

RELAÇÃO ENTRE OS QUANTIS DE CHUVAS MÁXIMAS E O TAMANHO DA AMOSTRA DE DADOS

Márcio Ferreira de Araujo Mendes¹; Cibele Pinheiro Martins Franco²;
Claudio Eduardo Vianna Silva³; Isabella Leôncio Guimarães⁴; Lílian Mara Sales Buonicontro⁵;
Luís Rogério Shimoda⁶; Luiz Adriano Torres Vieira⁷

Resumo – O presente trabalho teve como objetivo verificar a influência do número de anos de dados de chuva na inferência dos seus respectivos quantis de valores extremos. Foram selecionadas 8 (oito) estações pluviométricas para que suas amostras de chuva diária máxima compusessem os dados básicos deste trabalho. As distribuições de probabilidades empregadas nas análises do presente estudo e que se destacam na modelação de eventos máximos anuais de variáveis hidrológicas foram: Normal, Log-Normal, Gumbel, GEV (Generalized Extreme Value), Pearson Tipo III e Log-Pearson Tipo III. Para a verificação do ajuste da distribuição de probabilidade, foram utilizados os testes de aderência Kolmogorov-Smirnov e o Teste de Qui-Quadrado. Ao analisar a variação do Erro Relativo com o tamanho da amostra e com o Tempo de Recorrência, para as 8 (oito) estações selecionadas, verificou-se que as curvas de tendências apresentaram comportamentos semelhantes, ou seja, o Erro Relativo varia inversamente com o tamanho da amostra e diretamente proporcional ao Tempo de Recorrência.

Palavras-Chave – Hidrologia, Análise de Frequência.

RELATIONSHIP BETWEEN THE MAXIMUM RAINFALL QUANTIS AND DATA SAMPLE SIZE

Abstract – The present study aimed to investigate the influence of the number of years of rainfall data in the inference of their respective quantis of extreme values. We selected eight (8) rainfall stations to their maximum daily precipitation samples compose the basic data for this research. The probability distributions used in the analysis of this study and who excel at modeling events maximum annual hydrological variables were: Normal, Log-Normal, Gumbel, GEV (Generalized Extreme Value), Pearson Type III and Log-Pearson Type III. The adjustment for the probability distribution tests were used to grip Kolmogorov-Smirnov and Chi-Square. By analyzing the variation of relative error on the sample size and the recurrence time for the eight (8) selected stations, it was found that curves showed similar trends, and the relative error varies inversely with sample size is directly proportional to the time of recurrence.

Keywords – Hydrology, Frequency Analysis.

¹ Engenheiro Civil (UFMG), MSc em Recursos Hídricos (COPPE/UFRJ), professor da PUC Minas. e-mail: mmpucbh@gmail.com

² Aluna do curso de Eng. Civil da PUC Minas. e-mail: belpmf@yahoo.com.br

³ Aluno do curso de Eng. Civil da PUC Minas. e-mail: cevianna@gmail.com

⁴ Aluna do curso de Eng. Civil da PUC Minas. e-mail: isabela.leoncio@gmail.com

⁵ Engenheira Civil (PUC Minas). e-mail: lilianbuonicontro@yahoo.com.br

⁶ Aluno do curso de Eng. Civil da PUC Minas. e-mail: lrshimoda@yahoo.com.br

⁷ Aluno do curso de Eng. Civil da PUC Minas. e-mail: luiz.vieira@pbh.gov.br

1. INTRODUÇÃO

A precipitação é um processo aleatório e sua previsão, na maioria das vezes, é realizada com base na estatística de eventos passados. Na prática, o conhecimento estatístico das precipitações é de interesse da ordem técnica, por sua frequente aplicação nos projetos associados a aproveitamentos de recursos hídricos. Os estudos estatísticos permitem verificar com que frequência as precipitações ocorreram com dada magnitude, estimando as probabilidades teóricas de ocorrência das mesmas (Tucci, 2007).

Além disso, o estudo do comportamento da precipitação diária máxima anual é fundamental para a compreensão dos processos hidrológicos numa determinada região, principalmente para se conhecer a vazão de enchente de uma bacia, pois, através dos dados de precipitação máxima, é possível inferir os valores de vazão.

Em suma, os objetivos principais deste trabalho são: verificar a influência do número de anos de dados de chuva na inferência dos seus respectivos quantis de valores extremos, o que se justifica diante do fato de que muitos postos de medição de chuva possuem poucos anos de atividade, gerando amostras de pequena dimensão. Desta forma, pretende-se analisar oito estações pluviométricas com mais de 50 anos de operação; avaliar o erro cometido, dependendo da quantidade de anos de dados de chuva disponível, ou seja, o tamanho da amostra, para inferência de chuvas extremas.

2. DADOS BÁSICOS

A partir de consulta ao inventário de estações pluviométricas (BRASIL, 2009) selecionaram-se algumas estações com período de medição superior a 50 anos. Os dados de precipitação máxima diária de cada uma dessas estações foram obtidos através do portal HIDROWEB e, então, feita uma primeira análise para se determinar a série histórica de precipitações máximas diárias. A partir dessa análise foram selecionadas 8 (oito) estações pluviométricas para que suas amostras de chuva diária máxima, consistidas conforme indicado pela ANA, compusessem os dados básicos para esse trabalho (Tabela 1 e Figura 1).

Tabela 1 – Estações Pluviométricas Selecionadas

Código	Estação	Município	Latitude	Longitude	Altitude (m)	Extensão da Série (anos)
1544012	São Francisco	São Francisco - MG	-15° 56' 58"	-44° 52' 05"	448	57
1842005	Coroaci	Coroaci - MG	-18° 36' 43"	-42° 16' 43"	530	59
1943000	Mineração Morro Velho	Nova Lima - MG	-19° 58' 45"	-43° 51' 00"	770	58
1943023	Taquaraçu	Taquaraçu de Minas - MG	-19° 39' 50"	-43° 41' 17"	710	53
1944004	Ponte Nova do Paraopeba	Betim - MG	-19° 57' 20"	-44° 18' 24"	721	59
2247055	Jaguariuna	Jaguariuna - SP	-22° 42' 26"	-46° 59' 12"	570	60
2346095	Atibaia	Atibaia - SP	-23° 06' 20"	-46° 33' 26"	730	59
2349007	Piraju	Piraju - SP	-23° 11' 08"	-49° 23' 27"	500	63

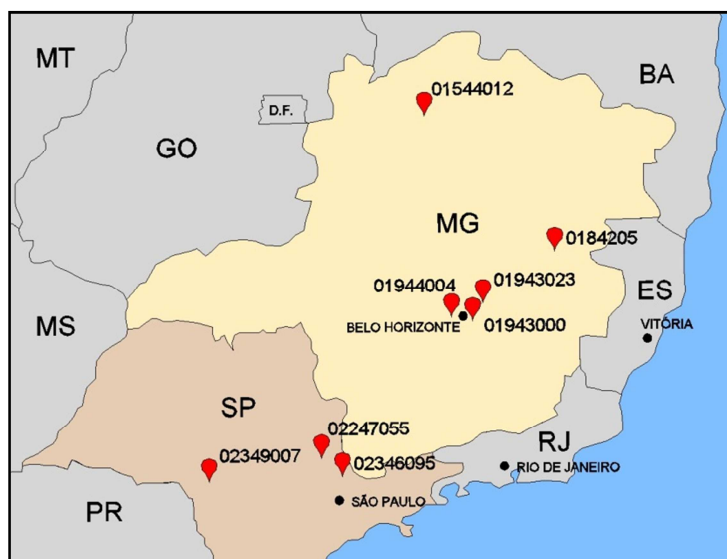


Figura 1– Mapa de Localização das Estações Pluviométricas Seleccionadas

Para cada estação escolhida, os dados de precipitação diária máxima foram separados por ano hidrológico, e selecionados o valor máximo de cada ano para compor a amostra de precipitações diárias máximas anuais de forma a subsidiar as análises para a determinação dos quantis de valores extremos. Além disso, quando a estação apresentava falha em um determinado ano hidrológico, este ano era desconsiderado para a composição da amostra.

3. METODOLOGIA

A metodologia empregada para trabalhar os dados de precipitação está apresentada resumidamente no fluxograma da Figura 2.

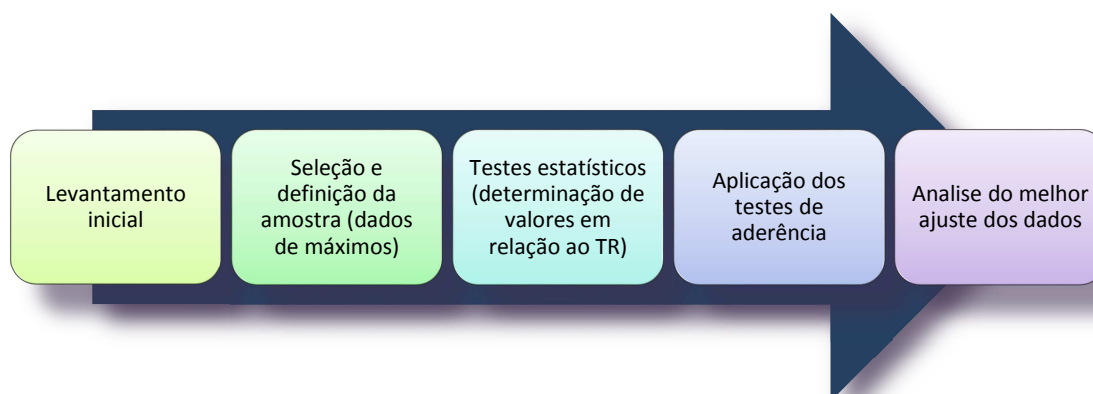


Figura 2– Fluxograma da metodologia

3.1 Metodologia para Determinação dos Quantis de Precipitação Máxima

Cada uma das amostras de dados de precipitações diárias máximas anuais de cada estação selecionada foi analisada estatisticamente para se determinar os quantis de precipitações extremas.

As séries de máximos anuais empregam apenas o maior evento de cada ano hidrológico, não considerando que o segundo maior evento de um ano pode ser superior aos picos de outros anos – daí a maior objeção à sua utilização. (Naghetini e Pinto, 2007)

Muito embora a análise de séries com duração parcial evite este problema, apresenta-se aí uma dificuldade extra, referente à definição de critérios que identifiquem somente aqueles eventos superiores ao limite estabelecido e que sejam independentes, “com a garantia de que não sejam utilizadas duas ou mais ocorrências que tenham como origem, o mesmo mecanismo ou evento causal” (Naghetini e Pinto, 2007, p.339). Eis o motivo de a análise com séries de duração parcial ser mais trabalhosa e difícil.

Portanto, optou-se por analisar as séries de precipitação diárias máximas anuais de dados obtidos de máximos anuais, ou seja, séries de dados obtidos a partir do maior evento de cada ano hidrológico.

As séries amostrais foram inicialmente testadas para a determinação de valores atípicos ou *outliers*, conforme o teste de Grubbs e Beck (Gubs e Beck, 1972 *apud* Naghetini e Pinto, 2007). Nesse teste, determinam-se os limites superiores (x_{sup}) e inferiores (x_{inf}), com certa significância, através da estatística de Grubbs e Beck, em seguida observações que eventualmente se posicionem acima do limite superior ou abaixo do inferior são consideradas atípicas ou *outliers* e excluídas antes da realização de análises de frequências.

Em seguida, foram ajustadas distribuições de probabilidades teóricas (Normal, Log-Normal, Gumbel, GEV – *Generalized Extreme Value*, Pearson Tipo III e Log-Pearson Tipo III) com parâmetros estimados pelo Método dos Momentos.

Para a escolha da melhor distribuição de probabilidade, levou-se em conta, além da análise visual dos gráficos de ajuste, os testes de aderência. Segundo Naghetini e Pinto (2007, p. 286), “os testes de aderência, como quaisquer testes de hipóteses, têm o objetivo de verificar se há uma diferença estatisticamente significativa entre as observações e as supostas realizações”. Neste trabalho, os testes de aderência que foram aplicados são o Teste de Kolmogorov-Smirnov e o Teste de Qui-Quadrado, por se tratarem de testes apropriados para análises de variáveis contínuas.

O Teste de Kolmogorov-Smirnov baseia-se na comparação do máximo da diferença absoluta das frequências empírica e teórica com um valor tabelado, função do tamanho da amostra e do grau de significância a ser considerado. Já o Teste de Qui-Quadrado baseia-se na diferença quadrática entre as realizações da variável aleatória e o valor esperado para essas realizações.

3.2 Metodologia para Verificação da Influência do Tamanho da Amostra

A partir dos dados básicos das oito estações previamente escolhidas, selecionaram-se os valores máximos anuais (ano hidrológico), formando-se, assim, a chamada amostra de referência de precipitações máximas anuais.

A essas séries, compostas por valores que foram considerados como referências, empreenderam-se análises de frequência e testes de aderência, conforme descritos anteriormente, para escolher a distribuição de probabilidade que melhor se ajustou à amostra de dados.

Os quantis de precipitações extremas, determinados a partir das amostras totais de dados, foram considerados como valores de referência para cada estação pluviométrica, respectivamente.

Em seguida, para a verificação da influência do tamanho da amostra na análise estatística, foram descartados os últimos anos das séries até que todas as estações pluviométricas em estudo possuíssem 50 anos. Este procedimento foi repetido para 40, 30, 20, 10 e 5 anos de dados amostrais.

Para cada amostra com tamanho entre 5 a 50 anos de dados empreendeu-se uma análise estatística para a determinação do melhor ajuste de distribuição de probabilidade, conforme já descrito, e, assim, definiu-se os quantis de precipitação máxima para cada estação e tamanho de amostra.

Admitindo-se os quantis de chuva máxima para cada período de retorno, considerando a amostra com o maior número de anos de dados, como valor de referência ($V_{\text{referência}}$), e os quantis de chuva máxima, para as amostras com os demais tamanhos, como valores estimados (V_{estimado}), determinou-se o Erro Relativo (ER) para os quantis de chuva máxima estimados através da seguinte equação (WIKIPÉDIA, 2012):

$$ER = \frac{|V_{\text{estimado}} - V_{\text{referência}}|}{V_{\text{referência}}} \quad (9)$$

Com os resultados pode-se verificar se o tamanho da amostra influencia na inferência de valores extremos de precipitação e como isso ocorre.

4. RESULTADOS OBTIDOS

Com base na metodologia para determinação dos quantis de precipitação máxima, escolheu-se a distribuição de probabilidade que melhor se ajustou a cada uma das amostras de dados analisadas, conforme apresentado na Tabela 2, a seguir, onde o de tamanho >50 anos corresponde ao tamanho máximo de dados para cada estação estudada.

A predominância para melhor ajuste da distribuição Pearson Tipo III, mas também com melhor ajuste para as distribuições Log-Normal, Gumbel, GEV e Log-Pearson Tipo III demonstra que as amostras de dados possuem características assimétricas.

De posse dos resultados estatísticos das amostras, foram desenvolvidas análises para diferentes tempos de retorno (2, 10, 50, 100, 500, 1.000 e 10.000 anos), procurando relacionar o Erro Relativo com o tamanho da amostra de dados, e pôde-se perceber que os resultados médios obtidos para a variação do Erro Relativo com o tamanho da amostra apresentam uma curva assintótica com a horizontal ($ER = 0$).

Tabela 2 – Distribuição de Probabilidade Escolhida para Ajuste

Estação (código)	Tamanho da Amostra (anos)						
	> 50	50	40	30	20	10	5
01544012	P III	P III	LP	P III	P III	P III	LP
01842005	G	P III	P III	P III	G	P III	P III
01943000	LP	LP	G	G	LP	LN	P III
01943023	P III	P III	GEV	G	G	LN	LP
01944004	LP	P III	LN	P III	P III	P III	P III
02247055	P III	P III	LP	LP	LP	P III	LP
02346095	G	G	LP	LP	P III	P III	G
02349007	P III	P III	PIII	LP	LP	P III	P III

Legenda:
LN - Dist. Log-Normal; G - Dist. Gumbel; GEV - Dist. GEV; P III - Dist. Pearson III; LP - Dist. Log-Pearson III

Para melhorar a visualização dos resultados obtidos, agruparam-se na Figura 3 **Erro! Fonte de referência não encontrada.**, a seguir, as curvas de tendência, a partir das médias dos resultados das estações, para cada tamanho de amostra e para cada Tempo de Recorrência.

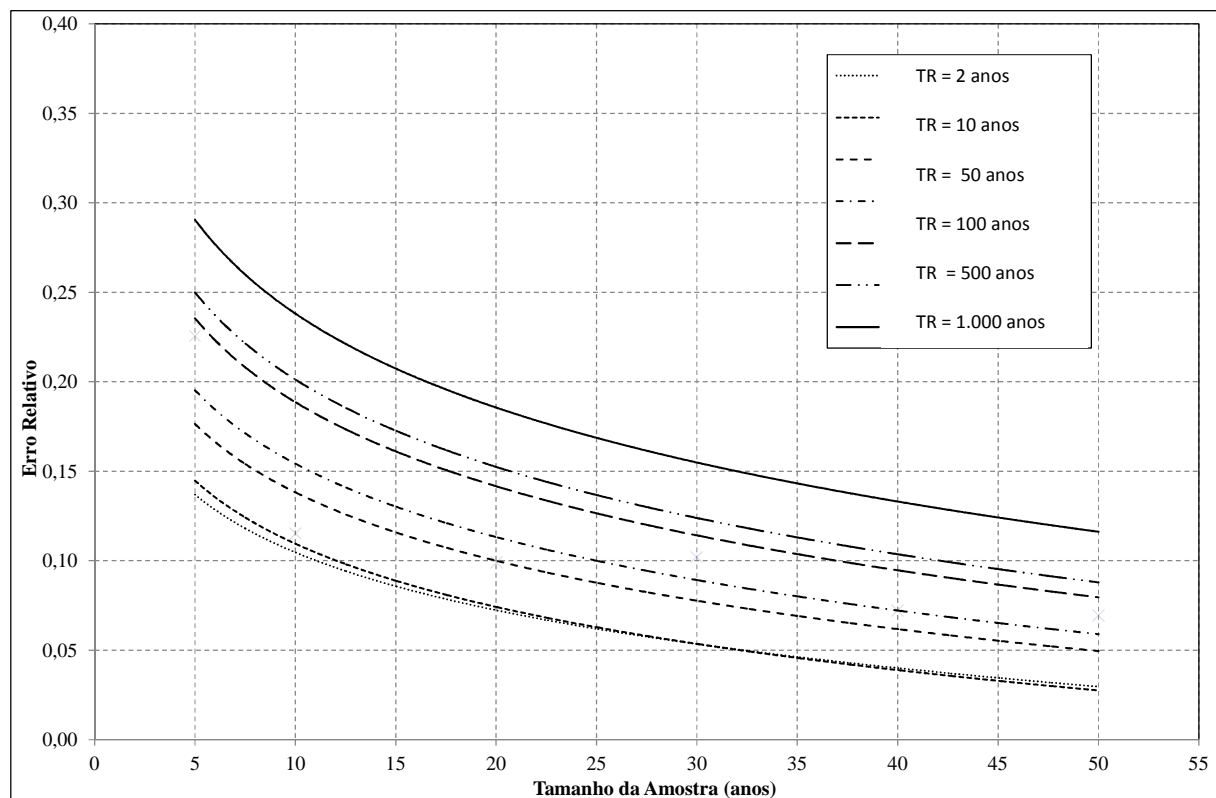


Figura 3 – Resumo das Tendências Médias das Variações dos Erros Relativos com o Tamanho das Amostras

5. CONCLUSÕES

A partir da análise dos resultados apresentados, é possível concluir que a inferência de quantis de chuva máxima sofre grande influência do tamanho da amostra de dados disponível.

Ao analisar a variação do Erro Relativo com o tamanho da amostra e com o Tempo de Recorrência, para as 8 (oito) estações selecionadas, verificou-se que as curvas de tendências apresentaram comportamentos semelhantes, e pode-se concluir que, além do tamanho da amostra, o erro varia também de acordo com o Tempo de Recorrência.

Resumidamente pode-se dizer que:

- a) O Erro Relativo, fixado o tempo de retorno, é inversamente proporcional ao tamanho da amostra, ou seja, decresce com o aumento da amostra disponível;
- b) O Erro Relativo tende para 0 (zero) à medida que o tamanho da amostra cresce, ou seja, a medida que a amostra tende para a população.
- c) O Erro Relativo, fixado o tamanho da amostra, é diretamente proporcional ao Tempo de Recorrência, ou seja, cresce com o aumento do Tempo de Recorrência.

Da análise dos resultados apresentados nos gráficos desse trabalho, pode-se perceber que à medida que se aumenta o Tempo de Recorrência relativo ao quantil de chuva máxima, o Erro Relativo aumenta. Além disso, curiosamente, as curvas de tendência das médias da variação do erro com o tamanho da amostra, para cada Tempo de Recorrência, apresentam um comportamento semelhante.

Isto pode ser explicado, pois a incerteza na determinação do quantil de chuva máxima é maior quanto mais futura for à inferência (maior Tempo de Recorrência) e quanto menor for o tempo de observação (menor tamanho da amostra).

Pôde-se perceber nos resultados obtidos que nem todas as amostras de dados das estações apresentaram resultados seguindo a mesma tendência média de variação do Erro Relativo com o tamanho da amostra, conforme supracitado. Não foi possível analisar tal fato profundamente, porém acredita-se que isto seja decorrente de possíveis inconsistências que ainda existam nos dados utilizados.

Após o estudo realizado, sugere-se que, para a inferência de quantis de chuva máxima, seja utilizado o maior número de dados de medição de chuva possível e consistidos, pois, dessa forma o erro é reduzido e pode tender a zero.

Para análises que buscam quantis com um Tempo de Recorrência grande, a atenção para o número de dados de medição de chuva consistidos da estação estudada deverá ser maior, pois como explicado anteriormente, quanto maior o Tempo de Recorrência, maior é o Erro Relativo para uma mesma quantidade de dados.

Infelizmente, no Brasil, a quantidade de dados consistidos e os postos com estes dados disponíveis são poucos e bastante heterogêneos (existência de falhas) se comparados à quantidade de postos de medição existentes.

Pode-se, então, concluir que é de extrema importância manter uma constante e duradoura coleta de dados confiáveis de precipitações nas estações pluviométricas, essencial para a elaboração de qualquer estudo do comportamento hidrológico de uma região.

REFERÊNCIAS

BRASIL. *Inventário das Estações Pluviométricas*. 2 ed. Brasília: Ministério do Meio Ambiente, ANA, SGH, 2009. 196 p.

TUCCI, Carlos E. M. (Org.). *Hidrologia: ciência e aplicação*. 4 ed. Porto Alegre: UFRGS, ABRH, 2007. 943 p.

NAGHETTINI, Mauro; PINTO, Éber José de Andrade. *Hidrologia Estatística*. Belo Horizonte: CPRM, 2007. 552 p. Disponível em: <<http://www.cprm.gov.br/publique/cgi/cgilua.exe/sys/start.htm?infoid=981&sid=36>> Acessado em: 12 de mar de 2012.

WIKIPÉDIA. *Teoria dos Erros*. Portal Wikipédia – A Enciclopédia Livre. Disponível em: <http://pt.wikipedia.org/wiki/Teoria_dos_erros> Acessado em: 04 de out de 2012.