático médio. Já o Método da Máxima Verossimilhança estimou com o maior viés, porém com a menor variância, gerando o menor EQM entre os métodos. Foi observado, também, que o método convergiu bem, não convergindo em apenas 35 das 1000 amostras, e a mesma taxa pequena de não-convergência foi observada para as demais simulações Pearson tipo III – Weibull. Além disso, verificou-se que os picos das distribuições dos quantis estimados por todos os métodos, como observado na Figura 6.5 ficou bastante

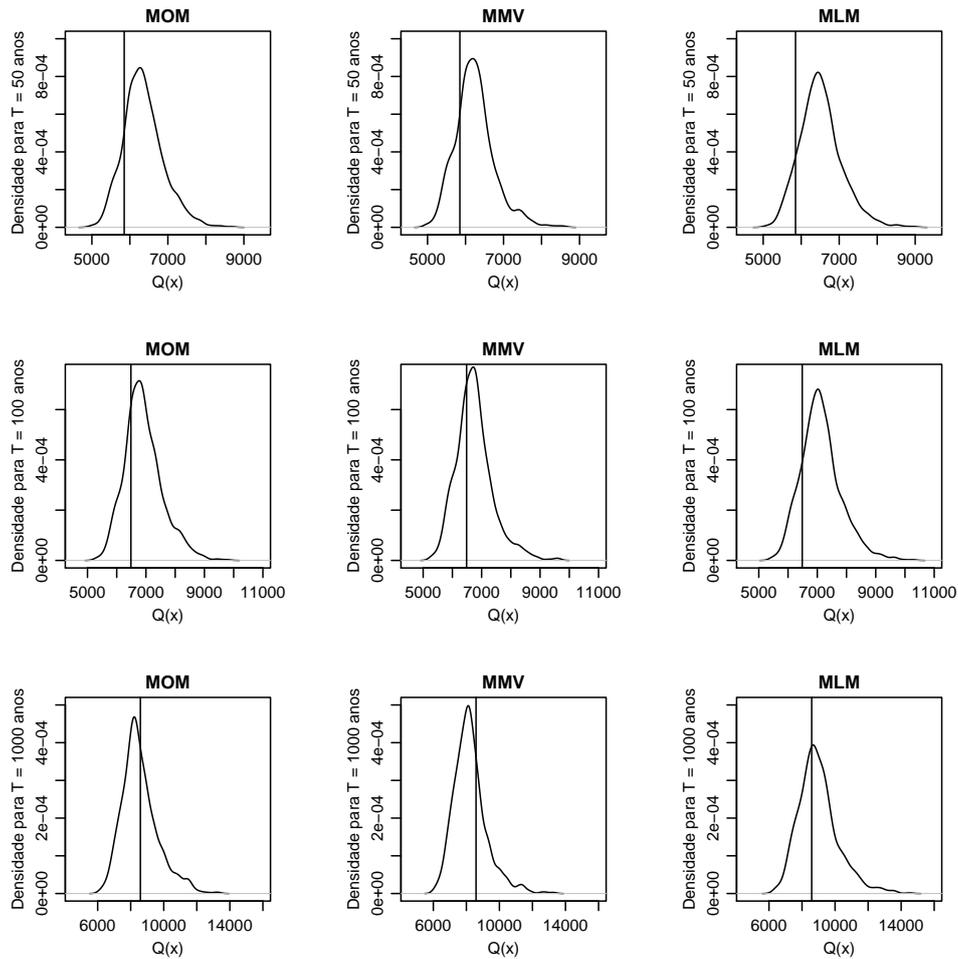


Figura 6.4: Densidade das vazões para tempo de retorno de 50, 100 e 1000 anos estimadas pelo Método dos Momentos (MOM), pelo Método da Máxima Verossimilhança (MMV) e pelo Método dos Momentos-L (MLM), para a simulação Gumbel – Weibull, $n = 50$, com contaminação. A linha vertical representa o valor da vazão para a população em cada caso.

próximo do valor real das vazões para os tempos de retorno de 50, 100 e 1000 anos.

Neste caso, também foi observado que a contaminação gerou viés bastante alto na estimação dos três métodos, como observado nas simulações com apenas uma distribuição. Ademais, nestes casos, o Método da Máxima Verossimilhança foi o melhor estimador, estimando os quantis com menor viés e variância entre os três métodos.

6.2.4 Weibull – Gumbel, $n = 30$, com contaminação

Nesta simulação a distribuição de Weibull gerou os dados, e a distribuição de Gumbel foi ajustada às amostras, que continham 30 dados, sendo 5 destes extremos à

Tabela 6.7: Viés, variância e erro quadrático médio (EQM) das vazões para os tempos de retorno de 50, 100 e 1000 anos estimadas pelos Métodos dos Momentos (MOM), da Máxima Verossimilhança (MMV) e dos Momentos-L (MLM), para a simulação Pearson tipo III – Weibull, $n = 30$

Tempo de retorno	Medida	Método		
		MOM	MMV	MLM
Q50	Viés	-83,77	-167,17	-44,44
	Variância	$2,99 \times 10^5$	$2,79 \times 10^5$	$3,19 \times 10^5$
	EQM	$3,07 \times 10^5$	$3,07 \times 10^5$	$3,21 \times 10^5$
Q100	Viés	-124,92	-228,79	-76,45
	Variância	$4,09 \times 10^5$	$3,72 \times 10^5$	$4,41 \times 10^5$
	EQM	$4,25 \times 10^5$	$4,24 \times 10^5$	$4,47 \times 10^5$
Q1000	Viés	-283,86	-451,65	-206,54
	Variância	$8,95 \times 10^5$	$7,65 \times 10^5$	$9,81 \times 10^5$
	EQM	$9,76 \times 10^5$	$9,69 \times 10^5$	$1,02 \times 10^6$

direita. Como pode-se observar na Tabela 6.8, todos os métodos tiveram desempenho bastante ruins, com viés extremamente alto. O Método dos Momentos, que foi o que estimou com o menor viés e a menor variância entre os três, estimou, por exemplo, a vazão para o tempo de retorno de 1000 anos com desvio médio de $3000 \text{ m}^3/\text{s}$, o que é cerca de 50% maior do que o valor real, de aproximadamente $6000 \text{ m}^3/\text{s}$, como pode ser visto na Figura 6.6. Além disso, o Método da Máxima Verossimilhança apresentou desempenho aquém dos demais, com o dobro do erro quadrático médio da estimação do Método dos Momentos. O Método dos Momentos-L também estimou com mais viés e variância do que o Método dos Momentos, ainda que tenha tido desempenho melhor do que o da Máxima Verossimilhança.

Este comportamento foi observado em todas as simulações em que a distribuição de Weibull foi a que gerou a população de dados, independentemente do tamanho da amostra ou se houve ou não contaminação, ainda que, nos casos em que não houve contaminação, o viés na estimação dos métodos foi menor, como era esperado.

6.2.5 Discussão

O que se destacou, nas simulações em que as distribuições ajustadas são diferentes das que geraram a população de dados de vazão máxima anual, foi que nos casos em que a distribuição de Gumbel foi ajustada, o viés e a variância das estimações feitas pelos três métodos são bastante altos. Isto se deve ao fato de a distribuição de Gumbel ser limitada

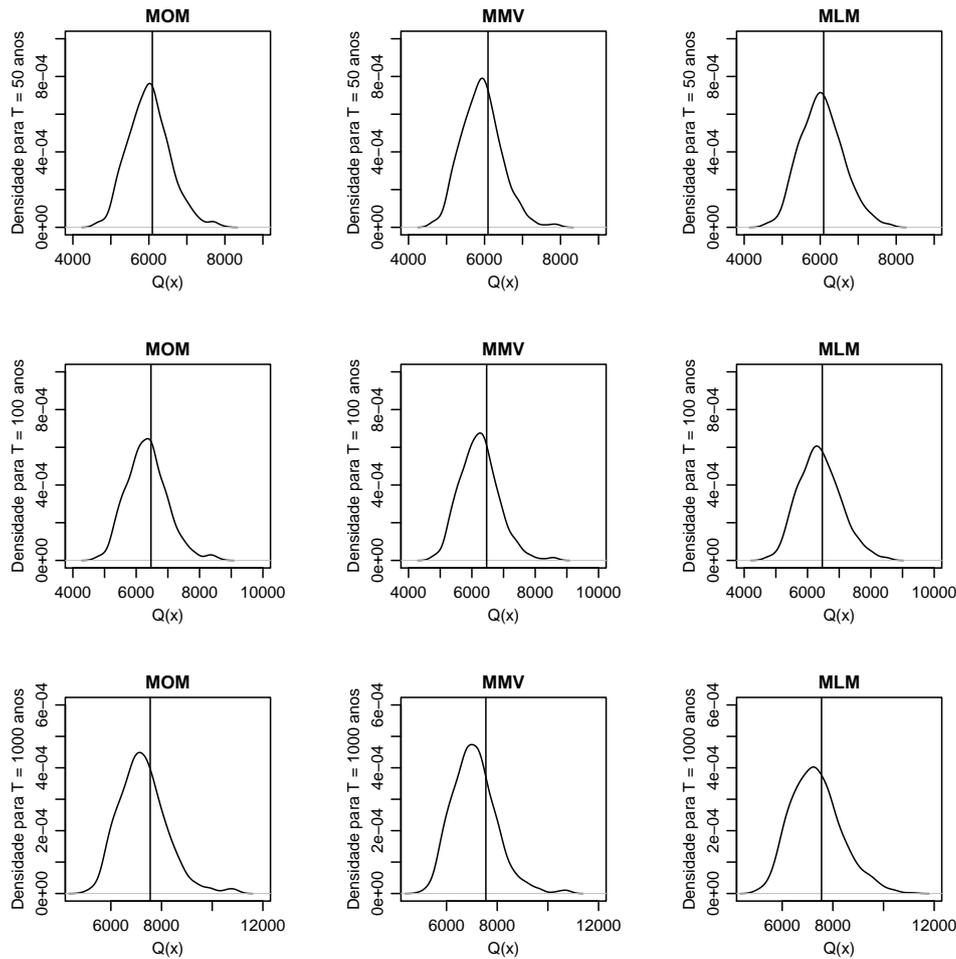


Figura 6.5: Densidade das vazões para tempo de retorno de 50, 100 e 1000 anos estimadas pelo Método dos Momentos (MOM), pelo Método da Máxima Verossimilhança (MMV) e pelo Método dos Momentos-L (MLM), para a simulação Pearson tipo III – Weibull, $n = 30$. A linha vertical representa o valor da vazão para a população em cada caso.

tanto quanto à obliquidade dos dados (a distribuição só pode ter obliquidade positiva, e Eletrobrás (2000), por exemplo, recomenda o seu uso apenas para valores de $\gamma < 1,5$), quanto ao número de parâmetros: um de localidade e um de escala, ou seja, a distribuição tem uma forma definida, à qual os dados podem ou não aderir bem. O Método da Máxima Verossimilhança, em particular, apresentou desempenho bastante ruim, não convergindo em muitas amostras e estimando os quantis com viés e variância bastante elevados.

Na maior parte dos casos em que as distribuições de Pearson tipo III ou de Weibull foram ajustadas, quando não houve contaminação dos dados à direita, a estimação dos três métodos foi satisfatória. O Método dos Momentos foi, contudo, o que se sobressaiu

Tabela 6.8: Viés, variância e erro quadrático médio (EQM) das vazões para os tempos de retorno de 50, 100 e 1000 anos estimadas pelos Métodos dos Momentos (MOM), da Máxima Verossimilhança (MMV) e dos Momentos-L (MLM), para a simulação Weibull – Gumbel, $n = 30$, com contaminação

Tempo de retorno	Medida	Método		
		MOM	MMV	MLM
Q50	Viés	1266,82	1812,13	1480,86
	Variância	$1,01 \times 10^5$	$2,11 \times 10^5$	$1,17 \times 10^5$
	EQM	$1,71 \times 10^6$	$3,49 \times 10^6$	$2,31 \times 10^6$
Q100	Viés	1642,98	2293,9	1901,97
	Variância	$1,39 \times 10^5$	$2,88 \times 10^5$	$1,63 \times 10^5$
	EQM	$2,84 \times 10^6$	$5,55 \times 10^6$	$3,78 \times 10^6$
Q1000	Viés	3049,01	4048,93	3456,53
	Variância	$3,21 \times 10^5$	$6,42 \times 10^5$	$3,82 \times 10^5$
	EQM	$9,62 \times 10^6$	$1,7 \times 10^7$	$1,23 \times 10^7$

na análise das medidas de erro nas estimações: embora o Método dos Momentos-L, em parte das simulações, tenha estimado com o menor viés entre os três métodos, o Método dos Momentos convencional foi o que teve menor variância na estimação, além de baixo viés, o que gerou o menor erro quadrático médio entre os métodos na maioria das estimações. O Método da Máxima Verossimilhança mostrou-se o mais indicado para os casos em que a amostra continha 50 dados, e também nos casos em que houve contaminação da amostra com dados extremos à direita, quando a distribuição de Weibull foi ajustada.

Nota-se que nos casos em que a distribuição de Gumbel foi a geradora, à medida em que o tempo de retorno aumenta, ou seja, que o quantil da distribuição "*caminha*" para a direita (ou que a probabilidade de não-excedência aumenta), o viés tende a diminuir para as estimações de todos os métodos, e para o maior tempo de retorno utilizado neste trabalho, de 1000 anos, as vazões estimadas pelos três métodos foram, na maior parte das vezes, bem próximas do valor teórico. Nestas simulações, observou-se que o Método da Máxima Verossimilhança também teve problemas de convergência, principalmente quando a distribuição de Weibull foi ajustada, falhando em estimar os parâmetros, na média, em 270 das 1000 amostras. Este método foi, contudo, o que estimou com menor EQM nos casos em que houve contaminação das amostras, apesar de ter sido desconsiderado devido ao grande número de falhas na estimação. Em todos estes, novamente, o Método dos Momentos estimou os quantis que menor EQM do que o Método dos Momentos-L.

Observou-se aqui, assim como nos casos em que usou-se apenas uma distribuição,

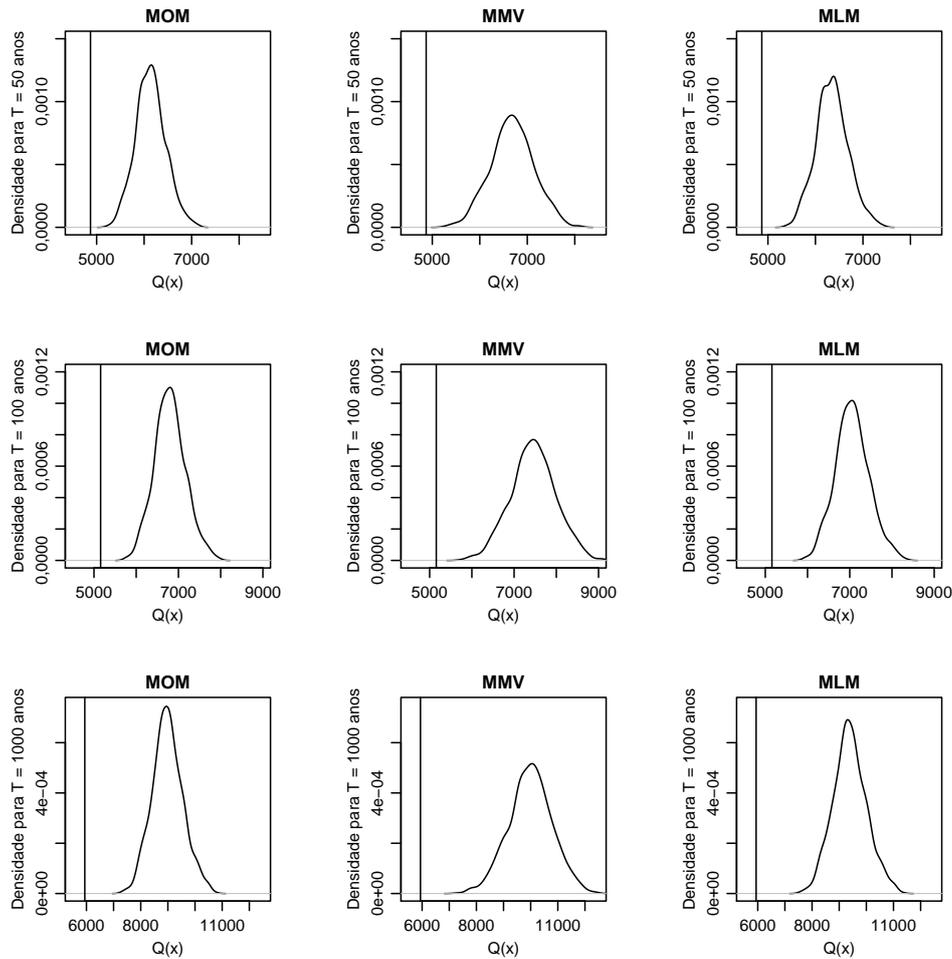


Figura 6.6: Densidade das vazões para tempo de retorno de 50, 100 e 1000 anos estimadas pelo Método dos Momentos (MOM), pelo Método da Máxima Verossimilhança (MMV) e pelo Método dos Momentos-L (MLM), para a simulação Weibull – Gumbel, $n = 30$, com contaminação. A linha vertical representa o valor da vazão para a população em cada caso.

que o viés das estimações dos três métodos aumentou significativamente quando houve contaminação da amostra, em especial no caso em que a distribuição de Pearson tipo III foi a geradora e o ajuste foi feito com a distribuição de Weibull. É importante ressaltar, como feito na seção anterior, que estes dados extremos devem ser estudados cuidadosamente, para evitar erros de estimação de vazões de projeto em estudos hidrológicos diversos.

Com relação aos métodos, é notável que o Método dos Momentos-L, ainda que em algumas simulações tenha estimado com menos viés do que os demais, em nenhuma simulação foi o método que estimou com variância ou com menor erro quadrático médio: na maior parte das simulações em que a amostra era de 50 dados, e em algumas em que a

amostra era menor mas continha menos dados, o Método da Máxima Verossimilhança foi o melhor estimador; nos demais casos, o Método dos Momentos foi superior. É interessante destacar, inclusive, que de todas as 24 simulações, em apenas uma o Método dos Momentos-L estimou com menor EQM do que o Método dos Momentos convencional, resultado similar ao encontrado nas simulações utilizando apenas uma distribuição, resultado semelhante ao encontrado por Oztekin (2005), por exemplo.

Ressalta-se, também, o contraste dos vieses e das variâncias das estimações nestas simulações com os das simulações utilizando apenas uma distribuição: comparando as Figuras 6.1 e 6.5, e as Tabelas 6.2 e 6.7, por exemplo, observa-se que o viés chega a ser o dobro, quando a distribuição ajustada é diferente da distribuição que originou os dados. Mesmos nos casos em que a distribuição ajustada foi a de Pearson tipo III ou a de Weibull, que são distribuições de três parâmetros e que podem assumir qualquer valor, positivo ou negativo, de obliquidade, os ajustes não foram tão bons como nos casos em que usou-se a mesma distribuição que gerou os dados da população. Isso reforça a importância que deve ser dada à escolha da distribuição que será ajustada aos dados em estudos hidrológicos, utilizando ferramentas como gráficos de obliquidade–curtose (Cullen e Frey, 1999), de $\log - \log$ ou de Razão da máxima soma (Beirlant et al., 2004), por exemplo.

7. Conclusão

Neste trabalho, foram avaliados os desempenhos dos Métodos dos Momentos, da Máxima Verossimilhança e dos Momentos-L, na estimativa de parâmetros das distribuições de Gumbel, de Pearson tipo III e de Weibull a dados de vazão máxima anual, tanto a dados reais quanto a dados simulados. Para os dados reais, quando o tamanho das amostras é considerado grande (maior do que 50 dados), como no caso do rio dos Patos, os três métodos estimam bem os parâmetros, e as diferenças entre os quantis é pequena. Para os casos em que a amostra é menor, como para os rios São Salvador e Paraguai, o Método da Máxima Verossimilhança foi o que ajustou as distribuições que melhor aderiram aos dados, e também fez estimativas maiores das vazões para os tempos de retorno de 50, 100 e 1000 anos, o que seria preferido em estudos hidrológicos para construção de obras hidráulicas, embora as diferenças relativas não tenha sido grandes. Além disso, o *outlier* observado na vazão máxima do rio São Salvador em 1981, sendo quase o dobro da média dos restantes das máximas anuais, mudou consideravelmente as estimativas feitas pelos três métodos, em especial o Método dos Momentos, que foi bastante sensível e estimou uma distribuição aparentemente muito aderente a estes dados. Aconselha-se, então, em casos como este, alternativas quanto ao uso direto de dados deste tipo em estudos hidrológicos.

Quanto às simulações, tanto nos casos em que a distribuição ajustada era a mesma que gerou a população de dados de vazão máxima anual, quanto nos casos em que eram diferentes, o Método dos Momentos-L nunca foi, com exceção de uma simulação, o estimador com menor erro quadrático médio, embora em algumas simulações ele tenha estimado com o menor viés entre os três métodos. Para as simulações com amostras de 50 dados, e em alguns casos em que as amostras continham 30 dados, mas com contaminação, na maior parte dos casos, o Método da Máxima Verossimilhança estimou com menor viés e erro quadrático médio, sendo o melhor estimador; nas demais simulações, o Método

dos Momentos foi superior. O Método da Máxima Verossimilhança, contudo, apresentou problemas de convergência nas estimações dos parâmetros para algumas amostras, para as distribuições de Weibull, em que foi utilizado um maximizador computacionalmente, e de Pearson tipo III, em que foi utilizado o Método de Newton-Raphson. Em estudos hidrológicos individuais (para apenas um rio ou bacia), porém, se este problema de convergência pudesse ser de alguma forma contornado, este método seria o mais indicado; caso contrário, o Método dos Momentos seria o mais aconselhável.

Foi observado, também, que a contaminação das amostras com dados extremos à direita teve um impacto bastante grande nas estimações das vazões para os tempos de retorno utilizados neste estudo, gerando bastante viés; isto deve ser considerado em estudos hidrológicos reais pois, assim como no caso do rio São Salvador, estes dados extremos gerar bastante viés. Estes dados devem ser trabalhados com cuidado, como, por exemplo, atribuindo pesos aos dados atípicos da amostra. Além disso, nestes estudos as distribuições devem ser cuidadosamente escolhidas de acordo com as características que o conjunto de dados apresenta, como a obliquidade e a cauda à direita, pois uma escolha ruim pode gerar baixa acurácia na estimação das vazões de projeto, o que gera, por sua vez, superdimensionamento de estruturas e conseqüente aumento do custo, quando em estudos de obras hidráulicas, como usinas hidrelétricas.

Referências Bibliográficas

- Agência Nacional de Águas – ANA (2005). *Hidroweb: sistema de informações hidrológicas*. Disponível em: <<http://hidroweb.ana.gov.br/>>. Acesso em 22 de novembro de 2015.
- Akram, M. e Hayat, A. (2014). Comparison of estimators of the Weibull distribution. *Journal of Statistical Theory and Practice*, 8(2):238–259.
- Aucoin, F. (2015). *FAdist: Distributions that are Sometimes Used in Hydrology*. R package version 2.1.
- Bai, J., Jakeman, A. J., e Taylor, J. A. (1990). Percentile estimation of the three-parameter gamma and lognormal distributions: method of moments versus maximum likelihood. *Mathematics and Computers in Simulation*, 32:167–172.
- Becker, M. e Klößner, S. (2013). *PearsonDS: Pearson Distribution System*. R package version 0.97.
- Beirlant, J., Goegebeur, Y., Segers, J., e Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*. Wiley.
- Bobée, B. B. e Robitaille, R. (1977). The use of Pearson type 3 and log-Pearson type 3 distribution revised. *Water Resources Research*, 13(2):427–442.
- Chow, V. T., Maidment, D. R., e Mays, L. W. (1988). *Applied Hydrology*. McGraw-Hill.
- Cousineau, D. (2009). Fitting the three-parameter Weibull distribution: review and evaluation of existing and new methods. *IEEE Transactions on Dielectrics and Electrical Insulation*, 16(2):281–288.
- Cullen, A. C. e Frey, H. C. (1999). *Probabilistic techniques in exposure assessment*. Plenum Press.

- Delicado, P. e Gorla, M. N. (2008). A small sample comparison of maximum likelihood, moments and L-moments methods for the asymmetric exponential power distribution. *Computational Statistics and Data Analysis*, 52(3):1661–1673.
- Eletrobrás (2000). *Diretrizes para estudos e projetos de Pequenas Centrais Hidrelétricas*.
- Fisher, R. A. (1912). On an absolute criterion for fitting frequency curves. *Messenger of Mathematics*, 41:155–160.
- Fisher, R. A. (1921). On the “probable error” of a coefficient of correlation deduced from a small sample. *Metron*, 1:3–32.
- Fisher, R. A. (1937). Professor Karl Pearson and the Method of Moments. *Annals of Eugenics*, 7(4):303–318.
- Goda, Y., Kudaka, M., e Kawai, H. (2010). Incorporation of Weibull distribution in L-moments method for regional frequency analysis of peaks-over-threshold wave heights. Em *Proceedings of the 32nd International Conference on Coastal Engineering*. Shanghai, China.
- Greenwood, J. A., Landwehr, J. M., Matalas, N. C., e Wallis, J. R. (1979). Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Water Resources Research*, 15(5):1049–1054.
- Gumbel, E. J. (1935). Les valeurs extrêmes des distributions statistiques. *Annales de l’Institut Henri Poincaré*, 5(2):115–158.
- Gumbel, E. J. (1941). The return period of flood flows. *The Annals of Mathematical Statistics*, 12:163–190.
- Hald, A. (1999). On the history of maximum likelihood in relation to inverse probability and least squares. *Statistical Science*, 14(2):214–222.
- Hosking, J. R. M. (1989). Some theoretical results concerning L-moments. Relatório técnico, IBM Research.
- Hosking, J. R. M. (1990). L-Moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society*, 52(1):105–124.

- Hosking, J. R. M. (2015). *L-moments*. R package, version 2.5.
- Hosking, J. R. M. e Wallis, J. R. (1997). *Regional frequency analysis: an approach based on L-moments*. Cambridge University Press.
- Interagency Advisory Committee on Water Data (1982). *Guidelines for determining flood flow frequency*. Hydrology Subcommittee, Office of Water Data Coordination, U.S. Geological Survey, Reston, VA.
- Komsta, L. e Novomestky, F. (2015). *moments: Moments, cumulants, skewness, kurtosis and related tests*. R package version 0.14.
- Landwehr, J. R., Matalas, N. C., e Wallis, J. R. (1979). Probability-weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. *Water Resources Research*, 14:902–920.
- Mahdi, S. e Cenac, M. (2005). Estimating parameters of Gumbel distribution using the methods of moments, probability weighted moments and maximum likelihood. *Revista de Matemática: Teoría y aplicaciones*, 12(1 & 2).
- Oztekin, T. (2005). Comparison of parameter estimation methods for the three-parameter Generalized Pareto Distribution. *Turkish Journal of Agriculture and Forestry*, 29:419–428.
- Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society A*, 185:343–414.
- Pearson, K. (1895). Contributions to the mathematical theory of evolution, II: skew variation in homogeneous material. *Philosophical Transactions of the Royal Society A*, 186:71–110.
- R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Serrano, S. E. (1997). *Hydrology for engineers, geologists and environmental professionals*. HydroScience Inc.

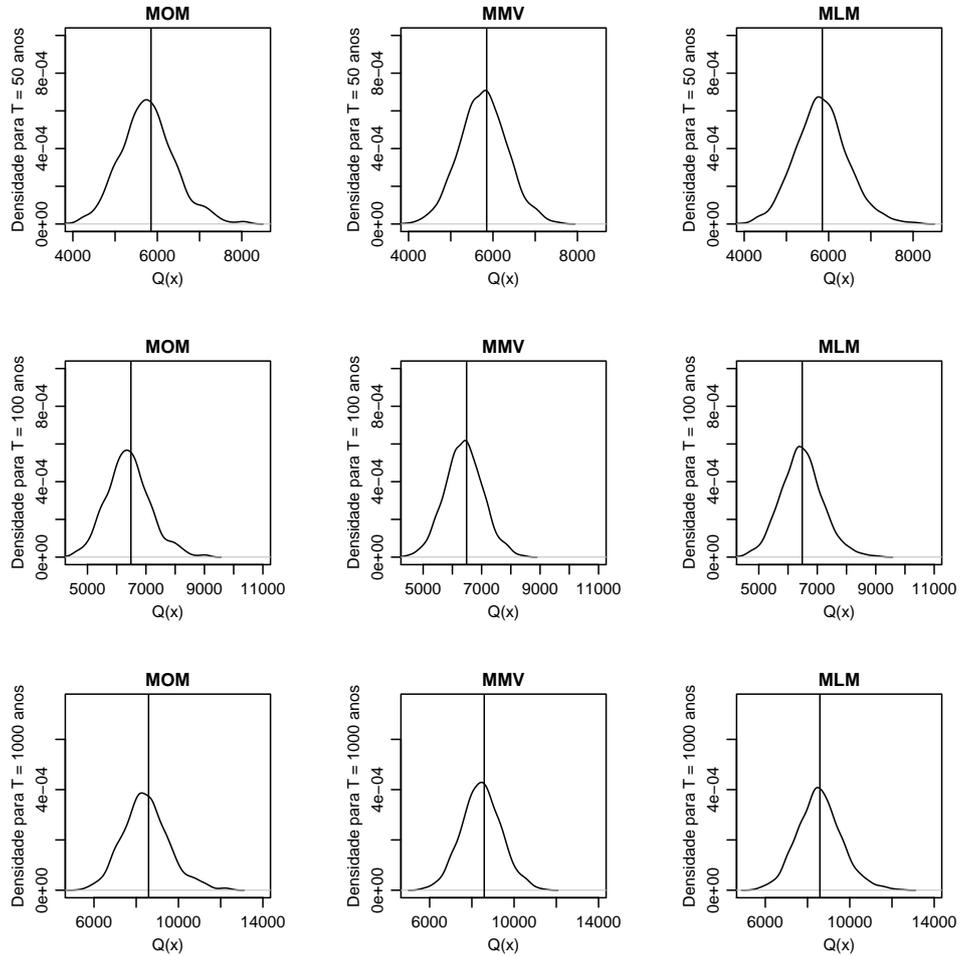
Soetaert, K. (2009). *rootSolve: Nonlinear root finding, equilibrium and steady-state analysis of ordinary differential equations*. R package version 1.6.

Venables, W. N. e Ripley, B. D. (2002). *Modern Applied Statistics with S*. New York, fourth edio. ISBN 0-387-95457-0.

Weibull, W. (1951). A statistical distribution function of wide applicability. *Journal of Applied Mechanics Transactions ASME*, 18(3):293–297.

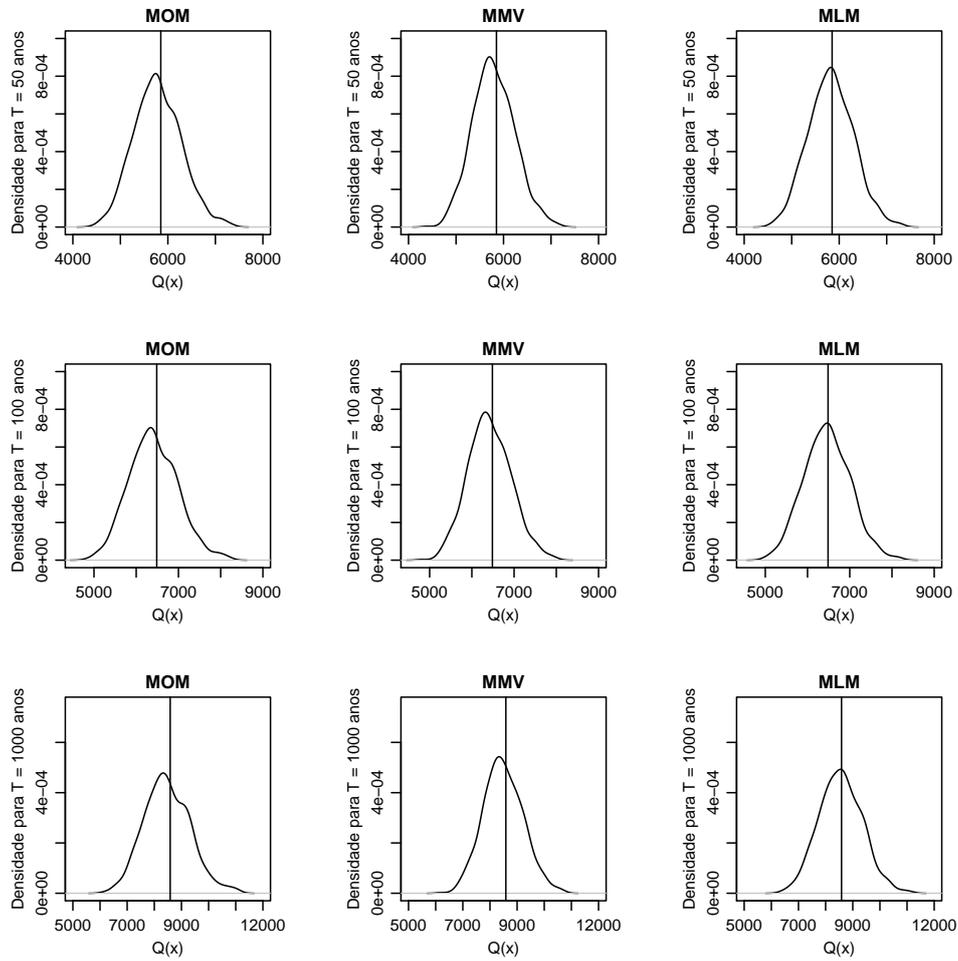
Anexos

Gumbel – Gumbel, n = 30



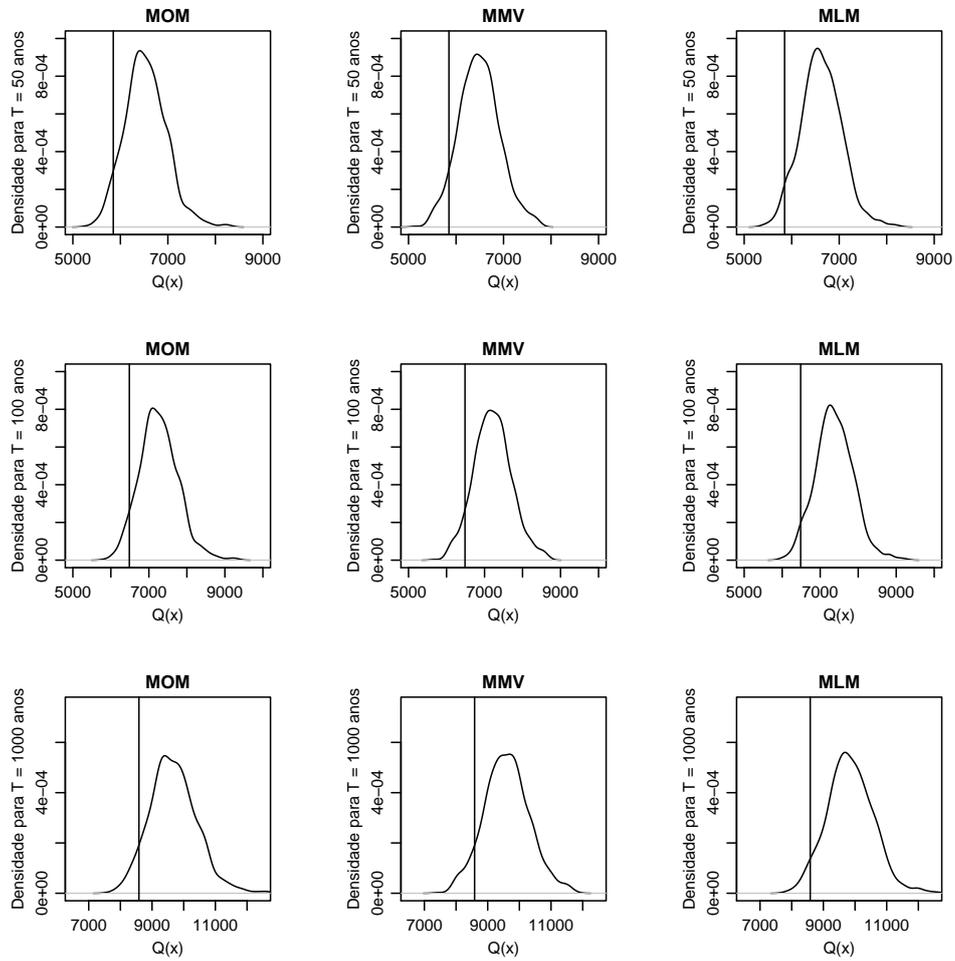
Tempo de retorno	Medida	Método		
		MOM	MMV	MLM
Q50	Viés	-72,9	-78,75	-13,12
	Variância	$4,19 \times 10^5$	$3,19 \times 10^5$	$3,87 \times 10^5$
	EQM	$4,24 \times 10^5$	$3,25 \times 10^5$	$3,88 \times 10^5$
Q100	Viés	-85,32	-92,3	-12,98
	Variância	$5,64 \times 10^5$	$4,23 \times 10^5$	$5,19 \times 10^5$
	EQM	$5,72 \times 10^5$	$4,32 \times 10^5$	$5,19 \times 10^5$
Q1000	Viés	-126,34	-137,07	-12,53
	Variância	$1,21 \times 10^6$	$8,79 \times 10^5$	$1,1 \times 10^6$
	EQM	$1,23 \times 10^6$	$8,98 \times 10^5$	$1,1 \times 10^6$

Gumbel – Gumbel, n = 50



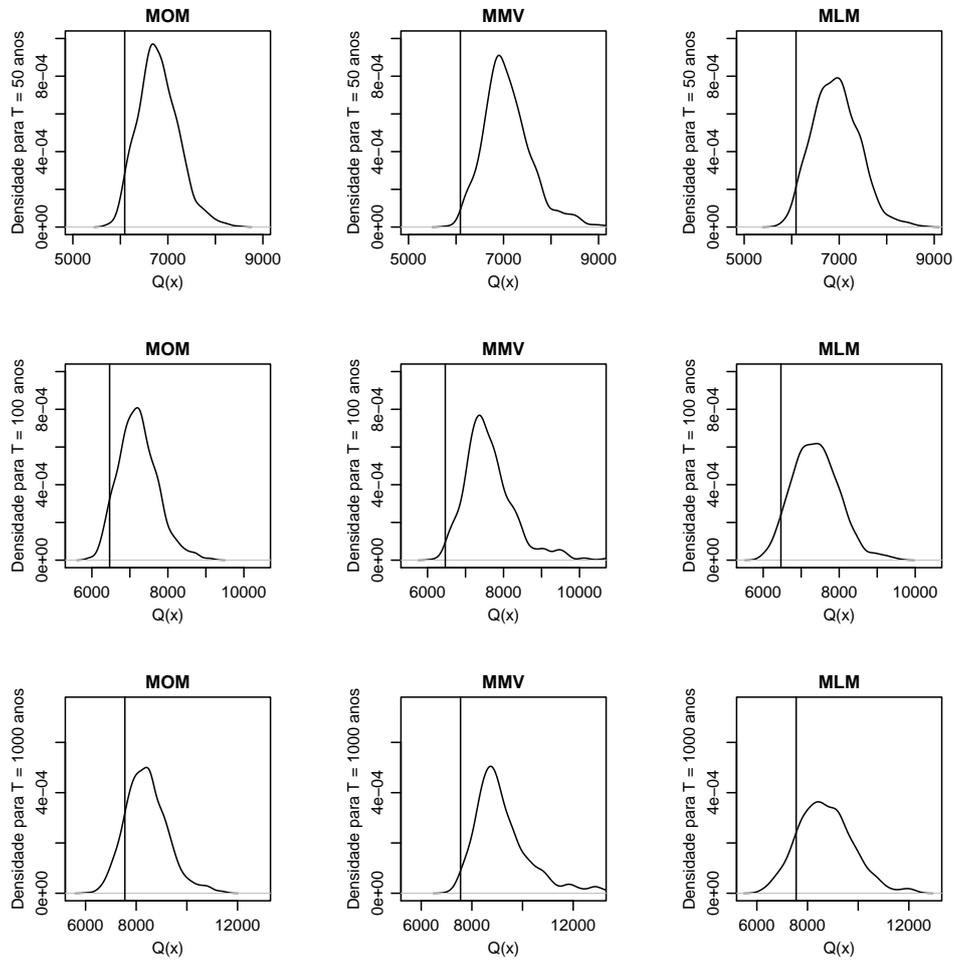
Tempo de retorno	Medida	Método		
		MOM	MMV	MLM
Q50	Viés	-66,67	-57,05	-22,72
	Variância	$2,48 \times 10^5$	$1,96 \times 10^5$	$2,25 \times 10^5$
	EQM	$2,53 \times 10^5$	2×10^5	$2,26 \times 10^5$
Q100	Viés	-77,92	-66,33	-24,74
	Variância	$3,34 \times 10^5$	$2,59 \times 10^5$	$3,01 \times 10^5$
	EQM	$3,41 \times 10^5$	$2,64 \times 10^5$	$3,01 \times 10^5$
Q1000	Viés	-115,12	-97	-31,43
	Variância	$7,16 \times 10^5$	$5,34 \times 10^5$	$6,33 \times 10^5$
	EQM	$7,3 \times 10^5$	$5,44 \times 10^5$	$6,34 \times 10^5$

Gumbel – Gumbel, n = 50, com contaminação



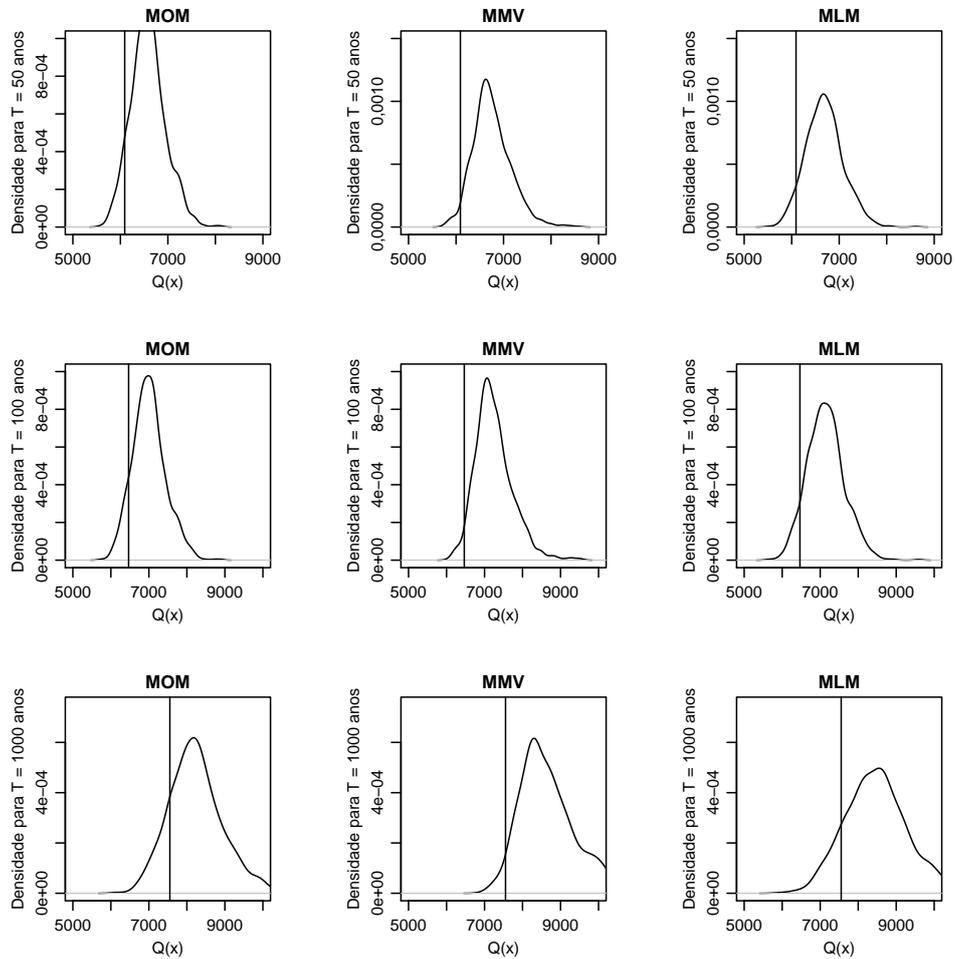
Tempo de retorno	Medida	Método		
		MOM	MMV	MLM
Q50	Viés	682,01	637,93	771,87
	Variância	$1,92 \times 10^5$	$1,85 \times 10^5$	$1,86 \times 10^5$
	EQM	$6,58 \times 10^5$	$5,92 \times 10^5$	$7,82 \times 10^5$
Q100	Viés	776,64	725,42	885,36
	Variância	$2,59 \times 10^5$	$2,45 \times 10^5$	$2,49 \times 10^5$
	EQM	$8,62 \times 10^5$	$7,71 \times 10^5$	$1,03 \times 10^6$
Q1000	Viés	1089,32	1014,52	1260,4
	Variância	$5,59 \times 10^5$	$5,07 \times 10^5$	$5,28 \times 10^5$
	EQM	$1,75 \times 10^6$	$1,54 \times 10^6$	$2,12 \times 10^6$

Pearson tipo III – Pearson tipo III, n = 30, com contaminação



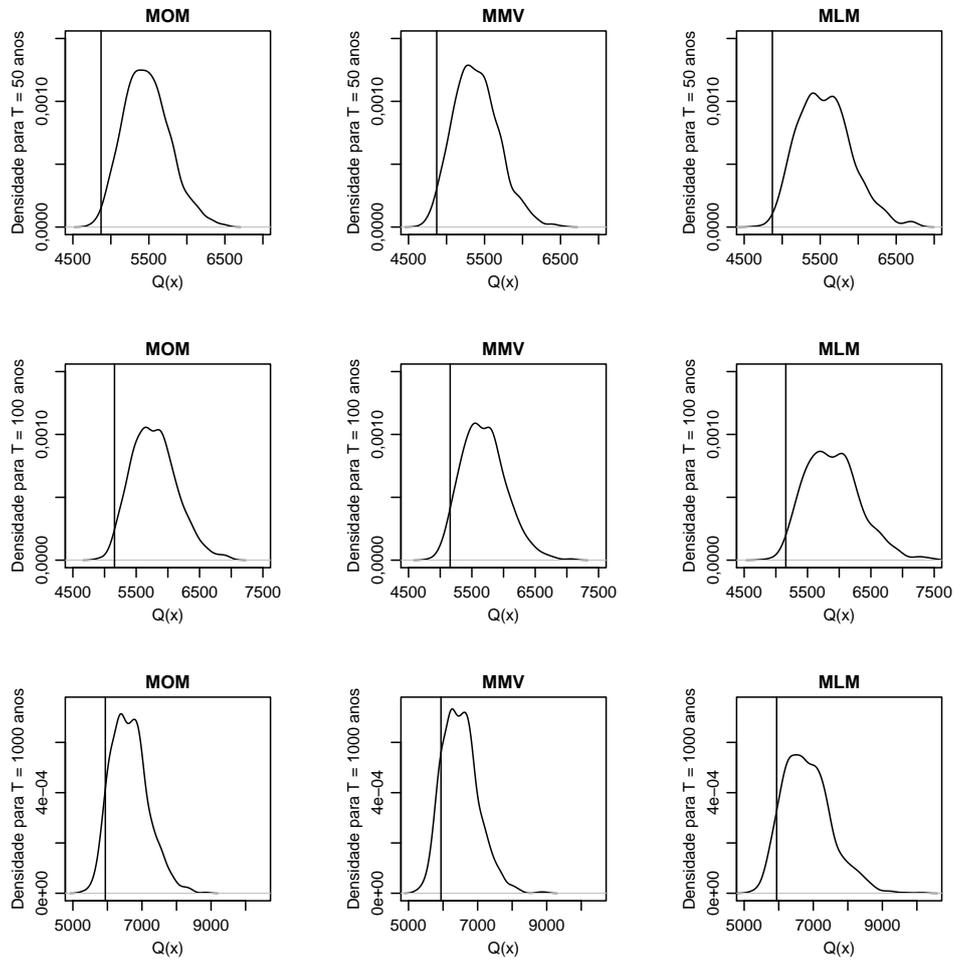
Tempo de retorno	Medida	Método		
		MOM	MMV	MLM
Q50	Viés	709,2	1020,75	842,51
	Variância	$1,82 \times 10^5$	$3,29 \times 10^5$	$2,34 \times 10^5$
	EQM	$6,85 \times 10^5$	$1,37 \times 10^6$	$9,44 \times 10^5$
Q100	Viés	749,64	1181,15	919,89
	Variância	$2,62 \times 10^5$	$5,28 \times 10^5$	$3,68 \times 10^5$
	EQM	$8,24 \times 10^5$	$1,92 \times 10^6$	$1,21 \times 10^6$
Q1000	Viés	857,95	1712,32	1155,01
	Variância	$6,81 \times 10^5$	$1,65 \times 10^6$	$1,13 \times 10^6$
	EQM	$1,42 \times 10^6$	$4,58 \times 10^6$	$2,46 \times 10^6$

Pearson tipo III – Pearson tipo III, n = 50, com contaminação



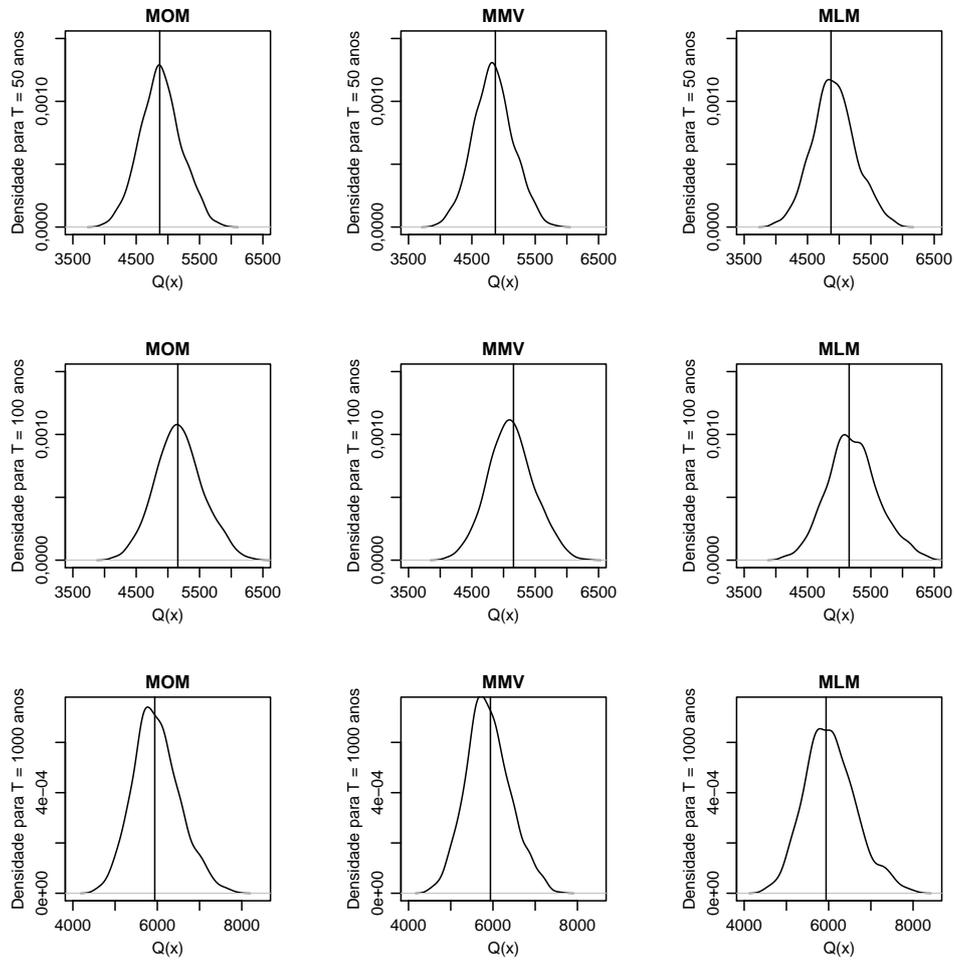
Tempo de retorno	Medida	Método		
		MOM	MMV	MLM
Q50	Viés	491,47	665,46	598,03
	Variância	$1,32 \times 10^5$	$1,61 \times 10^5$	$1,56 \times 10^5$
	EQM	$3,74 \times 10^5$	$6,04 \times 10^5$	$5,13 \times 10^5$
Q100	Viés	538,97	781,2	677,72
	Variância	$1,92 \times 10^5$	$2,43 \times 10^5$	$2,39 \times 10^5$
	EQM	$4,82 \times 10^5$	$8,53 \times 10^5$	$6,98 \times 10^5$
Q1000	Viés	682,66	1165,94	932,23
	Variância	$5,09 \times 10^5$	$6,93 \times 10^5$	$7,05 \times 10^5$
	EQM	$9,75 \times 10^5$	$2,05 \times 10^6$	$1,57 \times 10^6$

Weibull – Weibull, n = 30, com contaminação



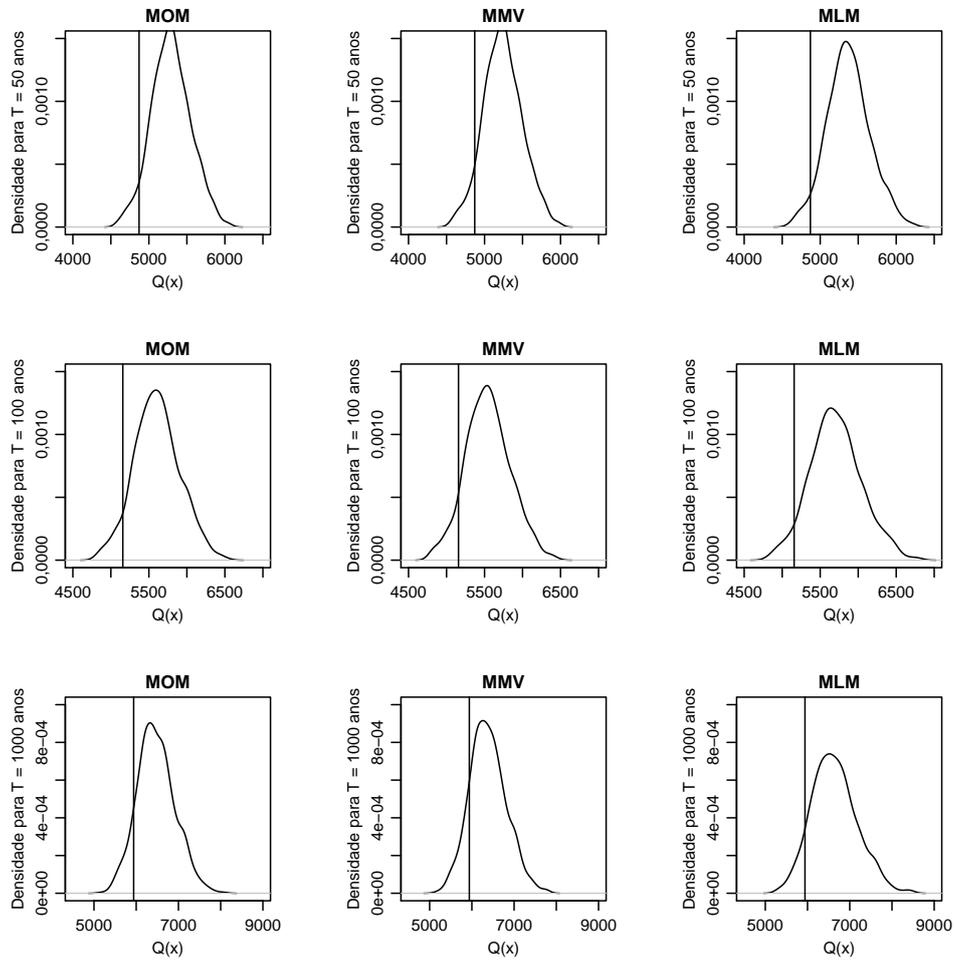
Tempo de retorno	Medida	Método		
		MOM	MMV	MLM
Q50	Viés	603,77	523,33	697,13
	Variância	$9,12 \times 10^4$	$8,82 \times 10^4$	$1,23 \times 10^5$
	EQM	$4,56 \times 10^5$	$3,62 \times 10^5$	$6,09 \times 10^5$
Q100	Viés	630,55	535,92	746,42
	Variância	$1,26 \times 10^5$	$1,21 \times 10^5$	$1,85 \times 10^5$
	EQM	$5,23 \times 10^5$	$4,09 \times 10^5$	$7,42 \times 10^5$
Q1000	Viés	700,58	565,14	888,04
	Variância	$2,84 \times 10^5$	$2,72 \times 10^5$	$4,89 \times 10^5$
	EQM	$7,74 \times 10^5$	$5,91 \times 10^5$	$1,28 \times 10^6$

Weibull – Weibull, n = 50



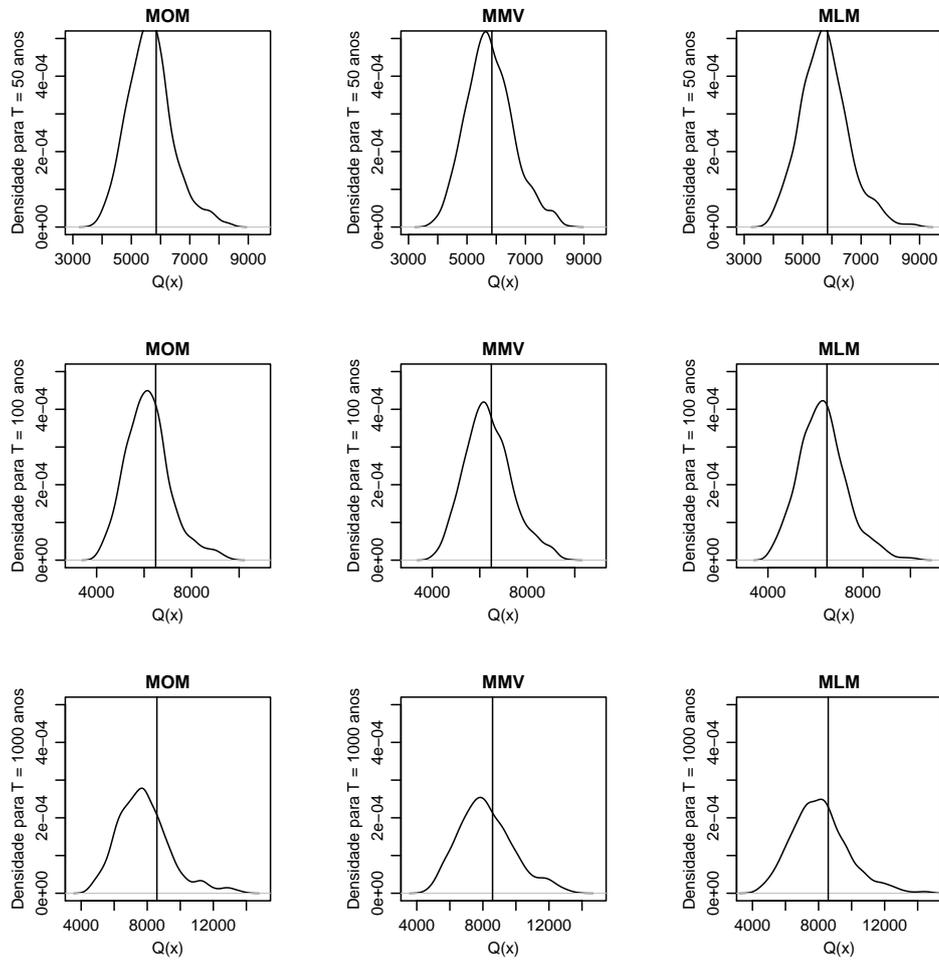
Tempo de retorno	Medida	Método		
		MOM	MMV	MLM
Q50	Viés	17,12	-31,72	48,41
	Variância	$1,05 \times 10^5$	$9,93 \times 10^4$	$1,19 \times 10^5$
	EQM	$1,05 \times 10^5$	10^5	$1,22 \times 10^5$
Q100	Viés	22,12	-37,6	60,81
	Variância	$1,42 \times 10^5$	$1,31 \times 10^5$	$1,64 \times 10^5$
	EQM	$1,42 \times 10^5$	$1,33 \times 10^5$	$1,68 \times 10^5$
Q1000	Viés	41,24	-51,78	103,33
	Variância	$3,05 \times 10^5$	$2,7 \times 10^5$	$3,69 \times 10^5$
	EQM	$3,06 \times 10^5$	$2,73 \times 10^5$	$3,79 \times 10^5$

Weibull – Weibull, n = 50, com contaminação



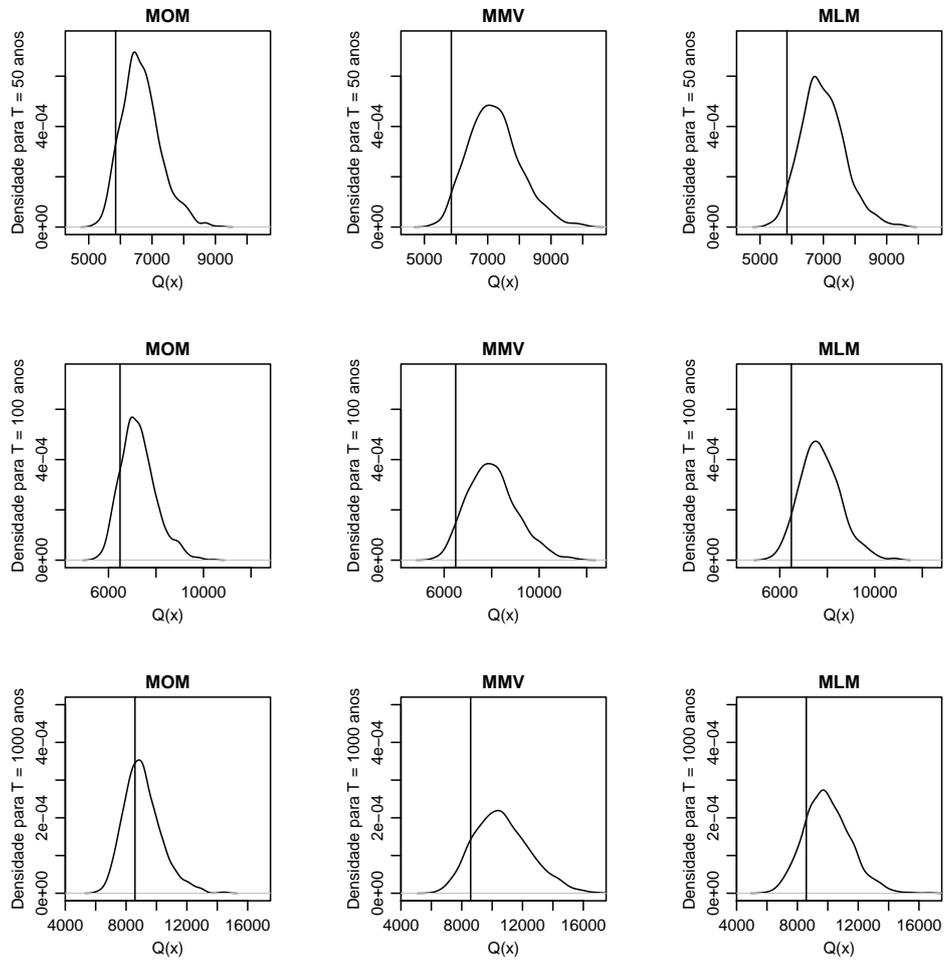
Tempo de retorno	Medida	Método		
		MOM	MMV	MLM
Q50	Viés	421,06	369,95	499,72
	Variância	$6,66 \times 10^4$	$6,4 \times 10^4$	$7,98 \times 10^4$
	EQM	$2,44 \times 10^5$	$2,01 \times 10^5$	$3,3 \times 10^5$
Q100	Viés	450,12	389,02	549,23
	Variância	$9,05 \times 10^4$	$8,65 \times 10^4$	$1,15 \times 10^5$
	EQM	$2,93 \times 10^5$	$2,38 \times 10^5$	$4,17 \times 10^5$
Q1000	Viés	531,95	441,35	695,73
	Variância	$2,01 \times 10^5$	$1,89 \times 10^5$	$2,88 \times 10^5$
	EQM	$4,84 \times 10^5$	$3,84 \times 10^5$	$7,72 \times 10^5$

Gumbel – Pearson tipo III, n = 30



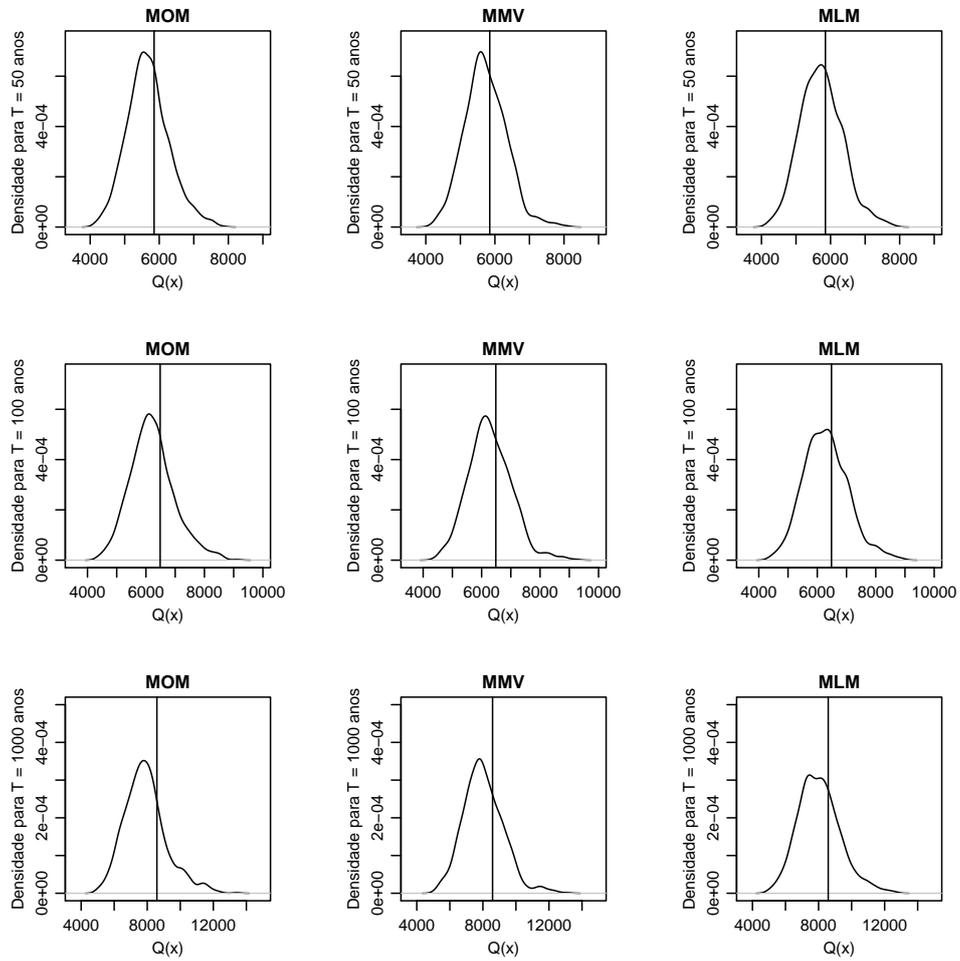
Tempo de retorno	Medida	Método		
		MOM	MMV	MLM
Q50	Viés	-205,27	-59,76	-95,47
	Variância	$5,99 \times 10^5$	$6,53 \times 10^5$	$6,65 \times 10^5$
	EQM	$6,41 \times 10^5$	$6,56 \times 10^5$	$6,74 \times 10^5$
Q100	Viés	-316,94	-113,62	-168,11
	Variância	$9,06 \times 10^5$	$9,89 \times 10^5$	$1,01 \times 10^6$
	EQM	$1,01 \times 10^6$	10^6	$1,04 \times 10^6$
Q1000	Viés	-780,64	-370,44	-493,51
	Variância	$2,53 \times 10^6$	$2,73 \times 10^6$	$2,85 \times 10^6$
	EQM	$3,14 \times 10^6$	$2,87 \times 10^6$	$3,09 \times 10^6$

Gumbel – Pearson tipo III, $n = 30$, com contaminação



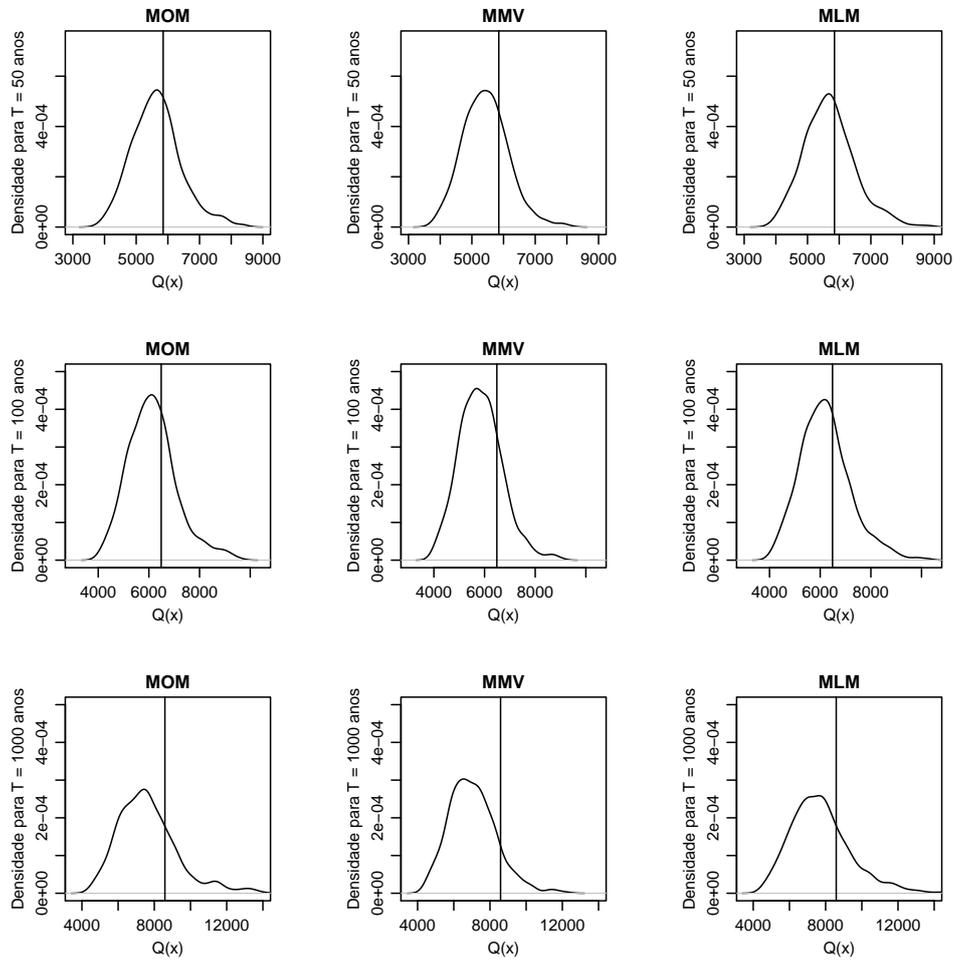
Tempo de retorno	Medida	Método		
		MOM	MMV	MLM
Q50	Viés	810,49	1371,35	1139,76
	Variância	$3,77 \times 10^5$	$6,66 \times 10^5$	$4,66 \times 10^5$
	EQM	$1,03 \times 10^6$	$2,55 \times 10^6$	$1,77 \times 10^6$
Q100	Viés	771,82	1552,95	1223,28
	Variância	$5,58 \times 10^5$	$1,07 \times 10^6$	$7,36 \times 10^5$
	EQM	$1,15 \times 10^6$	$3,48 \times 10^6$	$2,23 \times 10^6$
Q1000	Viés	501,07	2071,74	1387,32
	Variância	$1,5 \times 10^6$	$3,31 \times 10^6$	$2,25 \times 10^6$
	EQM	$1,75 \times 10^6$	$7,61 \times 10^6$	$4,18 \times 10^6$

Gumbel – Pearson tipo III, n = 50



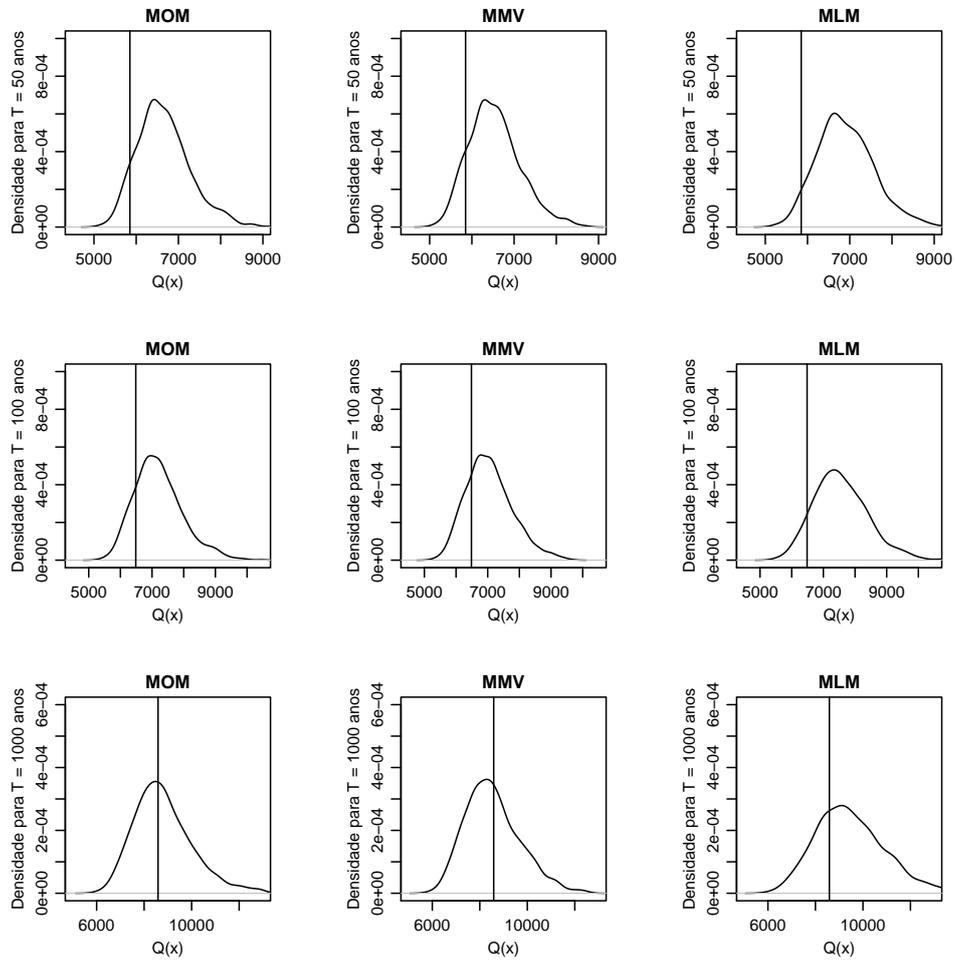
Tempo de retorno	Medida	Método		
		MOM	MMV	MLM
Q50	Viés	-170,2	-121,69	-112,37
	Variância	$3,71 \times 10^5$	$3,63 \times 10^5$	$3,76 \times 10^5$
	EQM	4×10^5	$3,77 \times 10^5$	$3,89 \times 10^5$
Q100	Viés	-269,87	-201,87	-191,7
	Variância	$5,7 \times 10^5$	$5,42 \times 10^5$	$5,73 \times 10^5$
	EQM	$6,43 \times 10^5$	$5,83 \times 10^5$	$6,1 \times 10^5$
Q1000	Viés	-691,85	-554,27	-541,93
	Variância	$1,64 \times 10^6$	$1,46 \times 10^6$	$1,61 \times 10^6$
	EQM	$2,12 \times 10^6$	$1,77 \times 10^6$	$1,9 \times 10^6$

Gumbel – Weibull, n = 30



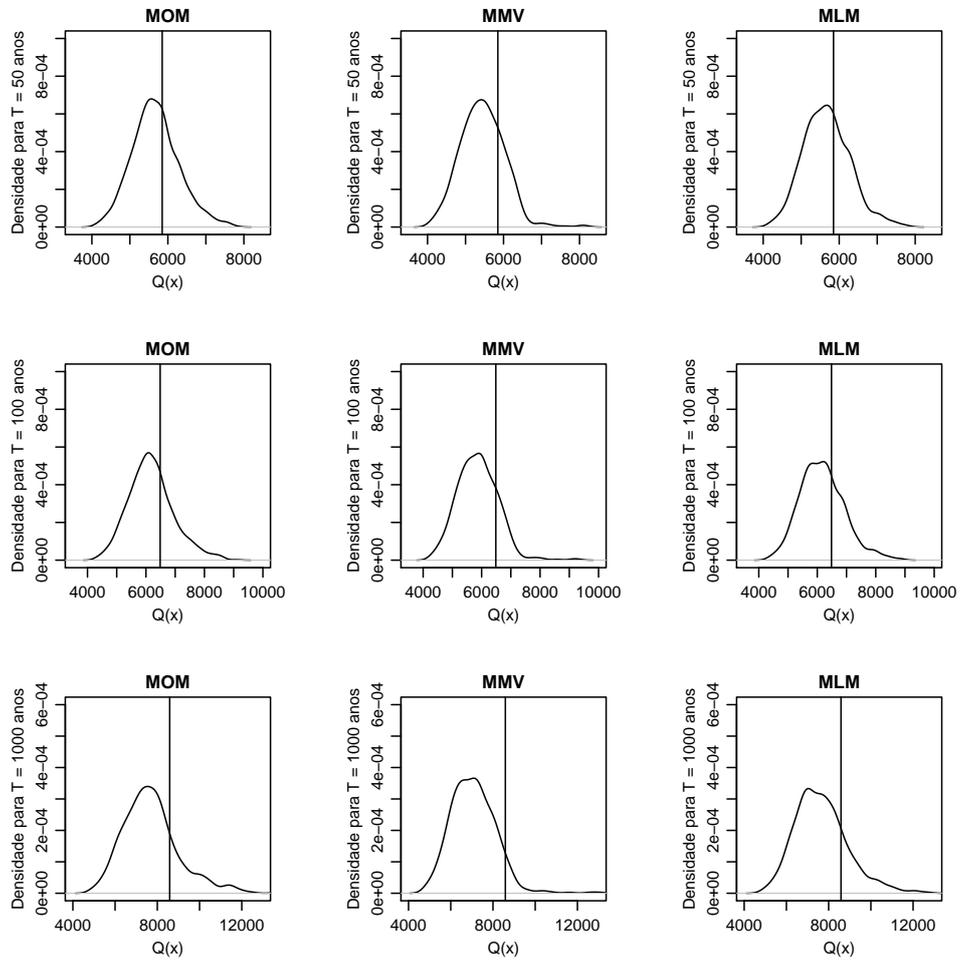
Tempo de retorno	Medida	Método		
		MOM	MMV	MLM
Q50	Viés	-205,9	-414,78	-154,65
	Variância	$6,19 \times 10^5$	$5,17 \times 10^5$	$6,65 \times 10^5$
	EQM	$6,62 \times 10^5$	$6,89 \times 10^5$	$6,89 \times 10^5$
Q100	Viés	-346,12	-636,98	-279,97
	Variância	$9,43 \times 10^5$	$7,28 \times 10^5$	$1,01 \times 10^6$
	EQM	$1,06 \times 10^6$	$1,13 \times 10^6$	$1,09 \times 10^6$
Q1000	Viés	-953,11	-1533,64	-838,47
	Variância	$2,69 \times 10^6$	$1,7 \times 10^6$	$2,84 \times 10^6$
	EQM	$3,6 \times 10^6$	$4,05 \times 10^6$	$3,54 \times 10^6$

Gumbel – Weibull, n = 30, com contaminação



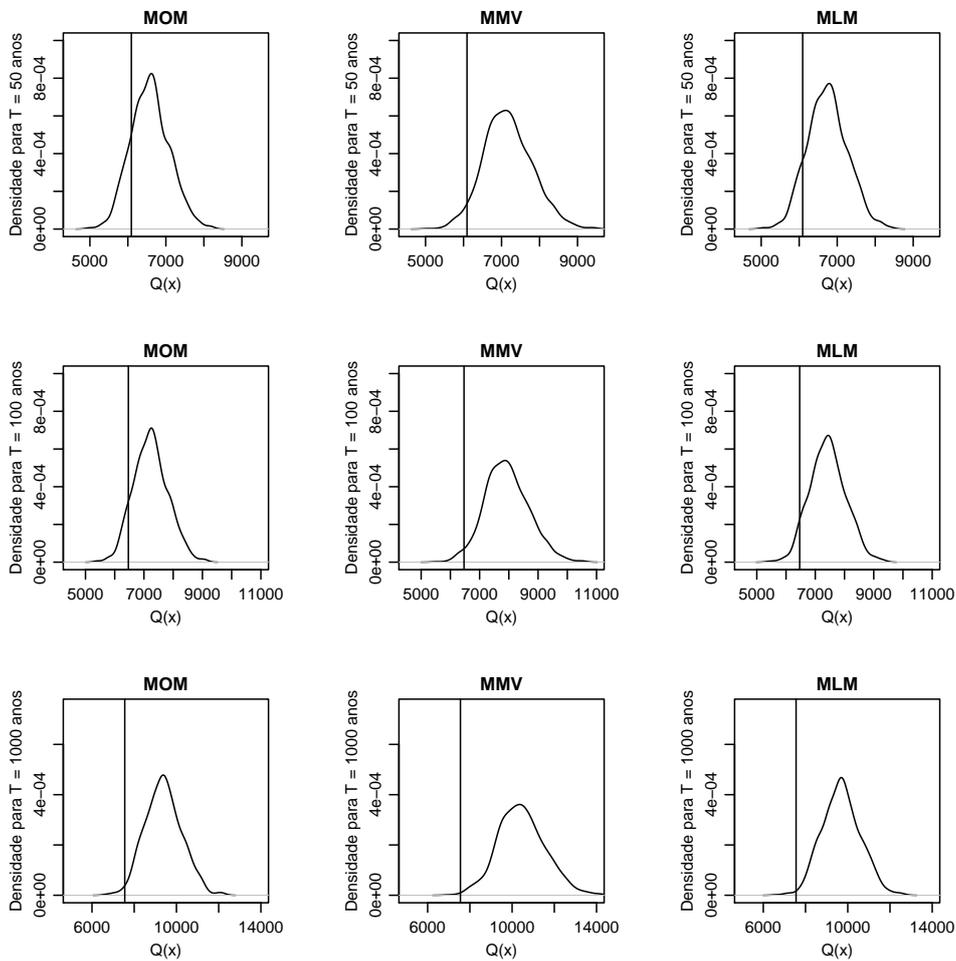
Tempo de retorno	Medida	Método		
		MOM	MMV	MLM
Q50	Viés	800,6	686,7	1058,46
	Variância	$3,96 \times 10^5$	$3,72 \times 10^5$	$4,68 \times 10^5$
	EQM	$1,04 \times 10^6$	$8,44 \times 10^5$	$1,59 \times 10^6$
Q100	Viés	716,43	566,41	1069,1
	Variância	$5,86 \times 10^5$	$5,36 \times 10^5$	$7,37 \times 10^5$
	EQM	$1,1 \times 10^6$	$8,56 \times 10^5$	$1,88 \times 10^6$
Q1000	Viés	226,04	-42,68	909,7
	Variância	$1,55 \times 10^6$	$1,32 \times 10^6$	$2,24 \times 10^6$
	EQM	$1,6 \times 10^6$	$1,32 \times 10^6$	$3,07 \times 10^6$

Gumbel – Weibull, n = 50



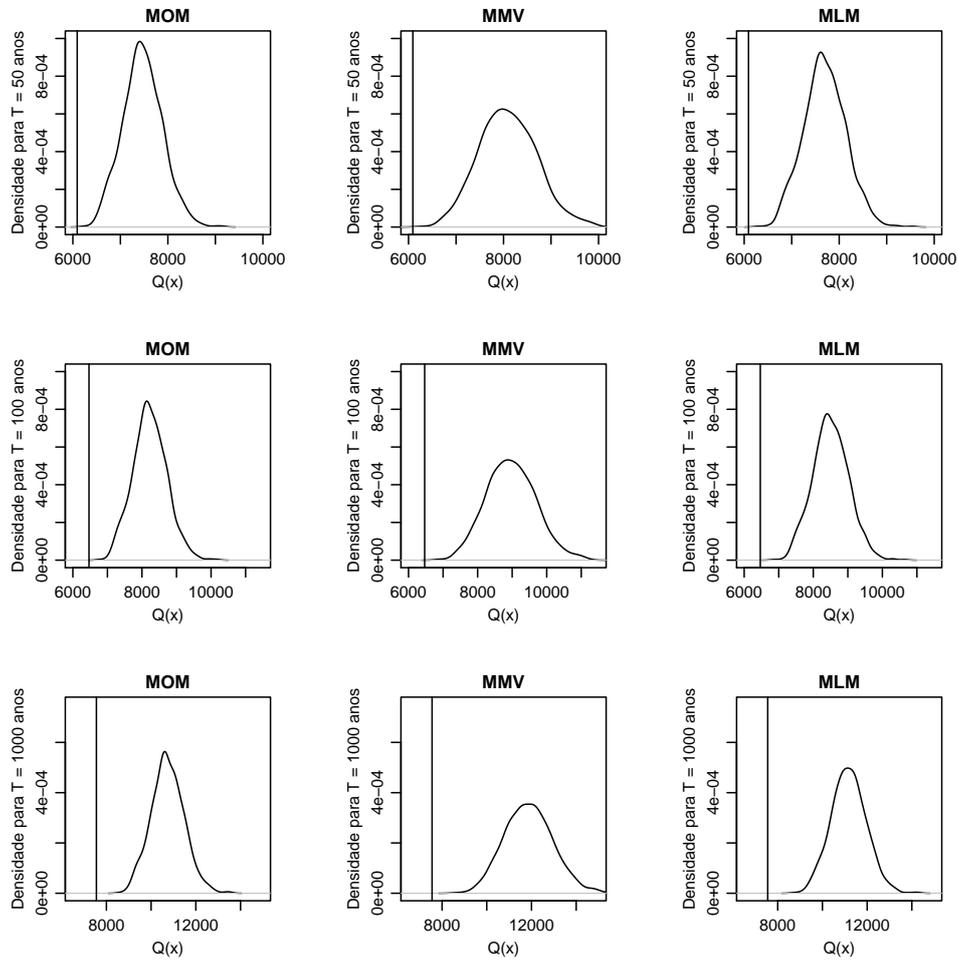
Tempo de retorno	Medida	Método		
		MOM	MMV	MLM
Q50	Viés	-171,5	-403,62	-177,31
	Variância	$3,85 \times 10^5$	$3,3 \times 10^5$	$3,77 \times 10^5$
	EQM	$4,15 \times 10^5$	$4,93 \times 10^5$	$4,09 \times 10^5$
Q100	Viés	-302,83	-623,76	-314,94
	Variância	$5,98 \times 10^5$	$4,7 \times 10^5$	$5,72 \times 10^5$
	EQM	$6,9 \times 10^5$	$8,59 \times 10^5$	$6,71 \times 10^5$
Q1000	Viés	-884,94	-1517,96	-924,06
	Variância	$1,76 \times 10^6$	$1,12 \times 10^6$	$1,58 \times 10^6$
	EQM	$2,55 \times 10^6$	$3,42 \times 10^6$	$2,44 \times 10^6$

Pearson tipo III – Gumbel, n = 30



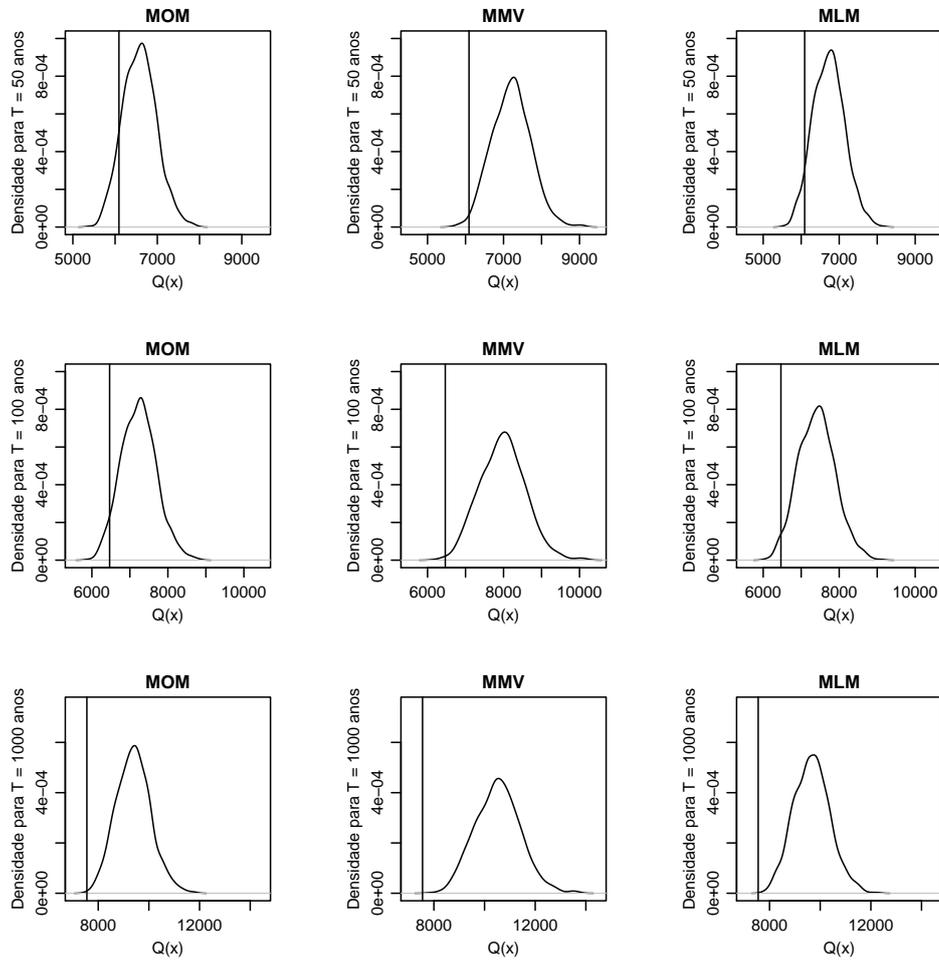
Tempo de retorno	Medida	Método		
		MOM	MMV	MLM
Q50	Viés	492,37	1080,31	653,85
	Variância	$2,58 \times 10^5$	$4,09 \times 10^5$	$2,79 \times 10^5$
	EQM	$5,01 \times 10^5$	$1,58 \times 10^6$	$7,06 \times 10^5$
Q100	Viés	767,53	1467,87	962,91
	Variância	$3,46 \times 10^5$	$5,58 \times 10^5$	$3,76 \times 10^5$
	EQM	$9,35 \times 10^5$	$2,71 \times 10^6$	$1,3 \times 10^6$
Q1000	Viés	1831,73	2903,48	2139,17
	Variância	$7,47 \times 10^5$	$1,23 \times 10^6$	$8,2 \times 10^5$
	EQM	$4,1 \times 10^6$	$9,66 \times 10^6$	$5,4 \times 10^6$

Pearson tipo III – Gumbel, $n = 30$, com contaminação



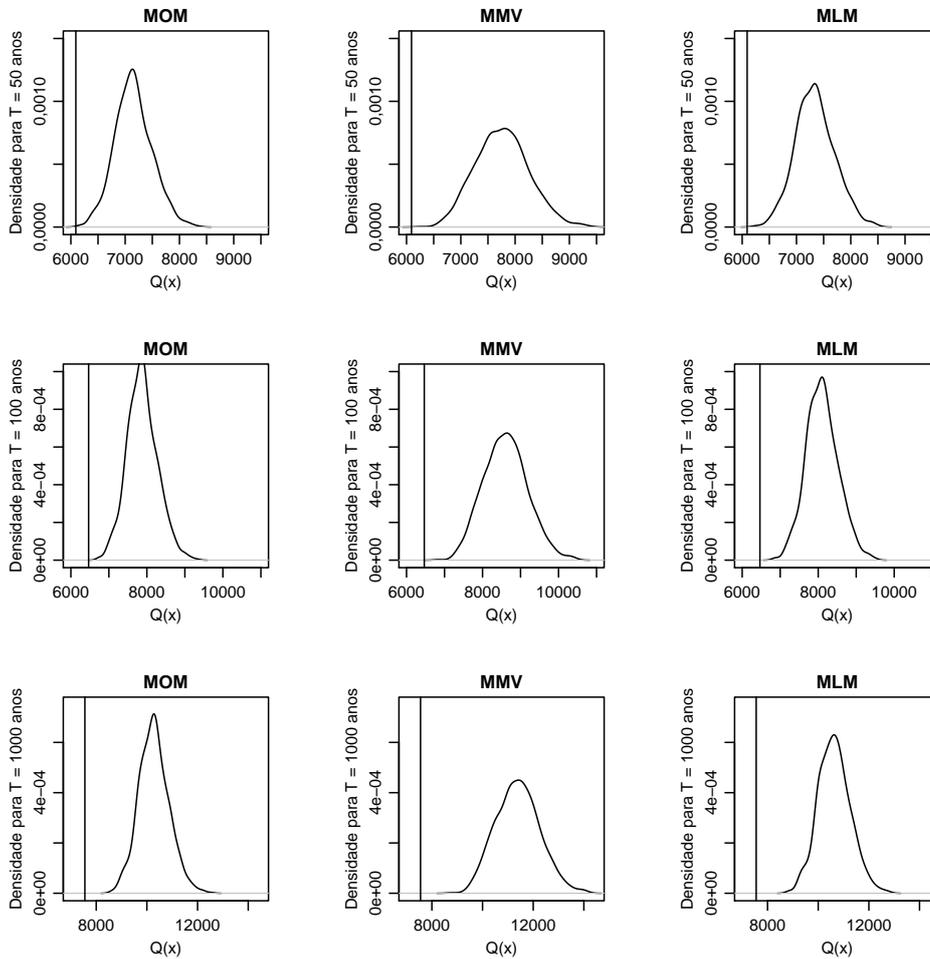
Tempo de retorno	Medida	Método		
		MOM	MMV	MLM
Q50	Viés	1377,22	1998,56	1605,18
	Variância	$1,76 \times 10^5$	$3,69 \times 10^5$	$1,96 \times 10^5$
	EQM	$2,07 \times 10^6$	$4,36 \times 10^6$	$2,77 \times 10^6$
Q100	Viés	1761,3	2502,41	2037,13
	Variância	$2,41 \times 10^5$	$5,09 \times 10^5$	$2,7 \times 10^5$
	EQM	$3,34 \times 10^6$	$6,77 \times 10^6$	$4,42 \times 10^6$
Q1000	Viés	3185,43	4322,33	3619,44
	Variância	$5,47 \times 10^5$	$1,15 \times 10^6$	$6,2 \times 10^5$
	EQM	$1,07 \times 10^7$	$1,98 \times 10^7$	$1,37 \times 10^7$

Pearson tipo III – Gumbel, n = 50



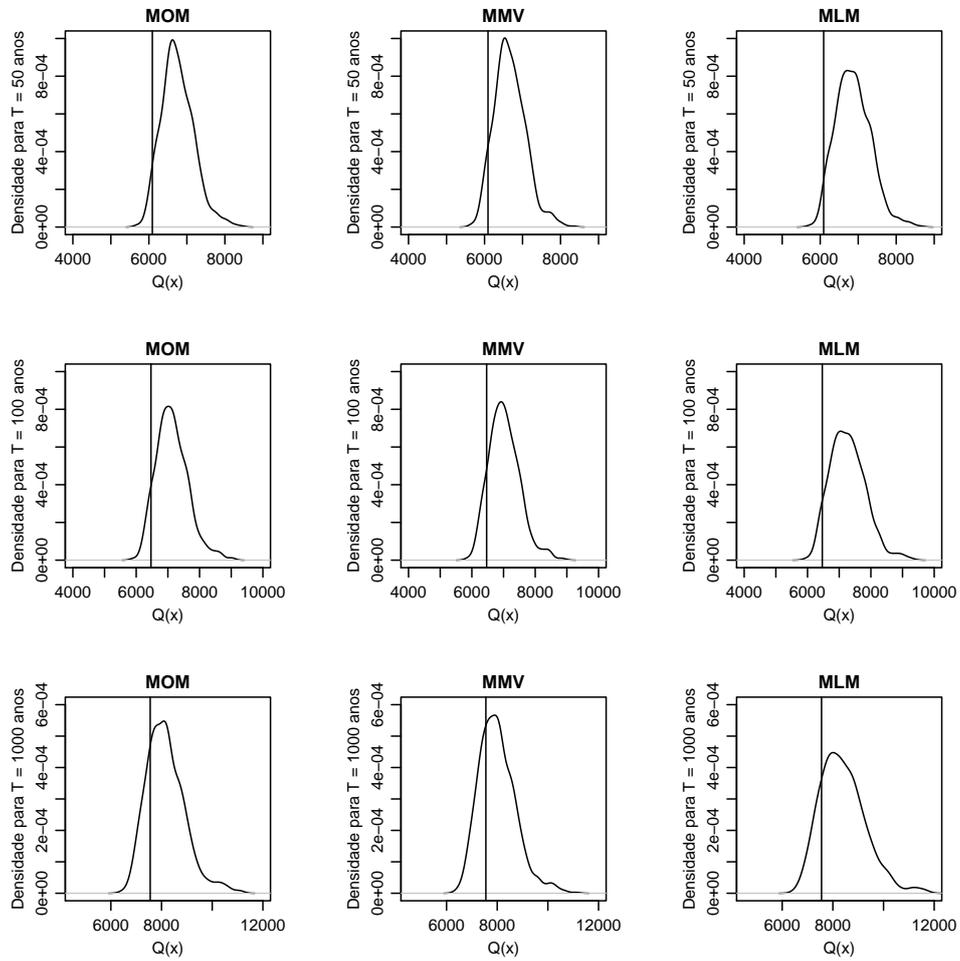
Tempo de retorno	Medida	Método		
		MOM	MMV	MLM
Q50	Viés	490,75	1113,95	637,59
	Variância	$1,59 \times 10^5$	$2,61 \times 10^5$	$1,71 \times 10^5$
	EQM	$3,99 \times 10^5$	$1,5 \times 10^6$	$5,77 \times 10^5$
Q100	Viés	767,38	1509,75	945,06
	Variância	$2,13 \times 10^5$	$3,57 \times 10^5$	$2,31 \times 10^5$
	EQM	$8,02 \times 10^5$	$2,64 \times 10^6$	$1,12 \times 10^6$
Q1000	Viés	1836,47	2972,59	2116,05
	Variância	$4,62 \times 10^5$	$7,93 \times 10^5$	$5,05 \times 10^5$
	EQM	$3,83 \times 10^6$	$9,63 \times 10^6$	$4,98 \times 10^6$

Pearson tipo III – Gumbel, n = 50, com contaminação



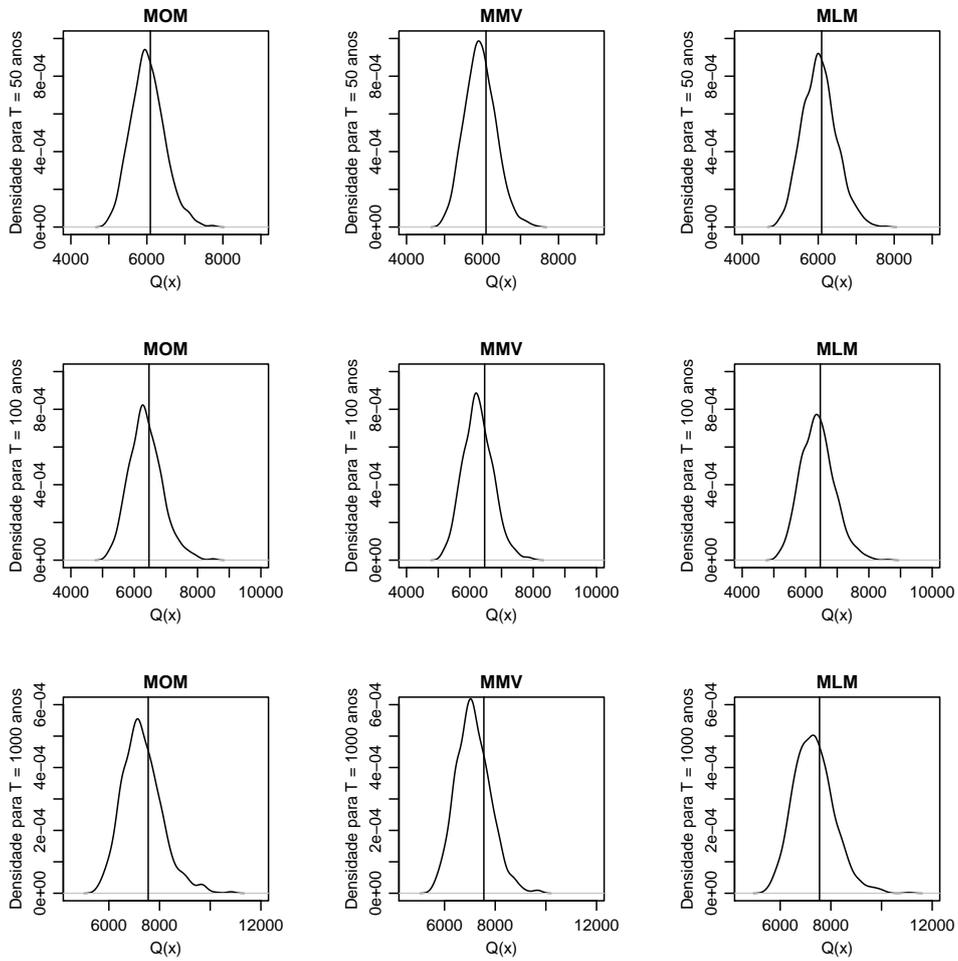
Tempo de retorno	Medida	Método		
		MOM	MMV	MLM
Q50	Viés	1062,33	1670,35	1249,36
	Variância	$1,19 \times 10^5$	$2,39 \times 10^5$	$1,32 \times 10^5$
	EQM	$1,25 \times 10^6$	$3,03 \times 10^6$	$1,69 \times 10^6$
Q100	Viés	1412,96	2137,93	1639,26
	Variância	$1,61 \times 10^5$	$3,28 \times 10^5$	$1,81 \times 10^5$
	EQM	$2,16 \times 10^6$	$4,9 \times 10^6$	$2,87 \times 10^6$
Q1000	Viés	2726,56	3837,98	3082,65
	Variância	$3,6 \times 10^5$	$7,37 \times 10^5$	$4,06 \times 10^5$
	EQM	$7,79 \times 10^6$	$1,55 \times 10^7$	$9,91 \times 10^6$

Pearson tipo III – Weibull, n = 30, com contaminação



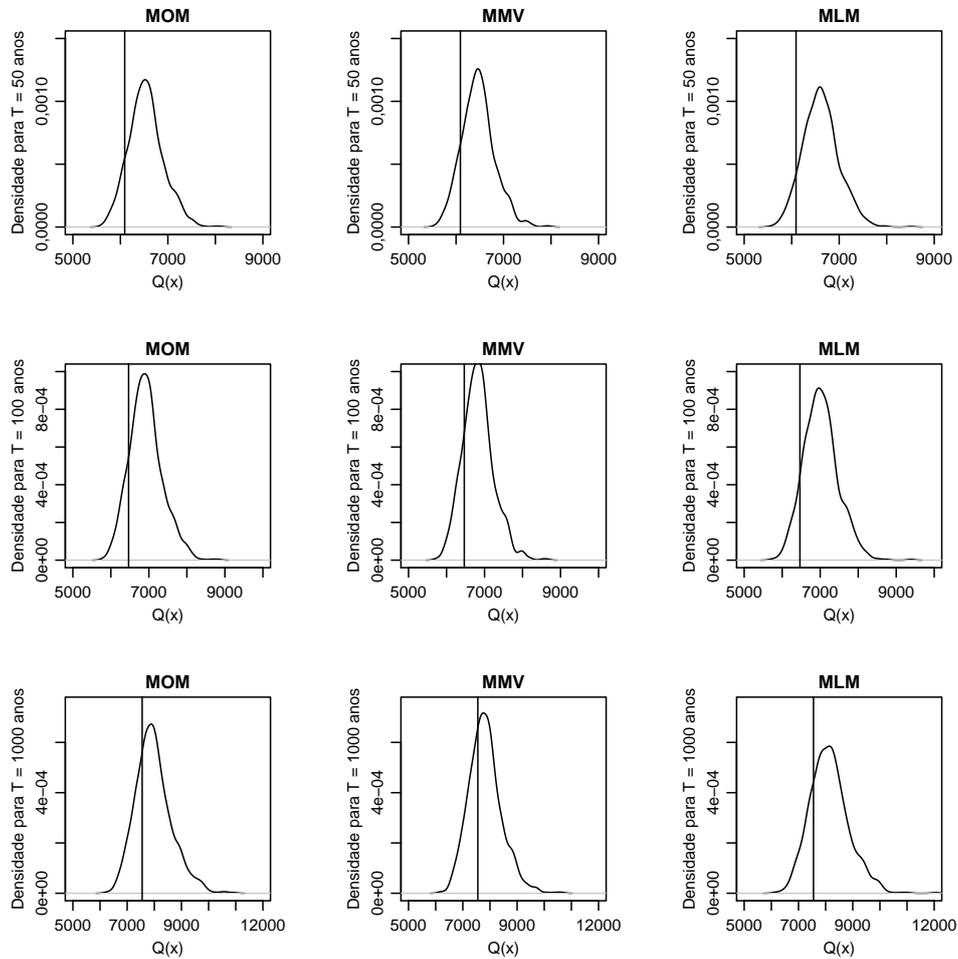
Tempo de retorno	Medida	Método		
		MOM	MMV	MLM
Q50	Viés	667,79	577,88	778,39
	Variância	$1,83 \times 10^5$	$1,71 \times 10^5$	$2,11 \times 10^5$
	EQM	$6,29 \times 10^5$	$5,05 \times 10^5$	$8,16 \times 10^5$
Q100	Viés	667,76	560,83	806,99
	Variância	$2,54 \times 10^5$	$2,35 \times 10^5$	$3,09 \times 10^5$
	EQM	$6,99 \times 10^5$	$5,5 \times 10^5$	$9,61 \times 10^5$
Q1000	Viés	613,24	456,01	843,93
	Variância	$5,77 \times 10^5$	$5,23 \times 10^5$	$7,91 \times 10^5$
	EQM	$9,53 \times 10^5$	$7,31 \times 10^5$	$1,5 \times 10^6$

Pearson tipo III – Weibull, n = 50



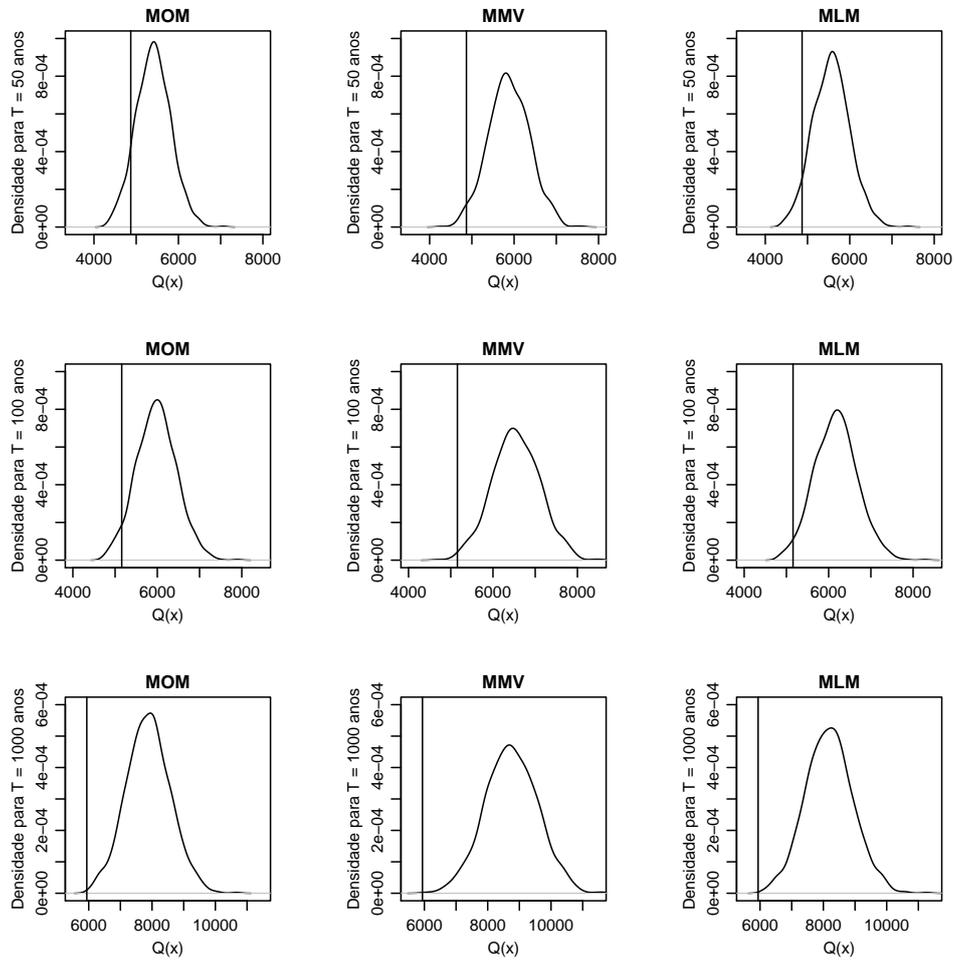
Tempo de retorno	Medida	Método		
		MOM	MMV	MLM
Q50	Viés	-65,14	-150,21	-54,44
	Variância	$1,96 \times 10^5$	$1,66 \times 10^5$	$1,99 \times 10^5$
	EQM	2×10^5	$1,88 \times 10^5$	$2,02 \times 10^5$
Q100	Viés	-99,46	-206,56	-86,57
	Variância	$2,73 \times 10^5$	$2,21 \times 10^5$	$2,76 \times 10^5$
	EQM	$2,83 \times 10^5$	$2,64 \times 10^5$	$2,84 \times 10^5$
Q1000	Viés	-239,09	-416,13	-219,68
	Variância	$6,18 \times 10^5$	$4,57 \times 10^5$	$6,18 \times 10^5$
	EQM	$6,75 \times 10^5$	$6,3 \times 10^5$	$6,67 \times 10^5$

Pearson tipo III – Weibull, n = 50, com contaminação



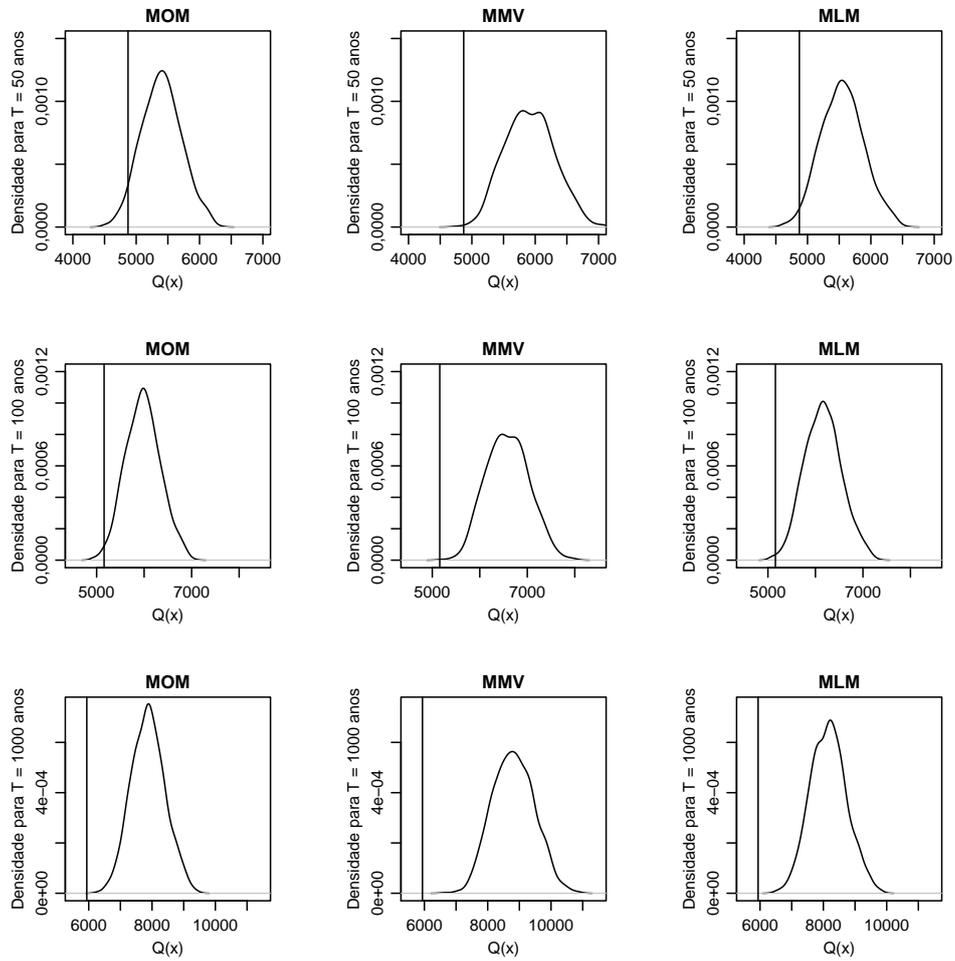
Tempo de retorno	Medida	Método		
		MOM	MMV	MLM
Q50	Viés	449,39	375,85	525,02
	Variância	$1,35 \times 10^5$	$1,22 \times 10^5$	$1,43 \times 10^5$
	EQM	$3,37 \times 10^5$	$2,63 \times 10^5$	$4,19 \times 10^5$
Q100	Viés	453,52	362,64	550,38
	Variância	$1,89 \times 10^5$	$1,67 \times 10^5$	$2,07 \times 10^5$
	EQM	$3,95 \times 10^5$	$2,99 \times 10^5$	$5,1 \times 10^5$
Q1000	Viés	418,31	273,64	582,94
	Variância	$4,43 \times 10^5$	$3,68 \times 10^5$	$5,11 \times 10^5$
	EQM	$6,18 \times 10^5$	$4,43 \times 10^5$	$8,51 \times 10^5$

Weibull – Gumbel, n = 30



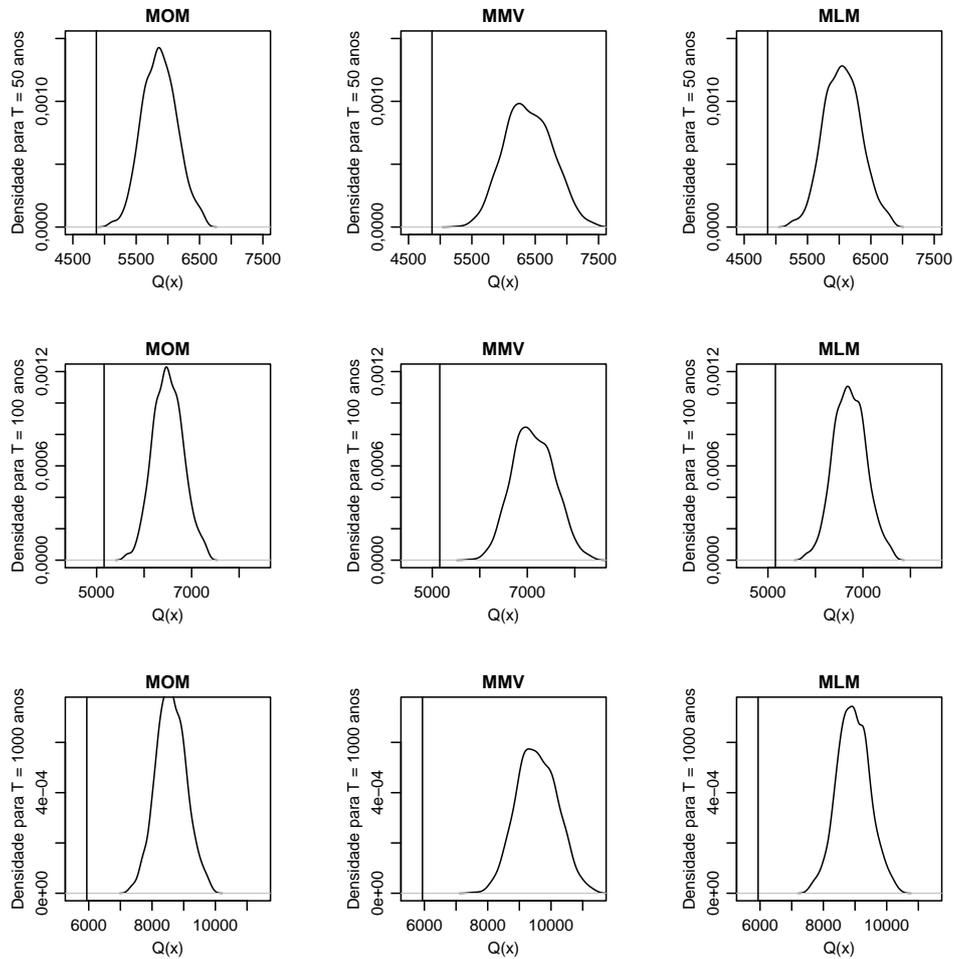
Tempo de retorno	Medida	Método		
		MOM	MMV	MLM
Q50	Viés	534,27	1007,7	692,76
	Variância	$1,67 \times 10^5$	$2,36 \times 10^5$	$1,89 \times 10^5$
	EQM	$4,53 \times 10^5$	$1,25 \times 10^6$	$6,69 \times 10^5$
Q100	Viés	823,35	1387,8	1015,12
	Variância	$2,24 \times 10^5$	$3,17 \times 10^5$	$2,55 \times 10^5$
	EQM	$9,02 \times 10^5$	$2,24 \times 10^6$	$1,29 \times 10^6$
Q1000	Viés	1941,64	2806,83	2243,39
	Variância	$4,84 \times 10^5$	$6,84 \times 10^5$	$5,58 \times 10^5$
	EQM	$4,25 \times 10^6$	$8,56 \times 10^6$	$5,59 \times 10^6$

Weibull – Gumbel, n = 50



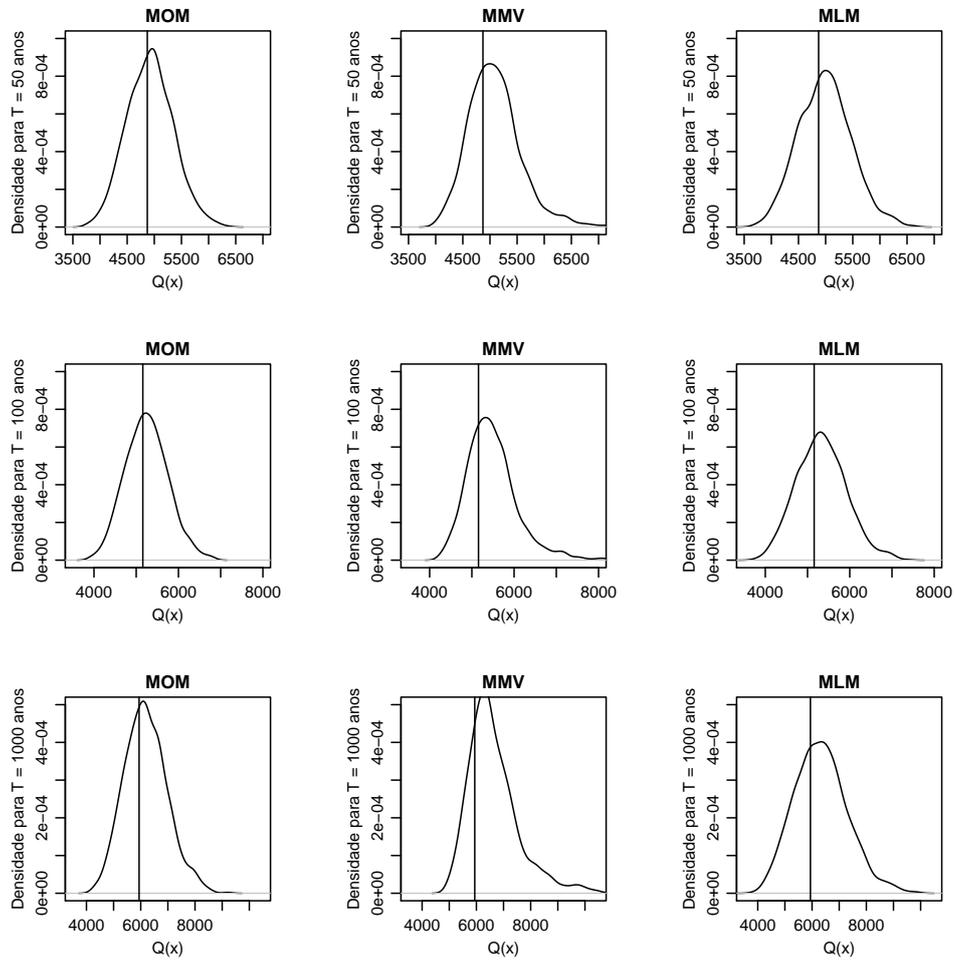
Tempo de retorno	Medida	Método		
		MOM	MMV	MLM
Q50	Viés	530,3	1042,41	680,5
	Variância	10^5	$1,51 \times 10^5$	$1,12 \times 10^5$
	EQM	$3,81 \times 10^5$	$1,24 \times 10^6$	$5,75 \times 10^5$
Q100	Viés	818,92	1429,64	1000,68
	Variância	$1,33 \times 10^5$	$2,02 \times 10^5$	$1,5 \times 10^5$
	EQM	$8,04 \times 10^5$	$2,25 \times 10^6$	$1,15 \times 10^6$
Q1000	Viés	1935,72	2872,24	2221,7
	Variância	$2,84 \times 10^5$	$4,3 \times 10^5$	$3,25 \times 10^5$
	EQM	$4,03 \times 10^6$	$8,68 \times 10^6$	$5,26 \times 10^6$

Weibull – Gumbel, n = 50, com contaminação



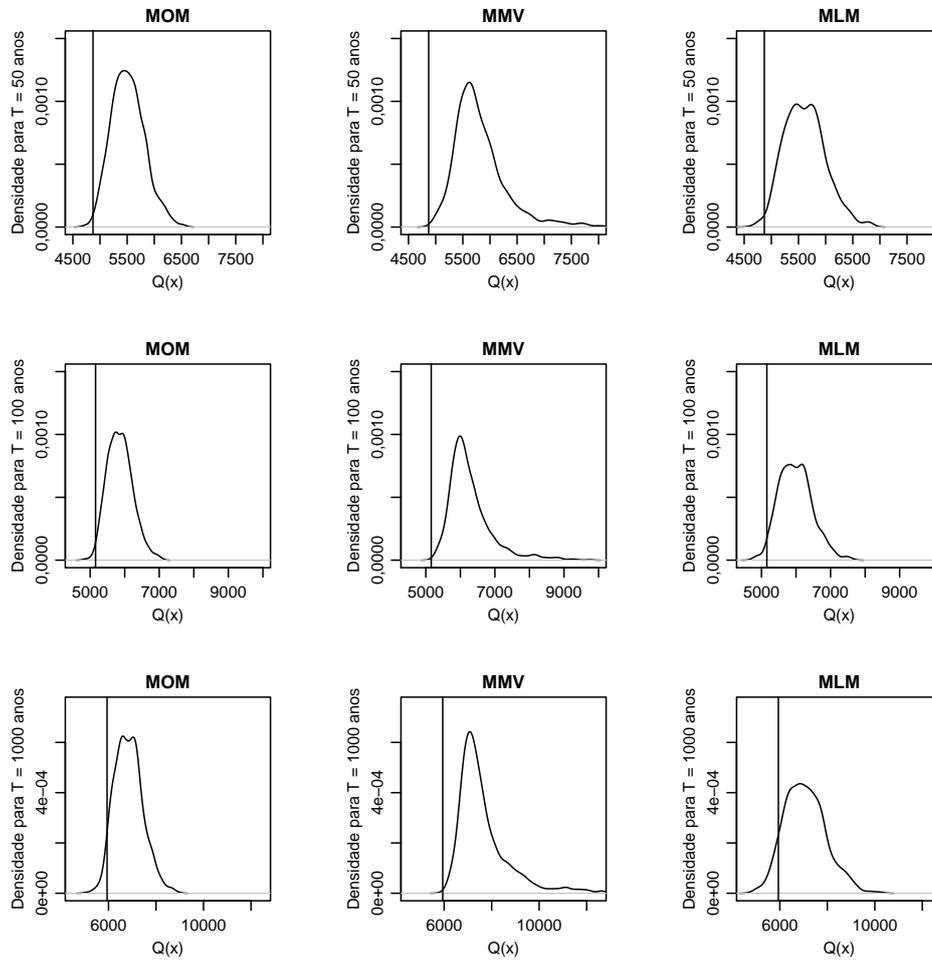
Tempo de retorno	Medida	Método		
		MOM	MMV	MLM
Q50	Viés	1000,45	1523,6	1184,26
	Variância	$7,32 \times 10^4$	$1,4 \times 10^5$	$8,43 \times 10^4$
	EQM	$1,07 \times 10^6$	$2,46 \times 10^6$	$1,49 \times 10^6$
Q100	Viés	1347,78	1972,2	1570,19
	Variância	$9,85 \times 10^4$	$1,88 \times 10^5$	$1,15 \times 10^5$
	EQM	$1,92 \times 10^6$	$4,08 \times 10^6$	$2,58 \times 10^6$
Q1000	Viés	2658,55	3617,6	3008,49
	Variância	$2,17 \times 10^5$	$4,04 \times 10^5$	$2,57 \times 10^5$
	EQM	$7,29 \times 10^6$	$1,35 \times 10^7$	$9,31 \times 10^6$

Weibull – Pearson tipo III, n = 30



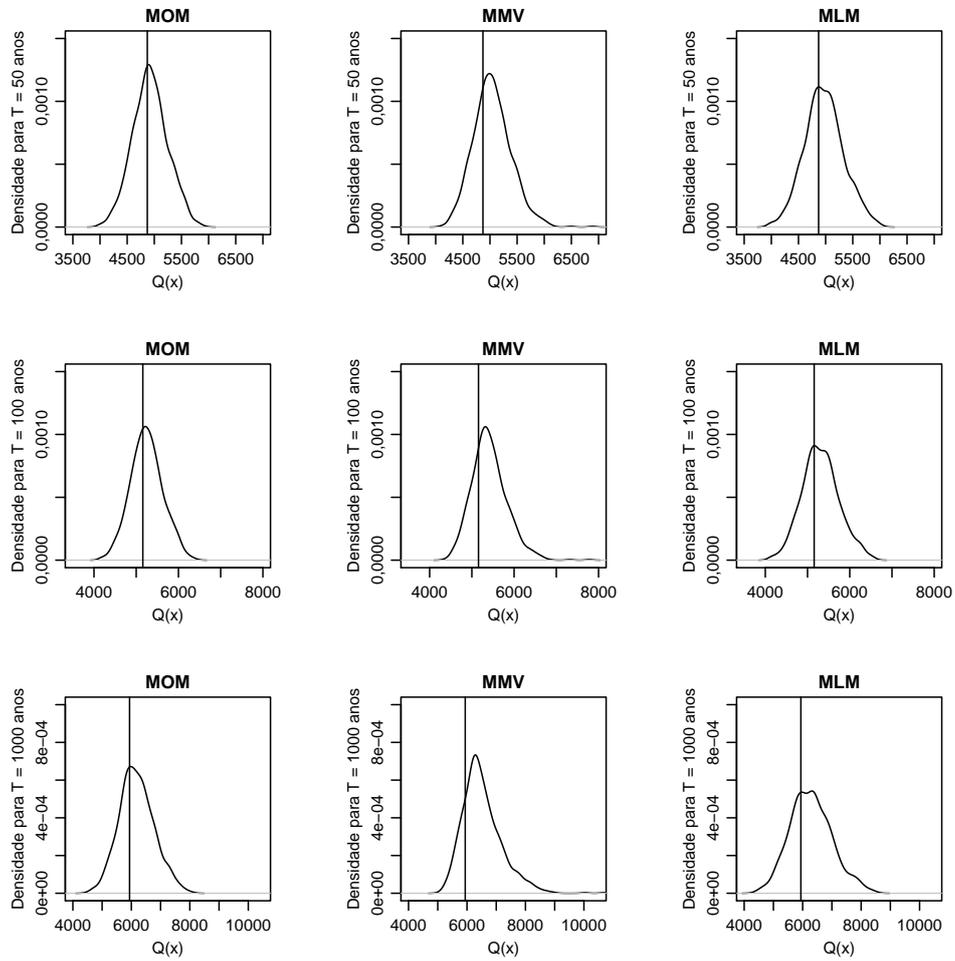
Tempo de retorno	Medida	Método		
		MOM	MMV	MLM
Q50	Viés	51,54	224,74	118,06
	Variância	$1,8 \times 10^5$	$2,56 \times 10^5$	$2,32 \times 10^5$
	EQM	$1,83 \times 10^5$	$3,06 \times 10^5$	$2,46 \times 10^5$
Q100	Viés	89,98	332,02	173,57
	Variância	$2,52 \times 10^5$	$3,84 \times 10^5$	$3,45 \times 10^5$
	EQM	$2,6 \times 10^5$	$4,94 \times 10^5$	$3,75 \times 10^5$
Q1000	Viés	245,9	731,16	388,29
	Variância	$6,13 \times 10^5$	$1,07 \times 10^6$	$9,4 \times 10^5$
	EQM	$6,73 \times 10^5$	$1,6 \times 10^6$	$1,09 \times 10^6$

Weibull – Pearson tipo III, n = 30, com contaminação



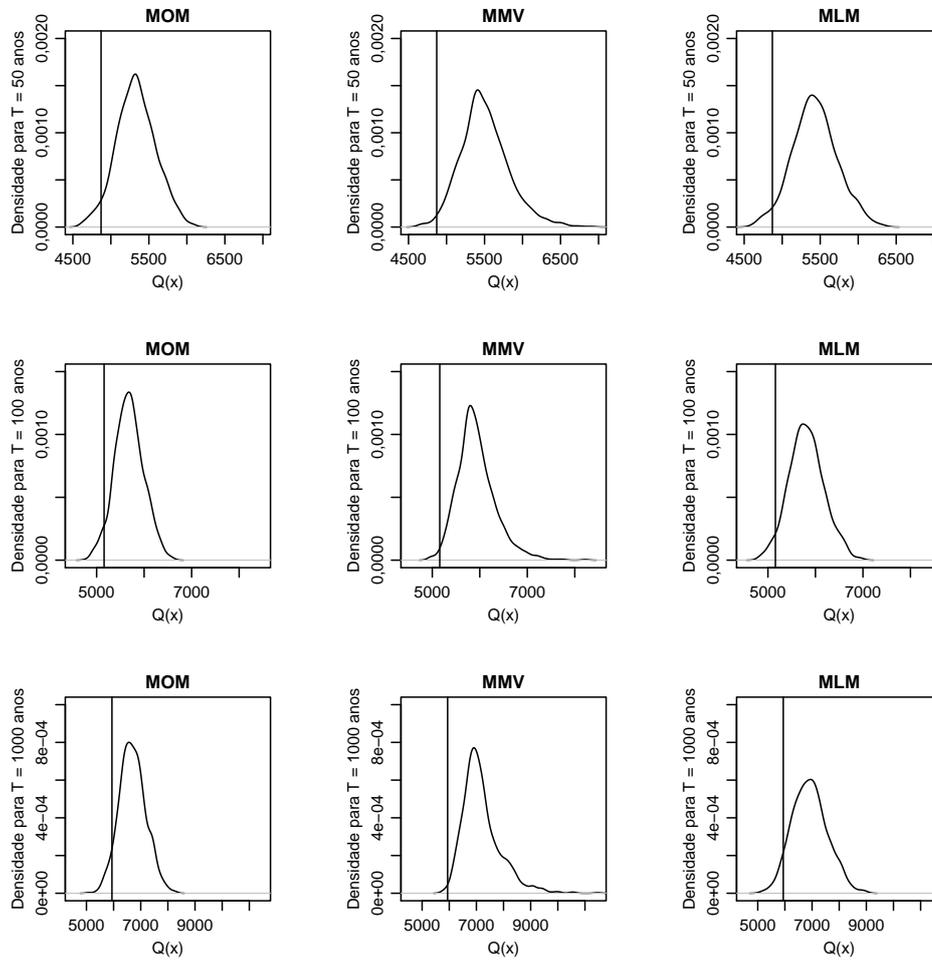
Tempo de retorno	Medida	Método		
		MOM	MMV	MLM
Q50	Viés	645,04	944,11	748,78
	Variância	$9,18 \times 10^4$	$2,38 \times 10^5$	$1,4 \times 10^5$
	EQM	$5,08 \times 10^5$	$1,13 \times 10^6$	$7,01 \times 10^5$
Q100	Viés	705,79	1119,63	835,09
	Variância	$1,32 \times 10^5$	$3,96 \times 10^5$	$2,29 \times 10^5$
	EQM	$6,3 \times 10^5$	$1,65 \times 10^6$	$9,26 \times 10^5$
Q1000	Viés	908,65	1726,98	1125,01
	Variância	$3,51 \times 10^5$	$1,31 \times 10^6$	$7,4 \times 10^5$
	EQM	$1,18 \times 10^6$	$4,29 \times 10^6$	$2,01 \times 10^6$

Weibull – Pearson tipo III, n = 50



Tempo de retorno	Medida	Método		
		MOM	MMV	MLM
Q50	Viés	51,87	170,73	101,26
	Variância	$1,05 \times 10^5$	$1,21 \times 10^5$	$1,3 \times 10^5$
	EQM	$1,07 \times 10^5$	$1,5 \times 10^5$	$1,4 \times 10^5$
Q100	Viés	90,09	256,04	151,55
	Variância	$1,46 \times 10^5$	$1,74 \times 10^5$	$1,9 \times 10^5$
	EQM	$1,54 \times 10^5$	$2,4 \times 10^5$	$2,13 \times 10^5$
Q1000	Viés	243,99	575,95	346,15
	Variância	$3,56 \times 10^5$	$4,5 \times 10^5$	$5,13 \times 10^5$
	EQM	$4,16 \times 10^5$	$7,82 \times 10^5$	$6,33 \times 10^5$

Weibull – Pearson tipo III, n = 50, com contaminação



Tempo de retorno	Medida	Método		
		MOM	MMV	MLM
Q50	Viés	461,57	637,36	560,7
	Variância	$6,66 \times 10^4$	$1,05 \times 10^5$	$8,75 \times 10^4$
	EQM	$2,8 \times 10^5$	$5,11 \times 10^5$	$4,02 \times 10^5$
Q100	Viés	527,67	772,21	654,25
	Variância	$9,37 \times 10^4$	$1,64 \times 10^5$	$1,36 \times 10^5$
	EQM	$3,72 \times 10^5$	$7,6 \times 10^5$	$5,64 \times 10^5$
Q1000	Viés	758,08	1245,43	978,12
	Variância	$2,41 \times 10^5$	$4,98 \times 10^5$	$4,18 \times 10^5$
	EQM	$8,16 \times 10^5$	$2,05 \times 10^6$	$1,37 \times 10^6$