https://doi.org/10.1590/1678-4499.20220061

AGROMETEOROLOGY
Article

Is it possible to apply the regional frequency analysis to daily extreme air temperature data?

Letícia Lopes Martins[1], Julia Camila de Souza[1], Graciela da Rocha Sobierajski[2], Gabriel Constantino Blain[3,*]

1. Instituto Agronômico – Programa de Pós-Graduação em Gestão de Recursos Ambientais – Campinas (SP), Brazil.
2. Instituto Agronômico – Centro de Fruticultura – Jundiaí (SP), Brazil.
3. Instituto Agronômico – Centro de Biossistemas Agrícolas e Pós-Colheita – Campinas (SP), Brazil.

Received: Mar. 23, 2022 | **Accepted:** Jul. 22, 2022

Section Editor: Patrícia Cia

***Corresponding author:** gabriel.blain@sp.gov.br

How to cite: Martins, L. L., Souza, J. C., Sobierajski, G. R. and Blain, G. C. (2022). Is it possible to apply the regional frequency analysis to daily extreme air temperature data? Bragantia, 81, e4322. https://doi.org/10.1590/1678-4499.20220061

ABSTRACT: The improvement of probabilistic assessments of extreme air temperature events is a major goal for agrometeorological studies. The regional frequency analysis based on L-moments (RFA-Lmom) has been successfully used to improve the study of hydrometeorological variables such as extreme rainfall events. This study investigated the hypothesis that the RFA-Lmom can be applied to extreme maximum (Tmax) and minimum (Tmin) air temperature data. The RFA-Lmon was calculated considering its original algorithm (multiplicative approach) and a new procedure referred as to additive approach. The suitability of both approaches was evaluated through Monte Carlo experiments, which simulated homogeneous and heterogeneous groups of Tmin and Tmax series, and through a case study based on weather stations situated in the state of São Paulo, Brazil. The results found in this study indicated that the RFA-Lmom can be applied to Tmax and Tmin data in tropical/subtropical regions such as the state of São Paulo. In addition, the additive approach consistently outperformed the multiplicative approach. Both discordance and heterogeneous measures presented their best performances when calculated under this new approach. The original goodness-of-fit measure may also be replaced by its bivariate extension when the group is formed by more than 15 series.

Key words: probabilistic assessment, L-moments, heatwaves, frost events.

INTRODUCTION

The assessment of the probability of extreme air temperature events is a major goal for several agrometeorological studies because this element is one of the main weather variables that affects agriculture around the world (Alvarez et al. 2012). For instance, heatwaves (Vogel et al. 2019) and frost events (Blain 2011) may lead to harvest failures, threatening food security in countries worldwide. These events may not only affect the regions where they occurred, since other parts of the world may experience indirect effects such as reduced exports of agricultural products and higher food prices (GFSP 2015, Puma et al. 2015). This is the case of several Brazilian regions, which rank at the top-five world producer of agricultural commodities such as coffee, citrus, soybean, sugarcane, among many others.

Estimating the probability of extreme meteorological events is a difficult task because, by definition, they occur at long-return periods (> 100 years), which usually surpass the available length of meteorological records (Goudenhoofdt et al. 2017, Blain et al. 2021). Although this difficulty is a major challenge for any study addressing weather extremes, it is particularly relevant for tropical and subtropical regions where the availability and quality of long-running air temperature series are a matter of great concern. Therefore, the use of any statistical technique capable of improving the probabilistic assessment of extreme air temperature data may be regarded as a key-step for agrometeorological studies addressing the effect of this weather variable on agricultural production.

The statistical technique regional frequency analysis based on L-moments (RFA-Lmom; Hosking and Wallis 1993, 1997) was originally designed to improve the probabilistic assessment of hydrometeorological variables such as extreme rainfall data (Caporali et al. 2008). A key-step of the RFA is to verify if a pre-specified group of weather stations or meteorological series can be deemed as "acceptable homogeneous" (Hosking and Wallis 1993, 1997). The frequency distributions of the data series forming a homogeneous group are expected to share a unique shape, apart from a local/at-site scaling factor (index flood; μ), which and usually is taken to be the population mean at each site and estimated by the sample mean of each series. Therefore, data from all weather stations in this group may be used to fit a unique parametric function, which is taken as "the regional distribution" (Hosking and Wallis 1993, 1997, Fowler and Kilsby 2003, Viglione et al. 2007, Caporali et al. 2008, Svensson and Jones 2010, Goudenhoofdt et al. 2017, Blain et al. 2021). In simple terms, a homogeneous group formed by 10 meteorological series with a length of records equal to 20 years is considered. Instead of fitting a parametric distribution to each one of these 20-year series (at-site approach), what may result in high levels of uncertainties due to the low data availability (lower than a 30-year normal period), the RFA enables combining summary statistics of each series to build regional estimates, which are potentially subjected to a lower degree of uncertainty. In addition, this technique enables estimation at ungauged sites by transferring information coming from nearby gauged sites (Renard 2011). Along with these desirable features, the RFA also evaluates the quality of the meteorological series forming the group (Blain et al. 2021). The RFA performs this quality assessment on a regional basis by means of the discordancy measure (d) that assesses the presence of discordant sites within a pre-delimitated group.

The RFA-Lmom is also based on other two statistics: the heterogeneity measure (H), that evaluates whether a pre-delimitated group can be deemed as homogeneous; and the goodness-of-fit measure (Z), that evaluates whether a candidate distribution can be regarded as the regional distribution for a homogeneous group of weather stations/sites. The regional distribution is then used to calculate regional quantiles [q(F), for $0 < F < 1$], which in turn define a dimensionless curve (regional growth curve) common to all sites. The quantile estimates at a particular i site [$Q_i(F)$] are calculated by multiplying q(F) by its corresponding index flood (Eq. 1).

$$[Q_i(F) = \mu_i q(F)] \tag{1}$$

The above-mentioned qualities emphasize the importance of applying the RFA-Lmom in any region of the world, including those where there is a lack of long-running meteorological records (Blain et al. 2021). Nevertheless, to the authors' best knowledge there is no study that applied the RFA-Lmom to extreme air temperature data. The application of this technique to air temperature data needs to verify if the RFA-Lmom algorithm can be applied to parametric distributions that have been used to model this environmental variable. More important, it should also be verified if the above-mentioned assumption, which leads to Eq. 1, holds for air temperature data. This issue arises because the natural base for temperature measurement is the Kelvin scale, in which the absolute zero (0 k) corresponds to -273.15°C or -459.67°F. Since the observed values for this meteorological element tend to lie in a narrow range compared to their difference from the absolute zero, it may be supposed that their dispersions tend to be significantly different from those observed for extreme rainfall data and may be taken to be considerable independent of the mean of their distributions. In this case, the Eq. 1, referred to as the multiplicative approach, may be replaced by Eq. 2, referred to as the additive approach.

$$Q_i(F) = \mu_i + q(F) \tag{2}$$

In this context, the main goal of this study was to investigate the hypothesis that the RFA-Lmom can be applied to both extreme maximum (Tmax) and minimum (Tmin) air temperature data. Thus, the remainder of this paper is organized as follows: Section "The regional frequency analysis based on the L-moments" presents the main steps of the original RFA-Lmom algorithm (multiplicative approach), and the adjustments required by the additive approach. This section also presents recent improvements in the RFA statistics, such as those suggested by Neykov et al. (2007), which proposed a more robust version of the d measure, and Kjeldsen and Prosdocimi (2015), which developed a bivariate extension of the Z measure. Section "Monte Carlo experiments" describes three Monte Carlo experiments designed to compare the performances of the multiplicative and the additive approaches and to evaluate the performance of the RFA-Lmom's measures when applied to Tmin and Tmax series. These controlled simulation experiments used the generalized extreme value (a three-parameter function) distributions to generate synthetic Tmin and Tmax data. Distinct levels of spatial dependence were also considered. This study also presents a case study that evaluated the suitability the RFA-Lmom for real-world Tmax and Tmin data. The testing area was the Eastern region of the state of São Paulo, Brazil, which is crossed by the Tropic of Capricorn. The remaining sections are results, discussions, and conclusions.

THEORETICAL BACKGROUND AND DATA

The regional frequency analysis based on the L-moments

As already described, the original algorithm of the RFA was designed assuming that the frequency distributions of the series situated within a homogeneous region/group are identical, apart from a site-specific scaling factor (Hosking and Wallis 1993). This factor, known as index flood because of its early applications in hydrology (e.g., Dalrymple 1960), may be estimated by the sample mean of each series (μ). Under the multiplicative approach, the reginal frequency distribution is defined to have mean equal to 1 and it may be used to specify dimensionless regional quantiles of non-exceedance probabilities F [q(F)] (Eq. 3).

$$Q_i\,(F) = \mu_i\, q(F) \quad \mathrm{i} = 1, \ldots \mathrm{N} \tag{3}$$

in which: Q(F) = the quantile of a non-exceedance probability F of series/site i.

The analysis of Eq. 3 indicates that identifying a homogeneous group of series and selecting a suitable model for specifying q(F) are key stages in a RFA. In this context, Hosking and Wallis (1993, 1997) proposed three L-moment based statistics (d, H and Z) that provide objective judgments to support these stages. These measures are described as follows.

The d measure was developed assuming that three L-moments ratios (L-CV, L-skewness, and L-kurtosis) of a particular series is represented as a point in a three-dimensional space. Consequently, the L-moments ratios of a group of series yields a cloud of points. Provided that the centre of this cloud is the group average values for L-CV, L-skewness, and L-kurtosis, those points lying far from this centre are flagged as "discordant sites" (Hosking and Wallis 1997, Neykov et al. 2007, Blain et al. 2021). This measure assumes that the three-dimensional vector formed by the L-moments ratios came from a multivariate normal distribution and its critical limits can be found in Hosking and Wallis (1997).

The d measure can be used at two stages of the RFA. At the outset of the analysis, i.e., before the groups have been pre-delimitated. At this stage, the discordance measure is applied to all available series so that it might flag those with gross errors. After that (second application), this statistic may be used along with the H measure to discriminate those series preventing a pre-delimitated group to be deemed as acceptable homogeneous. Algebraically, the original d measure is based on the sample mean and the sample covariance matrix. These two statistics may be greatly affected by the presence of few outliers (Rousseeuw and Leroy 1987, Hubert et al. 2005, Neykov et al. 2007). To overcome this disadvantage, Neykov et al. (2007) proposed using the minimum covariance determinant estimator to improve the estimates of the sample mean and the sample covariance matrix of the group of series. Further information on this adapted/improved version of the discordance measure (referred in this study as to RD) can be found in Neykov et al. (2007). The RD is expected to outperform the original d measure (Neykov et al. 2007).

Within the framework of the RFA-Lmom, cluster analysis methods are often used to pre-delimitate a group of series (Wang et al. 2017). The H measure is applied to this pre-delimitated group to verify the hypothesis that it can be deemed as acceptable homogeneous. As suggested by Hosking and Wallis (1997), this hypothesis can only be accepted when the H measure of a group of series is equal to or lower than 1. Whenever a group presents H > 1, those series exhibiting the highest values for the discordance measure are removed from the group, and the H measure is applied to the remaining series. Within a pre-delimited group, the H measure compares the between-site variability of L-moments ratios with that expected from a truly homogeneous group. This truly homogeneous group is often simulated from multi-parameter distributions such as the four-parameter Kappa or the five-parameter Wakeby distributions. The between-site variation of the L-moments ratios of the observed series is calculated according to Eqs. 4 and 5.

$$t^R = \frac{\sum_{i=1}^{N} n_i t^{(i)}}{\sum_{i=1}^{N} n_i} \quad (4)$$

in which: $t^{(i)}$ = the sample L-CV; n_i = the length of record of each i site.

$$V = \frac{\sum_{i=1}^{N} n_i (t^{(i)} - t^R)^2}{\sum_{i=1}^{N} n_i} \quad (5)$$

The between-site variation of the L-moments ratios of the series forming a truly homogeneous group may be calculated according to the following computational algorithm (developed in the R-software environment), which is based on the studies of Hosking and Wallis (1988, 1997), Collier (2011)[1], Blain et al. (2021):

- Step 1: Fit a highly flexible distribution to the regional average L-moments ratios. "Highly flexible" Hosking and Wallis (1997) means a function capable of mimicking a range of different distributions, which are usually used to assess the probability of occurrence of the variable under analysis. As previously described, the four-parameter Kappa distributions is frequently used for such a purpose (Hosking and Wallis 1997, Bradley 1998, Santos et al. 2011, Basu and Srinivas 2013, Sung et al. 2018);
- Step 2: Transform the data used in step 1 into normally distributed data;
- Step 3: Calculate the cross-correlation of the normalized data, which will yield a correlation matrix;
- Step 4: Simulate cross-correlated series from a multivariate normal distribution. The number of simulated series must be the same as those of the sites/weather stations forming the original group (step 1);
- Step 5: Use the highly flexible distribution fitted in step 1 to perform the inverse transformation of the data generated in step 4. The length of these latter series must be the same as those of the series forming the original group. Provided that all synthetic series were generated from the same distribution, they form, by definition, a homogeneous group;
- Step 6: Apply Eqs. 4 and 5 to these synthetic series (V.synthetic);
- Step 7: repeat steps 2 to 6 a large number of times (e.g., 10,000 times);
- Step 8: Calculate the mean ($\mu_{synthetic}$) and the standard deviation ($\sigma_{synthetic}$) of all V.synthetic values. The H measure is then calculated by Eq. 6.

$$H = \frac{V - \mu_{synthetic}}{\sigma_{\mathrm{synthetic}}} \quad (6)$$

As recommended by Hosking and Wallis (1997), a group of series may be deemed as acceptably homogeneous if H ≤ 1, possibly heterogeneous if 1 < H < 2, and definitely heterogeneous if H ≥ 2 (Hosking and Wallis 1997, Viglione et al. 2007, Sung et al. 2018, Blain et al. 2021). As described in Castellarin et al. (2008), the H ≤ 1 may be regarded as a strict limit, since the value of this statistic corresponding to the 10% significance level is H = 1.28 (Hosking and Wallis 1997). These two limits (1 and 1.28) were evaluated in this study.

The Z measure was originally developed to evaluate if three-, four- or five-parameter distributions approach the true underlying frequency distribution of the series forming a homogeneous group. The Z measure calculation algorithm is also based on the above-described algorithm (steps 4 to 7) and is calculated as follows (Eqs. 7, 8 and 9).

[1] Collier, A. J. (2011). Extreme value analysis of non-stationary processes: a study of extreme rainfall under changing climate. Newcastle University (Thesis).

$$Z = \frac{\tau_4 - t_4^R + B_4}{\sigma_4} \tag{7}$$

$$B_4 = N_{sim}^{-1} \sum_{m=1}^{N_{sim}} \left(t_4^{[m]} - t_4^R \right) \tag{8}$$

$$\sigma_4 = \left[(N_{sim} - 1)^{-1} \left\{ \sum_{m=1}^{N_{sim}} \left(t_4^{[m]} - t_4^R \right)^2 - N_{sim} B_4^2 \right\} \right]^{1/2} \tag{9}$$

in which: t_4^R = the group average L-Kurtosis; τ_4 = the L-Kurtosis of the candidate distribution; t_4^m = the group average L-Kurtosis of them simulated group; N_{sim} = the number of simulations (step 7).

The null hypothesis of the Z measure assumes that a particular candidate distribution may be used as the true regional distribution. Such hypothesis is often accepted when |Z| ≤ 1.64. Although Hosking and Wallis (1997) suggested that the RFA statistic should not be used as formal significance tests in which the probabilities of type I errors are pre-specified, this critical limit has often taken to correspond to the 10% significance level (Sung et al. 2018). When distinct probability functions presents |Z| ≤ 1.64, the distribution presenting the smallest absolute Z value is often regarded as the best fitting distribution among the candidates (e.g., Hosking and Wallis 1997, Blain et al. 2021).

As can be noted from Eq. 7, the Z measure considers only the difference between the group average L-Kurtosis and the L-Kurtosis of the candidate distribution. Aiming at improving the performance of the original Z measure, Kjeldsen and Prosdocimi (2015) proposed a bivariate extension of this goodness-of-fit measure (Z.bivar) that is based on the individual and the joint sampling variability of the regional L-Skewness and L-Kurtosis. When applied to three-parameter distributions, this bivariate extension consistently outperformed its original version (Kjeldsen and Prosdocimi 2015). This bivariate extension may use the same synthetic series used to calculate H and Z measures to estimate the individual and the joint sampling variability of the regional L-Skewness and L-Kurtosis.

As previously described, because of the dispersion of their distributions, Eq. 3 (multiplicative approach) may not be a suitable choice for pooling information of a homogeneous group of air temperature series. In this case, an additive approach (Eq. 10) that assumes that the frequency distributions of the series situated within a homogeneous region/group are identical apart from an additive site-specific scaling factor emerges as an interesting alternative.

$$Q_i\,(F) = \mu_i + q(F) \text{ for i=1, … N} \tag{10}$$

Under this additive approach, the discordance and heterogeneous measures should be calculated using the L-moment scale measure l2 rather than L-CV. The regional frequency distribution should also be defined to have its mean value equal to zero rather than 1. Both Z and Z.bivar measures – which are based on L-Skewness and L-Kurtosis ratios – are calculated in the same way as in the multiplicative approach.

Monte Carlo experiments

The performances of both multiplicative and additive approaches for pooling information from distinct Tmax or Tmim series belonging to a homogeneous group of air temperature series were evaluated through the first Monte Carlo experiment proposed in this study. Based on a generalized extreme value (GEV) distribution, this controlled experiment simulated 10,000 homogeneous groups of Tmax and Tmim series and used the multiplicative and the additive approach to calculate Q(F) at each site. The non-exceedance probabilities F were set to the empirical cumulative probabilities of each Tmax or Tmin simulated values of each series, so that the Q(F) values obtained from Eqs. 3 and 8 could be compared with its corresponding simulated Tmax or Tmin value. These comparisons were based on the average squared difference between each Q(F) value and its corresponding Tmax or Tmin simulated values (MSE). The MSE were expressed as their squared root; RMSE=√MSE (Eq. 11).

$$\sqrt{\left\{\frac{1}{n1} * \sum_{m=1}^{n1}\left[\frac{T - Q_i(F)}{T}\right]^2\right\}} \tag{11}$$

in which: T = the synthetic Tmax or Tmin values; Qi(F) = their corresponding quantile estimates according to the multiplicative and additive approaches.

The parameters (Φ) of the GEV distribution were set to values observed in distinct locations of the state of São Paulo, which is situated within the coordinates 19°S and 26°S latitude and 53°W and 44°W longitude (a tropical-subtropical region crossed by the Tropic of Capricorn). As can be observed in Table 1, these parameters represent distinct climate conditions ranging from frequency distributions with mean 2.8°C and coefficient of variation equal to 68%, where negative values are observed (Monte Alegre do Sul, one of the highest locations in the state), to frequency distributions with mean 36.9°C and coefficient of variation equal to 3.8% (Ubatuba, situated in the coastal region). The sample sizes (n) of all simulated series were set to 30 years (normal period), the number of sites delimitating a region ranged from 7 to 30, and the average cross-correlation was set to 0 (no spatial dependence), 0.4, and 0.8 (strong spatial dependence).

Table 1. Parameters of the generalized extreme value (GEV) used to simulate extreme maximum (Tmax) and minimum (Tmin) series in the state of São Paulo, Brazil.

Monte Alegre do Sul (-22.67°S, -46.67°W, 785 m)	
GEV (2.1, 2, 0.37)	GEV (34, 0.92, 0.05)
Ribeirão Preto (-21.21°S, -47.90°W, 547 m)	
GEV (3.3, 2.60, 0.34)	GEV (36.1, 1.30, 0.35)
Ubatuba (-23.50°S, -45.0°W, 3 m)	
GEV (6.3, 1.7, 0.25)	GEV (36.4, 1.4, 0.24)

In order to evaluate the performance of the RFA-Lmon statistics when applied to Tmax and Tmin series, the GEV distribution (Table 1) were used to generate the synthetic series. Considering the low density of long-running air temperature records in South America, the sample sizes (n) of all simulated series were set to 30 (normal period) and 20 years. Figure 1 depicts the major steps of this second Monte Carlo experiment. As can be noted (Fig. 1), the performance of Z and Z.bivar measures were evaluated regarding their ability to correctly accept the GEV as a parent distribution and to discriminate it, among the other candidates, as the best fitting function. The other candidate distributions are presented in the next section.

The third Monte Carlo experiment performed in this study was designed to evaluate the performance of d, RD, and H measures when applied to heterogeneous groups of air temperature series. Similar to previous studies (e.g., Neykov et al. 2007, Viglione et al. 2007), the heterogeneous groups (Het.Group) comprised two homogeneous groups (Hom.Group1, and Hom.Group2) that were generated from distributions with different parameters (Het.Group = Hom.Group1 + Hom. Group2). Considering the parameters presented in Table 1 for the Tmin series, Hom.Group1 was generated from the GEV Φ values of the location of Monte Alegre do Sul (mean equal to 2.8, and coefficient of variation equal to 68%), while Hom. Group2 was generated from Φ values of the location of Ribeirão Preto (mean equal to 4, and coefficient of variation equal to 62%). The decision of simulating discordant series presenting probability density functions considerable close to the others forming the group allowed us to evaluate the ability of the discordance measures to flag discordant series which may not be easily detected by other forms of consistence analysis. The number of series of Hom.Group1 were always larger than those of Hom.Group2, so that percentages of discordant series in the Het.Groups varied from ~3 to ~30%. This controlled simulation experiment adopted the same n, N and spatial correlation levels considered in the first Monte Carlo experiment (step 1), and it also used the steps 2 to 5 of the first Monte Carlo experiment. The Z and Z.bivar measures were not calculated since the groups of series are heterogeneous, by definition.

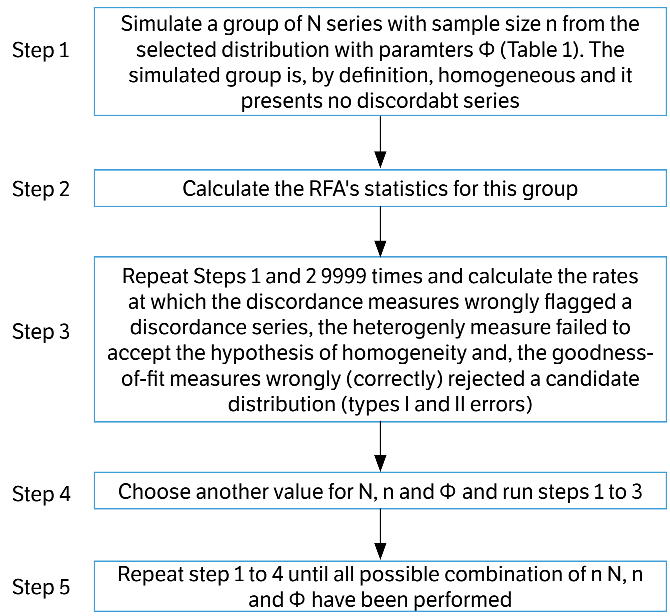
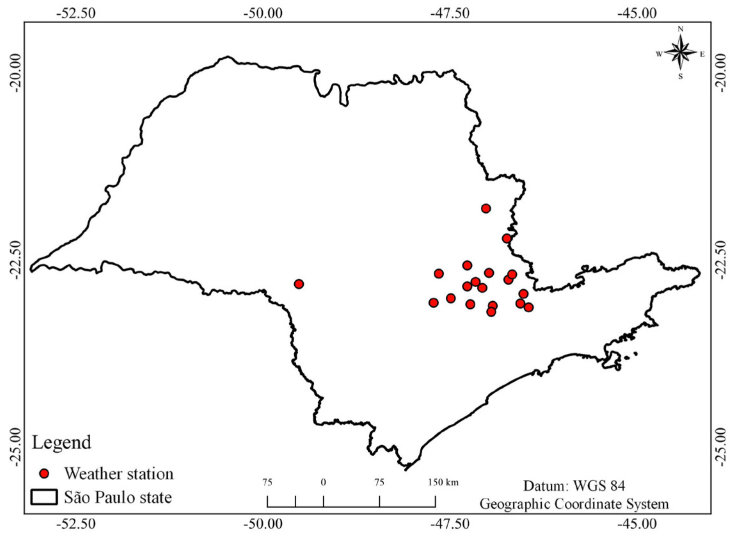


Figure 1. Major steps of a Monte Carlo simulation designed to evaluate the performance of the regional frequency analysis based on L-moments applied to homogeneous groups of extreme air temperature series.

Case study

The RFA-Lmom was applied to Tmax and Tmin data series from 19 weather stations belonging to the Agronomic Institute of Campinas (IAC/APTA/SAA, state of São Paulo). This state is situated at the boundaries of the tropical region (19°S and 26°S latitude and 53°W and 44°W longitude), where frost events leading to agricultural damages are, by definition, rare, but severe. Of these 19 weather stations, 18 are situated in the Eastern region of the state of São Paulo (Fig. 2). The other weather station is situated in the Midwest region of the state. The length of the selected air temperature series varied from 10 to 30 years (1991-2021). These series have been routinely used in both academic and operational modes (Blain et al. 2018, Blain et al. 2021). The percentage of missing records in each series is no greater than 5%. The Pearson's correlation coefficient was calculated for the Tmax and Tmin series observed in each location. Considering that we found no significant value for these coefficients, we assumed that minimum and maximum temperatures in the state of São Paulo can be modelled independently.

With regards to the definition of extreme values, the block maxima (and block minima) approach, as described in several studies (e.g., Coles 2001), was adopted in this study. Each block corresponded to one-year period. Thus, any year presenting a data gap were excluded from this approach. The use of this approach is based on the extremal types theorem, that postulates that the probability function towards which the sampling distribution of the X largest values of independent and identically distributed (iid) observations converges as X increases is the GEV distribution (e.g., Coles 2001, Wilks 2011). This theorem can also be applicable to distributions of extreme minima (Coles 2001, Wilks 2011, Fontolan et al. 2019), which are concerned with the X smallest values of iid observations. Despite this theoretical basis, in practical applications, the iid assumption is rarely met, and the GEV may not be the most suitable distribution to represent a set of extreme-value data (Wilks 2011). Therefore, instead of relying on this theoretical basis, the suitability of the GEV is often empirically evaluated along with other distributions (Wilks 2011). This study adopted this latter strategy and considered the following two- and three-parameter functions as candidates to represent the regional distribution: normal, Gumbel, GEV, generalized logistic (GLO), Pearson type III (PE3) and generalized Pareto (GP). These parametric functions are frequently used in extreme weather studies.



Local	Lat (S)	Long (W)	Local	Lat (S)	Long (W)
Amparo	-22.76	-46.72	Jundiaí-ETEC	-23.19	-46.95
Atibaia	-23.08	-46.56	Limeira	-22.57	-47.27
Bom Jesus dos Perdões	-23.13	-46.45	Monte Alegre	-22.69	-46.67
Bragança	-22.95	-46.52	Paulínia	-22.79	-47.16
Campinas	-22.87	-47.07	Piracicaba	-22.68	-47.65
Capivari	-23.01	-47.49	Santa Cruz do Rio Pardo	-22.82	-49.52
Espírito Santo do Pinhal	-22.21	-46.74	Sumaré	-22.85	-47.27
Indaiatuba	-23.09	-47.23	Tietê	-23.07	-47.72
Jaguariúna	-22.67	-46.98	Vargem	-21.81	-47.02
Jundiaí	-23.11	-46.93			

Figure 2. Air temperature series. State of São Paulo, Brazil.

The RFA analysis started by applying the discordance measures to all Tmax and Tmin series considered in this case study. The series that presented values for this measure larger than 3 were considered as possible discordant and were subjected to close examination. The H measure was applied to the remaining series to check the hypothesis of homogeneity, and the goodness-of-fit measures were applied to select the best fitting distribution among the candidate distributions (GEV, GLO, PE3, GP, normal and Gumbel). The selected probability function was regarded as the regional distribution for the group and it was used to construct the regional growth curve and the quantile function of each series forming the group. The additive approach was adopted in this case study because, as extensively already demonstrated, it consistently outperformed the multiplicative approach. The accuracy of the Q(F) estimates (Eq. 10) was evaluated using the approaches suggested by Hosking and Wallis (1997), which are based on forming a synthetic region/group with the same number of series, record length and cross-correlation as those of the homogeneous group of real-world series. The degree of heterogeneity that must be included in the synthetic group (Hosking and Wallis 1997) was estimated according to the procedure adopted in Hansen (2015). In this controlled simulation procedure, quantiles estimates were calculated for several F values, which varied from 0.10 to 0.95, so that the square root of the mean square errors (RMSE) were calculated according to Eq. 12.

$$\sqrt{\left\{\frac{1}{M} * \sum_{m=1}^{M} \left[\frac{Q_i^m(F) - Q_i(F)}{Q_i(F)}\right]^2\right\}} \qquad (12)$$

in which: M = the number of simulations; $Q_i^m(F)$ = the quantile estimated in the m-simulation; $Q_i(F)$ = obtained from Eq. 9 for a i-series.

RESULTS AND DISCUSSION

The results of the first Monte Carlo experiment indicated that Eq. 10 (additive approach) outperformed Eq. 3 (multiplicative approach) for pooling information from distinct Tmax or Tmim series. The RMSE values obtained from the additive approach were lower than or (at most) equal to those obtained from the multiplicative approach (Fig. 3). This statement holds particularly true for Tmin series.

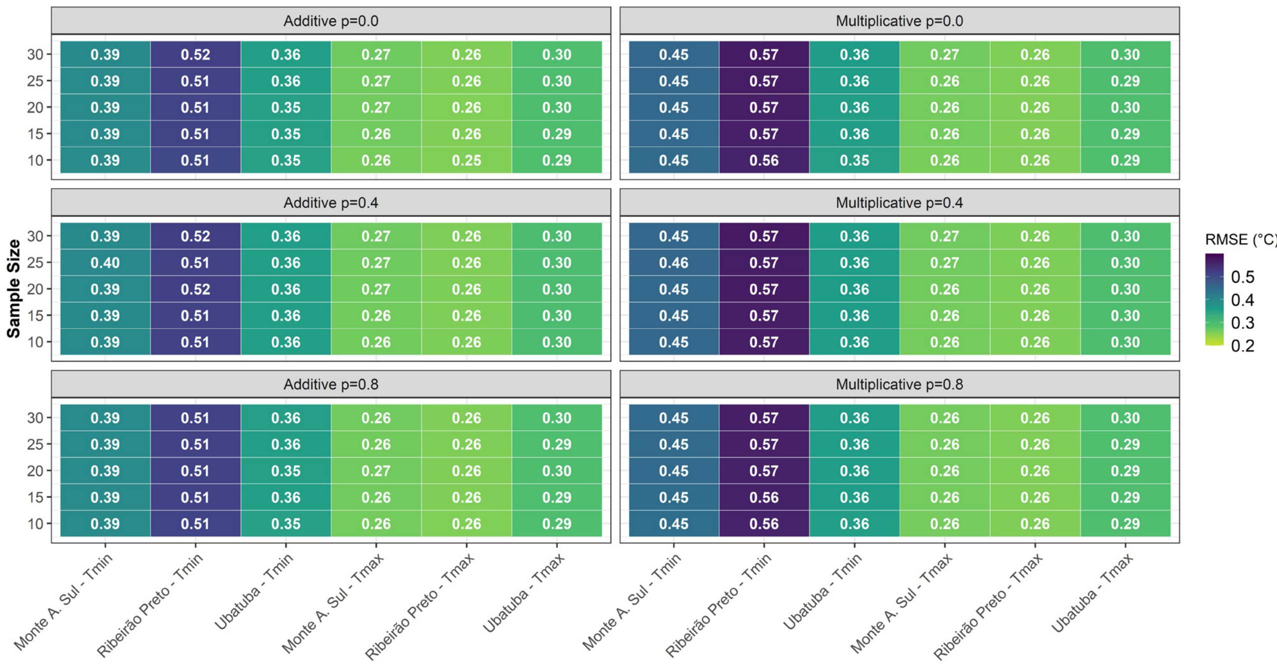


Figure 3. Root-mean squared errors (RMSE; °C) between Q(F) values (Eq. 3; right side and Eq. 10, left side) and their corresponding simulated air temperature values. The RMSE values presented for each generalized extreme value distributions, sample size and cross-correlation (ρ) is the average values of 10,000 Monte Carlo simulations.

Second Monte Carlo experiment: homogeneous groups

As previously described, Hosking and Wallis (1997) suggested the RFA's statistics should not be used as formal significance tests. This statement is based on the fact that accurate significance levels can only be obtained if the series that form the homogeneous group came from a population with the same frequency distribution used in step 1 of the algorithm already described. However, studies such as Masselot et al. (2017), Sung et al. (2018), Blain et al. (2021) used probability levels of falsely rejecting a true null hypothesis equal to 10 or 5% as baselines to evaluate the performance of the RFA's statistics. In this study, the performance of the RFA's measures were compared with those expected from a hypothesis test performed at the 10% significance level, and also with the results found in previous studies (e.g., Hosking and Wallis 1997, Viglione et al. 2007, Sung et al. 2018, Blain et al. 2021) that applied the RFA to variables that it was originally designed to (e.g., extreme rainfall values). This strategy is also in line with the goal of this study that investigates the hypothesis that the RFA-Lmom, which was originally designed to assess the probability of hydrometeorological variables such as extreme rainfall, can be applied to extreme Tmax and Tmin series.

With regards to the discordance measures (d and RD), the additive approach outperformed the multiplicative approach by showing rejection rates closer to the baselines of 10 or 5% (Fig. 4). Considering the multiplicative approach, the rejection rates presented by the original d measure were lower than 3% in all simulation rounds. This suggests that this statistic may present a behaviour similar to a conservative test that rejects a null hypothesis at rates considerably lower than the pre-specified significance level. Considering the conceptual trade-off between type I and II errors (Wilks 2011), this seemingly desirable feature also suggests that the original d measure may fail to correctly flag a discordant series at unacceptably high rates. On the other hand, the RD measure (multiplicative approach) could meet this pre-determined 10% chance of falsely flagging a series only for groups presenting N ≥ 20. For groups with N < 20, the RD measure presented a behaviour similar to a liberal test that falsely rejects a null hypothesis at high rates (~25% when N was set to 7). Under the additive approach, the performances of the d and RD measures were similar to each other, and they were only affected by the number of series forming the groups. The type I error rates presented by both measures decreased from values slightly higher than 10% to, approximately, 5% as N increased. This latter result is in line with those found by Santos Júnior et al. (2022) that applied the original RFA algorithm to monthly rainfall data. With regards to the sample size, the rejection rates observed when n was set to 20 were similar to those observed when n = 30 years (not shown here for the sake of brevity). This may be regarded as the first evidence suggesting that the RFA-Lmom may overcame limitations imposed by the low density of long-running maximum and minimum air temperature records in South America.

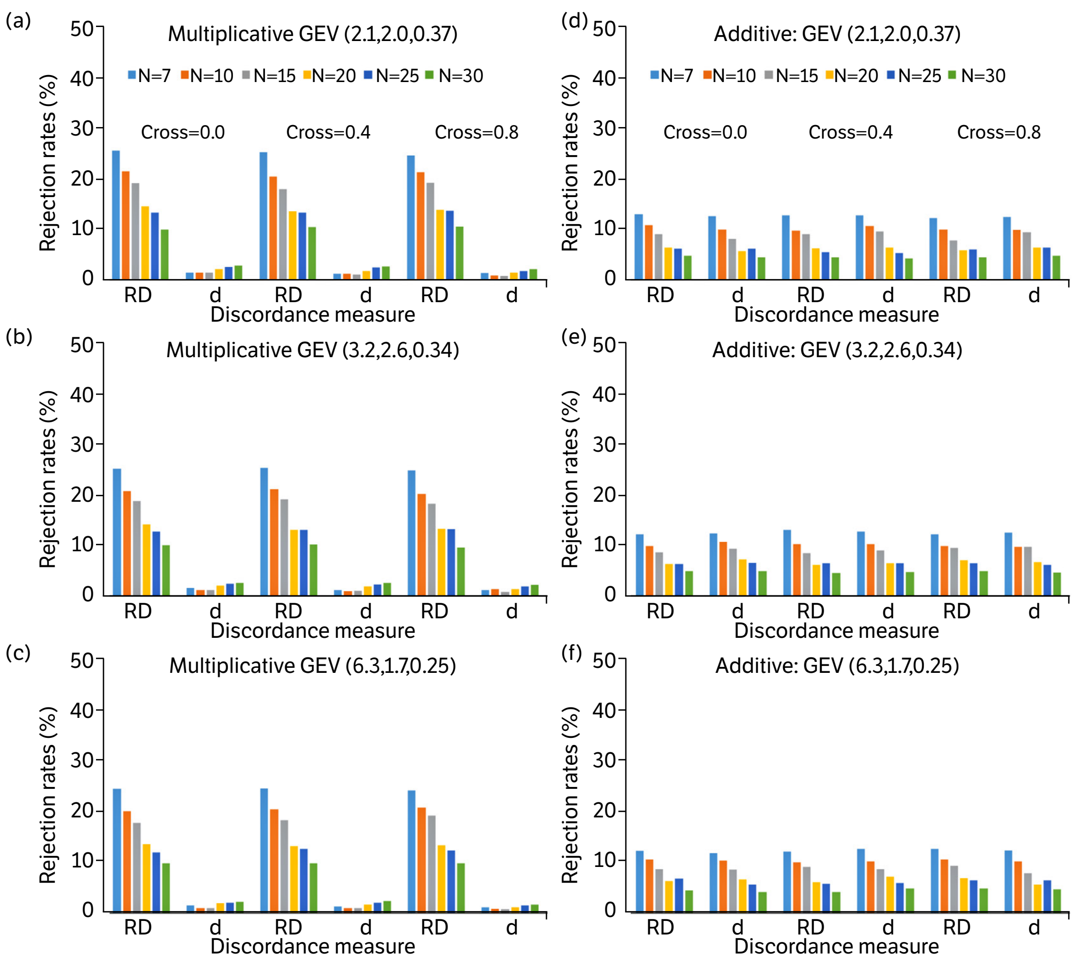


Figure 4. Rate at which both discordance (RD) and d measures, calculated under the multiplicative (a to c) and additive (d to f) approaches, falsely flagged synthetic air temperature series as discordant. The homogeneous groups were generated from the generalized extreme value (GEV). The cross-correlation varied from zero (cross = 0) to 0.8 (cross = 0.8). The length of record is 20 years.

The results depicted in Fig. 5 suggest that the H measure can be used to identify homogeneous groups of Tmax and Tmin series. Considering the critical limit H ≤ 1.0, the success rates shown by this measure remained above 80% for all factors considered in this first Monte Carlo experiment (n, N, spatial dependence levels, and distinct distributions). In other words, different from what was observed for the discordance measure, the results of Fig. 5 indicate that both additive and multiplicative approach are able to properly quantify the between-site variation of the L-moments ratios of the series that form a truly homogeneous group of air temperature series. As described in section "The regional frequency analysis based on the L-moments", the H ≤ 1.0 may be regarded as a strict limit (Castellarin et al. 2008), since the value of this statistic corresponding to the 10% significance level is H = 1.28 (Hosking and Wallis

1997). Naturally, the adoption of this latter value increased the success rates drawn from Monte Carlo experiment 2. However, as demonstrated in the next section, this larger critical limit (1.28 instead of 1) reduced the power of the H measure to correctly identify heterogeneous groups. As observed for the discordance measures, the success rates observed when n was set to the normal period (30 years) were similar to those observed when n = 20 years (not shown here for the sake of brevity). This may be regarded as another evidence suggesting that the RFA-Lmom may overcame limitations imposed by the low density of long-running air temperature records in South America.

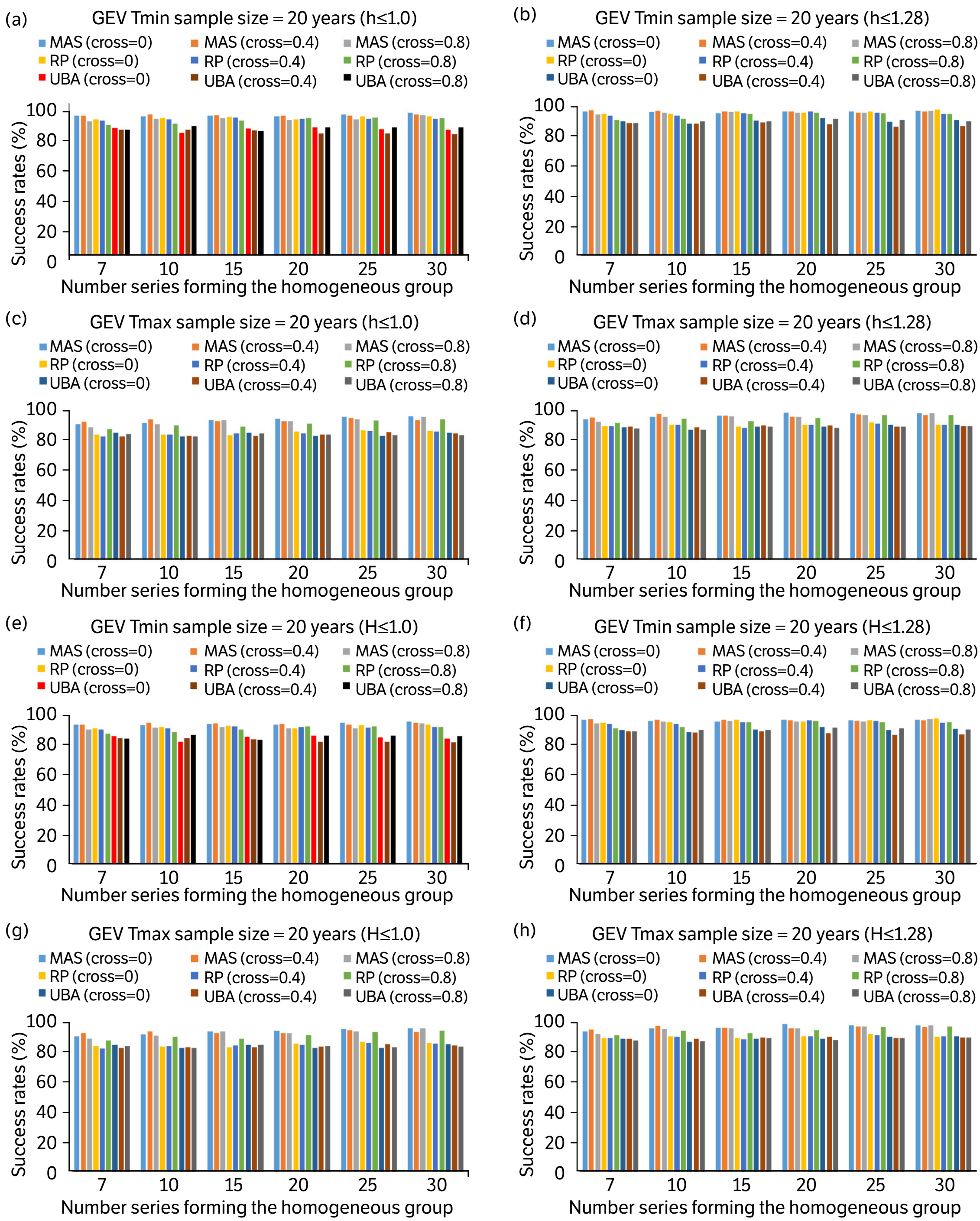


Figure 5. Rates at which the H measure, calculated under the multiplicative (a to d) and additive (e to h) approaches, correctly accepted the hypothesis of homogeneity. The homogeneous groups were generated from the generalized extreme value (GEV), considering the parameters shown in Table 1: Monte Alegre do Sul (MAS), Ribeirão Preto (RP) and Ubatuba (UBA). The cross-correlation varied from zero (cross = 0) to 0.8 (cross = 0.8). The length of record is 20 years.

The rates at which the original Z measure correctly accepted the probability function used in step 1 of Monte Carlo experiment 1 as the parent distribution of the series forming the groups remained close to 90% in all simulations rounds (Table 2). These success rates are comparable to those found in studies that applied the RFA to extreme rainfall values (Hosking and Wallis 1997, Kjeldsen and Prosdocimi 2015, Sung et al. 2018, Blain et al. 2021) and are also consistent with the rates at which type errors I are expected from a hypothesis test performed at the 10% significance level. The rates at which the Z.bivar measure correctly accepted the GEV distribution as the parent distribution remained at considerable low levels whenever N was set to its smallest values (N = 7 and 10; Table 2). This result limits the application of this latter statistic to groups formed by N < = 15 series. Considering the groups formed by 20 or more series, the success rates presented by this latter bivariate measure got closer to those presented by the original Z measure as N increased.

Table 2. Rates at which the original Z measure and its bivariate extension (Z.bivar) correctly accepted a probability function as the parent distribution of air temperature series forming homogeneous groups. The groups were generated from the generalized extreme value (GEV). The cross-correlation varied from zero (cross = 0) to 0.8 (cross = 0.8). The length of record is 20 years.

GEV (2.1, 2, 0.37)							GEV (34, 0.92, 0.05)						
N	Cross = 0		Cross = 0.4		Cross = 0.8		N	Cross = 0		Cross = 0.4		Cross = 0.8	
	Z	Zbivar	Z	Zbivar	Z	Zbivar		Z	Zbivar	Z	Zbivar	Z	Zbivar
7	86.8	65	85.6	62.2	89.6	67.4	7	90	60.4	90.4	61.6	89.2	67
10	87.8	77.4	87.4	75.2	88	73.6	10	89.2	64.4	90.8	64.2	90	72
15	87.2	86.8	85.2	86	88.4	86.6	15	89.4	77.4	91.6	73	90.8	75.6
20	87	92	87.8	92.6	87	92.6	20	89.4	86.8	90	84.2	91	86
25	88.8	95.4	87.4	95.8	87.4	95.2	25	90.4	92.2	91	88.8	92	90.4
30	88	97.4	88.6	97.2	88.4	98.2	30	90	95.2	89.2	91.6	91.4	94
GEV (3.2, 2.6, 0.34)							GEV (36.1, 1.30, 0.35)						
7	89	69.8	91	68	85.8	68	7	86.2	65	90	68.2	89.6	68.2
10	90.2	80.4	89.2	79.2	88	75	10	85.8	76.8	89.4	79	92.4	81.4
15	91.4	91.8	87.8	87.4	88	87.6	15	87.4	87.4	90.4	88.6	89.6	88
20	89.6	95.4	87.2	92.2	89.4	93.2	20	88.4	94	88.4	94	89.2	93.2
25	89.6	97	89.4	96.8	87.8	96.2	25	85	95.2	90	96	88.8	95.8
30	89.2	97.6	88	97.2	89.6	98	30	84.8	96.2	88.8	98.2	90	97.8
GEV (6.3, 1.7, 0.25)							GEV (36.4, 1.4, 0.24)						
7	89.2	55.6	90.8	54.4	91.2	64.8	7	88	70.2	85.6	65.6	90	70.4
10	86.2	63.8	91.6	62.4	91.2	71.2	10	88	79.8	86.4	74.4	88.2	79.4
15	88.6	77.8	90.2	78	92.8	78.2	15	87.6	87.8	87.6	87.6	89.6	88
20	90.2	86.2	89.2	86.4	92.2	87.2	20	88	93.6	85	92	87.6	93.4
25	88.4	91.4	87.2	90.4	92.2	92	25	87.6	96.8	86.2	94.6	86.8	96.6
30	89.2	93	87.4	93.8	92	94.8	30	87.6	97.2	86.6	96.6	86	97.6

The results presented in Table 3 indicate that the rates at which both original Z and Z.bivar measures chose the GEV model as the best fitting distribution ranged from 55 to 80%. These rates are comparable to those found in previous studies (e.g., Hosking and Wallis 1997, Kjeldsen and Prosdocimi 2015). The rates at which the Z.bivar measure correctly selected the best fitting distribution were comparable with those showed by the original Z measure. Therefore, the results presented in Tables 2 and 3 suggest that the original Z measure may be replaced by its bivariate extension only when a homogenous group is formed by N ≥ 15 Tmax or Tmin series. The results presented in Tables 2 and 3 also indicate that both Z and Z.bivar did not lose power when the spatial correlation was set to its highest level (0.8). Similar results were observed by Blain et al. (2021), and it indicates that the algorithm may be explained by the fact that the presence of spatial dependence among the series forming a group increases the variance of the regional L-moment ratios (Hosking and Wallis 1988, Castellarin et al. 2008, Kjeldsen and Jones 2006). As for the RD and H measures, the success rates observed when n was set to the normal period (30 years) were similar to those observed when n = 20 years (not shown because of the sake of brevity).

Table 3. Rates at which the original Z measure and its bivariation extension (Z.bivar) correctly select a probability function as the best fitting distribution of air temperature series forming homogeneous groups. The groups were generated from the generalized extreme value (GEV). The cross-correlation varied from zero (cross = 0) to 0.8 (cross = 0.8). The length of record is 20 years.

	GEV (2.1, 2, 0.37)							GEV (34, 0.92, 0.05)					
N	Cross = 0		Cross = 0.4		Cross = 0.8		N	Cross = 0		Cross = 0.4		Cross = 0.8	
	Z	Zbivar	Z	Zbivar	Z	Zbivar		Z	Zbivar	Z	Zbivar	Z	Zbivar
7	57.4	58.8	55.6	58.2	48.8	51.8	7	78.4	75	80.4	78.2	79.8	84.2
10	61.4	62.8	61.2	63.4	54.6	56.6	10	77.2	68	81.2	74	81.4	79.4
15	69.2	70	66.8	68.8	60.2	59.8	15	74.2	59.8	78.2	62.2	80.6	73.2
20	71	71	68.6	68	65.2	64.6	20	76.4	52.6	78.2	58.4	79.8	68.4
25	69.2	69.8	72.2	72	62.4	62.6	25	71	48.2	74.2	49.8	78.2	66
30	72.6	72.2	73.4	73.4	66.2	65.8	30	70.2	44.6	72.8	47.4	77.2	62.2
	GEV (3.2, 2.6, 0.34)							GEV (36.1, 1.30, 0.35)					
7	57.6	59.8	60.2	61	49.4	52.2	7	40.6	39.2	41	39.2	49.8	46
10	60.4	61.8	62	63.2	54	54.6	10	41.8	39	39	37.2	44.6	42.6
15	66.8	66.2	67.6	67.4	59.6	61.8	15	34.4	35.6	34.6	34	37	35.8
20	66.6	67	69.6	69.2	65	63.4	20	34	34.6	32.6	30.4	36.2	34.8
25	67.8	67.2	70.8	70.4	68	68.8	25	32.8	32.6	30	28.4	33.8	33.8
30	66.4	66.4	71.2	71.2	67.4	67.2	30	28.2	28.4	26.8	26.8	32	32
	GEV (6.3, 1.7, 0.25)							GEV (36.4, 1.4, 0.24)					
7	53	57.4	55.8	60.6	51.8	57.2	7	48.2	46.4	49.8	47.2	58.6	55.8
10	57.6	64.4	57	62.6	54.2	61.2	10	41	40.8	46.2	44.4	54.4	51
15	60.2	65	62	66.8	58.8	61.4	15	38.4	38.2	40.6	40.6	46.6	43.2
20	61.8	64.2	64.8	65.8	60.4	60	20	36	36.2	36.2	37.6	43.4	42.4
25	63.6	67	67.8	69.6	62	61.8	25	32.2	32	33.4	34.4	41.6	41.2
30	65.8	67	68.2	70.2	64.6	65	30	30.4	29.8	35	35	41.4	40.8

Third Monte Carlo experiment: heterogeneous groups

Similar to what was observed for the homogeneous groups, the success rates obtained from the multiplicative approach (Fig. 6) were consistently lower than those obtained from the additive approach (Fig. 7). Under this latter approach (Fig. 7), the rates at which both d and RD measures correctly flagged the discordant series were consistently higher than those obtained under the multiplicative approach (Fig. 6). These latter results may be regarded as another evidence favouring the use of the additive approach.

Considering only the additive approach, the d and RD measures showed similar performances. The rates at which both measures correctly flagged the discordant series (success rates) decreased as the percentage of discordant series in the groups increased. This behaviour is in line with results found by Neykov et al. (2007) and Santos Júnior et al. (2022) and is a consequence of the concept adopted by these measures to flag discordant series. They consider the L-moments or L-moments ratios of each series forming a group as a point in a three-dimensional space and flags those points lying far from the center of the cloud of points, which is the group average values for the L-moments ratios (Hosking and Wallis 1997, Neykov et al. 2007, Blain et al. 2021). As described in section "Monte Carlo experiments", the discordant series are those coming from Hom.Group2 (recall that Het.Group = Hom.Group1 + Hom.Group2). As the number of discordant series within this latter group increases, their effect (weight) on the average value for the entire group (Het.Group) also increases. Consequently, the centre of the cloud moves towards Hom.Group2 limiting the ability of the discordance measures to flag discordant series in Het.Group.

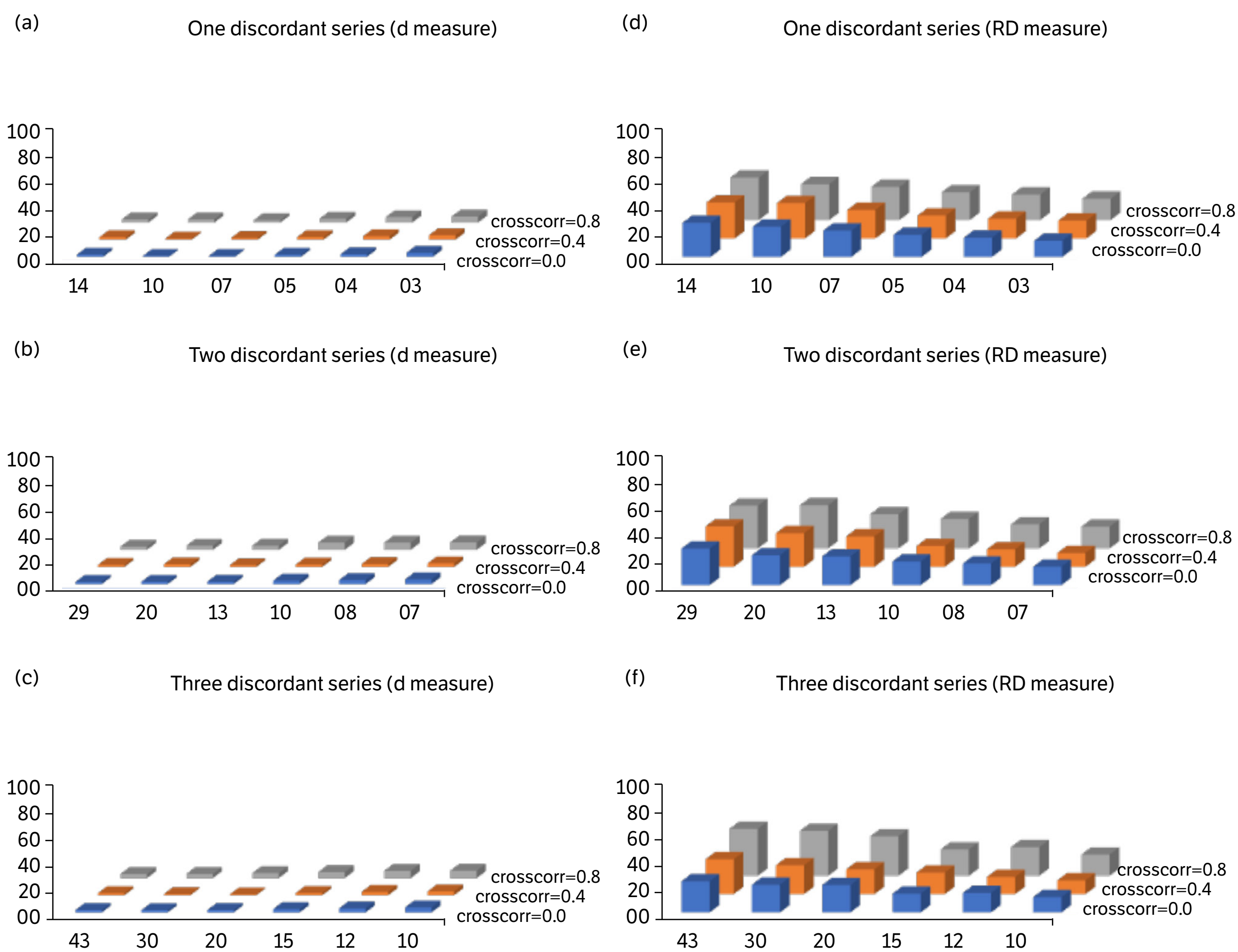


Figure 6. Rate at which both d (a to c) and discordance (RD) (d to f) measures correctly flagged synthetic air temperature series as discordant (multiplicative approach). The heterogeneous groups were generated from the generalized extreme value distribution – and comprised two homogeneous groups (Hom.Group1 and Hom.Group2). Hom.Group1 was generated from the parameters of the location of Monte Alegre do Sul (Table 1; minimum air temperature), while Hom.Group2 was generated from the parameters of the location of Ribeirão Preto (Table 1; minimum air temperature). The cross-correlation varied from zero (cross = 0) to 0.8 (cross = 0.8). The length of record is 20 years.

As observed for the discordance measure, the rejection rates showed by the H measure calculated under the additive approach were consistently higher than those observed under the multiplicative approach (Fig. 8). In other words, the results of this Monte Carlo experiment are in line with those of the first and second experiments, also favouring the use of the additive approach. Considering only this latter approach, the performance of the H measure was only affected by the number of discordant series within the group. For instance, consider those groups with 10% of discordant series. The success rates increased from ~ 75% (one discordant series; Fig. 8d) to ~ 95% (three discordant series; Fig. 8f). A similar relationship between the number of discordant series and the number of series forming the group can also be observed in previous studies (e.g., Hosking and Wallis 1997, Neykov et al. 2007, Requena et al. 2016, Blain et al. 2021). In addition, the opposite behaviour between the discordance and heterogeneity measures is consistent with the strategy suggested by Hosking and Wallis (1997) and adopted by Blain et al. (2021) of using the discordance and heterogeneity measure in an integrated manner, that is: once a group of series shows H > 1, the series with the largest RD value is removed from the group and the H measure is again estimated. This procedure is repeated until H ≤ 1.

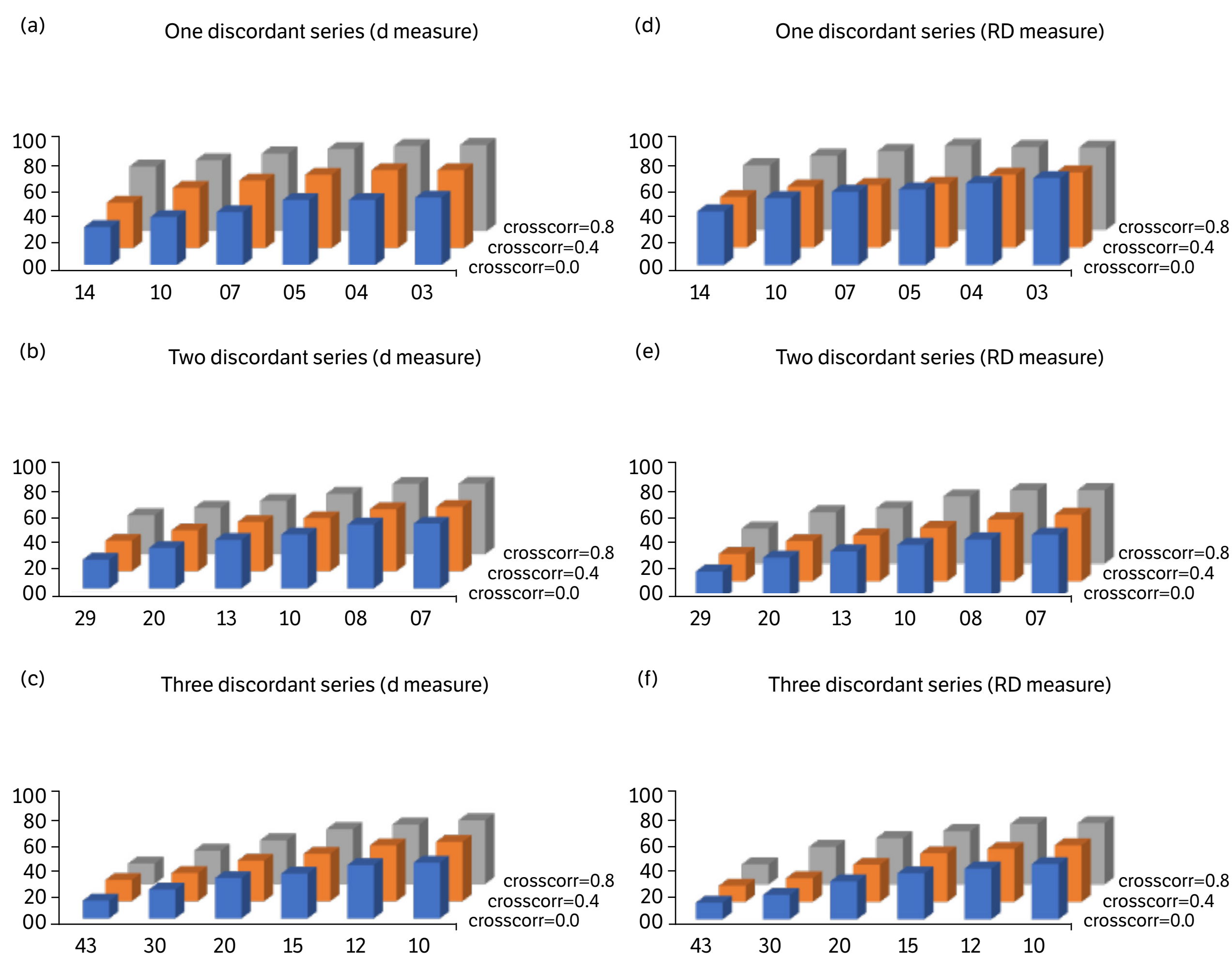


Figure 7. Rate at which both d (a to c) and discordance (RD) (d to f) measures correctly flagged synthetic air temperature series as discordant (additive approach). The heterogeneous groups were generated from the generalized extreme value distribution (and comprised two homogeneous groups – Hom.Group1 and Hom.Group2). Hom.Group1 was generated from the parameters of the location of Monte Alegre do Sul (Table 1; minimum air temperature), while Hom.Group2 was generated from the parameters of the location of Ribeirão Preto (Table 1; minimum air temperature). The cross-correlation varied from zero (cross = 0) to 0.8 (cross = 0.8). The length of record is 20 years.

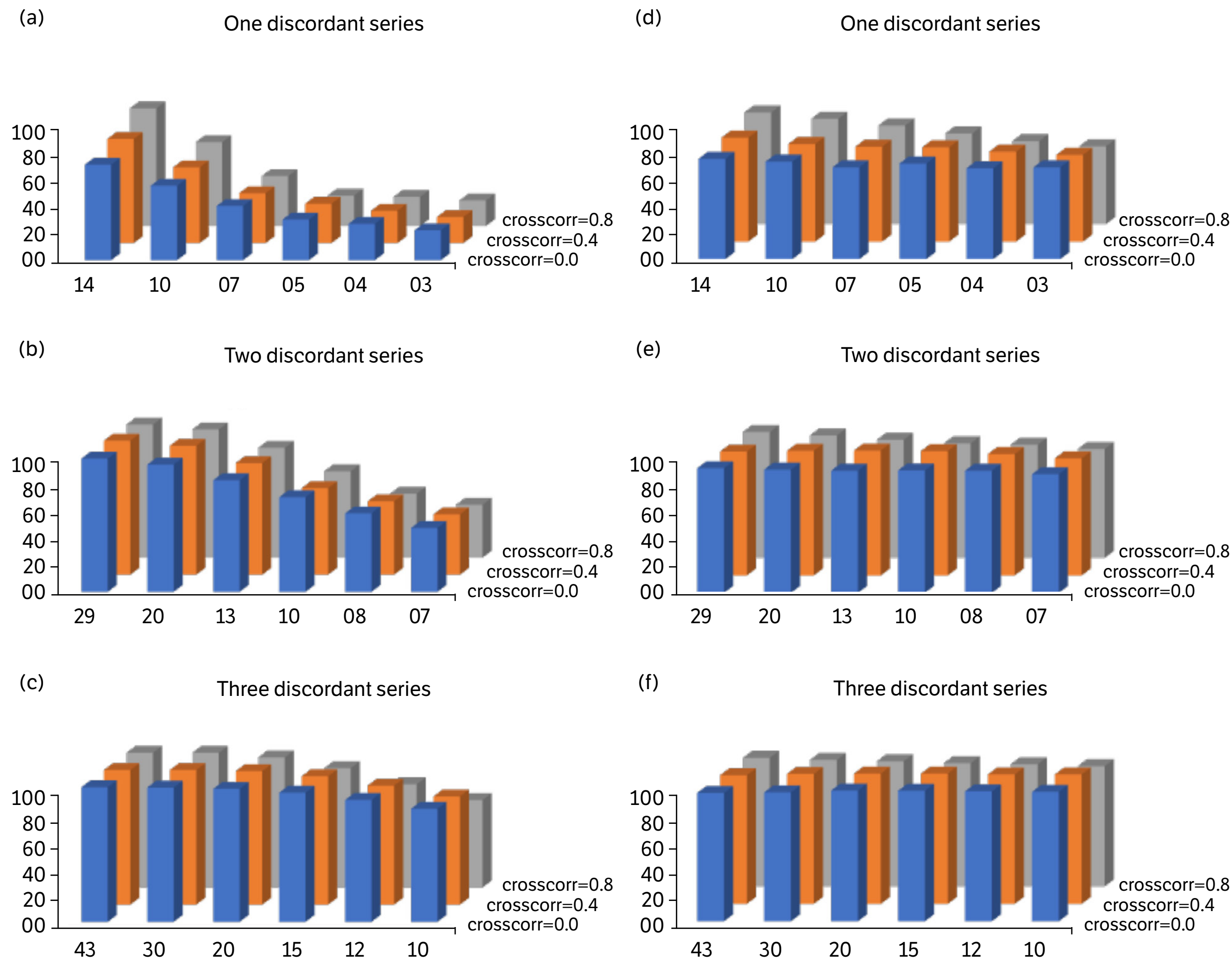


Figure 8. Rates at which the H measure, calculated under the multiplicative (a to c) and additive (d to f) approaches, correctly rejected the hypothesis of homogeneity. The heterogeneous groups were generated from the generalized extreme value and comprised two homogeneous groups (Hom.Group1 and Hom.Group2). Hom.Group1 was generated from the parameters of the location of Monte Alegre do Sul (Table 1; minimum air temperature), while Hom.Group2 was generated from the parameters of the location of Ribeirão Preto (Table 1; minimum air temperature). The cross-correlation varied from zero (cross = 0) to 0.8 (cross = 0.8). The length of record is 20 years.

Case study

The results drawn from the three Monte Carlo experiments support the hypothesis that the RFA-Lmom – calculated under the additive approach and considering the candidate distribution evaluated in this study – can be applied to both Tmax and Tmin series in the state of São Paulo. As observed in the previous sections the performances of the RFA statistics were comparable to those of previous studies that applied the original version of this regionalization technique (multiplicative approach) to hydrometeorological variables such as extreme rainfall data (e.g., Hosking and Wallis 1997, Neykov et al. 2007, Requena et al. 2016, Blain et al. 2021).

With regards to the Tmin series, only the location of Santa Cruz do Rio Pardo presented d and RD values close to the critical limit for this discordance measure (i.e., RDSantaCruzRioPardo = 2.9). However, it was not removed from the group because, according to the H measure (H = 0.10), this group of Tmin series can be deemed as acceptable homogeneous. With regards to the statistical properties of the series forming this homogenous group, the L-scale values (l2; equivalent to dispersion measures such as variance and standard deviations) ranged from 0.80 to 1.41. Almost all series presented negative L-Skewness values, which indicate the data are skewed to left. The values for L-kurtosis (which is analogous to the traditional kurtosis) ranged from 0.02 to 0.32. The original Z measure indicated that the GLO (Z = 0.45) and the Gumbel (Z = -0.71) can be accepted as the underlying distributions of these series. The Z.bivar measure indicated only the GLO distribution as an acceptable candidate (Z.bivar = 0.22). All other probability functions were rejected by both Z and Z.bivar measures. This may be regarded as an unexpected result since this latter parametric function is rarely considered in studies addressing the probability of extreme air temperature events in the state of São Paulo (e.g., Sansigolo 2008; Blain and Lulu 2011). However, this result is in line with Nidhin and Chandran (2013), that proved the theoretical importance of this parametric function in extreme value modelling, including extreme minima.

The group/regional parameters of the GLO are 0.15 (location), 1.05 (scale) and 0.08 (shape), and the regional growth curve is depicted in Fig. 9. Considering the mean of each Tmin series as the index flood (Eq. 10), Fig. 10 exemplifies the use of Eq. 10 by showing the quantiles of non-exceedance probabilities [Q(F)] for four locations. Monte Alegre do Sul is the most susceptible location to air temperature values that may lead to agricultural losses. Considering that air temperature values equal to or lower than 2°C may lead to agricultural losses in the state of São Paulo (e.g., Sentelhas et al. 1995, Blain 2011), Tmin ≤ 2°C is experienced in Monte Alegre do Sul at a probability slightly higher to 30%. In Campinas, this probability decreases to ~5%. In the locations of Bragança and Jundiaí, Tmin ≤ 2°C are observed at a probability ~15%. Finally, the results of Eq. 12 supported the use of the RFA-Lmom to improve the probabilistic assessments of extreme minimum air temperature in the state of São Paulo. Considering the smallest non-exceedance probabilities (≤ 0.30), the RMSE values obtained using this regionalization technique (Fig. 11, blue bars) were consistently lower than those obtained from the at-site approach (Fig. 11, red bars).

With regards to the Tmax series, the locations of Santa Cruz do Rio Pardo and Amparo presented d and RD values higher than the critical limit (i.e., 4.6 and 5.6, respectively). However, these locations were not removed from the group because, according to the H measure, this group of Tmax series can be deemed as acceptable homogeneous (H = -0.80). With regards to the statistical properties of the series forming this homogenous group, the L-scale values ranged from 0.63 to 1, and, apart from Amparo, all series presented positive L-Skewness values, which indicates the data are skewed to right. The values for L-kurtosis ranged from -0.15 to 0.39. Both Z and Z.bivar measures indicated the GEV as the best fitting distribution for the series forming the group (Z = -0.07 and Z.bivar = 0.11). This result is in line with previous studies that evaluated the use of this latter distribution to assess the probability of extreme Tmax in the state of São Paulo. The Z measure also indicated the normal, gum and PE3 distributions as acceptable candidates. The Z.bivar measure indicated the gum and PE3 as acceptable candidates.

The group/regional parameters of the GEV distribution are -0.67 (location), 1.25 (scale) and 0.05 (shape). The regional growth curve for this variable is depicted in Fig. 9. Considering the mean of each Tmax series as the index flood, Fig. 10 also exemplifies the use of Eq. 10 for this variable. The results of Eq. 12 also supported the use of the RFA-Lmom to improve the probabilistic assessments of extreme maximum air temperature in the state of São Paulo. Considering the highest non-exceedance probabilities (≥ 0.65), the RMSE values obtained using this regionalization technique (Fig. 11, blue bars) were consistently lower than those obtained from the at-site approach (Fig. 11, red bars).

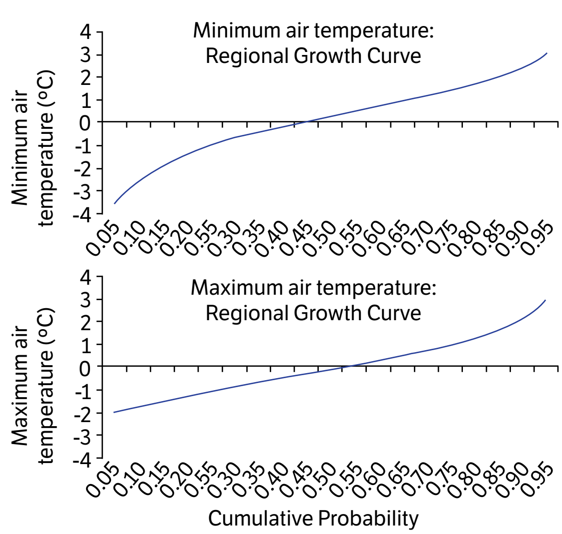


Figure 9. Regional growth curves for daily-extremes values of minimum and maximum air temperature. Both plots were constructed under the regional approach considered in the case study.

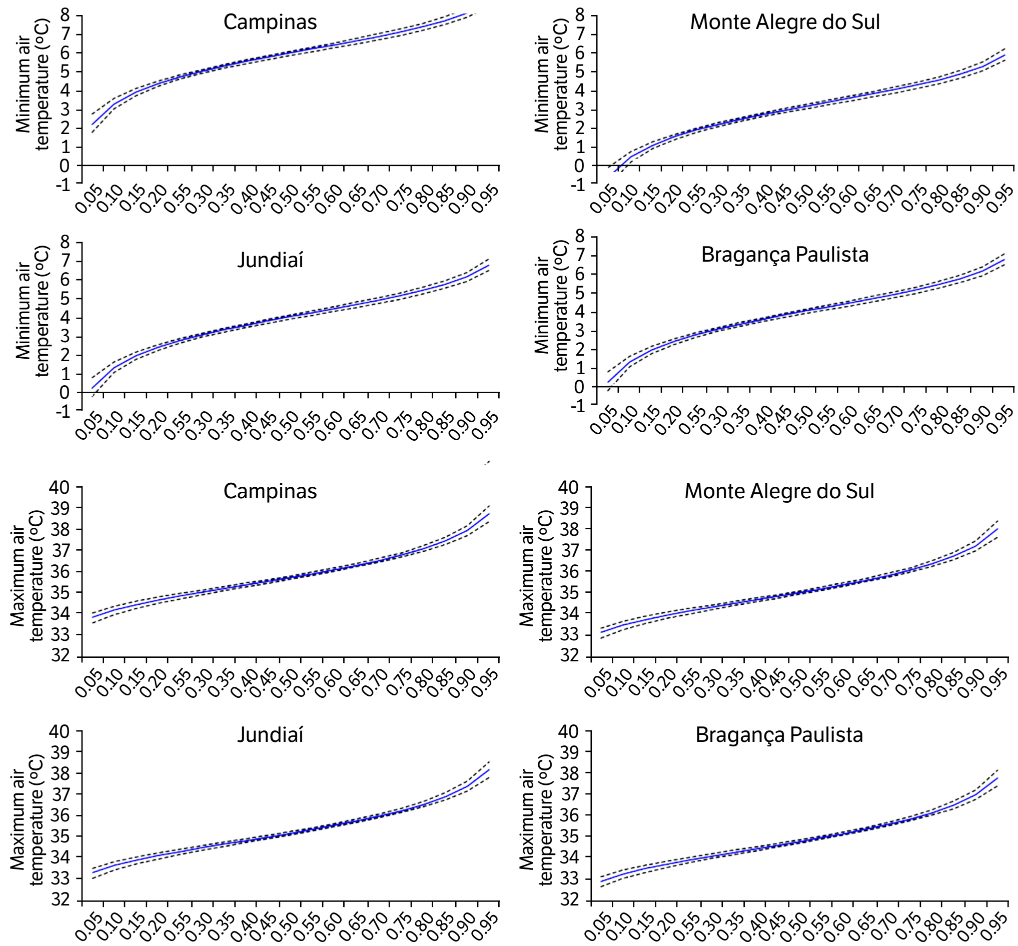


Figure 10. Quantiles of non-exceedance probabilities [Q(F)] for four locations of the state of São Paulo, Brazil. The [Q(F)] values (blue line) were calculated through the index flood approach applied to a generalized logistic distribution (GLO = 1.03, 023, 0.09; minimum air temperature) and to a generalized extreme value distribution (GEV = -0.66, 1.25, 0.05); maximum air temperature). The index flood assumed the sample mean value of each air temperature series and was calculated under the additive approach.

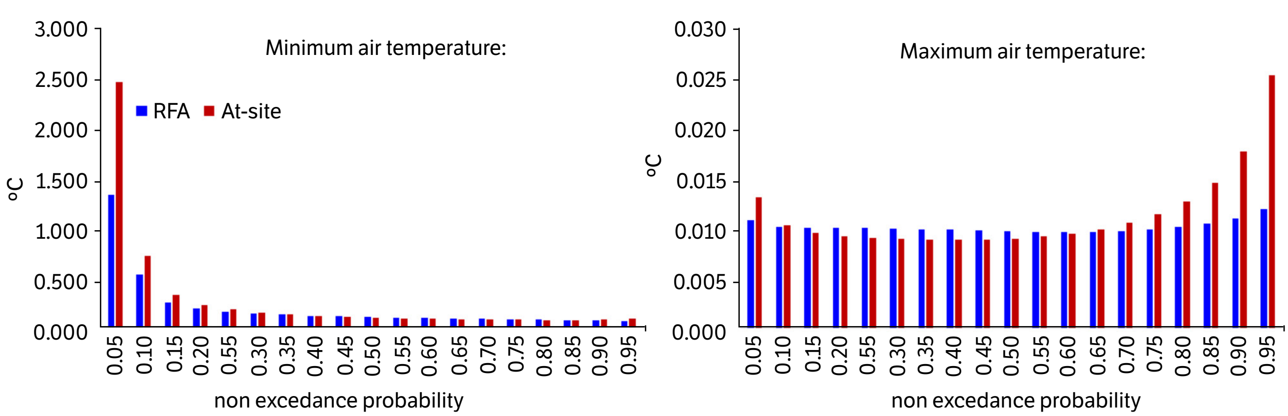


Figure 11. Root-mean squared errors (RMSE; °C) between Q(F) values (calculated from Eq. 10) and quantiles estimates calculated from the at-site approach. Case study, state of São Paulo, Brazil.

CONCLUSION

Considering that the improvement of the probabilistic assessment of extreme air temperature events is a major goal for several environmental studies, we investigated the hypothesis that the RFA-Lmom can be applied to both extreme maximum and minimum air temperature data. More specifically, this study investigated if the multiplicative approach – as described in the introduction of this study – holds for air temperature data. This issue arose because the observed values for air temperature data tend to lie in a narrow range compared to their difference from 0 Kelvin (the only natural basis for measuring this meteorological element). In this context, an additive approach was proposed as a strategy to adapt the RFA-Lmon to air temperature data. Results from three Monte Carlo experiments and a case study indicated that the additive approach consistently outperformed the multiplicative approach. Both discordance and heterogeneous measures presented their best performances when calculated under this new procedure. This statement holds for both maximum and minimum air temperature series. The results of the case study also emphasized the capability of the RFA-Lmom to improve the probabilistic assessments of extreme air temperature events in the state of São Paulo. This improvement may contribute to reduce harvest failures caused by heat waves or frost events in this tropical/subtropical region. The quantile estimates obtained from this regionalization technique were consistently more accurate than those obtained from the at-site approach.

Considering the ability to select a suitable candidate parametric function to construct the regional growth curve and to choose the best fitting distribution, the original Z measure should be used when a homogeneous group is formed by $7 \leq N \leq 15$ series/sites. This measure may be replaced by its bivariate extension, proposed by Kjeldsen and Prosdocimi (2015), when $N > 15$. Under the additive approach, the discordance RD measure, proposed by Neykov et al. (2007), and the original d measure presented equivalent performances.

AUTHORS' CONTRIBUTION

Conceptualization: Martins, L. L., Souza, J. C., Sobierajski, G. R. and Blain, G. C.; **Investigation:** Martins, L. L., Souza, J. C., Sobierajski, G. R. and Blain, G. C.; **Methodology:** Martins, L. L., Souza, J. C., Sobierajski, G. R. and Blain, G. C.; **Formal analysis:** Sobierajski, G. R. and Blain, G. C.; **Data acquisition:** Blain, G. C.; Software: Blain, G. C.; **Validation:** Blain, G. C.; **Writing – original draft:** Martins, L. L., Souza, J. C., Sobierajski, G. R. and Blain, G. C.; **Writing – review & editing:** Martins, L. L., Souza, J. C., Sobierajski, G. R. and Blain, G. C.; **Visualization:** Blain, G. C.; **Supervision:** Blain, G. C.

DATA AVAILABILITY STATEMENT

The meteorological data used in this study belongs to the Agronomic Institute and is available by request (clima@iac.sp.gov.br).

FUNDING

Conselho Nacional de Desenvolvimento Científico e Tecnológico
https://doi.org/10.13039/501100003593
Grant No: 307616/2019-3

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
https://doi.org/10.13039/501100002322
Finance Code 001
Grant No. 33081018001P4 – Agricultura tropical e subtropical

ACKNOWLEDGMENTS

To the anonymous reviewer that suggested the use of the additive approach.

To the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

To the National Council for Scientific and Technological Development for providing the grant for the fourth author (CNPq; Process 307616/2019-3).

REFERENCES

Alvarez, E., Duque, A., Saldarriaga, J., Cabrera, K., Las Salas, G., Valle, I. D., Lema, A., Moreno, F., Orrego, S. and Rodriguez, L. (2012). Tree above-ground biomass allometries for carbon stocks estimation in the natural forests of Colombia. Forest Ecology and Management, 267, 297-308. https://doi.org/10.1016/j.foreco.2011.12.013

Basu, B. and Srinivas, V. V. (2013). Formulation of a mathematical approach to regional frequency analysis. Water Resources Research, 49, 6810-6833. https://doi.org/10.1002/wrcr.20540

Blain, G. C. (2011). Incorporating climate trends in the stochastic modeling of extreme minimum air temperature series of Campinas, state of São Paulo, Brazil. Bragantia, 70, 952-957. https://doi.org/10.1590/S0006-87052011000400031

Blain, G. C. and Lulu, J. (2011). Valores anuais absolutos de séries diárias de temperatura máxima e mínima do ar no estado de São Paulo: tendências, periodicidades e probabilidades associadas. Revista Brasileira de Meteorologia, 26, 29-40. https://doi.org/10.1590/S0102-77862011000100004

Blain, G. C., Avila, A. M. H. and Pereira, V. R. (2018). Using the normality assumption to calculate probability-based standardized drought indices: selection criteria with emphases on typical events. International Journal of Climatology, 38, e418-e436. https://doi.org/10.1002/joc.5381

Blain, G. C., Sobierajski, G. R., Xavier, A. C. F. and Carvalho, J. P. (2021). Regional Frequency Analysis applied to extreme rainfall events: evaluating its conceptual assumptions and constructing null distributions. Anais da Academia Brasileira de Ciências, 93, e20190406. https://doi.org/10.1590/0001-3765202120190406

Bradley, A. A. (1998). Regional frequency analysis methods for evaluating changes in hydrologic extremes. Water Resources Research, 34, 741-750. https://doi.org/10.1029/98WR00096

Caporali, E., Cavigli, E. and Petrucci, A. (2008). The index rainfall in the regional frequency analysis of extreme events in Tuscany (Italy). Environmetrics, 19, 714-724. https://doi.org/10.1002/env.949

Castellarin, A., Burn, D. H. and Brath, A. (2008). Homogeneity testing: how homogeneous do heterogeneous cross-correlated regions seem. Journal of Hydrology, 360, 67-76. https://doi.org/10.1016/j.jhydrol.2008.07.014

Coles, S. (2001). An introduction to statistical modeling of extreme value. London: Springer.

Dalrymple, T. (1960). Flood frequency analysis. Geological Survive Water Supply Paper, 1543-A, 11-51. https://doi.org/10.3133/wsp1543A

Fontolan, M., Xavier, A. C., Pereira, H. R. and Blain, G. C. (2019). Using climate change models to assess the probability of weather extremes events: a local scale study based on the generalized extreme value distribution. Bragantia, 78, 146-157. https://doi.org/10.1590/1678-4499.2018144

Fowler, H. J. and Kilsby, C. G. (2003). A regional frequency analysis of United Kingdom extreme rainfall from 1961 to 2000. International Journal of Climatology, 23, 1313-1334. https://doi.org/10.1002/joc.943

[GFSP] Global Food Security Programme. (2015). Extreme weather and resilience of the global food system. Final project report from the UK–US taskforce on extreme weather and global food system resilience. United Kingdom: The Global Food Security Programme.

Goudenhoofdt, E., Delobbe, L. and Willems, P. (2017). Regional frequency analysis of extreme rainfall in Belgium based on radar estimates. Hydrology and Earth System Sciences, 21, 5385-5399. https://doi.org/10.5194/hess-21-5385-2017

Hansen, C. R. (2015). Comparison of regional and at-site frequency analysis methods for the estimation of southern Alberta extreme rainfall. Canadian Water Resources Journal, 40, 325-342. https://doi.org/10.1080/07011784.2015.1060871

Hosking, J. R. M. and Wallis, J. R. (1988). The effect of inter-site dependence on regional flood frequency analysis. Water Resources Research, 24, 588-600. https://doi.org/10.1029/WR024i004p00588

Hosking, J. R. M. and Wallis, J. R. (1993). Some statistic useful in regional frequency analysis. Water Resources Research, 29, 271-281. https://doi.org/10.1029/92WR01980

Hosking, J. R. M. and Wallis, J. R. (1997). Regional frequency analysis: an approach based on L-moments. Cambridge: Cambridge University Press.

Hubert, M., Rousseeuw, P. J. and Vanden Branden, K. (2005). ROBPCA: a new approach to robust principal component analysis. Technometrics, 47, 64-79. https://doi.org/10.1198/004017004000000563

Kjeldsen, T. R. and Jones, D. A. (2006). Prediction uncertainty in a median-based index flood method using L moments. Water Resources Research, 42. https://doi.org/10.1029/2005WR004069

Kjeldsen, T. R. and Prosdocimi, I. (2015). A bivariate extension of the Hosking and Wallis goodness-of-fit measure for regional distributions. Water Resources Research, 51, 896-907. https://doi.org/10.1002/2014WR015912

Masselot, P., Chebana, F. and Ouarda, T. B. M. J. (2017). Fast and direct nonparametric procedures in the L-moment homogeneity test. Stochastic Environmental Research and Risk Assessment, 31, 509-522. https://doi.org/10.1007/s00477-016-1248-0

Neykov, N. M., Neytchev, P. N., Van Gelder, P. H. A. J. M. and Todorov, V. K. (2007). Robust detection of discordant sites in regional frequency analysis. Water Resources Research, 43. https://doi.org/10.1029/2006WR005322

Nidhin, K. and Chandran, C. (2013). Importance of generalized logistic distribution in extreme value modeling. Applied Mathematics, 4, 560-573. https://doi.org/10.4236/am.2013.43080

Puma, M. J., Bose, S., So, Y. C. and Cook, B. I. (2015). Assessing the evolving fragility of the global food system. Environmental Research Letters, 10, 024007. https://doi.org/10.1088/1748-9326/10/2/024007

Renard, B. (2011). A Bayesian hierarchical approach to regional frequency analysis. Water Resources Research, 47. https://doi.org/10.1029/2010WR010089

Requena, A. I., Chebana, F. and Mediero, L. (2016). A complete procedure for multivariate index-flood model application. Journal of Hydrology, 535, 559-580. https://doi.org/10.1016/j.jhydrol.2016.02.004

Rousseeuw, P. J. and Leroy, A. M. (1987). Robust regression and outliers detection. New York: John Wiley & Sons.

Sansigolo, C. A. (2008). Distribuições de extremos de precipitação diária, temperatura máxima e mínima e velocidade do vento em Piracicaba, SP (1917-2006). Revista Brasileira de Meteorologia, 23, 341-346. https://doi.org/10.1590/S0102-77862008000300009

Santos, J. F., Portela, M. M. and Pulido-Calvo, I. (2011). Regional Frequency Analysis of droughts in Portugal. Water Resources Management, 25, 3537. https://doi.org/10.1007/s11269-011-9869-z

Santos Júnior, E. P., Xavier, A. C. F., Martins, L. L., Sobierajaski, G. R. and Blain, G. C. (2022). Using a regional frequency analysis approach for calculating the Standardized Precipitation Index: an operational approach based on the two-parameter gamma distribution. Theoretical and Applied Cimatology, 48, 1199-1216. https://doi.org/10.1007/s00704-022-03989-7

Sentelhas, P. C., Ortolan, A. A. and Pezzopane, J. R. M. (1995). Estimativa da temperatura mínima de relva e da diferença de temperatura entre o abrigo e a relva em noites de geada. Bragantia, 54, 437-445. https://doi.org/10.1590/S0006-87051995000200023

Sung, J. H., Kim, Y. O. and Jeon, J. J. (2018). Application of distribution-free nonstationary regional frequency analysis based on L-moments. Theoretical and Applied Climatology, 133, 1219-1233. https://doi.org/10.1007/s00704-017-2249-8

Svensson, C. and Jones, D. A. (2010). Review of rainfall frequency estimation methods. Journal of Flood Risk Management, 3, 296-313. https://doi.org/10.1111/j.1753-318X.2010.01079.x

Viglione, A., Laio, F. and Claps, P. (2007). A comparison of homogeneity tests for regional frequency analysis. Water Resources Research, 43. https://doi.org/10.1029/2006WR005095

Vogel, E., Donat, G. M., Alexander, L. V., Meinshausen, M., Ray, D. K., Karoly, D., Meinshauesen, N. and Frieler, K. (2019). The effects of climate extremes on global agricultural yields. Environmental Research: Letters, 14, 054010. https://doi.org/10.1088/1748-9326/ab154b

Wang, Z., Zhaovang, Z., Lai, C., Lin, W., Wu, X. and Chen, X. (2017). A regional frequency analysis of precipitation extremes in Mainland China with fuzzy c-means and L-moments approaches. International Journal of Climatology, 37, 429-444. https://doi.org/10.1002/joc.5013

Wilks, D. S. (2011). Statistical methods in the atmospheric sciences. New York: Academic Press. https://doi.org/10.1016/C2017-0-03921-6