Adapting the regional frequency analysis based on L-moments to improve the standardized precipitation-evapotranspiration index

Letícia Lopes Martins[1] (iD), Graciela da Rocha Sobierajski[2] (iD), Gabriel Constantino Blain[3,*] (iD)

1. Instituto Agronômico ROR – Programa de Pós-graduação em Agricultura Tropical e Subtropical – Campinas (SP), Brazil.
2. Instituto Agronômico ROR – Centro de Fruticultura – Jundiaí (SP), Brazil.
3. Instituto Agronômico ROR – Centro de Biossistemas Agrícolas e Pós-Colheita – Campinas (SP), Brazil.

ABSTRACT: The standardized precipitation-evapotranspiration index (SPEI) is a widely used probability-based method that categorizes drought and wet events based on their expected frequency of occurrence. Regional frequency analysis based on L-moments (RFA-Lmom) is often employed to enhance probabilistic assessments of extreme hydrological data, which typically assume positive values. This study investigated whether applying the RFA-Lmom regionalization technique to the difference between precipitation and potential evapotranspiration (P-PE), the input variable for SPEI, can improve the index's ability to meet the normality assumption. We conducted analyses using Monte Carlo experiments, accounting for distinct climate conditions worldwide, and a case study in the watershed of the Piracicaba, Capivari and Jundiaí rivers, situated in the states of São Paulo and Minas Gerais, Brazil. In this region, P-PE frequency distributions may exhibit negative and positive sample means. Our findings suggested that modifying the RFA-Lmom method by replacing the L-moment ratio L-CV with the L-moment scale measure and using an additive model (instead of the original multiplicative procedure) for SPEI calculation allows the application of this regionalization technique to P-PE amounts. This adapted RFA-Lmom consistently enhanced the ability of SPEI frequency distributions to meet their normality assumption, thereby improving the quality of drought assessments based on this standardized index.

Key words: drought, additive approach, normality assumption, L-moments.

INTRODUCTION

The standardized precipitation-evapotranspiration index (SPEI), introduced by Vicente-Serrano et al. (2010), has gained widespread recognition for its utility in assessing drought frequency and intensity across various regions worldwide. In comparison to the standardized precipitation index (SPI), another widely used drought index developed by McKee et al. (1993)[1], the SPEI offers a more comprehensive assessment of water availability. This advantage arises from the fact that the SPEI takes into account not only precipitation (P), but also potential evapotranspiration rates (PE), making it a more encompassing measure of water availability (Vicente-Serrano et al. 2010, Beguería et al. 2014, Stagge et al. 2015, Pereira et al. 2018).

As a probability-based standardized drought index, the SPEI calculation algorithm relays upon two key steps:
- Fitting an appropriate parametric distribution to its input variable to estimate its cumulative probabilities;
- Transforming these probabilities into normally distributed estimates with zero mean and unit variance.

---

[1] Mckee, T. B., Doesken, N. J. and Kleist, J. (1993). The relationship of drought frequency and duration to time scales. In 8th Conference on Applied Climatology (p. 179-184). Boston: American Meteorological Society.

These two steps were proposed by McKee et al. (1993) as an attempt to ensure that the index estimates become normalized to a location and normalized in time. More specifically, while step 1 attempts to quantify the probability of the index's input variable, step 2 attempts to transform these probabilities into a standardized index that facilitates quantitative comparisons of drought occurrence at different locations and over different timescales (Lloyd-Hughes and Saunders 2002, Laimighofer and Laaha 2022). Any statistical method that enhances the outcomes of these steps holds the potential to improve the quality and accuracy of drought assessments based the SPEI (Guttman 1998, Wu et al. 2007, Stagge et al. 2015, Vicente-Serrano and Beguería 2016, Blain et al. 2018, Pieper et al. 2020, Blain et al. 2022, Santos Junior et al. 2022).

The regional frequency analysis based on L-moments (RFA-Lmom; Hosking and Wallis, 1993, 1997) has gained widespread use in improving the probabilistic assessment of extreme hydrological phenomena, such as rainfall and flood events (Caporali et al. 2008, Goudenhoofdt et al. 2017). Unlike the traditional 'at-site' approach, the RFA-Lmom allows for the regional pooling of data from multiple sites or weather stations. This regionalization technique enhances the accuracy of quantile estimates for extreme (hydro)meteorological events (Fowler and Kilsby 2003). Furthermore, RFA-Lmom offers advantages as the assessment of data quality by identifying anomalous series or sites within a region (Hosking and Wallis 1997, Neykov et al. 2007). This quality assessment relies on a discordance measure, which identifies sites with statistical properties significantly different from others in the same region or group. Additionally, the technique facilitates the estimation of the probability of a specific event at ungauged sites (target points) by leveraging information from nearby gauged sites (Renard 2011).

Motivated by these advantageous qualities, we hypothesized that the RFA-Lmom can be used to calculate the SPEI on a regional basis. In this context, the goal of this study was to investigate whether the RFA-Lmom may be applied to P-PE amounts and what its effect on the ability of the SPEI estimates to meet their normality assumption is. In this study, the SPEI calculated using the RFA-Lmom is referred to as RegionalSPEI. In order to facilitate the reproducibility of our study and results, we developed a set of computational functions capable of calculating all RFA's statistics under the additive approach, as well as the RegionalSPEI. These computational functions, developed in the R software environment, are available at https://github.com/gabrielblain/RegionalSPEI and can be freely installed using the following commands: library(devtools); install_github("gabrielblain/RegionalSPEI").

DATA AND METHODS

Data

This study made use of two distinct datasets (Fig. 1). The first, referred to as "database 1," consists of monthly P-PE amounts from 11 distinct global locations. These data were sourced from the R-package SPEI, developed by Beguería and Vicente-Serrano in 2017. The dataset spans from 1900 to 2020 and includes data series representing various climate types, encompassing: Tropical (Tampa/United States of America and São Paulo/Brazil), Monsoon (Indore/India), Mediterranean (Valencia/Spain), Semiarid (Albuquerque/United States of America), Continental (Wien/Austria), Cold (Punta Arenas/Chile), Oceanic (Abashiri/Japan), Kimberley (South Africa), Lahore (India), Punta Arenas (Chile), and Helsinki (Finland). These locations were selected to represent a diverse range of climatic conditions (Vicente-Serrano et al. 2010; Fig. 1a).

The second database (Fig. 1b), referred to as "database 2," consists of meteorological data from the watershed of the Piracicaba, Capivari, and Jundiaí (PCJ) rivers situated in the states of Minas Gerais and São Paulo, Brazil (Fig. 1b). The PCJ ranks among the most important basins in the state of São Paulo, covering 76 cities (15,320 km2) with ~5.8 million habitants (CBH-PCJ 2020). The region has intense economic, agricultural, and industrial activities (CBH-PCJ 2018), and the Cantareira System–one of the largest water supply systems in Brazil–is situated within this basin. This latter system supplies water to approximately 8.8 million people (CBH-PCJ 2020) living in the metropolitan region of São Paulo. Since 1961, this water reservoir has experienced an increase in the severity and frequency of droughts (Nobre 2016).

We calculated monthly P-PE data for the PCJ region using rainfall series of 34 gauges (Fig. 1b) belonging to Department of Water and Electricity (DAEE, 2023) and air temperature, incoming solar radiation, wind speed, and relative humidity data sourced from the Prediction of Worldwide Energy Resources project (NASA POWER, 2023). We downloaded these

gridded data using the R-package PowerSDI (Blain et al. 2024; version 1.0.0), which uses the NASA POWER API Client (Sparks 2023) to download the data. The PowerSDI package also calculates PE totals (FAO-56 Penman-Monteith method; Eq. 1; Allen et al. 1998) at the selected time scale. This gridded dataset covers a spatial extent of ½º latitude x 5/8º longitude.

$$PE = \frac{0.408\Delta(R_n - G) + \gamma\frac{900}{T + 273}u_2(e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \qquad (1)$$

where: R_n: the net radiation (MJ·m^{-2}·day^{-1}); G = the soil heat flux density (MJ·m^{-2}·day^{-1}); T = the daily mean air temperature (°C) at 2 m, based on the average of maximum and minimum temperatures; u_2 = the average wind speed at 2-m height (m·s^{-1}); e_s = the saturation vapor pressure (kPa); e_a = the actual vapor pressure (kPa); $(e_s - e_a)$: the saturation vapor pressure deficit ($\Delta e, kPa$) at temperature T; Δ = the slope of the saturated vapor pressure curve (kPa·°C^{-1}); γ = the psychometric constant (0.0677 kPa·°C^{-1}).
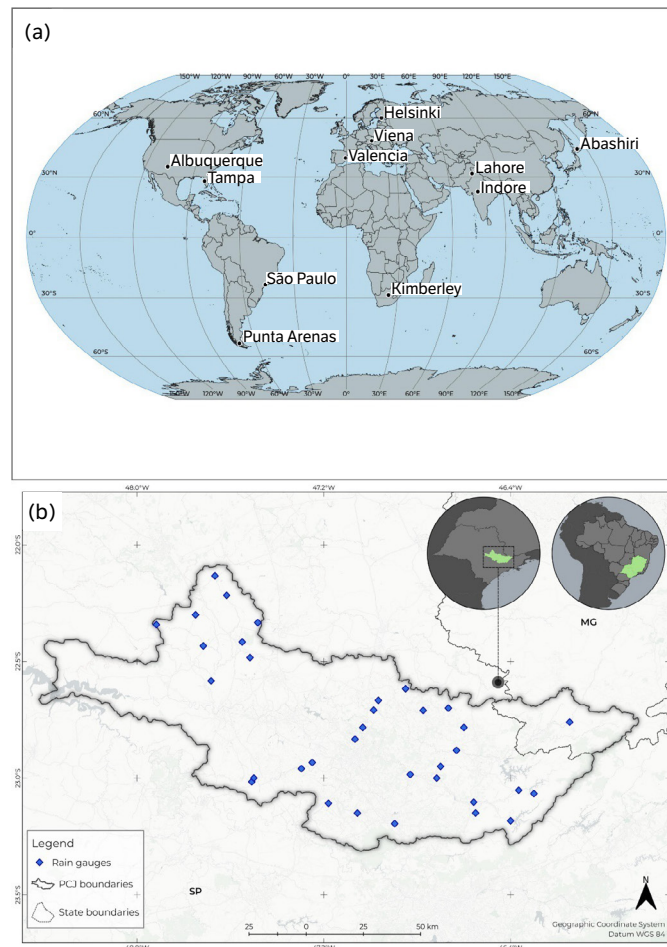
Figure 1. Location of the data used in (a) database 1 and (b) database 2.

Theoretical background: SPEI and RFA-Lmom algorithms

The SPEI and the importance of meeting the assumption of normality

Standardized drought indices, such as the SPEI, are probability-based algorithms designed to measure normalized anomalies in a specific variable accumulated over a defined time scale (Guttman 1998, Wu et al. 2007, Vicente-Serrano et al. 2010, Stagge et al. 2015, Blain et al. 2018, 2022, Santos Junior et al. 2022). The SPEI calculation begins by fitting a parametric

distribution to P-PE amounts accumulated at a predetermined time scale. Subsequently, the cumulative probability of each P-PE amount, denoted as H(x), is computed (Eq. 2), and Eqs. 3 and 4 are applied to H(x). The equiprobability transformation performed by Eqs. 3 and 4 (Abramowitz and Stegun 1965) is an approximation to the quantile function of the standard normal distribution, and it attempts to convert H(x) into standard normal variates [X~N(0,1)]. Like other standardized drought indices, the SPEI algorithm aims to normalize its input variable to facilitate both temporal and spatial comparisons of dry and wet events (Lloyd-Hughes and Saunders 2002) across distinct climate conditions. The SPEI calculation algorithm is described as follows (Eqs. 2 to 4).

$$H(x) = Pr\{x \leq X\} = F(x) \tag{2}$$

where: x = a given P-PE amount; $F(x)$ = the cumulative probability function of the distribution used to calculate the SPEI.

$$SPEI = -\left(t - \frac{c_0 + c_1 t + t^2 c_2}{1 + t d_1 + t^2 d_2 + t^3 d_3}\right) \text{ when } 0.0 < H(x) \leq 0.5 \tag{3}$$

$$SPEI = +\left(t - \frac{c_0 + c_1 t + t^2 c_2}{1 + t d_1 + t^2 d_2 + t^3 d_3}\right) \text{ when } 0.5 < H(x) \leq 1 \tag{3.1}$$

$$t = \sqrt{\left(ln\left(\frac{1}{(H(x))^2}\right)\right)} \text{ when } 0 < H(x) \leq 0.5 \tag{4}$$

$$t = \sqrt{\left(ln\left(\frac{1}{(1 - H(x))^2}\right)\right)} \text{ when } 0.5 < H(x) < 1.0 \tag{4.1}$$

where: c_0 = 2.515517; c_1 = 0.802853; c_2 = 0.010328; d_1 = 1.432788; d_2 = 0.189269; d_3 = 0.001308.

Upon achieving the equiprobabilistic transformation (Eqs. 3 and 4), the SPEI becomes invariant in both time and space domains. In other words, when the assumption of normality is met, the SPEI estimates occur at the frequencies outlined in Table 1, which categorizes the severity of dry and wet events based on their probability of occurrence (Blain et al. 2022). This explains why previous studies, such as Wu et al. (2007), Stagge et al. (2015), Blain et al. (2018), and Pieper et al. (2020), have employed various normality checking procedures to select suitable distributions for calculating standardized drought indices. Finding the distribution that best approaches the indices' frequency distributions to N(0,1) is equivalent to accurately estimating the cumulative probabilities H(x), which is equivalent to identifying the correct probability distribution of P minus PE. Therefore, it is essential to evaluate the impact of this regionalization technique on the SPEI series' ability to meet the normality assumption. This evaluation is elaborated upon in the next sections.

Table 1. Standardized precipitation-evapotranspiration index (SPEI) classification system.

SPEI Value	Category	Cumulative Probability	Expected frequency (%)
SPEI ≥ 2.00	Extreme wet	0.977–1.000	2.3
1.50 < SPEI ≤ 2.00	Severe wet	0.933–0.977	4.4
1.00 < SPEI ≤ 1.50	Moderate wet	0.841–0.933	9.2
-1.00 < SPEI ≤ 1.00	Near normal	0.159–0.841	68.2
-1.50 < SPEI ≤ -1.00	Moderate drought	0.067–0.159	9.2
-2.00 < SPEI ≤ -1.50	Severe drought	0.023–0.067	4.4
SPEI ≤ -2.00	Extreme drought	0.000–0.023	2.3

The RFA-Lmom: the multiplicative and the new additive approaches

The core concept behind regional frequency analysis is to leverage data from multiple sites to estimate the frequency distribution of a variable at a specific target point. In this context, a critical assumption of this technique is that the frequency distributions of the sites within a homogeneous region or group are essentially identical, differing only by a local scaling factor. This assumption forms the basis of the multiplicative approach, which is described by Eq. 5.

$$Q_i (F) = \mu_i q(F) \quad i=1, \ldots N \quad \text{and } 0 < F < 1 \tag{5}$$

where: Q(F) = the quantile of a non-exceedance probability F at site i; μ = the local scaling factor; q(F) = the regional quantiles, defining a dimensionless curve (regional growth curve) common to all sites.

The local scaling factor (μ), also known as the index flood due to its historical application in hydrological studies (e.g., Dalrymple 1960), plays a key role in the regional frequency analyses. Under the multiplicative approach, the values of each data series subjected to RFA-Lmom are divided by their respective index flood (e.g., their sample mean or median), ensuring that the resulting series has an average value approaching 1. The approach described by Eq. 5 has been widely used to enhance the probabilistic assessment of extreme rainfall data (e.g., Blain et al. 2021). However, Martins et al. (2022) identified a limitation with the assumption supporting Eq. 5. This limitation arose when applying this latter approach to air temperature series. The issue stemmed from the fact that observed values for this meteorological element tend to fall within a narrow range compared to their distance from absolute zero (0K corresponds to -273.15°C or -459.67°F). Consequently, their dispersions may be considered substantial and independent of their sample means. In response to this challenge, Martins et al. (2022) introduced an additive approach, which assumes that the frequency distributions of series within a homogeneous region or group are identical, differing only by an additive site-specific scaling factor. This alternative assumption (additive approach) gave rise to Eq. 6.

$$Q_i(F) = \mu_i + q(F) \quad i=1, \ldots N \text{ and } 0 < F < 1 \tag{6}$$

In the additive approach, the regional frequency distribution [q(F)] is characterized by a mean value of zero, in contrast to the multiplicative approach where it equals to 1. Another noteworthy aspect of this novel approach is the replacement of the L-CV ratio, a statistic analogous to the coefficient of variation, with the L-moment scale measure (l_2), which is analogous to the standard deviation. This substitution enables the adaptation of the method to variables that encompass both positive and negative values, offering a compelling alternative for customizing the RFA-Lmom algorithm for P-PE data. As pointed out by Hosking and Wallis (1997), any variable's location parameter can be used to define μ (Eqs. 5 and 6). In the context of our study, in which the SPEI attributes a zero value to observed P-PE amounts with a probability of occurrence equal to 0.5 (Eqs. 3 and 4), we have chosen to set μ as the median values of the P-PE series.

The analysis of Eq. 6 underscores the critical importance of delineating a homogeneous region and selecting an appropriate frequency distribution to specify q(F) within the RFA framework. In this regard, Hosking and Wallis (1991, 1997) have introduced three valuable statistics (discordance, heterogeneity, and goodness-of-fit), which offer objective support for making informed decisions at these pivotal stages.

The discordance measure (d) was designed to capture discrepancies arising from incorrect data values, outliers, and changes in the sample mean through the L-moments of a sample. Under the additive approach, this measure represents two L-moments ratios (L-skewness and L-kurtosis) and l_2 as coordinates in a three-dimensional plot. Each series contributes a point to this plot, forming a cloud of points in which the center corresponds to the group's average values for l_2, L-skewness, and L-kurtosis. The distance of a point from the center of this cloud determines its corresponding d value, with greater distances indicating higher discordance. Although Hosking and Wallis (1997) suggested critical limits, such as d = 1.917 (for groups with seven sites), d = 2.917 (for groups with 14 sites), and d = 3 (for groups with 15 or more sites), as a criterion for declaring a site as discordance, they emphasized that, when a pre-specified region or group is deemed as heterogeneous

by the H measure (described below), the series with the largest d values should be verified and possibly removed from the group (regardless of the magnitude of these d values). This strategy was adopted in studies such as Blain et al. (2021), Santos Junior et al. (2022), and in this study.

In the context of RFA-Lmom, cluster analysis methods are commonly employed for the initial grouping of series (Wang et al. 2017). When geographical information alone is used for this pre-delimitation, the heterogeneity measure (H) can serve as an independent test to validate whether the pre-delimited cluster can indeed be considered a homogeneous group (Hosking and Wallis 1997, Blain et al. 2021). A crucial step in calculating the H measure involves comparing the between-site variability of the L-moments ratios of a group of series (t^R and V; Eqs. 7 and 8) with what would be expected from a 'truly homogeneous group' ($V_{expected}$).

$$t^R = \frac{\sum_{i=1}^{N} n_i t^{(i)}}{\sum_{i=1}^{N} n_i} \tag{7}$$

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

$$V = \frac{\sum_{i=1}^{N} n_i (t^{(i)} - t^R)^2}{\sum_{i=1}^{N} n_i} \tag{8}$$

In this study, the $V_{expected}$ values were simulated from the following six-step computational algorithm, which was based on the studies of Hosking and Wallis (1997) and Collier (2011)[2].

- **Step 1:** Fit a "highly flexible" distribution to the regional average L-moments ratios, which were calculated from the observed data. As pointed out by Hosking and Wallis (1997), this "highly flexible" parametric function should be capable of mimicking the range of frequency distributions that the variable under analysis may exhibit. The five-parameter Wakeby distribution was used for such a purpose;
- **Step 2:** Calculate the correlation matrix of the data.
- **Step 3:** Use this correlation matrix to simulate cross-correlated data from a normal distribution. The number of simulated series and their sample size must be the same as those of the observed series (Step 1);
- **Step 4:** Apply the quantile function of the fitted Wakeby distribution to the data generated in Step 3 to obtain synthetic P-PE series. Considering that these series were generated from the same Wakeby distribution, they form a truly homogeneous group by construction;
- **Step 5:** Apply Eqs. 6 and 7 to this truly homogeneous group of synthetic series ($V_{synthetic}$);
- **Step 6:** Repeat Steps 2 to 5 a large number of times and calculate the mean ($\mu_{synthetic}$) and the standard deviation ($\sigma_{synthetic}$) of all $V_{synthetic}$ values. Then, the H measure is calculated according to Eq. 9.

$$H = \frac{V - \mu_{synthetic}}{\sigma_{synthetic}} \tag{9}$$

With regards to critical limits for the H measure, Hosking and Wallis (1997) indicated that the RFA's statistic should not be used as formal significance tests in which the probabilities of type I errors can be pre-specified. Rather, the reference values (e.g., H = 1) should be regarded only as guidelines values to deem a group of series as homogeneous. In this context, authors such as Núñez et al. (2011) regarded the H = 1 as a too strict limit and suggested using H = 2 instead. In fact, assuming that the frequency distributions of $V_{synthetic}$ approach the Gaussian shape, H = 1.64 would correspond to a 10% probability of falsely rejecting the assumption of homogeneity (Castellarin et al. 2008, Masselot et al. 2017, Martins et al. 2022). Considering the lack of previous studies which applied the RFA-Lmom to P-PE series, three distinct critical limits (1.00, 1.64, and 2.00) were considered in this study.

As previously described, selecting a frequency distribution to specify q(F) is a key stage of the RFA. The goodness-of-fit Z measure was proposed by Hosking and Wallis (1997) to verify if a particular parametric function (candidate distribution)

---

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

can be accepted as the true 'regional distribution' for a homogeneous group. This assumption often is accepted when $|Z| \leq 1.64$. Z is calculated by Eqs. 10 to 12.

$$Z = \frac{\tau_4 - t_4^R + B_4}{\sigma_4} \tag{10}$$

$$B_4 = N_{sim}^{-1} \sum_{m=1}^{N_{sim}} \left(t_4^{[m]} - t_4^R \right) \tag{11}$$

$$\sigma_4 = \left[\left(N_{sim} - 1 \right)^{-1} \left\{ \sum_{m=1}^{N_{sim}} \left(t_4^{[m]} - t_4^R \right)^2 - N_{sim} B_4^2 \right\} \right]^{1/2} \tag{12}$$

where: 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 the simulated group; N_{sim} = the number of simulations (t_4^m and N_{sim} are also obtained from the six-step computational algorithm).

In its original version, the Z measure considers only the difference between the group average value for the L-Kurtosis and the L-Kurtosis of the candidate distribution. In this context, Kjeldsen and Prosdocimi (2015) developed a bivariate extension of this goodness-of-fit measure (Z.bivar) that considers the approximate joint normal distribution of the regional L-skewness and L-kurtosis. The Z.bivar measure uses the same synthetic series generated in the six-step computational algorithm to calculate the individual and the joint sampling variability of the regional L-Skewness and L-Kurtosis. Critical levels for the Z.bivar are obtained from the quantile function of the χ^2 distribution with 2 degrees of freedom.

Applying the RFA-Lmom's statistics to P-PE

The first step for verifying if the RFA-lmom can be used to calculate the SPEI is to evaluate whether the statistics d, H, Z, and Z.bivar can be applied to P-PE totals. Therefore, we developed two Monte Carlo experiments to perform such evaluation under the several climate conditions represented by the weather stations of database 1 (Fig. 1a). For these two Monte Carlo experiments, we fitted generalized extreme value distributions (GEV) to each monthly P-PE series of database 1. Then, we used these fitted GEV distributions to generate groups of simulated P-PE series. These groups comprised series with a sample size equal to N = 30 (climatological normal period), varying numbers of series (N = 7, N = 14, and N = 21), and varying levels of spatial correlation representing weak, moderate, and strong dependence among the series ($\rho = 0.3$, $\rho = 0.6$ and, $\rho = 0.9$, respectively). Regarding the RFA-Lmom, Hosking and Wallis (1997) pointed out that d is likely to be useful only for regions with seven or more series, and the gain in the accuracy of the probabilistic assessments achieves its maximum for groups with approximately 20 series. This is the reason why the number of series within the groups varied from 7 to 21. However, we stress that regional heterogeneity can be reliably assessed by the H measure even for regions with N < 7 sites.

Finally, we used the GEV distributions in these two controlled experiments because previous studies (Kjeldsen and Prosdocimi 2015, Stagge et al. 2015, 2016, Vicente-Serrano and Beguería 2016, Blain et al. 2018) found it suitable to calculate the SPEI at distinct time scales. The candidate distributions used to calculate the Z and Z.bivar measures were the GEV (the true parent distribution of the series) and the generalized logistic, generalized normal, generalized Pareto and Pearson type III distributions (Vicente-Serrano et al. 2010, Stagge et al. 2015, 2016, Vicente-Serrano and Beguería 2016, Blain et al. 2018).

Naturally, the frequency distributions of the series forming each groups generated by the above-mentioned procedure are statistically identical. Therefore, under the RFA approach, these groups may be regarded as homogeneous by definition. On such background, in the first Monte Carlo experiment we applied all RFA-Lmom's statistics (d, H, Z and Z.bivar) to each set of homogeneous groups. A set of homogenous groups is that formed by groups generated from a particular weather station of database 1 (Fig. 1a), with the same number of series and level of spatial correlation. In this controlled experiment, we repeated the above-mentioned procedure 10,000 times for each set of homogeneous group and calculated the rate at which the discordance measure falsely detected discordance series (considering the critical limits suggested in Hosking

and Wallis 1997), the H measure falsely regarded the groups as heterogeneous, and the goodness-of-fit measure falsely rejected the use of the GEV distribution.

The second Monte Carlo experiment was similar to the first one. However, we inserted discordant series within each group so that they became heterogeneous by definition. Following Hosking and Wallis (1997), the discordant series had l2 values 30% higher than the others in the group. The other L-moments ratio remained the same. The decision of simulating the discordant series with probability density functions relatively close to the other group enabled us to evaluate the ability of the discordance measure to flag discordant series which may not be easily detected by other forms of consistency analysis (Martins et al. 2022).

Considering the critical limits described above, we calculated the rate at which the H and d measures correctly identified groups as heterogeneous and the series as discordant, respectively. According to the strategy suggested by Hosking and Wallis (1997), when a region is deemed heterogeneous, the discordant series are those that show the largest d values. Therefore, we also determined the rate at which the discordance measure correctly flagged the discordant series by exhibiting the highest values for them (regardless the critical limits). The Z and Z.bivar measures were not considered in this second experiment because they were designed to be applied to homogeneous groups.

On the normality of the RegionalSPEI: database 1

After verifying that the RFA-Lmom's statistics can be applied to P-PE totals, we evaluated the effect of this regionalization technique on the ability of the RegionalSPEI frequency distributions to meet their normality assumption. In other words, we performed a third Monte Carlo experiment designed to evaluate the ability of the RegionalSPEI to meet such a normality assumption.

In this controlled experiment, we calculated the RegionalSPEI to all P-PE series, which formed the homogeneous groups generated in the first Monte Carlo experiment, and we assessed the ability of these series to meet the normality assumption by using the Shapiro-Wilk's test as proposed in Stagge et al. (2015). According to this procedure, a SPEI series fails to meet its normality assumption if the p-value of this hypothesis test is lower than 0.05. This procedure was used in several studies such as Stagge et al. (2016), Vicente-Serrano and Beguería (2016), Blain et al. (2018), and Santos Junior et al. (2022). Similar to the other Monte Carlo experiments, we calculated the rate at which the p-values were lower than 0.05 at each location of database 1.

On the normality of the RegionalSPEI: database 2

The analysis started by applying the discordance measure to all 34 sites forming database 2 to detect those series with gross errors (Hosking and Wallis 1997). Since d flagged no series, we pre-specified homogeneous regions using cluster analyses (CA), which considered only the latitude and longitude of each site described in Fig. 1b. We used only these two geographical coordinates to facilitate the establishment of continuous regions/groups. The CA was based on the Ward's hierarchical method, described in several studies including Hosking and Wallis (1997), and the number of groups were set to six. Although this initial number is often an arbitrary choice, it was set in such a way as to form the highest possible number of homogeneous groups (according to the H measure) with at least three sites.

Following the results of dataset 1 (presented in section "results and discussion"), the H measure was applied to these groups, considering the critical level of H = 2. Within each homogeneous group, we applied the Z and Z.bivar measures to select among distinct candidate distributions the one that best fits the monthly P-PE series. The candidate distributions were the same as those considered in the Monte Carlo experiments. After select the regional frequency distributions, we estimated the RegionalSPI at each site and assessed the ability of its frequency distributions to meet the normality assumption using the Shapiro-Wilk's test as proposed in Stagge et al. (2015). We calculated the d, H, Z, Z.bivar, and RegionalSPI using the set of R functions available at https://github.com/gabrielblain/RegionalSPEI.

## RESULTS AND DISCUSSION

Database 1: homogeneous groups

The rates at which the original d measure wrongly flagged the P-PE series as discordant remained lower than the level of 3% originally considered in Hosking and Wallis (1997) in all trials (Fig. 2). This result indicated that when applied to P-PE series the original d measure tends to behave as a conservative test that rejects its null hypothesis at rates slightly lower than a pre-specified significance level. Martins et al. (2022) observed similar results when applying the RFA-Lmom to air temperature series in the state of São Paulo. These rates remained approximately constant across the locations. In addition, they were not significantly affected by the distinct levels of spatial dependence nor by the number of series within each homogeneous group. These results are in line with Martins et al. (2022) and Santos Junior et al. (2022).
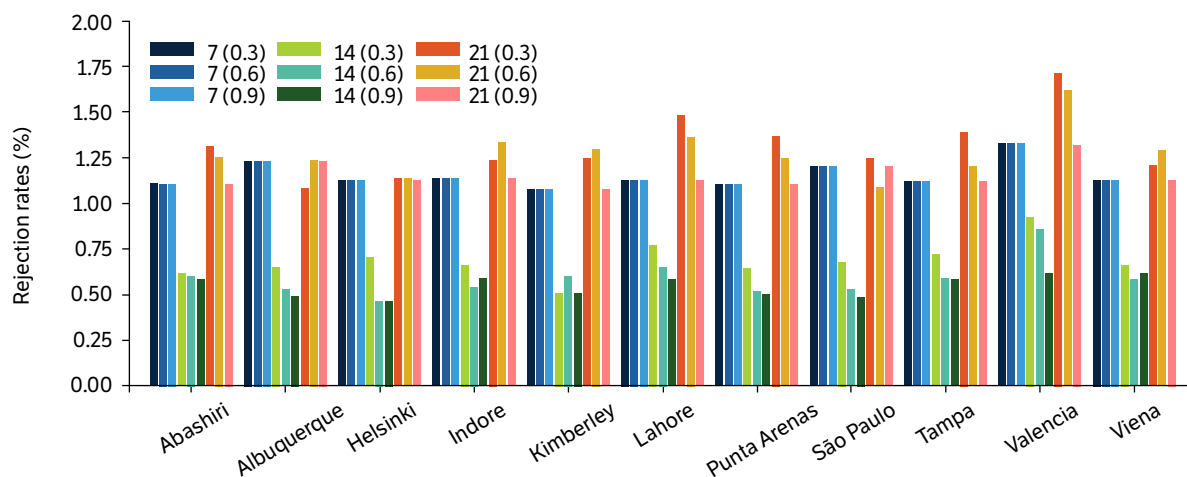
Figure 2. Rates at which the original d measure, calculated under the additive approach proposed in Martins et al. (2022), falsely flagged series of the difference between precipitation and potential evapotranspiration as discordant. 7, 14, and 21 are the number of series within each group, and 0.3, 0.6, and 0.9 are the level of spatial dependence.

The rates at which the H measure failed to accept the (true) assumption of homogeneity (Fig. 3) are comparable to those observed in previous studies that applied this technique to extreme precipitation series (e.g., Hosking and Wallis 1997, Neykov et al. 2007, Blain et al. 2021, among others), extreme air temperature series (Martins et al. 2022), and monthly precipitation series (Santos Junior et al. 2022). This statement holds for all climate conditions and levels of spatial dependence considered in this study (Fig. 3). Naturally, the highest rejection rates were observed when the critical limit was set to $H = 1$. Such rates reached values higher than 20% in some locations, suggesting that $H = 1$ may be regarded as a too strict limit (Castellarin et al. 2008, Masselot et al. 2017). The rejection rates observed when this critical value was set to 2 remained close to 5% in all locations. Thus, we may assume that the rejection rates presented in Fig. 3 favors the adoption of $H = 2$ as a critical limit for the H measure in this study.

With regards to the performances of the goodness-of-fit measures, the rejection rates presented by Z and Z. bivar were comparable to each other (Fig. 4), remaining close to 10% across all locations. These rates were not significantly affected by the distinct levels of spatial dependence nor by the number of series within each homogeneous group.
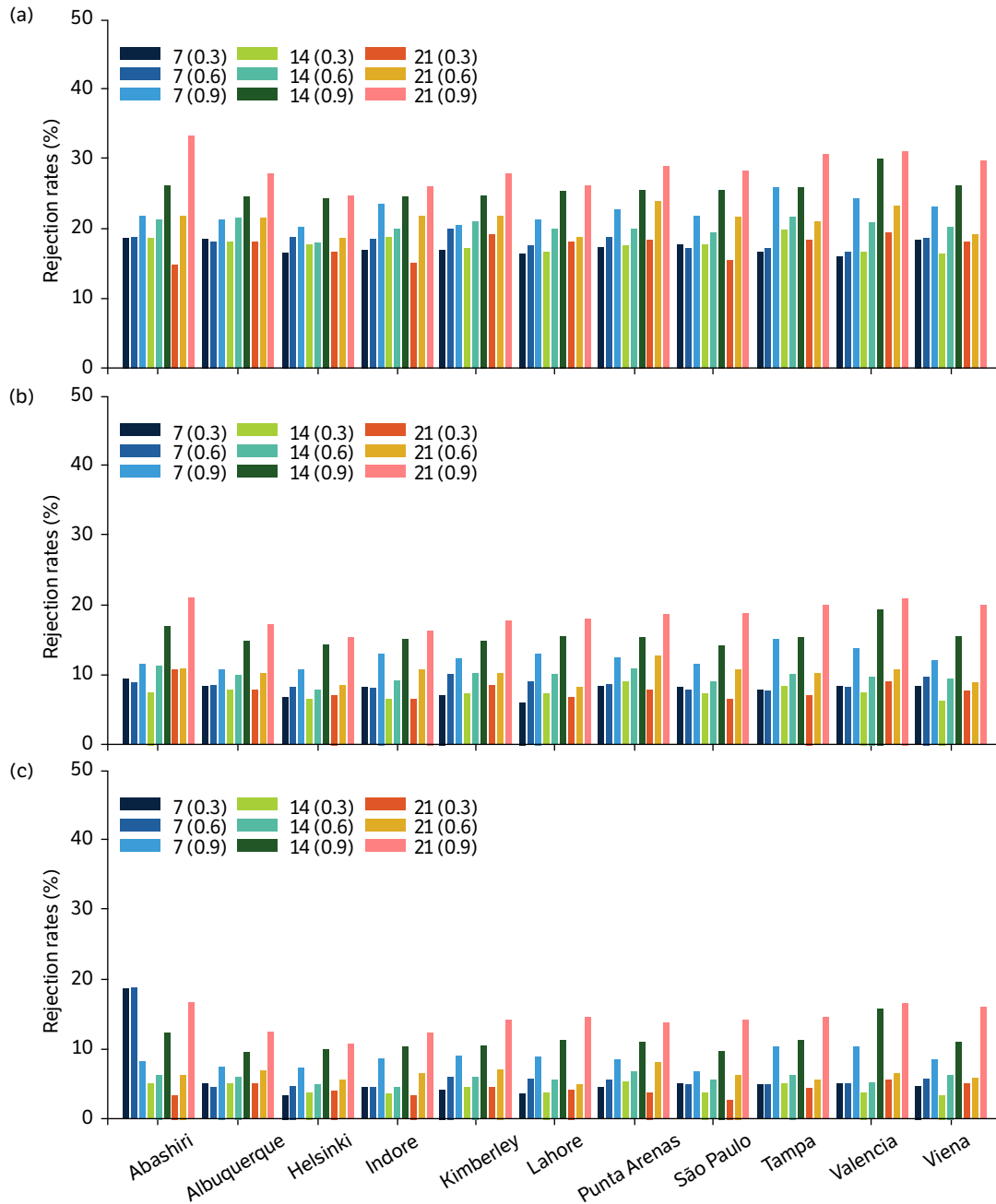
Figure 3. Rates at which the heterogeneity measure, calculated under the additive approach proposed in Martins et al. (2022), falsely rejected the assumption of homogeneity. The critical limits considered in this study were (a) 1, (b) 1.64, and (c) 2. 7, 14, and 21 are the number of series within each group, and 0.3, 0.6, and 0.9 are the level of spatial dependence.

Finally, the results of Fig. 5 indicate that the additional regional frequency analysis did not significantly affect the ability of the SPEI to meet the assumption of normality. As can be observed in Fig. 5, the rejection rates of the Shapiro-Wilk's test remained lower than the adopted significant level (5%) under all climate conditions, number of sites, and levels of spatial correlation considered in this experiment. Authors such as Stagge et al. (2015) and Blain et al. (2018) also considered rejection rates of the Shapiro-Wilk's test equal to or lower than 5% as a suitable threshold for calculating the SPEI in a particular region. In summary, while results of Figs. 2 to 4 indicate that the RFA-Lmom's statistics, calculated under the additional approach, can be applied to P-PE series, the results of Fig. 5 indicate that this technique can be used to calculate the RegionalSPEI.
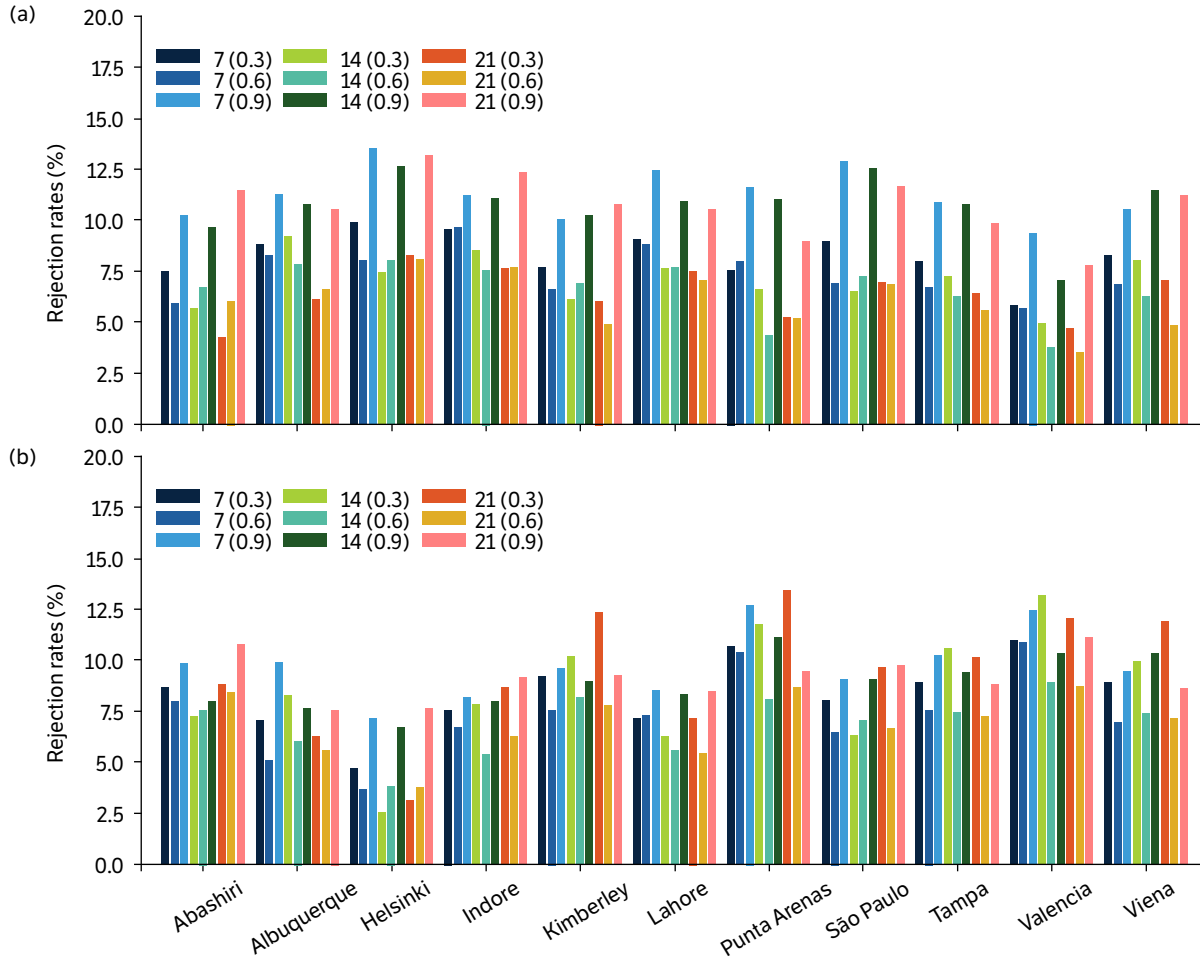
Figure 4. Rates at which wrongly rejected the generalized extreme value distribution as a suitable regional distribution. (a) Rates rejection of the goodness-of-fit measure (Z); (b) its bivariate extension (Z. bivar). 7, 14, and 21 are the number of series within each group, and 0.3, 0.6, and 0.9 are the level of spatial dependence.
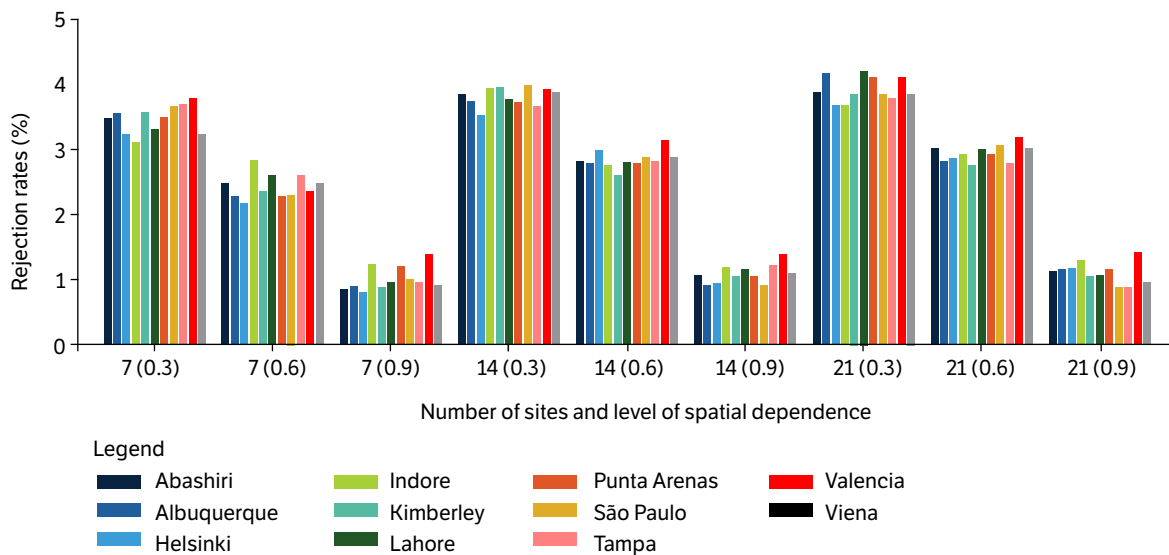
Figure 5. Rejection rates of the Shapiro-Wilk's test (significance level = 5%) under all climate conditions considered in this study. 7, 14, and 21 are the number of series within each group, and 0.3, 0.6, and 0.9 are the level of spatial dependence.

Database 1: heterogeneous groups

As expected, the highest success rates presented by the H measure were observed when the critical limit was set to H = 1. However, these rates were comparable to those obtained when this limit was set to H = 1.64 and H = 2 (Fig. 6). This statement holds for all climate conditions, number of discordant series, number of series forming the group, and levels of spatial dependence considered in this study. At this point, it becomes worth mentioning that Núñez et al. (2011) also adopted H = 2 as a critical limit for mapping droughts in Chile.
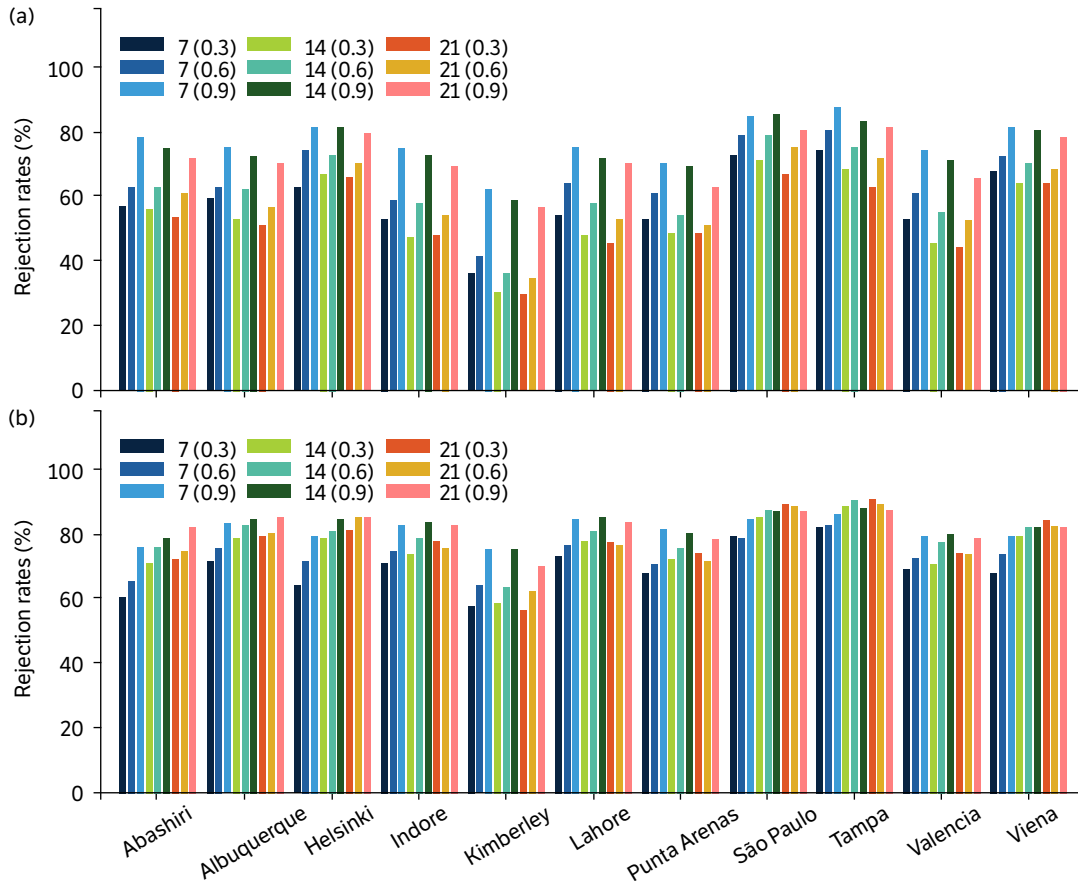
Figure 6. Rejection rates of the H measure (H = 2) for (a) one and (b) three discordant series. 7, 14, and 21 are the number of series within each group, and 0.3, 0.6, and 0.9 are the level of spatial dependence.

Regarding the d measure, the results of Fig. 7 indicate that its performance for exhibiting values higher than its critical thresholds (described in section "Data and methods") decreased as the percentage of discordant series in the groups increased. For instance, the highest rejection rates were often observed for groups with only one discordant series. On the other hand, the lowest rejection rates were observed for groups with the highest percentage of discordant series. Similar results were found by Neykov et al. (2007), Martins et al. (2022), and Santos Junior (2022). They may be explained by the fact that this measure represents l_2, L-Skewness, and L-Kurtosis of each series as a point in a three-dimensional plot (3D-plot). The points which lie far from the center of the cloud of points are flagged as discordant (the center of the cloud is the group average values for the L-moments; Hosking and Wallis 1997, Neykov et al. 2007, Blain et al. 2021). As the number of discordant series increases, their effect (weight) on the average value for the entire group also increases. Consequently, the center of the cloud moves towards them, decreasing the ability of the discordance measure to generate values higher than the critical levels. The performance of this measure was not affected by the level of spatial dependence among the series nor by the distinct climate conditions consider in this study. In this context, the strategy of removing

from the group the series with the largest d (Hosking and Wallis 1997) has shown to be more efficient than the use of the critical limits, since these discordant series often exhibited the highest d values.
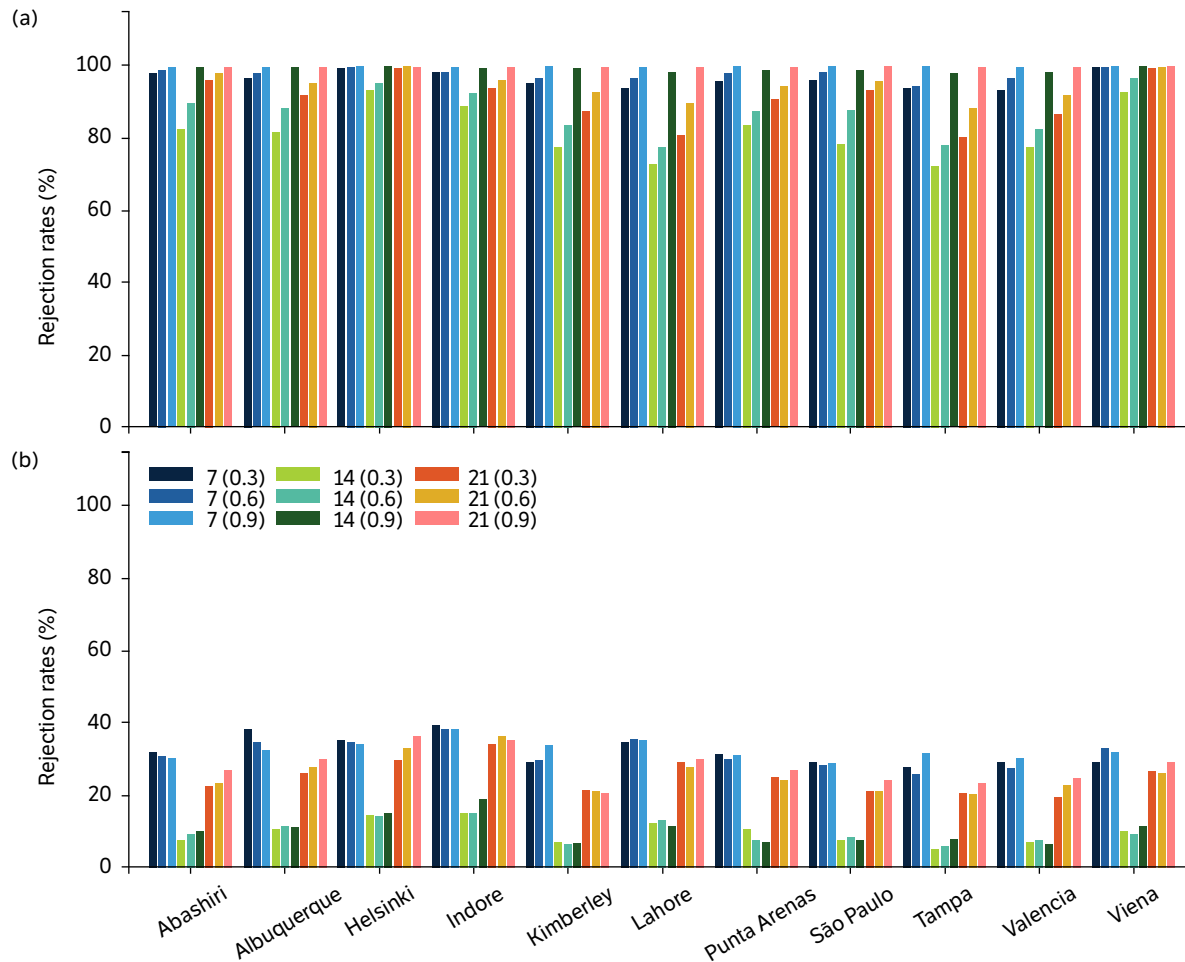

Figure 7. Rejection rates of the d measure for (a) 1 and (b) 3 discordant series. 7, 14, and 21 are the number of series within each group, and 0.3, 0.6, and 0.9 are the level of spatial dependence.

Database 2

Based on the findings from database 1, we established H = 2 as the critical limit for the heterogeneity measure. The homogeneous groups are illustrated in Fig. 8.

Table 2 shows the regional distribution for each group and month selected by the Z and Z.bivar measures. The results of Table 2 are consistent with the statement that even for a particular location the assumption that a single distribution is always able to fit all monthly meteorological series should be regarded with caution (Blain and Meschiatti 2015).

Regarding the ability of the RegionalSPEI series to meet the normality assumption, which is analogous to the capability of the analysis to identify an appropriate probability distribution for the underlying P minus PE data, the rejection rates of the Shapiro-Wilk's test within each were 0 (group 1), 5 (group 2), 2.8 (group 3), 1.9 (group 4), 4.2 (group 5), 1.7 (group 6), and 2.8 (group 7). As previously described, rejection rates of this normality test equal to or lower than 5% may be regarded as a suitable threshold for calculating the SPEI in a particular region (Stagge et al. 2015, Blain et al. 2018). This result further confirms that the addition-RFA does not significantly affect the capability of the SPEI frequency distributions to meet the normality assumption.
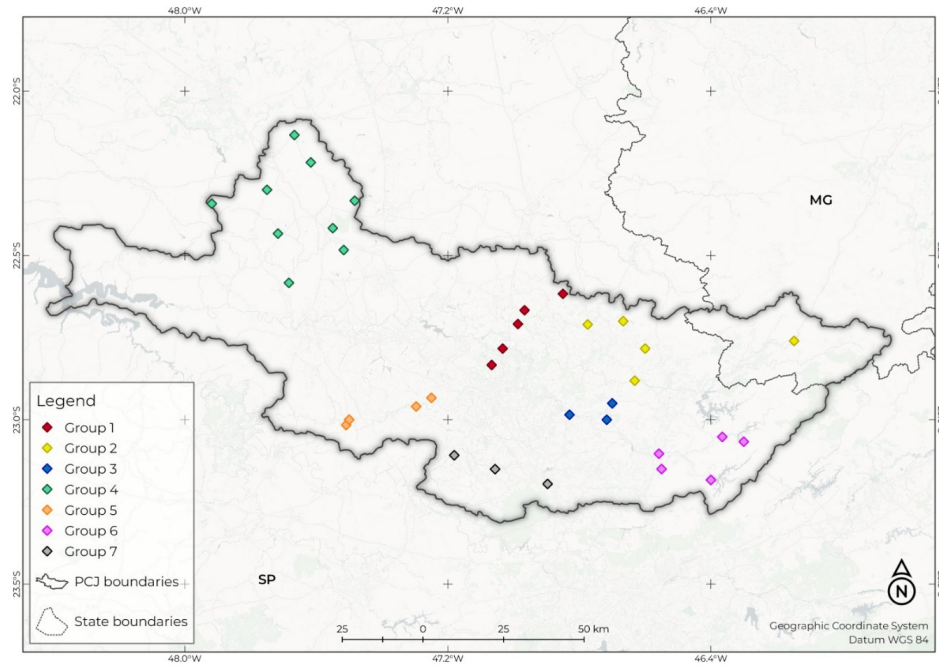
Figure 8. Homogeneous groups of monthly series of the difference between precipitation and potential evapotranspiration in the watershed of the Piracicaba, Capivari, and Jundiaí rivers.

Table 2. Regional frequency distributions for the three homogeneous groups in the watershed of the Piracicaba, Capivari, and Jundiaí rivers. Generalized extreme value (GEV), generalized logistic (GLO), generalized normal (GNO), generalized pareto (GPA) and Pearson type III (PE3) distributions.

Month	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7
Jan	GNO	GNO	GNO	GNO	GNO	GNO	GNO
Feb	GNO	GNO	GNO	GNO	GNO	GNO	GNO
Mar	GLO	GLO	GLO	GLO	GLO	GLO	GLO
Apr	PE3	PE3	PE3	PE3	PE3	PE3	PE3
May	GEV	GEV	GEV	GEV	GEV	GEV	GEV
Jun	GNO	GNO	GNO	GNO	GNO	GNO	GNO
Jul	GPA	GPA	GPA	GPA	GPA	GPA	GPA
Aug	GPA	GPA	GPA	GPA	GPA	GPA	GPA
Sep	PE3	PE3	PE3	PE3	PE3	PE3	PE3
Oct	GLO	GLO	GLO	GLO	GLO	GLO	GLO
Nov	PE3	PE3	PE3	PE3	PE3	PE3	PE3
Dec	GLO	GLO	GLO	GLO	GLO	GLO	GLO

CONCLUSION

The results found in this study support the hypothesis that the RFA-Lmom, calculated under the additive approach, can be applied to the difference between precipitation and evapotranspiration (an important agrometeorological parameter) and to calculate the widely used SPEI. The use of RFA-Lmom in the calculation of the SPEI did not significantly affect the ability of its at-site frequency distributions to meet the normality assumption. Accordingly, the additional regional frequency analysis can be used to identify an appropriate probability distribution for the underlying P minus PE data on the regional scale.

CONFLICT OF INTEREST

Nothing to declare.

AUTHORS' CONTRIBUTION

Conceptualization: Sobierajski, G. R. and Blain, G. C.; I**nvestigation:** Martins, L. L., Sobierajski, G. R. and Blain, G. C.; **Methodology:** 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:** Martins, L. L., Sobierajski, G. R. and Blain, G. C.; **Writing – original draft:** Sobierajski, G. R. and Blain, G. C.; **Writing – review & editing:** Martins, L. L., Sobierajski, G. R. and Blain, G. C.; **Visualization:** Blain, G. C.; **Supervision:** Blain, G. C.

DATA AVAILABILITY STATEMENT

The data used in this study is available at https://github.com/gabrielblain/PowerSDI and https://github.com/gabrielblain/SupplementalFiles_1.

REFERENCES

Abramowitz, M. and Stegun, I. A. (1965). Handbook of mathematical functions with formulas, graphs, and mathematical tables. New York: Dover Publications.

Allen, R. G., Pereira, L. S., Raes, D. and Smith, M. (1998). Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements. FAO Irrigation and Drainage Paper 56. Rome: FAO.

Beguería, S. and Vicente-Serrano, S. M. (2017). SPEI: Calculation of the Standardized Precipitation-Evapotranspiration Index. R package version 1.7. Available at: https://CRAN.R-project.org/package=SPEI. Accessed on: Aug. 10, 2023.

Beguería, S., Vicente-Serrano, S. M., Reig, F. and Latorre, B. (2014). Standardized precipitation evapotranspiration index (SPEI) revisited: parameter fitting, evapotranspiration models, tools, datasets and drought monitoring. International Journal of Climatology, 34, 3001-3023. https://doi.org/10.1002/joc.3887

Blain, G. C. and Meschiatti, M. C. (2015). Inadequacy of the gamma distribution to calculate the Standardized Precipitation Index. Revista Brasileira de Engenharia Agrícola e Ambiental, 19, 1129-1135. https://doi.org/10.1590/1807-1929/agriambi.v19n12p1129-1135

Blain, G. C., De 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., Martins L. L. and Sparks, A. H. (2024). PowerSDI: Calculate Standardised Drought Indices Using NASA POWER Data. R package version 1.0.0. Available at: https://github.com/gabrielblain/PowerSDI. Accessed on: Jan. 16, 2024.

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

Blain, G. C., Sobierajski, G. R., Weight, E., Martins, L. L. and Xavier, A. C. F. (2022). Improving the interpretation of standardizes precipitation index estimates to capture drought characteristics in changing climate conditions. International Journal of Climatology, 42, 5586-5608. https://doi.org/10.1002/joc.7550

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

[CBH-PCJ] Comitê das bacias hidrográficas dos rios Piracicaba, Capivari e Jundiaí (2018). Primeira Revisão do Plano das Bacias Hidrográficas dos rios Piracicaba, Capivari e Jundiaí 2010-2020 (Relatório final – revisão 5 – Tomo I). Piracicaba: Profill e Rhama.

[CBH-PCJ] Comitê das bacias hidrográficas dos rios Piracicaba, Capivari e Jundiaí (2020). Relatório da Situação dos Recursos Hídricos 2020 (ano base 2019); versão simplificada. Piracicaba: Agência das Bacias PCJ.

[DAEE] Departamento de Águas e Energia Elétrica. Banco de dados hidrológicos. DAEE. Available at: http://www.hidrologia.daee.sp.gov.br. Accessed on: Jul. 20, 2023.

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

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

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

Guttman, N. B. (1998). Comparing the Palmer drought index and the standardized precipitation index. Journal of the American Water Resources Association, 34, 113-121. https://doi.org/10.1111/j.1752-1688.1998.tb05964.x

Hosking, J. R. M. and Wallis, J. R. (1991). Some Statistics Useful in Regional Frequency Analysis, Res. Rep. RC 17096. New York: IBM Research Division.

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.

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

Laimighofer, J. and Laaha, G. (2022). How standard are standardized drought indices? Uncertainty components for the SPI & SPEI case. Journal of Hydrology, 613, 1285385. https://doi.org/10.1016/j.jhydrol.2022.128385

Lloyd-Hughes, B. and Saunders, M. A. (2002). A drought climatology for Europe. International Journal of Climatology, 22, 1571-1592. https://doi.org/10.1002/joc.846

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, 1-22. https://doi.org/10.1590/1678-4499.20220061

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

[NASA POWER] NASA Prediction of Worldwide Energy Resources. The Power Project. Available at: https://poert.larc.nasa.gov. Accessed on: Aug. 6, 2023.

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, W06417. https://doi.org/10.1029/2006WR005322

Nobre, C. A., Marengo, J. A., Seluchi, M. E., Cuartas, L. A. and Alves, L. M. (2016). Some Characteristics and Impacts of the Drought and Water Crisis in Southeastern Brazil during 2014 and 2015. Journal of Water Resource and Protection, 8, 252-262. https://doi.org/10.4236/jwarp.2016.82022

Núñez, J. H., Verbist, K., Wallis, J. R., Schaefer, M. G., Morales, L. and Cornelis, W. M. (2011). Regional Frequency analysis for mapping drought events in north-central Chile, Journal of Hydrology, 405, 352-366. https://doi.org/10.1016/j.jhydrol.2011.05.035

Pereira, V. R., Blain, G. C., Avila, A. M. H., Pires, R. C. M. and Silveira Pinto, H. (2018). Impacts of climate change on drought: changes to drier conditions at the beginning of the crop growing season in southern Brazil. Bragantia, 77, 201-211. https://doi.org/10.1590/1678-4499.2017007

Pieper, P., Düsterhus, A. and Baehr, J. (2020). A universal standardized precipitation index candidate distribution function for observations and simulations. Hydrology and Earth System Sciences, 24, 4541-4565. https://doi.org/10.5194/hess-24-4541-2020

Renard, B. (2011). A Bayesian hierarchical approach to regional frequency analysis. Water Resources Research, 47, W11513. https://doi.org/10.1029/2010WR010089

Santos Junior, E. P., Xavier, A. C. F., Martins, L. L., Sobierajski, 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 Climatology, 148, 1199-1216. https://doi.org/10.1007/s00704-022-03989-7

Sparks, A. (2023). nasapower: NASA-POWER Data from R. 10.5281/zenodo.1040727, R package version 4.0.7. Available at: https://CRAN.R-project.org/package=nasapower. Accessed on: Aug. 6, 2023.

Stagge, J. H., Tallaksen, L. M., Gudmundsson, L., Van Loon, A. F. and Stahl, K. (2015). Candidate distribution for climatological drought indices (SPI and SPEI). International Journal of Climatology, 35, 4027-4040. https://doi.org/10.1002/joc.4267

Stagge, J. H., Tallaksen, L. M., Gudmundsson, L., Van Loon, A. F. and Sthal, K. (2016). Response to comment on 'Candidate Distributions for Climatological Drought Indices (SPI and SPEI)'. International Journal of Climatology, 36, 2132-2136. https://doi.org/10.1002/joc.4564

Vicente-Serrano, S. M. and Beguería, S. (2016). Comment on 'candidate distributions for climatological drought indices (SPI and SPEI)' by James H. Stagge et al. International Journal of Climatology, 36, 2120-2131. https://doi.org/10.1002/joc.4474

Vicente-Serrano, S. M., Beguería, S. and Lopez-Moreno, J. I. (2010). A multiscalar drought index sensitive to global warming: the Standardized Precipitation Evapotranspiration Index. Journal of Climate, 23, 1696-1718. https://doi.org/10.1175/2009JCLI2909.1

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 cmeans and L moments approaches. International Journal of Climatology, 37, 429-444. https://doi.org/10.1002/joc.5013

Wu, H., Svoboda, M. D., Hayes, M. J., Wilhite, D. A. and Wen, F. (2007). Appropriate application of the standardized precipitation index in arid locations and dry seasons. International Journal of Climatology, 27, 65-79. https://doi.org/10.1002/joc.1371