# Distribuição Generalizada de Pareto aplicada à análise de eventos extremos de chuva em Uruguaiana-RS

Gilberto Rodrigues Liska; Amanda Larissa Alves Martins; Luiz Alberto Beijo; Marcelo Ângelo Cirillo; Fortunato Silva Menezes gilbertoliska@ufscar.br

UFSCar - CCA - DTAiSeR

V International Seminar on Statistics with R







- Introdução
- 2 Metodologia
- 3 Apresentação da Pesquisa e Análise dos Resultados
- 4 Considerações Finais
- 5 Referências

#### Introdução



- A chuva é vital para a vida na Terra;
- Sua elevada ocorrência com dias de chuva consecutivos e chuvas com grande intensidade;
- Consequências negativas.



Figura 1: Inundação na cidade de Uruguaiana

#### Análise de Eventos Extremos



- Distribuição Generalizada de Pareto (1975);
- Dado o uso de modelos probabilísticos, avaliar sua qualidade de ajuste é uma tarefa igualmente importante;
- O ajuste por meio de estimativas dos parâmetros das distribuições ajustadas pode levar à ocorrência de erro tipo II e para contornar esse fato, propõe-se um estudo de simulação.

## Disponibilidade dos materiais



Material: Todos os materiais (apresentação, scipt em R, arquivo de dados) estarão disponíveis em pasta virtual (V\_SER\_Gilberto)

```
Link da pasta virtual: https://ldrv.ms/u/s!
AvxsaQZPoPWd4TquBa0e7zxRvNvv?e=Z9Qm38
```



- Introdução
- Metodologia
- 3 Apresentação da Pesquisa e Análise dos Resultados
- 4 Considerações Finais
- 5 Referências

## Metodologia: Dados



- Dados climatológicos fornecidos pelo Banco de Dados Metereológicos de Ensino e Pesquisa (BDMEP) e pelo Instituto Nacional de Metereologia (INMET) de janeiro de 1961 a abril de 2019;
- Os dados foram agrupados em períodos mensais e em cada mês foi utilizado o método dos excedentes;
- Metodologia POT (PeaksOverThreshold).

## Metodologia: Distribuição Generalizada de Pareto



• Foi adotada a GPD, cuja função de distribuição é:

$$F(x|\xi,\sigma,u) = \begin{cases} 1 - \left[1 + \xi\left(\frac{x-u}{\sigma}\right)\right]^{-\frac{1}{\xi}}, & \xi \neq 0\\ 1 - \exp\left(-\frac{x-u}{\sigma}\right), & \xi \to 0 \end{cases}$$
(1)

onde u é o limiar,  $\sigma$  é o parâmetro de escala e  $\xi$  é o parâmetro de forma.

Através da GPD, podemos obter três classes de distribuições padrões que são as: Tipo I: Exponencial (lim <sub>ξ→0</sub> F (x | ξ, σ, u)), Tipo II: Pareto (ξ > 0) e Tipo III: Beta ou Pareto comum (ξ < 0).</li>

## Metodologia: Escolha do limiar



Para escolha de um limiar apropriado:

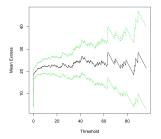


Figura 2: Gráfico de vida média residual para a escolha de um limiar

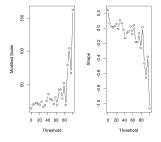


Figura 3: Gráfico de escolha de limiar para parâmetros estimados de escala e de forma

 Após a seleção do limiar, foram estimados os parâmetros da GPD pelo método da máxima verossimilhança.

## Metodologia: Testes de hipóteses



- A adequicidade do ajuste da GPD aos dados foi validada pelo teste de aderência Kolmogorov Smirnov(KS) e teste de independência de Ljung Box(LB);
- Foi aplicado o teste de Mann-Kendall;
- Teste de razão de verossimilhança (TRV), cuja estatística de teste é:

$$\Lambda = 2 \left[ I(\hat{\sigma}, \hat{\xi}) - I(\hat{\sigma}) \right], \tag{2}$$

Para todos os teste foi adotado 1% como nível de significância.

## Metodologia: Probabilidades e Nível de retorno



#### Cálculo de Probabilidades e Níveis de Retorno

• A probabilidade de que um nível seja excedido é dado por

$$\Pr[X > x] = \lambda \left[ 1 + \xi \left( \frac{x - u}{\sigma} \right) \right]^{-\frac{1}{\xi}} \tag{3}$$

onde  $\lambda=\Pr[X>u]$  e  $\xi$  e  $\sigma$  são as estimativas de máxima verossimilhança.

- O nível de retorno é o nível esperado para ser excedido uma vez a cada N anos.
- Além das estimativas do nível de retorno, foram construídos os intervalos de confiança associados aos períodos de retorno de 2, 5, 10, 30, 50 e 100 anos.

## Metodologia: Simulação computacional



Estudo de simulação para avaliar a qualidade do ajuste para distribuições de valores extremos

- Método de Monte Carlo:
- Série de treinamento (1961-1991);
- Série de teste (1992-2019);
- São considerados dois cenários: (1) gera amostras da distribuição Exponencial com os parâmetros estimados e; (2) gera amostras da GPD com os parâmetros estimados;

Cada cenário foi repetido 10.000 vezes. Depois disso, obteve-se a média de Monte Carlo do MAPE (erro pencentual médio absoluto) e RMSE (raiz quadrada do erro quadrático médio).

## Metodologia: Simulação computacional



Além disso, foram calculados os seguintes:

- Proporção da qual o TRV resultou em um p-valor superior ao nível de significância de 1% denotado de p̂<sub>TRV</sub>;
- Proporção da qual o MAPE da GPD é maior do que o MAPE da distribuição Exponencial, denotado por p̂<sub>MAPE</sub>;
- Proporção da qual a RMSE da GPD é maior do que a RMSE da distribuição Exponencial, denotada por p̂<sub>RMSE</sub>;

Ressalta-se que os tempos de retorno adotados compreendem o tempo da série de testes.

## Recursos Computacionais



 Para a realização dos testes de hipóteses, estimação dos parâmetros, cálculo das probabilidades, níveis de retorno e estudo de simulação, foi utilizado o Sistema Computacional Estatístico R, conforme RCoreTeam (2017) e pacote evd (2002).





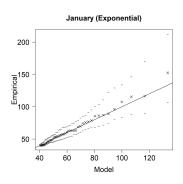
- Introdução
- Metodologia
- 3 Apresentação da Pesquisa e Análise dos Resultados
- 4 Considerações Finais
- 5 Referências

#### Apresentação da Pesquisa e Análise dos Resultados



Tabela 1: Limiar  $(\hat{u})$ , estimativas dos parâmetros e teste de Hipóteses (p-valor) das distribuições Generalizada de Pareto (GPD) e Exponencial para dados de chuva máxima mensal em Uruguaiana, RS.

Mês	Distribuição de Probabilidade	û	ô	$\hat{\xi}$	TRV	KS	LB	Mann Kendall
Janeiro	Exponencial	40	24,1319	-	0,883	0,774	0,242	0,7413
Janeno	GPD	40	24,6996	-0,0239	0,003	0,762	0,242	0,1413
Fevereiro	Exponencial	40	27,875	-	0,773	0,772	0,821	0.4654
	GPD	40	29,2355	-0,0486		0,723	-,	-,
Março	Exponencial	35	27,7219	-	0,377	0,717	0,266	0.0555
iviaiço	GPD	35	31,9006	-0,1463		0,459	0,200	0,0333
Abril	Exponencial	50	29,7306	-	0,376	0,951	0,374	0,4456
ADrii	GPD	50	35,3042	-0,1832		0,9		0,4430
Maio	Exponencial	45	30,9034	-	0,941	0,602	0,288	0,5735
	GPD	45	30,3336	0,019		0,637		
Junho	Exponencial	30	20,4484	-	0,051	0,113	0,222	0,1738
	GPD	30	14,1875	0,3039		0,527		
Julho	Exponencial	35	13,3269	-	0,788	0,848	0,099	0,0291
Junio	GPD	35	12,2025	0,087		0,889		0,0291
Agosto	Exponencial	35	16,8577	-	0,179	0,083	0,021	0,7407
Agosto	GPD	35	12,977	0,227		0,039		
Setembro	Exponencial	23	20,0213	-	0.216	0,378	0,675	0,2254
Setembro	GPD	23	24,3738	-0,2069		0,181		0,2254
Outubro	Exponencial	40	21,7176	-		0,478	0,27	0,6963
	GPD	40	19,4466	0,1053		0,752		0,0903
Novembro	Exponencial	40	20,4262	-	0.570	0,526	0,758	0.0415
	GPD	40	21,9398	-0,0755	0,578	0,684		0,0415
	Exponencial	40	21,1594	-	0.070	0,665	0.750	0.000
Dezembro	GPD	40	21,0502	0.0052	0,973	0,677	0,763	0,6555



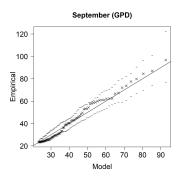


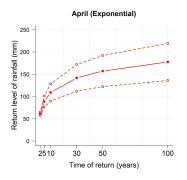
Figura 4: Gráficos de ajuste da melhor distribuição mostrada pela comparação das tabelas 4 e 5 para os dados de precipitação máxima mensal da cidade de Uruguaiana-RS.

Tabela 2: Probabilidades (%) de ocorrência de chuvas pelas distribuições de probabilidade dos dados mensais de precipitação máxima do município de Uruguaiana, RS.

• • •	Distribuição de	Quantidade de chuva (mm)					
Mês	Probabilidade	50	75	100	125	150	
1	Exponencial	66,07	23,45	8,32	2,95	1,05	
Janeiro	GPD	66,58	23,65	8,19	2,76	0,9	
Fevereiro	Exponencial	69,86	28,49	11,62	4,74	1,93	
revereiro	GPD	70,83	29,13	11,51	4,35	1,57	
	Exponencial	58,21	23,62	9,59	3,89	1,58	
Março	GPD	61,44	25,03	8,9	2,63	0,6	
Abril	Exponencial	100	43,13	18,6	8,02	3,46	
ADrii	GPD	100	46,84	19,4	6,78	1,84	
	Exponencial	85,06	37,88	16,87	7,51	3,35	
Maio	GPD	84,83	37,54	16,82	7,63	3,5	
	Exponencial	37,6	11,07	3,26	0,96	0,28	
Junho	GPD	30,93	10,85	4,91	2,59	1,52	
Julho	Exponencial	32,45	4,97	0,76	0,12	0,02	
Juino	GPD	31,1	5,59	1,26	0,34	0,1	
	Exponencial	41,07	9,32	2,12	0,48	0,11	
Agosto	GPD	35,83	9,66	3,52	1,55	0,78	
C	Exponencial	25,96	7,45	2,14	0,61	0,18	
Setembro	GPD	28,42	5,99	0,6	0,01	0	
Outubro	Exponencial	63,1	19,96	6,31	2	0,63	
Outubro	GPD	60,61	19,24	6,91	2,74	1,18	
Novembro	Exponencial	61,29	18,02	5,3	1,56	0,46	
ivovembro	GPD	62,89	18,27	4,67	1,02	0,18	
	Exponencial	62,34	19,13	5,87	1,8	0,55	
Dezembro	GPD	62,22	19,1	5,9	1,84	0,58	

Tabela 3: Estimativas dos níveis de retorno (mm) pelas distribuições de probabilidade dos dados mensais de precipitação máxima do munícipio de Uruguaiana, RS.

Mês	Distribuição de	Tempo de Retorno (anos)						
ivies	Probabilidade	2	5	10	30	50	100	
Janeiro	Exponencial	56,01	78,12	94,85	121,36	133,69	150,41	
	GPD	56,27	78,28	94,73	120,24	132,01	148,07	
Fevereiro	Exponencial	67,97	93,51	112,83	143,46	157,69	177,02	
Fevereiro	GPD	68,65	93,58	111,76	139,79	152,49	171,05	
	Exponencial	64,78	90,18	109,39	139,85	154,01	173,23	
Março	GPD	66,73	90,18	105,82	127,88	136,67	149,79	
Abril	Exponencial	61,26	88,5	109,11	141,77	156,96	177,57	
Abrii	GPD	62,94	90,69	109,09	133,4	143,2	155,29	
	Exponencial	49,55	77,86	99,28	133,23	149,02	170,44	
Maio	GPD	49,46	77,58	99,22	133,71	149,47	173,03	
Junho	Exponencial	34,11	52,85	67,02	89,49	99,94	114,11	
	GPD	32,94	48,88	64,2	96,2	114,48	131,16	
	Exponencial	35,63	47,84	57,08	71,72	78,52	87,76	
Julho	GPD	35,57	47,25	56,73	72,96	80,89	92,18	
	Exponencial	35,88	51,33	63,01	81,53	90,14	101,83	
Agosto	GPD	35,68	49,05	61,19	84,6	97,89	117,3	
	Exponencial	45,29	63,64	77,51	99,51	109,74	123,62	
Setembro	GPD	47,24	63,4	73,73	87,43	92,8	99,41	
0.1.	Exponencial	56,28	76,18	91,24	115,1	126,19	141,24	
Outubro	GPD	55,18	75,43	92,09	120,66	134,59	152,27	
Name	Exponencial	51,14	69,85	84,01	106,45	116,88	131,04	
Novembro	GPD	51,72	70,37	83,64	103,28	112,03	123,35	
	Exponencial	60,54	79,93	94,6	117,84	128,65	143,32	
Dezembro	GPD	60,49	79,93	94,6	117,85	128,65	143,32	



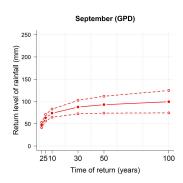


Figura 5: Gráficos de níveis de retorno (em anos) e intervalos de confiança para os dados de precipitação máxima mensal do município de Uruguaiana-RS. As linhas tracejadas representam o intervalo de confiança de 95% e as linhas sólidas representam o nível de retorno estimado pela melhor distribuição mostrada pela comparação das tabelas 4 e 5.

Tabela 4: Resultados de cenário 1 para a simulação de Monte Carlo em 10000 repetições para cada mês do ano para as distribuições Exponencial e GPD dos dados mensais de precipitação máxima em Uruguaiana-RS.

Mês	Distribuição de Probabilidade*	MAPE	RMSE	ρ̂ <sub>ΜΑΡΕ</sub> (%)	ρ̂ <sub>RMSE</sub> (%)	ρ̂ <sub>TRV</sub> (%)
Janeiro	Exponencial GPD	23,99 24,05	37,11 37,23	54,31	54,31	94,42
Fevereiro	Exponencial GPD	33,18 33,19	20,69 20,72	51,04	49,32	94,80
Março	Exponencial GPD	20,92 21,01	27,70 27,79	54,70	54,76	94,58
Abril	Exponencial GPD	22,59 22,68	44,44 44,50	54,17	56,19	94,70
Maio	Exponencial GPD	15,14 15,00	19,57 19,38	44,69	44,69	94,55
Junho	Exponencial GPD	56,70 56,74	105,47 105,54	54,57	54,57	94,86
Julho	Exponencial GPD	53,30 53,19	25,42 25,36	45,96	45,97	94,94
Agosto	Exponencial GPD	89,39 89,38	25,72 25,73	50,06	48,36	94,80
Setembro	Exponencial GPD	25,32 25,20	15,25 15,16	45,91	45,91	94,79
Outubro	Exponencial GPD	31,36 31,41	52,15 52,26	53,28	53,28	94,64
Novembro	Exponencial GPD	13,19 13,15	12,65 12,61	46,00	45,92	94,72
Dezembro	Exponencial GPD	20,77 20,86	18,48 18,59	54,47	54,47	94,56

<sup>\*</sup> A distribuição de probabilidade em itálico indica que a distribuição exponencial é melhor em relação ao mês. A distribuição de probabilidade em sublinhado indica que a GPD é melhor em relação ao mês. MAPE: Erro percentual médio absoluto. RMSE: Raiz quadrada do erro quadrático médio.  $\hat{p}_{TRV}$ : proporção da qual o TRV resultou em um p-valor maior que o nível de significância de 1%.  $\hat{p}_{MAPE}$ : proporção da qual o MAPE da GPD é maior do que o MAPE da distribuição Exponencial.  $\hat{p}_{RMSE}$ : proporção em que o RMSE da GPD é maior que RMSE da distribuição Exponencial. 4 - 1 4 - 4 - 4 - 5 + 4 - 5 +

Tabela 5: Resultados do cenário 2 para simulação de Monte Carlo em 10000 repetições para cada mês do ano para as distribuições Exponencial e GPD dos dados mensais de precipitação máxima em Uruguaiana-RS.

Mês	Distribuição de Probabilidade*	MAPE	RMSE	ρ <sub>ΜΑΡΕ</sub> (%)	ρ̂ <sub>RMSE</sub> (%)	ρ̂ <sub>TRV</sub> (%)
Janeiro	Exponencial GPD	24,03 24,46	37,19 37,88	77,86	77,87	89,19
Fevereiro	Exponencial GPD	33,17 32,90	20,68 20,41	22,36	16,95	76,05
Março	Exponencial GPD	20,15 22,88	26,97 28,96	100,00	100,00	1,26
Abril	Exponencial GPD	22,45 25,64	44,30 45,86	100,00	100,00	0,06
Maio	Exponencial GPD	15,16 15.61	19,61 20.23	65,69	65,68	92,11
Junho	Exponencial GPD	63,92 63,58	117,87 117,32	11,86	11,86	73,48
Julho	Exponencial GPD	53,59 55,74	25,56 26,91	97,94	97,95	39,20
Agosto	Exponencial GPD	89,19 89,49	25,66 26,25	70,84	91,60	0,20
Setembro	Exponencial GPD	25,85 20,31	15,63 10,93	0,00	0,00	0,01
Outubro	Exponencial GPD	31,34 30,03	42,09 49,72	1,09	1,09	26,54
Novembro	Exponencial GPD	13,14 11,72	12,61 11,54	1,69	1,34	48,83
Dezembro	Exponencial GPD	20,77 20,74	18,48 18,45	48,16	48,16	94,71

<sup>\*</sup>A distribuição de probabilidade em itálico indica que a distribuição exponencial é melhor em relação ao mês. A distribuição de probabilidade em sublinhado indica que a GPD é melhor em relação ao mês. MAPE: Erro percentual médio absoluto. RMSE: Raiz quadrada do erro quadrático médio.  $\hat{p}_{TRV}$ : proporção da qual o TRV resultou em um p-valor maior que o nível de significância de 1%.  $\hat{p}_{MAPE}$ : proporção da qual o MAPE da GPD é maior do que o MAPE da distribuição Exponencial.  $\hat{p}_{RMSE}$ : proporção em que o RMSE da GPD é maior que RMSE da distribuição Exponencial.



- Introdução
- Metodologia
- 3 Apresentação da Pesquisa e Análise dos Resultados
- 4 Considerações Finais
- 5 Referências

## Considerações Finais



- A distribuição generalizada de Pareto foi ajustada satisfatoriamente em todos os meses e pode ser usada para fornecer níveis extremos de precipitação máxima;
- Foram calculadas as estimativas de precipitação dos meses de janeiro a dezembro para os períodos de retorno de 2, 5, 10, 30, 50 e 100 anos. A maior estimativa foi observada em abril e o menor nível de retorno foi em julho;
- Não foi encontrada tendência dos máximos mensais de precipitação pluviométrica pelo teste de Mann-Kendall;
- As estimativas intervalares fornecem valores mais seguros de precipitações máximas prováveis, e podem ser usadas para auxiliar no planejamento urbano e rural na cidade de Uruguaiana-RS;
- Ao comparar as distribuições por simulação computacional, foi possível identificar a verdadeira distribuição de probabilidade dos valores extremos do excesso de um limiar;



- Introdução
- 2 Metodologia
- 3 Apresentação da Pesquisa e Análise dos Resultados
- 4 Considerações Finais
- 6 Referências

#### Referências



- Beijo, L. A.; Muniz, J. A. & Castro Neto, P. Maximum rainfall return period by extreme values type I distribution in Lavras, Minas Gerais state, Brazil. Ciência e Agrotecnologia, 2005, 29, 657-667
- Blain, G. C. Dry months in the agricultural region of Ribeirão Preto, state of São Paulo-Brazil: an study based on the extreme value theory. Engenharia Agrícola, 2014, 34, 992-1000
- 3 da Pereira Britto, F. D. R.; Barletta, R. & Mendonça, M. Regionalização sazonal e mensal da precipitação pluvial máxima no estado do Rio Grande do Sul. Revista Brasileira de Climatologia, 2006, 2, 35-52
- 4 Coles, S. An introduction to statistical modeling of extreme values. Great Britain: Springer (2001)
- G1. Deslizamentos de terra em Rolante atingiram 230 hectares. Avaiable in: http://glo.bo/2kvctLE. Acessed in 20 may 2020
- 6 Herrmann, M. L. P. Atlas de desastres naturais do Estado de Santa Catarina. Florianópolis, IOESC, 2006, 1, 146
- Mendes, B. V. DE M. Introdução à análise de eventos extremos. Rio de Janeiro: E-papers Serviços Editoriais Ltda (2004)
- 8 R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

#### Referências



- 1 Rizzo, M. L. Statistical Computing with R Chapman and Hall/CRC, 2007, 416
- Sá, E. A. S.; de Moura, C. N.; Padilha, V. L. & Campos, C. G. C. Trends in daily precipitation in highlands region of Santa Catarina, southern Brazil. Ambiente e Agua An Interdisciplinary Journal of Applied Science, 2018, 13, 1-13
- Salviano, M. F.; Groppo, J. D. & Pellegrino, G. Q. Análise de Tendências em Dados de Precipitação e Temperatura no Brasil. Revista Brasileira de Meteorologia, 2016, 31, 64-73
- 4 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, 2013, 34, 5-12
- 5 Silva, D. D. da; Gomes Filho, R. R.; Pruski, F. F.; Pereira, Sí. B. & de Novaes, L. F. Chuvas intensas no Estado da Bahia. Revista Brasileira de Engenharia Agrícola e Ambiental, 2002, 6, 362-367
- 6 Sthephenson, A. G. evd: Extreme Value Distributions. R News, 2002, 2, 31-32
- 7 Thomas, M. et al. Applications of extreme value theory in public health. PLoS ONE, p. 1-7 (2016)
- (3) Wishcmeier, W. H.; Smith, D. D. Predicting rainfall erosion losses: a guide to conservation planning. U.S. Departament of Agriculture, p. 67, (1978)

Obrigado pela atenção!

artigo original em https://doi.org/10.1007/s42452-020-03199-8