Research Article

Generalized Pareto distribution applied to the analysis of maximum rainfall events in Uruguaiana, RS, Brazil

Amanda Larissa Alves Martins[1] · Gilberto Rodrigues Liska[2] · Luiz Alberto Beijo[3] · Fortunato Silva de Menezes[4] · Marcelo Ângelo Cirillo[4]

Abstract

The rainfall monitoring allows us to understand the hydrological cycle that not only influences the ecological and environmental dynamics, but also affects the economic and social activities. These sectors are greatly affected when rainfall occurs in amounts greater than the average, called extreme event; moreover, statistical methodologies based on the mean occurrence of these events are inadequate to analyze these extreme events. The Extreme Values Theory provides adequate theoretical models for this type of event; therefore, the Generalized Pareto Distribution (Henceforth GPD) is used to analyze the extreme events that exceed a threshold. The present work has applied both the GPD and its nested version, the Exponential Distribution, in monthly rainfall data from the city of Uruguaiana, in the state of Rio Grande do Sul in Brazil, which calculates the return levels and probabilities for some events of practical interest. To support the results, the goodness of fit criteria is used, and a Monte Carlo simulation procedure is proposed to detect the true probability distribution in each month analyzed. The results show that the GPD and Exponential Distribution fits to the data in all months. Through the simulation study, we perceive that the GPD is more suitable in the months of September and November. However, in January, March, April, and August the, Exponential Distribution is more appropriate, and in the other months, we can use either one.

Keywords Extreme value theory · Probability distribution · Rain amount · Inundation · Environmental concern

1 Introduction

Rainfall is vital for life on Earth [1], but its occurrence in high magnitude can cause damage and losses, usually causing flooding, destruction of buildings and crops, soil erosion, breaches of dikes and dams, among others [2, 3]. Damage in cities tends to be more severe because of the rapid urbanization and installation of complex infrastructure [4]. In addition, the frequency of extreme weather events has shown an increasing trend in various regions of the planet [2, 5]. In addition, the frequency of extreme weather events has shown an increasing trend in several regions of the planet [6–8], and the southern region of Brazil has suffered from the occurrence of these events [2, 5].

To minimize negative impacts or avoid economic, social and environmental losses, it is necessary to plan activities and constructions based on the probabilistic forecast of the occurrence of maximum precipitation in a given location [9]. For the forecasting process the fit of mathematical statistical models to the data, which can study the phenomena with different approaches, as well as the occurrence of extreme values, temporal distribution, spatial

✉ Gilberto Rodrigues Liska, gilbertoliska@ufscar.br; Amanda Larissa Alves Martins, amandamartins2503@gmail.com; Luiz Alberto Beijo, prof.beijo@gmail.com; Fortunato Silva de Menezes, fmenezes@dfi.ufla.br; Marcelo Ângelo Cirillo, macufla@des.ufla.br | [1]Academic of the Interdisciplinary Degree in Science and Technology, Federal University of Pampa, Itaqui, RS, Brazil. [2]Department of Agroindustrial Technology and Rural Socioeconomics, Federal University of São Carlos, Araras, SP, Brazil. [3]Federal University of Alfenas, Alfenas, MG, Brazil. [4]Federal University of Lavras, Lavras, MG, Brazil.

distribution, the intensity of the phenomenon, among others [10–12].

Statistical approaches based on the analysis of extreme values have shown promising results in the forecasting of these events in several areas of science [13–16]. One of the models extensively employed, for this purpose, in various scientific fields such as insurance, finance, meteorology, and the environment is the Generalized Pareto Distribution [17, 18].

Given the use of probabilistic models, assessing their goodness of fit is an equally important task. In the analysis of extreme events, this stage is practically not taken into account, even when it is a very consolidated methodology. Goodness of fit tests such as Kolmogorov-smirnov, chi-squared, and likelihood ratio are widely used [17, 19, 20]. However, as recommended by [21], the fit of the distributions using estimates of the parameters of the fitted distributions can lead to the occurrence of type II error, and, to circumvent this fact, proposes a simulation study. In general, these simulation studies are based on Monte Carlo procedures [22, 23].

Hence, the present work aims to fit the Generalized Pareto Distribution to the maximum monthly rainfall in the city of Uruguaiana, Rio Grande do Sul state, Brazil, as well as to calculate the probability of some extreme events occurring, calculate return levels of extreme rainfall events and its confidence intervals in periods of 2, 5, 10, 30, 50 and 100 years.

2 Methodology

The data set was obtained from the meteorological database for teaching and research (BDMEP), from January 1961 to April 2019, made available by the National Institute of Meteorology (INMET) and registered at the Uruguaiana—Rio Grande do Sul state weather station. The data are grouped in monthly periods and in each month the threshold method is used. Consequently, the highest values of rainfall above a sufficiently high threshold have been estimated according to the POT (peaks over threshold) methodology. As a result, they are analyzed by Generalized Pareto Distribution.

According to Coles [24], as well as Generalized Extreme Values (Henceforth GEV) distribution is the limit distribution of the block maxima, and the GPD appears as the parametric form for limit distribution for threshold excesses, whose probability density function is given by

$$f(x|\xi,\sigma,u) = \begin{cases} \frac{1}{\sigma}\left[1+\xi\left(\frac{x-u}{\sigma}\right)\right]^{-\left(1+\frac{1}{\xi}\right)}, \ x \ge 0 \ \text{if} \ \xi \ne 0 \\ \frac{1}{\sigma}\exp\left(-\frac{x-u}{\sigma}\right), \ 0 \le x \le \frac{1}{|\xi|} \ \text{if} \ \xi \to 0 \end{cases} \quad (1)$$

The distribution function is given by

$$F(x|\xi,\sigma,u) = \begin{cases} 1-\left[1+\xi\left(\frac{x-u}{\sigma}\right)\right]^{-\frac{1}{\xi}}, \ \xi \ne 0 \\ 1-\exp\left(-\frac{x-u}{\sigma}\right), \ \xi \to 0 \end{cases} \quad (2)$$

where u is the threshold, σ is the scale parameter and ξ the shape parameter. In priori, the threshold should be known and it is described in Sect. 2.1. The parameters σ and ξ must be estimated from the data and it is described in Sect. 2.2. Through the GPD distribution, three classes of standard distributions can be obtained: Type I: Exponential ($\lim_{\xi \to 0} F(x|\xi,\sigma,u)$), Type II: Pareto ($\xi > 0$) and Type III: Beta or ordinary Pareto ($\xi < 0$).

2.1 Threshold selection

To choose the appropriate threshold value, an exploratory graphical tool was used based on the linearity of the mean excesses function. This plot consists of the mean excesses above several thresholds with the threshold itself (Fig. 1). This plot is also known as mean residual life plot [25].

On the other hand, the mean residual life plot can be difficult to interpret as a threshold selection method. A complementary technique is employed, and it is based on fitting the GPD at a variety of thresholds, and on looking at the stability of the parameter estimates [24]. This plot is known as threshold choice plot (Fig. 2).
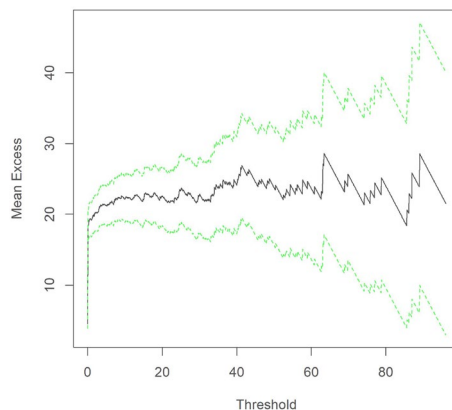

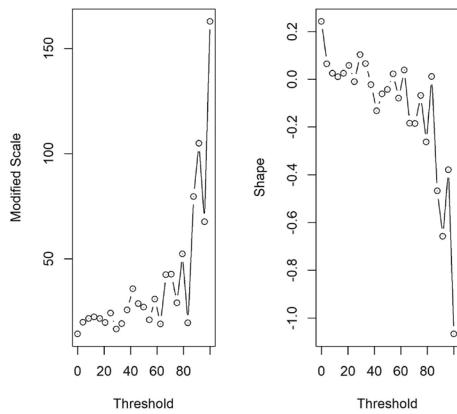**Fig. 1** Mean residual life plot for choosing a threshold

Fig. 2 Threshold choice plot for scale and shape estimated parameters

The choice of the very high threshold may result in a small number of observations, influencing the variance of the estimators. However, a threshold that does not satisfy the theoretical assumptions may result in distorted estimates. Thus, one should choose the threshold that makes the mean residual life plot and the functions of the parameters σ and ξ more or less linear [26].

2.2 Parameter estimation

After selection of the threshold, the GPD parameters were estimated by the maximum likelihood method. The maximum likelihood estimators maximize the log-likelihood function. Suppose $y_1, \ldots, y_k$ are the k excesses of a threshold u [24]. For $\xi \neq 0$

$$l(\sigma, \xi) = -k \log(\sigma) - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{k} \log\left(1 + \xi \frac{x_i}{\sigma}\right), \quad (3)$$

where $(1 + \sigma^{-1} \xi x_i) > 0$ for $i = 1, \ldots, k$; in other way, $l(\sigma, \xi) = -\infty$. In the $\xi \to 0$ case, the log-likelihood function is given by

$$l(\sigma) = -k \log(\sigma) - \frac{1}{\sigma} \sum_{i=1}^{k} x_i. \quad (4)$$

The maximum likelihood estimators of parameters σ and ξ are obtained through the solution of the homogeneous equations, given by partial derivatives of log-likelihood with respect each parameter. The estimation of σ and ξ requires the use of a numerical maximization, usually any method for this works, like Newton–Raphson, Simulated Annealing, Fisher's scoring or its variations [27].

2.3 Hypothesis testing

With the parameters estimated, goodness of fit criteria of the GPD model were evaluated. The Kolmogorov Smirnov (KS) test was used to compare the theoretical cumulative distribution and the empirical cumulative distribution [28]. The Ljung Box (LB) independence test, whose statistics are compared with the α-th quantile of the chi-squared distribution with one degree of freedom. The Mann-Kendall test was used to determine if the series has a statistically significant time trend [29]. When very small values of p-value are found, it indicates evidence in favor of the alternative hypothesis, that is, there is some tendency to modify the behavior of the analyzed series.

For the maximum likelihood estimates, one can test if ξ is statistically null. Then, to test the null hypothesis that the extremes distributions is exponential, we use the likelihood ratio test (LT), whose test statistic is

$$\Lambda = 2\left[l(\hat{\sigma}, \hat{\xi}) - l(\hat{\sigma})\right], \quad (5)$$

where $l(\hat{\sigma})$ and $l(\hat{\sigma}, \hat{\xi})$ represent the log-likelihoods respectively using the Exponential and GPD densities with the respective maximum likelihood estimates [26]. Thus, the null hypothesis that $\xi = 0$ is rejected if Λ is greater than the α-th quantile of the chi-squared distribution with 1 degree of freedom. Alternatively, if the p-value of the test is less than the significance level, the null hypothesis is rejected. For all tests we adopt 1% as significance level

2.4 Probability of excesses and return levels

According to Eq. 2 in the $\xi \neq 0$ case, to estimate the probability of occurrence of precipitation above a threshold, we have that

$$\Pr\left[X > x | X > u\right] = \left[1 + \xi\left(\frac{x - u}{\sigma}\right)\right]^{-\frac{1}{\xi}}. \quad (6)$$

However, in equation 6 it calculates the probability of occurrence of a given maximum precipitation that is higher than the adopted threshold. It is desired to calculate the probability of occurrence of precipitation above a maximum value. Therefore, equation 6 is simplified in

$$\Pr\left[X > x\right] = \lambda\left[1 + \xi\left(\frac{x - u}{\sigma}\right)\right]^{-\frac{1}{\xi}}, \quad (7)$$

where $\lambda = \Pr\left[X > u\right]$. Hence, the level x_m that is exceeded on average once every m observations is the solution of

$$\lambda\left[1 + \xi\left(\frac{x_m - u}{\sigma}\right)\right]^{-\frac{1}{\xi}} = \frac{1}{m}. \quad (8)$$

Therefore, the equation 8 leads to the m-observation return level. For representation, it is often more convenient to give return levels on an annual scale, so that the N-year return level is the level expected to be exceeded once every N years. If there are n_x observations per year, this corresponds to the m-observations return level, where $m = N \times n_x$ [24]. Hence, the N-Year return level is defined by

$$\hat{z}_N = \hat{u} + \frac{\hat{\sigma}}{\hat{\xi}}\left[\left(Nn_x\hat{\lambda}\right)^{\hat{\xi}} - 1\right] \tag{9}$$

where n_x is the number of days to be analyzed. We analyzed monthly rainfall data, so $n_x = 31, 30, 28$ days, according to month. If $\xi \to 0$, the return level is defined by

$$\hat{z}_n = \hat{u} + \hat{\sigma} \log\left(Nn_x\hat{\lambda}\right). \tag{10}$$

For the estimates of return level, we need to know the estimates of the parameters of the GPD. As a result, to estimate the probabilities and return level, the maximum likelihood estimates will be used, as described in the previous sections. Thus, an estimate for λ is required, which has the following natural estimator

$$\hat{\lambda} = \frac{k}{n} \tag{11}$$

corresponding to the proportion of the sample points exceeding u. In addition to the return level estimates, the confidence intervals with confidence coefficient $(1 - \alpha) \times 100\%$, associated with the return periods of 2, 5, 10, 30, 50 and 100 years, were constructed using the delta method, as described in Coles [24]. Since the number of excesses of u follows a binomial distribution, $\hat{\lambda}$ is also the maximum likelihood estimate of λ. The confidence intervals for $\hat{z}_N$ can be obtained by the delta method, but the uncertainty in the estimate of $\hat{\lambda}$ should also be included in the calculation. From the standard properties of the binomial distribution, $Var\left(\hat{\lambda}\right) \approx \hat{\lambda}(1 - \hat{\lambda})/n$, then the complete variance-covariance matrix is approximately

$$V = \begin{bmatrix} \hat{\lambda}(1 - \hat{\lambda})/n & 0 & 0 \\ 0 & v_{1,1} & v_{1,2} \\ 0 & v_{2,1} & v_{2,2} \end{bmatrix} \tag{12}$$

where $v_{i,j}$, represents the term (i, j) of the variance-covariance matrix of $\hat{\sigma}$ and $\hat{\xi}$. Thus by the delta method,

$$Var\left(\hat{z}_N\right) \approx \nabla z_N^T \, V \, \nabla z_N \tag{13}$$

where

$$\nabla z_N^T = \left[\frac{\partial z_N}{\partial \lambda}, \frac{\partial z_N}{\partial \sigma}, \frac{\partial z_N}{\partial \xi}\right] \tag{14}$$

evaluated in $\left(\hat{\lambda}, \hat{\sigma}, \hat{\xi}\right)$. Therefore, the confidence interval $(1 - \alpha) \times 100\%$ for $\hat{z}_N$ is given by

$$CI_{(1-\alpha)\times100\%}\left(\hat{z}_N\right) = \hat{z}_N \pm z_{\frac{\alpha}{2}}\sqrt{Var\left(\hat{z}_N\right)}, \tag{15}$$

where $z_{\frac{\alpha}{2}}$ is the $\frac{\alpha}{2}$-th quantile of the standard normal distribution.

2.5 Simulation study to evaluate goodness of fit for extreme values distributions

A computational simulation study was conducted with the purpose of evaluating the performance of the distributions in each month. For this, the Monte Carlo simulation method was used, which consists of making several achievements of a phenomenon according to pre-established parameters. At the end of these simulations, we can calculate the mean and standard deviation of the simulations and these represent measures of accuracy and precision, respectively [30, 31]. For each month, the data series was divided into a training series, comprising 30 years (1961–1991), and a test series, comprising 29 years (1992–2019). Thus, two scenarios are considered: (1) the first scenario generates samples of the Exponential distribution with the estimated parameters, and (2) the second scenario generates samples of the GPD with the estimated parameters.

Each scenario [$(k = (1), (2)$] is repeated 10000 times, according to the Monte Carlo simulation procedure, following the steps described below:

(i) With the training sample, generate a sample of the same size (n) according to the probability distribution of scenario k;

(ii) Estimate the parameters of the Exponential and GPD distributions using the maximum likelihood method, described in Sect. 2.2;

(iii) Perform the likelihood ratio test of step (ii);

(iv) For the return periods of 2, 5, 10, 15, 20, 25, 28 years, calculate the respective return level with the probability distributions and their respective parameters estimated in step (ii);

(v) With the test sample, obtain the observed return levels for the return periods of 2, 5, 10, 15, 20, 25, 28 years. Calculate the Mean Absolute Percentage Error (MAPE) and the Root Mean Squared Error (RMSE), given by equations 16 and 17, respectively.

$$RMSE = \sqrt{\frac{\sum_{i=i}^{n_z}\left(z_{N_i} - \hat{z}_{N_i}\right)^2}{n_z}} \tag{16}$$

$$MAPE = \frac{1}{n_z} \sum_{i=1}^{n_z} \left| \frac{z_{N_i} - \hat{z}_{N_i}}{z_{N_i}} \right| \times 100 \qquad (17)$$

Steps from (i) to (v) are repeated 10000 times. After that, we obtain the Monte Carlos average from *MAPE* and *RMSE*. In addition, the following were calculated: the proportion of which the LT, in step (iii), resulted in a *p*-value higher than the significance level of 1%, denoted by $\hat{p}_{LT}$; the proportion of which the MAPE of the GPD is greater than the MAPE of the Exponential distribution, denoted by $\hat{p}_{MAPE}$; and the proportion of which the RMSE of the GPD is greater than the RMSE of the Exponential distribution, denoted by $\hat{p}_{RMSE}$. It should be noted that the adopted return times, 2, 5, 10, 15, 20, 25, 28 years, $n_z = 7$, comprise the time of the test series.

Finalizing the proposed methodology, we used the R software [32] and the *evd* package [33].

3 Discussion and results

Table 1 shows that in all months the exponential distribution ($\xi \to 0$) performs better by the likelihood ratio test. The Mann-Kendall test indicated no trend in all months of the year, since the *p*-values showed results higher than 0.01. That is, there are statistical indications that each series of monthly rainfall ceilings does not have a trend over the years. Furthermore, the series of monthly highs are independent, with 1% level of significance. We should highlight that we have used these tests to verify the assumptions of the Extreme Value Theory models, but that they could be used for other interests, such as [2, 29, 34] in the trend analysis of hydro-climatic series. In addition, the Kolmogorov-Smirnov test states that both distributions were fitted in all months and the QQ plots corroborate the results (Fig. 3). Satisfactory adjustment of the GPD distribution was also found by Lazoglou [35], Salleh and Hassan [36], Wan et al. [37], Zahid et al. [38].

From the fit of the exponential distribution, we verify, in Table 2 that in the months of October to February and April to May, amounts of rainfall above 50 mm are recorded, with a probability of occurrence greater than

Table 1 Threshold ($\hat{u}$) selected by procedure described in Sect. 2.1, parameter estimates and Hypothesis tests (*p*-value) of the Generalized Pareto (GPD) and Exponential distributions for monthly maximum rainfall data of the city of Uruguaiana, RS, Brazil

| Month	Probability distribution	$\hat{u}$	$\hat{\sigma}$	$\hat{\xi}$	LT	KS	LB	Mann Kendall
January	Exponential	40	24.1319	–	0.883	0.774	0.242	0.7413
	GPD	40	24.6996	−0.0239		0.762		
February	Exponential	40	27.875	–	0.773	0.772	0.821	0.4654
	GPD	40	29.2355	−0.0486		0.723		
March	Exponential	35	27.7219	–	0.377	0.717	0.266	0.0555
	GPD	35	31.9006	−0.1463		0.459		
April	Exponential	50	29.7306	–	0.376	0.951	0.374	0.4456
	GPD	50	35.3042	−0.1832		0.9		
May	Exponential	45	30.9034	–	0.941	0.602	0.288	0.5735
	GPD	45	30.3336	0.019		0.637		
June	Exponential	30	20.4484	–	0.051	0.113	0.222	0.1738
	GPD	30	14.1875	0.3039		0.527		
July	Exponential	35	13.3269	–	0.788	0.848	0.099	0.0291
	GPD	35	12.2025	0.087		0.889		
August	Exponential	35	16.8577	–	0.179	0.083	0.021	0.7407
	GPD	35	12.977	0.227		0.039		
September	Exponential	23	20.0213	–	0.216	0.378	0.675	0.2254
	GPD	23	24.3738	−0.2069		0.181		
October	Exponential	40	21.7176	–	0.482	0.478	0.27	0.6963
	GPD	40	19.4466	0.1053		0.752		
November	Exponential	40	20.4262	–	0.578	0.526	0.758	0.0415
	GPD	40	21.9398	−0.0755		0.684		
December	Exponential	40	21.1594	–	0.973	0.665	0.763	0.6555
	GPD	40	21.0502	0.0052		0.677		
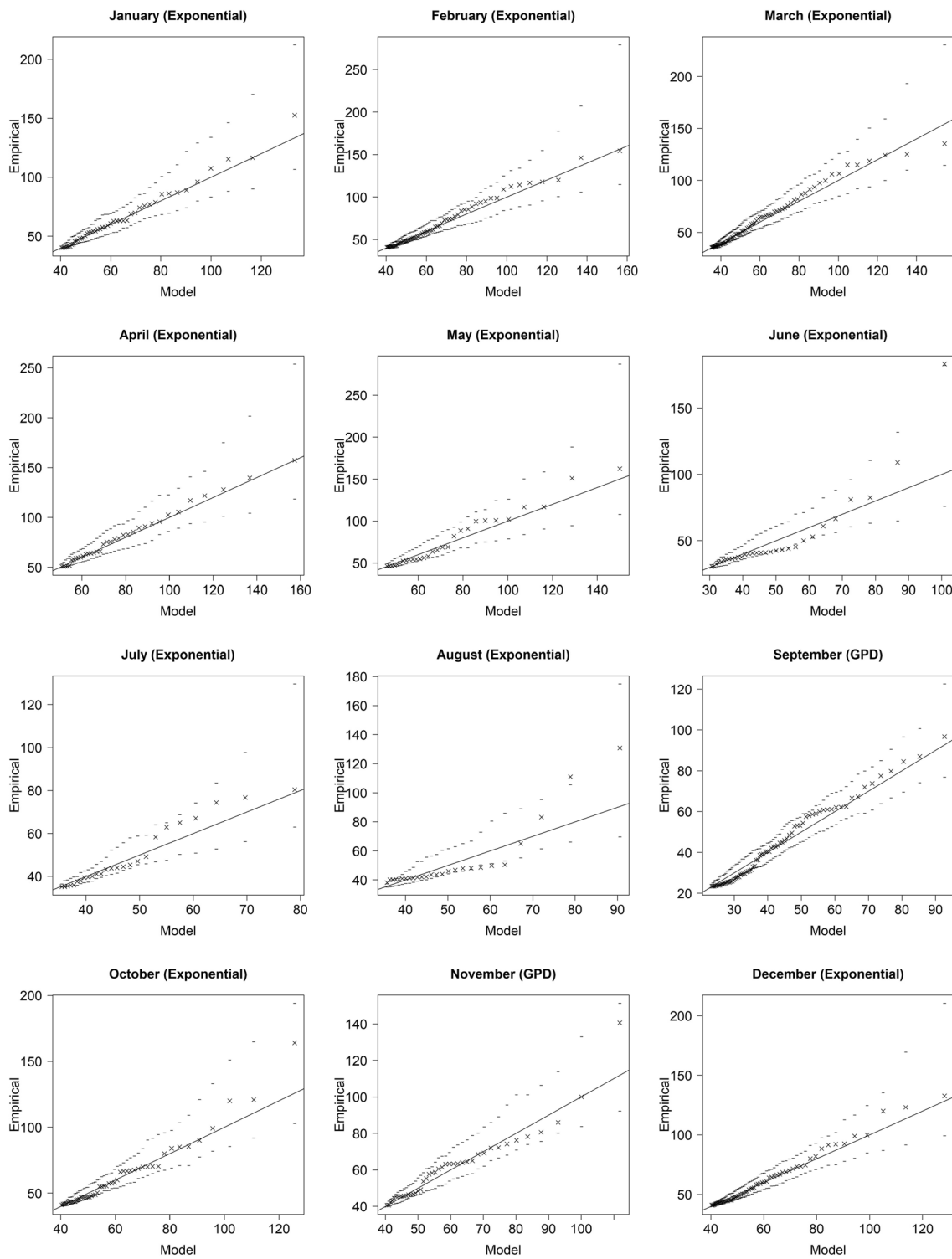
Fig. 3 Q-Q plots of the best fitted distribution showed by the comparison of Tables 4 and 5 for monthly maximum rainfall data of the city of Uruguaiana, RS, Brazil. Dashed lines represent the 95% confident interval, the points represent the empirical return level and estimated by the fitted model, solid line represent the 1 to 1 relation by return levels

60%. The table also shows that the probability of rainfall above 150 mm is higher in April and May than in other months of the year.

Rain volumes between 100 mm and 180 mm in a few hours can lead to landslides and flooding. One example occurred in the city of Rolante, metropolitan region of

Table 2 Probability (%) of rainfall occurrence by the probability distributions for monthly maximum rainfall data of the city of Uruguaiana, RS, Brazil

Month	Probability Distribution	Amount of precipitation (mm)				
		50	75	100	125	150
January	Exponential	66.07	23.45	8.32	2.95	1.05
	GPD	66.58	23.65	8.19	2.76	0.9
February	Exponential	69.86	28.49	11.62	4.74	1.93
	GPD	70.83	29.13	11.51	4.35	1.57
March	Exponential	58.21	23.62	9.59	3.89	1.58
	GPD	61.44	25.03	8.9	2.63	0.6
April	Exponential	100	43.13	18.6	8.02	3.46
	GPD	100	46.84	19.4	6.78	1.84
May	Exponential	85.06	37.88	16.87	7.51	3.35
	GPD	84.83	37.54	16.82	7.63	3.5
June	Exponential	37.6	11.07	3.26	0.96	0.28
	GPD	30.93	10.85	4.91	2.59	1.52
July	Exponential	32.45	4.97	0.76	0.12	0.02
	GPD	31.1	5.59	1.26	0.34	0.1
August	Exponential	41.07	9.32	2.12	0.48	0.11
	GPD	35.83	9.66	3.52	1.55	0.78
September	Exponential	25.96	7.45	2.14	0.61	0.18
	GPD	28.42	5.99	0.6	0.01	0
October	Exponential	63.1	19.96	6.31	2	0.63
	GPD	60.61	19.24	6.91	2.74	1.18
November	Exponential	61.29	18.02	5.3	1.56	0.46
	GPD	62.89	18.27	4.67	1.02	0.18
December	Exponential	62.34	19.13	5.87	1.8	0.55
	GPD	62.22	19.1	5.9	1.84	0.58

Porto Alegre, which has an average rainfall of 180 mm accumulated. landslides caused by a flood reached an area of 230 hectares and more than 6,600 inhabitants, and mud were dragged by the river, causing a cutoff of the water supply in eight municipalities of the region [39].

Herrmann [40] reported that in November 1991 there was precipitation in only two days with accumulated above 400 mm in São José / SC. There were numerous landslides and deaths in the eastern mountain range of Santa Catarina since houses crashed down and several sections of the highway BR 101 were blocked by the collapse of barriers. In December 1995, heavy rainfall resulted in 29 deaths, causing 29 municipalities in the mesoregion of southern Santa Catarina to declare a state of calamity.

Table 3 presents estimates of maximum rainfall return levels for periods of 2 to 100 years for each month. We monitored that by means of the fit of the GPD and exponential distributions, that the precipitation estimates increase as the time of return increases. This fact is already expected and is in agreement with Zahid et al. [38].

In the period from September to May, rainfall above 50 mm is recorded, which depending on the hourly intensity may cause erosive processes in the soil, which can

become harmful in order to contribute to the removal of essential nutrients for the development of the crop [25].

In March, it is expected that the maximum rainfall return level of 154.01 mm is exceeded once in 50 years by the Exponential distribution. Medeiros et al. [41] found for the same month a return level of 124.33 mm by the Gumbel distribution in the municipality of Jataí-Goiás and report that high levels of precipitation daily can cause intense rainfall and that estimates of precipitation in different return periods can be useful for assist professionals involved with planning and execution of hydraulic structure projects in decision making in control of floods.

Zahid et al. [38] conducted a study on temperatures return levels in concluded that extreme temperatures can affect yields. The crops are very sensitive to temperature variations in the order of 1 °C, according to Hatfield & Prueger [42]. Every harvest has a certain temperature tolerance limit. When the temperature exceeds this limit, the yield of the harvest is drastically reduced. The same goes for extreme rainfall.

The results indicate that the month of April presented the highest rainfall return levels, whose expected level is 156.96 mm in an average period of 50 years. As a way of providing greater precision in the results, Beijo et al. [43]

Table 3 Return Levels estimates (mm) by the probability distributions for monthly maximum rainfall data of the city of Uruguaiana, RS, Brazil

Month	Probability distribution	Time of return (years)					
		2	5	10	30	50	100
January	Exponential	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
February	Exponential	67.97	93.51	112.83	143.46	157.69	177.02
	GPD	68.65	93.58	111.76	139.79	152.49	171.05
March	Exponential	64.78	90.18	109.39	139.85	154.01	173.23
	GPD	66.73	90.18	105.82	127.88	136.67	149.79
April	Exponential	61.26	88.5	109.11	141.77	156.96	177.57
	GPD	62.94	90.69	109.09	133.4	143.2	155.29
May	Exponential	49.55	77.86	99.28	133.23	149.02	170.44
	GPD	49.46	77.58	99.22	133.71	149.47	173.03
June	Exponential	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
July	Exponential	35.63	47.84	57.08	71.72	78.52	87.76
	GPD	35.57	47.25	56.73	72.96	80.89	92.18
August	Exponential	35.88	51.33	63.01	81.53	90.14	101.83
	GPD	35.68	49.05	61.19	84.6	97.89	117.3
September	Exponential	45.29	63.64	77.51	99.51	109.74	123.62
	GPD	47.24	63.4	73.73	87.43	92.8	99.41
October	Exponential	56.28	76.18	91.24	115.1	126.19	141.24
	GPD	55.18	75.43	92.09	120.66	134.59	152.27
November	Exponential	51.14	69.85	84.01	106.45	116.88	131.04
	GPD	51.72	70.37	83.64	103.28	112.03	123.35
December	Exponential	60.54	79.93	94.6	117.84	128.65	143.32
	GPD	60.49	79.93	94.6	117.85	128.65	143.32

calculated the maximum rainfall return levels in Lavras, Minas Gerais state, by type I extreme values distribution (Gumbel), and found that for an average period of 50 years, expected level is 148 mm and with a 95% confidence that varies between 131 mm and 164 mm. These authors also recommend that, in the analysis of maximum precipitation, if the interest is in the maximum extreme event, it is suggested that the upper limit of the interval be used as a reference value. In this sense, the Fig. 4 shows the behavior of the return levels and their 95 % confidence intervals.

Rain shall be considered erosive and individual as long as they are greater than or equal to 10 mm or greater or equal to 6.0 mm, provided that they occur in a maximum of 15 minutes and separated from each other by a period of at least six hours with a rainfall of 1.0 mm or less [44].

As seen in Table 1, the likelihood ratio test attests that the Exponential distribution is sufficient to model rainfall data and in a few months the Kolmogorov Smirnov test indicated that the GPD distribution is more appropriate, by comparing its *p*-values. If two probability distributions from the same family fit a set of data, the one with the least number of parameters is preferable [45]. This fact is important when there are problems in estimating the parameters of models, which can occur in methods based on

likelihood [13, 46, 47]. In our study, this fact did not occur, which allows us to conduct the simulation study referred to in the Sect. 2.5. We conclude that there are months in which the Exponential distribution is more adequate, as in the months of January, March, April and August, since most of the comparison criteria used are favorable to this distribution. In September and November, most criteria indicated that the GPD distribution is more appropriate (Tables 4 and 5).

In the months of February, May, June, July, October and December, the result was inconclusive, as there was no unanimity between the two distributions in the two scenarios evaluated (Tables 4 and 5). In that case, we can use any of the distributions. We should emphasize that the Exponential distribution is expected to present a better result in the first scenario and GPD in the second scenario. When this does not occur, there is a strong indication that the true distribution in that month is that which was unanimously elected by the adopted criteria.

Regarding simulation studies involving distributions of extreme values, Xavier et al. [48] have reported in their simulation studies involving the generalized extreme values distribution, in the presence of covariates to model trend or temporal effect. The one that is more parsimonious is
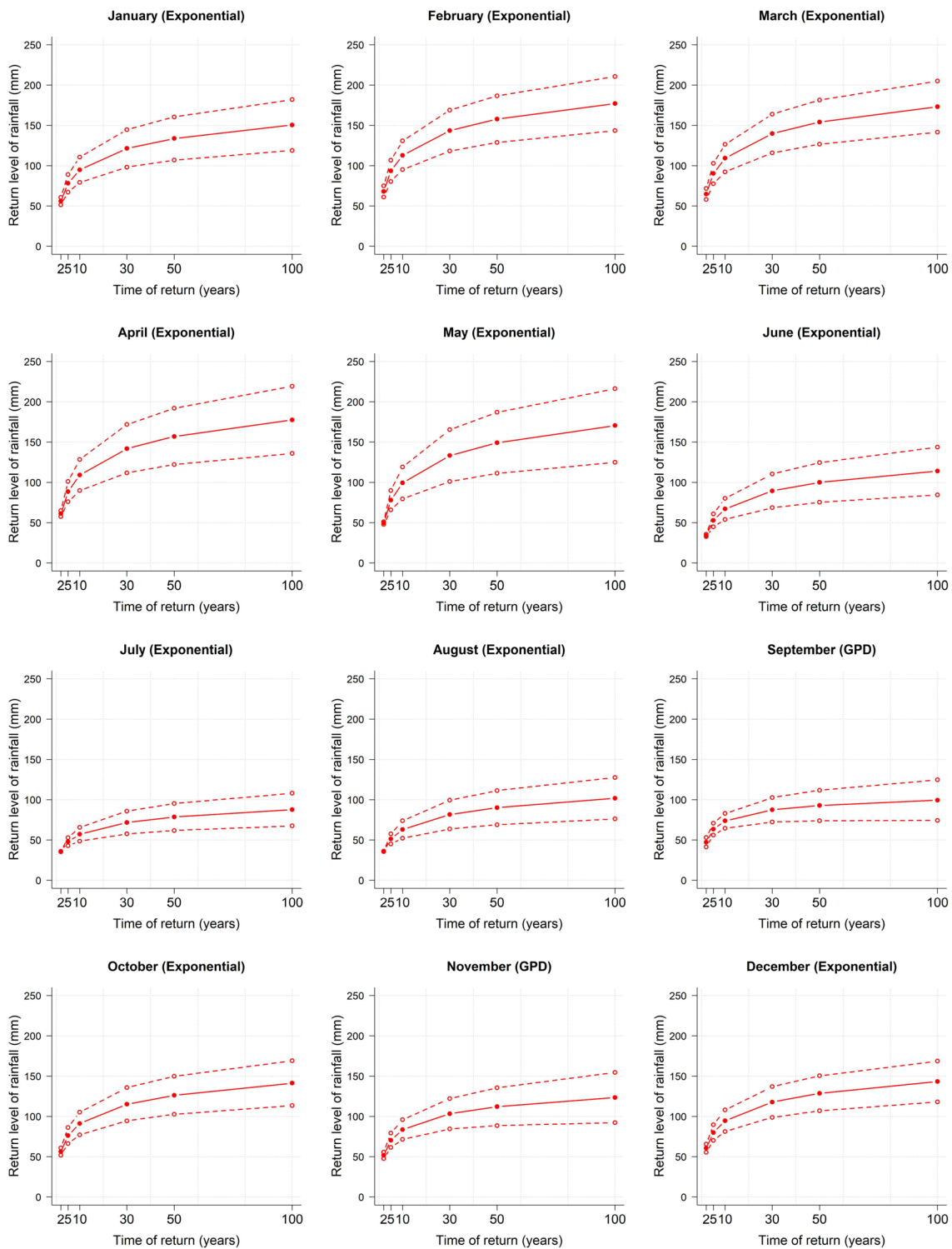
Fig. 4 Return level plot (in years) and confidence intervals for monthly maximum rainfall data of the city of Uruguaiana, RS, Brazil. Dashed lines represent the 95% confident interval and solid line represent estimated return level by the best distribution showed by the comparison of Tables 4 and 5

preferable and that according to the subject of study, the method used to select models is an important issue. In the same sense, Kim et al. [49] have showed by Monte Carlos simulation that the model comparison methods behave differently in the evaluation of stationary and nonstationary GEV models. For the nonstationary case,

Table 4 Results of scenario 1 for the Monte Carlo simulation in 10000 replicates for each month of the year for the Exponential and GPD distributions of monthly maximum rainfall data in Uruguaiana-RS

Month	Probability distribution*	MAPE	RMSE	$\hat{p}_{MAPE}$(%)	$\hat{p}_{RMSE}$(%)	$\hat{p}_{LT}$(%)
January	*Exponential*	23.99	37.11	54.31	54.31	94.42
	GPD	24.05	37.23			
February	*Exponential*	33.18	20.69	51.04	49.32	94.80
	GPD	33.19	20.72			
March	*Exponential*	20.92	27.70	54.70	54.76	94.58
	GPD	21.01	27.79			
April	*Exponential*	22.59	44.44	54.17	56.19	94.70
	GPD	22.68	44.50			
May	Exponential	15.14	19.57	44.69	44.69	94.55
	<u>GPD</u>	15.00	19.38			
June	*Exponential*	56.70	105.47	54.57	54.57	94.86
	GPD	56.74	105.54			
July	Exponential	53.30	25.42	45.96	45.97	94.94
	<u>GPD</u>	53.19	25.36			
August	*Exponential*	89.39	25.72	50.06	48.36	94.80
	GPD	89.38	25.73			
September	Exponential	25.32	15.25	45.91	45.91	94.79
	<u>GPD</u>	25.20	15.16			
October	*Exponential*	31.36	52.15	53.28	53.28	94.64
	GPD	31.41	52.26			
November	Exponential	13.19	12.65	46.00	45.92	94.72
	<u>GPD</u>	13.15	12.61			
December	*Exponential*	20.77	18.48	54.47	54.47	94.56
	GPD	20.86	18.59			

*Probability distribution in italic indicates that the Exponential distribution is better in respect month. Probability distribution in underline indicates that the GPD is better in respect month, *MAPE* Mean absolute percentage error, *RMSE* Root mean squared error. $\hat{p}_{LT}$; proportion of which the LT resulted in a *p*-value higher than the significance level of 1%, $\hat{p}_{MAPE}$; proportion of which the MAPE of the GPD is greater than the MAPE of the Exponential distribution, $\hat{p}_{RMSE}$; proportion of which the RMSE of the GPD is greater than the RMSE of the Exponential distribution

the Akaike information criteria showed better results and in the stationary case the likelihood ratio test was superior in detecting the most appropriate model. Our study used stationary GPD and we showed by Monte Carlo simulation that there are months when the most adequate distribution is different from that chosen in the Table 1. We intend to extend this study to other probability distributions.

Beijo et al. [50] stresses the importance of obtaining accurate estimates for rainfall. From a practical point of view, accuracy is important in terms of safety and economy, because when, in a shorter period, there is greater rainfall than expected, this can cause serious damage. In the case of the construction of a contour line, it would not support the volume of water and, consequently, would cause soil erosion and burial of plantations, causing serious damage to the environment and to the owners. Thus, and in accordance with the results of the Tables 4 and 5 , we provide the QQplots and confidence intervals for

return levels according to the most accurate probability distribution.

4 Conclusions

The Generalized Pareto distribution was satisfactorily fitted in all months and can be used to provide maximum rainfall extreme levels. No positive trend and temporal dependence of monthly maximum rainfall was found.

The rainfall estimates from January to December were calculated for the return periods of 2, 5, 10, 30, 50 and 100 years. The highest estimate was observed in April (with rainfall above 170 mm every 100 years and with 95% confident interval of 140 mm to 220 mm, approximately) and the lowest return level was in July (with rainfall near from 90 mm every 100 years).

Table 5 Results of scenario 2 for the Monte Carlo simulation in 10000 replicates for each month of the year for the Exponential and GPD distributions of monthly maximum rainfall data in Uruguaiana-RS

Month	Probability distribution*	MAPE	RMSE	$\hat{p}_{MAPE}$(%)	$\hat{p}_{RMSE}$(%)	$\hat{p}_{LT}$(%)
January	*Exponential*	24.03	37.19	77.86	77.87	89.19
	GPD	24.46	37.88			
February	Exponential	33.17	20.68	22.36	16.95	76.05
	<u>GPD</u>	32.90	20.41			
March	*Exponential*	20.15	26.97	100.00	100.00	1.26
	GPD	22.88	28.96			
April	*Exponential*	22.45	44.30	100.00	100.00	0.06
	GPD	25.64	45.86			
May	*Exponential*	15.16	19.61	65.69	65.68	92.11
	GPD	15.61	20.23			
June	Exponential	63.92	117.87	11.86	11.86	73.48
	<u>GPD</u>	63.58	117.32			
July	*Exponential*	53.59	25.56	97.94	97.95	39.20
	GPD	55.74	26.91			
August	*Exponential*	89.19	25.66	70.84	91.60	0.20
	GPD	89.49	26.25			
September	Exponential	25.85	15.63	0.00	0.00	0.01
	<u>GPD</u>	20.31	10.93			
October	Exponential	31.34	42.09	1.09	1.09	26.54
	<u>GPD</u>	30.03	49.72			
November	Exponential	13.14	12.61	1.69	1.34	48.83
	<u>GPD</u>	11.72	11.54			
December	Exponential	20.77	18.48	48.16	48.16	94.71
	<u>GPD</u>	20.74	18.45			

*Probability distribution in italic indicates that the Exponential distribution is better in respect month. Probability distribution in underline indicates that the GPD is better in respect month. *MAPE* Mean absolute percentage error, *RMSE* Root mean squared error. $\hat{p}_{LT}$; proportion of which the LT resulted in a *p*-value higher than the significance level of 1%, $\hat{p}_{MAPE}$; proportion of which the MAPE of the GPD is greater than the MAPE of the Exponential distribution, $\hat{p}_{RMSE}$; proportion of which the RMSE of the GPD is greater than the RMSE of the Exponential distribution

By comparing the distributions by computer simulation, it was possible to identify the true probability distribution of extreme values of the excess of a threshold. We chose three measures of fit quality to make the comparisons, and the measures $\hat{p}_{MAPE}$ and $\hat{p}_{RMSE}$ are obtained as a result. The proposed algorithm could be adapted for other measures of fit quality, such as the Akaike (AIC) information criterion, its corrected version (AICc), or Bayesian (BIC), among others. The length of the training and testing series is another issue that can be discussed. The original series should be as large as possible, but not less than 30 years. It is essential to have a balance between the sizes of the training and test series, so that if the training series is very long, the adjusted model can generalize well and, if the test set is long, the sample used to fit the model may be insufficient to reproduce the test series. In our work, for simulation, we divided the series into 30 years to adjust the model and 29 years to carry out the calculations of the appropriate quality measures, totaling 59 years of time series. The more extended set allows greater flexibility between the training and test series, and care has to be taken for short series, usually less than 30 years.

The results have practical implications for assessing the risk of extreme rain events in Uruguaiana, Brazil. The graphics are prepared to guide the local administration to support adaptations, such as the preparation of baseline contingency plans to deal with the maximum rainfall based on the current climatology. Studies like this are not yet available in this municipality. Our results will contribute to regional planning and may also be useful for ongoing economic and environmental projects in southern Brazil, as well as for a better understanding of the Pampa biome.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. da Pereira Britto FDR, Barletta R, Mendonça M (2006) Regionalização sazonal e mensal da precipitação pluvial máxima no estado do Rio Grande do Sul. Rev Bras de Climatol 2:35–52

2. Sá EAS, de Moura CN, Padilha VL, Campos CGC (2018) Trends in daily precipitation in highlands region of Santa Catarina, southern Brazil. Ambiente e Agua—an Interdiscip J Appl Sci 13:1–13

3. Batool N, Shah SA, Dar SN, Skinder S (2019) Rainfall variability and dynamics of cropping pattern in Kashmir Himalayas: a case study of climate change and agriculture. SN Appl Sci 1:606

4. Willems P, Arnbjerg-Nielsen K, Olsson J, Nguyen V (2012) Climate change impact assessment on urban rainfall extremes and urban drainage: methods and shortcomings. Atmos Res 103:106–118

5. Berlato MA, Cordeiro APA (2018) Signs of global and regional climate changes, projections for the twenty-first century and trends observed in Rio Grande do Sul state, Brazil: a review. Agrometeoros 25:273–302

6. Stocker T, Qin D, Plattner G-K, Tignor M, Allen S, Boschung J, Nauels A, Xia Y, Bex V, Midgley PIPCC (2013) Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change Cambridge University Press 2018:1585

7. Asadieh B, Krakauer NY (2015) Global trends in extreme precipitation: climate models versus observations. Hydrol Earth Syst Sci 19:877–891

8. Worku G, Teferi E, Bantider A, Dile YT (2019) Observed changes in extremes of daily rainfall and temperature in Jemma Sub-Basin, Upper Blue Nile Basin. Ethiopia. Theor Appl Climatol 135:839–854

9. Beharry SL, Gabriels D, Lobo D, Clarke RM (2019) A 35-year meteorological drought analysis in the Caribbean Region: case study of the small island state of Trinidad and Tobago. SN Appl Sci 1:1256

10. Ferreira HA, Liska GR, Cirillo MA, Borém FM, Ribeiro DE, Cortez RM, Guiraldeli CH (2016) Selecting a probabilistic model applied to the sensory analysis of specialty coffees performed with Consumer. IEEE Lat Am Trans 14:1507–1512

11. Sansigolo CA (2008) Distribuiç ões de extremos de precipitação diária, temperatura máxima e mínima e velocidade do vento em Piracicaba, SP (1917–2006). Rev Brasil de Meteorol 23:341–346

12. Pereira Britto F, Barletta R, Mendonça M (2008) Variabilidade espacial e temporal da precipitação pluvial no rio grande do sul: influência do fenômeno el niño oscilação sul. Rev Bras de Climatol 3:37–48

13. Butturi-Gomes D, Beijo LA, Avelar FG (2019) On modeling the maximum duration of dry spells: a simulation study under a Bayesian approach. Theor Appl Climatol 137:1337–1346

14. Byström HN (2005) Extreme value theory and extremely large electricity price changes. Int Rev Econ Finance 14:41–55

15. Cotta HHA, Correa WDSC, Albuquerque TTdA (2016) Gumbel distribution application for values of extreme precipitation in municipality of Vitória-ES. Rev Bras de Climatol 19:203–217

16. Thomas M et al (2016) Applications of extreme value theory in public health. PLoS One 11:1–7

17. Alam M, Emura K, Farnham C, Yuan J (2018) Best-fit probability distributions and return periods for maximum monthly rainfall in bangladesh. Climate 6:9

18. Salles TT et al (2019) Bayesian approach and extreme value theory in economic analysis of forestry projects. For Policy Econ 105:64–71

19. Yuan J, Emura K, Farnham C, Alam MA (2018) Frequency analysis of annual maximum hourly precipitation and determination of best fit probability distribution for regions in Japan. Urban Clim 24:276–286

20. dos Reis CJ, Beijo LA, Avelar FG (2017) Temperatura mínima esperada para Piracicaba-SP via distribuições de valores extremos. Rev Bras de Agric Irrig 11:1639–1650

21. Bautista EAL, Zocchi SS, Angelocci LR (2004) A distribuição generalizada de valores extremos aplicada ao ajuste dos dados de velocidade máxima do vento em Piracicaba, São Paulo. Brasil. Rev Mat Estat 22:95–111

22. Blain GC (2014) Revisiting the critical values of the Lilliefors test: towards the correct agrometeorological use of the Kolmogorov-Smirnov framework. Bragantia 73:192–202

23. Yao L, Dongxiao W, Zhenwei Z, Weihong H, Hui S (2014) A Monte Carlo simulation of multivariate general Pareto distribution and its application. Ocean Sci Discuss 11:2733–2753

24. Coles S (2001) An introduction to statistical modeling of extreme values. Springer, Great Britain

25. Silva AT, Portela MM, Naghettini M (2013) 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. Rev Recur Hídricos 34:5–12

26. DE Mendes BV M (2004) Introdução à análise de eventos extremos. E-papers Serviços Editoriais Ltda, Rio de Janeiro

27. Castillo E, Hadi AS, Balakrishnan N, Sarabia JM (2004) Extreme value and related models with applications in engineering and science. Wiley, New York, p 362

28. Blain GC (2014) Dry months in the agricultural region of Ribeirão Preto, state of São Paulo-Brazil: an study based on the extreme value theory. Eng Agrícola 34:992–1000

29. Salviano MF, Groppo JD, Pellegrino GQ (2016) Análise de tendências em Dados de precipitação e temperatura no Brasil. Rev Bras de Meteorol 31:64–73

30. Rizzo ML (2007) Statistical computing with R. Chapman and Hall, CRC, p 416

31. Robert C, Casella G (2010) Introducing Monte Carlo methods with R. Springer, New York

32. R Core Team (2019) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

33. Sthephenson AG (2002) Evd: extreme value distributions. R News 2:31–32

34. Tan ML, Samat N, Chan NW, Lee AJ, Li C (2019) Analysis of precipitation and temperature extremes over the Muda River Basin. Malaysia. Water 11:283

35. Lazoglou, G. & Anagnostopoulou, C. An Overview of Statistical Methods for Studying the Extreme Rainfalls in Mediterranean. Proceedings, 2017, 1, 681

36. Salleh NHM, Hasan H (2018) Generalized Pareto distribution for extreme temperatures in peninsular Malaysia. Sci Int (Lahore) 30:63–67

37. Wan Zin WZ, Jemain AA, Ibrahim K (2009) The best fitting distribution of annual maximum rainfall in Peninsular Malaysia

based on methods of L-moment and LQ-moment. Theor Appl Climatol 96:337–344

38. Zahid M, Blender R, Lucarini V, Bramati MC (2017) Return levels of temperature extremes in southern Pakistan. Earth Syst Dynam 8:1263–1278

39. G1. Deslizamentos de terra em Rolante atingiram 230 hectares. Avaiable in:http://glo.bo/2kvctLE. Accessed 20 May 2020

40. Herrmann MLP (2006) Atlas de desastres naturais do Estado de Santa Catarina. IOESC, Florianópolis, 1, 146

41. de Medeiros ES, Alves MA, de Souza SA (2019) Return level of estimation of maximum daily precipitation in the municipality of Jataí. Goiás. Ciência e Nat 41:e36

42. Hatfield JL, Prueger JH (2015) Temperature extremes: effect on plant growth and development. Weather Clim Extremes 10:4–10

43. Beijo LA, Muniz JA, Castro Neto P (2005) Maximum rainfall return period by extreme values type I distribution in Lavras, Minas Gerais state. Brazil Ciência e Agrotecnol 29:657–667

44. Wishcmeier WH, Smith DD (1978) Predicting rainfall erosion losses: a guide to conservation planning. Departament of Agriculture, U.S, p 67

45. Emiliano PC, MáJ Vivanco, de Menezes FS (2014) Information criteria: How do they behave in different models? Comput Stat Data Anal 69:141–153

46. Delicado P, Goria M (2008) A small sample comparison of maximum likelihood, moments and L-moments methods for the asymmetric exponential power distribution. Comput Stat Data Anal 52:1661–1673

47. Šimková T, Picek J (2017) A comparison of L-, LQ-, TL-moment and maximum likelihood high quantile estimates of the GPD and GEV distribution. Commun Stat Simul Comput 46:5991–6010

48. Xavier ACF, Blain GC, de Morais MVB, Sobierajski GdR (2019) Selecting "the best" nonstationary generalized extreme value (GEV) distribution: on the influence of different numbers of GEV-models. Bragantia 78:606–621

49. Kim H, Kim S, Shin H, Heo J-H (2017) Appropriate model selection methods for nonstationary generalized extreme value models. J Hydrol 547:557–574

50. Beijo LA, Vivanco MJF, Muniz JA (2009) Bayesian analysis for estimating the return period of maximum precipitation at Jaboticabal São Paulo state. Brazil. Ciência e Agrotecnol 33:261–270