

DISTRIBUIÇÃO GENERALIZADA DE PARETO APLICADA À ANÁLISE DE EVENTOS EXTREMOS DE PRECIPITAÇÃO PLUVIOMÉTRICA EM URUGUAIANA-RS

Amanda Larissa Alves Martins¹

Gilberto Rodrigues Liska²

Resumo:

A precipitação pluviométrica é o processo pelo qual a água condensada na atmosfera atinge gravitacionalmente a superfície terrestre. Ela é vital no ciclo da água que permite vida na Terra, nesse sentido a precipitação é um dos elementos climáticos que possuem grande impacto, pois está relacionado com diversos setores da sociedade, podendo afetar a economia, o meio ambiente e a própria sociedade. Prever a probabilidade de acontecimento desses eventos é importante para auxiliar o planejamento urbano e rural. A Teoria de Valores Extremos fundamenta a modelagem de eventos associados a probabilidades muito pequenas ou eventos raros. Uma das grandes preocupações é desenvolver técnicas para prever a ocorrência desses eventos. Essa teoria tem como uma das suas distribuições de probabilidade a distribuição Generalizada de Pareto (GPD) que é conhecida como limite da amostra de excessos ou excedências acima de um limiar suficientemente alto. Nesse sentido o presente trabalho tem como objetivo geral: analisar eventos extremos de precipitação por meio da teoria de valores extremos em Uruguaiana, RS. Para tal, utilizou-se os dados climatológicos fornecidos pelo Instituto Nacional de Meteorologia, os quais foram analisados sob a distribuição Generalizada de Pareto. As estimativas dos parâmetros foram obtidas pelo método da máxima verossimilhança e a aderência da distribuição foi avaliada pelos teste de hipóteses de Kolmogorov-Smirnov e gráficos Quantil-Quantil (QQplot). A independência das séries foi verificada pelo teste de Ljung Box e o teste de razão de verossimilhança foi utilizado para analisar a nulidade do parâmetro de forma da GPD. Os resultados alcançados evidenciam que a Distribuição Generalizada de Pareto se ajusta satisfatoriamente aos dados e pode ser utilizadas para fornecer níveis extremos de precipitação pluvial máxima.

Palavras-chave: Teoria de Valores Extremos; Probabilidade; Quantidade de chuva; Inundação.

Modalidade de Participação: Iniciação Científica

DISTRIBUIÇÃO GENERALIZADA DE PARETO APLICADA À ANÁLISE DE EVENTOS EXTREMOS DE PRECIPITAÇÃO PLUVIOMÉTRICA EM URUGUAIANA-RS

¹ Aluno de graduação. amandamartins2503@gmail.com. Autor principal

² Docente. gilbertoliska@unipampa.edu.br. Orientador

DISTRIBUIÇÃO GENERALIZADA DE PARETO APLICADA À ANÁLISE DE EVENTOS EXTREMOS DE PRECIPITAÇÃO PLUVIOMÉTRICA EM URUGUAIANA-RS

1 INTRODUÇÃO

A precipitação na hidrologia é entendida como a água proveniente do vapor de água na atmosfera depositada na superfície terrestre. (PINTO et al., 1976) A chuva, o granizo e a neve são exemplos de precipitação. Ela constitui um importante papel no ciclo hidrológico que por consequência influencia na qualidade do meio ambiente, além de afetar de forma direta e indireta a vida econômica e social da população, como na geração de energia e nas atividades agrícolas (BRITTO; BARLETTA; MENDONÇA, 2006).

Entende-se que a precipitação é vital para a vida na Terra, entretanto sua elevada ocorrência com dias de chuva consecutivos e chuvas com grande intensidade, podem acarretar em consequências negativas. Conforme Sá et al., (2018), essas consequências podem afetar os segmentos socioeconômicos de energia, abastecimento de água e a ocorrência de desastres naturais, como inundações e erosões no solo.

A possibilidade de minimizar, os riscos provenientes de eventos extremos com maior segurança, é de extrema importância para o auxílio ao planejamento urbano e rural. Abordagens estatísticas focadas em valores extremos mostram resultados promissores na previsão de eventos incomuns em ciências da terra, genética e finanças (THOMAS et al., 2016). Uma forma de modelar esses eventos é a Teoria de Valores Extremos, esta foi desenvolvida na década de 1920 e tem sido usada para prever a ocorrência de eventos tão variados como secas e inundações. Uma de suas distribuições é a Distribuição Generalizada de Pareto, que foi introduzida na literatura por Pickands em 1975 como a distribuição limite da amostra de excessos ou excedências acima de um limiar suficientemente alto. Sua aplicação tem sido feita extensivamente em diversos campos científicos, sendo mais referidas as aplicações nas áreas do seguros, finanças, meteorologia e ambiente.

Dado o exposto, nesse trabalho será usada a Distribuição Generalizada de Pareto para analisar as séries históricas de precipitação que abrangem o período de Janeiro de 1961 a Dezembro de 2015, esses dados foram agrupados em períodos mensais e a cada mês foi utilizado o método dos excedentes, isto é, os maiores valores de um conjunto de dados acima de um limiar suficientemente alto. O ajuste foi feito pelo método da máxima verossimilhança e foi realizado os testes de hipóteses de Kolmogorov-Smirnov, Ljung Box e o teste de razão de verossimilhança.

O objetivo deste trabalho é estudar o ajuste da Distribuição Generalizada de Pareto aos dados de precipitação diária da cidade de Uruguaiiana-RS, assim como calcular a probabilidade de um determinado evento extremo ocorrer e calcular níveis de retorno de eventos extremos de chuva.

2 METODOLOGIA

Os dados utilizados neste trabalho são de registros históricos de precipitação diários, no período de janeiro de 1961 a dezembro de 2015, disponibilizados pelo Instituto Nacional de Meteorologia (INMET) e registradas na estação meteorológica de Uruguaiiana – RS. Os dados foram agrupados em períodos mensais e a cada mês foi utilizado o método dos excedentes, isto é, os maiores valores de um conjunto de dados acima de um limiar suficientemente alto. Os valores dos excedentes são determinados de acordo com um limiar adotado, conforme a metodologia POT (*peaks over threshold*) e foram modelados pela Distribuição Generalizada de Pareto (GPD).

Conforme (COLES, 2001), assim como as Distribuições Generalizadas de Valores Extremos (GEV) são as distribuições limite para o máximo, as do tipo GPD são as formas

paramétricas para distribuições limite de excessos (Teorema de Balkema-Haan), ou seja, a GPD aparece como a distribuição limite dos excessos além de um limiar. A função de densidade da distribuição Generalizada de Pareto é denotada por:

$$f_{\xi,\sigma,\mu}(y) = \begin{cases} \frac{1}{\sigma} \left(1 + \xi \left(\frac{y-\mu}{\sigma} \right) \right)^{-(1+1/\xi)} & \text{para } 0 \leq y, \text{ se } \xi \neq 0 \\ \frac{1}{\sigma} \exp \left\{ -\frac{y-\mu}{\sigma} \right\} & \text{para } 0 \leq y \leq \frac{1}{|\xi|}, \text{ se } \xi = 0 \end{cases}$$

A Função de distribuição é dada por:

$$F_{\xi,\sigma,\mu}(y) = \begin{cases} 1 - \left(\frac{\xi(y-\mu)}{\sigma} \right)^{-1/\xi} & \text{para } \xi \neq 0 \\ 1 - \exp \left(-\frac{y-\mu}{\sigma} \right) & \text{para } \xi = 0 \end{cases}$$

Através da distribuição GPD podem-se obter três classes de distribuições padrões que são as: Tipo I: Exponencial ($\xi = 0$), Tipo II: Pareto ($\xi > 0$) e Tipo III: Beta ou Pareto Comum ($\xi < 0$).

Para escolha do valor limiar apropriado foi utilizada uma ferramenta gráfica exploratória baseada na linearidade da função média dos excessos que consiste na análise visual do gráfico de média dos excedentes (*mean residual life plot*). Conforme Silva, Portela e Naghettini (2013), esse gráfico relaciona a excedência média acima de cada limiar com o próprio limiar. Como o gráfico de média dos excedentes pode ser difícil de interpretar como um método de seleção do limiar. Uma técnica complementar foi usada para ajustar a distribuição generalizada de Pareto em uma variedade de limiares, e olhar para a estabilidade das estimativas dos parâmetros (COLES, 2001). O gráfico resultante desse procedimento é denominado de *threshold choice plot* (MENDES, 2004).

Após a seleção do limiar, foram estimados os parâmetros da GPD pelo método da máxima verossimilhança. Mais detalhes sobre esse método podem ser vistos em Casella e Berger (2001).

Com os parâmetros definidos verificou-se a adequação da GPD os dados. Aplicou-se o teste de aderência Kolmogorov Smirnov (KS) e teste de independência Ljung Box (LB), cuja estatística é comparada com o quantil da distribuição ao qui-quadrado com um grau de liberdade, com um nível de 5% de significância (BLAIN, 2014).

Para as estimativas por máxima verossimilhança, pode-se testar se ξ é estatisticamente zero. Então, para testar a hipótese nula de que a distribuição dos extremos é exponencial, foi utilizado o teste da razão de verossimilhança cuja estatística de teste é

$$TRV = -2(\mathcal{L}_{EXPONENCIAL} - \mathcal{L}_{GPD})$$

onde $\mathcal{L}_{EXPONENCIAL}$ e $\mathcal{L}_{GPD}$ representam as log-verossimilhanças usando respectivamente as densidade de Exponencial e da GPD com as respectivas estimativas por máxima verossimilhança (MENDES, 2004).

Para estimar as probabilidades de ocorrência de precipitação acima de um valor máximo foi utilizada a expressão:

$$\Pr\{X > x\} = \lambda \left[1 + \hat{\xi} \left(\frac{x-u}{\hat{\sigma}} \right) \right]^{-1/\hat{\xi}}$$

onde $\lambda = \Pr\{X > u\}$ e $\hat{\xi}$ e $\hat{\sigma}$ são as estimativas de máxima verossimilhança. O nível de retorno é o nível esperado para ser excedido uma vez a cada N anos e depende tempo de retorno. Dessa forma, o nível de retorno é definido por:

$$\hat{z}_n = u + \frac{\hat{\sigma}}{\hat{\xi}} \left[(Nn_y \hat{\lambda})^{\hat{\xi}} - 1 \right]$$

a menos que $\xi = 0$, resultando em:

$$\hat{z}_n = u + \hat{\sigma} \log(Nn_y \hat{\lambda}).$$

Finalizando a metodologia, para a realização dos testes de hipóteses, estimação dos parâmetros e cálculo das probabilidades, foi utilizado o Sistema Computacional Estatístico R, conforme R Core Team (2017) e pacote *evd* (STEPHENSON, 2002).

3 DISCUSSÃO E RESULTADOS

Na tabela 1, pode-se verificar que em todos os meses foi ajustada a distribuição exponencial ($\xi = 0$), pelo Teste de Razão de Verossimilhança. Na tabela também encontramos os resultados do teste de KS e LB para cada um dos meses.

Tabela 1: Estimativas dos parâmetros e Testes de Hipóteses da distribuição de probabilidade generalizada de Pareto para os dados de precipitação diária da cidade de Uruguaiana-RS.

Mês	Distribuição de Probabilidade	μ	σ	ξ	TRV	KS	LB
Janeiro	Exponencial	40	24,56	0	0,7293	0,6051	0,3279
Fevereiro	Exponencial	40	28,34	0	0,7815	0,8244	0,5872
Março	Exponencial	35	27,39	0	0,3871	0,9913	0,3941
Abril	Exponencial	40	23,9	0	0,4151	0,6001	0,2829
Mai	Exponencial	45	30,1	0	0,535	0,8096	0,5067
Junho	Exponencial	30	15,31	0	0,7583	0,825	0,5872
Julho	Exponencial	30	14,2	0	0,7469	0,9126	0,1415
Agosto	Exponencial	35	12,03	0	0,5827	0,0214	0,953
Setembro	Exponencial	40	15,86	0	0,2292	0,6723	0,3928
Outubro	Exponencial	40	22,64	0	0,3497	0,2687	0,253
Novembro	Exponencial	35	17,54	0	0,2788	0,6199	0,582
Dezembro	Exponencial	40	23,06	0	0,8383	0,8162	0,7695

Percebe-se na tabela 1 que em todos os meses os ajuste da distribuição exponencial às séries de precipitação foram satisfatórios e as séries de máximos mensais são independentes, com 5% de significância.

Uma crítica que surge no uso do teste Kolmogorov-Smirnov é que ele é insensível nos extremos da distribuição e testes alternativos satisfatórios não existem. Para contornar esse problema, gráficos QQ, mostrados na figura 1, foram complementarmente usados para avaliar o comportamento das distribuições nos seus extremos. Os gráficos de quantil-quantil para o diagnóstico da distribuição GPD comparam os quantis da distribuição teórica com os quantis da distribuição ajustada. O ajuste ocorre se o par ordenado formado por esses quantis estiverem em sua grande maioria contidos nas bandas de 95% de confiança. Dessa forma, os gráficos corroboram o ajuste das distribuições na Tabela 1.

A tabela 2 apresenta as estimativas dos níveis de retorno de precipitação para os períodos de retorno de 2 a 100 anos para todos os meses. Analisando os resultados apresentados na tabela 2, vê-se que por meio do ajuste da distribuição exponencial, que as estimativas de precipitação aumentam à medida que se aumenta o tempo de retorno.

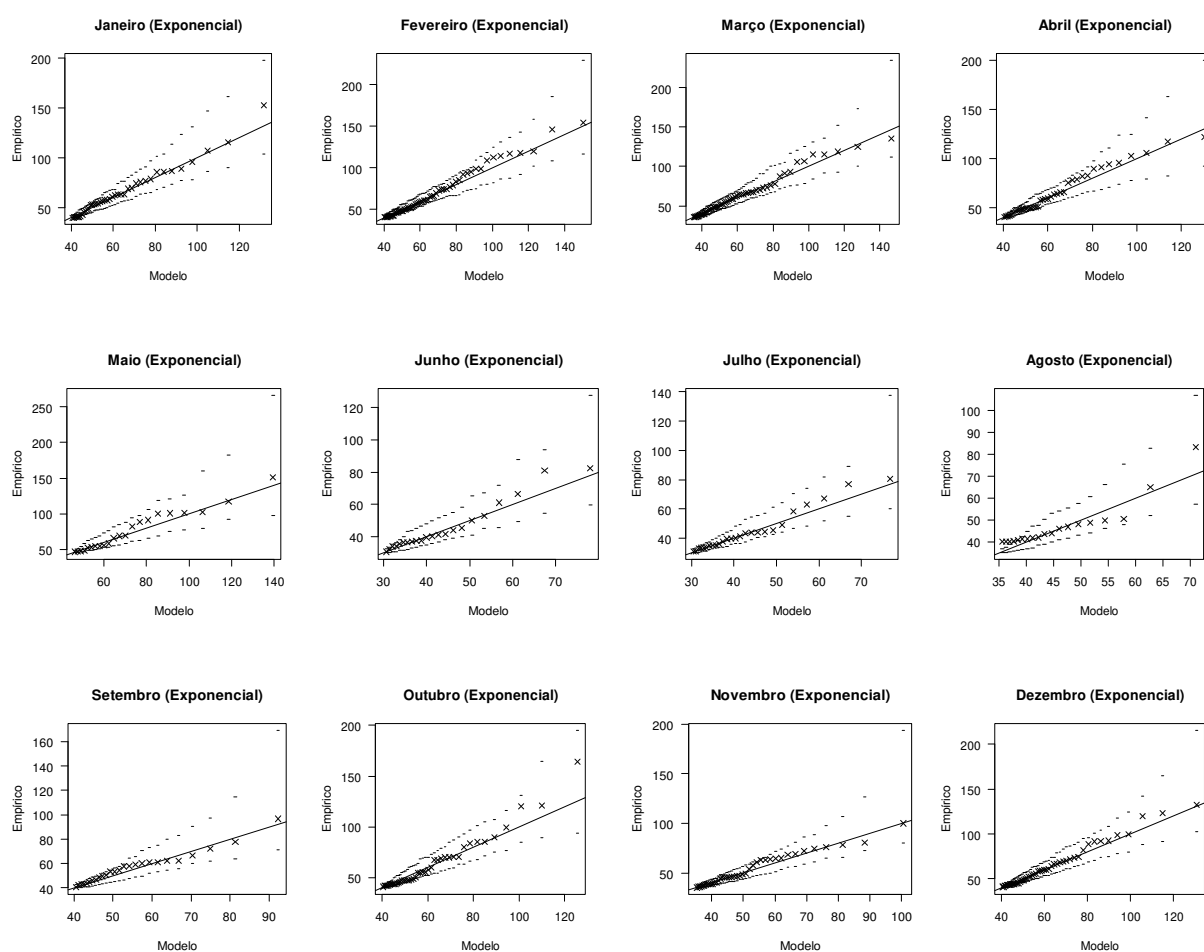


Figura 1: Gráficos QQplot para adequação de ajuste da GPD aos dados de precipitação máxima mensal da cidade de Uruguaiana-RS.

No período de Setembro a Maio, registram-se chuvas acima de 50 mm, que dependendo da intensidade horária poderá causar processos erosivos no solo, o que se torna prejudicial no sentido de contribuir a retirada de nutrientes essenciais para o desenvolvimento da cultura, conforme Silva, Portela e Naghettini (2013). Essa tabela mostra ainda, que as estimativas para o nível de retorno de precipitação são maiores nos meses referidos, em comparação ao resto do ano.

Tabela 2: Estimativas de Níveis de Retorno utilizando a distribuição exponencial para os dados de precipitação diária da cidade de Uruguaiana-RS.

Mês	Distribuição de Probabilidade	Tempos de Retorno					
		2	5	10	30	50	100
Janeiro	Exponencial	57,03	79,53	96,56	123,54	136,10	153,12
Fevereiro	Exponencial	59,64	85,61	105,25	136,38	150,86	170,50
Março	Exponencial	53,98	79,08	98,06	128,15	142,14	161,12
Abril	Exponencial	56,56	78,46	95,02	121,28	133,49	150,05
Maio	Exponencial	65,86	93,45	114,32	147,39	162,77	183,63

(Continuação da Tabela 2)

Mês	Distribuição de Probabilidade	Tempos de Retorno					
		2	5	10	30	50	100
Junho	Exponencial	40,61	54,64	65,26	82,08	89,90	100,52
Julho	Exponencial	39,84	52,86	62,70	78,31	85,56	95,41
Agosto	Exponencial	43,33	54,35	62,69	75,90	82,04	90,38
Setembro	Exponencial	50,99	65,53	76,52	93,94	102,05	113,04
Outubro	Exponencial	55,69	76,44	92,14	117,01	128,58	144,28
Novembro	Exponencial	47,16	63,23	75,39	94,67	103,63	115,79
Dezembro	Exponencial	55,98	77,11	93,09	118,43	130,21	146,19

4 CONSIDERAÇÕES FINAIS

A distribuição Generalizada de Pareto com parâmetro de forma nulo (exponencial) se ajustou satisfatoriamente em todos os meses estudados e pode ser utilizada para fornecer níveis extremos de precipitação pluvial máxima. Como passo seguinte ao trabalho, serão calculadas as probabilidades de certos níveis de precipitação serem superados e os resultados serão comparados à metodologia dos blocos máximos, uma vez que o presente estudo utilizou a metodologia *peaks over threshold* (POT).

REFERÊNCIAS

- BLAIN, G. C. Dry months in the agricultural region of Ribeirão Preto, state of São Paulo-Brasil: an study based on the extreme value theory. **Engenharia Agrícola**, v. 35, n. 5, p. 992–1000, 2014.
- BRITTO, F. P.; BARLETTA, R.; MENDONÇA, M. Regionalização sazonal e mensal da precipitação pluvial máxima no Estado do Rio Grande do Sul. **Revista Brasileira de Climatologia**, v. 2, p. 35–51, 2006.
- CASELLA, G.; BERGER, R. R. **Statistical Inference**. 2. ed. [s.l.] Thomson Learning, 2001.
- COLES, S. **An Introduction to Statistical Modeling of Extreme Values**. London: Springer London, 2001.
- MENDES, B. V. DE M. **Introdução à análise de eventos extremos**. Rio de Janeiro: E-papers Serviços Editoriais Ltda, 2004.
- PINTO, N. L. DE S. et al. **Hidrologia básica**. Editora Bl ed. [s.l.: s.n.].
- R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing**. Vienna, Austria R Foundation for Statistical Computing, , 2015.
- SÁ, E. A. S. et al. Trends in daily precipitation in highlands region of Santa Catarina, southern Brazil. **Ambiente e Agua - An Interdisciplinary Journal of Applied Science**, 2018.
- SILVA, A. T.; PORTELA, M. M.; NAGHETTINI, M. Análise de frequência de máximos anuais baseada em séries de duração parcial. Combinação das distribuições de Poisson inflacionada de zeros e generalizada de Pareto, modelo ZIP-GP. **Revista Recursos Hídricos**, v. 34, n. 2, p. 5–12, 2013.
- STHEPHENSON, A. G. evd: Extreme Value Distributions. **R News**, v. 2, n. 2, p. 31–32, 2002.
- THOMAS, M. et al. Applications of extreme value theory in public health. **PLoS ONE**, p. 1–7, 2016.