The PowerSDI: an R-package for implementing and calculating the SPI and SPEI using data from the NASA-POWER project

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

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

ABSTRACT: The standardized precipitation (SPI) and standardized precipitation-evapotranspiration (SPEI) indices are important tools for monitoring drought events, but the low density of weather station networks limits their use in many regions. To address this issue, we developed the PowerSDI R-package, which calculates these two indices using gridded-data from the NASA-POWER project (NASA-SPI and NASA-SPEI). Different from other packages, the PowerSDI package has two modes: the scientific, and the operational. In the scientific mode, the users may assess the quality of the indices estimates through their agreement with a reference/observed series and through the evaluation of how well these estimates meet the conceptual assumptions required for calculating both SPI and SPEI. This evaluation is based on measure of accuracy (e.g., Willmott index of agreement), goodness-of-fit tests (e.g., Anderson-Darling), and normality tests (e.g., Shapiro-Wilk's test), which are calculated by the ScientSDI.R, Reference.R, and Accuracy.R functions. In the operational mode, users can calculate both indices routinely using the OperatSDI.R function. The package also uses a quasi-weekly time scale, allowing for index calculations four times a month. The OperatSDI.R enables users to download NASA-POWER data for all available period or only for the quasi-week they intend to monitor (reducing the function's running time). In short, the PowerSDI facilitates the routine use of these two widely used drought indices and, unlike others existing software, it provides a solid scientific basis for using NASA-POWER data in drought monitoring systems, which can help improve drought preparedness and response efforts worldwide. The package is freely available at two repositories: Github (https://github.com/gabrielblain/PowerSDI), and CRAN (https://CRAN.R-project.org/package=PowerSDI).

Key words: standardized drought indices, drought monitoring, R-software.

INTRODUCTION

Drought is a slow-moving hazard that affects both human and natural ecosystems. From an operational viewpoint, this phenomenon is often defined as a departure in the current climate conditions with respect to a normal or appropriate threshold, which is frequently taken as the sample mean, median, or another statistical measure of a particular variable (Mishra and Singh 2010, Dai 2011, Blain et al. 2022, Santos Junior et al. 2022).

In this context, distinct probability-based drought indices, such as the standardized precipitation index (SPI; McKee et al. 1993)[1] and the standardized precipitation-evapotranspiration index (SPEI; Vincente-Serrano et al. 2010), have been

---

[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. Boston: American Meteorological Society, 179-184.

widely used by drought monitoring systems throughout the globe and in several academic studies (Guttman 1999, Wu et al. 2005, Russo et al. 2013, Beguería et al. 2014, Li et al. 2015, Stagge et al. 2015, Blain et al. 2018, Rashid and Beecham 2019, Pieper et al. 2020, Blain et al. 2021, Blain et al. 2022, Santos Junior et al. 2022, Martins et al. 2023). The SPI, which requires only rainfall data as its input variable, is also recommended by the World Meteorological Organization as a starting point for meteorological drought monitoring (Hayes et al. 2011, Hao et al. 2017). The SPEI uses both rainfall and potential evapotranspiration data as its input variables. Thus, it may provide a broader description of the drought conditions than the SPI (Vincente-Serrano et al. 2010, Beguería et al. 2014, Stagge et al. 2015, Pereira et al. 2018).

The calculation algorithm of these two standardized drought indices (SDI) relay on two steps. The first step involves calculating the cumulative probabilities of the input variable by fitting a parametric distribution. In the second step, these probabilities are transformed into normally distributed estimates with a 0 mean and unit variance. These two steps can be regarded as an effort to normalize the indices' estimates both in location and over time (Guttman 1999, Wu et al. 2007, Stagge et al. 2015, Blain et al. 2018, Pieper et al. 2020, Blain et al. 2022, Santos Junior et al. 2022).

Despite this widespread use, the low density of weather station networks and data quality issues are the most significant limiting factors for calculating these indices, especially in developing countries (Bardin-Camparotto et al. 2013, Meschiatti and Blain 2016). Among all strategies designed to overcome this difficulty, the use of remote sensing data has emerged as one of the best options. In this context, the NASA-POWER project (https://power.larc.nasa.gov/) has gained popularity as a source for weather data input (Bai et al. 2010, Monteiro et al. 2018, Duarte and Sentelhas 2020).

Unlike other gridded databases, the NASA-POWER provides meteorological and agrometeorological data as early as 1981 or 1991, depending on the variable. Therefore, this project is capable of meeting the 30-year continuous records required for calculating standardized drought indices (McKee et al. 1993). NASA-POWER data can be freely downloaded at https://power.larc.nasa.gov/. Additionally, researchers can use the R-package NASA POWER API Client ('nasapower'), version 4.0.10 (Sparks 2023), available at https://cran.r-project.org/web/packages/nasapower/index.html, to download NASA-POWER data directly within an R session.

The 'nasapower' package enables users to retrieve multiple meteorological and radiation datasets simultaneously, presented as a data frame tibble object. Consequently, the data obtained through this latter package are readily applicable in a diverse range of statistical modelling approaches (Sparks 2023), including estimations of drought indices. Finally, data from the NASA-POWER project has shown good performance in estimating rainfall and potential evapotranspiration data in several regions of the world.

Rodrigues and Braga (2021) assessed the performance of NASA-POWER reanalysis data for estimating daily potential evapotranspiration (PE) data in Alentejo Region, Southern Portugal. They observed a good accuracy ($R^2 > 0.70$) between PE estimated from ground weather stations and PE estimated from row NASA-POWER data (with no bias correction). Al-Kilani et al. (2021) evaluated the performance of this reanalysis dataset for estimating the SPI across Jordan. They found relatively high correlations between rainfall data observed at ground weather stations and those from NASA-POWER ($0.67 \leq R^2 \leq 0.91$). However, they also indicated that further studies, which compare NASA-POWER data with reference/observed data, are required to improve the performance of NASA-POWER data for estimating the SPI.

In this study, we assumed that the accessibility of friendly-use computational packages designed to calculate these two drought indices from data provided by the NASA-POWER project is a key point for the improvement of drought monitoring programs. This improvement is of particular relevance in regions where the availability of weather station data is a matter of concern. We also assumed that these packages should be capable of assessing the quality of the indices estimates by evaluating their agreement with a reference/observed series and by verifying how well these estimates meet their conceptual assumptions (described in details in the next section). Finally, considering that such computational codes may be importante for developing countries, they should be developed in license-free software environments.

In this context, we developed the PowerSDI, which is an R-software package (https://CRAN.R-project.org/package=PowerSDI) capable of calculating the SPI and SPEI from NASA-POWER data (NASA-SPI and NASA-SPEI). The package is based on five major user-friendly R-functions designed to calculate these two indices in both scientific and operational or routine modes. More specifically, the functions ScientSDI.R, Accuracy.R, Reference.R, and PlotData.R may be used to assess, among other features, the ability of the SPI and SPEI frequency distributions to meet the normality

assumption, and how well NASA-POWER estimates represent "real-world" data. Additionally, the OperatSDI.R function facilitates calculating these two indices in an operational or routine mode. The PowerSDI uses the 'nasapower' package (Sparks 2018, 2023) for downloading NASA-POWER daily data and the 'lmom' package (Hosking 2022) for calculating the distribution parameters.

The remainder of the paper is organized as follow: "SPI and SPEI: calculation algorithm and assumptions" describes the SPI and SPEI calculation algorithms, highlighting their conceptual assumptions, which must be taken into account when implementing these indices in a particular region and at a particular time scale. This section also describes two methods for estimating the potential evapotranspiration amounts as required by the SPEI calculation algorithm. Further information regarding NASA-POWER data is also provided. Section "Goodness-of-fit tests, normality-checking procedures and other model performance checking-methods" presents several procedures designed to assess how well the NASA-SPI and NASA-SPEI meet the conceptual. This section also describes model performance checking-methods that are used by the PowerSDI package to verify how well the NASA-POWER data actually represent "real-world/observed" data. Section "The POWERSDI R-package" shows how the PowerSDI package can be used to calculate all methods described in the previous sections. Section "Case studies applications" presents two case studies that evaluated the applicability of the package under distinct climate conditions. While the first case performed a detailed evaluation of the package in the state of São Paulo, Brazil, the second case assessed its applicability in entire Brazil. The last section presents the final remarks of this study, including suggestions for future development (versions) of the Package.

SPI AND SPEI: CALCULATION ALGORITHM AND ASSUMPTIONS

The SPI and SPEI are standardized drought indices that share the same multi-scalar calculation algorithm and were designed to be normalized in both time and space domains (Wu et al. 2007, Vicente-Serrano et al. 2010, Stagge et al. 2015, Blain et al. 2018). Accordingly, the first step of the calculation algorithm of both indices is to fit a parametric distribution to their input data accumulated at specific time scale (Guttman 1999). Theoretically, the SPI may be calculated at time scales as short as one week (Wu et al. 2007). However, time scales ranging from one to 24 months are often used (Wu et al. 2007, Vicente-Serrano et al. 2010, Stagge et al. 2015, Blain et al. 2018). Aiming at enhancing its own flexibility, the PowerSDI adopted a basic time scale that splits each month into four sub-periods: days 1 to 7, days 8 to 14, days 15 to 21, and days 22 to 28, 29, 30, or 31 depending on the month. For instance, if TS = 4, the time scale corresponds to a moving window with a one-month length that is calculated four times each month. If TS = 48, the time scale corresponds to a moving window with a 12-month length that is calculated four times each month. This time scale is referred as to quart.month.

The quart.month time scale is similar to that one adopted in Vicente-Serrano et al. (2022). As pointed out by these authors, standardized drought indices are relative metrics that require homogeneous periods. Therefore, using calendar weeks as the reference periods can be challenging since the first day of each year can fall on different days, causing inconsistency throughout the year. Leap years also add difficulties to this comparison (Vicente-Serrano et al. 2022). Additionally, a quart. month time scale of TS = 4 precisely aligns with the one-month time scale adopted in several studies that use these SDI. This alignment would have been impossible if a fixed time scale (e.g., seven days) had been chosen. Considering that the reliability of both SPI and SPEI estimates is an increasing function of the length of records available for their calculation (Guttman 1999) and that the longer the time scale, the smaller the length of records available for calculating these two SDI, the PowerSDI limited the time scales to values ranging from 1- to 96-quart.month.

With regard to the SPI, the fitted distribution is then used to estimate the cumulative probabilities of rainfall amounts. Although distinct probability functions may be used for such purpose (Guttman 1999), the two-parameter gamma distribution has been widely used to estimate the SPI (e.g., Hayes et al. 1999, Wu et al. 2005, Wu et al. 2007, Stagge et al. 2015, Blain et al. 2018, Blain et al. 2022). Thus, this latter distribution was adopted by the PowerSDI package. Finally, since the rainfall frequency distributions are 0-bounded, a mixed function that joins the probabilities of P = 0 and P > 0 must be applied (Thom 1951; Eq. 1).

$$H(X) = q + (1 + q)G(P > 0; \alpha, \beta) \tag{1}$$

where: $G(P > 0; \alpha, \beta)$: the two-parameter gamma distribution; α: its shape parameter; β: its scale parameter; q: the probability of P = 0.

The PowerSDI estimates q through the Weibull plotting position function, as suggested by Solakova et al. (2014), Stagge et al. (2015), Blain et al. (2018) and Blain et al. (2022) (Eq. 2).

$$q = \frac{n_z + 1}{2(n + 1)} \tag{2}$$

where: n_z: the number of 0; n: the sample size.

Regarding the SPEI, the fitted distribution is used to estimate the cumulative probabilities of the difference between rainfall and potential evapotranspiration amounts (PPE). Thus, this index requires the selection of a method to estimate PE. Three methods are often considered for such a purpose (Beguería et al. 2014): the Thornthwaite method (Thornthwaite 1948), the FAO-56 Penman-Monteith method (Allen et al. 1998), and the Hargreaves & Samani method (Hargreaves and Samani 1985).

The Thornthwaite method is a temperature-based model developed on a monthly basis, with equations derived from monthly air temperature means and the maximum number of sun hours in each month. The adoption of this method would prevent calculating the SPEI on weekly basis. The FAO-56 Penman-Monteith method (PE.PM) is recommended by the Food and Agriculture Organization (FAO) as the standard method to estimate this climatic variable and has been extensively validated throughout the globe. The only drawback of the PE.PM method is that it requires extensive data (Eq. 3). In this context, Droogers and Allen (2002) stated that, if the reliability or availability of the data at hand is a matter of concern, the replacement of the PE.PM model by a simpler method, which requires a smaller number of input data, should be considered. As pointed out by Rodrigues and Braga (2021), the Hargreaves & Samani method (PE.HS) has been widely used in regions where only maximum and minimum air temperature data are available. The PE.HS method (Eq. 4) also requires estimating daily insolation values on a horizontal surface at the top of atmosphere (R_a), which can be easily calculated as a function of the latitude and day of the year. The NASA-POWER project also provides daily R_a values. Thus, the PowerSDI was developed in such a way to calculate both PE.PM and/or PE.HS amounts.

$$PE.PM = \frac{0.408\Delta(R_n - G) + \gamma \dfrac{900}{T + 273} u_2(e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \tag{3}$$

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 (Δe, 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).

$$PE.HS = 0.0223 \times 0.4081633 R_a \times (T_{max} - T_{min})^{0.5} \times (T_{avg} + 17.8) \tag{4}$$

where: R_a: the extraterrestrial radiation (MJ·m⁻² day⁻¹); 0.0223: a factor conversion from America to the International System of Units; T_{avg}: the average air temperature (°C); T_{max}: the maximum air temperature (°C); T_{min}: the minimum air temperature (°C).

After estimating PE, the cumulative probability of PPE values may be then estimated. The SPEI algorithm often uses the generalized extreme value (GEV; Eq. 5.1) or the generalized logistic (GLO; Eq. 5.2) distributions for such a purpose (Vicente-Serrano et al. 2010, Beguería et al. 2014, Stagge et al. 2015, Stagge et al. 2016, Vicente-Serrano and Beguería 2016, Blain et al. 2018). A review of these studies suggests that the performance of these two distributions for calculating the SPEI

tend to be similar to each other over most of the range of the index possible values (e.g., -2.0:2:0). However, these studies also found significant differences between the two probability functions in the lower and upper tails (Vicente-Serrano and Beguería 2016).

In this context, we assumed that the PowerSDI package should also allow the users to choose between these two models when calculating the SPEI. As suggested by Vicente-Serrano and Beguería (2016) and Blain et al. (2018), the L-moments approach (Hosking 1990) was adopted to estimate the distributions' parameters. Finally, Stagge et al. (2015) pointed out that the SPI and SPEI frequency distributions should be truncated to the range between [-3.0:3.0] to avoid the high levels of uncertainties associated with estimates outside these limits. Considering the relatively limited length of the NASA POWER records, the PowerSDI adopted such bounds.

$$H(X) = GEV (PPE; \mu, \sigma, \tau) \tag{5.1}$$

$$H(X) = GLO (PPE; \mu, \sigma, \tau) \tag{5.2}$$

where: μ: the location parameter of the generalized extreme value or generalized logistic distributions. σ: the scale parameter of the generalized extreme value or generalized logistic distributions. τ: the shape parameter of the generalized extreme value or generalized logistic distributions.

The final step of these two SDIs calculation algorithms (Eq. 6) is to transform H(X) into normally distributed variables (standard normal; 0 mean and unit variance; Φ) so that they become normalized to a location and normalized in time (Wu et al. 2007).

$$SDI = \Phi^{-1}[H(x)] \tag{6}$$

Once the assumption of normality is actually met, the distinct SPI or SPEI values occur at the frequencies presented in Table 1, regardless of the region, period of the year or time scale. This standardized nature facilitates quantitative comparisons of drought occurrence at different locations and over different time scales (Lloyd-Hughes and Saunders 2002). However, depending on the climate conditions (e.g., arid climates or regions with a distinct dry season), the SPI at short-time scales may fail to meet the normally assumption and, consequently, fail to properly quantify drought conditions (Wu et al. 2007). This is the reason why normality-checking procedures have been used to evaluate the quality of SPI and SPEI estimates and to select appropriate time scale for their calculation (e.g., Wu et al. 2007, Stagge et al. 2015, Blain et al. 2018, Pieper et al. 2020). As described in the next sections, the PowerSDI proposes using these normality-checking procedure to evaluate the suitability of the NASA-POWER data for calculating both SPI and SPEI in a particular location and at a given time scale.

Table 1. Standardized drought indices (SDI) classification system.

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

As previously described, the NASA-POWER project provides all the data required for calculating both SPI and SPEI. The meteorological data have spatial resolution of 0.5 × 0.625 degrees, while the radiation data have resolution of 1 × 1 degree (using the WGS84 grid reference system). Considering that the PowerSDI package calculates Ra as a function of latitude

and the day of the year, it enables the calculation of standardized drought indices at the finer resolution (0.5 × 0.625) when the Hargreaves & Samani method is used for computing the SPEI. However, if the PM method is employed, the spatial resolution should be at least 1 × 1 degree.

As previously mentioned, the PM method requires additional variables such as net radiation. The PowerSDI package imports the function get_power() from the 'nasapower' package to download the multiple variables required for calculating the indices. This function allows users to download data for a single point, a specific region, or even obtain global coverage (for more details, see https://cran.r-project.org/web/packages/nasapower/nasapower.pdf). The PowerSDI uses this function to acquire data at a particular point or cell. The longitude and latitude values for this cell may vary by 0.5 × 0.5 decimal degrees (Sparks 2023).

In terms of temporal resolution, the NASA-POWER data covers a range from hourly to annual intervals. With the help of the function get_power(), the PowerSDI package specifically downloads NASA-POWER data at the daily time scale from the agroclimatology community. Subsequently, the PowerSDI aggregates the daily data at the quart.month time scale, as chosen by the users. Additional information about the data sources used by the NASA-POWER project, including their latency time, can be found at https://power.larc.nasa.gov/docs/methodology/data/sources/ and https://power.larc.nasa.gov/docs/methodology/data/processing/.

GOODNESS-OF-FIT TESTS, NORMALITY-CHECKING PROCEDURES AND OTHER MODEL PERFORMANCE CHECKING-METHODS

Since the calculation algorithm of the SPI and SPEI relies on fitting a parametric distribution to their input data, the PowerSDI employs the ScientSDI.R function to assess the fit of the two-parameter gamma and GEV/GLO distributions to rainfall and PPE amounts, respectively. To accomplish this, the PowerSDI uses two goodness-of-fit tests previously applied in studies such as Blain and Meschiatti (2014), Stagge et al. (2015) and Blain et al. (2018), aimed at selecting suitable distributions for calculating these indices. These tests are the Kolmogorov-Smirnov/Lilliefors (Lilliefors 1967; Eq. 7) and Anderson-Darling (Anderson and Darling 1954; Eq. 8) tests.

$$D_n = |F_n(x) - F(x)| \tag{7}$$

where: $F_n(x)$: the empirical cumulative probability, estimated as $F_n(x_i) = i/n$ for the ith smallest data; $F(x)$: the theoretical cumulative probability.

$$AD^2 = -n - S \tag{8}$$

where: AD^2: the Anderson-Darling statistics; S is estimated as Eq. 8.1:

$$S = \sum_{i=1}^{n} \frac{(2a-1)}{n} \left[ln \, ln \, F(Y_i) + ln \, ln \, (1 - F(Y_{n+1-i})) \right] \tag{8.1}$$

As highlighted by various studies, including Vlček and Huth (2009), Wilks (2011), Stagge et al. (2015), and Blain et al. (2018), the Kolmogorov-Smirnov/Lilliefors test (referred to as the Lilliefors test) considers only the maximum difference between the empirical and theoretical cumulative probability functions (Eq. 7). On the other hand, in comparison to the Lilliefors test, the Anderson-Darling test (referred to as AD) places greater emphasis on the distribution tails (Stagge et al., 2015). This distinction is why the PowerSDI offers users the possibility to calculate both of these tests.

When calculating these two tests, it is essential to note that, similar to the approach in the majority of (hydro)meteorological studies, the parameters of these distributions are fitted using all available data from the NASA-POWER project. In other words, the goodness-of-fit tests are applied to the same data sample used to fit the distributions' parameters. Consequently,

Eqs. 7 and 8 cannot be used as distribution-free tests (Wilks 2011, Blain 2014, Stagge et al. 2015, Blain et al. 2018). According to Lilliefors' studies (1967), critical values for Eqs. 7 and 8 should be specified through statistical simulations.

The PowerSDI package calculates these critical values using the following procedure, as described in Wilks (2011), Blain (2014), Stagge et al. (2015), and Blain et al. (2018):

- A large number of samples (2,000) from the fitted gamma, GEV, or GLO distributions are generated;
- From each of these synthetic samples, the gamma, GEV, or GLO parameters are calculated, and synthetic Dn (Eq. 7) and AD2 (Eq. 8) values are obtained.

As the null hypothesis of both tests assumes that the data were drawn from the candidate parametric distribution (true by construction), the collection of the 2,000 synthetic D_n and AD^2 values represents the null distributions for these tests. Therefore, an α-level critical value can be approximated as the (1-α) quantile of the null distributions. The function ScientSDI.R allows users to select the 5 or 10% significance levels. Stagge et al. (2015) and Blain et al. (2018) applied this procedure to evaluate the fit of several distributions (including gamma, GEV, and GLO) to the SPI and SPEI input variables.

Considering step 2 of the SPI and SPEI calculation algorithms, the function ScientSDI.R also applies two normality-checking procedures to the NASA-SPI and NASA-SPEI series. The first one is based on Wu et al. (2007). According to this first procedure, a NASA-SPI and/or NASA-SPEI frequency-distribution is regarded as non-normal when the following three criteria are simultaneously:

- Shapiro–Wilk's statistic (W) less than 0.960;
- The corresponding p-values less than 0.10;
- The absolute value of the median greater than 0.05.

Otherwise, the distribution is normal.

The second one is based on the studies of Stagge et al. (2015) and Vincente-Serrano and Beguería (2016) according to which a NASA-SPI and/or NASA-SPEI frequency distribution is regarded as non-normal when the p-value of W is smaller than 0.05.

As previously described, the PowerSDI–through the function Accuracy–is also capable of verifying how well NASA-POWER data actually represent real-world/observed data. Thus, this function calculates two scalar measures of accuracy: the absolute mean error (AME; Eq. 9) and the root-mean-square error (RMSE; Eq. 10). As pointed out by Wilks (2011), the AME and RMSE define accuracy as the average correspondence between individual predictions and observed/reference data.

$$AME = \frac{\sum_{i=1}^{n}|O_i - P_i|}{n} \tag{9}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}|O_i - P_i|^2}{n}} \tag{10}$$

where: *P*: predicted values. *O*: observed values.

In this study, data coming from the NASA POWER was considered as the predicted data. The function Accuracy.R also calculates the original, modified, and refined Willmott's indices of agreement (d_{orig}, d_{mod} and d_{ref}, respectively; Willmott et al. 1985, Willmott et al. 2012). Both d and d_{mod} assume their maximum value (equals to 1) when there is a perfect agreement between estimates and observations. The smallest possible value for both indices is 0. The advantage of the modified version over the original index is that errors and differences are given their appropriate weighting factors (e.g., Willmott et al. 1985, Willmott et al. 2009). Thus, the d_{mod} is often regarded as a more rigorous method than d_{orig} (Legates and McCabe Jr. 1999, Willmott et al. 2012, Bardin-Camparotto et al. 2013, Martins et al. 2023).

Although both d and d_{mod} had led to remarkable improvements in model-checking methods (Willmott et al. (2012) stated that the overall range of these two indices [0:1] may not adequately represent the great variety of forms that predicted/simulated values can differ from observed/reference data. In this context, they proposed the d_{ref}, which assumes the -1 value when there is no agreement between observed and predicted data and the 1 value when there is a perfect agreement. These three indices are calculated as follows (Eqs. 11 to 13).

$$d_{orig} = 1 - \frac{\sum_{i=1}^{n}(P_i - O_i)^2}{\sum_{i=1}^{n}(|P_i - \bar{O}| + |O_i - \bar{O}|)^2} \tag{11}$$

$$d_{mod} = 1 - \frac{\sum_{i=1}^{n}(P_i - O_i)}{\sum_{i=1}^{n}(|P_i - \bar{O}| + |O_i - \bar{O}|)} \tag{12}$$

$$d_{ref} = \begin{cases} 1 - \dfrac{\sum_{i=1}^{n}|P_i - O_i|}{2\sum_{i=1}^{n}|O_i - \bar{O}|}, when \\[2ex] \sum_{i=1}^{n}|P_i - O_i| \le c \sum_{i=1}^{n}|O_i - \bar{O}| \\[2ex] \dfrac{2\sum_{i=1}^{n}|O_i - \bar{O}|}{\sum_{i=1}^{n}|P_i - O_i|} - 1, when \\[2ex] \sum_{i=1}^{n}|P_i - O_i| > 2 \sum_{i=1}^{n}|O_i - \bar{O}| \end{cases} \tag{13}$$

where: P: the predicted values; O: the observed values.

Willmott et al. (1985) suggested using bootstrap techniques (resampling with replacement) for specifying confidence intervals to the model-checking methods described before. Thus, the function Accuracy.R may also provide confidence intervals for AME, RMSE, d_{orig}, d_{mod} and d_{ref} by generating a large number of bootstrapped samples from the original pairs of observed and predicted values. Due to its widespread use, the function Accuracy.R also calculates the Pearson's determination coefficient (R^2).

As emphasized by Willmott et al. (1985), the confidence interval (CI) specifies a range of values within which the AME, RMSE, d_{orig}, d_{mod}, and d_{ref} are expected to vary by chance. Consequently, users can interpret the magnitude of the CI as an indicator of the reliability of the estimated values for the comparison metrics (Willmott et al. 1985).

Finally, the scientific mode of the PowerSDI package also presents another function (Reference.R) that calculates both SPI and SPEI from daily data obtained from a ground weather station or any other reference source. The outputs of this function also include rainfall, PE and PPE values accumulated at the time scales chosen by the user. Therefore, while the ScientSDI.R function is capable of providing variables (rain, PE, and PPE) and indices (SPI and SPEI) from NASA-POWER data, the Reference.R function is capable of providing the same variables and indices from a reference source. Therefore, these two functions may provide inputs for the function Accuracy.R.

As further described in the next section, the input file for the Accuracy function is a two-column matrix with reference and observed data, respectively. This function makes no temporal aggregation of this input data and returns single AME, RMSE, dorig, dmod and dref values for each input file.

THE POWERSDI R-PACKAGE

As described in the previous sections, the PowerSDI was designed to facilitate using NASA-POWER data in drought assessments and monitoring systems. Thus, its scientific mode (functions ScientSDI.R, Reference.R, Accuracy.R, and PlotData.R) enables the users to apply all statistical methods described in this study. In other words, these four functions help the users to verify if the NASA-SPI and NASA-SPEI can be applied in a particular region and at a particular time scale. The ScientSDI.R function also calculates the parameters of the parametric distributions required for calculating the NASA-SPI and NASA-SPEI (Eqs. 2, 5.1, and 5.2). Completed this verification step and parametric fit, the operational mode of the PowerSDI package (function OperatSDI.R) can be used to generate routine operational NASA-SPI and NASA-SPEI estimates in several regions and at distinct time scales.

Furthermore, the OperatSDI.R enables users to download NASA-POWER data for all available period or only for the quasi-week they intend to monitor. This approach naturally reduces the function's running time and facilitates its use in routine drought monitoring efforts. Figure 1 provides an overview of the PowerSDI package.
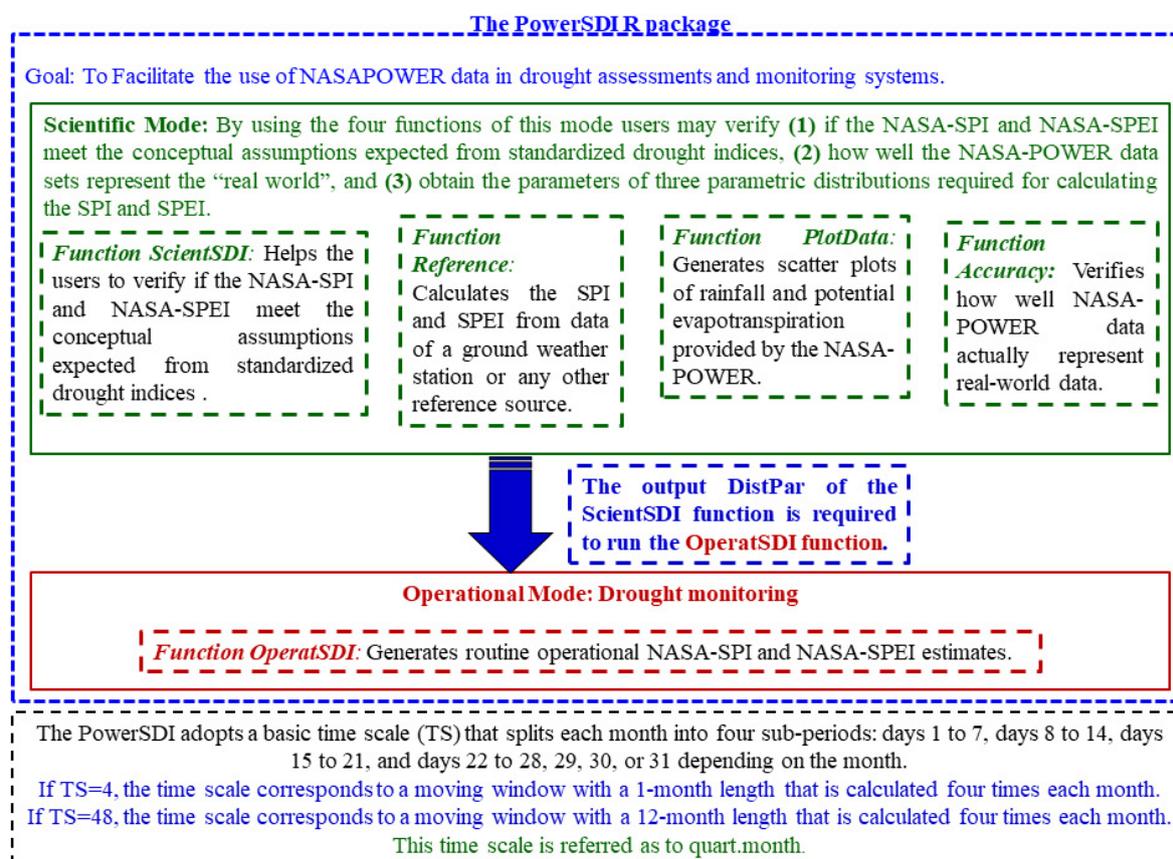

Figure 1. Overview of the PowerSDI package. Scientific and operational modes. The package also has two custom functions (print.PowerSDI. Accuracy, and plot.PowerSDI.Accuracy) for Accuracy function.

Detailing the functions

Supplementary Table 1 (available at https://github.com/gabrielblain/SupplementalFiles_1) presents the basic instructions for the five functions of the PowerSDI package. It is essential to mention that, as described before, the outcomes of the ScientSDI.R function (Suppl. Table 1) can assist users in selecting an appropriate PE estimation method and time scales for calculating the SPI and SPEI in their region of interest.

As highlighted by Wu et al. (2007), for arid climates or those with dry seasons, the SPI at short-time scales may fail to meet its normality assumption. This deviation from normality often occurs due to the relatively high number of 0 in the rainfall series accumulated at short-time scales (Wu et al. 2007, Blain et al. 2018). Since the ScientSDI.R function applies two normality-checking procedures, users can use it to verify if, at a particular short-time scale, the number of non-normally distributed SPI series is unacceptably high. If this is the case, users can consider adopting a larger time scale and use the ScientSDI.R function to verify if the number of non-normally distributed SPI series decreases.

To illustrate this statement better, consider the application of the ScientSDI function calculated at two distinct time scales and using the two distributions to calculate the SPEI. Figure 2 presents an R-script that can be used to perform these calculation in Campinas, state of São Paulo.

```
Fig2.R* ×
        Source on Save        Q                              Run        Source
1  TS1.GEV <- ScientSDI(lon=-47.3,lat=-22.87,start.date="1993-01-01",end.date="2022-12-31",
2                       distr="GEV",TS=1,Good="yes")
3  TS4.GEV <- ScientSDI(lon=-47.3,lat=-22.87,start.date="1993-01-01",end.date="2022-12-31",
4                       distr="GEV",TS=4,Good="yes")
5  TS1.GLO <- ScientSDI(lon=-47.3,lat=-22.87,start.date="1993-01-01",end.date="2022-12-31",
6                       distr="GLO",TS=1,Good="yes")
7  TS4.GLO <- ScientSDI(lon=-47.3,lat=-22.87,start.date="1993-01-01",end.date="2022-12-31",
8                       distr="GLO",TS=4,Good="yes")
```

Figure 2. Using the ScientSDI function in Campinas, state of São Paulo, Brazil.

The results of the normality tests obtained from the R-script depicted in Fig. 2 are presented in Table 2. According to the normality checking procedure proposed by Wu et al. (2007), eight out of the 48-quart.month NASA-SPI series, calculated at the 1-quart.month time scale, could not be considered as normally distributed, resulting in an acceptance rate of 83.3% (100 × (48-8)/48). In contrast, the same analyses applied to the NASA-SPI series calculated at the 4-quart. month time scale, indicated that only four out of the 48 series were considered as non-normal, resulting in an acceptance rate of 91.7% (Table 2).

Considering the results of the other normality checking procedure (proposed by Stagge et al. 2015), the user verifies that the acceptance rate obtained at the 1-quart.month time scale was 89.6% (Table 2). However, when time scale was set to 4 (4-quart.month time scale), only one series failed to meet the normality assumption, resulting in an acceptance rate of 97.9%.

Based on the studies of Blain et al. (2018), it is reasonable to assume that failure to meet the normality assumptions at rates (rejection rates) close to or lower than 10% is an acceptable threshold for calculating an SDI at the selected time scale and with a pre-defined parametric distribution. Examining the findings from Table 2, it can be inferred that the ScientSDI function enabled us to conclude that the 4-quart.month is an appropriate time scale for calculating the NASA-SPI in Campinas, whereas the 1-quart.month is not suitable.

Based on the studies of Stagge et al. (2015), Vincente-Serrano and Beguería (2016), and Blain et al. (2018), users may also infer that the best distribution for calculating the SPEI is the one that leads to the highest number of normally-distributed series. In this context, users may verify that the acceptance rates of the NASA-SPEI calculated with the GEV were slightly but consistently higher than those obtained when the GLO was used (Table 2). Similar inferences can be made regarding the PE estimation method. As presented in Table 2, the PM method did not lead to a higher number of normally-distributed NASA-SPEI series when compared to those obtained when using the HS method. In this case, users may decide to adopt the simplest EP estimation method in Campinas.

Table 2. Results of the normality-checking procedures applied by the ScientSDI.R (PowerSDI package) in the location of Campinas, state of São Paulo, Brazil. The function was applied considering two-time scales (1-quart.month and 4-quart.month) and two distributions generalized extreme value (GEV) and generalized logistic (GLO). The SPEI was also calculated using two potential evapotranspiration estimation methods: Hargreaves and Samani (SPEI.HS) and FAO-56 Penman-Monteith (SPEI.PM).

| Time scale and distribution | Normality checking procedures used by the PowerSDI (acceptance rates*; %) | | | | | |
| | Wu et al. (2007) | | | Stagge et al. (2015) | | |
	SPI	SPEI.HS	SPEI.PM	SPI	SPEI.HS	SPEI.PM
1-quart.month: GLO	83.3	85.4	87.5	89.6	89.6	89.6
4-quart.month: GLO	91.7	97.9	95.8	97.9	100.0	100.0
1-quart.month: GEV	83.3	79.2	79.2	89.6	83.3	85.4
4-quart.month: GEV	91.7	95.8	93.8	97.9	97.9	95.8

*Calculated dividing the number of normally-distributed series by 48; SPI: standardized precipitation; SPEI: standardized precipitation-evapotranspiration.

CASE STUDIES APPLICATIONS

As previously described, the first case study evaluated the applicability of the PowerSDI package in the state of São Paulo (Fig. 3). The state has more than 41 million inhabitants, which represents approximately 22% of Brazilian population. São Paulo has the highest gross domestic product in Brazil, representing around 30% of the total wealth produced in the country (http://www.fearp.usp.br). In this state, the wet season occurs during the austral summer, when the monthly rainfall amounts are usually larger than the PE totals (Blain et al. 2018). December and January are the rainiest months of the year, presenting rainfall frequency distributions that approach the Gaussian shape (Blain et al. 2007). As pointed out by Ben-Gai et al. (1998), bell-shaped rainfall frequency distributions are often observed in equatorial climates. The state also presents a distinct dry season (July and August), when the monthly rainfall amounts are smaller than the potential evapotranspiration totals and the rainfall frequency distributions assume the exponential shape (Blain et al. 2007). This latter distribution shape is usually observed in semi-arid or arid locations (Ben-Gai et al. 1998).

The analyses started within the scientific mode of the PowerSDI package, and we verified if the NASA-SPI and the NASA-SPEI met the conceptual assumptions described before. Within this mode, we also evaluated how well NASA-POWER data represented the "real-world" conditions of the state. All analyses were carried out at the 4-quart.month time scale due to the results found for Campinas (Fig. 2 and Table 2). In addition, monthly time scales are often used in drought monitoring systems in Brazil. Daily rainfall and air temperature data from seven weather stations situated at experimental farms of the Department of Agriculture of the State of São Paulo (1991–2022; red dots in Fig. 3) were taken as the reference data because of their widespread use in scientific studies (Blain et al. 2018).
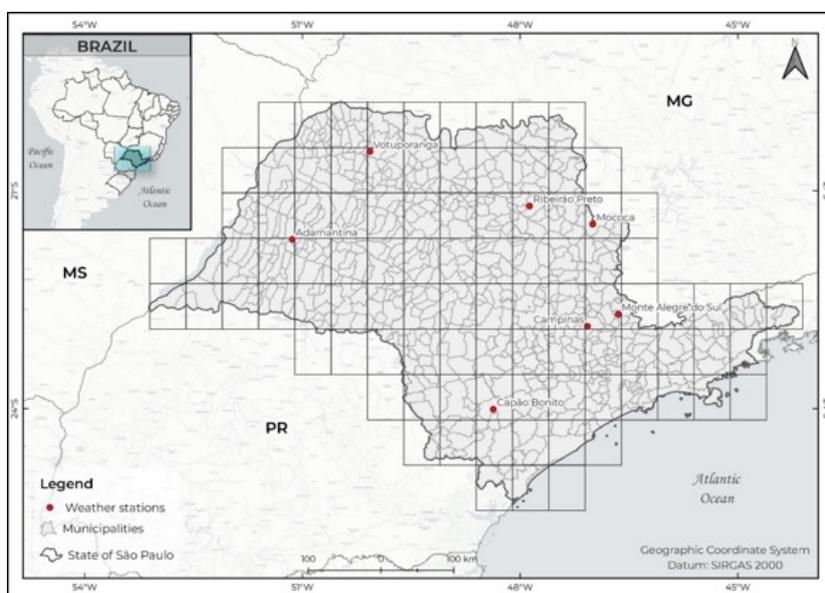

Figure 3. Weather stations (red dots) situated in the state of São Paulo, Brazil. The black solid lines are NASA-POWER pixels (0.500º × 0.625º).

Implementing the NASA-SPI and NASA-SPEI in the state of São Paulo

First, we carried out a visual inspection of the NASA-POWER data in order to detect suspicious values. Thus, we applied the function PlotData.R to each cell corresponding to the seven locations depicted in Fig. 3 in order to generate plots of NASA-Rain and NASA-PE.HS for each of these locations. The HS method was adopted because the reference sources (ground weather stations) cannot provide all variables required by the PM method (a common situation in developing countries).

The plots generated by the PlotData.R function revealed suspicious NASA-Rain data larger than 250 mm for Campinas and Monte Alegre do Sul. As exemplified in Fig. 4 for the weather station of Monte Alegre do Sul, these suspicious values

are also considerably larger than any other rainfall records of the series. Considering that these suspicious data represent less than 1% of each data sample, they were simple replaced by 250 mm. The PE.HS showed no suspicious data for any location. As expected, the longest quart.month (days 22 to 31) are those that present the highest accumulated values for PE.
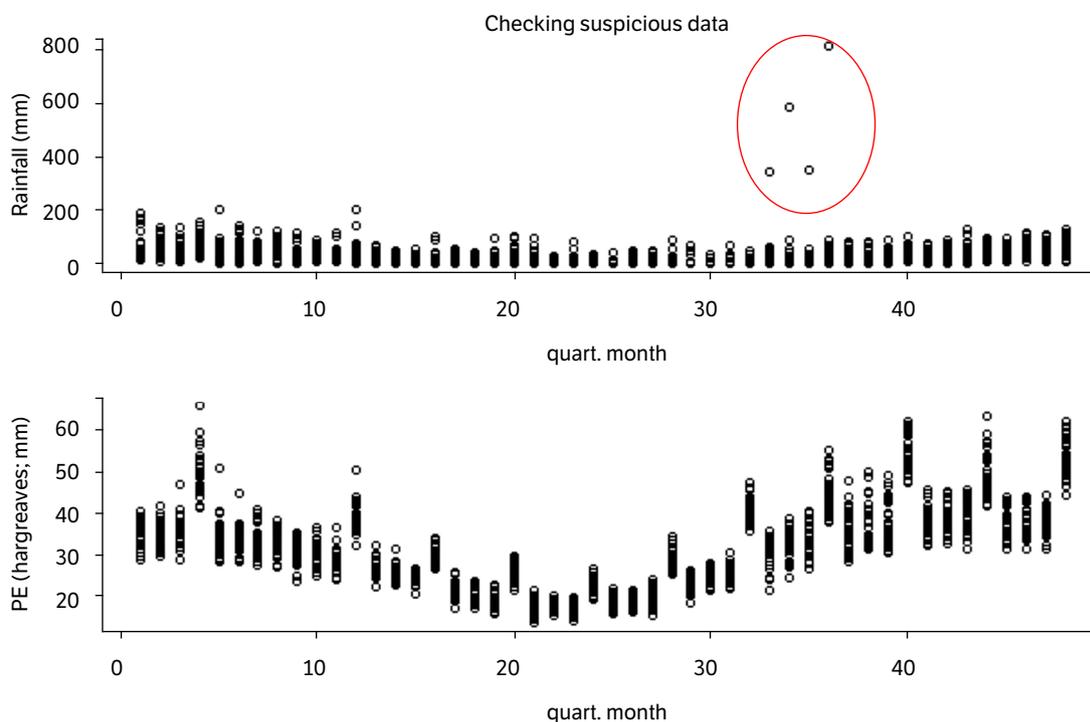

Figure 4. Rainfall and potential evapotranspiration (PE) plots generated by the PlotData.R function. The time scale is 1-month.quart month; Monte Alegre do Sul, state of São Paulo, Brazil (1991–2022).

We then applied the Reference.R function to generate the reference values for rain, PE.HS, PPE-HS, SPI and SPEI. Because we needed to replace the above-mentioned suspicious data with 250 mm, we ran the ScientSDI.R for the locations of Campinas and Monte Alegre do Sul with the argument RainUplim set to 250 mm (RainUplim = 250). For the other locations, this argument was set to its default value (NULL). The acceptance rates (calculated as described in Table 2) of the goodness-of-fit tests and normality-checking procedures generated by the ScientSDI function are shown in Table 3. The Lilliefors and Anderson-Darling tests indicated that the NASA-Rain frequency distributions can have their probabilistic structures described by the two-parameter gamma. These tests also indicated that the GEV and GLO distributions can be used to represent the NASA-PPE-HS frequency distributions.

As can be noted, the null hypothesis of these two goodness-of-fit tests were accepted at rates higher than 81% (Table 3) at all locations. As expected, the acceptance rates obtained using the GEV distribution were close to those obtained from the GLO, with the GEV model showing slightly higher rates (Stagge et al. 2015, Stagge et al. 2016, Vicente-Serrano and Beguería 2016, Blain et al. 2018). Similar results were found for the normality checking procedures (Table 3). The rates at which the NASA-SPI and NASA-SPEI frequency distributions (calculated with the GEV) met the assumption of normality were always higher than 89%. These results are in line with the study of Blain et al. (2018), that recommended the two-parameter gamma and the GEV distribution to calculate, respectively, the SPI and SPEI in the state of São Paulo. The GEV distribution was adopted in this case study.

After verifying that the NASA-SPI and NASA-SPEI met the conceptual assumptions expected from standardized drought indices, we applied the Accuracy.R function to compare each pair of NASA-POWER and reference data. The outcomes of the Accuracy.R function are shown in Suppl. Table 2 (available at https://github.com/gabrielblain/SupplementalFiles_1), and they indicate that the NASA-Rain, NASA-PE, NASA-PPE-HS, NASA-SPI, and NASA-SPEI, accumulated at TS = 4,

can be used to represent the "real-world" conditions of each location considered in this case study. As presented in Suppl. Table 2, the lowest value of the Willmott's indices was d.mod = 0.69 (SPI), and the R^2 remained equal to or larger than 0.70 for the three variables. It is also noteworthy that the confidence intervals of all comparison metrics are narrow, favouring the reliability of the AME, RMSE, d_{orig}, d_{mod}, d_{ref}, and R^2 estimates (Willmott et al. 1985).

Table 3. Results of the goodness-of-fit tests and normality-checking procedures applied by the ScientSDI.R function of the PowerSDI package. The locations Adamantina (Adm), Campinas (Cps), Capão Bonito (Cap), Mococa (Moc), Monte Alegre do Sul (MteAle), Ribeirão Preto (Rib), and Votuporanga (Vot) belong to the state of São Paulo, Brazil. The acceptance rates presented for each local were calculated dividing the number of times the null hypothesis of each test was accepted by 48 (in percentage).

Locals	Goodness-of-fit tests: Acceptance rates (%)					
	Gamma		GLO		GEV	
	Rainfall		PPE		PPE	
	Lilliefors	AD	Lilliefors	AD	Lilliefors	AD
Adm	97.9	89.6	89.6	97.9	91.7	95.8
Cps	97.9	89.6	89.6	97.9	91.7	95.8
Cap	100.0	97.9	89.6	87.5	95.8	100.0
Moc	95.8	95.8	95.8	95.8	93.8	100.0
MteAle	93.8	93.8	85.4	81.3	85.4	89.6
Rib	95.8	89.6	91.7	83.3	91.7	87.5
Vot	89.6	81.3	87.5	93.8	91.7	91.7
Locals	Normality-checking procedures: Acceptance rates (%)					
	Gamma		GLO		GEV	
	SPI		SPEI		SPEI	
	Test I	Test II	Test I	Test II	Test I	Test II
Adm	100.0	100.0	97.9	97.9	100.0	100.0
Cps	100.0	100.0	97.9	97.9	100.0	100.0
Cap	97.9	100.0	95.8	93.8	100.0	100.0
Moc	95.8	97.9	95.8	95.8	97.9	97.9
MteAle	97.9	95.8	93.8	91.7	95.8	89.6
Rib	89.6	93.8	85.4	83.3	93.8	91.7
Vot	87.5	97.9	91.7	91.7	97.9	97.9

GLO: generalized logistic; GEV: generalized extreme value; PPE: potential evapotranspiration amounts; SPI: standardized precipitation; SPEI: standardized precipitation-evapotranspiration.

The results of Table 3 and Suppl. Table 2 indicate that the NASA-SPI and NASA-SPEI can be used to assess and monitor drought events in the state of São Paulo. The plot.PowerSDI.Accuracy function, which generated scatter plots between observed and reference data (Suppl. Fig. 1, https://github.com/gabrielblain/SupplementalFiles_1), is also in line with this latter statement.

Thus, the last step performed in the scientific mode of the PowerSDI package was to run again the ScientSDI.R function in order to obtain the parameters of the gamma and GEV distributions (the output DistPar) for the entire state. Considering the spatial resolution of the NASA-POWER data, the DistPar were generated for 69 pixels (black solid lines of Fig. 2).

In order to facilitate the reproducibility of our results, we made available the coordinates of these 69 points at https://github.com/gabrielblain/SupplementalFiles_1 (grid_sp.csv). The ScientSDI functions was applied as described in Suppl. Table 3 (https://github.com/gabrielblain/SupplementalFiles_1). The script of Suppl. Table 3 took approximately 1 hour and 30 minutes to perform all calculations. It allowed us to calculate the acceptance rates at which both NASA-SPI and NASA-SPEI met their conceptual assumptions at each 69 points in the state. These acceptance rates are depicted in Fig. 5.

The maps presented in this study were plotted using the following R-packages 'ggplot2' (Wickham et al. 2023), 'sp' (Pebesma et al. 2023), 'RColorBrewer' (Neuwirth 2022), and 'sf' (Pebesma et al. 2023).
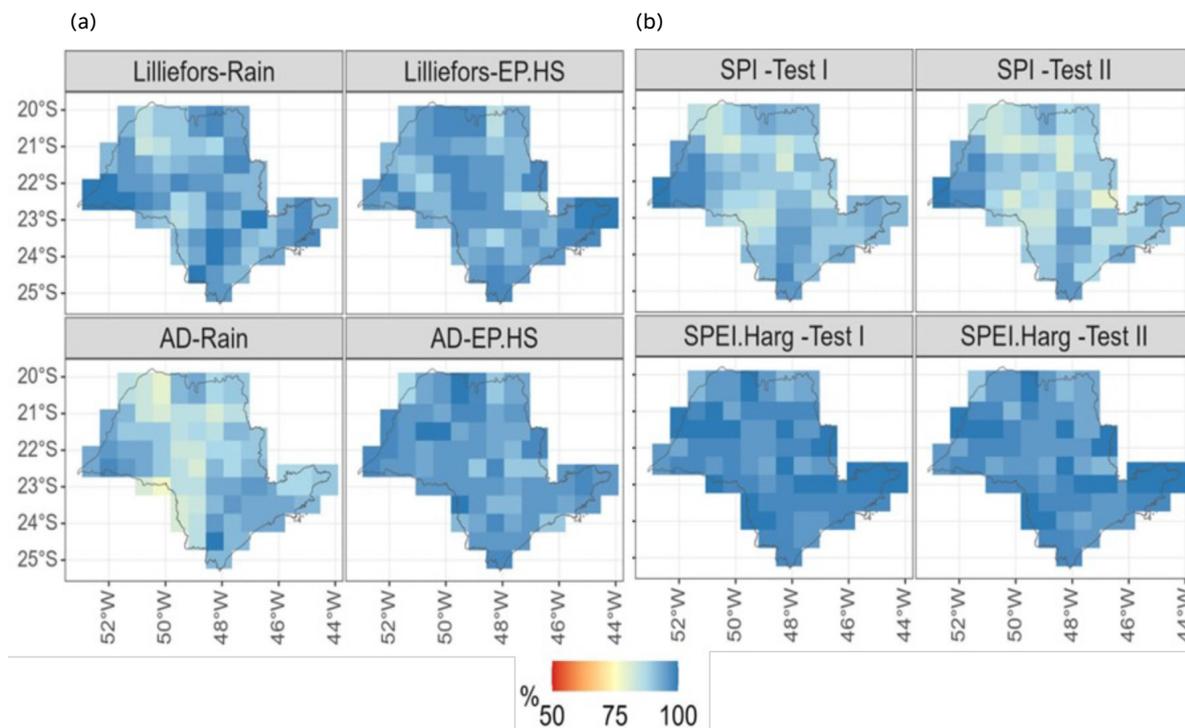

Figure 5. Implementing the NASA-SPI and NASA-SPEI in the state of São Paulo, Brazil. Acceptance rates of the (a) Lilliefors and Anderson-Darling (AD) goodness-of-fit tests, and (b) the two normality-checking procedure applied by the ScientSDI.R function of the PowerSDI package. The acceptance rates presented for each local were calculated dividing the number of times the null hypothesis of each test was accepted by 48 (in percentage). EP.HS is the potential evapotranspiration calculated through the Hargreaves and Samani (Harg) method.

The analysis of Fig. 5, along with the results and Table 3 and Suppl. Table 2, may be regarded as a solid scientific basis supporting the implementation and use of the NASA-SPI and NASA-SPEI in the state of São Paulo, Brazil. Thus, these two indices can now be routinely calculated in operational mode to monitor drought conditions in the State.

Drought monitoring in the state of São Paulo using NASA-POWER data (PowerSDI package)

Since the parameters of the gamma and GEV distributions were previously estimated by the ScientSDI.R function, users are able to use the OperatSDI.R function to download NASA-POWER data for any monitoring period. We chose to demonstrate the function's capabilities by analysing January 2014, which was one of the driest years on record in the state. The OperatSDI functions was applied as follow (Suppl. Table 4, https://github.com/gabrielblain/ SupplementalFiles_1).

The script presented in Table S4 took less than 3 minutes to calculate the NASA-SPI and NASA-SPEI for the entire state. This relative short running time may be regarded as a desirable feature of the OperatSDI function, which was designed to be used in a routine/operational mode. Regarding its outputs (Fig. 6), Nobre et al. (2016) attributed the meteorological causes of this extreme drought event to changes in regional circulation. A mid-troposphere blocking high occurred over 45 days throughout South-Eastern Brazil. Both NASA-SPI and NASA-SPEI captured this extreme dry condition, showing negative values across almost the entire state (Fig. 6). This result, along with all others found in this case study, indicates that the NASA-SPI and NASA-SPEI can be used in drought assessments and monitoring in the state of São Paulo.
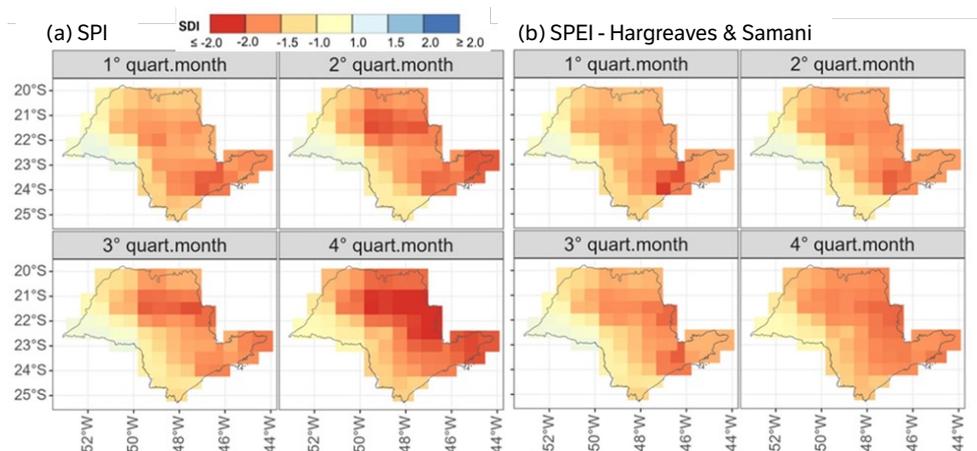
Figure 6. The standardized precipitation (SPI) and (b) standardized precipitation-evapotranspiration (SPEI) calculated using NASA-POWER data. The drought indices were obtained through the OperatSDI.R function of the PowerSDI package.

Case study 2: drought monitoring in entire Brazil

As previously described, the second case study evaluated the applicability of the PowerSDI package in entire Brazil. In order to facilitate the reproducibility of our results, we made available the 2,841 coordinates (grid_Br.csv) at which the package's functions ScientSDI and OperatSDI were applied, considering the 4-quart.month time (1993–2022). While the ScientSDI was applied to estimate the parameters of the gamma and GEV distributions, the OperatSDI was used to exemplify the routine use of the package during July 2023 (Fig. 7). The OperatSDI function took less than 50 minutes to calculate PE, PPE, NASA-SPI, and NASA-SPEI for the entire country, which covers an area of approximately 8.5 million square kilometers (almost as large as continental Europe). The NASA-SPEI estimates (Fig. 7) indicated some areas in the Amazon rainforest experiencing moderate, severe, and extreme dry conditions. Thus, these estimates described the onset of the record-breaking drought that the Amazon rainforest has faced since November 2023.
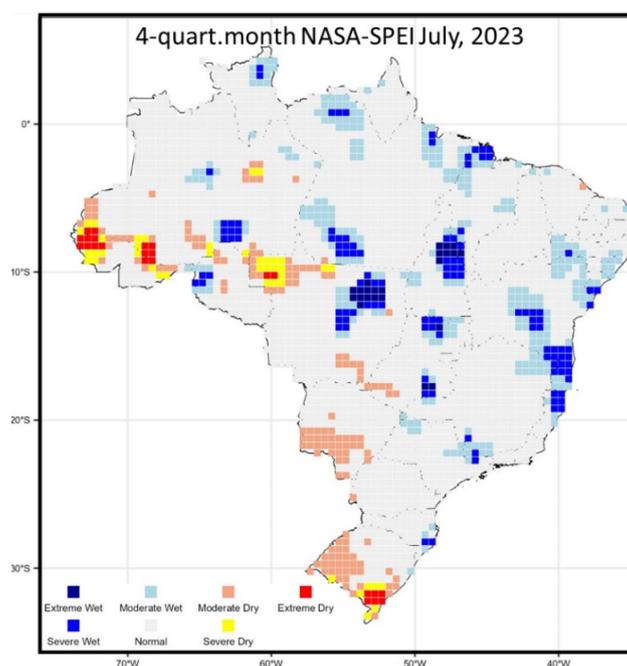

Figure 7. The standardized precipitation-evapotranspiration calculated using NASA-POWER data. The drought index was obtained through the OperatSDI.R function of the PowerSDI package.

FINAL REMARKS

The SPI and SPEI have been widely used to assess and monitor drought events throughout the globe. However, the low density of weather station networks and data quality issues limit their use in several regions of the world. The NASA-POWER project has emerged as an interesting data source capable of overcoming this difficulty. Unlike other gridded-databases, this data source can meet the 30-year continuous records required for calculating these two drought indices.

In this context, we developed the PowerSDI package that calculates these two drought indices using data provided by the NASA-POWER project. Unlike other existing software, the PowerSDI package has two modes: the scientific and the operational. In the scientific mode, the users may assess the quality of the indices estimates through their agreement with a reference/observed series and through the evaluation of how well these estimates meet the conceptual assumptions required for calculating both SPI and SPEI. In the operational mode, users can calculate both indices routinely using the OperatSDI.R function. This function enables users to download NASA-POWER data for all available series or only for the period they intend to monitor, reducing the function's running time. Unlike other existing packages, the PowerSDI adopts a quasi-weekly time scale, allowing for index calculations four times a month. The package is freely available at two repositories, Github (https://github.com/gabrielblain/PowerSDI), and CRAN (https://CRAN.R-project. org/package=PowerSDI).

Regarding future studies and improvements, we highlight that there is still no consensus in the literature on the best distribution for calculating the SPI and SPEI (Guttman 1999, Stagge et al. 2015, Stagge et al. 2016, Vicente-Serrano and Beguería 2016, Blain et al. 2018, Pieper et al. 2020). This version of the PowerSDI package attempted to address this issue by providing distinct goodness-of-fit tests and normality-checking procedures and by allowing the users to select between the GEV and GLO distributions to calculate the SPEI. Future updates of this package may further address this question by providing alternative distributions to calculate the SPI and SPEI, and other normality-checking procedures that directly compare the indices values estimated from the candidate distributions with their corresponding theoretical values derived from the standard normal distribution (Pieper et al. 2020).

CONFLICT OF INTEREST

Nothing to declare.

AUTHORS' CONTRIBUTION

Conceptualization: Sobierajski, G. R. and Blain, G. C.; **Investigation:** 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.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Al-Kilani, M. R., Rahbeh, M., Al-Bakri, J., Tadesse, T. and Knutsom, C. (2021). Evaluation of remotely sensed precipitation estimates from the NASA POWER project for drought detection over Jordan. Earth Systems and Environment, 5, 561-573. https://doi.org/10.1007/s41748-021-00245-2

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.

Anderson, T. W. and Darling, D. A. (1954). A Test of goodness of fit. Journal of the American Statistical Association, 49, 765-769. https://doi.org/10.2307/2281537

Bai, J., Chen, X., Dobermann, A., Yang, H., Cassman, K. G. and Zhang, F. (2010). Evaluation of NASA satellite- and model-derived weather data for simulation of maize yield potential in China. Agronomy Journal, 102, 9-16. https://doi.org/10.2134/agronj2009.0085

Bardin-Camparotto, L., Blain, G. C., Giarolla, A., Adami, M. and Camargo, M. B. P. (2013). Validation of temperature and rainfall data obtained by remote sensing for the State of Sao Paulo, Brazil. Revista Brasileira de Engenharia Agrícola e Ambiental, 17, 665-671. https://doi.org/10.1590/S1415-43662013000600013

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

Ben-Gai, T., Bitan, A., Manes, A., Alpert, P. and Rubin, S. (1998). Spatial and temporal changes in rainfall frequency distribution patterns in Israel. Theoretical and Applied Climatology, 61, 177-190. https://doi.org/10.1007/s007040050062

Blain, G. C. (2014). Revisiting the critical values of the Lilliefors test: towards the correct agrometeorological use of the Kolmogorov-Smirnov framework. Bragantia, 73, 192-202. https://doi.org/10.1590/brag.2014.015

Blain, G. C. and Meschiatti, M. C. (2014). Using multi-parameters distributions to assess the probability of occurrence of extreme rainfall data. Revista Brasileira de Engenharia Agrícola e Ambiental, 18, 307-313. https://doi.org/10.1590/S1415-43662014000300010

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., Piedade, S. M. S., Camargo, M. B. P. and Giarolla, A. (2007). Monthly rainfall temporal distribution observed in the Agronomic Institute Weather Station at Campinas, São Paulo State, Brazil. Bragantia, 66, 347-355. https://doi.org/10.1590/S0006-87052007000200019

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

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

Dai, A. (2011). Drought under global warming: a review. WIREs Climate Change, 2, 45-65. https://doi.org/10.1002/wcc.81

Droogers, P. and Allen, R. G. (2002). Estimating Reference Evapotranspiration Under Inaccurate Data Conditions. Irrigation and Drainage Systems, 16, 33-45. https://doi.org/10.1023/A:1015508322413

Duarte, Y. C. N. and Sentelhas, P. C. (2020). NASA/POWER and Daily Gridded weather datasets- how good they are for estimating maize yields in Brazil? International Journal of Biometeorology, 64, 319-329. https://doi.org/10.1007/s00484-019-01810-1

Guttman, N. B. (1999). Accepting the Standardized Precipitation Index: a calculation algorithm. Journal of the American Water Resources Association, 35, 311-322. https://doi.org/10.1111/j.1752-1688.1999.tb03592.x

Hao, Z., Yuan, X., Xia, Y., Hao, F. and Singh, V. P. (2017). An overview of drought monitoring and prediction systems at regional and global scales. Bulletin of the American Meteorological Society, 98, 1879-1896. https://doi.org/10.1175/BAMS-D-15-00149.1

Hargreaves, G. H. and Samani, Z. A. (1985). Reference crop evapotranspiration from temperature. Applied Engineering Agriculture, 1, 96-99. https://doi.org/10.13031/2013.26773

Hayes, M. J., Svoboda, M. D., Wall, N. and Widhalm, M. (2011). The Lincoln declaration on drought indices – universal meteorological drought index recommended. Bulletin of the American Meteorological Society, 92, 485-488. https://doi.org/10.1175/2010BAMS3103.1

Hayes, M. J., Svoboda, M. D., Wilhite, D. A. and Vanyarkho, O. V. (1999). Monitoring the 1996 drought using the standardized precipitation index. Bulletin of the American Meteorological Society, 80, 429-438. https://doi.org/10.1175/1520-0477(1999)080<0429:MTDUTS>2.0.CO;2

Hosking, J. R. M. (1990). L-Moments: Analysis and Estimation of distributions Using Linear Combinations of Order Statistics. Journal of the Royal Statistical Society, 52, 105-124.

Hosking, J. R. M. (2022). L-Moments. R package, version 2.9. Available at: https://CRAN.R-project.org/package=lmom. Accessed on: Dec. 4, 2023.

Legates, D. R. and McCabe Jr., G. J. (1999). Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water Resources Research, 35, 233-241. https://doi.org/10.1029/1998WR900018

Li, J. Z., Wang, Y. X., Li, S. F. and Hu, R. (2015). A nonstationary standardized precipitation index incorporating climate indices as covariates. Journal of Geophysical Research: Atmospheres, 120, 12082-12095. https://doi.org/10.1002/2015JD023920

Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. Journal of the American Statistical Association, 62, 399-402. https://doi.org/10.1080/01621459.1967.10482916

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., Martins, W. A., Rodrigues, I. C., Xavier, A. C. F., Moraes, J. F. and Blain, G. C. (2023). Gap-filling of daily precipitation and streamflow time series: a method comparison at random and sequential gaps. Hydrological Sciences Journal, 68, 148-160. https://doi.org/10.1080/02626667.2022.2145200

Meschiatti, M. C. and Blain, G. C. (2016). Increasing the regional availability of the Standardized Precipitation Index: an operational approach. Bragantia, 75, 507-521. https://doi.org/10.1590/1678-4499.478

Mishra, A. K. and Singh, P. V. (2010). A review of drought concepts. Journal of Hydrology, 391, 202-216. https://doi.org/10.1016/j.jhydrol.2010.07.012

Monteiro, L. A., Sentelhas, P. C. and Pedra, G. U. (2018). Assessment of NASA/POWER satellite-based weather system for Brazilian conditions and its impact on sugarcane yield simulation. International Journal of Climatology, 38, 1571-1581. https://doi.org/10.1002/joc.5282

Neuwirth, E. (2022). Package 'RColorBrewer'. Version 1.1-3. Available at: https://CRAN.R-project.org/package=RColorBrewer. Accessed on: Dec. 4, 2023.

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

Pebesma, E, Bivand, R., Racine, E., Sumner, M., Cook, I., Keitt, T., Lovelace, R., Wickham, H., Ooms, J., Müller, K., Pedersen, T. L., Baston, D. and Dunnington, D. (2023). Package 'sf'. Version 1.0-14. Available at: https://cran.r-project.org/web/packages/sf/sf.pdf. Accessed on: Dec. 4, 2023.

Pereira, V. R., Blain, G. C., Avila, A. M. H., Pires, R. C. M. and Pinto, H. S. (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

Rashid, M. M. and Beecham, S. (2019). Development of a non-stationary standardized precipitation index and its application to a south Australian climate. Science of the Total Environment, 657, 882-892. https://doi.org/10.1016/j.scitotenv.2018.12.052

Rodrigues, G. C. and Braga, R. P. (2021). Estimation of Daily Reference Evapotranspiration from NASA POWER Reanalysis Products in a Hot Summer Mediterranean Climate. Agronomy, 11, 2077. https://doi.org/10.3390/agronomy11102077

Russo, S., Dosio, A., Sterl, A., Barbosa, P. and Vogt, J. (2013). Projection of occurrence of extreme dry-wet years and seasons in Europe with stationary and nonstationary standardized precipitation indices. Journal of Geophysical Research: Atmospheres, 118, 7628-7639. https://doi.org/10.1002/jgrd.50571

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

Solakova, T., De Michele, C. and Vezzoli, R. (2014). Comparison between parametric and nonparametric approaches for the calculation of two drought indices: SPI and SSI. Journal of Hydrologic Engineering, 19, 4014010. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000942

Sparks, A. (2018). Nasapower: A NASA POWER Global Meteorology, Surface Solar Energy and Climatology Data Client for R. Journal of Open Source Software, 3, 1035. https://doi.org/10.21105/joss.01035

Sparks, A. (2023). Nasapower: nasa-power data from R. https://doi.org/10.5281/zenodo.1040727

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

Thom, H. C. S. (1951). A frequency distribution for precipitation. Bulletin of The American Meteorological Society, 32, 397.

Thornthwaite, C. W. (1948). An approach toward a rational classification of climate. Geographical Review, 38, 55-94. https://doi.org/10.2307/210739

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

Vicente-Serrano, S. M., Domínguez-Castro, F., Reig, F., Beguería, S., Tomas-Burguera, M., Latorre, M., Peña-Ângulo, D., Noguera, I., Rabanaque, I., Luna, Y., Morata, A. and El Kenawy, A. (2022). A near real-time drought monitoring system for Spain using automatic weather station network. Atmospheric Research, 271, 106095. https://doi.org/10.1016/j.atmosres.2022.106095

Vlček, O. and Huth, R. (2009). Is daily precipitation Gamma-distributed? Adverse effects of an incorrect use of the Kolmogorov-Smirnov test. Atmospheric Research, 93, 759-766. https://doi.org/10.1016/j.atmosres.2009.03.005

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H. and Dunnington, D. (2023). Package 'ggplot2'. Version 3.4.4. Available at: https://CRAN.R-project.org/package=ggplot2. Accessed on: Dec. 4, 2023.

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

Willmott, C. J., Ackleson, S. G., Davis, R. E., Feddema, J. J., Klink, K. M., Legates, D. R., O'Donnell, J. and Rowe, C. M. (1985). Statistics for the evaluation of model performance. Journal of Geophysical Research, 90, 8995-9005. https://doi.org/10.1029/JC090iC05p08995

Willmott, C. J., Matsuura, K. and Robeson, S. M. (2009). Ambiguities inherent in sums-of-squares-based error statistics. Atmospheric Environment, 43, 749-752. https://doi.org/10.1016/j.atmosenv.2008.10.005

Willmott, C. J., Robeson, S. M. and Matsuura, K. A. (2012). A refined index of model performance. International Journal of Climatology, 32, 2088-2094. https://doi.org/10.1002/joc.2419

Wu, H., Hayes, M. J., Wilhite, D. A. and Svoboda, M. D. (2005). The effect of the length of record on the standardized precipitation index calculation. International Journal of Climatology, 25, 505-520. https://doi.org/10.1002/joc.1142

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