



## ENGINEERING SCIENCES

# Regional Frequency Analysis applied to extreme rainfall events: Evaluating its conceptual assumptions and constructing null distributions

GABRIEL C. BLAIN, GRACIELA DA R. SOBIERAJSKI, ANA CAROLINA F. XAVIER & JOÃO PAULO DE CARVALHO

**Abstract:** Besides increasing the amount of data that can be used in a fitting process, the Regional Frequency Analysis (RFA) also assesses the quality of weather station networks. This technique assumes that it is possible to form homogeneous groups of meteorological series presenting independent and identically distributed data. Based on the hypothesis that such homogeneous groups can be formed under tropical-subtropical conditions, this study applied the RFA to assess the probability of one-day annual maximum rainfall in the State of São Paulo, Brazil. Critical limits used in previous studies to declare a region/group as 'acceptable homogeneous' ( $H \leq 1.00$ ) or to select a distribution ( $|Z| \leq 1.64$ ) were evaluated through Monte Carlo simulations. While the limit  $H \leq 1$  is appropriate, the limit  $|Z| \leq 1.64$  may lead to unacceptably high rates of rejecting a true null hypothesis. This statement is particularly true for the general logistic distribution. A computational algorithm allowing the selection of critical limits corresponding to pre-specified probabilities of rejecting a true null hypothesis is provided. Considering the new critical limits, data from one of the largest weather station networks of the State have been pooled into four homogeneous groups. Both generalized logistic and extreme value distributions are recommended for the probabilistic assessment of such groups.

**Key words:** Extreme value, fitting process, generalized logistic, homogeneous groups, tropical-subtropical region.

## INTRODUCTION

Improving the probabilistic assessment of extreme rainfall events has been a common goal for many statistical studies because such events pose serious hazards to human activities, human health, and the environment. However, estimating the probability of such events is a difficult task because - by definition - they occur at long return periods ( $>100$  years), which usually surpass the available length of at-sites rainfall records (Goudenhoofd et al. 2017). On such context, parametric distributions, such as the generalized extreme value distribution (GEV), the Pearson type III distribution (PE3) and the

generalized logistic (GLO), are frequently used to estimate the probability of such extremes (Khan et al. 2017). The parameters of these probabilistic functions are usually estimated from rainfall data recorded at individual weather stations, the so-called 'at-site approach'.

Regardless the parameter estimation procedure (e.g. maximum likelihood or L-moments), the above-mentioned approach is associated with relatively high levels of uncertainties that varies with the data availability. In general, the smaller the data availability, the larger the uncertainty level of both parameters and quantile estimates. Thus, increasing the amount of rainfall data that can be used to fit

the parameters of such parametric distributions is a key-step to improve the assessment of the probability of extreme rainfall events. On such context, the pooling of several at-sites rainfall series, which transfers information from nearby weather stations/sites to a target point, is an interesting alternative to reduce the uncertainty associated with the limited number of data at a single site (Svensson & Jones 2010, Goudenhoofdt et al. 2017). In other words, it is now well accepted that such a pooling – known as regional frequency analyses (RFA) – improves the evaluation of extreme weather events in respect to those based on at-site records because it increases the amount of data available for a fitting process.

Different statistical procedures, such as the index flood (Dalrymple 1960, Guttman 1993), may be used within the framework of RFA (Bradley 1998). This procedure establishes homogeneous groups, in which all sites/weather stations forming such groups are expected to present identical probability distributions, apart from a site-specific scaling factor (Hosking & Wallis 1997). These homogeneous groups can be established by means of group/regional L-moments, which are calculated as weighted average of sample L-moments for the sites forming each group (e.g. Hosking & Wallis 1997). By forming these homogeneous set of sites, a pooling of information (data) from all weather stations forming the groups can be gathered together in order to allow the estimation of the probability of rainfall events when individual series are too short to provide their reliable assessment (Fowler & Kilsby 2003). The distribution fitted from the group/regional L-moments is referred as to ‘the regional distribution’.

For identifying a homogeneous group under the framework of the RFA, initial clusters of sites must be proposed in such a way as to allow the calculation of the following summary statistics

(Hosking & Wallis 1997): Discordancy measure (D), heterogeneity measure (H) and goodness-of-fit measure (Z). Several approaches can be used for such purpose (Basu & Srinivas 2013); among then, the cluster analysis is frequently applied (Hosking & Wallis 1997). The initial clusters can be established taking into account only geographical information (e.g. latitude, longitude and altitude; Sung et al. 2018), descriptive statistics (e.g. mean, median or standard deviation), or both. When the cluster analysis is based only on geographical information, the above-mentioned summary statistics – particularly the H measure – may be used as an independent test of group homogeneity. In other words, while the cluster analysis can be used to set the initial number of groups, both D and H measures can be used to objectively evaluate the outcomes of this cluster analysis (Hosking & Wallis 1997) and to set the final number of homogeneous groups (this procedure is further described in next sections). The Z measure is used to select candidate distribution capable of describing the true underlying frequency distribution for a homogeneous group (Hosking & Wallis 1997).

From the statistical viewpoint, the RFA is carried out under the assumption that the data to be used to fit the regional distributions are independent and identically distributed (iid data; Buishand 1984). In other words, the RFA assumes that (i) the at-site records are randomly distributed (e.g. there is no serial correlation nor trends) and that (ii) there is no significant spatial correlation/dependence among the extreme data obtained from the sites forming a particular homogeneous group (Basu & Srinivas 2013). Regarding the RFA calculation algorithm, a group of weather station/sites is deemed homogeneous when the between site variability of the L-moments is statistically equal to that expected from homogeneous groups. In other

words, apart from a scaling factor, the sample L-moments of the sites forming a homogeneous group can only differ from each other due to sample variabilities. More specifically, a group of sites has been considered homogeneous when its H measure is equal to or lower than 1 (e.g. Hosking & Wallis 1997, Bradley 1998, Basu & Srinivas 2013, Sung et al. 2018). This critical limit ( $H \leq 1$ ) was initially proposed by Hosking & Wallis (1993). It was based on data from a climatic division of the continental US (North Cascades). However, such a limit has been used with little or no evaluation of its suitability in the different regions of the world. This latter statement also holds true for the Z measure.

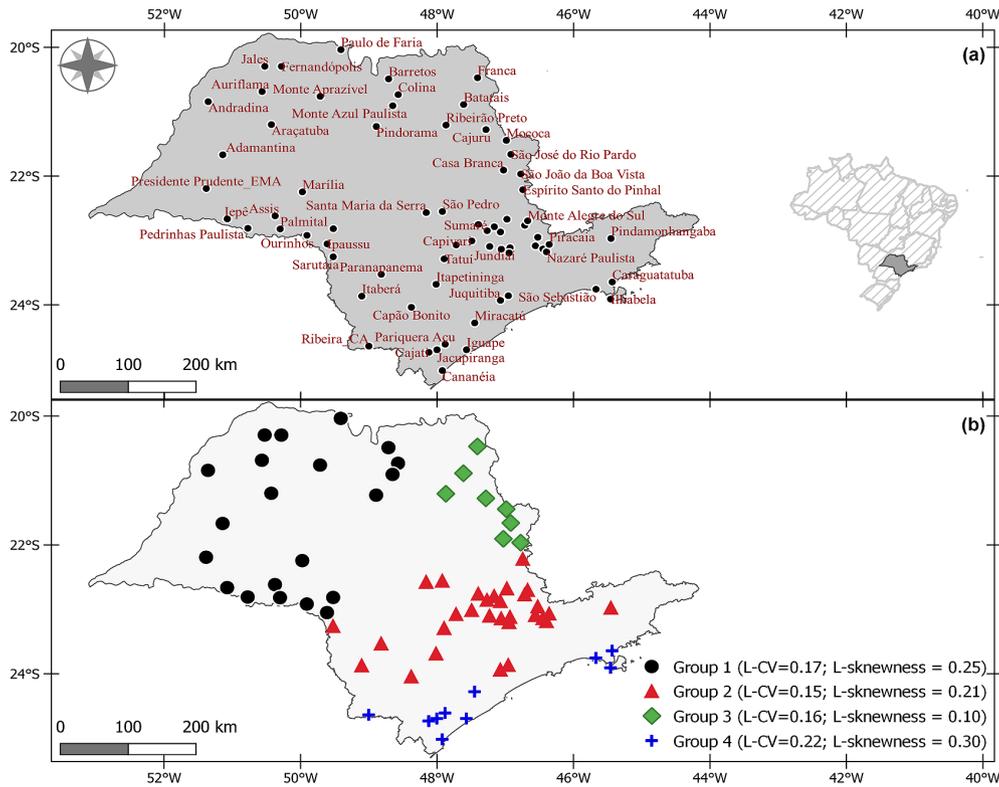
Besides increasing the amount of data that can be used in a fitting process, the RFA also assesses on a regional basis the quality of data obtained from each weather station (see next section). Naturally, this latter feature emphasizes the importance of applying the RFA in regions where the availability and quality of historical records are a matter of concern. On such context, a review on the RFA literature indicates that although this technique has been widely used in several parts of the World (Burn 1990, Hosking & Wallis 1997, Bradley 1998, Fowler & Kilsby 2003, Santos et al. 2011, Basu & Srinivas 2013, Sung et al. 2018), its application on tropical or subtropical regions of Brazil is yet to be properly evaluated.

Based on the hypothesis that under South America tropical-subtropical climate conditions it is possible to establish homogeneous groups in which the above-mentioned iid assumption is met, this study applied the RFA to assess the probability of occurrence of one-day annual maximum rainfall in the State of São Paulo, Brazil. Considering that there is no universally superior approach for setting the initial number of clusters (Hosking & Wallis 1997), this study proposed an interactive technique that sets

such a number by calculating two statistics of the RFA (D and H) along with a cluster analysis based only on geographical information. Finally, the critical limits often used to declare a group as homogeneous ( $H \leq 1.0$ ) and to select a candidate distribution ( $|Z| \leq 1.64$ ) have been evaluated through Monte Carlo simulations experiments. A computational algorithm that allows the users to select new critical limits for H and Z measures according to a pre-specified probability of rejecting a true null hypothesis (type I error) has been provided.

## MATERIALS AND METHODS

The state of São Paulo is situated within the coordinates 19°S and 26°S latitude and 53°W and 44°W longitude (crossed by the Tropic of Capricorn). The rainy season occurs during the austral summer and it is associated with the South Atlantic Convergence Zone (SACZ). In the winter season, the South Atlantic high-pressure system predominates (Vera et al. 2006). The daily rainfall data have been obtained from a weather station network belonging to the Centre of Agrometeorological information (CIIAGRO) of the Agronomic Institute of Campinas (IAC/APTA/SAA). Only series presenting more than 10 years of continuous records have been selected. Consequently, although the CIIAGRO's network presents more than 150 weather stations, only 84 have been considered in this study (Figure 1a). The length of the selected rainfall series varied from 10 to 60 years. The series presenting length of records equal to 60 years are Campinas, Jundiaí, Mococa, Monte Alegre do Sul, Paríquera\_Açu, Pindorama, Ribeirão Preto (Table I). These weather stations are situated at experimental farms of the Department of Agriculture of the State of São Paulo (APTA/SAA) and have been routinely used in both academic



**Figure 1. a) Weather Stations belonging to the Agronomic Institute of Campinas and subjected to the regional frequency analysis. b) The weather stations forming four homogeneous groups. Within each group, the one-day annual maximum rainfall observed at each site present the same frequency distribution apart from a scaling factor. Statistics describing, in terms of regional L-moments, the coefficient of variation (L-CV) and the Skewness (L-skewness) of the pooling of series forming each group, is also presented.**

and operational mode (Blain et al. 2017). Any missing daily record has been replaced by data extracted manually from pluviographs or from automatic weather stations situated at the same site. The percentage of missing records in each of these eight series is no greater than 2%. The other series (Figure 1a) present length of record ranging from 10 to 27 years. For each of these 77 series, we selected the largest possible continuous period of record presenting no more than 2% of missing data. From each rainfall series, the one-day maxima observed in each year have been collected generating 84 block maxima series. In other words, the so-called block maxima approach, as described in several studies (e.g. Coles 2001), was adopted in this study. As further described, all block maxima series have been subjected to a data quality

assessment based on the discordance measure. Therefore, we decided not to fill the missing records of the series presenting length of record between 10 and 27 years.

As previously described, this study used parametric distributions to define ‘rainfall extremes’ associated with a particular return period (Fowler & Kilsby 2003). As previously described, the block maxima approach has been used so that the parametric distributions were fitted from one-day annual maxima series. This latter approach tends to remove the influence of auto-correlation and seasonality in the modelling of extreme events. Further information on this approach, including its advantages and drawbacks, can be found in several studies such as Coles (2001).

**Table 1. Period of record (years), mean value (mm) and discordance measure (Di) of daily-extreme rainfall series of the State of São Paulo, Brazil. The Z measure of four candidate distribution – generalized logistic (GLO), generalized extreme value (GEV), generalized normal (GNO), Pearson Type III and generalized Pareto (GP) is also presented.**

Group 1 (H=0.22)				Group 2 (H=0.23)				Group 4 (H=0.37)			
Location	years	mean	D(i)	Location	years	mean	D(i)	Location	Years	mean	D(i)
Adamantina	1993-2017	84.3	0.74	Amparo	2009-2017	71.7	1.45	Cananéia	2000-2017	124.1	0.42
Andradina	1995-2017	87.2	0.29	Atibaia	2001-2017	68.7	0.23	Caraguatatuba	2007-2017	105.6	0.89
Araçatuba	1997-2017	99	0.18	BomJesusPerdões	2009-2017	63	0.52	Iguape	2004-2013	121.8	1.37
Assis	1991-2017	84.2	0.19	BragançaPaulista	2001-2017	68.1	0.17	Ilhabela	2006-2017	120.2	0.99
Auriflamma	1997-2017	79.7	0.52	Campinas	1958-2017	82.3	0.03	Jacupiranga	2000-2011	129.9	2.47
Barretos	1998-2017	80.2	2.38	Capão Bonito	1993-2017	81.2	0.29	Miracatú	2001-2017	91.8	0.24
Colina	2008-2017	68.7	1.57	Capivari	2001-2017	78.2	2.18	PariqueraAçú	1958-2017	112.8	0.56
Fernandópolis	2008-2017	76.9	1.27	EspíritoSãoPinhal	2002-2017	66	0.36	Ribeira	2006-2017	68.4	1.67
Iepê	2001-2017	84.9	0.26	Indaiatuba	2008-2017	64.4	0.89	São Sebastião	2007-2017	126.7	0.58
Ipaussu	2000-2017	80.5	0.24	Itaberá	2008-2017	75.3	0.99				
Jales	1994-2017	77.8	0.99	Itapetininga	1995-2010	89.7	1.58				
Marília	1993-2012	89.6	0.18	Itapeva	1995-2017	82.2	1.89				
Monte Aprazível	2003-2017	77.5	1.36	Jaguariúna	2008-2017	70.9	0.94				
Monte APaulista	2008-2017	67.5	2.5	Jundiaí	1958-2017	70.2	0.3	Zmeasure	Group 1	Group 2	
Ourinhos	2001-2017	82.3	0.63	Jundiaí-ETEC	2008-2017	79.4	2.05	GLO	-0.06	0.33	
Palmital	1994-2013	87.7	1.81	Juquitiba	2008-2017	69.3	0.91	GEV	-1.21	-1.18	

Table I. Continuation

Paulo de Faria	2008-2017	90.7	1.63	MonteAlegreSul	1958-2017	78	1.05	GNO	-1.75	-1.62	
Pedrinhas Paulista	1993-2014	105.9	0.46	Nazaré Paulista	2008-2017	591	2	PE3	-2.73	-2.48	
Pindorama	1993-2014	84.3	0.11	Parapanema	2008-2017	76.8	0.67	GP	-4.1	-4.76	
Pres.Prudente	2008-2017	84.2	2.35	Pau línia	1994-2017	73.6	1.82	Zmeasure	Group 3	Group 4	
StacruzRPardo	1993-2014	97.1	1.34	Pindamonhangaba	1993-2014	70.6	2.44	GLO	1.92	-0.12	
Group 3 ( <b>H=0.59</b> )											
Location	years	mean	D(i)	Piracaia	2001-2017	68.2	1.06	GEV	0.59	-0.69	
Batatais	2003-2017	68.7	0.82	StabárbaraOeste	2001-2017	67.3	0.29	GNO	0.67	-1.09	
Cajuru	2008-2017	62.7	1.36	StaMariadaSerra	2001-2017	93.8	2.09	PE3	0.52	-1.8	
Casa Branca	1996-2017	79.6	0.43	SlourençoSerra	2008-2017	74.7	0.81	GP	-2.12	-2.21	
Franca	2007-2017	80.5	1.21	São Pedro	2001-2017	91.7	0.52				
Mococa	1958-2017	84.7	0.22	Sarutaia	2008-2017	85.7	1.27				
Ribeirão Preto	1958-2017	79.3	1.46	Sumaré	2001-2017	72.4	0.58				
SIBoaVista	2006-2017	95.9	1.22	Tatuí	1993-2013	78	0.34				
SIRioPardo	2008-2017	77.7	1.28	Tietê	1993-2017	72.9	0.27				

### Trends and auto-correlation of the at-site series.

Independently and identically distributed data are expected to present significant trends nor serial correlations (Chandler & Scott 2011). Therefore, the Mann-Kendall and Wald-Wolfowitz Run test were applied to all extreme rainfall series. While the former test is widely used to check for the presence of trends in environmental series, the latter test is used to check for randomness or serial correlation. In this study, only series presenting no serial correlation nor trend were subjected to the RFA.

### Regional Frequency analysis: Summary statistics and initial clusters

#### Discordancy Measure

Within a pre-specified region or group, the D measure can be used to identify sites presenting a vector of L-moments ratios ( $u_i$ ; L-CV, L-skewness or L-kurtosis) significantly different from the pooling group (Hosking & Wallis 1997). The D measure is calculated through the following steps:

1<sup>st</sup> - Calculate the regional L-moments, which is an unweighted group average ( $U$ ) of the vectors  $u_i$  obtained from each site of the group.

$$U = N^{-1} \sum_{i=1}^N u_i \quad (1)$$

Where in  $N$  is the number of  $i$  sites or weather station forming the group.

2<sup>nd</sup> - Calculate the sample covariate matrix according to equation 2.

$$S = (N-1)^{-1} \sum_{i=1}^N (u_i - U)(u_i - U)^T \quad (2)$$

Finally, the discordance measure for each  $i$  site can be calculated by equation 3.

$$D_i = \frac{1}{3} (u_i - U)^T S^{-1} (u_i - U) \quad (3)$$

Considering that the  $U_i$  values were drawn from independent and identical multivariate normal distributions, critical values for  $D$  ( $D_{crit}$ ) can be approximated by equation 4.

$$D_{crit} = \min \left\{ 3, \frac{(N-1)\alpha^*}{(N-4+3\alpha^*)} \right\} \quad (4)$$

Wherein  $\alpha^*$  is the upper ( $100\alpha/N$ ) percentage of the F distribution with 3 and  $N-4$  degrees of freedom. The D measure was applied in two steps of the study. First, at the outset of the study (i.e. before the initial clusters have been delimited) the D measure was used to flag those sites that may present gross errors (Hosking & Wallis 1997). Later, the D measure was again applied within each pre-delimited group so that it was used to identify those weather stations that are discordant with the homogeneous group as a whole. Finally, it has to be emphasized that the D measure can only be properly interpreted for groups formed by seven or more sites (e.g. Santos et al. 2011).

#### Heterogeneity Measure and initial clusters

As previously described, within a homogeneous group, all sites are expected to present identical probability distributions, apart from a site-specific scaling factor. Therefore, the sample L-moments of the sites forming such a group are expected to differ from each other only because of sample variabilities (Hosking & Wallis 1997). On such background, a statistical test may be built by comparing, in terms of L-moments, the between-site variation with that expected for a homogeneous group. More specifically, consider a group with  $N$  weather stations. Also consider that each  $i^{th}$  weather station/site has record length equal to  $n_i$  and at site sample L-CV denoted by  $t^{(i)}_3, t^{(i)}_4$ . The group average L-moment ratios (L-CV, L-skewness and L-kurtosis) can be calculated by equation 5. Note

that the contribution of each  $t^{(i)}$  to the final value of  $t^r$  is weighted according to each  $n_i$  value.

$$t^R = \frac{\sum_{i=1}^N n_i t^{(i)}}{\sum_{i=1}^N n_i} \quad (5)$$

A measure of the between-site variation, in terms of L-CV, can be estimated by equation 6.

$$V = \frac{\sum_{i=1}^N n_i (t^{(i)} - t^R)^2}{\sum_{i=1}^N n_i} \quad (6)$$

Hosking & Wallis (1997) recommended the following steps to compare the between-site variation with that expected for a real homogeneous group.

Step 1: Fit a “highly flexible” distribution to the regional average L-moments ratios. With “highly flexible” Hosking & Wallis (1997) meant a function capable of mimicking a range of different distributions, such as the generalized extreme value, the Pearson type III and the generalized logistic distributions, which are usually used to fit extreme rainfall data. As previously described, the four-parameter Kappa distribution fitted through the L-moments method is widely used for such a purpose (Hosking & Wallis 1997, Bradley 1998, Santos et al. 2011, Basu & Srinivas 2013, Sung et al. 2018).

Step 2: Use the Kappa distribution to simulate a large number of extreme rainfall series. The number and length of the simulated series must be same as those of the sites/ weather stations forming the group. Since all synthetic series were generated by the same (Kappa) distribution their pooling is, by definition, homogeneous. They also present no serial nor cross correlations.

Step 3: Use equation 6 to calculate the mean ( $\mu_{v.sim}$ ) and the standard deviation ( $\sigma_{v.sim}$ ) of all values of the simulated group of series. The V.SIM values are obtained from equation 6. The H measure is then calculated by equation 7.

$$H = \frac{V - \mu_v}{\sigma_v} \quad (7)$$

A group of series is regarded ‘acceptably homogeneous’ when  $H < 1.0$ ; ‘possible heterogeneous’ if  $1 < H < 2$  and ‘definitely heterogeneous’ when  $H > 2$  (Hosking & Wallis 1997). Although it is possible to use the H statistic as a formal hypothesis test, accurate significance levels can only be obtained if the iid assumption of the series is met and the true ‘regional distribution’ is Kappa (Hosking & Wallis 1997). As previously described, the critical limit  $H \leq 1.0$  has been largely used to declare a region or group as ‘acceptable homogeneous’. The suitability of such a limit has been evaluated through Monte Carlo Simulation (described below). Finally, alternative measure of between-site dispersion can also be used. For instance, equation 7 may also be based on the relationship between L-CV and L-skewness or between L-skewness and L-kurtosis. However, such alternative measures lack power to discriminate between heterogeneous/homogeneous groups (Hosking & Wallis 1997), hence they were not considered in this study.

The initial number of groups were established through an interactive procedure that took into account cluster analyses (CA) and both D and H measures. The CA has been performed by using latitude, longitude and altitude of each weather station and the Ward’s hierarchical method (Hosking & Wallis 1997). All geographical information has been normalized. Finally, the spatial dependence of the sites forming the homogeneous groups has been assessed for all pairs of sites through Kendall’s tau (Genest & Favre 2007).

### Goodness-of-fit measure (Z)

Within a homogeneous group, the Z measure was used to evaluate the hypothesis that a candidate

distribution is the true underlying frequency distribution for such a group. Considering a three-parameter distribution fitted by the L-moments method the quality of such a fit may be evaluated by the following equation.

$$Z = \frac{\tau_4 - t_4^R + B_4}{\sigma_4} \quad (8)$$

Where  $t_4^R$  is the group average L-Kurtosis; is the L-Kurtosis of the candidate distribution and is the bias of  $t_4^R$ .

Once the iid assumption is met, the Z measure is expected to present a standard normal distribution under the hypothesis that the candidate model is the true 'regional distribution' (Hosking & Wallis 1997, Sung et al. 2018). Therefore, a candidate distribution is frequently selected when  $|Z| \leq 1.64$ , which is expected to correspond to the 10% significance level (Sung et al. 2018). The following widely used three-parameter distributions have been considered in this study: Generalized Extreme Value distribution (GEV), the Pearson type III distribution (PE3) the Generalized Logistic (GLO) and the Generalized Pareto (GPA). The assumption that, under  $H_0$ , the distribution of Z follows the standard normal distribution was also evaluated through sets of Monte Carlo simulations experiments.

### Assessing uncertainties and constructing the 'regional/group growth curve'

Although the Z measure had been designed to identify candidate distributions capable of yielding accurate quantile estimates for the sites composing a homogeneous group, it does not give an estimate of uncertainties in quantile/return-period predictions of the 'pooled growth curve'. Thus, an estimate of uncertainty in return-period predictions is required for providing a measure of confidence in the use of the 'regional distribution' (Fowler & Kilsby 2003).

The pooled uncertainty measure (PU; Robson & Reed 1999) quantifies the differences between each site growth curve factor and the pooled growth factor (equation 9). In this study, the PU has been calculated for the following return-periods: 2, 5, 10, 20, 50, 100 and 1000 years. It was also multiplied by the group average of the AM series to reflect the group average uncertainty in millimetres (Fowler & Kilsby 2003) (equation 9).

$$PU_T = \sqrt{\frac{\sum_{i=1}^N n_i (\ln x_T - \ln x_T^p)^2}{\sum_{i=1}^N n_i}} \quad (9)$$

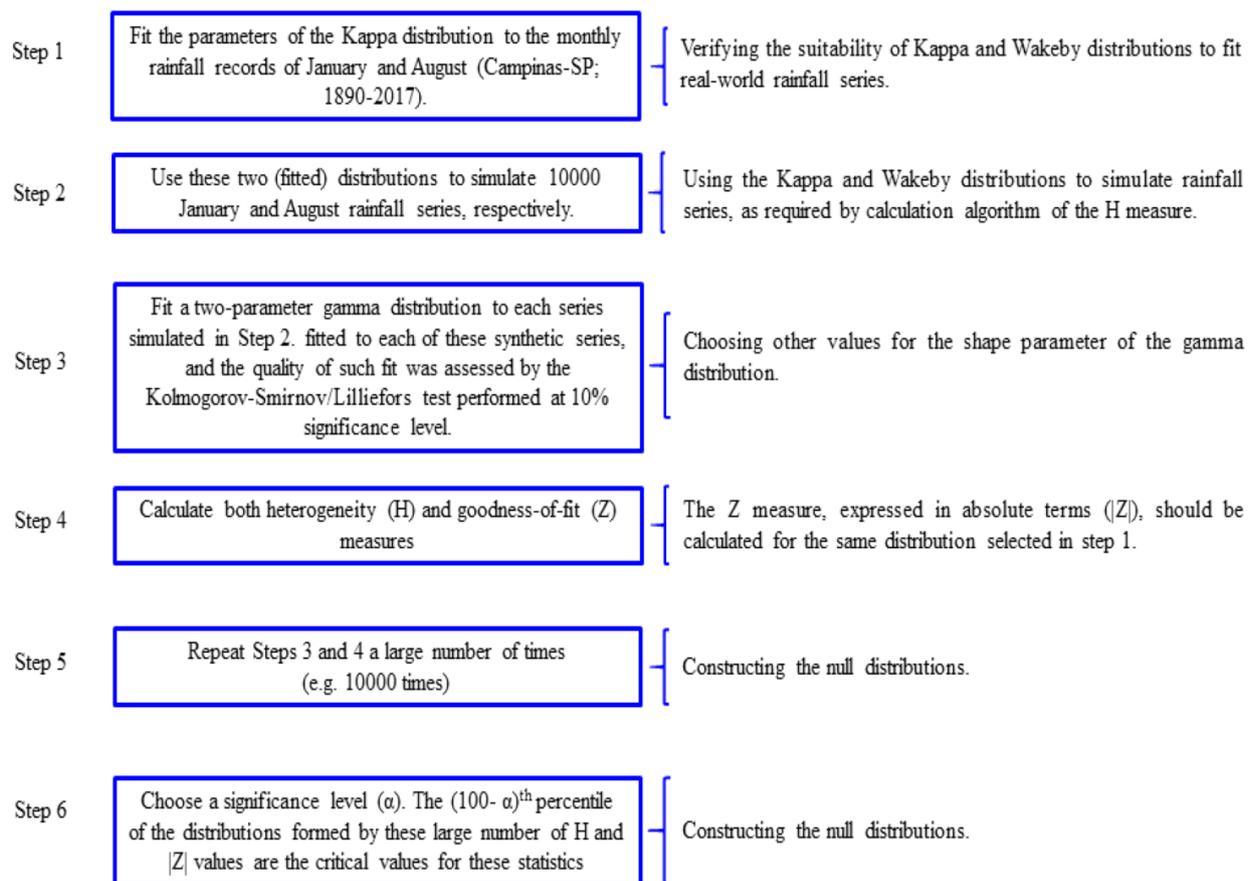
Where  $i$  is the  $i^{\text{th}}$  site,  $N$  is the number of sites within the homogeneous group,  $n_i$  is the record length (years),  $X_T$  is the T-year growth factor and  $X_T^p$  is the T-year pooled growth factor. By definition, the quantile function  $q(F)$  of the 'regional distribution' – called regional growth curve – may be regarded as a dimensionless function common to all sites forming the homogeneous group. The quantile function of each site  $i$ , can be calculated by:

$$Q_i(F) = \mu_i q(F) \quad (10)$$

Where  $\mu$  is the index flood of the  $i^{\text{th}}$  site, which usually assumes the value of the corresponding sample mean.

### Null-distributions of the H and Z measures: Monte Carlo simulations and R-codes

As previously described, the critical limit  $H \leq 1.0$  for declaring a group as 'acceptable homogeneous' was initially proposed by Hosking & Wallis (1993) and it has been applied to different regions of the world with little or no evaluation of its suitability. This lack of evaluations is also true for the assumption that under  $H_0$  the distribution of the Z measure approaches the standard normal distribution. Therefore, the following Monte Carlo simulation experiment



**Figure 2. Computational algorithm allowing the selection of critical limits corresponding to pre-specified probabilities ( $\alpha$ ) of rejecting a true null hypothesis.**

is proposed to evaluate the suitability of both critical limits by constructing null distributions for H and Z measures.

Step 1: Select one of the three-parameter distribution used in this study.

Step 2: Use the selected distribution to simulate homogeneous groups of at-site extreme rainfall series. The number of homogeneous groups, the number of at-site series forming the groups and the length of each series considered in the sets of Monte Carlo simulations were the same as those obtained from the CIAGRO (Figure 1b, next section).

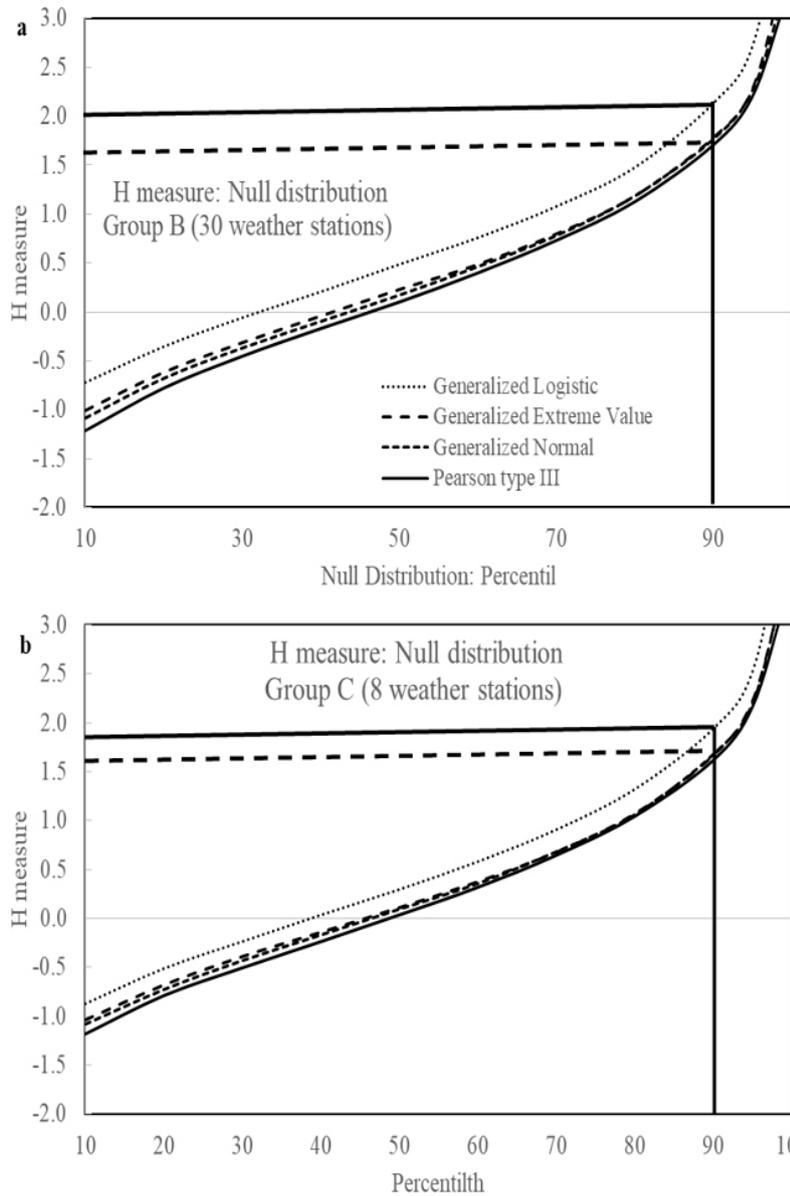
Step 3: Calculate both H and Z measures for each homogeneous groups.

Step 4: Repeat steps 2 and 3 a large number of times (e.g. 10000 trials in each Monte Carlo experiment).

Step 5: From the results of step 4 calculate the percentage in which  $H < 1.0$  and the percentage in which Z selected the same distribution used in step 2.

Step 6: Select another three-parameter distribution and repeat steps 2 to 5.

The use of the information depicted in Figure 1b in the step 3 of the Monte Carlo simulation facilitates the comparison between the simulated results and those found from the real data of the State of São Paulo. The simulated group that used the information from group 1 (2, 3 or 4) was named group A (B, C or D). Q-Q plots (Wilks 2011) have also been



**Figure 3.** Null Distribution of the Heterogeneity Measure (H). The vertical and horizontal lines indicate the 90th percentile of the null distributions, which in a formal hypothesis test would correspond to 10% a significance level. The null distributions have been constructed considering four different three-parameter distributions and different number of weather stations forming the homogeneous groups.

constructed to evaluate the departures of the null distribution of the Z measure from the normal standard distribution. Finally, since these controlled simulations follow the main steps proposed by Hosking & Wallis (1997), they can be easily applied in other regions or to new datasets. Thus, we present an R-code (Figure 2 and Supplementary Material - Appendix) that may be used to perform all steps of the Monte Carlo simulations and, consequently, to

construct null-distributions for both H and Z measures. The R-software is a free widely used computational environment ([www.r-project.org](http://www.r-project.org)).

**RESULTS AND DISCUSSION**

**Trends, serial correlations and the first use of the D measure**

The Mann-Kendall test has been widely used in both meteorological and agrometeorological

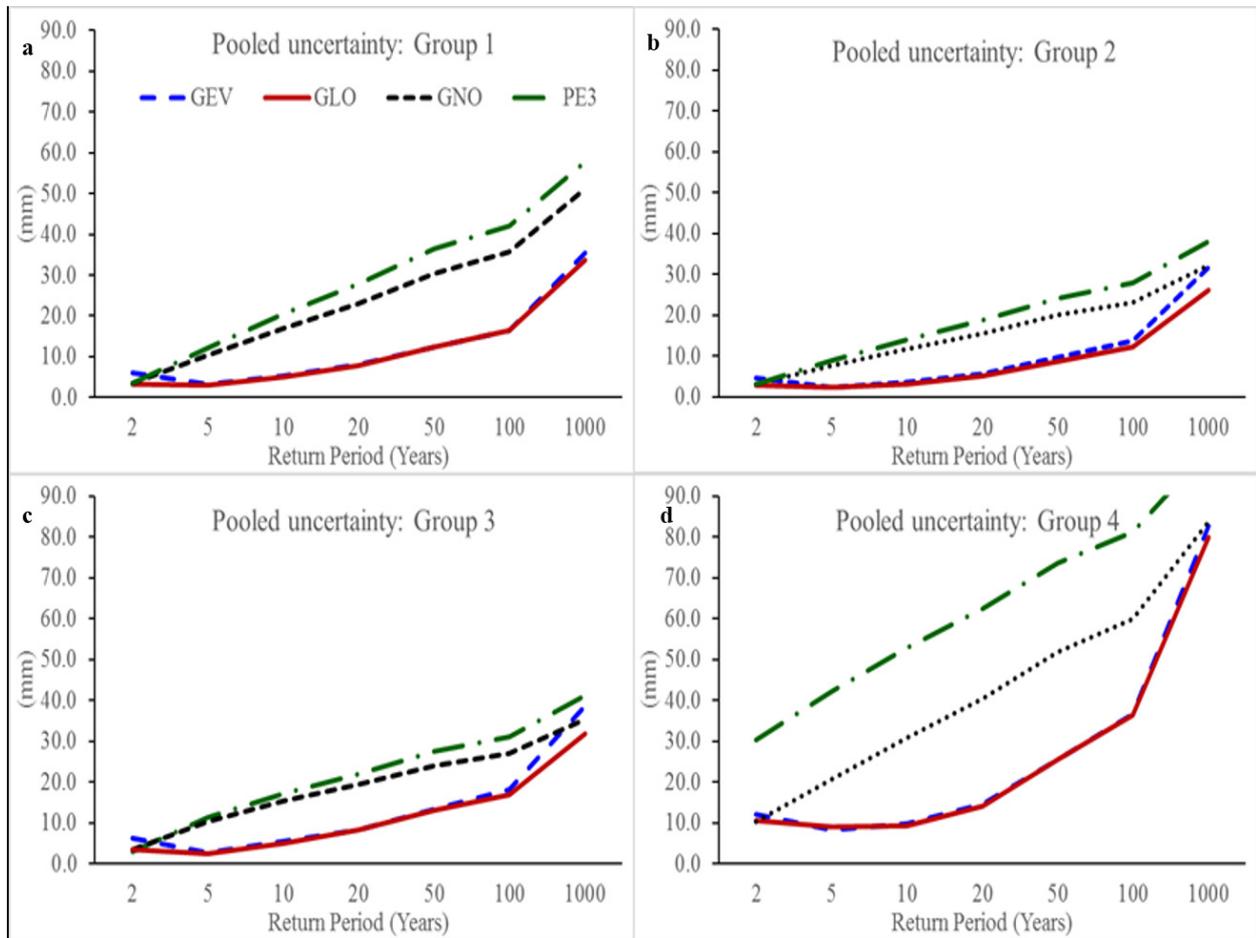
studies to detect trends (Chandler & Scott 2011). However, from a strictly mathematical standpoint, the rejection of its null hypothesis can only be taken as an evidence that the data are not independent and identically distributed (Chandler & Scott 2011). Therefore, of the 84 series depicted in Figure 1a, 16 series (~15%) did not meet the iid assumption, according to the MK test. On the other hand, and considering that block maxima series are expected to present no serial correlation nor seasonality, of the 84 series depicted in Figure 1a, 68 can be regarded as free from trends at both 5 and 10% significance level (Chandler & Scott 2011). These 68 series can also be regarded as free from serial correlation, since the null hypothesis of the Wald-Wolfowitz Run test could not be rejected for any of them.

As previously described, the D measure can also be applied before forming the homogeneous groups to flag those sites presenting gross errors (Hosking & Wallis 1997). Two out of these 70 series presented D measures larger than the critical value. As previously described, when applied at the outset of the analysis, i.e. before forming the homogeneous groups, this latter measure of discordance can be used to flag sites, which may present gross errors. Hence the RFA were applied to the remaining 70 series (Figure 1b; next section).

The initial number of groups required to initialize a cluster analyses (CA) is always an arbitrary choice. In this study, this initial number was set to seven. Nevertheless, as described below, the RFA can objectively address the outcome of the CA. Thus, it provides a statistical basis for setting the final number of clusters/groups in which the weather stations will be grouped. This initial number (seven) formed groups representing less than seven weather stations. Since the summary statistics of the RFA should only be applied to groups/regions with more than seven weather stations (Santos et al.

2011), the CA was again initialized with the rainfall series clustered into six groups. Again, this latter number formed groups with less than seven weather stations. The CA was again performed with the rainfall series clustered into five groups. Since all clusters presented more than seven weather stations both D and H measures could be applied to objectively evaluate the outcomes of the CA. Although all groups could be regarded as 'acceptably homogeneous', two groups presented negative H values (Group 1: 12 sites  $H=-0.33$ ; Group 2: 30 sites  $H=0.23$ ; Group 3: 9 sites  $H=-0.15$ ; Group 4: 7 sites  $H=0.60$  and Group 5: 10 sites  $H=0.36$ ). This suggests that there is less dispersion among at-site sample L-CVs values than what is expected from homogeneous groups composed by iid series (Hosking & Wallis 1997). Thus, the CA was again performed with the rainfall series clustered into four groups. All clusters presented more than seven weather (Table I) and all groups may be regarded as 'acceptably homogeneous' since all of them presented  $H < 1$  (positives values). The D measure identified no discordance site within the groups (Table I). Therefore, testing the validity of the critical limit ( $H < 1$ ) proposed by Hosking & Wallis (1993) becomes the final step to declared these four groups as 'acceptable homogeneous'.

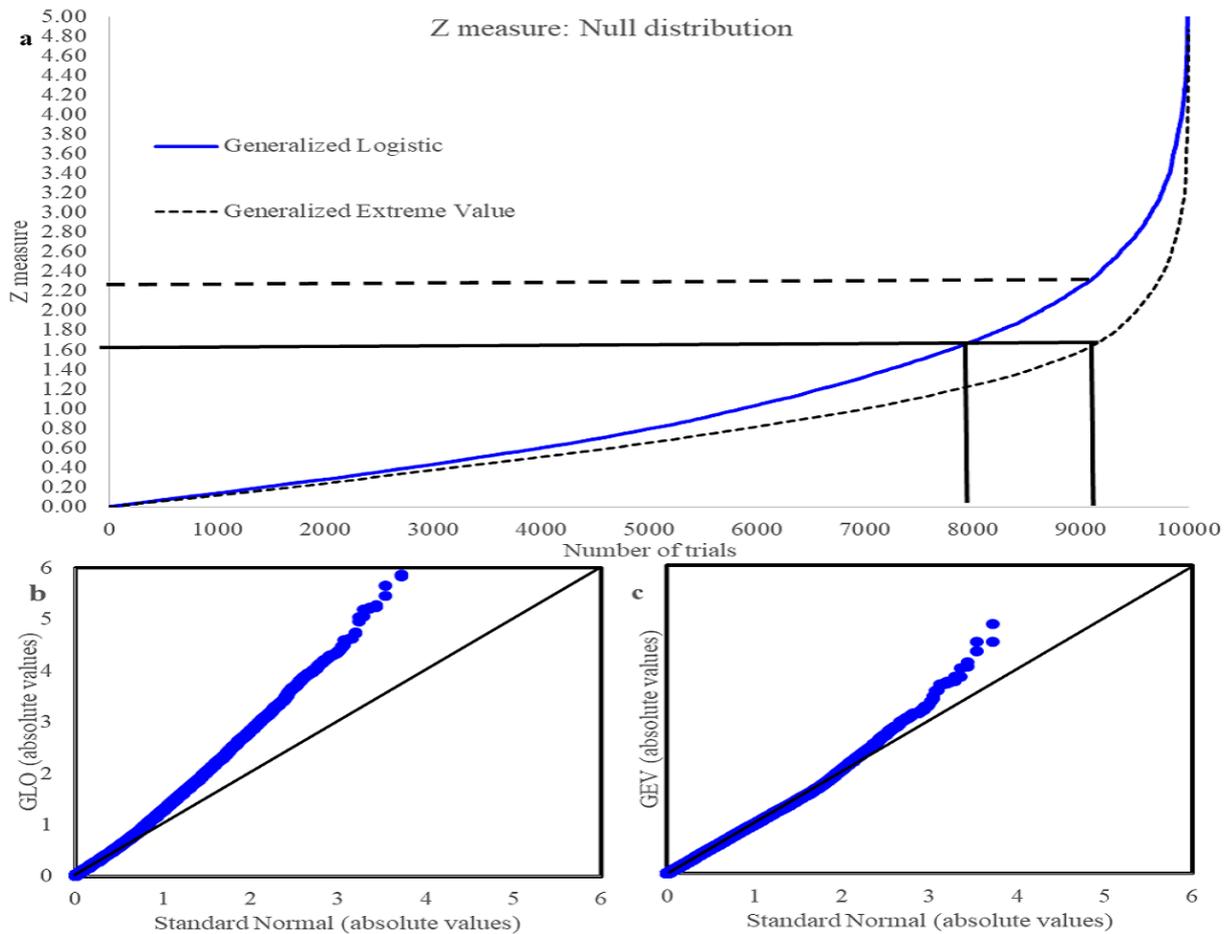
As previously described the suitability of the  $H \leq 1$  limit was assessed by constructing null distributions for this heterogeneity measure. As exemplified in Figure 3a, b for groups B (that presents the highest number of weather stations; 30) and C (that presents the lowest number of weather stations; 8), the critical limit  $H \leq 1$  is, in general terms, a strict threshold regardless the true underlying regional distribution. This statement is supported by the fact that such critical value always fell before commonly used probability levels of rejecting a true null hypothesis - the 90<sup>th</sup> or 95<sup>th</sup> percentile of the null distributions, which in a formal hypothesis



**Figure 4.** Pooled uncertainty (mm) for different return periods and for different parametric distributions: generalized extreme value (GEV), generalized logistic (GLO), generalized normal (GNO) and Pearson type III (PE3). Groups 1 to 4 are presented in Figure 1.

test would correspond to 10 or 5% significance level (Wilks 2011; Figure 3a, b). In addition, the probability of falsely rejecting the hypothesis of homogeneity reaches its highest levels when the series forming the homogeneous group comes from a GLO distribution (Figure 3a, b). For this latter distribution, the null hypothesis was falsely rejected at rates of 32, 36, 28 and 30% ( $H > 1.0$ ; Groups A to D; respectively). These rates are higher than those observed for the other distributions, such as the GEV: 27, 28, 21, and 25% (GEV; Groups A to D; respectively). Since the results depicted in Figure 3a, b also holds true for the other two groups (A and D), we may assume that, at least for the state of

São Paulo, the critical limit  $H \leq 1$  is a conservative threshold for declaring a group of extreme-rainfall series as ‘acceptably homogeneous’. It is also worth mentioning that the  $H$  measures of the four groups (Figure 1b and Table I) fell close to the central part of the null distribution regardless the three-parameter function used to simulate the homogeneous groups. Naturally, this latter feature has increased our confidence that the groups presented in Figure 1b are indeed homogeneous. Finally, the Kendall’s tau correlation indicated the presence of no spatial dependence among the sites forming each of the four groups. Therefore, we may assume that the series depicted in Figure 1b meet the



**Figure 5.** Null Distribution of the goodness-of-fit (Z) Measure. a) The vertical and horizontal solid lines represent the percentiles of the null distributions associated with the original critical value (1.64). The vertical and horizontal dashed lines represents the 90th percentile of the null distribution, which in a formal hypothesis test would correspond to 10% a significance level. The null distributions have been constructed considering two different three-parameter distributions; b) and c) depict the Q-Q plot for Z values drawn from these two distributions.

iid assumption (Basu & Srinivas 2013) and, considering both D and H measures, they can be pooled together into four homogeneous groups (Figure 1b).

With regards to regional L-moments, the analysis of Figure 1b also suggests that the temporal variability of the one-day extreme rainfall data of groups 1, 2 and 3 are similar to each other, since the L-CV values found for these three groups are close to each other (0.17, 0.15, 0.16, respectively). However, among these three groups, group 1 experienced more intense

rainfall events than groups 2 and 3. This latter statement is based on the fact that, among these three groups, the highest L-skewness value (0.25) was found in group 1 (Fowler & Kilsby 2003). As expected, the group formed by weather stations situated in the coastal area of the State presented, among the four groups, the highest L-CV and L-skewness value. In other words, this latter region experienced the highest temporal variability of the one-day extremes with the most intense rainfall events.

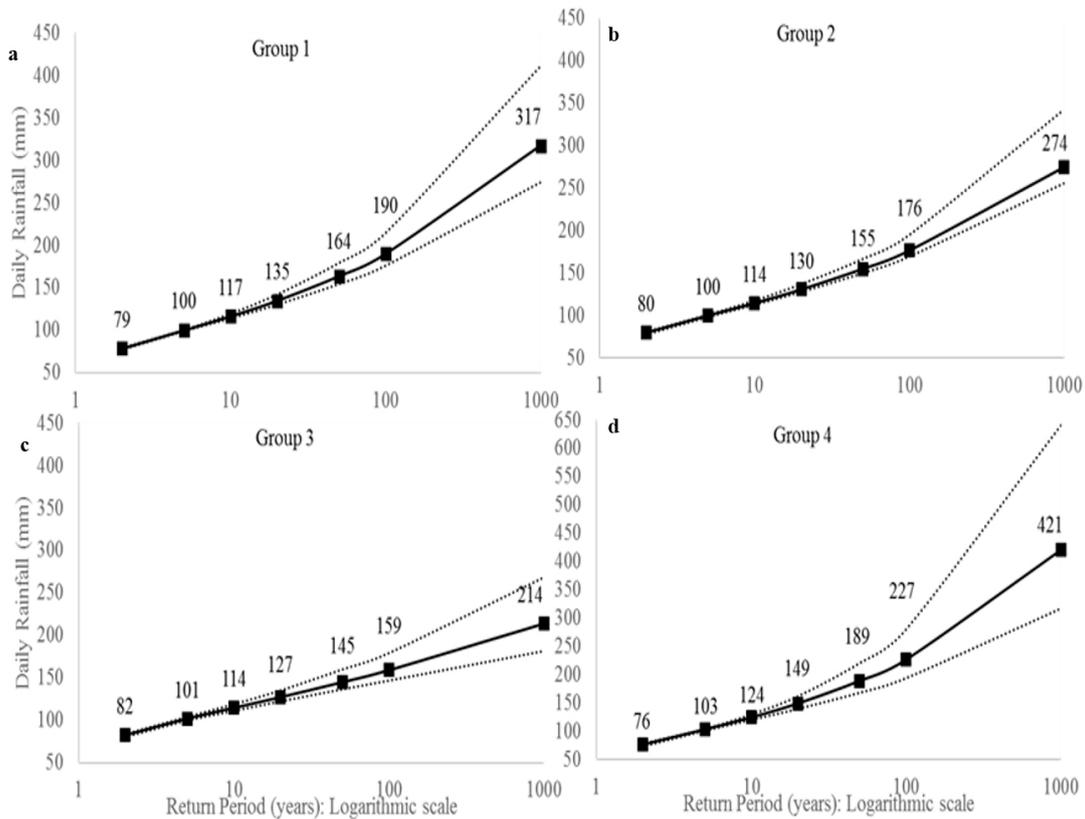
### Regional distribution: the null distribution of the Z measure

As previously described, once the iid assumption of the series forming a homogeneous group is met, the Z measure is expected to follow the standard normal distribution under  $H_0$  (Sung et al. 2018). Consequently, the largely used  $|Z| \leq 1.64$  critical limit is expected to represent the acceptance of the candidate distributions at the 10% significance level (Sung et al. 2018). According to this original limit the GEV is the only distribution that can be used to assess the probability of the rainfall series in the four groups (Table I). However, the GLO presented the lowest Z values for three groups (1, 2 and 4). In addition, these two distributions (GEV and GLO) presented similar performances in respect to their pooled uncertainties measures in all groups (Figure 4a-d). Moreover, the GLO consistently presented the lowest PU values in all groups (including group 3 in which such distribution has been rejected by the original limit  $|Z| \leq 1.64$ ). Naturally, this strongly suggests that the assumption that the Z measure always follows the standard normal distribution may not hold true. In other words, this feature suggests that the critical value ( $Z=1.64$ ) may not be a suitable limit for evaluating the fit of the GLO distribution within the framework of the RFA. On the other hand, the PU values depicted in Figure 4a-d also suggests that such a critical limit is suitable for the other three distributions (GEV, GNO and PE3). This inference is based on the fact that the highest PU values is always observed for those distributions presenting  $|Z| > 1.64$ . Finally, as observed in studies such as Ghiaei et al. (2018) and Sung et al. (2018), the GPA distribution performed poorly for all groups and was no longer considered in this study.

Figure 5a depicts the null distribution of the Z measure for both GEV and GLO distributions. Although Figure 5a has been construct using

data from group A, its results are equivalent to those found for the other three groups. As can be noted for the GLO distribution, the  $|Z| \leq 1.64$  limit leads to a number of false rejections considerably high (~ 2000 trials or 20%) than what would be expected for an experiment carried out at the 10% significance level (1000 trials). On the other hand, by using the same limit  $|Z| \leq 1.64$ , the other three distributions (GEV, GLO and PE3) were rejected at rates consistent with such a significant level (~1000 trails or 10%; Figure 5a). These results can be further evaluated by a Q-Q plot constructed from  $|Z|$  values drawn from the GLO and GEV (Figure 5b, c). For the GLO distribution, the Q-Q plot clearly describes a remarkable departure from the quantiles of the standard normal distribution (Figure 5b, also in absolute values). Such a departure is particularly notable for  $|Z| \leq 0.80$ . However, when the  $|Z|$  values drawn from the GEV are used to construct the Q-Q plot (Figure 5c), this departure becomes notable only for  $|Z| \geq 2.50$ . This latter feature can also be observed for the other two distributions (GNO and PE3; not shown here for the sake of brevity) and it explains why the critical limit initially proposed by Hosking & Wallis (1993)  $|Z|=1.64$  led to rejection rates close to 90% (95%) only for the GEV, GNO and PE3 distributions. This latter feature strongly suggests that the type of the distribution (GLO or GEV, GLO or GNO and GLO or PE3) is the key-factor defining the null distribution of the Z values.

Finally, it becomes worth emphasizing that the null distributions depicted in Figure 5a - which presents a departure from the normally assumption - were constructed by means of the R-code described in this study. Therefore, such a code can be used to draw critical values for both H and Z measure replacing the original values suggested by Hosking & Wallis (1993). By using this code, the new critical limit for the GLO distribution -associated with a 10% probability



**Figure 6.** Group quantile function fitted to weather stations situated in the State of São Paulo, Brazil. The groups have been formed under the framework of the regional frequency analysis with critical values adapted to this subtropical region. The one-day rainfall values presented in each graph correspond to the following return period (in years): 2, 5, 10, 20, 50, 100 and 1000. The dashed lines represent the 95% confidence interval estimated by means of a parametric bootstrap procedure, based on the generalized logistic distribution.

of falsely rejecting a true  $H_0$  - is  $\sim 2.25$ . Therefore, we assume that both GEV and GLO can be used to assess the probability of the rainfall series in the four groups depicted in Figure 1b, with the GLO presenting the best performance. Naturally, the iid assumption has always to be meet.

The GEV-parameters (location, scale and shape) of each ‘regional distribution’ are, respectively: 0.85, 0.22 and -0.12 (group 1); 0.86, 0.21 and -0.06 (group 2); 0.88, 0.25 and 0.10 (group 3) and, 0.81, 0.26 and -0.15 (group 4). Apart from group 3, all groups presented negative shape parameters, which describe heavy-tailed distributions common to hydrological data (Gilroy & Mccuen 2012). The shape parameter assumed its lowest value within group 4 what

indicates that, among all groups, this latter group presents the largest quantiles as the cumulative probabilities approach 1 (Wilks 2011). As also expected for this pool of weather stations situated in the coastal area of the State, the scale parameter, which is analogue to the variance of a normal distribution (Delgado et al. 2010), assumed its highest value (0.26). Naturally, the parameters of the GLO distribution describe similar features for the four groups. For instance, the growth curve of group 4 (3) presents the highest (lowest) quantiles for each corresponding return-period. The growth curves (Figure 6) are also consistent with the L-CV and L-Skewness values presented in Figure 1b. For instance, the group presenting the highest

L-Skewness value (0.30; Group 4; Figure 6d) is that which presents, in general, the highest one-day rainfall totals for each return period considered in Figure 6. In addition, the range of the confidence intervals (CI) observed in each group do not vary monotonically among the different return periods ( $T_R$ ). For instance, considering groups 1 to 4, the range of the CI for  $T_R = 100$  years are: 39 mm [176: 215] (Figure 6a), 25 mm [169: 194] (Figure 6b), 33 mm [146: 179] (Figure 6c) and 87 mm [194: 281] (Figure 6d), respectively. For  $T_R = 1000$ , these ranges are 137 mm [274: 411] (Figure 6a), 87 mm [255: 342] (Figure 6b), 87 mm [181: 268] (Figure 6c) and 324 mm [317: 641] (Figure 6d). Future studies may evaluate the reasons leading for such behaviour of the CI values. Factors related to the different number of weather stations with different length of records (Table I), distinct inter-station distances and, naturally, the rainfall patterns observed within each group, may provide some explanation for such behaviour. The average (avg) and standard deviation (st) of the inter-stations distances observed within each group are: group 1 (avg=201 km, st=69 km), group 2 (avg=243 km, st=59 km), group 3 (avg=119 km, st=36 km), group 4 (avg=203 km, st=109 km).

As previously described, the growth curve for each weather station can be constructed by using the information depicted in Figure 6a-d [ $q(F)$ ] in equation 10 (the index flood are presented in Table 1; mean value of each at-site series). The GLO-parameters are: 0.93, 0.15, -0.25 (group 1); 0.95, 0.15, -0.21 (group 2), 0.97, 0.15, -0.10 (group 3) and 0.90, 0.19, -0.29 (group 4).

## CONCLUSION

The Regional Frequency Analysis can be used to assess the probability of daily-extremes of rainfall events in the State of São Paulo, Brazil.

By means of this technique, data belonging to one of the largest weather station networks of the State have been pooled together into four homogeneous groups. The Generalized Extreme Value and the Generalized Logistic distributions can be used to assess the probability of such extremes within the four groups.

Since there were a lack of studies addressing the use of the RFA under the climate conditions of the State of São Paulo, the suitability of the critical limits often used to declare a region/group as 'acceptable homogeneous' ( $H \leq 1.00$ ) or to select a candidate distribution ( $|Z| \leq 1.64$ ) has been evaluated. The limit  $H \leq 1.00$  is an appropriate (conservative) threshold for declaring a group of extreme-rainfall series as 'acceptably homogeneous' in the State of São Paulo. However, particularly for the Generalized Logistic distribution, the limit  $|Z| \leq 1.64$ , which is expected to correspond to the 10% significance level, led to rejection rates of a true null hypothesis considerable higher than 20%. A further investigation of this feature indicated that Z values drawn from this latter distribution present a remarkable departure from the quantiles of the standard normal distribution, which in turn has been used to drawn critical values for this heterogeneity measure. Therefore, a new critical limit ( $|Z| \leq 2.25$ ) – corresponding to a probability of 10% of falsely reject a true null hypothesis – has been drawn from null distributions constructed from sets of Monte Carlo simulations. These simulations were based on the climatic conditions of the State of São Paulo. For the other three distributions subjected to the above-mentioned investigation (Generalized Extreme Value, Generalized Normal and Pearson Type III), the limit  $|Z| \leq 1.64$  leads to rejection rates of a true null hypothesis close to 10%. The Z values drawn from each of these distributions approach the quantiles of the standard normal distribution for  $|Z| \leq 2.50$ . In

other to facilitate the use of the RFA in future studies, an R-code (R-software; [www.r-project.org](http://www.r-project.org)) capable of building null distributions for both H and Z measures has also been provided. Once the iid assumption is met, such a code may be used to draw critical values for both H and Z measures derived from three-parameter distributions such as Generalized Logistic, Generalized Extreme Value, Generalized Normal Pearson Type III and Generalized Pareto.

### Acknowledgments

To Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), for fellowship to the first author (process number 307616/2019-3).

### REFERENCES

- BASU B & SRINIVAS VV. 2013. Formulation of a mathematical approach to regional frequency analysis. *Water R Res* 49: 6810-6833.
- BLAIN GC, AVILA AMH & PEREIRA VR. 2017. Using the normality assumption to calculate probability-based standardized drought indices: selection criteria with emphases on typical events. *Int J Climatol* 38: e418-e436.
- BRADLEY AA. 1998. Regional frequency analysis methods for evaluating changes in hydrologic extremes. *Water R Res* 34: 741-750.
- BUISHAND TA. 1984. Bivariate extreme-value data and the station-year method. *J Hydrology* 69: 77-95.
- BURN DH. 1990. Evaluation of regional flood frequency analysis with a region of influence approach. *Water R Res* 26: 2257-2265.
- CHANDLER RE & SCOTT ME. 2011. *Statistical methods for trend detection and analysis in the environmental analysis*. 1<sup>st</sup> ed., Chichester: J Wiley & Sons, 388 p.
- COLES S. 2001. *An introduction to statistical modeling of extreme value*. 1<sup>st</sup> ed., London: Springer, 183 p.
- DALRYMPLE T. 1960. Flood frequency analysis. U.S. Geol Surv Water Supply Pap 1543-A: 11-51.
- DELGADO JM, APEL H & MERZ B. 2010. Flood trends and variability in the Mekong River. *Hydrol and Earth Syst Sci* 14: 407-418.
- FOWLER HJ & KILSBY CG. 2003. A regional frequency analysis of United Kingdom extreme rainfall from 1961 to 2000. *Int J Climatol* 23: 1313-1334.
- GENEST C & FAVRE A. 2007. Everything you always wanted to know about copula modeling but were afraid to Ask. *J Hydrol Eng* 12: 347-368.
- GHIAEI F, KANKAL M, ANILAN T & YUKSEK O. 2018. Regional intensity-duration-frequency analysis in the Eastern. *Theor Appl Climatol* 131: 245-257.
- GILROY KL & MCCUEN RH. 2012. A nonstationary flood frequency analysis method to adjust for future climate change and urbanization. *J Hydrol* 414: 40-48.
- GOUDENHOOFDT E, DELOBBE L & WILLEMS P. 2017. Regional frequency analysis of extreme rainfall in Belgium based on radar estimates. *Hydrol Earth Syst Sci* 21: 5385-5399.
- GUTTMAN NB. 1993. The Use of L-Moments in the Determination of Regional Precipitation Climates. *J Climate* 6: 2310-2325.
- HOSKING JRM & WALLIS JR. 1993. Some statistic useful in Regional frequency analysis. *Water R Res* 29: 271-281.
- HOSKING JRM & WALLIS JR. 1997. *Regional Frequency Analysis: An approach based on L-moments*. 1<sup>st</sup> ed., London: Cambridge University Press, 224 p.
- KHAN SA, HUSSAIN I, HUSSAIN T, FAISAL M, MUHAMMAD YS & SHOUKRY AM. 2017. Regional Frequency Analysis of Extremes Precipitation Using L-Moments and Partial L-Moments. *Advances in Meteorology* 2017: 1-20.
- ROBSON A & REED D. 1999. *Flood estimation handbook*. 1<sup>st</sup> ed., Wallingford: UK, Centre for Ecology and Hydrology, 584 p.
- SANTOS JF, PORTELA MM & PULIDO-CALVO I. 2011. Regional Frequency Analysis of Droughts in Portugal. *Water Resour Manage* 14: 3525-3537.
- SUNG JH, KIM YO & JEON JJ. 2018. Application of distribution-free nonstationary regional frequency analysis based on L-moments. *Theor Appl Climatol* 3(4): 1219-1219.
- SVENSSON C & JONES DA. 2010. Review of rainfall frequency estimation methods. *J Flood Risk Manag* 3: 296-313.
- VERA C ET AL. 2006. Toward a Unified View of the American Monsoon Systems. *J Clim* 19: 4977-5000.
- WILKS DS. 2011. *Statistical methods in the atmospheric sciences*. 3<sup>rd</sup> ed., San Diego: Academic Press, 702 p.

### SUPPLEMENTARY MATERIAL

#### Appendix.

**How to cite**

BLAIN GB, SOBIERAJSKI GR, XAVIER ACF & CARVALHO JP. 2021. Regional Frequency Analysis applied to extreme rainfall events: Evaluating its conceptual assumptions and constructing null distributions. *An Acad Bras Cienc* 93: e2019406. DOI 10.1590/0001-3765202120190406.

*Manuscript received on April 8, 2019;  
accepted for publication on February 11, 2020*

**GABRIEL C. BLAIN<sup>1</sup>**

<https://orcid.org/0000-0001-8832-7734>

**GRACIELA DA R. SOBIERAJSKI<sup>2</sup>**

<https://orcid.org/0000-0002-7211-9268>

**ANA CAROLINA F. XAVIER<sup>3</sup>**

<https://orcid.org/0000-0002-9732-8779>

**JOÃO PAULO DE CARVALHO<sup>4</sup>**

<https://orcid.org/0000-0001-5202-7457>

<sup>1</sup> Instituto Agronômico (IAC), Centro de Ecofisiologia e Biofísica, Avenida Barão de Itapura, 1481, Jardim Guanabara, 13012-970 Campinas, SP, Brazil

<sup>2</sup> Instituto Agronômico (IAC), Centro de Frutas, Avenida Luiz Pereira dos Santos, 1500, Jardim Currupira, 13214-820 Jundiaí, SP, Brazil

<sup>3</sup> Programa de Pós-Graduação em Agricultura Tropical e Subtropical, Instituto Agronômico (IAC), Avenida Barão de Itapura, 1481, Jardim Guanabara, 13012-970 Campinas, SP, Brazil

<sup>4</sup> Instituto Agronômico (IAC), Avenida Barão de Itapura, 1480, Jardim Guanabara, 13012-970 Campinas, SP, Brazil

Correspondence to: **Gabriel Constantino Blain**

*E-mail: gabriel@iac.sp.gov.br*

**Author contributions**

Gabriel C. Blain: conceptualization, methodology, R-code development, validation, formal analysis, investigation, data acquisition, original draft preparation, reviewing and editing, visualization and supervision. Graciela da R. Sobierajski: conceptualization, methodology, formal analysis, investigation, original draft preparation, reviewing and editing. Ana Carolina F. Xavier: validation, investigation and original draft preparation. João Paulo de Carvalho: validation, data acquisition and original draft preparation. All authors have read and agreed to the published version of the manuscript.

