Earth Science Informatics (2025) 18:159
https://doi.org/10.1007/s12145-024-01687-8

SOFTWARE

DroughtStats: A comprehensive software for drought monitoring and analysis

Tolga Barış Terzi[1] · **Bihrat Önöz**[2]

Received: 27 September 2024 / Accepted: 27 December 2024 / Published online: 9 January 2025

Abstract
The significance of drought monitoring and prediction systems has grown substantially due to the escalating impacts of climate change. However, existing tools for drought analysis face several limitations, including restricted functionality to single-variable indices, reliance on predefined probability distributions, lack of flexibility in choosing distributions, and the need for advanced programming expertise. These constraints hinder comprehensive and accurate drought assessments. This study introduces DroughtStats, a novel, user-friendly software designed to overcome these challenges and enhance drought analysis capabilities. DroughtStats integrates advanced statistical tools to analyze hydrometeorological data, compute both single-variable and multivariable drought indices using empirical and parametric methods, and evaluate drought characteristics with improved accuracy. Notably, it supports a broader range of probability distributions, performs copula-based analyses, and estimates potential evapotranspiration using multiple methods, including Penman–Monteith. Additionally, DroughtStats can analyze the relationship between different datasets using techniques like copula-based Kendall's tau. By addressing the limitations of existing tools, DroughtStats provides a more flexible and comprehensive approach to drought monitoring. Its versatility and global applicability are demonstrated through a case study in Turkey's Çoruh River Basin (CRB), where drought indices based on precipitation and streamflow are calculated to characterize drought conditions. The results show that DroughtStats can successfully identify and characterize drought events at various time scales, providing valuable insights into drought severity, frequency, and recovery, and offering a reliable tool for ongoing drought monitoring and management.

Keywords Drought analysis · Drought monitoring · Drought index · Copula · Multivariate standardized drought index

Introduction

The observed global warming to the date has been reassessed using improved observational data. According to the Intergovernmental Panel on Climate Change (IPCC) Assessment Report 5 (AR5), the rate of global surface temperature increase was smaller than the rate calculated since 1951 (IPCC 2014). However, Assessment Report 6 (AR6) indicates that the slower rate of global surface temperature rise reported in AR5 was a temporary phenomenon. The reassessment of global warming to the date, utilizing updated observational data, underscores a more pronounced trend in global warming. Specifically, for the decade spanning 2011–2020, the global surface temperature increase relative to 1850–1900 is now assessed at 1.09 °C (IPCC 2023). This intensification in global warming contributes to alterations in the water and energy balance, which in turn exacerbates the frequency and severity of extreme weather events, such as droughts (Alessi et al. 2022; Tan et al. 2023).

Among the extreme weather events intensified by global warming, drought stands out as a unique natural hazard characterized by its slow onset, gradual development, and extensive spatial impact. Unlike sudden disasters such as floods or hurricanes, the creeping nature of drought makes it harder to detect, monitor, and address effectively, often leaving its

Communicated by Hassan Babaie

✉ Tolga Barış Terzi
tolgabaristerzi@ktu.edu.tr

Bihrat Önöz
bihrat.onoz@isikun.edu.tr

[1] Department of Civil Engineering, Faculty of Engineering, Karadeniz Technical University, Trabzon, Turkey

[2] Department of Civil Engineering, Faculty of Engineering and Natural Sciences, Işık University, Istanbul, Turkey

impacts underestimated until they become severe. Similar to other natural hazards, drought affects economic, environmental, and social aspects of life, disrupting agricultural production, depleting water resources, and threatening ecosystems (Shaikh et al. 2021; Leng et al. 2022; Zeng et al. 2023). These impacts are compounded by the growing demand for water driven by population growth, urbanization, and agricultural expansion. As climate change intensifies and global water demand increases, the detrimental effects of droughts are exacerbated, complicating drought mitigation efforts (AghaKouchak et al. 2014; Walker and Van Loon 2023). According to European Commission: Joint Research Centre et al. (2020), the economic damage of droughts in Europe is projected to increase by 170% with a 1.5 °C rise in warming, underscoring the urgent need for improved drought monitoring and prediction systems. Accurate and comprehensive monitoring not only supports timely response and resource allocation but also aids in understanding the diverse manifestations of drought across different sectors.

With the addition of ecological drought as a distinct category in the AR6, droughts are now classified into five types based on their impacts: meteorological, hydrological, agricultural, ecological, and socioeconomic (Zhang et al. 2022). Typically serving as the first indication of a drought event, meteorological drought refers to a prolonged period of below-average precipitation (Li et al. 2022). Prolonged precipitation deficits may result in hydrological drought, which is characterized by reduced water availability in rivers, reservoirs, and aquifers. These conditions can lead to agricultural drought, defined by insufficient soil moisture for crop and livestock production (Liu et al. 2021). Furthermore, water scarcity may disrupt ecosystems, causing ecological drought, which can lead to habitat loss and reduced biodiversity. Ultimately, the combined impacts of these drought types can manifest as socioeconomic drought, where water shortages affect livelihoods, economic stability, and social well-being. While different types of droughts are interconnected and interact in complex ways, each drought type also exhibits distinct characteristics that necessitate tailored approaches to their assessment (Vorobevskii et al. 2022). To effectively capture these characteristics and quantify drought severity, various drought indices have been developed. These indices provide a systematic framework for monitoring droughts and understanding their progression across different temporal and spatial scales.

Drought indices, which quantify drought conditions in numerical terms, have been developed to assess the severity and duration of different types of droughts over the past few decades. Among the most commonly used are the Standardized Precipitation Index (SPI) (McKee et al. 1993), the Standardized Precipitation Evapotranspiration Index (SPEI) (Vicente-Serrano et al. 2010), the Palmer Drought Severity Index (PDSI) (Palmer 1965), the Standardized Runoff Index (SRI) (Shukla and Wood 2008), and the Standardized Soil Moisture Index (SSI) (Xu et al. 2018). These indices are favored in the literature for their simplicity, broad applicability across regions, and their capacity to accurately represent drought conditions (Çavuş et al. 2023; Gümüş, 2023; Lorenzo et al. 2024; Sun et al. 2023). In addition to univariate drought indices, multivariate drought indices that combine multiple drought types, typically by determining the joint probability between hydrometeorological variables, have also been developed and utilized in past studies (Hao and AghaKouchak 2013; Farahmand and AghaKouchak 2015; Li et al. 2021; Zhang et al. 2019).

While drought indices provide essential quantitative assessments, their practical application often requires computational tools for efficient calculation and analysis. These tools vary in complexity and regional applicability, either manually applying the relevant equations or automating the process. Widely-used tools for drought analysis include the 'SPI Program' (National Drought Mitigation Center 2024), the 'SPEI' package in R (Beguería and Vicente-Serrano 2013), the 'drought' package in R (Hao et al. 2017), 'SDAT' (Farahmand and AghaKouchak 2015), and 'DrinC' (Tigkas et al. 2014, 2022). Additionally, to meet more specific regional needs, specialized software tools have been developed. For example, the UniSQ Drought Monitor (Guillory et al. 2023) assesses drought conditions in Australia by utilizing the Australian Combined Drought Indicator (CDI), which integrates the SPI, Normalized Difference Vegetation Index (NDVI), evapotranspiration, and soil moisture data. Similarly, Pywr-DRB (Hamilton et al., 2024) analyzes drought scenarios in the Delaware River Basin (DRB), focusing on reservoir operations, transbasin water diversions, and maintaining minimum flow targets. While these specialized tools are often more sophisticated, they are region-specific and lack the versatility needed for global application.

The past decade has witnessed significant advancements in drought monitoring and analysis, introducing increasingly sophisticated methods that provide deeper insights into drought phenomena (AghaKouchak et al. 2015; Hao and Singh 2015; Mukherjee et al. 2018; Van Loon 2015;). While existing tools for calculating drought indices have contributed to these advancements, they still face several limitations. Tools like the 'SPEI' and 'drought' R packages require a certain degree of knowledge of the programming language they were developed in, thus restricting their accessibility for non-expert users. Although user-friendly tools such as the 'SPI Program' and 'DrinC' are still widely used in the literature, they do not fully address the contemporary needs of drought analysis. A major limitation of these tools is their lack of flexibility in selecting probability distribution functions (PDF) for calculating drought indices. This lack of flexibility is problematic, as drought indices are highly

sensitive to the choice of PDFs (Tijdeman et al. 2020), a factor often overlooked in the literature. One possible reason this issue persists could be the lack of tools offering a broader range of PDFs for drought index calculation. Moreover, while the 'drought' package allows for the calculation of a nonparametric multivariate standardized drought index, none of these tools support the computation of multivariate drought indices using copula functions, an essential tool for modelling the joint behavior of hydrometeorological variables and offering a more integrated view of drought dynamics. Despite their widespread use in comprehensive drought evaluation, copula functions remain unsupported by most existing tools (Tian et al. 2023; Wang et al. 2024; Won et al. 2020; Zavareh et al. 2023).

To address these challenges, this study introduces DroughtStats, a novel, user-friendly software tool designed to overcome these limitations in drought analysis. DroughtStats integrates advanced statistical methods for both single-variable and multivariable drought indices, including copula-based analyses and a diverse range of probability distributions. Unlike existing tools, DroughtStats balances accessibility with sophisticated analytical features, allowing users of all expertise levels to perform detailed drought analyses at both global and regional scales. The software also includes essential features such as homogeneity and trend tests, potential evapotranspiration (PET) estimation using multiple methods, extraction of drought characteristics, and the ability to analyze inter-dataset relationships through techniques like copula-based Kendall's tau.

This paper outlines the objectives, functionalities, and real-world applications of DroughtStats. The Çoruh River Basin (CRB) in Turkey is selected as a case study to illustrate its practical utility, where the software is used to calculate both empirical and parametric drought indices and characterize drought conditions effectively.

Overview of the software

DroughtStats is a user-friendly software tool developed to analyze drought characteristics through both univariate and multivariate drought indices. It provides comprehensive statistical analyses, including homogeneity testing, statistical summaries of the dataset, and trend analysis. In addition to traditional drought indices, DroughtStats integrates advanced techniques for multivariate drought analysis, such as copula-based methods, offering a more nuanced understanding of drought dynamics. The software is designed to be accessible to users with varying levels of expertise, making it suitable for both researchers and practitioners. The general structure and workflow of DroughtStats are outlined in the flowchart presented in Fig. 1.

Statistical analysis of hydrometeorological data

DroughtStats provides comprehensive statistical analysis of the given hydrometeorological dataset, offering key descriptive statistics such as variance, coefficient of variation, skewness, kurtosis, and autocorrelation. Additionally, it includes tools for performing homogeneity and trend tests to ensure the reliability and examine temporal changes in the data.

Homogeneity tests

DroughtStats applies various homogeneity tests to assess the consistency and stability of the dataset over time. These include Buishand's test, Pettitt's test, the Standard Normal Homogeneity Test (SNHT), and the von Neumann ratio test, which help identify any significant shifts or changes in the data series.

The von Nuemann ratio test (Von Neumann 1941) is frequently used in literature to detect non-homogeneity in time series. However, the test does not provide any information about the location of break point for homogeneity. The von Neumann ratio is shown in the Eq. (1).

$$N = \frac{\sum_{i=1}^{n-1} \left(Y_i - Y_{i+1}\right)^2}{\sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2} \tag{1}$$

where $\overline{Y}$ represents the average of the Y values and N represents the mean squared proportion of the variation of the variance in succession. A time series is considered homogeneous when the N value calculated from Eq. (1) is equal to 2. However, if there is a break in the data, N tends to be lower than 2 (Buishand 1982). Critical values of N are given in Winjgaard et al. (2003) Table IV.

Buishand (1982) states, homogeneity tests can be based on adjusted partial sums or cumulative deviations from the mean. The adjusted partial sums are defined in Eq. (2).

$$S^* = 0; S_k^* = \sum_{i=1}^{k} \left(Y_i - \overline{Y}\right), k = 1, 2, \ldots, N \tag{2}$$

When a time series is homogeneous, Y_i values will not systematically deviate from their own mean, and thus, S_k^* values are expected to fluctuate around zero. However, if there is a break in the data, then S_k^* is expected to reach either a maximum or a minimum near the point of break in the data. The rescaled adjusted range, R, which is the difference between the maximum and minimum of the S_k^* values scaled by the standard deviation, is used to test the significance of this shift in S_k^*.

$$R = \frac{\left(max_{0 \le k \le n} S_k^* - min_{0 \le k \le n} S_k^*\right)}{s} \tag{3}$$

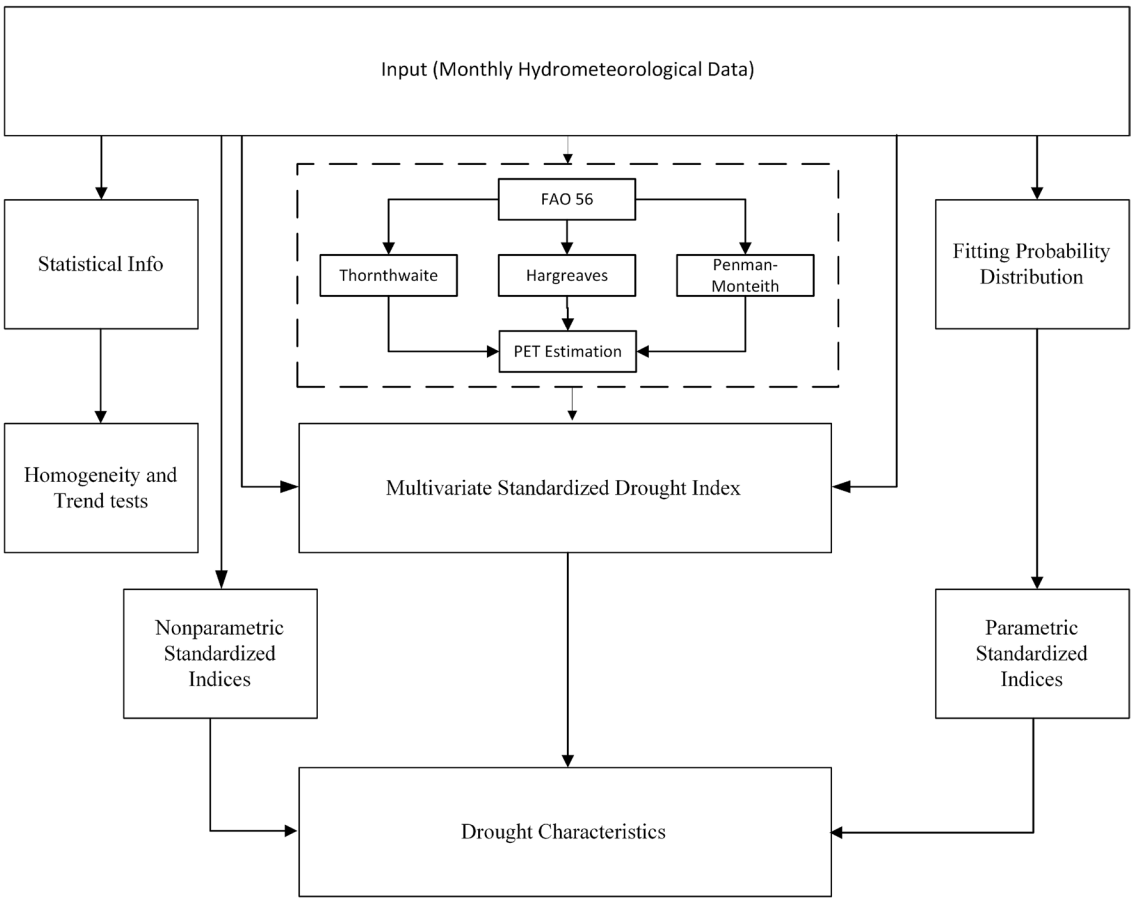

Fig. 1 The general structure of the software

The Pettitt test is a nonparametric method that detects the points of change in a time series. The ranks $(r_1, r_2, \ldots, r_n)$ of the data $(Y_1, Y_2, \ldots, Y_n)$ are used to calculate the statistics (Pettitt 1979):

$$X_k = 2\sum_{i=1}^{k} r_i - k(n+1), k = 1,2,\ldots,n \tag{4}$$

If there is a break in year K, the statistical maximal or minimal near the year K could be defined as:

$$X_K = max_{1\le k\le n}|X_k| \tag{5}$$

The critical values for the Pettitt test are given in Pettitt (1979). A time series is homogeneous when the X_K values do not exceed the critical values.

The standard normal homogeneity test, developed by Alexandersson (1986), is frequently used in hydrological studies. This method divides the examined dataset into two subsets at a specified point and compares the means of these subsets. If the specified point is k, then the $T(k)$ could be defined as:

$$T(k) = k\bar{z}_1^2 + (n-k)\bar{z}_2^2, k = 1,2,\ldots,n \tag{6}$$

where,

$$\bar{z}_1 = \frac{1}{k}\frac{\sum_{i=1}^{k}\left(Y_i - \overline{Y}\right)}{s} \tag{7}$$

$$\bar{z}_2 = \frac{1}{n-k}\frac{\sum_{i=k+1}^{n}\left(Y_i - \overline{Y}\right)}{s} \tag{8}$$

If there is a break point, then $T(k)$ would reach a maximum value near the break point. The test statistic T_0 could be defined as:

$$T_0 = max_{1\le k\le n}T(k) \tag{9}$$

Jarušková (1996) further studied this method and developed a test statistic (T_n) based on T_0, which could be defined as:

$$T_0 = \frac{n(T(n))^2}{n-2+(T(n))^2} \tag{10}$$

If T_0 exceeds a certain threshold the null hypothesis will be rejected. Critical values can be found in Winjgaard et al. (2003) Table III.

Trend analysis

DroughtStats incorporates a comprehensive suite of trend tests to identify temporal trends in hydrometeorological datasets. These tests are implemented using the open-source pymannkendall library (Hussain and Mahmud 2019), a widely recognized tool for nonparametric trend analysis. The included methods are the Mann–Kendall trend test (Kendall 1955; Mann 1945), the Modified Mann–Kendall trend test (Hamed and Ramachandra Rao 1998), which accounts for autocorrelation effects, the Seasonal Mann–Kendall trend test (Hirsch et al. 1982), designed for data with periodic variability, and Sen's slope test (Sen 1968), which quantifies the magnitude of detected trends. The use of pymannkendall library ensures methodological rigor and reproducibility in trend analysis.

Mann–Kendall trend test is a nonparametric and distribution free test that that performs well even in the presence of outliers (Önöz and Bayazıt 2003). The test is recommended by the World Meteorological Organization and has been widely used in numerous hydrological studies. The trend statistic S is calculated as shown in Eq. (11).

$$S = \sum_{i<j} sgn\left(x_j - x_i\right) \tag{11}$$

where,

$$sgn\left(x_j - x_i\right) = \begin{cases} 1 x_j > x_i \\ 0 x_j = x_i \\ -1 x_j < x_i \end{cases} \tag{12}$$

The variance of trend statistic S is calculated as shown in Eq. (13).

$$Var(S) = \frac{n(n-1)(2n-5)}{18} \tag{13}$$

The standardized test statistic Z can then be calculated based on trend statistic S, the variance of trend statistic $Var(S)$ and the specified significance level α.

$$Z = \frac{S}{Var(S)^{\alpha}} \tag{14}$$

Significant autocorrelations within a time series can interfere with accurately assessing the variance of trend statistic S. To address this issue, Hamed and Ramachandra Rao (1998), suggested that a nonparametric trend estimator $V*(S)$ can be derived directly from the original time series.

$$V^*(S) = Var(S)Cor \tag{15}$$

where,

$$Cor = 1 + \frac{2}{n(n-1)(n-2)} \sum_{i=1}^{n-1} (n-1)(n-i-1)(n-i-2)\rho_s(i) \tag{16}$$

In Eq. (16), n represents the total number of observations in the time series and serves to normalize the correction factor, ensuring that the adjustment for autocorrelation is appropriately scaled relative to the dataset size. Larger datasets reduce the relative influence of individual correlations, as the weight assigned to each lag diminishes with increasing n. The term $\rho_s(i)$ denotes the autocorrelation at lag i, which quantifies the dependence between data points separated by i time steps. Positive values of $\rho_s(i)$ indicate a direct relationship between observations at that lag, whereas negative values suggest an inverse relationship. These lag-specific correlations are crucial for accounting for the internal structure of the time series, ensuring that the variance adjustment reflects the true level of independence in the dataset.

Seasonal Mann–Kendall test is a special variation of Mann–Kendall trend test that takes seasonality into account. Hirsch et al. (1982) suggest that dataset X is made up of subsamples X_i and H_0' hypothesis is that X is a sample of independent random variables (x_{ij}) and that X_i is a subsample of independent and identically distributed random variables $i = 1, 2, \ldots, 12$. Then the test statistic S_i can be defined as:

$$S_i = \sum_{k=1}^{n_i-1} \sum_{j=k+1}^{n_i} sgn\left(x_{ij} - x_{ik}\right) \tag{17}$$

Then, variance of the test statistic S' can be defined as:

$$Var(S\prime) = \sum_{i=1}^{12} Var\left(S_i\right) = \sum_{i=1}^{12} \frac{n_i\left(n_i-1\right)\left(2n_i+5\right)}{18} \tag{18}$$

And the standard normal deviate Z is then defined as:

$$Z = \begin{cases} \frac{S\prime-1}{(Var(S\prime))^{\frac{1}{2}}} if S\prime > 0 \\ 0 if S\prime = 0 \\ \frac{S\prime+1}{(Var(S\prime))^{\frac{1}{2}}} if S\prime < 0 \end{cases} \tag{19}$$

Sen's slope is a method developed by Sen (1968) that determines the magnitude of a trend in time series. The slope is calculated as shown in Eq. (20).

$$S_{ij} = \frac{y_j - y_i}{x_j - x_i}, 1 \le i \le j \le n \tag{20}$$

Sen's slope estimator $\widehat{\beta}$ then can be defined as shown in Eq. (21).

$$\widehat{\beta} = median\left(S_{ij}\right) \tag{21}$$

Sen's slope estimator $\widehat{\beta}$ then can be used to calculate the intercept $\widehat{\alpha}$ as shown in Eq. (22).

$$\widehat{\alpha} = median\left(y_i - \widehat{\beta}x_i\right) \tag{22}$$

DroughtStats integrates these trend tests into a user-friendly framework, leveraging the pymannkendall library for robust computation. Results are exported into Excel files, enabling users to analyze trends efficiently and effectively.

Probability distributions

While the primary focus of DroughtStats is drought analysis, the accurate selection of appropriate PDFs is crucial for reliable drought analysis. DroughtStats enables users to fit their data to a range of PDFs, which are an essential component of drought indices and their calculation. The software supports a total of 31 different PDFs including the Beta Prime, Cauchy, Exponential, Exponential Normal (2p), Exponential Normal (3p), Fisk, Gamma (3p), Gamma (2p), Generalized Extreme, Generalized Exponential, Generalized Inverse Gaussian, Generalized Logistic, Generalized Gamma, Generalized Pareto, Gumbel (Left), Gumbel (Right), Inverse Gamma, Inverse Gaussian, Johnson SB, Logistic, Log-Normal (3p), Log-Normal (2p), Log Uniform, Mielke (3p), Normal Inverse Gaussian, Pareto (2p), Pearson 3, Rayleigh, Uniform, Weibull Max, and Weibull Min (2p) distributions. For the sake of simplicity, this paper will explain the calculation process using only the Gamma distribution.

The 2-parameter Gamma distribution function is defined by its shape parameter k and scale parameter θ. The probability density function of 2-parameter Gamma distribution can be defined as:

$$f(x;k,\theta) = \frac{x^{k-1}e^{-\frac{x}{\theta}}}{\theta^k \Gamma(k)} \tag{23}$$

where $\Gamma(k)$ is,

$$\Gamma(k) = \int_0^\infty t^{k-1}e^{-t}dt \tag{24}$$

Fitting a distribution function to a dataset involves estimating the distribution parameters from the sample data. In DroughtStats, the Maximum Likelihood Estimation (MLE) method is used to estimate these parameters. The MLE method aims to identify the values of parameters k and θ that maximize the log-likelihood function, which quantifies the likelihood of observing the given sample under the assumed probability distribution.

The log-likelihood function of Gamma distribution is defined as shown in Eq. (25).

$$l : l(k,\theta) = \sum_{i=1}^{n}[(k-1)lnx_i - \frac{x_i}{\theta} - kln\theta - ln\Gamma(k)] \tag{25}$$

where $\Gamma(k)$ denotes the Gamma function.

To estimate the parameters, the partial derivatives of the log-likelihood function with respect to k and θ are computed and set equal to zero. Solving these equations provides the parameter estimates that maximize the likelihood of observing the given data under the Gamma distribution.

In DroughtStats, probability distribution functions are fitted to the data, and the Kolmogorov–Smirnov (K-S) statistic is used to assess the goodness-of-fit. The K-S statistic measures the maximum distance between the empirical distribution function (EDF) of a sample and the cumulative distribution function (CDF) of a reference distribution. The K-S statistic is defined as (Massey 1951):

$$D_n = sup_x\left|F_n(x) - F(x)\right| \tag{26}$$

where $F_n(x)$ is the EDF of the sample data, and $F(x)$ is the theoretical CDF of the fitted distribution function. Figure 2 illustrates an example of streamflow data fitted using the DroughtStats software.

Univariate drought indices

A univariate drought index is typically derived from a single hydrometeorological variable, such as the Standardized Precipitation Index (SPI) (McKee et al. 1993), which is based solely on precipitation. Drought indices are commonly developed using percentiles, anomaly, or standardization techniques to accurately characterize drought conditions. The methodology employed to calculate the SPI can also be applied to other variables, including soil moisture, streamflow, runoff, groundwater, and snowmelt, to create various types of Standardized Drought Indices.

DroughtStats offers a range of drought indices, including those based on the methodology of the SPI and non-parametric methods derived from plotting positions, such as the Gringorten (1963) plotting position formula. These indices include, but are not limited to, the SPI, Standardized Precipitation Evapotranspiration Index (SPEI), Standardized Runoff Index (SRI), Streamflow Drought Index (SDI), Standardized Soil Moisture Index (SSI), and Standardized Groundwater Index (SGI). (Bloomfield and Marchant 2013; Nalbantis and Tsakiris 2008; Shukla and Wood 2008; Vicente-Serrano et al. 2010).

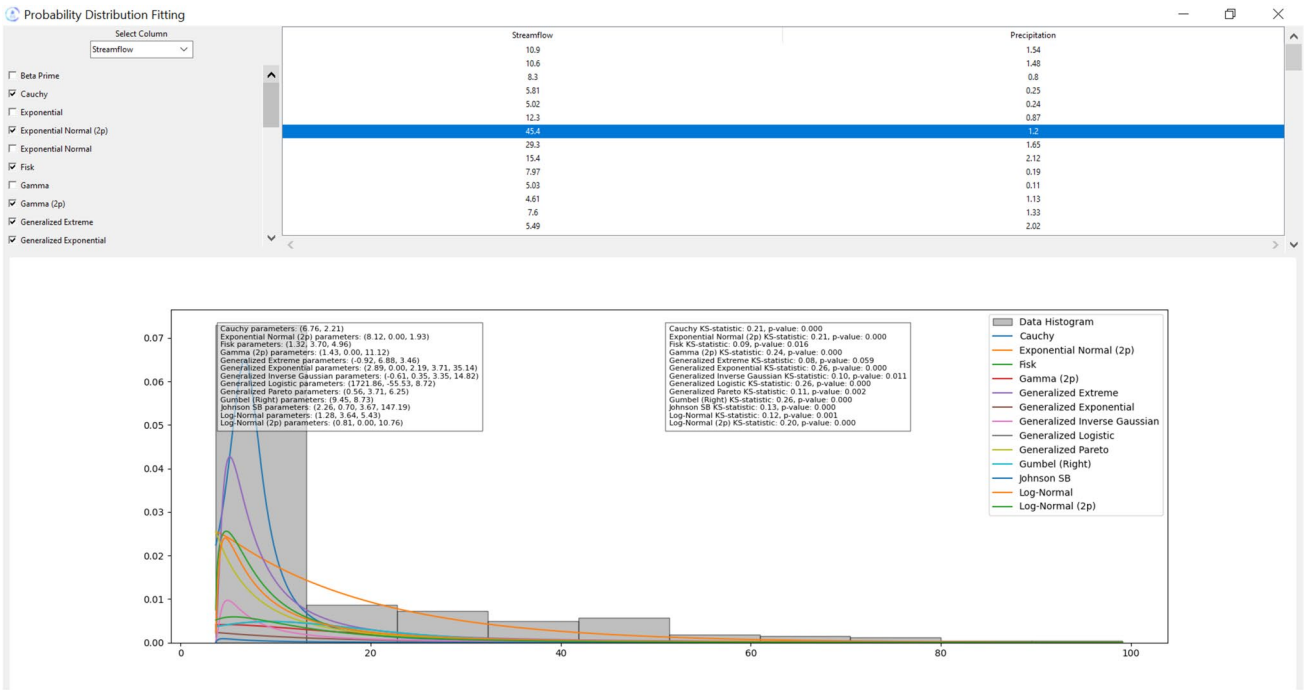

Fig. 2 Fitting probability distribution functions to data using DroughtStats

Nonparametric standardized drought indices

Farahmand and AghaKouchak (2015) suggested that, rather than using parametric probability distribution functions, the empirical probability could be employed to derive a nonparametric standardized drought index. By applying the Gringorten (1963) plotting position formula, the marginal probability of hydrometeorological variables can be defined as:

$$p(x_i) = \frac{i - 0.44}{n + 0.12} \tag{27}$$

In Eq. (27) n denotes the number of observed data while i represents the rank of precipitation data from the smallest. The output of the Eq. (27) is then used as an input in Eq. (28) to calculate the nonparametric Standardized Index (SI):

$$SIe = \phi^{-1}(p) \tag{28}$$

The symbol ϕ denotes the standard normal distribution function, while p represents the calculated joint probability, as determined using Eq. (27).

Drought indices are highly dependent on the choice of PDFs, and the results can vary significantly based on the selected distribution. This is particularly problematic when the chosen PDF does not adequately represent the underlying characteristics of the hydrometeorological variable, leading to potentially inaccurate assessments. Nonparametric standardized drought indices address this issue by eliminating the need for predefined PDFs. These indices rely on empirical data, making them more flexible and robust, particularly in regions or conditions where the appropriate distribution is uncertain or difficult to determine. As such, nonparametric approaches provide a reliable and globally applicable method for assessing drought conditions, avoiding the biases that can arise from inappropriate distribution choices.

Parametric standardized drought indices

Parametric standardized drought indices can be calculated using various probability distribution functions. However, for simplicity, this paper will focus on explaining the calculation process of the parametric Standardized Precipitation Index (SPI), assuming the use of the Gamma distribution. As discussed in Sect. 2.2, DroughtStats enables users to fit different distribution functions to the data and estimate the required parameters. Once the parameters are estimated, the cumulative distribution function (CDF) of the Gamma distribution can be defined as:

$$F(x) = \frac{1}{\Gamma(k)}\gamma\left(\mathrm{k}, \frac{\mathrm{x}}{\theta}\right) \tag{29}$$

where $\Gamma(k)$ is the gamma function evaluated at k, k is the shape parameter and θ is the scale parameter. Then $H(x)$, cumulative probability from the Gamma distribution, can be defined as:

$$H(x) = F(x) \tag{30}$$

Using the inverse of standard normal cumulative distribution function, cumulative probability from the Gamma distribution can be converted into SPI values as shown in Eq. (31).

$$SPI = \phi^{-1}(H(x)) \tag{31}$$

where ϕ represents the standard normal cumulative distribution function.

In case of zero precipitation values in the dataset, probability of zero precipitation (q) is incorporated into the Gamma distribution.

$$H(x) = q + (1 - q)F(x) \tag{32}$$

In the case of Eq. (32), $H(x)$ incorporates both the probability of zero precipitation and the Gamma CDF for non-zero precipitation. To ensure that the CDF values remain within the valid range of [0, 1], DroughtStats introduces a constant. This adjustment addresses the issue encountered with the inverse of a standard normal distribution, where the values asymptotically approach $-\infty$ and $+\infty$ as the CDF values approach 0 and 1, respectively. By incorporating this constant, the CDF is effectively constrained, preventing these extreme values and maintaining consistency in the calculation of the standardized drought index.

Figure 3 illustrates the calculation window for parametric standardized drought indices. After the calculations are completed, the results are exported and saved into an Excel file for further analysis.

$SI_{seasonality}$

Based on the framework developed by Terzi and Önöz (2024a), DroughtStats incorporates the calculation of $SI_{seasonality}$, which provides a methodology for computing standardized drought indices that account for seasonal variability within the data. This approach differs from traditional methods that utilize a single best-fit probability distribution for the entire dataset. Instead, it identifies and applies best-fit probability distributions for each subgroup of data throughout the calculation process. By automatically selecting the optimal probability distribution for each subgroup, DroughtStats calculates the indices accordingly and reconstructs the index time series, offering a refined and contextually sensitive measure of drought conditions.

The calculation steps of $SI_{seasonality}$ are as follows:

1. The selected data are aggregated over the specified time scale.
2. The aggregated data are divided into twelve distinct groups to account for seasonal variability.
3. Goodness-of-fit tests are used to determine the most suitable PDF for each group.

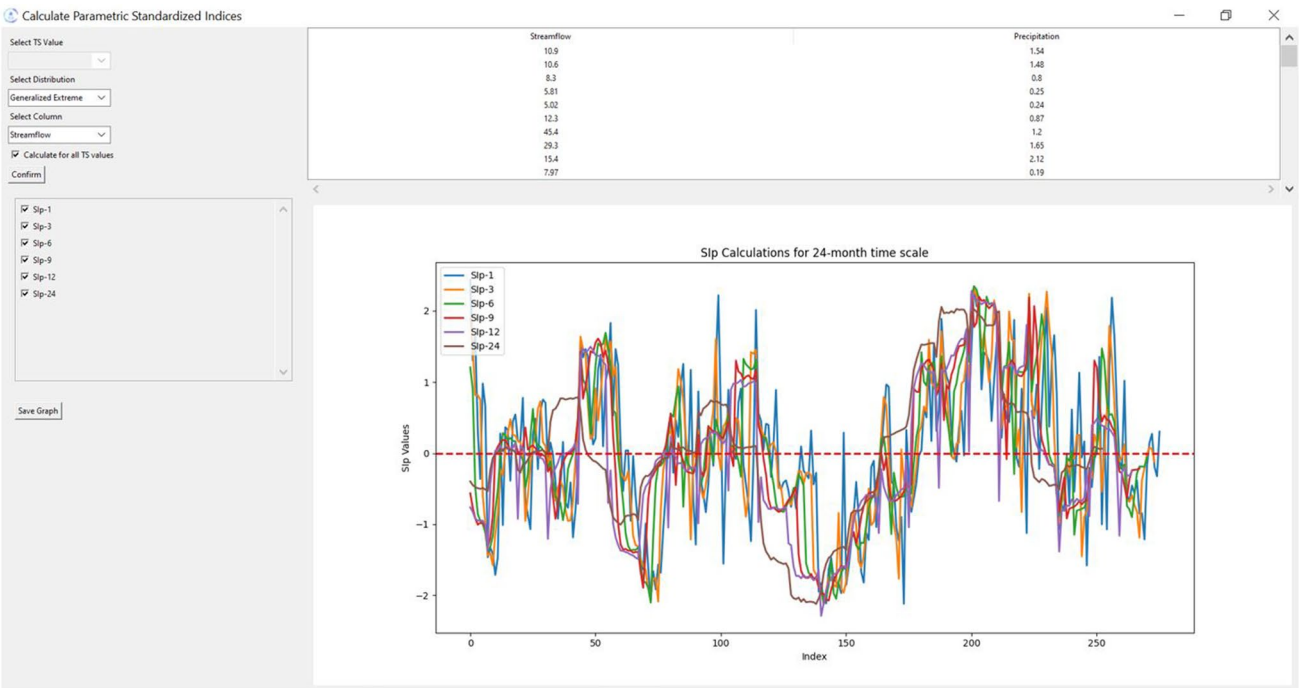

Fig. 3 Calculation of parametric standardized drought indices in DroughtStats

4. Each group is individually fitted with its best-fitting PDF, as identified by the goodness-of-fit tests.
5. A CDF is constructed for each fitted PDF to compute the cumulative probability for each group.
6. These cumulative probabilities are standardized by applying the inverse of the standard normal distribution function to each group's CDF.
7. The standardized probabilities are then combined to reconstruct the data into a unified time series.

Reconnaissance drought index

Tsakiris and Vangelis (2005) proposed a meteorological drought index that incorporates both precipitation and evapotranspiration. The proposed index, also known as the Reconnaissance Drought Index (RDI), is based on the cumulative values of precipitation and potential evapotranspiration.

The starting value of the index for a specific period, denoted by a particular month (k) within a year, is determined as follows:

$$a_k = \frac{\sum_{j=1}^{j=k} Pj}{\sum_{j=1}^{j=k} PET_j} \tag{33}$$

Two forms of RDI, Normalized (RDI_n) and Standardized (RDI_{st}), then can be calculated based on a_k as follows:

$$RDI_n(k) = \frac{a_k}{\bar{a}_k} - 1 \tag{34}$$

$$RDI_{st}(k) = \frac{y_k - \bar{y}_k}{\hat{\sigma}_k} \tag{35}$$

where, y_k is the $\ln(a_k)$, $\bar{y}_k$ represents the mean and $\hat{\sigma}_k$ is the standard deviation.

Deciles index

Deciles Index (DI) is one of the simplest drought indices in the literature (Gibbs and Maher 1967). Hydrometeorological data is ranked in ascending order and divided into ten equal parts, each representing %10 of the data. The deciles are then group into five classes as shown in Table 1.

Table 1 Classification of decile index

Decile Class	Description
deciles 1–2: lowest 20%	much below normal
deciles 3–4: next lowest 20%	below normal
deciles 5–6: middle 20%	near normal
deciles 7–8: next highest 20%	above normal
deciles 9–10: highest 20%	much above normal

Multivariate drought indices

DroughtStats enables users to calculate both nonparametric and parametric multivariate drought indices, with the latter based on various Copula families including Gaussian Copula, T Copula, Clayton Copula, Gumbel Copula and Frank Copula.

Nonparametric multivariate standardized drought index

The joint probability of selected precipitation and streamflow data is described in the Eq. (36), precipitation and streamflow are represented by X and Y respectively.

$$P(X \le x, Y \le y) = p \tag{36}$$

In the Eq. (36) p denotes the joint probability of the selected data. MSDI then can be defined on joint probability of selected hydrometeorological variables by utilizing the inverse of a standard normal distribution as shown in the Eq. (37) (Hao and AghaKouchak 2013):

$$MSDIe = \phi^{-1}(p) \tag{37}$$

MSDI, calculated based on the joint probability of the X and Y using the standard normal distribution function (ϕ), provides insights into drought conditions across time spans of 1, 3, 6, and 12 months (Hao and AghaKouchak 2014).

In the case of bivariate analysis, the empirical joint probability can be estimated using a modified version of the Gringorten plotting position formula, as proposed by Gringorten in 1963:

$$P(x_{k,}y_k) = \frac{(m_k - 0.44)}{(n + 0.12)} \tag{38}$$

In Eq. (38), n denotes the number of observed precipitation and streamflow data while m_k represents the number of occurrences of the (x_i, y_i) pair such that both x_i is less than or equal to x_k and y_i is less than or equal to y_k. The output of the Eq. (38) will then be used as an input for Eq. (37) to calculate the MSDI series.

Copula-based multivariate standardized drought index

In hydrological studies, copulas are used to model the joint distribution of multiple hydrometeorological variables, such as precipitation, temperature, and streamflow, which are essential for assessing drought conditions. If the hydrometeorological variables are considered as random variables X and Y, the joint distribution with the cumulative joint probability p can be represented using a copula C (Nelsen 2006).

$$P(X \le x, Y \le y) = C[F(X), G(Y)] = p \tag{39}$$

where C is the copula function, $F(X)$ and $G(Y)$ are the marginal cumulative distributions of random variables X and Y, respectively. A Copula C provides the ability to build the joint distribution of random variables based on their individual marginal distributions. In hydrological studies, the application of copulas to model nonlinear dependence structures within multivariate data has become increasingly prevalent (Ballarin et al. 2021; Hao and AghaKouchak 2013; Varol et al. 2023; Xiang et al. 2024; Zhang et al. 2023).

DroughtStats allows users to compute parametric multivariate drought indices using different copula families. For the sake of simplicity, this paper will focus solely on explaining the calculation process for the Clayton copula (Clayton 1978).

The Clayton copula for the bivariate case is defined as follows:

$$C_\theta(u, v) = \left[\max\left(u^{-\theta} + v^{-\theta} - 1; 0\right)\right]^{-\frac{1}{\theta}} \tag{40}$$

where u and v are marginal cumulative distribution functions of the random variables and θ is a parameter that controls the strength of the dependence between the variables. The Clayton copula is particularly useful for modeling variables that exhibit strong lower-tail dependence, which is often seen in drought conditions where extreme low values are correlated.

To fit a copula function to a dataset, the following steps are employed:

1. First, the marginal cumulative distribution functions (CDFs) for each variable are computed using the empirical data. In DroughtStats, the Gringorten Plotting Position Formula (1963) is used to derive the empirical CDFs. This formula provides a non-parametric way to estimate the cumulative distribution based on the ranks of observed data points, ensuring that each observation is mapped to its corresponding cumulative probability.
2. Once the marginal distributions are determined, Maximum Likelihood Estimation (MLE) is used to estimate the copula parameters. MLE involves maximizing the likelihood function, which represents the probability of observing the given data under the selected copula model. The parameters are estimated by minimizing the negative log-likelihood function.
3. After the copula parameters are estimated, the joint CDF for the specified copula family is constructed. The copula function combines the marginal distributions with the estimated parameters to model the multivariate dependence between the variables.

The MSDI is then computed using the joint CDF derived from the copula. The MSDI values are obtained by applying the inverse of the standard normal cumulative distribution function to the copula values as shown in Eq. (41).

$$MSDI = \phi^{-1}\left(C_\theta(u, v)\right) \tag{41}$$

where ϕ^{-1} denotes the inverse of the standard normal CDF, converting the copula values into the standardized drought index.

To identify the most suitable copula model for the data, DroughtStats uses two common information criteria: the Akaike Information Criterion (AIC) (Akaike 1988) and the Bayesian Information Criterion (BIC) (Schwarz 1978). These criteria balance the goodness of fit of the model with its complexity to avoid overfitting. The AIC can be defined as follows:

$$AIC = 2k - 2\ln\left(\widehat{L}\right) \tag{42}$$

where k is the number of estimated parameters and $\widehat{L}$ is the maximized value of the likelihood function.

The BIC is defined as:

$$BIC = kln(n) - 2\ln\left(\widehat{L}\right) \tag{43}$$

where n is the number of observations, k is the number of estimated parameters and $\widehat{L}$ is the maximized value of the likelihood function.

By comparing the AIC and BIC values for different copula families, DroughtStats selects the copula that best fits the data and models the multivariate dependence structure. Once the MSDI values are calculated, DroughtStats saves the results, including the monthly AIC and BIC values, into an Excel file (.xlsx) chosen by the user. This allows for easy analysis and further processing of the multivariate drought indices.

Figure 4 illustrates the process of MSDI calculation in DroughtStats, providing a visual representation of the copula-based methodology.

Multivariate standardized drought index based on different Copula functions

Many hydrological studies in the literature employ copula functions to assess their suitability for specific datasets, often focusing on a single copula family. Based on the methodology introduced by Terzi and Önöz (2024b), DroughtStats takes a more comprehensive approach by evaluating the fit of multiple copula families for each month, accounting for the seasonality of hydrometeorological data. This method allows for the selection of the most appropriate copula family for each month and the calculation of Multivariate Standardized

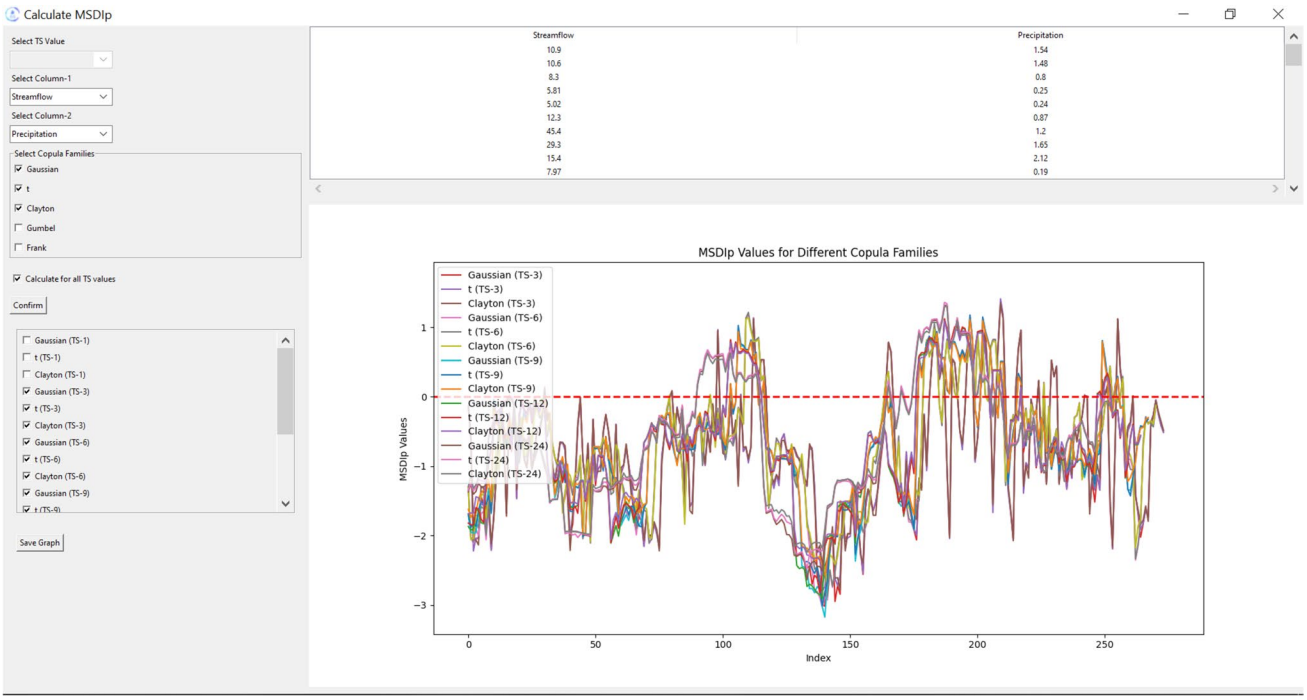

Fig. 4 Calculation of parametric MSDI in DroughtStats

Drought Index (MSDI) values based on the best-fit copula. Calculation window of MSDI using the newly developed methodology is shown in Fig. 5.

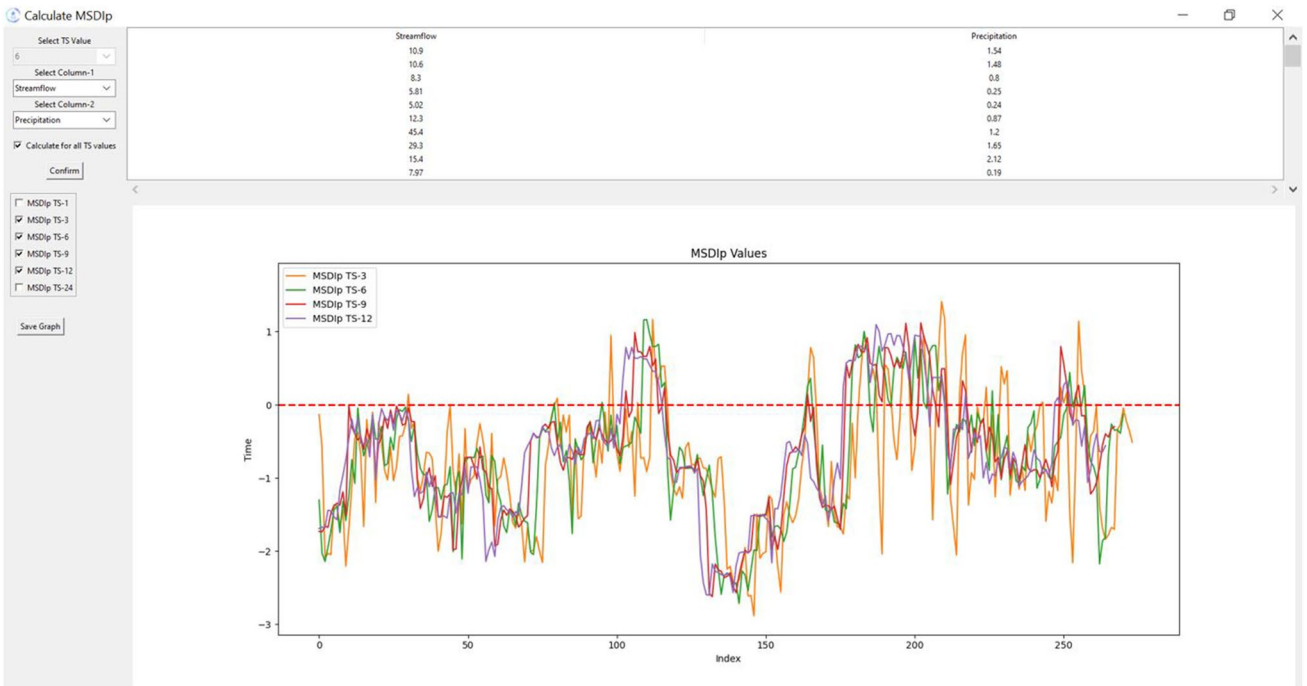

Fig. 5 Calculation of MSDI based on different Copula families in DroughtStats

Estimation of potential evapotranspiration

DroughtStats estimates Potential Evapotranspiration (PET) using three widely recognized methods: the Thornthwaite equation, the Hargreaves method, and the Penman–Monteith equation. For brevity, the detailed equations for these methods are not presented here, but references to the original sources are provided for further reading.

1. The Thornthwaite equation estimates PET based on temperature data. This method, described in detail by Thornthwaite (1948), is widely used due to its simplicity and applicability across various climates.
2. The Hargreaves method is an empirical formula that estimates PET using temperature data, and is particularly useful in regions with limited meteorological data. The detailed equations and assumptions behind this method can be found in Hargreaves et al. (1985).
3. The Penman–Monteith equation is considered the most accurate method for PET estimation and incorporates a wide range of meteorological variables, including temperature, radiation, humidity, and wind speed. A detailed explanation of the method can be found in Allen et al. (1998), Chapter 2.

DroughtStats supports these methods and allows for the estimation of PET based on the user's dataset, including necessary inputs such as temperature, latitude, and altitude. The software also provides flexibility by calculating actual vapor pressure based on available data, such as relative humidity.

Drought characteristics

DroughtStats calculates drought characteristics based on the principles of Yevjevich's (1967) run theory, which is widely used in drought analysis. The software identifies the onset and conclusion of drought events and calculates their severity, intensity, frequency, and recovery duration. This method is highly flexible, allowing users to define key parameters that best suit the specific context of their analysis. Since drought thresholds and the minimum duration required to recognize a drought event can vary across different studies and regions, DroughtStats provides a user-defined approach to identifying and evaluating drought events.

The calculation of drought characteristics relies on three main input parameters: the drought threshold, the recovery threshold, and the minimum duration. The drought threshold is the level below which a drought event is considered to have started and ended. Once the drought index falls below this threshold, the system marks the onset of the drought and subsequently its conclusion once the index rises back above it. The recovery threshold, on the other hand, defines the level at which the region is considered to have recovered from the drought. Recovery occurs when the drought index exceeds this threshold, indicating that the drought's impacts are no longer present.

The minimum duration is the third crucial parameter, specifying the minimum number of months the drought index must remain below the drought threshold for an event to be recognized as a drought. This ensures that short-term fluctuations do not erroneously trigger drought events, providing a more accurate representation of sustained drought conditions.

DroughtStats uses these parameters to identify separate drought events, calculate their start and end points and assess their severity, duration, and intensity. Additionally, the software determines whether the region has fully recovered from the drought. In cases where a drought event ends but another event occurs before the recovery threshold is reached, the area is considered not to have recovered from the initial drought, and the region is marked as having "no recovery". Figure 6 provides a visual representation of this process, illustrating how DroughtStats calculates and evaluates drought characteristics based on the input parameters.

The calculated drought characteristics are then exported and saved into an Excel file in tabular format for further analysis.

Data relations

DroughtStats provides users with the capability to explore the relationship between two time series through the 'Data Relations' window. This section outlines the specific dependence and correlation measures employed to analyze and quantify the relationships between the selected time series.

Spearman's rank correlation coefficient (ρ)

Spearman's rank correlation coefficient (ρ) is a nonparametric statistic used to measure the rank-based relationship between two variables (Spearman 1904). It evaluates how well the relationship between the variables can be represented by a monotonic function, indicating that as one variable increases, the other consistently either increases or decreases. The formula for calculating Spearman's rank correlation coefficient is as follows:

$$\rho = 1 - \frac{6 \sum d_i^2}{n\left(n^2 - 1\right)} \tag{44}$$

In Eq. (44), d_i represents the difference between the ranks of each observation, and n denotes the total number of observations.

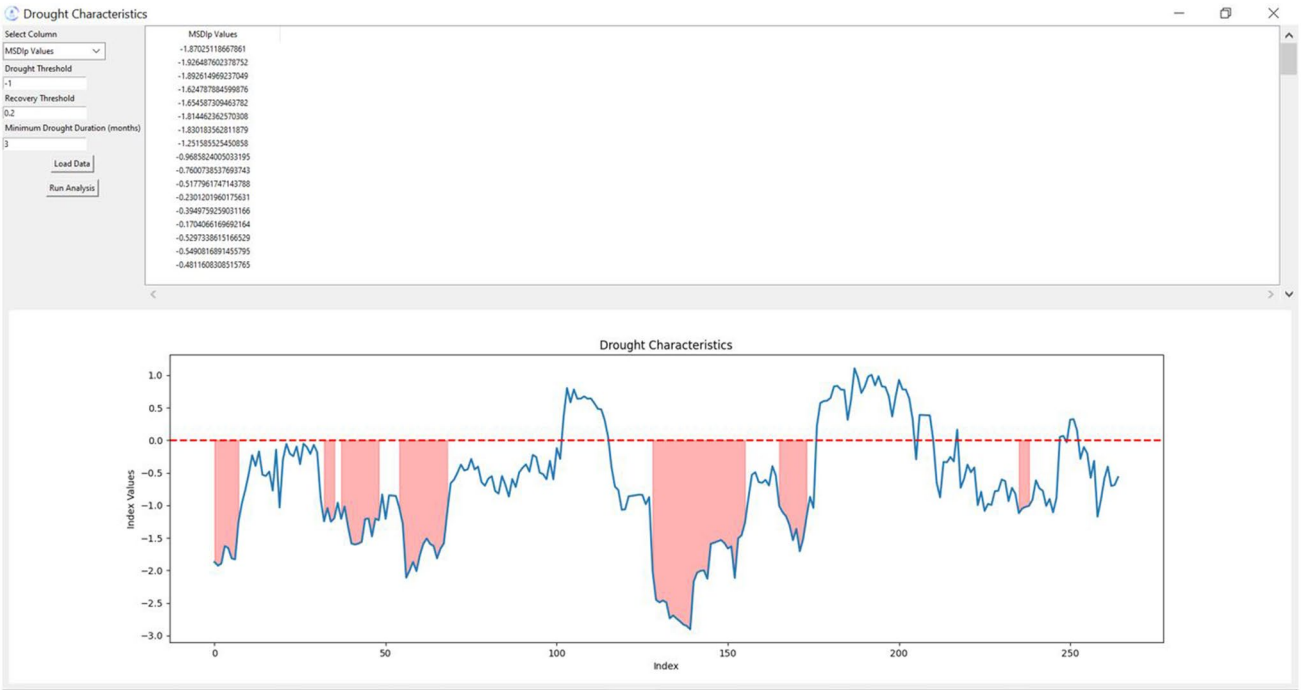

Fig. 6 Calculation of drought characteristics in DroughtStats

Kendall's rank correlation coefficient (T)

Kendall's rank correlation coefficient (T) is a non-parametric statistic used to measure the ordinal association between two variables (Kendall 1938). Similar to Spearman's rank correlation, it evaluates the strength and direction of a monotonic relationship between the variables; however, it specifically considers the relative ordering of pairs of observations. Kendall's T is based on the concepts of concordant and discordant pairs: a pair is concordant if the ranks of both variables align, while it is discordant if the ranks of the variables disagree. The calculation of Kendall's T is defined as follows:

$$\mathrm{T} = \frac{N_C - N_D}{\frac{1}{2}n(n-1)} \tag{45}$$

where N_C is the number of concordant pairs, N_D is the number of discordant pairs and n is the number of observations.

Mutual information (MI)

Mutual information (MI) quantifies the amount of information gained about one random variable through the observation of another random variable. It serves as a measure of the mutual dependence between two variables, effectively capturing both linear and non-linear relationships between them (Shannon 1948). The MI between two continuous random variables is calculated as follows:

$$I(X;Y) = \iint P_{(X,Y)}(x,y)\log\left(\frac{P_{(X,Y)}(x,y)}{P_X(x)P_Y(y)}\right)dxdy \tag{46}$$

where P_X and P_Y are the marginal probability density functions of X and Y respectively and $P_{(X,Y)}$ is the joint probability density function of X and Y.

Distance correlation (dCor)

Distance correlation measures the dependence between two random variables by capturing both linear and non-linear associations (Székely et al. 2007). It extends the traditional concept of correlation by focusing on the distances between pairs of observations rather than their values or ranks. Distance covariance quantifies this dependence by evaluating the distances between pairs of observations. The formula for calculating distance covariance is as follows:

$$dCov^2(X,Y) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\widetilde{D}_x(i,j) * \widetilde{D}_y(i,j) \tag{47}$$

where $\widetilde{D}_x(i,j)$ and $\widetilde{D}_y(i,j)$ are the centered version of distance matrices for the variables X and Y, respectively.

Distance correlation is the normalized form of distance covariance, and is calculated as follows:

$$dCor(X,Y) = \frac{dCov(X,Y)}{\sqrt{dVar(X) * dVar(Y)}} \quad (48)$$

where $dVar(X)$ and $dVar(Y)$ are the measures of dispersion of the distances in X and Y, respectively.

Copula-based Kendall's tau (T_c)

Schweizer and Wolff (1981) introduced copula-based Kendall's T, which employs copula functions to quantify the dependence between two random variables. The formula for copula-based Kendall's T is defined as follows:

$$T = 4\int_0^1\int_0^1 C(u,v)dC(u,v) - 1 \quad (49)$$

where $C(u,v)$ is the copula function that represents the joint cumulative distribution of the random variables after transforming them into their uniform marginals.

In copula theory, Kendall's T is a measure of concordance that ranges from 1 to −1, where 1 signifies perfect positive dependence and −1 denotes perfect negative dependence. By employing copulas, Kendall's T can effectively capture the true dependence structure between variables, even when the relationship is non-linear or characterized by complex dependencies.

Application

The Çoruh River Basin, located in Turkey, serves as the study area for demonstrating the application of the proposed software. The CRB is highly significant because of its crucial contribution to Turkey's economic and social development (Terzi 2024). This analysis utilizes monthly precipitation and streamflow data spanning from 1989 to 2011, collected from a total of 8 hydrological and 8 meteorological stations distributed throughout the basin. For the purpose of this demonstration, the calculations will be presented for a single representative station, simplifying the exposition while maintaining methodological rigor.

Statistical analysis

Prior to conducting the drought analysis, a comprehensive statistical evaluation of the precipitation and streamflow data was conducted. This preliminary assessment included the calculation of various statistical metrics, which provide essential insights into the data and serve as a foundational step for subsequent drought analysis. The statistical metrics for precipitation and streamflow data are presented in Table 2 and Table 3, respectively.

Homogeneity tests indicate that both the streamflow and precipitation data are homogeneous, and no significant trends were observed in either dataset.

Calculation of univariate indices

The most suitable probability distribution functions for both datasets were determined to be the Generalized Extreme Value (GEV) distribution. Although the GEV distribution exhibits a lower K-S value compared to the Gamma distribution, Gamma distribution is traditionally recommended for the calculation of SPI due to its effective representation of precipitation characteristics. This study employs the GEV distribution for the calculation of SSFI and two-parameter Gamma distribution for the calculation of SPI.

The SPI and SSFI indices are calculated using both nonparametric and parametric methods across 1, 3, 6, 9, and 12-month time scales. The results for the SPI are shown in Fig. 7, while the SSFI results are presented in Fig. 8. Although parametric and empirical indices share similarities, as highlighted by Terzi and Önöz (2023), significant divergences can be observed at the distribution tails.

Calculation of multivariate standardized drought index

In this study, the Multivariate Standardized Drought Index (MSDI) is calculated in three distinct ways: as a nonparametric MSDI using the Gringorten (1963) plotting position formula modified for the bivariate case; as a parametric MSDI using different copula families; and, as detailed in Sect. 2.4.3, by applying optimal the copula family for each corresponding month. The calculation of the parametric

Table 2 Statistical metrics of precipitation data for station E2304 in the CRB

Mean	Stdev	Variance	Median	Mode	Skewness	Kurtosis	Coefficient of Variation	Autocorrelation
1.29	0.88	0.77	1.16	1.55	1.00	1.15	0.68	0.22

Table 3 Statistical metrics of streamflow data for station E2304 in the CRB

Mean	Stdev	Variance	Median	Mode	Skewness	Kurtosis	Coefficient of Variation	Autocorrelation
15.88	16.98	288.49	7.61	10.4	2.08	4.24	1.07	0.58

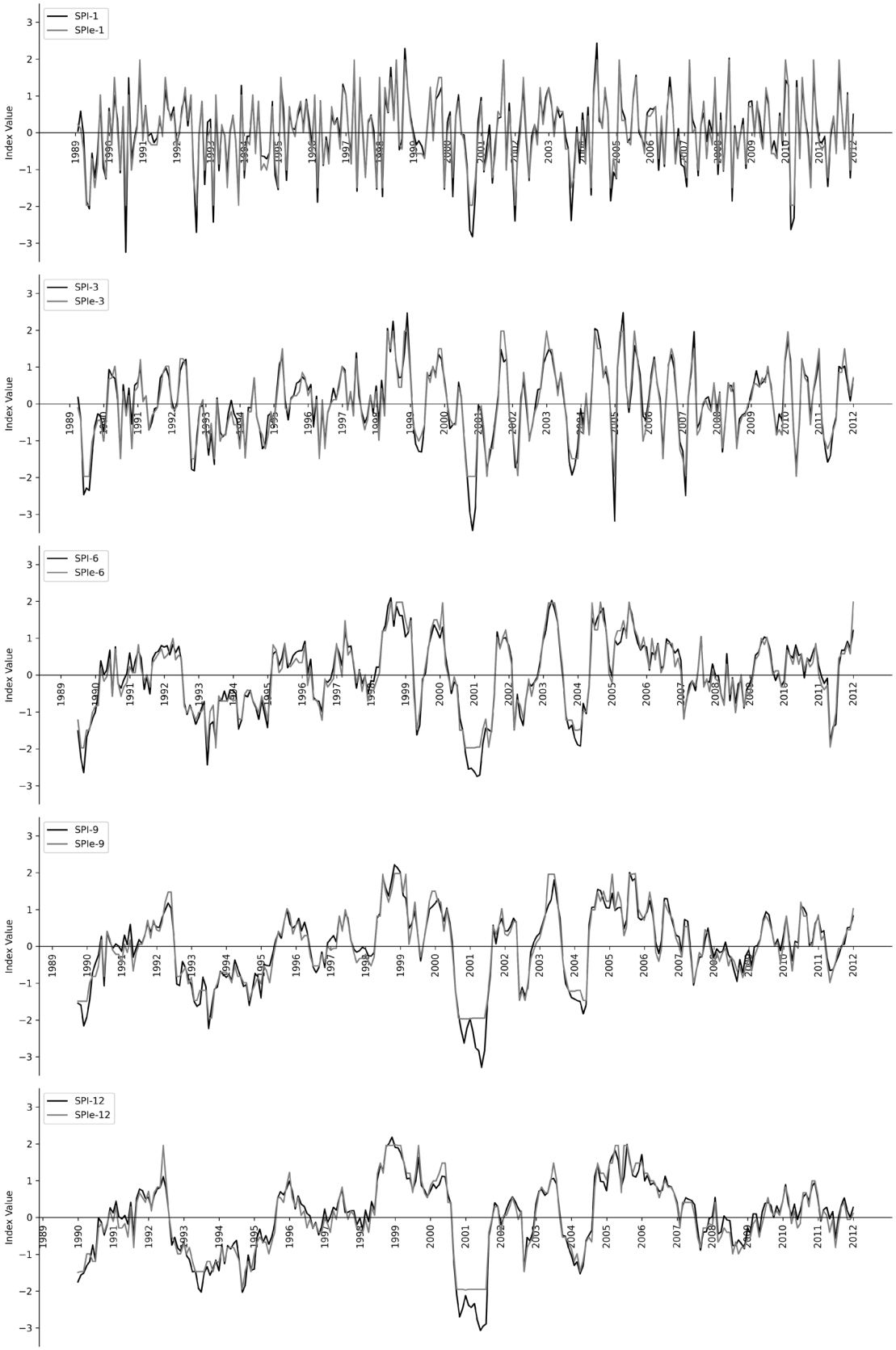

Fig. 7 Calculated parametric and nonparametric SPI values

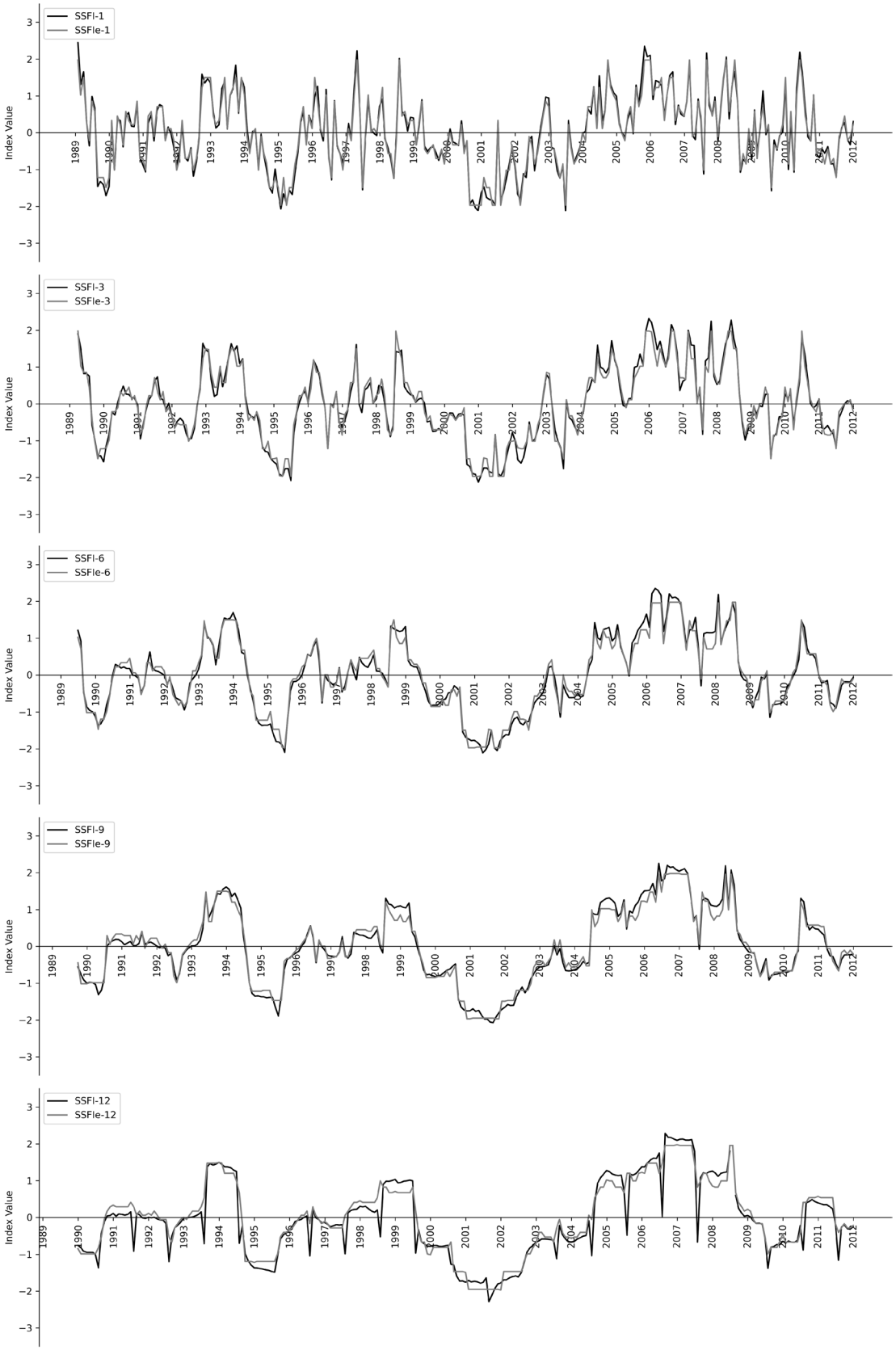

Fig. 8 Calculated parametric and nonparametric SSFI values

Table 4 Monthly AIC and BIC values for different copula families

	Gaussian		T		Clayton		Gumbel		Frank	
Month	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
Oct	**1.55**	2.69	3.40	5.67	2.00	3.14	2.00	3.14	1.75	2.89
Nov	1.94	3.07	3.20	5.47	2.00	3.14	**0.63**	1.76	2.00	3.13
Dec	2.00	3.14	3.80	6.07	**1.71**	2.84	2.00	3.14	1.74	2.88
Jan	2.00	3.14	3.64	5.92	2.00	3.14	**1.99**	3.13	2.00	3.14
Feb	0.28	1.41	2.27	4.54	2.00	3.14	2.00	3.14	**−0.79**	0.34
Mar	0.40	1.53	2.39	4.66	1.37	2.51	0.60	1.74	**−0.20**	0.93
Apr	1.86	3.00	3.86	6.13	2.00	3.14	**1.76**	2.90	1.94	3.08
May	1.01	2.14	−0.64	1.63	**−2.80**	−1.66	1.60	2.74	0.43	1.56
Jun	−0.90	0.24	1.09	3.36	**−3.93**	−2.79	1.22	2.35	−0.02	1.11
July	1.80	2.94	3.40	5.68	2.00	3.13	2.00	3.14	**1.56**	2.69
Aug	0.60	1.73	2.58	4.86	**0.23**	1.37	1.20	2.34	0.30	1.44
Sep	0.77	1.91	2.54	4.81	1.40	2.54	**0.56**	1.69	1.23	2.37

MSDI involves determining the best-fit copula families based on monthly AIC and BIC values. Table 4 shows the monthly AIC and BIC values.

The minimum AIC values in Table 4 are highlighted as they indicate the best-fitting copula family for each month. Parametric MSDI values are calculated using two approaches: first, by selecting a best-fit copula family for the entire dataset, and second, by selecting the best-fit copula family for each month and applying these monthly copula fits to the same dataset. This distinction underscores the difference between using a single copula family across the entire dataset versus adapting the copula family monthly based on the best fit.

Figure 9 illustrates the calculated MSDI values at different time scales. While the MSDI values exhibit similar patterns and high correlation, discrepancies become more apparent at the extreme highs and lows, indicating that improper selection of copula families may lead to inaccuracies in determining drought characteristics.

Correlation coefficients between single variable indices and the MSDI are shown in Table 5.

The data suggests that the use of the MSDI for drought characterization and monitoring is highly reliable. As shown in Table 5, all correlation coefficients are above 0.70, with the majority exceeding 0.75, indicating that the developed MSDI indices exhibit substantial reliability compared to other drought indices. It is important to note that the correlation coefficients between single-variable indices range from 0.10 to 0.37 across all time scales. In contrast, the correlation coefficients between single-variable indices and the MSDI values are significantly higher, ranging from 0.75 to 0.81. This suggests that the relationship between single-variable indices is considerably weaker than the relationship between these indices and the MSDI values across all time scales. An improvement in correlation is observed when transitioning from single-variable indices to the MSDI, as noted by Yisehak and Zenebe (2020).

Drought characteristics

Drought characteristics for the calculated indices are assessed as explained in Sect. 2.6. For the sake of simplicity, this paper will focus on illustrating the drought characteristics of the $MSDI_{monthly}$, as detailed in Sect. 2.4.3, at 9-month time scale. In this study, a drought event is defined as MSDI values remaining below −0.8 for more than three consecutive months. A drought is considered fully recovered once MSDI values exceed −0.2 (Mo 2011). The period between full recovery and the end of the drought is referred to as the drought recovery period. Drought characteristics for the MSDI at 9-month scale are provided in Table 6.

The start date in Table 6 refers to the index of the data, with the understanding that MSDI values are calculated on a 9-month time scale and the water year begins in October. Therefore, an index of 0 corresponds to June. Since DroughtStats does not require specific start dates, it uses indices instead of actual dates when necessary.

The drought characteristics derived from the monthly MSDI at the 9-month time scale reveal important insights into the nature of drought events in the study area. The results indicate significant variability in drought severity, duration, and recovery. Notably, between 2000 and 2005, the region experienced the longest and most severe drought event, which can be classified as an extreme drought, with MSDI-9 values below −2. This highlights the severity of the drought conditions in the basin during this period. Another major drought event occurred between 1992 and 1996, during which two separate drought events were identified, both with similar severity. This suggests that the region experienced recurrent droughts within a relatively short time frame. While some drought events exhibit high

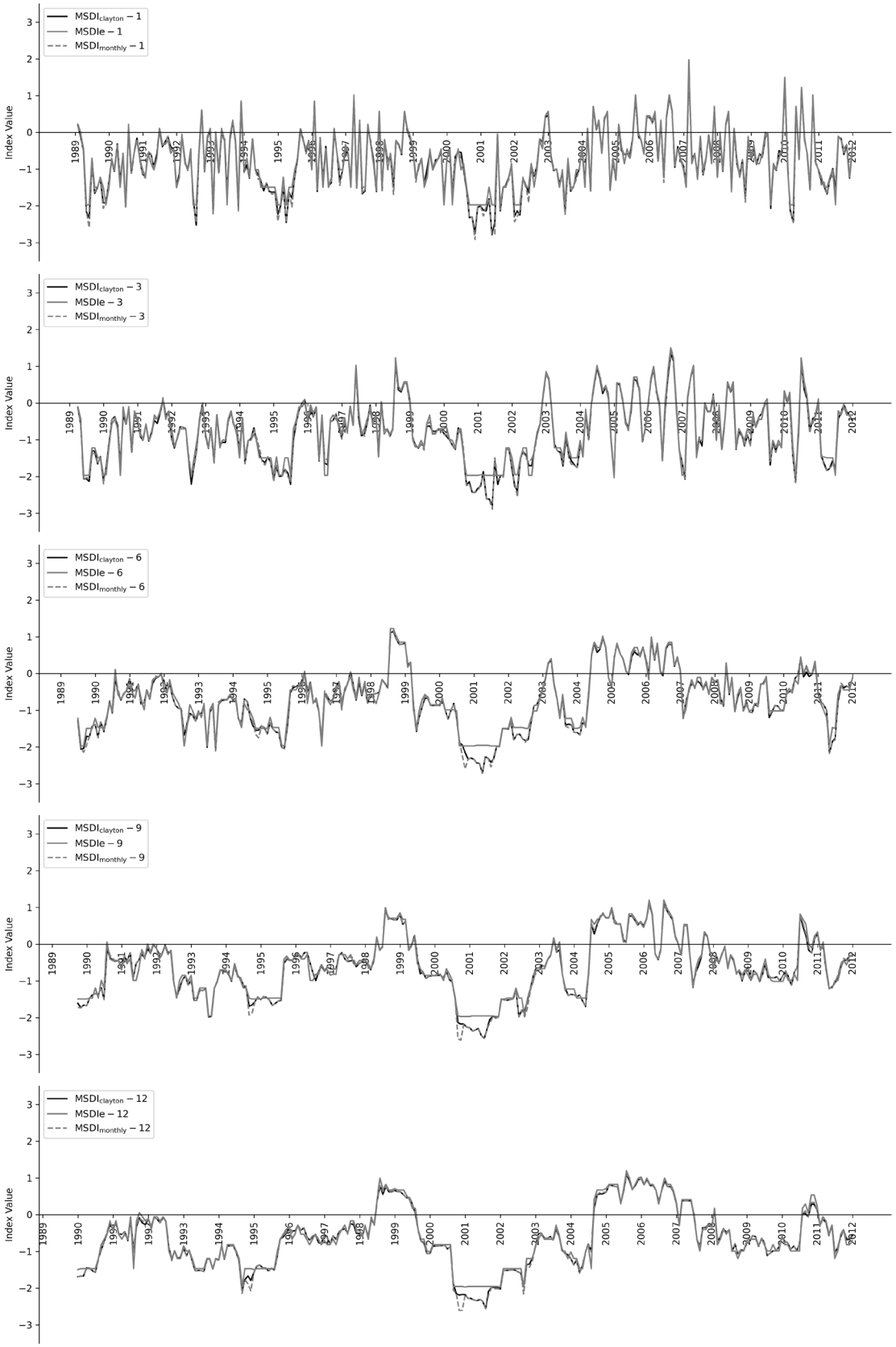

Fig. 9 Calculated MSDI values

Table 5 Correlation coefficients between single variable indices and the MSDI

Correlation	MSDIe	MSDIp Gaussian	MSDIp t	MSDIp Clayton	MSDIp Gumbel	MSDIp Frank	MSDIp Monthly
SSFI	0.76	0.77	0.77	0.77	0.77	0.77	0.77
SSFIe	0.75	0.78	0.77	0.77	0.78	0.78	0.77
SPI	0.77	0.81	0.80	0.80	0.82	0.81	0.80
SPIe	0.77	0.79	0.78	0.78	0.79	0.79	0.78

Table 6 Drought characteristics of monthly MSDIp at 9-month scale

Start Date	Duration (month)	Severity	Frequency of Occurrence	Recovery Duration (month)
0	10	−15.08	3.73	No recovery
33	16	−20.69	5.97	No recovery
55	16	−23.20	5.97	No recovery
120	7	−6.15	2.61	No recovery
128	30	−56.35	11.19	7
167	9	−12.31	3.36	2
234	3	−2.64	1.12	No recovery
238	6	−5.44	2.24	No recovery
259	4	−4.29	1.49	No recovery

severity and prolonged durations, others are less intense but more frequent. Many drought events do not show full recovery within the analyzed period, indicating persistent or recurring drought conditions. These findings highlight the complexities of drought dynamics, underscoring the need for long-term monitoring and effective management strategies to address the challenges posed by ongoing and future drought risks.

Discussion

This study introduces DroughtStats, a user-friendly software tool designed to enhance drought analysis and monitoring. By addressing limitations in existing tools, such as reliance on single-variable indices, fixed probability distributions, and limited flexibility, DroughtStats provides a versatile framework for analyzing hydrometeorological data in drought analysis. It supports the calculation of single-variable and multivariable drought indices using both parametric and nonparametric methods, offering improved accuracy in characterizing droughts across diverse contexts.

The case study conducted in Turkey's Çoruh River Basin (CRB) highlights the effectiveness of DroughtStats. Calculations of SPI and SSFI indices across multiple time scales demonstrated the software's ability to capture short-term and long-term drought events. Additionally, the calculated MSDI values revealed strong correlations with single-variable indices while providing a more integrated perspective of drought conditions. This comprehensive approach underscores the importance of considering interactions between multiple hydrometeorological variables, such as precipitation and streamflow, to achieve a nuanced understanding of drought dynamics.

The drought characteristics analysis identified key patterns in the CRB. The longest and most severe drought event, classified as an extreme drought event, occurred between 2000 and 2004, with MSDI-9 values exceeding −2. Another major observation was the presence of two distinct drought events between 1992 and 1994, each with similar severity. These findings reflect DroughtStats' ability to accurately identify and distinguish consecutive drought events. Importantly, these results are consistent with those of Simsek et al. (2024), who identified 1992–1996 and 2000–2004 as significant drought periods in the CRB using SPI at meteorological station 17089, which is located near station E2304 used in this study. This consistency validates DroughtStats' capability to capture meteorological drought events with SPI and to extend the analysis to multivariable droughts using MSDI, which integrates both meteorological and hydrological data.

Future improvements to DroughtStats could include the integration of additional drought indices, such as trivariate drought indices, to offer a more comprehensive understanding of drought conditions. The implementation of frameworks for drought propagation analysis would further enhance its capability to trace the evolution of drought across different sectors. Additionally, incorporating frequency analysis and developing Intensity–Duration–Frequency (IDF) curves would broaden its applicability, particularly in drought risk assessment and management strategies.

In conclusion, DroughtStats represents a significant advancement in drought analysis and monitoring. Its integration of single-variable and multivariable indices, coupled with advanced statistical tools, offers a robust and flexible framework for assessing drought characteristics. These capabilities, alongside its user-friendly interface, make DroughtStats a valuable tool for improving drought characterization and decision-making processes.

Conclusion

This study introduces DroughtStats, a comprehensive software tool designed to enhance drought analysis by overcoming the limitations of existing tools. It offers flexibility by integrating both single-variable and multivariable drought indices, utilizing copula-based methods, and providing a broader range of probability distributions. DroughtStats enables detailed and accurate analysis of drought conditions at global and regional scales, making it accessible for users at all expertise levels.

The case study conducted in Turkey's Çoruh River Basin demonstrated DroughtStats' effectiveness in capturing drought events across various time scales, with the software successfully identifying and characterizing drought severity, frequency, and recovery. This provides valuable insights for drought monitoring and management.

In conclusion, DroughtStats represents a significant advancement in the field of drought analysis, providing a versatile, user-friendly, and reliable tool that supports more precise drought monitoring, ultimately contributing to more informed decision-making and better management strategies in the face of increasing drought risks.

Authors contribution T.B.T. wrote the main manuscript text, developed the software, and contributed to the methodology, conceptualization, and data curation. B.Ö. provided supervision, resources, and validation. Both authors reviewed and edited the manuscript.

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK).

Data availability The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethical approval Not applicable.

Consent for publication Not applicable.

Conflict of interest The authors declare no competing interests.

References

AghaKouchak A, Cheng L, Mazdiyasni O, Farahmand A (2014) Global warming and changes in risk of concurrent climate extremes: Insights from the 2014 California drought. Geophys Res Lett 41:8847–8852. https://doi.org/10.1002/2014gl062308

AghaKouchak A, Farahmand A, Melton FS, Teixeira J, Anderson MC, Wardlow BD, Hain CR (2015) Remote sensing of drought: Progress, challenges and opportunities. Rev Geophys 53:452–480. https://doi.org/10.1002/2014rg000456

Akaike H (1998) Information Theory and an Extension of the Maximum Likelihood Principle. Springer Series in Statistics 199–213. https://doi.org/10.1007/978-1-4612-1694-0_15

Alessi MJ, Herrera DA, Evans CP, DeGaetano AT, Ault TR (2022) Soil Moisture Conditions Determine Land-Atmosphere Coupling and Drought Risk in the Northeastern United States. JGR Atmos 127. https://doi.org/10.1029/2021jd034740

Alexandersson H (1986) A homogeneity test applied to precipitation data. J Climatol 6:661–675. https://doi.org/10.1002/joc.3370060607

Allen, RG, Pereira, LS, Raes, D, Smith, M (1998) Crop evapotranspiration: Guidelines for computing crop water requirements. FAO Irrigation and Drainage Paper No. 56. Food and Agriculture Organization of the United Nations (FAO), Rome

Ballarin AS, Barros GL, Cabrera MCM, Wendland EC (2021) A copula-based drought assessment framework considering global simulation models. J Hydrol Reg Stud 38:100970. https://doi.org/10.1016/j.ejrh.2021.100970

Beguería, S, Vicente-Serrano, SM (2023) SPEI: Calculation of the Standardized Precipitation-Evapotranspiration Index. https://spei.csic.es. Accessed 6 December 2024

Bloomfield JP, Marchant BP (2013) Analysis of groundwater drought building on the standardised precipitation index approach. Hydrol Earth Syst Sci 17:4769–4787. https://doi.org/10.5194/hess-17-4769-2013

Buishand TA (1982) Some methods for testing the homogeneity of rainfall records. J Hydrol 58:11–27. https://doi.org/10.1016/0022-1694(82)90066-x

European Commission: Joint Research Centre, Cammalleri, C, Naumann, G, Mentaschi, L, Formetta, G, Forzieri, G, Gosling, S, Bisselink, B, De Roo, A, Feyen, L (2020) Global warming and drought impacts in the EU: JRC PESETA IV project: Task 7. Publications Office. https://doi.org/10.2760/597045

Cavus Y, Stahl K, Aksoy H (2023) Drought intensity–duration–frequency curves based on deficit in precipitation and streamflow for water resources management. Hydrol Earth Syst Sci 27:3427–3445. https://doi.org/10.5194/hess-27-3427-2023

Clayton DG (1978) A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biom 65:141–151. https://doi.org/10.1093/biomet/65.1.141

Farahmand A, AghaKouchak A (2015) A generalized framework for deriving nonparametric standardized drought indicators. Adv Water Res 76:140–145. https://doi.org/10.1016/j.advwatres.2014.11.012

Gibbs WJ, Maher JV (1967) Rainfall deciles as drought indicators. Bureau of Meteorology Bulletin No. 48. Commonwealth of Australia, Melbourne

Gringorten II (1963) A plotting rule for extreme probability paper. J Geophys Res 68:813–814. https://doi.org/10.1029/jz068i003p00813

Guillory L, Pudmenzky C, Nguyen-Huy T, Cobon D, Stone R (2023) A drought monitor for Australia. Environ Model Softw 170:105852. https://doi.org/10.1016/j.envsoft.2023.105852

Gumus V (2023) Evaluating the effect of the SPI and SPEI methods on drought monitoring over Turkey. J Hydrol 626:130386. https://doi.org/10.1016/j.jhydrol.2023.130386
Hamed KH, Ramachandra Rao A (1998) A modified Mann-Kendall trend test for autocorrelated data. J Hydrol 204:182–196. https://doi.org/10.1016/j.envsoft.2024.106185
Hao Z, AghaKouchak A (2013) Multivariate Standardized Drought Index: A parametric multi-index model. Adv Water Res 57:12–18. https://doi.org/10.1016/j.advwatres.2013.03.009
Hao Z, AghaKouchak A (2014) A Nonparametric Multivariate Multi-Index Drought Monitoring Framework. J Hydrometeorol 15:89–101. https://doi.org/10.1175/jhm-d-12-0160.1
Hao Z, Singh VP (2015) Drought characterization from a multivariate perspective: A review. J Hydrol 527:668–678. https://doi.org/10.1016/j.jhydrol.2015.05.031
Hao Z, Hao F, Singh VP, Ouyang W, Cheng H (2017) An integrated package for drought monitoring, prediction and analysis to aid drought modeling and assessment. Environ Model Softw 91:199–209. https://doi.org/10.1016/j.envsoft.2017.02.008
Hirsch RM, Slack JR, Smith RA (1982) Techniques of trend analysis for monthly water quality data. Water Resour Res 18:107–121. https://doi.org/10.1029/wr018i001p00107
Hussain Md, Mahmud I (2019) pyMannKendall: a python package for non parametric Mann Kendall family of trend tests. JOSS 4:1556. https://doi.org/10.21105/joss.01556
IPCC (2014) Climate change 2014: synthesis report. Contribution of Working Groups I, II, and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Pachauri RK, Meyer LA (eds). IPCC, Geneva
IPCC (2023) Summary for policymakers. In: Lee H, Romero J (eds) Climate change 2023: synthesis report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, pp 1–34. IPCC, Geneva. https://doi.org/10.59327/IPCC/AR6-9789291691647.001
Jarušková D (1996) Change-Point Detection in Meteorological Measurement. Mon Wea Rev 124:1535–1543. https://doi.org/10.1175/1520-0493(1996)124/3C1535:CPDIMM/3E2.0.CO;2
Kendall MG (1938) A new measure of rank correlation. Biom 30:81–93. https://doi.org/10.1093/biomet/30.1-2.81
Kendall MG (1955) Rank correlation methods. Charles Griffin, London
Leng S, Huete A, Cleverly J, Gao S, Yu Q, Meng X, Qi J, Zhang R, Wang Q (2022) Assessing the Impact of Extreme Droughts on Dryland Vegetation by Multi-Satellite Solar-Induced Chlorophyll Fluorescence. Remote Sens 14:1581. https://doi.org/10.3390/rs14071581
Li Y, Gong Y, Huang C (2021) Construction of combined drought index based on bivariate joint distribution. Alexandria Eng J 60:2825–2833. https://doi.org/10.1016/j.aej.2021.01.006
Li P, Huang Q, Huang S, Leng G, Peng J, Wang H, Zheng X, Li Y, Fang W (2022) Various maize yield losses and their dynamics triggered by drought thresholds based on Copula-Bayesian conditional probabilities. Agric Water Manag 261:107391. https://doi.org/10.1016/j.agwat.2021.107391
Liu C, Yang C, Yang Q, Wang J (2021) Spatiotemporal drought analysis by the standardized precipitation index (SPI) and standardized precipitation evapotranspiration index (SPEI) in Sichuan Province, China. Sci Rep 11. https://doi.org/10.1038/s41598-020-80527-3
Lorenzo MN, Pereira H, Alvarez I, Dias JM (2024) Standardized Precipitation Index (SPI) evolution over the Iberian Peninsula during the 21st century. Atmos Res 297:107132. https://doi.org/10.1016/j.atmosres.2023.107132
Mann HB (1945) Nonparametric Tests Against Trend. Econom 13:245. https://doi.org/10.2307/1907187
Massey FJ Jr (1951) The Kolmogorov-Smirnov Test for Goodness of Fit. J Am Stat Assoc 46:68–78. https://doi.org/10.1080/01621459.1951.10500769
McKee TB, Doesken NJ, Kleist J (1993) The relationship of drought frequency and duration to time scales. 8th Conference on Applied Climatology, Anaheim, pp 179–184.
Mishra AK, Singh VP (2010) A review of drought concepts. J Hydrol 391:202–216. https://doi.org/10.1016/j.jhydrol.2010.07.012
Mishra AK, Singh VP (2011) Drought modeling – A review. J Hydrol 403:157–175. https://doi.org/10.1016/j.jhydrol.2011.03.049
Mo KC (2011) Drought onset and recovery over the United States. J Geophys Res 116. https://doi.org/10.1029/2011jd016168
Mukherjee S, Mishra A, Trenberth KE (2018) Climate Change and Drought: a Perspective on Drought Indices. Curr Clim Change Rep 4:145–163. https://doi.org/10.1007/s40641-018-0098-x
Nalbantis I, Tsakiris G (2008) Assessment of Hydrological Drought Revisited. Water Resour Manage 23:881–897. https://doi.org/10.1007/s11269-008-9305-1
National Drought Mitigation Center (2024) Program to calculate standardized precipitation index. https://drought.unl.edu/monitoring/SPI/SPIProgram.aspx Accessed 6 December 2024
Nelsen RB (2006) An introduction to copulas. Springer, New York
Önöz B, Bayazıt M (2003) The power of statistical tests for trend detection. Turk J Eng Environ Sci 27(4):247–251
Palmer WC (1965) Meteorological drought (Research Paper No. 45). U.S. Weather Bureau.
Pettitt AN (1979) A non-parametric approach to the change-point problem. Appl Stat 28:126. https://doi.org/10.2307/2346729
Schwarz G (1978) Estimating the dimension of a model. Ann Stat 6(2):461–464. http://www.jstor.org/stable/2958889 Accessed 6 December 2024
Schweizer B, Wolff EF (1981) On nonparametric measures of dependence for random variables. Ann Stat 9(4):879–885
Sen PK (1968) Estimates of the Regression Coefficient Based on Kendall's Tau. J Am Stat Assoc 63:1379–1389. https://doi.org/10.1080/01621459.1968.10480934
Shannon CE (1948) A Mathematical Theory of Communication. Bell Syst Tech J 27:379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x
Shukla S, Wood AW (2008) Use of a standardized runoff index for characterizing hydrologic drought. Geophys Res Lett 35. https://doi.org/10.1029/2007gl032487
Şimşek O, Soydan Oksal NG, Uncu EM, Gümüş V, Şeker M (2024) SYİ Yöntemiyle Çoruh Havzası Uzun Dönem (1969–2020) Meteorolojik Kuraklığının Analizi. Politeknik Dergisi 27:1553–1564. https://doi.org/10.2339/politeknik.1295259
Spearman C (1904) The proof and measurement of association between two things. Am J Psychol 15(1):72–101. https://doi.org/10.2307/1412159
Sun P, Liu R, Yao R, Shen H, Bian Y (2023) Responses of agricultural drought to meteorological drought under different climatic zones and vegetation types. J Hydrol 619:129305. https://doi.org/10.1016/j.jhydrol.2023.129305
Székely GJ, Rizzo ML, Bakirov NK (2007) Measuring and testing dependence by correlation of distances. Ann Stat 35(6). https://doi.org/10.1214/009053607000000505
Tan X, Wu X, Huang Z, Fu J, Tan X, Deng S, Liu Y, Gan TY, Liu B (2023) Increasing global precipitation whiplash due to anthropogenic greenhouse gas emissions. Nat Commun 14. https://doi.org/10.1038/s41467-023-38510-9
Terzi TB, Önöz B (2023) Drought analysis in the Coruh River Basin based on standardized streamflow index. In: Özkök A, Kekeçoğlu M (eds) Proceedings of the 8th International Congress on Environmental Research and Technology, pp 22–29.
Terzi TB, Önöz B (2024a) Drought analysis in the Seyhan River Basin based on standardized drought indices using a new approach considering seasonality. Environ Earth Sci 84. https://doi.org/10.1007/s12665-024-12039-6

Terzi TB, Önöz B (2024b) Advanced drought analysis using a novel copula-based multivariate index: A case study of the Ceyhan River Basin. Manuscript submitted for publication, Sustain Water Resour Manag

Terzi TB (2024) Çoruh, Seyhan ve Ceyhan havzalarında kuraklık analizi / Drought analysis in Çoruh, Seyhan and Ceyhan river basins. Master's thesis, Istanbul Technical University

Tian Q, Wang F, Tian Y, Jiang Y, Weng P, Li J (2023) Copula-based comprehensive drought identification and evaluation over the Xijiang River Basin in South China. Ecol Indic 154:110503. https://doi.org/10.1016/j.ecolind.2023.110503

Tigkas D, Vangelis H, Tsakiris G (2014) DrinC: a software for drought analysis based on drought indices. Earth Sci Inform 8:697–709. https://doi.org/10.1007/s12145-014-0178-y

Tigkas D, Vangelis H, Proutsos N, Tsakiris G (2022) Incorporating aSPI and eRDI in Drought Indices Calculator (DrinC) Software for agricultural drought characterisation and monitoring. Hydrol 9:100. https://doi.org/10.3390/hydrology9060100

Tijdeman E, Stahl K, Tallaksen LM (2020) Drought characteristics derived based on the standardized streamflow index: A large sample comparison for parametric and nonparametric methods. Water Resour Res 56. https://doi.org/10.1029/2019wr026315

Tsakiris G, Vangelis H (2005) Establishing a drought index incorporating evapotranspiration. Eur Water 9(10):3–11

Van Loon AF (2015) Hydrological drought explained. WIREs. Water 2:359–392. https://doi.org/10.1002/wat2.1085

Varol T, Atesoglu A, Ozel HB, Cetin M (2023) Copula-based multivariate standardized drought index (MSDI) and length, severity, and frequency of hydrological drought in the Upper Sakarya Basin, Turkey. Nat Hazards 116:3669–3683. https://doi.org/10.1007/s11069-023-05830-4

Vicente-Serrano SM, Beguería S, López-Moreno JI (2010) A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index. J Clim 23:1696–1718. https://doi.org/10.1175/2009jcli2909.1

von Neumann J (1941) Distribution of the Ratio of the Mean Square Successive Difference to the Variance. Ann Math Statist 12:367–395. https://doi.org/10.1214/aoms/1177731677

Vorobevskii I, Kronenberg R, Bernhofer C (2022) Linking different drought types in a small catchment from a statistical perspective – Case study of the Wernersbach catchment. Germany J Hydrol X 15:100122. https://doi.org/10.1016/j.hydroa.2022.100122

Walker DW, Van Loon AF (2023) Droughts are coming on faster. Science 380:130–132. https://doi.org/10.1126/science.adh3097

Wang W, Yang H, Huang S, Wang Z, Liang Q, Chen S (2024) Trivariate copula functions for constructing a comprehensive atmosphere-land surface-hydrology drought index: A case study in the Yellow River basin. J Hydrol 642:131784. https://doi.org/10.1016/j.jhydrol.2024.131784

Wijngaard JB, Klein Tank AMG, Können GP (2003) Homogeneity of 20th century European daily temperature and precipitation series. Intl J Climatol 23:679–692. https://doi.org/10.1002/joc.906

Won J, Choi J, Lee O, Kim S (2020) Copula-based joint drought index using SPI and EDDI and its application to climate change. Sci Total Environ 744:140701. https://doi.org/10.1016/j.scitotenv.2020.140701

Xiang X, Li Y, Wu X, Cao H, Lin X (2024) Delineating the predominant impact of rising temperature on the enhancement of severity in compound drought-hot events in China: An empirical Copula and path analysis-based approach. J Hydrol Reg Stud 53:101769. https://doi.org/10.1016/j.ejrh.2024.101769

Xu Y, Wang L, Ross K, Liu C, Berry K (2018) Standardized soil moisture index for drought monitoring based on Soil Moisture Active Passive observations and 36 years of North American Land Data Assimilation System data: A case study in the Southeast United States. Remote Sens 10:301. https://doi.org/10.3390/rs10020301

Yevjevich V (1967) An objective approach to definitions and investigations of continental hydrologic droughts. Hydrology Paper No. 23. Colorado State University, Fort Collins

Yisehak B, Zenebe A (2020) Modeling multivariate standardized drought index based on the drought information from precipitation and runoff: a case study of Hare watershed of Southern Ethiopian Rift Valley Basin. Model Earth Syst Environ 7:1005–1017. https://doi.org/10.1007/s40808-020-00923-6

Zavareh MMJ, Mahjouri N, Rahimzadegan M, Rahimpour M (2023) A drought index based on groundwater quantity and quality: Application of multivariate copula analysis. J Clean Prod 417:137959. https://doi.org/10.1016/j.jclepro.2023.137959

Zeng J, Zhou T, Qu Y, Bento VA, Qi J, Xu Y, Li Y, Wang Q (2023) An improved global vegetation health index dataset in detecting vegetation drought. Sci Data 10. https://doi.org/10.1038/s41597-023-02255-3

Zhang Y, Huang S, Huang Q, Leng G, Wang H, Wang L (2019) Assessment of drought evolution characteristics based on a nonparametric and trivariate integrated drought index. J Hydrol 579:124230. https://doi.org/10.1016/j.jhydrol.2019.124230

Zhang X, Hao Z, Singh VP, Zhang Y, Feng S, Xu Y, Hao F (2022) Drought propagation under global warming: Characteristics, approaches, processes, and controlling factors. Sci Total Environ 838:156021. https://doi.org/10.1016/j.scitotenv.2022.156021

Zhang W, Feng K, Wang F, Wang W, Zhang Z, Wang Y, Huang S (2023) Dynamic Evolution and Copula-Based Multivariable Frequency Analysis of Meteorological Drought Considering the Spatiotemporal Variability in Northwestern China. Water 15:3861. https://doi.org/10.3390/w15213861