

Water Resources Research



RESEARCH ARTICLE

10.1029/2019WR026315

Key Points:

- The Standardized Streamflow Index (SSI) is a probability-based index used to express the streamflow drought hazard as an anomaly
- We show a sensitivity of drought characteristics to the method used to derive the SSI as well as to low- and high-end sample properties
- We evaluate the suitability of different methods to derive the SSI and discuss implications for drought assessments and operational use

Correspondence to:

E. Tjeldeman,
erik.tjeldeman@uni-heidelberg.de

Citation:

Tjeldeman, E., Stahl, K., & Tallaksen, L. M. (2020). Drought characteristics derived based on the Standardized Streamflow Index: A large sample comparison for parametric and nonparametric methods. *Water Resources Research*, 56, e2019WR026315. <https://doi.org/10.1029/2019WR026315>

Received 9 SEP 2019

Accepted 13 SEP 2020

Accepted article online 17 SEP 2020

Drought Characteristics Derived Based on the Standardized Streamflow Index: A Large Sample Comparison for Parametric and Nonparametric Methods

E. Tjeldeman¹ , K. Stahl² , and L. M. Tallaksen³

¹Institute of Geography, Professorship of Hydrology and Climatology, Heidelberg University, Heidelberg, Germany,

²Faculty of Environment and Natural Resources, University of Freiburg, Freiburg, Germany, ³Department of Geosciences, University of Oslo, Oslo, Norway

Abstract The streamflow drought hazard can be characterized in a variety of ways, including using different indices. Traditionally, percentile-based indices, such as Q95 (the flow exceeded 95% of time), have been used by the hydrological community. Recently, the use of anomaly indices such as the Standardized Streamflow Index (SSI), a probability index-based approach adopted from the climatological community, has increased in popularity. The SSI can be calculated based on various (non)parametric methods. Up to now, there is no consensus which method to use. This study aims to raise awareness how the inherent sensitivity of the SSI to the used method influences derived drought characteristics. We compared SSI time series computed with seven different probability distributions and two fitting methods as well as with different nonparametric methods for 369 rivers across Europe. Results showed that SSI time series and associated drought characteristics are indeed sensitive to the method of choice. A resampling experiment demonstrated the sensitivity of the parametric SSI to properties of both the low and high end of the sample. Such sensitivities might hinder a fair comparison of drought in space and time and highlight the need for a clear recommendation which method to use. We could recommend overall suitable methods, for example, from the parametric approaches, the Tweedie distribution has several advantageous properties such as a low rejection rate (2%) and a lower bound at zero. However, the most suitable method depends on the used evaluation criteria. Rather, we stress that shown approach-specific sensitivities and uncertainties should be carefully considered.

1. Introduction

1.1. Drought Definition and Identification

Droughts pose a threat in virtually all climate zones around the world (Mishra & Singh, 2010). They can be defined as a “sustained and regionally extensive occurrence of below average natural water availability” (Tallaksen & Van Lanen, 2004), where the average depends on the location and time of the year. Drought is thus not to be confused with aridity, which refers to permanently dry conditions (Wilhite, 1992). In addition, streamflow droughts are distinct from low flows, which are commonly considered as a seasonally re-occurring feature (e.g., Smakhtin, 2001). Deviations from the normal situation are often characterized by drought indices that express the anomaly of a given hydroclimatological variable with respect to a reference period. Such drought indices are also used in drought monitoring and early warning systems to depict regions at drought risk. Furthermore, hydrometeorological drought indices are used to compare the characteristics of historical drought events over time (e.g., Vicente-Serrano et al., 2014), between regions (e.g., Andreadis et al., 2005; Ionita et al., 2017; Sheffield et al., 2009) or across climates (e.g., Van Loon et al., 2014).

Hydrometeorological drought indices express anomalies in different ways. They can be defined based on absolute values or deviations, for example, the percentage deviation from a predefined threshold, such as annual average precipitation, streamflow, or any other variable. Alternatively, they can be defined in a relative way, expressing drought severity based on percentiles or ranks of historical data, for example, the streamflow drought monitoring application of USGS WaterWatch (<https://waterwatch.usgs.gov>). There are different ways to express these relative drought indices. Options include the use of empirical percentiles of the variable of interest or the use of standardized drought indices. Empirical percentiles, such as

©2020. The Authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

streamflow percentiles, express the variable as a historical non-exceedance frequency within a certain period, whereas standardized drought indices, such as the Standardized Precipitation Index (SPI; McKee et al., 1993), express the variable as a non-exceedance probability. In the latter case, the variable is transferred to the probability space by fitting a parametric or nonparametric probability distribution to the variable of interest and then transforming the derived probabilities of the variable to the Standard Normal distribution with an assumed mean of zero and standard deviation of one. The probability of non-exceedance of a certain quantity of the variable of interest is then reflected by the number of standard deviations from zero. For example, a value of -1 indicates one standard deviation below zero, which has a corresponding non-exceedance probability of around 15.9%, whereas a value of -2 indicates two standard deviations below zero and has a corresponding non-exceedance probability of around 2.3%.

The use of standardized drought indices has increased in popularity over the past decades (Bachmair et al., 2016), initiated by the development of the well-known SPI (McKee et al., 1993). The SPI, a standardized anomaly index solely based on aggregated precipitation, is currently the most commonly used index in drought monitoring and early warning systems around the world (Bachmair et al., 2016). The popularity of the SPI is related to the fact that it allows for a fair and consistent comparison of drought over space and time as compared with using absolute values, for example, precipitation in millimeter. One limitation of the SPI is that it only looks at anomalies in precipitation and thus ignores other meteorological variables and processes that are important in controlling the terrestrial water balance, such as evapotranspiration and snowmelt and accumulation. Recent research efforts have focused on including these processes in similar standardized meteorological drought indices, including the Standardized Precipitation and Evapotranspiration Index (SPEI, Vicente-Serrano et al., 2010) and the Standardized Melt and Rainfall Index (SMRI, Staudinger et al., 2014). The concept of standardizing drought indices has recently also become popular within the hydrological community for hydrological variables such as runoff (e.g., Shukla & Wood, 2008) and streamflow (e.g., Vicente-Serrano et al., 2012).

All indices described in the previous paragraphs can be calculated for different, user-preferred, accumulation periods, normally given in months. For standardized meteorological indices, such as the SPI, accumulation periods longer than 1 month are often used as a proxy to reflect longer-term memory within the hydrological cycle. For example, soil moisture deficits, important for, for example, agriculture, might be related to meteorological deficits of a few months (or shorter). On the other hand, streamflow deficits in catchments with large groundwater stores are related to long-term meteorological deficits of typically 12 or more months (e.g., Barker et al., 2016). Time series of soil moisture or streamflow already implicitly incorporate such time lags in precipitation as any value reflects what happened in a certain time window prior to the time of interest, where this time window represents the memory of the system. In such cases, the accumulation period should be chosen similar to the original or smoothed time series, which is rarely longer than 1 month.

The Standardized Streamflow Index (SSI) is used to characterize anomalies in observed streamflow (e.g., Modarres, 2007; Svensson et al., 2017; Vicente-Serrano et al., 2012; Zaidman et al., 2002). Hence, The SSI is not to be confused with the Standardized Runoff Index (SRI), which is used to reflect anomalies in modeled runoff per unit area (Shukla & Wood, 2008). The SSI is often calculated for a monthly accumulation period; however, some studies use longer accumulation periods, for example, 3–12 months, to track the cumulative water deficit over the hydrological year (e.g., Nalbantis & Tsakiris, 2009). According to a survey by Bachmair et al. (2016), the SSI is currently used operationally in some drought monitoring and early warning systems. Further, it has been named as a trigger in drought plans in Chile (Núñez et al., 2014). The SSI is the only recommended drought index for streamflow drought monitoring and early warning in the current version of the Handbook of Drought Indicators and Indices (WMO & GWP, 2016). Besides the use of the SSI for monitoring streamflow droughts, the SSI has been used to define characteristics of streamflow drought events in various studies (e.g., Barker et al., 2016; Lorenzo-Lacruz et al., 2010; Van Oel et al., 2018). The threshold level method, originally designed to be applied on raw streamflow data (Zelenhasi & Salvai, 1987), has been applied to SSI time series using a given SSI value as threshold (e.g., -0.84 corresponding to the commonly used 20 percentile threshold). Accordingly, drought characteristics, such as duration (the time the SSI time series is below the threshold), deficit volume (summed difference between SSI time series and threshold over the duration of a drought event), the number of SSI occurrences below a certain threshold, or monthly minimum SSI, can be derived.

1.2. Computation of the SSI

Nowadays, it is relatively easy to compute a standardized drought index given the availability of several software packages. However, accurate computation of the SSI, and any other standardized drought index, involves several additional steps beyond applying such a software package (Tallaksen & Van Lanen, 2004). First, a (set of) sample(s) needs to be chosen, including the selection of catchments, the period over which the sample should be standardized (observation period), and the period that serves as reference for this standardization (reference period). When deriving the SSI in a parametric way, the parameters of the population distribution, of which the sample is assumed to be a random subset, can be estimated by fitting a representative parametric distribution to the sample data. Accordingly, a parameter estimation method (or fitting method) as well as a parametric distribution need to be selected. In the case of a regional data set, it is important to choose the same reference period and preferably the same distribution and fitting method to allow a consistent comparison across catchments.

Three used fitting methods to derive the parametric SSI are product moments, L-moments (Lmom, Hosking, 1990) and maximum likelihood estimation (MLE). Each fitting method has its own advantages and disadvantages (Tallaksen & Van Lanen, 2004). The Lmom method is less sensitive to outliers than the product moment method (Tallaksen & Van Lanen, 2004). The MLE approach is the most theoretically sound and flexible distribution fitting method (Stagge et al., 2015; Tallaksen & Van Lanen, 2004), but it is less stable; that is, a solution cannot always be found. Furthermore, MLE can be computationally demanding. A combination of the MLE and Lmom approach was applied by Stagge et al. (2015) to combine the flexibility of the MLE approach with the robustness of the Lmom approach. In this case, the distribution properties derived with the Lmom approach are used as starting values for estimating the distribution properties using MLE. All the three above fitting methods have been applied for the SSI, for example, the Lmom method for a set of catchments in Spain (Vicente-Serrano et al., 2012), the MLE method for several case studies around the world (e.g., Van Oel et al., 2018; Wen et al., 2011; Zhang et al., 2014), and the combined MLE and Lmom approach for catchments in the United Kingdom (Svensson et al., 2017).

The choice of distribution to derive the parametric SSI, ranging from simple and robust two-parameter distributions to more complex, but flexible three- and four-parameter distribution (e.g., Svensson et al., 2017), should be justified by rigorous testing of which distribution fits the sample well, or which distribution has the lowest rejection rate for a set of samples. Several goodness of fit tests exist to identify the best fitting or least rejected distributions (e.g., Tallaksen & Van Lanen, 2004) and different studies have used different tests. Vicente-Serrano et al. (2012) tested six different three-parameter probability distributions to compute the SSI for a set of rivers across the Iberian Peninsula. They conclude that, if one had to choose one, the Generalized Extreme Value (GEV) or Generalized Logistic distribution was most suitable, one reason being that the observed number of extreme events showed the highest agreement with respect to the expected occurrence of extreme events. In other words, the derived SSI exhibits the characteristics expected from a probabilistic index. However, they recommend using the best fitting distribution for each catchment or calendar month, which might result in the use of different probability distributions. Svensson et al. (2017) tested a variety of two-, three-, and four-parameter distributions to fit to streamflow data for a river data set from the United Kingdom. They found that the rejection rate was generally lower for the three- and four-parameter probability distributions, which can be expected given their higher flexibility. However, commonly used three-parameter distributions, such as the Pearson-III or GEV distribution, are not bounded at zero, resulting in occasional probabilities of unrealistic negative streamflow values. To overcome these problems, they suggested using the Tweedie distribution, which has a lower bound at zero.

An alternative, but less often used, approach to derive standardized drought indices, is the use of a nonparametric method. Different nonparametric methods have been used in drought literature, including the transformation of plotting positions (PP) to the Standard Normal distribution (Farahmand & AghaKouchak, 2015) or the use of kernel density estimation (KDE, Vidal et al., 2010). PP are rank-based and thereby flexible to directly describe any sample at hand independent of its properties. Accordingly, PP can be used without testing the goodness of fit (Farahmand & AghaKouchak, 2015). Many different PP formulas exist, of which those of Weibull and Gringorten have been used to derive standardized drought indices (e.g., Farahmand & AghaKouchak, 2015; Sořáková et al., 2014). In addition, empirical PP, that is, the non-exceedance frequency of the flow duration curve, are commonly used in threshold-level based drought identification methods.

Further remarks on the suitability of these (and other) PP formulas for data stemming from different types of population distributions can be found in, for example, Cunnane (1978). KDE is a nonparametric smoothing procedure (Wilks, 2011), where smoothing refers to the estimation of a “smoothed” continuous density function for the sample at hand (as opposed to the histogram). The kernel density function, which can be used as estimate of the population distribution, is derived by summing up a set of single kernels drawn around each individual observation. As kernels are directly drawn around each observation, KDE is able to deal with more complex, for example, multimodal, distributions. KDE requires several methodological choices, including the selection of the kernel shape (e.g., Gaussian, rectangular, etc.) and an estimation of the kernel bandwidth, where the choice of bandwidth or bandwidth estimation method often has the largest effect on the estimated kernel density function (Wilks, 2011). Only a few studies have used KDE to derive standardized drought indices (including Kumar et al., 2016; Samaniego et al., 2012; Vergni et al., 2017; Vidal et al., 2010), of which only the study of Vidal et al. (2010) considered streamflow. The latter studies all used Gaussian-shaped kernels, of which the bandwidths were either estimated using unbiased cross validation (ucv; numerical optimization of the bandwidth via cross validation) or plug-in selectors (referring to the plugging in of estimates into the bandwidth estimation formula).

1.3. Limitations of the SSI

The use of the SSI has several limitations, especially when applied in a large-scale setting with diverse streamflow regimes. When using a parametric approach, there is seldom a parametric probability distribution that fits all monthly streamflow samples equally well; for example, sample properties of streamflow in January might be very different from those of August, and sample properties of a highly variable snow dominated streamflow regime might be very different from those of a stable groundwater-fed streamflow regime. Therefore, there is no universally recommended distribution and different studies use different, best fitting or least often rejected, distributions for their study catchment(s). One option is to use the best fitting parametric probability distribution for each calendar month and catchment (in the case of a regional assessment), as recommended by Vicente-Serrano et al. (2012). However, this might cause discrepancies in space and time because of distributional differences, potentially hindering a fair comparison between studies, catchments, and calendar months. Another option is to have the same, least often rejected, distribution for the entire data set as it assures that the properties of the parametric distribution are the same for each catchment and calendar month. However, limiting to one distribution might result in SSI values derived from poorly fitting distributions for some catchments and calendar months. Nonparametric approaches offer more flexibility, for example, PP directly describe the ranking of the sample at hand and KDE is based on kernels drawn around each individual observation. However, PP can be more susceptible to sampling uncertainty, especially in the tails of the distribution (e.g., Folland & Anderson, 2002), whereas KDE can be prone to over- or under-smoothing. Further, the above-described KDE approaches and especially PP are less suitable to assign probabilities to flows beyond the range spanned by the observations.

Even if there was a common agreement about the most suitable method to derive the SSI, there are inherited uncertainties due to the often relatively short sample and choice of reference period (i.e., the sample properties; Hong et al., 2015; Núñez et al., 2014). The properties of the probability distribution are sensitive to the sample data set (it being a single site or a regional data set) and whether it captures the natural variability of the time series; for example, it might stem from a rather wet or dry period or a period with extreme floods or droughts.

1.4. Aim and Outline

With the increasing popularity of the use of standardized probability indices, such as the SSI, for both drought research and drought monitoring and early warning, there is the need to raise awareness of how the characteristics of the SSI as well as the various methodological choices needed to compute the SSI affect the properties of SSI time series and drought characteristics derived from these time series. A thorough analysis of the sensitivity of the SSI, focusing on the most often used parametric approach, to the choice of probability distribution, fitting method as well as the sample properties and related uncertainties, is required. This involves:

- testing the sensitivity of the parametric SSI to the probability distribution, fitting method, and sample properties, both in the low and high end of the sample, and

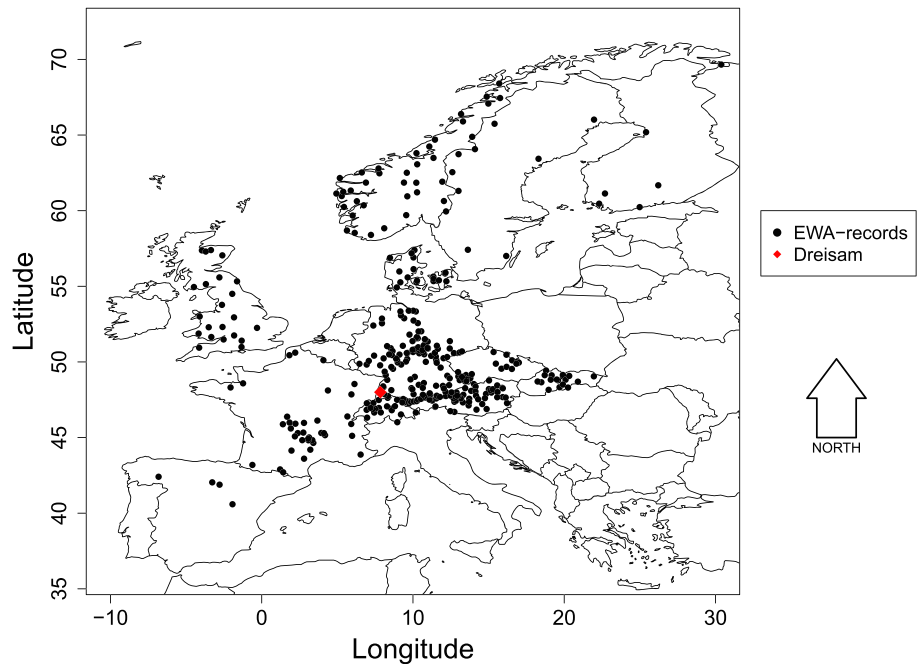


Figure 1. Location of gauging stations from the European Water Archive rivers (black dots) and the case study river Dreisam (red diamond).

- evaluating the implications for drought monitoring, reflected by differences in the exceedance of trigger levels, and drought characterization, reflected by differences in drought characteristics derived from SSI time series.

An overall assessment of the SSI benefits from the consideration of nonparametric methods. Therefore, we further compare the properties of the nonparametric SSI with those of the parametric SSI and discuss implications for drought monitoring and research. Finally, we discuss the applicability of the SSI obtained using both parametric and nonparametric methods as compared to empirical percentiles. Here, we specifically interpret empirical percentiles as indices describing the historical non-exceedance frequency within a certain period as opposed to the non-exceedance probability given by the SSI. This discussion reveals some points for consideration when choosing one approach over the other.

2. Data and Methods

2.1. Streamflow Data

Streamflow data stem from two different sources: (1) a case study, the river Dreisam (station Ebnet), and (2) a series of catchments across Europe with near-natural flow from the European Water Archive (EWA; Stahl et al., 2010). All data were analyzed in the same manner. The river Dreisam is located in southwestern Germany, whereas EWA catchments cover different climatic regions across Europe and reflect a variety of streamflow regimes (Figure 1). For all rivers, continuous (no gaps) daily streamflow series were accumulated into monthly average streamflow records covering the 40-year period 1970 to 2009 for the river Dreisam and 1965 to 2004 for rivers from the EWA. These two periods were also used as reference periods for the SSI computation. The number of zero-flow occurrences in the two data sets was limited; monthly average flow of the river Dreisam was above zero for all months and only six out of the 4,416 monthly flow series from the EWA (12 months times 368 catchments) had 1 month with zero flow. Thus, our results are not representative of arid regions or intermittent streams, for which other distributions or drought identification approaches may be more suited (an example of an approach to deal with zero values is presented in Stagge et al., 2015).

2.2. Parametric SSI Calculation

Parametric SSI time series were calculated by fitting a set of parametric probability distributions to streamflow samples for all catchments and months (12 monthly streamflow samples for each catchment). The

Table 1
SSI Values, Corresponding Expected Non-Exceedance Frequency (in Percent and Number of Events for a 40-Year Sample) and the Theoretical Return Period

SSI value	Expected non-exceedance frequency (percent)	Expected non-exceedance frequency for a 40-year sample	Theoretical return period (years, rounded)
−0.68	25	10	4
−0.84	20	8	5
−1.28	10	4	10
−1.96	2.5	1	40
−3	0.14	0	741

properties of the fitted parametric distributions for each calendar month were then used to transfer monthly streamflow values of the given calendar month to the probability space and finally to the Standard Normal distribution or SSI values (similar to the approach first described in McKee et al., 1993). In this study, seven probability distributions were considered, namely, the (1) Normal, (2) Log-normal, (3) Gamma, (4) Pearson-III, (5) GEV, (6) Generalized Logistic (Genlog), and (7) Tweedie distribution. Of these distributions, the Normal, Log-normal, and Gamma distribution have two parameters, whereas the remaining four distributions are three-parameter distributions. Only the Log-normal, Gamma, and Tweedie distributions are bounded below at zero. The distributions were selected based on their previous use for SSI calculation. The probability density functions of these distributions are given in, for example, Stagge et al. (2015) and Svensson et al. (2017). For the first six distributions, this study used two common approaches to estimate the parameters: the Lmom approach and the combined MLE and Lmom approach (Stagge et al., 2015), in the following referred to as just MLE for simplicity. The parameters of the Tweedie distribution were estimated following the approach described in Svensson et al. (2017). It partly deviates from the MLE approach, as the Tweedie distribution only has a closed form for some special cases. All other cases require numerical parameter estimation methods and the Lmom method could not be used to derive starting values for MLE. Instead, the numerical estimates of the parameters of the Tweedie distribution were used as starting values for MLE.

In total, 13 different parametric SSI time series were derived (two probability distribution fitting methods and seven parametric distributions for each catchment and calendar month). The parameters of the Tweedie distribution could only be estimated using MLE. Similar to Stagge et al. (2015), we placed a lower limit on SSI values; values smaller than -3 are associated with a non-exceedance probability of $\approx 0.14\%$ or a theoretical return period that exceeds 741 years (Table 1) and were retained at -3 due to the increasing uncertainty for these lower values (Stagge et al., 2016). The Shapiro-Wilk test (Shapiro & Wilk, 1965) was used to test if the calculated SSI for each catchment and calendar month followed the expected Normal distribution (using a test criterion of $p < 0.05$, similar to, e.g., Svensson et al., 2017). SSI series for certain catchments or calendar months were not considered for further analyses for p values lower than 0.05, unless specifically mentioned. Note that passing this test does not mean the data follow the Normal distribution perfectly; it only means that the null hypothesis of the Shapiro-Wilk test that the data follow the Normal distribution cannot be rejected.

The SSI was also calculated for bootstrapped streamflow samples (resampling with replacement), illustrating the sensitivity to sample properties and related uncertainties for the SSI. Two different resampling strategies were used:

1. full resampling (with replacement), where the parameters of the population distribution were estimated from 40 random samples of the original sample ($n = 40$), and
2. partial resampling, where the parameters of the population distribution were estimated from fixed below normal flow values (10 lowest values of the original sample) and 30 random samples of the 30 highest values of the original sample (with replacement).

Each resampling procedure was repeated 1,000 times. The parameters of the monthly flow distribution were derived for each subsample. These parameters of the resampled time series were then used to calculate the SSI for the original (not resampled) time series, that is, using the parameters of the resampled monthly flow distribution to estimate the non-exceedance probability (or SSI) of the original

observations. The spread in SSI for below normal flow derived from the distribution parameters of the fully resampled samples ($SSI_{full.res}$) indicates the uncertainty of the SSI. The spread in SSI values derived from the distribution parameters of the partially resampled samples ($SSI_{partial.res}$) shows the sensitivity of the SSI for below normal flow to changes in the normal- and high-end flow range.

2.3. Evaluation and Sensitivity of the Parametric SSI

The parametric SSI was evaluated based on its suitability to express the streamflow drought hazard in a probabilistic way. Two criteria were used:

1. goodness of fit according to the rejection rate and p value of the Shapiro-Wilk test and
2. whether the expected spread in minimum SSI values matches the observed spread.

With the first criterion, we aim to assess how well different parametric distributions fit the considered streamflow samples and thus result in SSI time series that closely resemble the expected Standard Normal distribution. With the second criterion, we aim to investigate whether the derived SSI time series exhibit the expected behavior from a probabilistic index. Given the probabilistic nature of the SSI, it can be expected that low SSI values occur for a certain percentage of samples, especially when considering a large set of samples. For instance, a binomial distribution can be used to calculate that an SSI value of -2.5 (return period of 161 years) or worse in a sample of length 40 has an occurrence probability of around 22% and is therefore expected to appear in 22% of the samples. In a similar way, the expected spread in minimum SSI, for example, the values that are expected to appear in 5–95% of the samples, can be derived.

The sensitivity of the parametric SSI to the used method was derived by comparing the spread among resulting SSI time series computed with

1. the same probability distribution, but with two different fitting methods;
2. the same fitting method, but with different probability distributions;
3. the same probability distribution and fitting method, across catchments and calendar months; and
4. different samples, that is, resampled SSI time series ($SSI_{full.res}$ and $SSI_{partial.res}$) to test the influence of sample properties.

For simplicity, we hereafter use the term “spread in SSI” when referring to spread in SSI for below normal flow anomalies, that is, in the dry range.

2.4. Implications for Drought Monitoring and Characterization

The number of SSI values below a given threshold was compared for the set of monthly parametric SSI time series obtained using different probability distributions, fitting methods, and samples. We used similar thresholds (TH) as in the streamflow drought monitoring application of the USGS WaterWatch (<https://waterwatch.usgs.gov>). This classification scheme is based on streamflow percentile categories “low” (defined in this study to be less than the 2.5th percentile), “much below normal” (below the 10th percentile), and “below normal” (below the 25th percentile), which correspond to SSI values of -1.96 (TH1), -1.28 (TH2), and -0.68 (TH3), respectively. The theoretical return period and expected historical non-exceedance frequency of these (and other) thresholds is presented in Table 1.

We further compared drought event characteristics—duration and (average) deficit volume—for SSIs derived with different parametric methods. For this comparison, SSIs calculated for each calendar month separately were merged to create continuous monthly time series from which drought event characteristics were extracted. We used an SSI threshold level of -0.84 , which corresponds to the 20th percentile threshold (Table 1), commonly used for streamflow drought analyses (e.g., Andreadis et al., 2005). The spread in drought characteristics derived from the set of SSI time series were assessed according to whether there were systematic differences in catchment average deficit volume, or not (referred to as unsystematic differences). Spearman’s rank correlation coefficient (Von Storch & Zwiers, 1999) was used as a measure of unsystematic differences (high = low degree of unsystematic differences, low = high degree of unsystematic differences).

2.5. Comparison With the Nonparametric SSI

Nonparametric SSI time series were derived using either PP of each monthly flow sample or KDE. We considered three different PP formulas based on their use in previous drought studies: Weibull (Equation 1), Gringorten (Equation 2), and empirical (Equation 3).

$$\text{rank}(m)/(n + 1) \quad (1)$$

$$\text{rank}(m - 0.44)/(n + 0.12) \quad (2)$$

$$\text{rank}(m)/n \quad (3)$$

where $\text{rank}(m)$ refers to the rank of each individual observation in the sample and n to the sample size; in this study, $n = 40$ (for an overview of different PP, see e.g., Cunnane, 1978). In case of ties in the sample, the average rank was assigned to the tied values; for example, when the two lowest values were the same, they were both assigned a rank of 1.5. The non-exceedance probabilities obtained using PP were transformed to the Standard Normal distribution in a similar way as was done for the parametric SSI (section 2.2).

KDE was done using a Gaussian kernel and two different bandwidth estimation methods used in previous studies to derive standardized drought indices: *ucv* and the direct plug-in method (Sheather & Jones, 1991; Venables & Ripley, 2002). KDE was applied on both original and \ln -transformed streamflow samples, where an \ln -transformation was applied to better handle often skewed monthly flow samples as well as to limit the assignment of probabilities to negative flow values. The integral of the derived kernel density functions was used to estimate non-exceedance probabilities of observed flow values, which were then transformed to the standard normal distribution. Next to the above-described KDE methods, there is a plethora of options to modify KDE by using, for example, different (adaptive) bandwidth estimation methods or different methods to impose a lower bound at zero. Here we limit our study to a few of the KDE options that have been considered in previous drought studies. A more thorough testing of the wide range of different KDE options that exist is beyond the scope of this study.

The nonparametric SSI was evaluated against the same criteria as the parametric SSI, that is, according to the goodness of fit and whether the derived minimum SSI values exhibit the spread expected from a probabilistic drought index (section 2.3). In addition, we derived the uncertainties of the nonparametric SSI derived with empirical PP and associated drought characteristics. A more comprehensive sensitivity analysis to the used nonparametric method, similar to the sensitivity analyses of the more often used parametric SSI, is outside the scope of the current study.

2.6. Empirical Streamflow Percentiles

We further calculated empirical streamflow percentiles, which express flow anomalies as the historical non-exceedance frequency within a given time period and not as a re-occurrence probability. The calculation of empirical streamflow percentiles is similar to the calculation of empirical PP (Equation 3), although their interpretation is notably different from the interpretation of the SSI. For instance, the interpretation that the flow in the current month of July was not exceeded for 20% of time in the past 40 years (empirical percentiles) is notably different from the interpretation that the flow in the current month of July has a non-exceedance probability of 20% (SSI).

3. Results

3.1. Evaluating the Parametric SSI

The fraction of parametric SSI time series rejected according to the Shapiro-Wilk test ($p < 0.05$) varied among probability distributions and distribution fitting methods (Figure 2a). As expected, time series calculated with less flexible two-parameter distributions were rejected more often than those calculated with three-parameter distributions. Of the three two-parameter distributions, the (symmetric) Normal distribution (not bounded at zero) shows the highest rejection rate (80% for both fitting methods), followed by the Gamma distribution (30% for Lmom and 29% for MLE) and the Log-normal distribution (17% for both methods). For SSI time series calculated with a three-parameter distribution, the Pearson-III distribution has a rejection rate of 24% using the Lmom approach and 7% using the MLE approach. Lowest rejection rates among parametric SSI time series were found for SSI time series derived based on the GEV distribution (5% for Lmom and 2% for MLE), Genlog distribution (9% for Lmom and 3% for MLE), and Tweedie distribution (2%). The general lower rejection rates of the MLE, as compared to the Lmom approach, are partially offset by the number of cases for which MLE did not converge to a

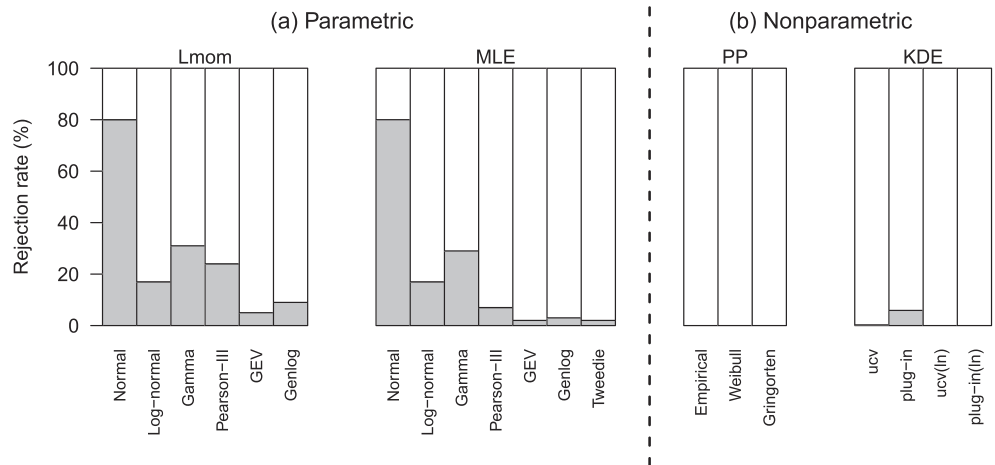


Figure 2. Rejection rate (across catchments and months) of different parametric (a) and nonparametric (b) SSI time series according to the Shapiro-Wilk test ($p < 0.05$).

solution for some distributions (not shown). The percentage of non-convergence was highest for the Pearson-III distribution (6%), notably lower for the GEV distribution (0.5%) and well below 0.5% for all other distributions.

Figure 3 shows the most suitable parametric distribution (derived with MLE) for all catchments of the EWA according to the rejection rate and p value of the Shapiro-Wilk test. A joint evaluation of the goodness of fit of all 12 flow samples of each catchment, based on the lowest number of rejected months and highest average p values (in case of ties in the lowest number of rejected months), reveals that the Tweedie distribution is overall most suitable for a large proportion of catchments, followed by the GEV, Genlog, and Pearson-III distributions (Figure 3a). The considered two-parameter distributions (Normal, Log-normal, and Gamma) were never labeled most suitable according to the overall assessment. A separate evaluation for each calendar month and catchment confirms that the Tweedie distribution is most suitable for a considerable proportion of the catchments across all calendar months (Figure 3b). However, a notable increase in variability is seen in the distribution of the most suitable parametric method when each calendar month is considered separately, and especially other three-parameter distributions were more often labeled as most suitable.

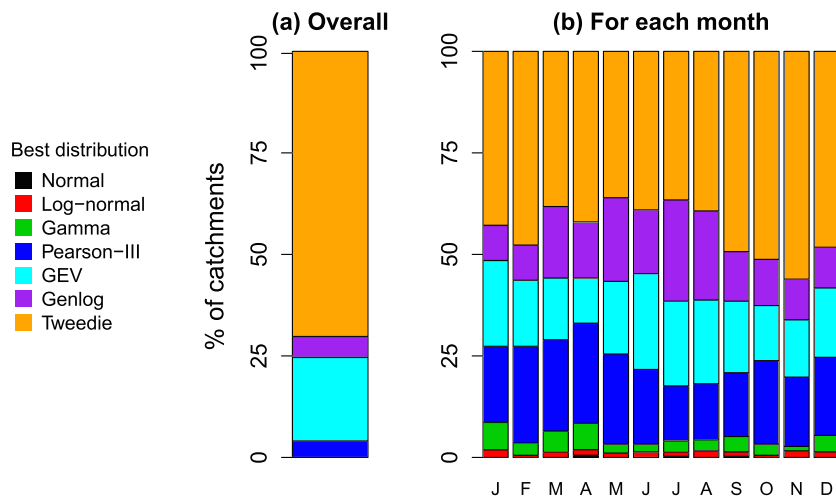


Figure 3. The best fitting distribution (using MLE) across catchments according to the rejection rate and p value of the Shapiro-Wilk test. Panel (a) shows the least often rejected distribution with the highest p value as a fraction (in percent) of catchments and (b) shows the distribution with the highest p value as a fraction of catchments for each calendar month.

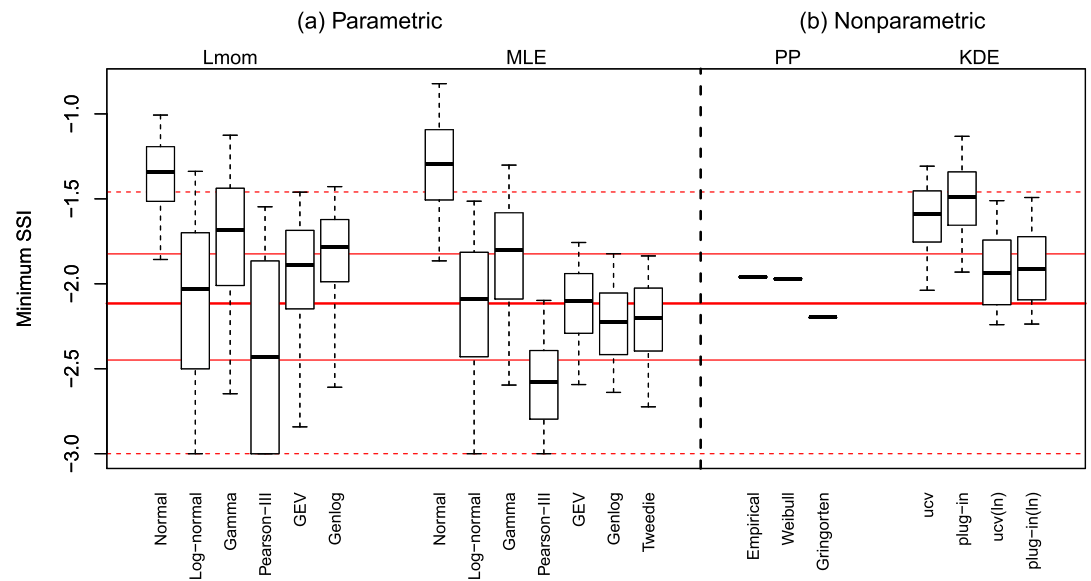


Figure 4. The spread in minimum SSI values (across all catchments and months) derived with different parametric (a) and nonparametric (b) approaches. Box: quantiles 25, 50, and 75. End of whiskers: quantiles 5 and 95. The expected spread is given for the same quantiles in red. Both rejected and non-rejected SSIs are considered.

Figure 4a shows the spread in minimum parametric SSI values across all monthly flow samples of the EWA catchments (in total >4,400 samples). A certain spread in minimum SSI values can be expected, given the large number of considered samples and the probabilistic nature of the SSI. This expected spread resembles most closely the SSI derived with the two-parameter Log-normal distribution (both MLE and Lmom). Some SSIs have a median minimum SSI value that is close to the expected median but have a lower than expected spread. This is the case for the parametric SSI derived with the GEV, Genlog, and Tweedie distributions (MLE). Compared to these distributions (MLE), the minimum SSI values derived with the GEV and Genlog distribution (Lmom) show a larger spread that more closely resembles the expected spread, but also a median that differs more from the expected median. Minimum SSI values from some other parametric distributions tend to either overestimate (Normal, Gamma) or underestimate (Pearson-III, especially when derived using MLE) minimum SSI values.

3.2. Sensitivity of the Parametric SSI

Figures 5a and 5c display the left tail of the cumulative parametric probability distribution functions fitted to average monthly streamflow as well as the empirical streamflow percentiles for the river Dreisam for the month July. The empirical distribution exhibits a typical stepwise feature; that is, flow for the lowest three percentile values is distinctly different from flow at higher percentile values. There is a clear spread among cumulative probability distributions and fitting methods, and none of the distributions matches the empirical data (sample) distribution particularly well. The spread is largest among the cumulative parametric probability functions when computed with the Lmom approach. This larger spread in distribution functions propagates to a larger spread in estimated SSI values of the lowest observed monthly flow (Figures 5b and 5d). Minimum SSI values range between -1.8 and -2.5 for the non-rejected probability distributions fitted with the Lmom approach and between -2 and -2.6 for the non-rejected distributions fitted with the MLE approach. Subsequently, this implies that the lowest threshold (TH1) may or may not be exceeded depending on a given combination of distribution and fitting method.

Figure 6 summarizes the spread (90% range) in non-rejected SSI values computed with the same probability distribution and fitting method across all catchments and calendar months. Although this spread is expected, given the probabilistic nature of the SSI (Figure 4, section 3.1), it might also be an aspect of standardized drought indices some end users are not aware of, as it contradicts in some way with “comparability in space and time.” In general, the spread is largest for the lowest streamflow percentiles. For these lowest

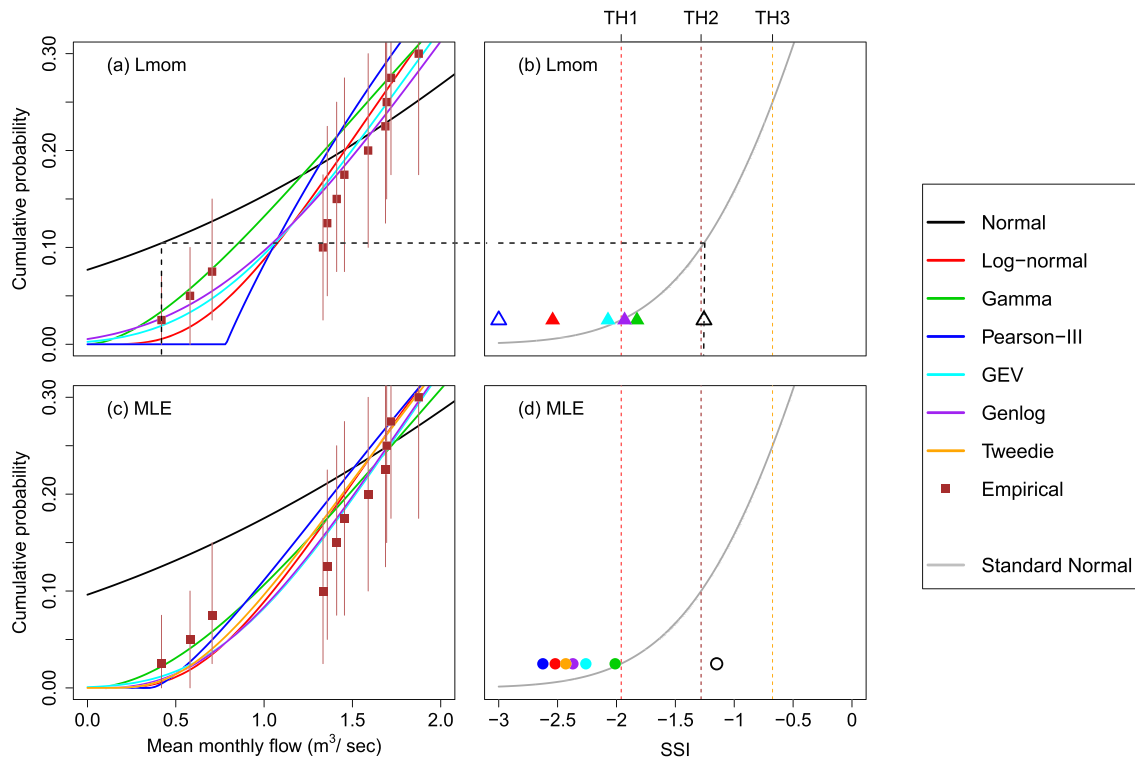


Figure 5. (a, c) Different cumulative probability distribution functions fitted to monthly average streamflow of the river Dreisam (July) computed with the Lmom (a) and MLE (c) fitting method. Dashed black line (a, b) visualizes the transformation from a monthly streamflow value to its corresponding SSI value according to the Normal distribution (Lmom). Empirical probabilities and their 90% uncertainty range are shown in brown. Corresponding minimum SSI values are shown in (b, d) where solid symbols show minimum SSI values derived from non-rejected SSI time series and open symbols values from rejected SSI time series (Shapiro-Wilk test: $p < 0.05$).

streamflow percentiles, SSI values range between -3 to -1.4 among catchments and calendar months. The spread in SSI further reveals that the same percentile value for different catchments and calendar months will be classified differently according to a given threshold level.

In Figures 7a, 7b, 7d, and 7e, the sensitivity in parametric SSI time series to the sample properties is shown for both the full and the partial resampling experiment. For brevity reasons, the results of the resampling experiments are only shown for the case study river Dreisam (July) using the two parametric distributions that were least often rejected (GEV and Tweedie, both estimated with the MLE approach). The 90% uncertainty range of the cumulative distribution function and associated range in $SSI_{full.res}$ is relatively wide (light blue band). Thus, whether or not the SSI falls below a certain threshold as well as how extreme river flow drought conditions are classified will vary due to sample uncertainty. For example, the lowest three events can be characterized as events with a very low non-exceedance probability ($SSI < -2.5$) as well as events that are less extreme ($SSI > -2$). As expected, the range in SSI is narrower for the partial resampling experiment ($SSI_{partial.res}$, dark blue band) as compared with the full resampling experiment ($SSI_{full.res}$). The smaller spread in $SSI_{partial.res}$ for the Tweedie distribution as compared to the GEV distribution implies that the probability of below normal flow is less affected by changes in normal and above normal flow for the sample at hand.

3.3. Evaluation of Implications for Drought Monitoring and Characterization

Figure 8 shows the number of times the parametric SSI fell below a certain threshold for the river Dreisam for SSI time series fitted with the seven probability distributions and two fitting methods. Note that not all derived occurrences below the threshold are shown because not all probability distributions resulted in non-rejected SSI time series for all calendar months. The number of occurrences below the threshold varies among probability distribution and fitting method, however, depicting also a clear seasonal pattern. There

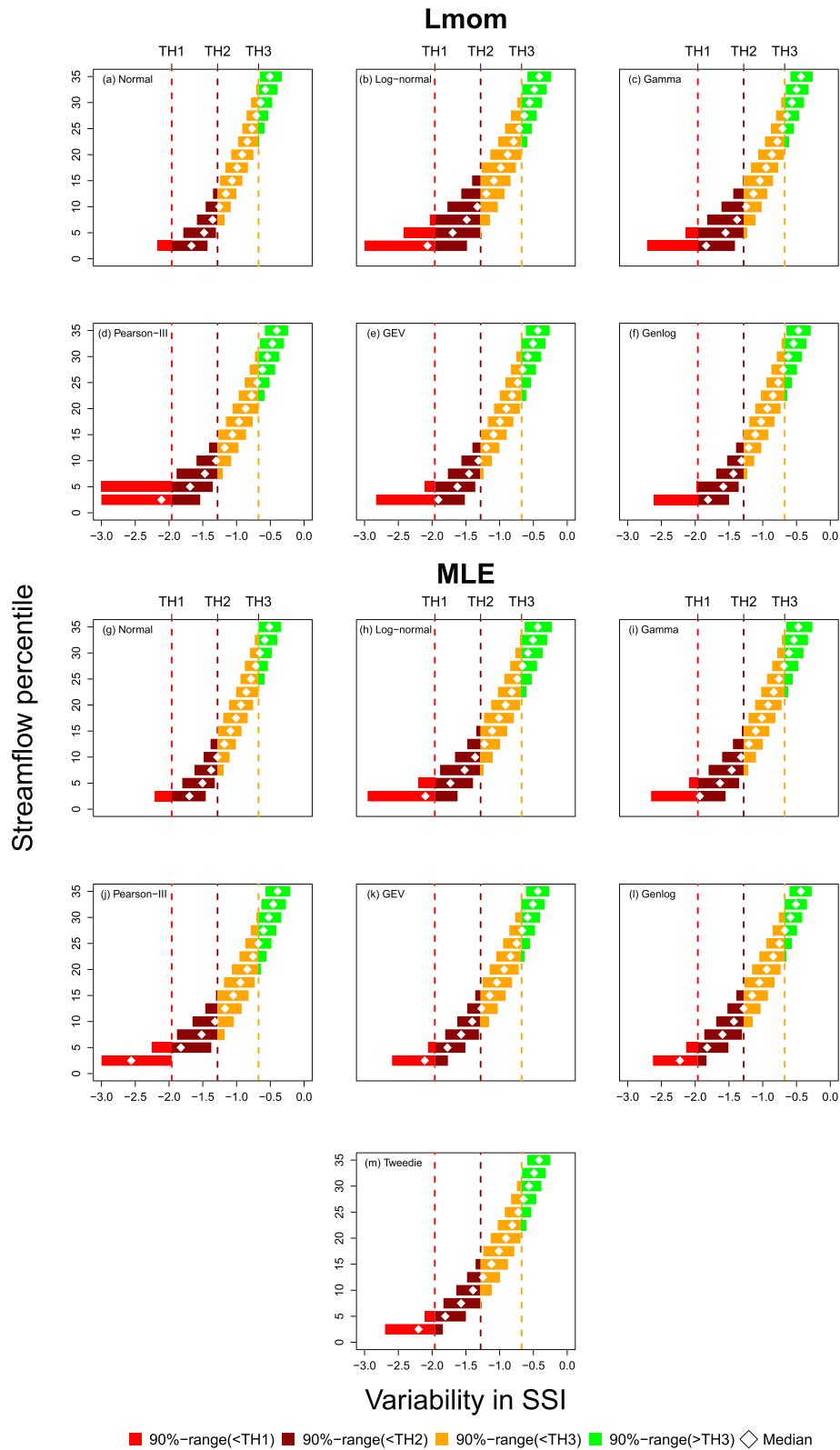


Figure 6. Spread in (non-rejected) SSI values computed based on seven probability distributions—Normal (a, g), Log-normal (b, h), Gamma (c, i), Pearson-III (d, j), GEV (e, k), Genlog (f, l), and Tweedie (m)—and two fitting methods—Lmom (a, f) and MLE (g, m)—across all catchments and calendar months. The median (triangle) and 90% range (box) are shown for different empirical streamflow percentiles. Colors of the box indicate whether the SSI is below a certain threshold (i.e., TH1, TH2, and TH3, section 2.5.1).

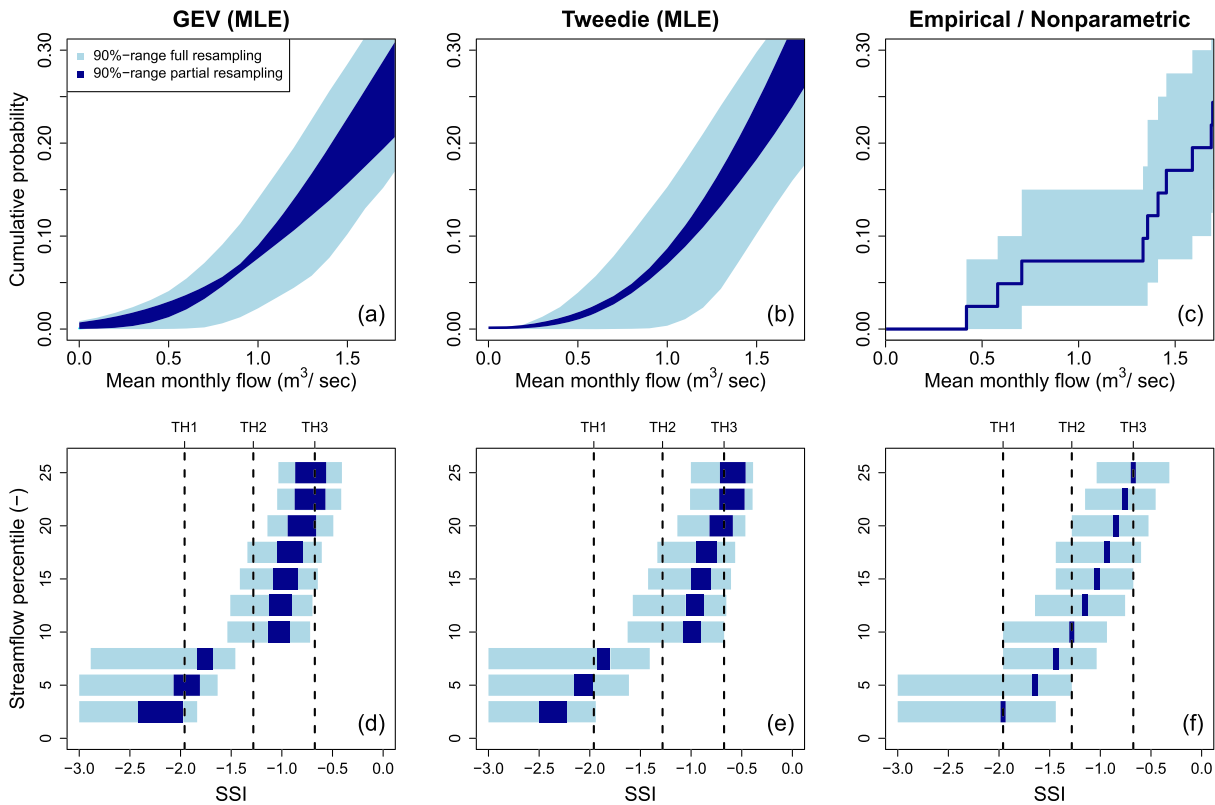


Figure 7. (a–c) The 90% range in the low end of the cumulative distribution of mean monthly flow for the river Dreisam (July) for three selected distributions—GEV (a), Tweedie (b), and empirical (c)—and (d–f) corresponding 90% ranges in SSI values. Ranges are derived based on full resampling (light blue) and partial resampling (dark blue).

are relatively few cases where SSI time series for a given month agree on the number of below threshold events. This number varies between 0 and 2 for TH1, between 1 and 7 for TH2, and between 7 and 14 for TH3. The number of, for example, below TH3 events is unevenly distributed over the year and is generally higher (above the expected historical non-exceedance frequency of 25%) in February, March, April, and November and lower (below the expected historical non-exceedance frequency) in summer. The spread in SSI values observed for catchments across Europe (shown in Figure 6) reveals that a similar variation in the number of occurrences below a certain threshold also exists within the entire European streamflow data set.

Figure 9 presents the non-rejected SSI time series of the river Dreisam fitted with different parametric probability distributions and fitting methods for the 2003 drought event. This drought event had the same duration (4 months) according to the SSI derived for all shown distributions. However, the deficit volume varied among SSI time series. For instance, the deficit volume of SSI time series computed with the Log-normal distribution (Lmom) is almost twice as high as the deficit volume of SSI time series computed with the Genlog distribution (Lmom).

Figure 10 presents the spread in average deficit volume derived from continuous parametric SSI time series of all EWA records without any rejected SSIs (colored symbols) as well as from SSI time series where at least 1 month contained a rejected SSI (gray symbols) according to different distributions and fitting methods. The diagonal of Figure 10 reveals the spread between fitting methods (Lmom and MLE) for the same probability distribution. Systematic differences between fitting methods are visible for the GEV and Genlog distribution; that is, average deficit volume is higher when computed with the MLE approach. Systematic differences were not found for any of the other distributions, with the exception of the rejected time series of the Pearson-III distribution. In general, high rank correlations were observed, indicating limited unsystematic differences between fitting methods.

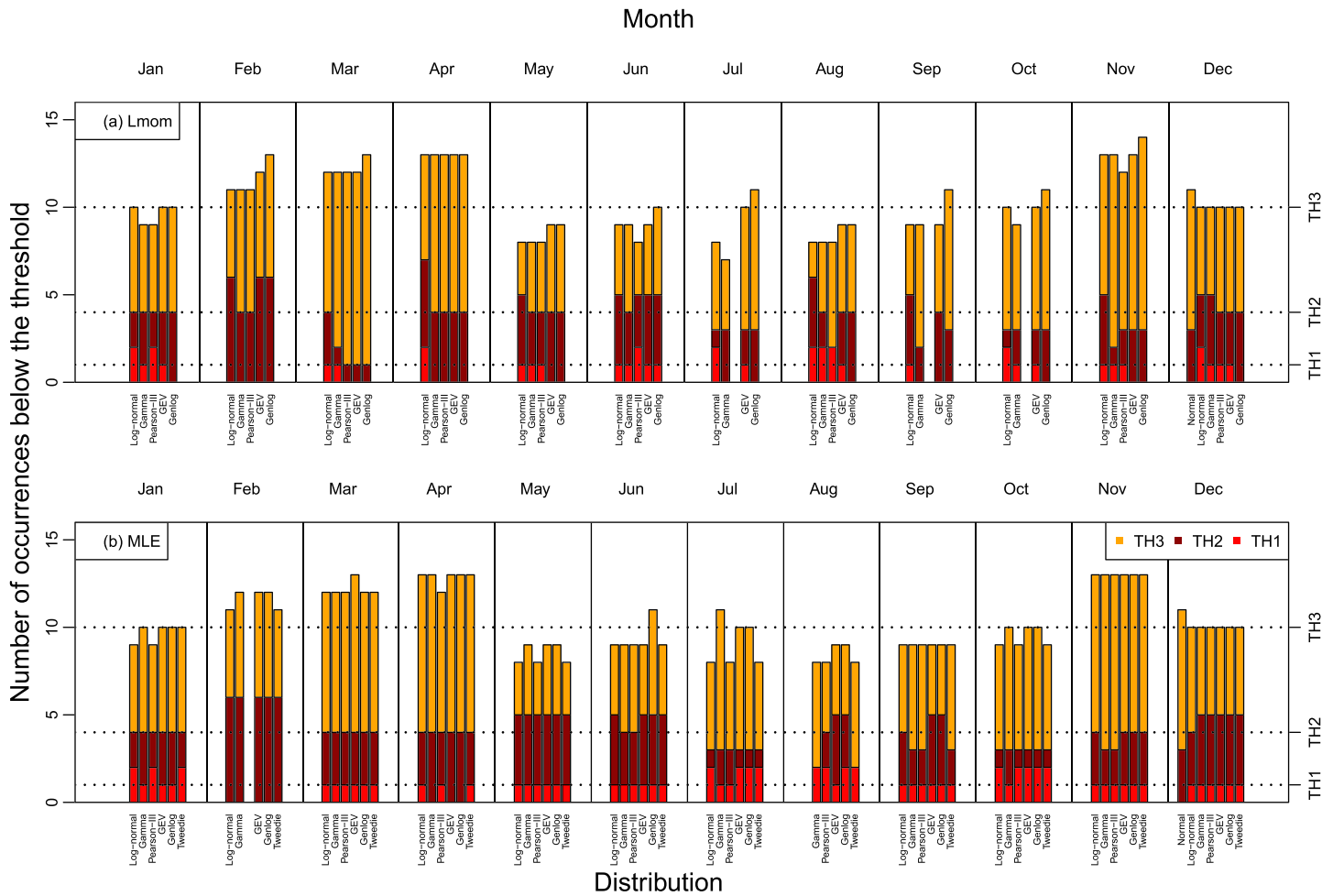


Figure 8. Seasonal variation in the number of times the SSI fell below a certain threshold for the river Dreisam derived from SSI time series computed with (a) the Lmom approach and (b) the MLE approach using different parametric probability distributions.

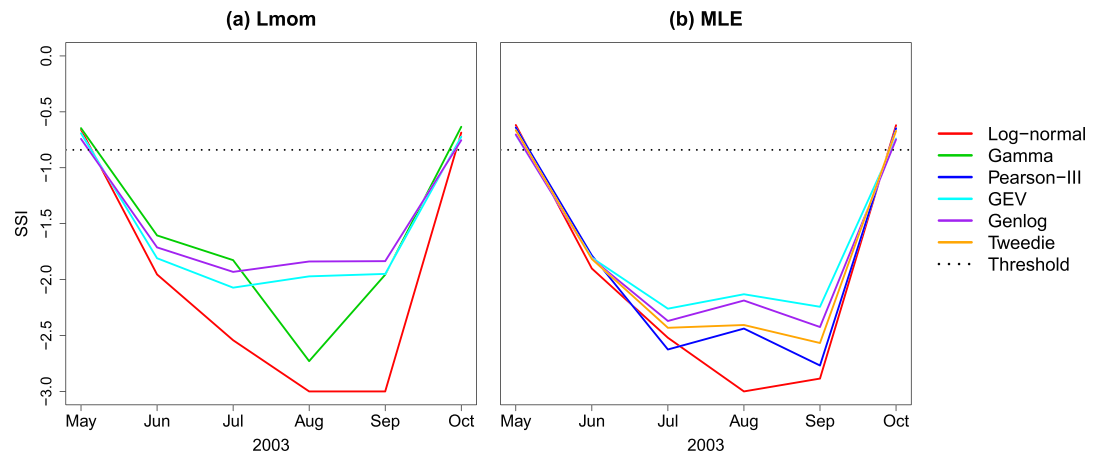


Figure 9. SSI time series for the drought of 2003 for the river Dreisam computed for different distributions and fitting methods: Lmom (a) and MLE (b).

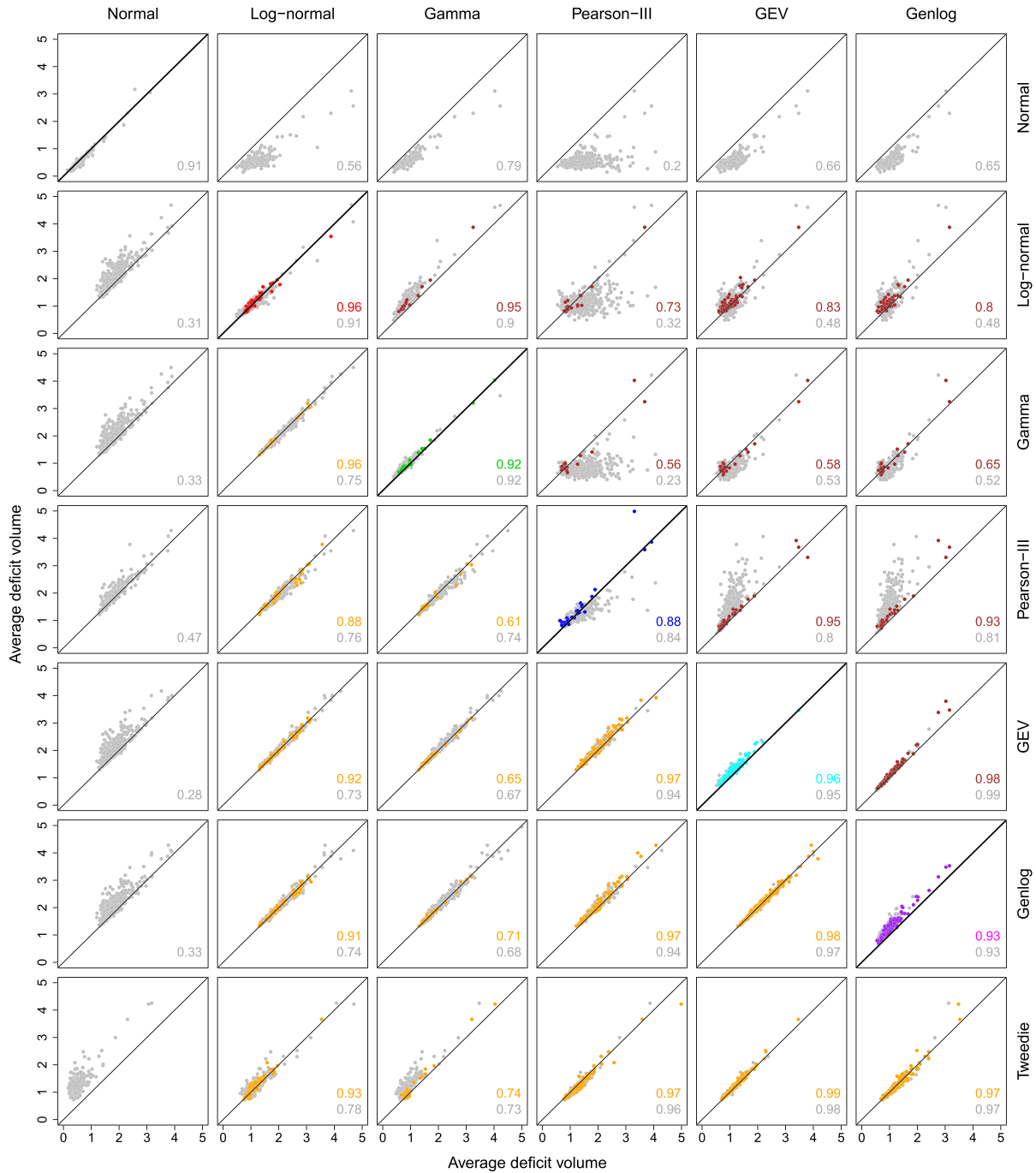


Figure 10. Sensitivity in average deficit volume derived from SSI time series to the used probability distribution and fitting method. Differences in average deficit volume among distributions are shown for both Lmom (plots above the diagonal) and MLE (plots below the diagonal). Differences between fitting methods (same distribution) are shown on the diagonal (y-axis = average deficit volume MLE, x-axis = average deficit volume Lmom). Colored points indicate that average deficit volume was derived from SSI time series with 12 non-rejected months; gray points indicate 1 or more months with rejected SSI time series; SSI of the rejected month(s) were included in the calculation of average deficit volume. Average deficit volume of SSI time series that contain one or more months in which MLE could not converge to a solution are not shown. Numbers in legend show Spearman's rank correlation coefficient for both the subsets with and without rejected SSIs (colored accordingly).

The plots above (Lmom) and below (MLE) the diagonal in Figure 10 present differences in average deficit volumes derived from SSI time series computed with different probability distributions, but the same fitting method. In general, there is a better agreement in deficit volume when it is derived from SSI time series

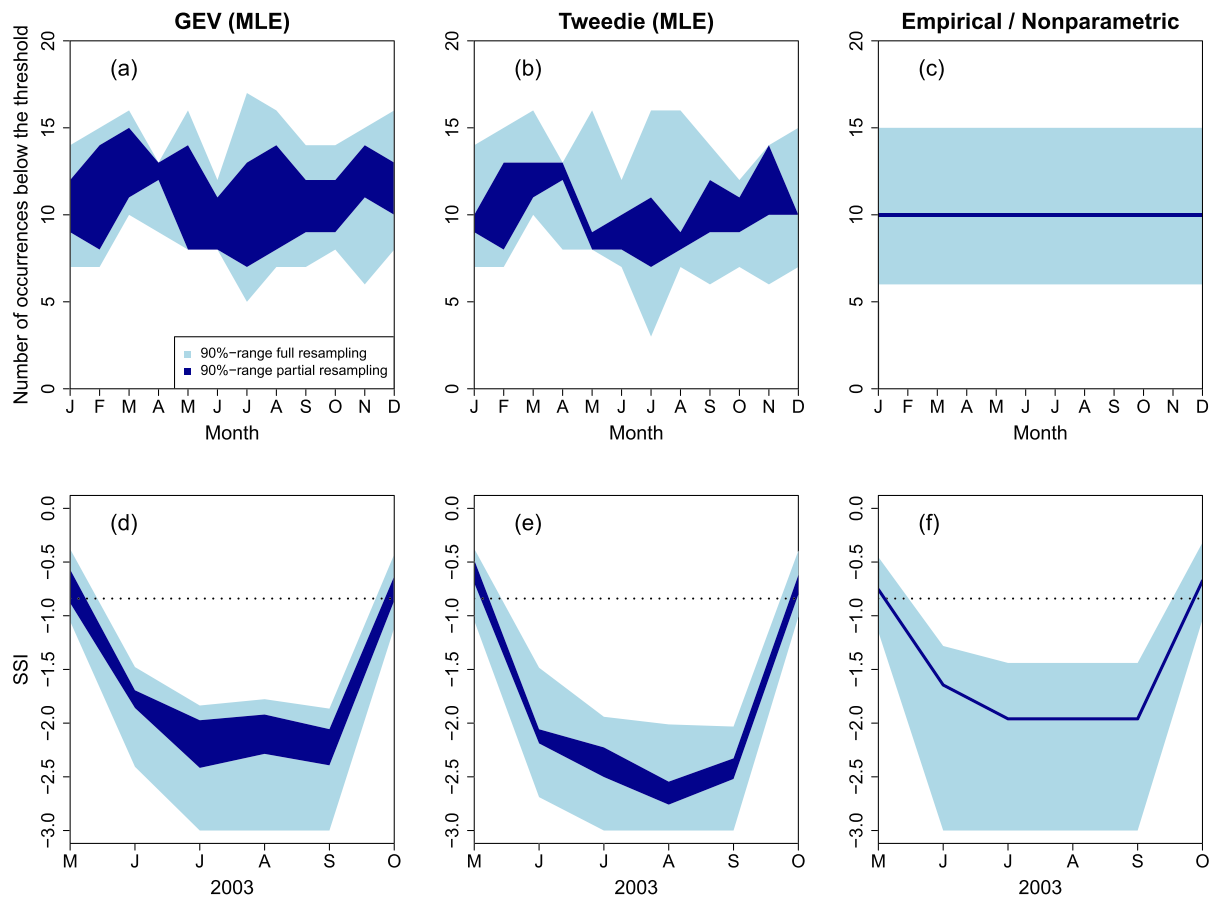


Figure 11. The 90% range in (a–c) the number of occurrences below the threshold (TH3) for the river Dreisam and (d–f) SSI time series of the river Dreisam in the year 2003. Ranges are derived from the resampled monthly flow samples; full resampling (light blue) and partial resampling (dark blue). Shown are results for the parametric SSI derived from the GEV (a, d) and Tweedie (b, e) distributions and the nonparametric SSI derived from empirical plotting positions (c, f).

without rejected months. Overall, lower average deficit volumes were found for the Normal distribution (although no cases exist where all calendar months had a non-rejected SSI distribution according to the Shapiro-Wilk test). The Gamma distribution revealed systematically lower average deficit volumes compared with the Log-normal distribution (both Lmom and MLE). Average deficit volume derived from time series computed with these two two-parameter probability distributions (Log-normal and Gamma) was distinctly different from those derived from SSI time series computed with any of the considered three-parameter probability distributions. These differences in average deficit volume were rather unsystematic, sometimes lower for one distribution and sometimes higher, as is also indicated by the lower Spearman's rank correlation coefficients, particularly when the SSI time series are computed with the Lmom approach. Average deficit volume from SSI time series derived with three-parameter distributions shows a strong agreement.

A similar spread in drought characteristics derived from resampled parametric SSI time series was found (Figures 11a, 11b, 11d, and 11e). For brevity reasons, this spread is only shown for the case study river Dreisam and the two distributions and fitting method that best described the data (GEV and Tweedie estimated with MLE). In general, there is a relatively large spread in both the number of occurrences below a certain threshold (Figures 11a and 11b) as well as in SSI time series (Figures 11d and 11e) and thus the associated drought characteristics such as deficit volume derived from those time series. The spread in $SSI_{full.res}$ is a measure of the uncertainty related to the sample properties. Accordingly, there will be a spread in drought characteristics. For instance, the SSI falls below TH3 between 4 and 16 times and also the deficit volume of the 2003 drought event varies clearly (Figures 11d and 11e). As expected, a smaller spread in

drought characteristics was found for the partially resampled SSI records ($SSI_{\text{partial.res}}$). Still, a sensitivity is seen in the below normal flow to changes in normal and above normal flows for the SSI derived from a parametric probability distribution (dark blue range in Figures 11a, 11b, 11d, and 11e).

3.4. Comparison With Nonparametric Approaches

The suitability of the nonparametric SSI was evaluated according to the goodness of fit and whether or not it exhibits the spread in minimum SSI values expected from a probabilistic index. For the nonparametric SSI, generally low rejection rates were found (Figure 2b), especially when compared to the rejection rates of some of the less suitable parametric distributions (Figure 2a). The nonparametric SSI derived from PP was never rejected, which is expected as PP directly describe the sample at hand. The nonparametric SSI derived with KDE was rejected for a small part of the samples when based on untransformed data, whereas KDE based on \ln -transformed data did not result in any rejected SSI series. On the other hand, the expected (spread in) minimum SSI values is notably underestimated by the considered nonparametric approaches (Figure 4b). The nonparametric SSI derived from PP shows a median minimum SSI value that is close to the expected median. However, no spread in minimum SSI values was observed. Minimum SSI values derived with KDE show some spread but overestimate the expected occurrence of low minimum SSI values, especially when KDE was applied on untransformed data.

The uncertainty of the nonparametric SSI was exemplified for the case study river Dreisam using the empirical PP method (Figures 7 and 11). When doing so, it should first be noted that the empirical distribution function (Figure 7c) is notably different from the two shown continuous distributions (GEV and Tweedie, Figures 7a and 7b) because of its stepwise shape and lack of information beyond the range spanned by the observations. A further notable difference between Figure 7c and Figures 7a and 7b are the larger uncertainty bounds in the left tail of the empirical percentile distribution. These larger uncertainties propagate to larger uncertainties in nonparametric SSI values derived from empirical PP (Figure 7f) as well as to larger uncertainties in drought characteristics derived from nonparametric SSI time series based on PP (Figures 11c and 11f). On the other hand, changes in normal and above normal sample properties, that is, the partial resampling experiment, do not propagate to uncertainties in the left tail of the empirical distribution function and the nonparametric SSI for below normal flow derived from empirical PP (Figure 7f). These normal- to high-end sample properties further do not have an effect on the derived drought characteristics (Figures 11c and 11f).

4. Discussion

4.1. SSI as a Probabilistic Drought Hazard Indicator

In this study, SSI time series were derived by fitting multiple candidate parametric probability distributions to monthly streamflow data using two fitting methods. Their performance was evaluated with respect to different criteria and moreover compared with nonparametric approaches based on PP and KDE. The main purpose of applying the SSI in drought studies is (i) to make streamflow time series comparable over space and time by standardizing often highly non-normal sample distributions for a given reference period and (ii) to allow estimating the re-occurrence probability of the below normal anomalies and extreme values in the tail of the distribution, where most SSI computation approaches enable extrapolation beyond the range spanned by the observations. Comparability over space and time can also be achieved using empirical streamflow percentiles that express anomalies as a historical non-exceedance frequency within a given period. The additional value of the SSI over empirical percentiles is that it transfers streamflow observations into probabilities based on the hypothesis that the variable of interest follows a specific distribution. The latter can provide valuable information for, for example, drought planning and management, but requires communication of the limitations and related uncertainties. In addition, if monthly flow is expressed as a non-exceedance probability, certain assumptions apply, including that the flow sample needs to be an independent and identically distributed random variable and that the distribution chosen satisfactory represents the monthly flow population.

Whether the chosen distribution can provide a satisfactory representation of the monthly flow population across the whole flow range is questionable. For example, for the case study river Dreisam, there is a clear mismatch between the empirical percentiles (sample) and the fitted parametric distributions, partly because

of the stepwise shape in the left tail of the distribution (Figure 5)—a common issue in low flow series. Part of the reason for this mismatch is related to the sample uncertainty; that is, the sample is not completely representative of the population. Furthermore, there are uncertainties associated with the empirical percentiles, being particularly high for the lowest/highest observed value (Figures 5, 7, and 11). However, the stepwise feature might also be part of the flow population. Hypothetical reasons could be that the lowest streamflow values are generated by other processes, such as base flow, whereas higher flow values might comprise a combination of baseflow and event flow. Another potential reason is that flow might cease from some tributaries during drought causing a step change in the flow behavior. In such cases, the considered parametric distributions might not be capable of capturing the complex mixture of processes in the left tail of the distribution. If the latter is the case, nonparametric approaches might give a better estimate of the tail of the population distribution. For instance, PP would directly describe the lowest three observations, whereas KDE would draw kernels around them. Alternatively, a mixed parametric distribution might be able to better reflect the stepwise shape in the left tail, but would also be less robust and more complex in its derivation as more parameters need to be estimated based on a rather limited data set. To conclude, we can only hypothesize about whether the stepwise feature is an element of the population distribution or not, and therefore, we do not know which approaches are more suitable to derive the SSI for this case. The latter stresses the importance of considering the uncertainty bounds; that is, although most parametric distributions do not directly describe such stepwise features, the observations will likely fall within their uncertainty bounds (e.g., Figure 7).

Whether monthly streamflow data are identically distributed is questionable as well. There are all kinds of possible non-stationarities in the streamflow record due to, for example, climate change, such as an increase in (potential) evaporation (e.g., Dai, 2011), or land cover changes, such as urbanization (e.g., Eng et al., 2013) or glacier retreat (e.g., Van Tiel et al., 2018). In addition, direct human influences (such as groundwater abstraction or reservoir operation) can greatly modify the flow record and are susceptible to changes over time (e.g., Tjeldeman et al., 2018), the latter not being a problem for the considered EWA catchments with near-natural flow in this study. Nevertheless, the above-described non-stationarities hinder a fair assignment of probabilities to streamflow, as flow values sampled in the past might not be representative for the current flow population.

4.2. Sensitivity to Parametric Probability Distributions and Fitting Methods

Based on a large European data set of monthly streamflow, this study revealed a sensitivity of the parametric SSI to both the probability distribution and fitting method. A similar sensitivity of the SSI to the choice of probability distribution was already shown in various previous studies, for example, for a set of streamflow records across the Iberian Peninsula by Vicente-Serrano et al. (2012) and across the United Kingdom by Svensson et al. (2017). The largest differences were found among SSI time series derived from two-parameter distributions as well as when comparing SSI time series of two-parameter distributions with those of three-parameter distributions. The SSI time series of the considered three-parameter distributions (besides Pearson-III) showed more similarity, mostly because these distributions often provided a good representation of the sample at hand. Our study further revealed differences in SSI time series computed with the same probability distribution, but using different distribution fitting methods, as also reported for the SPEI by Stagge et al. (2016).

The sensitivity of the SSI to the choice of distribution and fitting method has consequences for the application of the SSI for drought monitoring and early warning. Specifically, the number of occurrences below a given threshold will vary depending on the used distribution and fitting method (Figures 6 and 8). These discrepancies, if unaware of, can cause confusion if they are used as thresholds (triggers) in drought management plans or to portray the severity of drought conditions. The sensitivity of the SSI to the distribution and fitting method also has consequences for drought research aiming to characterize drought events; that is, using a different distribution might give a different account to the severity of historic drought events. Both systematic and unsystematic differences in drought deficit volume were found among drought characteristics derived from SSI time series computed with different distributions and fitting methods (Figure 10). Systematic differences are the least problematic for a comparison across catchments or climates, as another probability distribution or fitting method would likely give the same ranking of average deficit volume. More problematic are unsystematic differences, as the conclusion whether a catchment has higher or lower average deficit volumes than another catchment will depend on the method used.

4.3. Discrepancies Over Space and Time

The spread in SSI time series (across catchments and months) calculated with the same probability distribution and fitting method (Figure 6) is expected given the probabilistic nature of the SSI (section 3.1, Figure 4). At the same time, this spread complicates the interpretation and communication of the drought situation. For example, if unaware of, one might not expect that the lowest monthly flow value for some catchments or calendar months has an SSI value of -3 (non-exceedance probability of around 0.14%), whereas for other catchments and calendar months, the lowest average flow value has an SSI of -1.6 (or a non-exceedance probability of around 5.5%).

The spread found in SSI between catchments and calendar months further has implications for, for example, (large-scale) consistency in drought monitoring; that is, given thresholds (triggers) are passed more often for some catchments and calendar months (Figures 6 and 8). For example, if unaware of, one might not expect that for the Dreisam, the SSI falls below the 25th percentile threshold around 30% of time in March and 20% of time in August. A similar spread in the number of occurrences below the threshold between stations and climatic regions for the SPI in Texas led Quiring (2009) to suggest a set of objective thresholds based on empirical precipitation percentiles, that is, an SPI threshold between -1.5 and -1.64 depending on the climatic region, to match with the fixed (or predefined) percentile thresholds. By using the non-exceedance frequency directly as a threshold and a reference period that matches the observation period—instead of the SPI—the issues related to the spread in the historical number of occurrences below the threshold between stations and calendar months might have been avoided. However, it should be noted that the latter is only true if the reference period is equal to the observation period; that is, empirical percentiles are used to make inferences about the sample. A larger variability in the number of below threshold occurrences is expected when the empirical percentile distribution is used to make inferences about the flow population, for example, for observations outside of the reference period. This is confirmed by the often larger uncertainty bounds around (drought characteristics derived from) empirical streamflow percentiles (Figures 7 and 11).

4.4. Impact of Sample Properties and Uncertainty

The spread in resampled parametric SSI time series (full and partial) and drought characteristics derived from these resampled SSI time series (Figures 7 and 11) reveals relatively large uncertainty bounds. Similar uncertainties caused by sample differences were found in studies that investigated the impact of record length, reference period, and sample properties on the SSI (e.g., Hong et al., 2015; Núñez et al., 2014). Although the uncertainty in estimating SSI values is high, particularly for the lowest values, quantifying this uncertainty provides important information about the representativeness of the sample at hand as well as valuable, additional information for decision makers.

The partial resampling experiment demonstrated the degree to which the SSI for below normal flow anomalies is dependent on the properties of the (above) normal range of the sample. Although inherent in the nature of parametric statistical models, the fact that drought is getting more or less severe if normal and high flow magnitudes change raises several questions regarding comparability and communication, similar to the comment on flood frequency analysis made by Klemeš (2000). The dependency between normal and high flow magnitude and the SSI for below normal anomalies might further challenge drought research that aims to relate differences in drought characteristics across catchments to natural and human influences (e.g., Van Oel et al., 2018) or research that aims to compare the characteristics of SSI records derived from observations with those derived from simulations (e.g., Lai et al., 2019). Differences in SSI time series, or drought characteristics derived from those SSI time series, are not solely related to lower flows as the occurrence of more extreme flood events or human interventions that target high flow (such as flood control) can also modify the SSI and drought characteristics derived from those SSI records. The dependency between high flow magnitude and the SSI for low flow also challenges the assessment of changes in drought under future climate change (given that a new reference period is used to assess future drought conditions). The potential intensification of the hydrological cycle in some regions could result in more extreme wet and dry conditions, both affecting the probability and SSI for below normal flow. Accordingly, it is essential to evaluate the sensitivity to any changes in the flow regime when using the SSI.

Finally, it is worth noticing that bootstrap uncertainty ranges would be lower for longer records. However, considering longer record lengths could come at the cost of having a less representative sample as

catchments might have substantially changed over time. Further, the considered 40-year period in this study is comparable in length to the period used in numerous previous drought studies.

4.5. Comparison With the Nonparametric SSI

From a goodness of fit perspective, the nonparametric SSI often performs better than the parametric SSI. Only a small percentage of nonparametric SSI time series derived using KDE on untransformed data were rejected, whereas none of the other considered nonparametric approaches resulted in any rejected SSI series. However, a drawback of the nonparametric SSI is that it (severely) underestimates the expected occurrence of low SSI values. The minimum SSI value derived using PP is bounded and directly related to the considered number of observations (Equations 1–3). A similar lower bound is apparent for minimum SSI values derived using KDE, as at least part of the kernel drawn around the lowest observation contributes to the cumulative probability of the lowest observation (50% in case of a symmetric kernel). However, most of the time, more individual kernels contribute partly to the probability density function below the lowest observation, especially for the often positively skewed monthly flow distributions. This can inflate the left tail of the probability density function, especially when using untransformed data, increasing the magnitude of low (minimum) SSI values (Figure 4).

Although not directly shown, nonparametric SSI time series, and drought characteristics derived from those SSI time series, are sensitive to the used nonparametric method and can be variable over space and time. With regard to differences between methods, systematic differences are present among the nonparametric SSI derived with different PP formulas, purely because of differences in their definition (Equations 1–3). Figure 4 further suggests substantial differences among the nonparametric SSI derived with KDE based on either original or transformed data as well as differences among SSI time series derived with either PP or KDE. Discrepancies in space and time are also occurring for the nonparametric SSI derived with KDE (Figure 4). On the other hand, in the absence of ties in the sample and in the case that the reference period is equal to the observation period, the nonparametric SSI derived from PP would not show any discrepancies in space and time.

The nonparametric SSI derived from PP is associated with a relatively high uncertainty, especially for the extreme anomalies (see Figures 5, 7, and 11). On the other hand, the nonparametric SSI derived from PP for below normal flow is not sensitive to changes in the normal- or high-end of the sample. In the current study, we did not consider the uncertainty in the nonparametric SSI derived from KDE. The latter is done in Vergni et al. (2017) for the SPI and SPEI derived for a station in Italy, where generally smaller uncertainties were found when using KDE as compared to using a parametric approach. Nonetheless, the nonparametric SSI for below normal flow derived with the considered KDE methods in this study would be sensitive to changes in the normal- and high-end of the flow sample, as the bandwidth estimation is affected by properties of the entire flow range.

4.6. Streamflow Measurement Accuracy

Streamflow measurements are commonly derived from the stage-discharge relationship (rating curve). Uncertainties in the rating curve are typically high for both the high and the low flow end of the curve, due to, for example, measurement challenges. Measurement errors in both the high and low end of the flow spectrum affect the observations and thus the derivation of the parametric SSI, which again influence the use of the SSI as a drought index (below normal flow values). The nonparametric SSI derived from PP, on the other hand, is less affected by systematic errors in absolute values of flow magnitude, as it solely depends on the ranks of streamflow. An impact of measurement errors will only be observed if they imply a change in the order (ranks) of values. Furthermore, the nonparametric SSI values for below normal flow derived from PP are independent of the quantity of high flow magnitude, eliminating the potential impacts of measurement inaccuracies during high flow on the severity (probability) of below normal flow anomalies. The latter is not the case for the considered nonparametric SSI derived with KDE in this study (section 4.5).

4.7. Which Method to Use?

Today, it is relatively easy to derive an SSI time series, given that probability distribution fitting steps are automated in various software packages. For example, the Handbook of Drought Indicators and Indices from the World Meteorological Organization (WMO & GWP, 2016) recommends the use of the “SPI program” to compute the SSI. Notably, the latter handbook notes the main strength of the SSI as being “easy

to calculate” because it only requires running the “SPI program” of the National Drought Mitigation Centre, which uses by default the two-parameter Gamma distribution. Such recommendations are problematic, as accurate computation of the SSI requires a thorough investigation of, for example, the most suitable probability distribution for the variable at hand. The Gamma distribution has been found to be the most suited distribution for precipitation (SPI) in several studies (as confirmed in e.g., Stagge et al., 2015, for Europe as a whole). However, finding a regionally appropriate distribution for streamflow across different hydroclimatic regions and catchments is considered more challenging, and recent studies, including ours, suggest that the latter Gamma distribution is not appropriate. Accordingly, there is a need to raise awareness to the use of existing software—in this case developed for derivation of the SPI—to other variables and conditions without awareness of its assumptions and limitations. Further, clear recommendations on how to derive the SSI and its uncertainties as well as user-friendly guidance tools are required. Unfortunately, a clear recommendation how to derive the SSI is far from straightforward, as it depends partly on the used evaluation criteria including the goodness of fit, whether the SSI exhibits the expected characteristics of a probabilistic index, and the uncertainty.

From a goodness of fit perspective, nonparametric approaches are as expected more suitable, especially PP, as these approaches directly describe the sample at hand (Figure 2). Parametric approaches are based on sample statistics but do not directly describe each individual observation and therefore often have lower goodness of fit scores and higher rejection rates. However, the low rejection rates of especially the Tweedie, GEV, and Genlog distributions for the EWA data set suggest that these distributions are particularly suitable to derive the SSI (Figure 2), which is in line with findings of previous studies (e.g., Svensson et al., 2017; Vicente-Serrano et al., 2012). In addition, the general low rejection rates of the latter distributions suggest that using a best fitting distribution approach (as suggested in Vicente-Serrano et al., 2012) might not be needed, at least not for our set of samples. This would avoid any differences across SSI time series related to differences in distributional properties. From a goodness of fit perspective, the less flexible two-parameter distributions were more often rejected and therefore deemed less suitable.

From a probabilistic perspective, the SSI derived with the Log-normal distribution most accurately matches the (expected spread in) minimum SSI values (Figure 4). The SSI derived with other parametric distributions show either an overestimation or underestimation in (the spread of) minimum SSI values, where the overestimation or underestimation in minimum SSI values is generally less severe for the GEV, Genlog, and Tweedie distributions. On the other hand, the overestimation or underestimation is more severe for the two-parameter Normal and Gamma distribution and especially for the three-parameter Pearson-III distribution. The considered nonparametric approaches (severely) underestimate the magnitude and spread in minimum SSI values and are therefore less suitable to estimate probabilities of flow beyond the range spanned by the observations. This is an important drawback of the latter nonparametric approaches, as an advantage of probabilistic drought information is that it can be used to estimate probabilities of more extreme events, which, together with their uncertainties, can provide valuable information for, for example, drought planning and management.

From an uncertainty perspective, we showed that the parametric SSI derived with the GEV and Tweedie distribution generally had narrower uncertainty bounds as compared to the nonparametric SSI derived from empirical PP. On the other hand, the nonparametric SSI derived from PP for below normal flow anomalies is not sensitive to changes in the normal or high end of the sample. Overall, whether or not selecting the most suitable distribution based on low uncertainty bounds is a good selection criterion alone is questionable. Rather, it may be more important to select an appropriate population distribution and quantify the uncertainty.

When jointly considering all above evaluation criteria, one could argue that the parametric SSI derived with the Tweedie, GEV, or Genlog distribution are among the most suitable candidates. The parametric SSI derived with these distributions rarely performs best for any of the individual evaluation criteria but most importantly never performs badly either. A specific advantageous property of the Tweedie distribution is its lower bound at zero. However, using the latter distribution comes at the cost of higher computational demands. In the end, the potential user should carefully consider the performance of the considered methods using different evaluation criteria and decide which of those is most important for the desired

application. After choosing a method, the user should remain aware of the approach-specific strengths, limitations, and uncertainties and their impact on the derived results.

As a last consideration, we argue that, despite the benefit of probabilistic drought information as discussed above, it might not always be necessary to express the monthly streamflow drought hazard (SSI) in a probabilistic way, especially when considering the related assumptions, sensitivities, and uncertainties associated with its derivation. For some applications, for example, those where the key interest is to have anomaly information that is comparable over space and time, the use of empirical percentiles (section 2.6) expressing flow as a historical non-exceedance frequency in a given period might be preferred. For these applications, the simplicity, both computational and communication wise, are considered an asset, including (1) no assumptions associated with probabilities, (2) flexibility to describe any sample, (3) the insensitivity of empirical percentiles to flow magnitude (in the high end of the flow sample), and (4) the ease of calculation and interpretation (with regard to the last point, see also Shukla et al., 2011; Steinemann, 2014). However, it is important to stress that the use of empirical percentiles also requires consideration of their limitations and uncertainties. Even though there are no uncertainties associated to probability assumptions, there are still inherent uncertainties related to the used reference period; that is, percentiles only reveal sample statistics. Using percentiles to derive properties of the monthly flow population, for example, a certain drought threshold, is more uncertain. Finally, the fact that percentile rankings can be used without considering probability assumptions related to stationarity (section 4.1) does not imply that non-stationarities cannot be problematic in empirical streamflow percentile time series, as they can still cause a disconnect between the drought hazard and drought impact signal (Tijdeman et al., 2018).

5. Conclusion

Based on a diverse data set of (perennial) streamflow records across Europe, this study demonstrated the parametric SSI time series' sensitivity to the choice of probability distribution, fitting method, catchment, calendar month, and sample properties. It was revealed that SSI time series as well as drought characteristics derived from those time series, such as the number of below threshold occurrences or the deficit volume, varied considerably among the different methods. Further, the presented dependency of the SSI to both the low- and high-end sample properties shows the underlying uncertainties in SSI time series.

This sensitivity of the SSI to the used method and sample properties has implications for the use of the SSI in drought monitoring and early warning systems as well as in drought research, as major discrepancies found will give a rather different interpretation of a given drought. In particular, the discrepancy in the number of below threshold events among different (non)parametric methods is unwanted for drought monitoring and early warning applications as it could falsely initiate (or miss) triggers for actions in drought plans. Furthermore, systematic—and especially unsystematic—differences in drought deficit volumes between SSI time series computed with different probability distributions and fitting methods are unwanted in comparative drought research, as conclusions drawn from these studies might change depending on the method used. The dependency of the SSI (probabilities) for below normal flows to the normal and above normal end of the sample is potentially challenging to convey to users and decision makers. This dependency might further challenge drought research that aims to attribute differences in drought characteristics between catchments or periods, as these might not necessarily be solely related to differences in flow magnitude of dry anomalies. Overall, these sensitivities should be carefully evaluated when using the SSI in either drought monitoring and early warning applications or drought research.

Our results further confirm findings of previous studies that not all parametric distributions are equally suited to derive the SSI, one reason being that not all distributions are flexible enough to describe the flow population. With that regard, the use of nonparametric approaches, such as PP or KDE, can be attractive as these are flexible and directly describe the sample at hand. However, we exemplify that the nonparametric SSI derived from empirical PP has larger uncertainty bounds. We further show that the nonparametric SSI does not resemble the spread and magnitude of minimum SSI values expected from a probabilistic index. This can be a drawback, as quantification of these rarer events might be of particular interest for drought planning and management, especially when communicated together with the uncertainties.

Given the differences among the considered (non)parametric approaches and their varying performance according to different evaluation criteria, it was not straightforward to provide a clear recommendation

which method to use to derive the SSI. Nevertheless, an overall assessment of different evaluation criteria suggests the parametric SSI derived with the Tweedie, GEV, or Genlog distribution to be suitable candidates for the considered data set in this study, especially when taking into account the uncertainties. However, this should not be seen as a recommendation to use these distributions by default in future studies. Thorough testing of the suitability of the chosen method for the river flow data at hand remains of critical importance, in particular in case of a local case study where a good fit to a single time series is vital.

Both the SSI (parametric and nonparametric) and empirical streamflow percentiles are comparable over space and time. The advantage of the SSI over empirical flow percentiles is that it expresses flow as a non-exceedance probability, thereby allowing extrapolation beyond the period of record, rather than a historical non-exceedance frequency limited by a given time period. However, the additional value of expressing flow as a non-exceedance probability comes with various limitations and uncertainties that should be carefully considered. Probabilistic information and its uncertainty provide valuable information for stakeholders and decision makers, especially when it comes to planning for design extreme events, such as the 100-year event. Still, the use of empirical percentiles might be preferred in some cases where it is sufficient to express the streamflow situation as a historical non-exceedance frequency within a given period.

Data Availability Statement

Streamflow data from the European Water Archive (EWA) can be obtained (upon request) from the Global Runoff Data Centre (https://www.bafg.de/GRDC/EN/04_spcldtbss/42_EWA/ewa.html). Streamflow data from the case study river Dreisam were acquired from the regional state office for the environment of the federal state Baden-Württemberg (<https://www.lubw.badenwuerttemberg.de>). Parametric drought index time series derived in this study can be downloaded from the HeiDATA repository (<https://doi.org/10.11588/data/PFDJI1>). All analyses were carried out with the open-source software R (<https://www.r-project.org/>), partially using the packages “lmomco,” “tweedie,” “evd,” and “SCI.”

Acknowledgments

This work contributes to the DRIer project funded under the framework of the “Research Network Water” by the Ministry of Science, Research, and the Arts of the German Federal State of Baden-Württemberg. Funding from the German Research Foundation under DFG STA632/3-1 is acknowledged. This research further contributes to the UNESCO-IHP VIII FRIEND-Water program.

References

- Andreadis, K. M., Clark, E. A., Wood, A. W., Hamlet, A. F., & Lettenmaier, D. P. (2005). Twentieth-century drought in the conterminous United States. *Journal of Hydrometeorology*, 6(6), 985–1001. <https://doi.org/10.1175/JHM450.1>
- Bachmair, S., Stahl, K., Collins, K., Hannaford, J., Acreman, M., Svoboda, M., et al. (2016). Drought indicators revisited: The need for a wider consideration of environment and society. *Wiley Interdisciplinary Reviews Water*, 3(4), 516–536. <https://doi.org/10.1002/wat2.1154>
- Barker, L. J., Hannaford, J., Chiverton, A., & Svensson, C. (2016). From meteorological to hydrological drought using standardised indicators. *Hydrology and Earth System Sciences*, 20(6), 2483–2505. <https://doi.org/10.5194/hess-20-2483-2016>
- Cunnane, C. (1978). Unbiased plotting positions—A review. *Journal of Hydrology*, 37(3–4), 205–222. [https://doi.org/10.1016/0022-1694\(78\)90017-3](https://doi.org/10.1016/0022-1694(78)90017-3)
- Dai, A. (2011). Drought under global warming: A review. *Wiley Interdisciplinary Reviews: Climate Change*, 2(1), 45–65. <https://doi.org/10.1002/wcc.81>
- Eng, K., Wolock, D. M., & Carlisle, D. M. (2013). River flow changes related to land and water management practices across the conterminous United States. *Science of the Total Environment*, 463–464, 414–422. <https://doi.org/10.1016/j.scitotenv.2013.06.001>
- Farahmand, A., & AghaKouchak, A. (2015). A generalized framework for deriving nonparametric standardized drought indicators. *Advances in Water Resources*, 76, 140–145. <https://doi.org/10.1016/j.advwatres.2014.11.012>
- Folland, C., & Anderson, C. (2002). Estimating changing extremes using empirical ranking methods. *Journal of Climate*, 15(20), 2954–2960. [https://doi.org/10.1175/15200442\(2002\)015<2954:ECEUER>2.0.CO;2](https://doi.org/10.1175/15200442(2002)015<2954:ECEUER>2.0.CO;2)
- Hong, X., Guo, S., Zhou, Y., & Xiong, L. (2015). Uncertainties in assessing hydrological drought using streamflow drought index for the upper Yangtze River basin. *Stochastic Environmental Research and Risk Assessment*, 29(4), 1235–1247. <https://doi.org/10.1007/s00477-014-0949-5>
- 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(1), 105–124. <https://doi.org/10.1111/j.2517-6161.1990.tb01775.x>
- Ionita, M., Tallaksen, L. M., Kingston, D. G., Stagge, J. H., Laaha, G., Van Lanen, H. A. J., et al. (2017). The European 2015 drought from a climatological perspective. *Hydrology and Earth System Sciences*, 21(3), 1397–1419. <https://doi.org/10.5194/hess-21-13972017>
- Klemeš, V. (2000). Tall tales about tails of hydrological distributions II. *Journal of Hydrologic Engineering*, 5(3), 232–239. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2000\)5:3\(232\)](https://doi.org/10.1061/(ASCE)1084-0699(2000)5:3(232))
- Kumar, R., Musuuza, J. L., Van Loon, A. F., Teuling, A. J., Barthel, R., Ten Broek, J., et al. (2016). Multiscale evaluation of the Standardized Precipitation Index as a groundwater drought indicator. *Hydrology and Earth System Sciences*, 20, 1117–1131. <https://doi.org/10.5194/hess-20-1117-2016>
- Lai, C., Zhong, R., Wang, Z., Wu, X., Chen, X., Wang, P., & Lian, Y. (2019). Monitoring hydrological drought using long-term satellite-based precipitation data. *Science of the Total Environment*, 649, 1198–1208. <https://doi.org/10.1016/j.scitotenv.2018.08.245>
- Lorenzo-Lacruz, J., Vicente-Serrano, S. M., López-Moreno, J. I., Beguería, S., García-Ruiz, J. M., & Cuadrat, J. M. (2010). The impact of droughts and water management on various hydrological systems in the headwaters of the Tagus River (central Spain). *Journal of Hydrology*, 386(1–4), 13–26. <https://doi.org/10.1016/j.jhydrol.2010.01.001>
- McKee, T. B., Doesken, N. J., & Kleist, J. (1993). *The relationship of drought frequency and duration to time scales*. Paper presented at Proceedings of the 8th Conference on Applied Climatology, American Meteorological Society, Anaheim, USA.

- Mishra, A. K., & Singh, V. P. (2010). A review of drought concepts. *Journal of Hydrology*, 391(1-2), 202–216. <https://doi.org/10.1016/j.jhydrol.2010.07.012>
- Modarres, R. (2007). Streamflow drought time series forecasting. *Stochastic Environmental Research and Risk Assessment*, 21(3), 223–233. <https://doi.org/10.1007/s00477-006-0058-1>
- Nalbantis, I., & Tsakiris, G. (2009). Assessment of hydrological drought revisited. *Water Resources Management*, 23(5), 881–897. <https://doi.org/10.1007/s11269-008-9305-1>
- Núñez, J., Rivera, D., Oyarzún, R., & Arumí, J. L. (2014). On the use of standardized drought indices under decadal climate variability: Critical assessment and drought policy implications. *Journal of Hydrology*, 517, 458–470. <https://doi.org/10.1016/j.jhydrol.2014.05.038>
- Quiring, S. M. (2009). Developing objective operational definitions for monitoring drought. *Journal of Applied Meteorology and Climatology*, 48(6), 1217–1229. <https://doi.org/10.1175/2009JAMC2088.1>
- Samaniego, L., Kumar, R., & Zink, M. (2012). Implications of parameter uncertainty on soil moisture drought analysis in Germany. *Journal of Hydrometeorology*, 14(1), 47–68. <https://doi.org/10.1175/JHM-D-12-075.1>
- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3–4), 591–611. <https://doi.org/10.1093/biomet/52.3-4.591>
- Sheather, S. J., & Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society: Series B: Methodological*, 53(3), 683–690. <https://doi.org/10.1111/j.2517-6161.1991.tb01857.x>
- Sheffield, J., Andreadis, K. M., Wood, E. F., & Lettenmaier, D. P. (2009). Global and continental drought in the second half of the twentieth century: Severity-area-duration analysis and temporal variability of large-scale events. *Journal of Climate*, 22(8), 1962–1981. <https://doi.org/10.1175/2008JCLI2722.1>
- Shukla, S., Steinemann, A. C., & Lettenmaier, D. P. (2011). Drought monitoring for Washington State: Indicators and applications. *Journal of Hydrometeorology*, 12(1), 66–83. <https://doi.org/10.1175/2010JHM1307.1>
- Shukla, S., & Wood, A. W. (2008). Use of a standardized runoff index for characterizing hydrologic drought. *Geophysical Research Letters*, 35, L02405. <https://doi.org/10.1029/2007GL032487>
- Smakhtin, V. U. (2001). Low flow hydrology: A review. *Journal of Hydrology*, 240(3–4), 147–186. [https://doi.org/10.1016/S0022-1694\(00\)00340-1](https://doi.org/10.1016/S0022-1694(00)00340-1)
- Sořáková, T., De Michele, C., & Vezzoli, R. (2014). Comparison between parametric and nonparametric approaches for the calculation of two drought indices: SPI and SSI. *Journal of Hydrologic Engineering*, 19, 04014010. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000942](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000942)
- Stagge, J. H., Tallaksen, L. M., Gudmundsson, L., Van Loon, A. F., & Stahl, K. (2015). Candidate distributions for climatological drought indices (SPI and SPEI). *International Journal of Climatology*, 35(13), 4027–4040. <https://doi.org/10.1002/joc.4267>
- Stagge, J. H., Tallaksen, L. M., Gudmundsson, L., Van Loon, A. F., & Stahl, K. (2016). Response to comment on “Candidate Distributions for Climatological Drought Indices (SPI and SPEI)”. *International Journal of Climatology*, 36(4), 2132–2138. <https://doi.org/10.1002/joc.4564>
- Stahl, K., Hisdal, H., Hannaford, J., Tallaksen, L. M., van Lanen, H. A. J., Sauquet, E., et al. (2010). Streamflow trends in Europe: Evidence from a dataset of near-natural catchments. *Hydrology and Earth System Sciences*, 14(12), 2367–2382. <https://doi.org/10.5194/hess-14-2367-2010>
- Staudinger, M., Stahl, K., & Seibert, J. (2014). A drought index accounting for snow. *Water Resources Research*, 50, 7861–7872. <https://doi.org/10.1002/2013WR015143>
- Steinemann, A. (2014). Drought information for improving preparedness in the western states. *Bulletin of the American Meteorological Society*, 95(6), 843–847. <https://doi.org/10.1175/BAMS-D-13-00067.1>
- Svensson, C., Hannaford, J., & Prosdocimi, I. (2017). Statistical distributions for monthly aggregations of precipitation and streamflow in drought indicator applications. *Water Resources Research*, 53, 999–1018. <https://doi.org/10.1002/2016WR019276>
- Tallaksen, L. M., & Van Lanen, H. A. J. (2004). *Hydrological drought—Processes and estimation methods for streamflow and groundwater*. Amsterdam, The Netherlands: Elsevier B.V.
- Tijdeman, E., Barker, L. J., Svoboda, M. D., & Stahl, K. (2018). Natural and human influences on the link between meteorological and hydrological drought indices for a large set of catchments in the contiguous United States. *Water Resources Research*, 54, 6005–6023. <https://doi.org/10.1029/2017WR022412>
- Tijdeman, E., Hannaford, J., & Stahl, K. (2018). Human influences on streamflow drought characteristics in England and Wales. *Hydrology and Earth System Sciences*, 22(2), 1051–1064. <https://doi.org/10.5194/hess-22-1051-2018>
- Van Loon, A. F., Tijdeman, E., Wanders, N., Van-Lanen, H. A. J., Teuling, A. J., & Uijlenhoet, R. (2014). How climate seasonality modifies drought duration and deficit. *Journal of Geophysical Research: Atmospheres*, 119, 4640–4656. <https://doi.org/10.1002/2013JD020383>
- Van Oel, P. R., Martins, E. S. P. R., Costa, A. C., Wanders, N., & Van Lanen, H. A. J. (2018). Diagnosing drought using the downstreamness concept: The effect of reservoir networks on drought evolution. *Hydrological Sciences Journal*, 63(7), 979–990. <https://doi.org/10.1080/02626667.2018.1470632>
- Van Tiel, M., Teuling, A. J., Wanders, N., Vis, M. J. P., Stahl, K., & Van Loon, A. F. (2018). The role of glacier dynamics and threshold definition in the characterisation of future streamflow droughts in glacierised catchments. *Hydrology and Earth System Sciences*, 22(1), 463–485. <https://doi.org/10.5194/hess-22-463-2018>
- Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S*. New York, USA: Springer. <https://doi.org/10.1007/978-0-387-21706-2>
- Vergni, L., Todisco, F., & Mannocchi, F. (2017). Evaluating the uncertainty and reliability of standardized indices. *Hydrology Research*, 48(3), 701–713. <https://doi.org/10.2166/nh.2016.076>
- Vicente-Serrano, S. M., Beguería, S., & López-Moreno, J. I. (2010). A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index. *Journal of Climate*, 23(7), 1696–1718. <https://doi.org/10.1175/2009JCLI2909.1>
- Vicente-Serrano, S. M., López-Moreno, J. I., Beguería, S., Lorenzo-Lacruz, J., AzorinMolina, C., & Morán-Tejeda, E. (2012). Accurate computation of a streamflow drought index. *Journal of Hydrologic Engineering*, 17(2), 318–332. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000433](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000433)
- Vicente-Serrano, S. M., Lopez-Moreno, J. I., Beguería, S., Lorenzo-Lacruz, J., SanchezLorenzo, A., García-Ruiz, J. M., et al. (2014). Evidence of increasing drought severity caused by temperature rise in southern Europe. *Environmental Research Letters*, 9, 044001. <https://doi.org/10.1088/1748-9326/9/4/044001>
- Vidal, J.-P., Martin, E., Franchistéguy, L., Habets, F., Soubeyroux, J.-M., Blanchard, M., & Baillon, M. (2010). Multilevel and multiscale drought reanalysis over France with the Safran-IsbaModcou hydrometeorological suite. *Hydrology and Earth System Sciences*, 14, 459–478. <https://doi.org/10.5194/hess-14-459-2010>

- Von Storch, H., & Zwiers, F. W. (1999). *Statistical analysis in climate research*. Cambridge, UK: Cambridge University Press.
- Wen, L., Rogers, K., Ling, J., & Saintilan, N. (2011). The impacts of river regulation and water diversion on the hydrological drought characteristics in the Lower Murrumbidgee River, Australia. *Journal of Hydrology*, 405(3–4), 382–391. <https://doi.org/10.1016/j.jhydrol.2011.05.037>
- Wilhite, D. A. (1992). *Preparing for drought: A guidebook for developing countries*. Nairobi, Kenya: Diane Publishing.
- Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences*. London, England: Academic Press.
- WMO & GWP (2016). In M. Svoboda & B. A. Fuchs (Eds.), *Handbook of drought indicators and indices, Integrated Drought Management Tools and Guidelines Series 2*. Geneva, Switzerland: World Meteorological Organization (WMO) and Global Water Partnership (GWP).
- Zaidman, M. D., Rees, H. G., & Young, A. R. (2002). Spatio-temporal development of streamflow droughts in north-west Europe. *Hydrology and Earth System Sciences*, 6(4), 733–751. <https://doi.org/10.5194/hess-6-733-2002>
- Zelenhasić, E., & Salvai, A. (1987). A method of streamflow drought analysis. *Water Resources Research*, 23(1), 156–168. <https://doi.org/10.1029/WR023i001p00156>
- Zhang, R., Chen, X., Zhang, Z., & Shi, P. (2014). Evolution of hydrological drought under the regulation of two reservoirs in the headwater basin of the Huaihe River, China. *Stochastic Environmental Research and Risk Assessment*, 29(2), 487–499. <https://doi.org/10.1007/s00477-014-0987-z>