

# Candidate Distributions for Climatological Drought Indices (SPI and SPEI)

James H. Stagge,<sup>a\*</sup> Lena M. Tallaksen,<sup>a</sup> Lukas Gudmundsson,<sup>b</sup> Anne F. Van Loon<sup>c,d</sup> and Kerstin Stahl<sup>e</sup>

<sup>a</sup> Department of Geosciences, University of Oslo, Norway

<sup>b</sup> Institute for Atmospheric and Climate Science, ETH Zurich, Switzerland

<sup>c</sup> Hydrology and Quantitative Water Management Group, Wageningen University, The Netherlands

<sup>d</sup> School of Geography, Earth and Environmental Sciences, University of Birmingham, UK

<sup>e</sup> Institute of Hydrology, University of Freiburg, Germany

**ABSTRACT:** The Standardized Precipitation Index (SPI), a well-reviewed meteorological drought index recommended by the World Meteorological Organization (WMO), and its more recent climatic water balance variant, the Standardized Precipitation-Evapotranspiration Index (SPEI), both rely on selection of a univariate probability distribution to normalize the index, allowing for comparisons across climates. Choice of an improper probability distribution may impart bias to the index values, exaggerating or minimizing drought severity. This study compares a suite of candidate probability distributions for use in SPI and SPEI normalization using the  $0.5^\circ \times 0.5^\circ$  gridded Watch Forcing Dataset (WFD) at the continental scale, focusing on Europe. Several modifications to the SPI and SPEI methodology are proposed, as well as an updated procedure for evaluating SPI/SPEI goodness of fit based on the Shapiro–Wilk test. Candidate distributions for SPI organize into two groups based on their ability to model short-term accumulation (1–2 months) or long-term accumulation ( $>3$  months). The two-parameter gamma distribution is recommended for general use when calculating SPI across all accumulation periods and regions within Europe, in agreement with previous studies. The generalized extreme value distribution is recommended when computing the SPEI, in disagreement with previous recommendations.

KEY WORDS drought index; probability distribution; Standardized Precipitation Index; SPI; SPEI; potential evapotranspiration

Received 24 January 2014; Revised 24 November 2014; Accepted 6 January 2015

## 1. Introduction

Drought indices are used to objectively quantify and compare drought severity, duration, and extent across regions with varied climatic and hydrologic regimes. The Standardized Precipitation Index (SPI), outlined by McKee *et al.* (1993) and Guttman (1999), measures normalized anomalies in precipitation and has been recommended as a key drought indicator by the World Meteorological Organization (WMO, 2006) and a universal meteorological drought index by the Lincoln Declaration on Drought (Hayes *et al.*, 2011). In this way, accumulated precipitation can be compared objectively across locations with different climatologies and highly non-normal precipitation distributions. In addition, the more recently proposed Standardized Precipitation-Evapotranspiration Index (SPEI) (Vicente-Serrano *et al.*, 2010; Beguería *et al.*, 2013) utilizes a similar concept, but instead normalizes accumulated climatic water balance anomalies, defined as the difference between precipitation and potential evapotranspiration (PET). This retains the simplicity,

multitemporal nature, and statistical interpretability of the SPI, while producing a more comprehensive measure of water availability that accounts for atmospheric conditions that also affect drought severity such as temperature, wind speed, and humidity.

The process of transforming the often highly skewed distribution of accumulated precipitation (SPI) and climatic water balance (SPEI) to the standard normal distribution requires the choice and fitting of a univariate probability distribution. Selection of an appropriate parametric probability distribution is a key decision in calculating these drought indices, as selection of an inappropriate distribution can impart bias to the index values, exaggerating or minimizing drought severity (Sienz *et al.*, 2012).

The SPI is a well-tested and generally accepted index of drought severity, with several univariate probability distribution recommendations (Guttman, 1999; Lloyd-Hughes and Saunders, 2002; Giddings *et al.*, 2005; Vicente-Serrano *et al.*, 2010). However, these recommendations typically rely on the Kolmogorov–Smirnov (K–S) test (Lloyd-Hughes and Saunders, 2002; Vicente-Serrano *et al.*, 2010), which has been shown to be relatively insensitive in previous statistical analyses (Stephens, 1974; Mason and Schuenemeyer, 1983) and requires time-consuming Monte Carlo simulation of critical

\*Correspondence to: J. H. Stagge, Department of Geosciences, University of Oslo, P.O. Box 1047, Blindern, N-0316 Oslo, Norway. E-mail: j.h.stagge@geo.uio.no

values when applied to distributions derived from the data (Crutcher, 1975; Steinskog *et al.*, 2007). Probability distribution fitting for the SPI is also complicated by the presence of periods with zero precipitation, as described by Wu *et al.* (2007), and current procedures do not adequately capture the likelihood of zero precipitation events.

The SPEI has great promise as a drought index because it captures a broader measure of the available water (climatic water balance) and avoids issues inherent in the SPI, such as fitting periods with zero precipitation. However, as a newer index, it requires more rigorous testing with respect to its methodology and assumptions before it can gain widespread acceptance within the drought community. Index sensitivity to PET calculation method has already been addressed by Stagge *et al.* (2014) and Beguería *et al.* (2013), but to date there has been little testing of the univariate distributions used to normalize the SPEI. The original SPEI validation study (Vicente-Serrano *et al.*, 2010) selected a probability distribution based on monthly historical records from a limited number of sites around the world (11 in total). Recommendations from Vicente-Serrano *et al.* (2010) have formed the basis for several further studies (Beguería *et al.*, 2010; Vicente-Serrano *et al.*, 2011), without additional documented distribution testing.

Given these outstanding issues with implementing the SPI and SPEI, this study is designed to standardize the SPI and SPEI methodology, by introducing and employing a more sensitive statistical test for candidate probability distributions used to normalize these drought indices. The objectives are (1) to highlight the need for more rigorous testing of the assumed distributions for SPI and SPEI, (2) to introduce a methodology to perform this distribution testing, (3) to propose several improvements to SPI and SPEI methodology, and (4) to provide a relative comparison of several common distributions used to normalize the SPI and SPEI along with a general distribution recommendation.

This study uses Europe as a testing region, using maximum likelihood estimation (MLE) to fit distributions and normalize accumulated precipitation at the daily temporal and  $0.5^\circ \times 0.5^\circ$  spatial resolution. Candidate distributions are tested rigorously, first using the K-S and Anderson–Darling (A–D) tests as applied in previous studies, and then by a new method based on the Shapiro–Wilk (S–W) test for normality. Relative distribution rankings are determined by Akaike Information Criterion (AIC) as in Sienz *et al.* (2012), and all comparisons are evaluated with respect to accumulation period and location to determine if and why systematic biases occur. Using these tests, candidate distributions are recommended for the SPI and SPEI. While these distribution recommendations are shown to be successful general recommendations for a continental scale analysis of Europe, it is important to use the proposed testing methodology to verify goodness of fit whenever working in a new region or with a new data set.

## 2. Data and methods

### 2.1. Research extent

For the purpose of this research, the European extent is defined as the region between  $34^\circ\text{--}72^\circ\text{N}$  latitude and  $-13^\circ\text{--}32^\circ\text{E}$  longitude. This region is characterized by three major climate regions, described herein as the Southern Europe/Mediterranean, Central Europe, and Northern Europe. The Southern European/Mediterranean region is characterized by higher temperatures and a dry summer season, influenced by the Subtropical High Pressure Belt in the summer and mid-latitude westerlies during the winter (Stahl and Hisdal, 2004). Central Europe is characterized by a more temperate climate without a major summer dry season and is influenced by the westerlies of the mid-latitudes throughout the year. The Central European climate zone tends towards maritime climate in the west and continental climate patterns in the east. Northern Europe has a similar westerly influence and climatic pattern as Central Europe, but tends towards colder temperatures and significantly lower solar radiation during the winter (Stahl and Hisdal, 2004).

### 2.2. Climate data

All climate estimates were based on the Watch Forcing Dataset (WFD), a gridded historical climate data set based on ERA-40 reanalysis with  $0.5^\circ \times 0.5^\circ$  resolution (Weedon *et al.*, 2011). The WFD consists of subdaily forcing data spanning the time period 1 January 1958–31 December 2001 and employs bias correction for temperature and precipitation based on CRU-TS2.1 observations. Using the European extent defined above ( $34^\circ\text{--}72^\circ\text{N}$ ,  $-13^\circ\text{--}32^\circ\text{E}$ ) resulted in 3950 land grid cells based on the CRU land surface mask. Precipitation was calculated as the sum of rainfall and snowfall from the WFD data set. Other climate variables needed to calculate PET include daily average surface temperature (2 m), daily min/max surface temperature (2 m), and wind speed (10 m). Mean temperature and diurnal temperature range have been bias corrected based on CRU data (Weedon *et al.*, 2011), whereas WFD wind speed data received no bias correction (Haddeland *et al.*, 2011). For consistency, leap years were shortened to a 365-day calendar year by censoring the final day of the year.

When validated against Fluxnet observations (Weedon *et al.*, 2010, 2011), the WFD accurately reproduced temperature and precipitation, especially at the daily, monthly, and seasonal time scales. The WFD was found to be particularly accurate in Europe where CRU station coverage, used for bias correction, is most dense. The WFD has been shown to be accurate even in relatively data-sparse regions (van Huijgevoort *et al.*, 2011; Li *et al.*, 2013, 2014).

PET was calculated in this study using the Penman–Monteith equation with the Hargreaves–Samani modification (Hargreaves and Samani, 1985) as described in the FAO-56 (Allen *et al.*, 1998). This form of the Penman–Monteith equation was chosen because of its consistent performance in SPEI sensitivity tests (Beguería *et al.*, 2013; Stagge *et al.*, 2014), its relatively low data

requirements, and its recommendation as the standard for accurate PET estimation by both the WMO (WMO, 2006) and the FAO-UN (Allen *et al.*, 1998). Use of the Penman–Monteith equation differs from the original SPEI methodology (Vicente-Serrano *et al.*, 2010), which used the Thornthwaite method. The Thornthwaite equation has several well-documented limitations and biases (Jensen, 1973; Amatya *et al.*, 1995) and was found to produce SPEI index values that differed significantly from all other PET formulations tested by Stagge *et al.* (2014).

The modified form of the Penman–Monteith equation uses the daily difference between  $T_{\max}$  and  $T_{\min}$  as a proxy to estimate net radiation (Hargreaves and Samani, 1985), while retaining the full mass transfer term. This simplified method retains the physical foundation of the Penman–Monteith equation, while also largely avoiding concerns with mixing bias-corrected WFD temperature and precipitation with non-bias-corrected radiation (Haddeland *et al.*, 2011). Use of the modified Penman–Monteith equation for calculating SPEI is reasonable, as FAO notes that the modified equation is most accurate when daily estimates are averaged over several days or weeks (Allen *et al.*, 1998). For a more detailed explanation of the Penman–Monteith equation and assumptions used in this study, see Appendix A, which summarizes Allen *et al.* (1998).

### 2.3. SPI and SPEI calculation

SPI is typically computed by summing precipitation over  $k$  months, termed accumulation periods, and fitting these accumulated precipitation values to a parametric statistical distribution from which probabilities are transformed to the standard normal distribution ( $\mu = 0, \sigma = 1$ ) (McKee *et al.*, 1993; Guttman, 1999; Lloyd-Hughes and Saunders, 2002). SPEI is calculated in a similar fashion, but instead sums climatic water balance, defined as the difference between precipitation and PET (Vicente-Serrano *et al.*, 2010). Once accumulated precipitation has been transformed to probabilities, these probabilities are converted to the standard normal distribution to create the final drought index values. In this way, SPI and SPEI values are statistically interpretable, representing the number of standard deviations from typical accumulated precipitation, or climatic water balance, for a given location and time of year. All normalization was performed relative to the reference period 1970–1999, in accordance with standard 30-year reference periods.

In this study, SPI and SPEI are calculated at a daily temporal resolution, however retaining a monthly averaging period. This differs from previous studies that use monthly data, summing precipitation on the final day of each month. By using a daily time step, the drought index retains the higher resolution of the original climate data set, providing useful details for drought duration and climatic transitions within months. Accumulation periods considered in this study are the commonly used periods: 1, 2, 3, 6, 9, and 12 months, which are considered equivalent to 31, 61, 91, 183, 274, and 365 days, in the proposed daily

framework. For ease of interpretability and comparisons with past studies, we use here the previously established monthly nomenclature, for example SPI-6 corresponds to the SPI with a 6-month (183 day) accumulation period. Using a daily resolution requires fitting 365 parametric probability distributions, rather than 12 distributions, again providing more detail and smoother transitions.

### 2.4. Univariate distributions and fitting procedure

Care was taken to evaluate the most commonly used distributions and those recommended in past SPI/SPEI studies. Distributions were fit by MLE using the *fitdistrplus* package (Delignette-Muller *et al.*, 2013) in R, with L-moments used to establish initial values where possible. This combines the flexibility of MLE, used in McKee *et al.* (1993), Lloyd-Hughes and Saunders (2002), and Sienz *et al.* (2012), with the robustness of L-moments, recommended by Guttman (1999), Hosking and Wallis (2005), Vicente-Serrano *et al.* (2010), and Beguería *et al.* (2013).

Distributions considered for SPI include the (1) Gamma, (2) Gumbel, (3) Logistic, (4) Log-Logistic, (5) Lognormal, (6) Normal, and (7) Weibull distributions. Detailed probability density functions for all candidate distributions are provided in Appendix B. Only two parameter distributions were considered for use with SPI. While three parameter distributions that include a location parameter have been recommended by some (e.g. Guttman, 1999) because of their flexibility, these distributions permit negative values, introducing the problem of truncating the distribution at zero. The authors avoided comparing distributions with differing number of parameters (degrees of freedom) and are of the opinion that adding a location parameter is an unnecessary complication for the SPI, given relatively small historical sample sizes, previous recommendations of two parameter distributions (McKee *et al.*, 1993; Lloyd-Hughes and Saunders, 2002; Giddings *et al.*, 2005), and the positive results found herein (K–S rejection 5–9%, S–W rejection 3–13%).

Unlike distributions for SPI, candidate probability distributions for SPEI generally require a location parameter because climatic water balance is not bounded by zero and may take negative values if PET exceeds precipitation. Distributions considered for SPEI normalization include the (1) generalized logistic, (2) generalized extreme value (GEV), (3) normal, and (4) Pearson Type III distributions. A detailed explanation of these distributions is provided in Appendix C. The generalized logistic distribution is functionally equivalent to the recommended distribution in Vicente-Serrano *et al.* (2010), referred to there as the three-parameter log-logistic distribution. This distribution is referred to as the generalized logistic distribution herein to avoid confusion with the log transform of the logistic distribution and to be consistent with L-moment equations.

### 2.5. SPI and SPEI limits

SPI and SPEI values were limited to the range  $[-3,3]$  to ensure reasonableness. This issue has not been addressed explicitly in previous studies; however, the process of

calculating unbounded SPI and SPEI values based on the relatively limited historical record requires significant extrapolation and the associated uncertainty. For instance, by repeatedly sampling a typical SPI distribution based on 30 years of gamma-distributed precipitation, an SPI of  $-4$  has a 95% confidence interval of  $-5.3$  to  $-3.3$ , which equates to an event expected to occur 365 times in a period between either 2000 years or 15.5 million years. While this recurrence rate averages to a range of once in 5.7 to 47 000 years, this interpretation is potentially misleading. This is because the SPI/SPEI time series are highly auto-correlated, and extreme values tend to occur in multi-day or multi-month events. As such, the 365 occurrences (one for each day of the year) may very well occur during two or three drought events spread over the 2000 or 15.5 million year return period. This level of uncertainty (four orders of magnitude) inherent in an SPI estimate of  $-4$  is unreasonable. By contrast, estimates of less extreme drought events are more accurate because there is less extrapolation outside the existing historical data. Using the same precipitation distribution, the 95% confidence interval for an SPI of  $-1$  is  $-1.54$  to  $-0.67$ , a range of 4.0–16.2 years for 365 occurrences.

We therefore propose and implement bounds on the SPI and SPEI between  $-3$  and  $3$ . The likelihood of accumulated precipitation exceeding these values on a given day of the year is 0.14% or approximately once in 741 years. Given the 30-year reference period, this is a reasonable extrapolation limit. Events outside this range are not removed from the time series, but are designated  $<-3$  or  $>3$  to show that the event is extreme but cannot be accurately quantified. This is analogous to ‘below detection limit’ designations in laboratory experiments and has historical ties to the most extreme classes ( $<-2$  and  $>2$ ) proposed in McKee *et al.* (1993). True uncertainty bounds could be produced by bootstrapping; however, this is a computationally intensive task for large data sets. Our proposed limits are therefore a compromise between unrestricted extrapolation well outside the historical data and the computationally expensive task of computing bootstrap uncertainty estimates for all values of SPI or SPEI.

Placing reasonable limits on SPI/SPEI or providing uncertainty bounds is recommended for all future studies using these indices.

## 2.6. Zero precipitation

For SPI in regions with extremely low seasonal precipitation, there is potential for zero accumulated precipitation, particularly for short accumulation periods (1–3 months). In previous SPI studies (Lloyd-Hughes and Saunders, 2002; Wu *et al.*, 2007; Sienz *et al.*, 2012), periods with zero precipitation are handled by assigning SPI values based on the historical occurrence (%) of periods with zero precipitation:

$$p(x) = p_0 + (1 - p_0) F(x_{p>0}, \lambda) \quad (1)$$

where  $p$  represents the probability distribution for accumulated precipitation,  $x$ ,  $p_0$  equals the historical ratio of

periods with zero precipitation, and  $F(x, \lambda)$  represents the parametric univariate probability distribution fit to samples with detectable accumulated precipitation, using parameters  $\lambda$ . Using this method, hereafter referred to as the ‘former’ method, the  $p$  value for all periods with zero precipitation is set to the historical likelihood of zero precipitation, resulting in an SPI of  $\Phi^{-1}(p_0)$ , where  $\Phi^{-1}$  represents the inverse of the standard normal transform,  $\Phi$ .

The former method of handling zeros in the SPI is problematic because it assigns the maximum possible SPI value for the group of zeros, as shown in the left column of Figure 1. The SPI value assigned to all years with zero precipitation is shown by the large column immediately adjacent to the truncated normal distribution. As shown in Figure 1, using the former method truncates the SPI distribution at increasingly higher values and violates one of the primary definitions of the SPI: that its mean value is 0, corresponding to ‘typical’ conditions where 50% of years are wetter and 50% of years are drier. The resulting mean SPI for the former methodology is shown as a dashed line in Figure 1, which steadily increases above 0 as  $p_0$  increases. For extreme cases ( $p_0 > 50\%$ ), SPI will never fall below 0, creating the illogical situation that all periods with no detectable precipitation are ‘wetter’ than is typical for the region.

This methodological issue is addressed in this paper by assigning SPI values for zero precipitation based on the ‘centre of mass’ of the zero distribution rather than the maximum probability, thereby producing SPI values that maintain statistical interpretability (mean,  $\mu = 0$ ). The concept of ‘probability mass’ when normalizing zeros in the SPI was discussed in Solakova *et al.* (2014), although the implementation in this paper eventually used the former method discussed above. In the ‘proposed’ method for normalizing zero precipitation, the likelihood of zero precipitation is calculated based on the empirical cumulative distribution. The Weibull plotting position function has been shown to be the exact cumulative probability of non-exceedance in ranked observations (Makkonen, 2006; Makkonen, 2008), and takes the form:

$$p_0 = \frac{n_{p=0}}{n + 1} \quad (2)$$

where  $p_0$  represents the probability of zero precipitation,  $n_{p=0}$  represents the number of samples in the reference period where total precipitation equals zero, and  $n$  represents the total number of samples in the reference period. Because the precipitation is equal for all zero precipitation periods, and without additional information regarding drought severity, then the drought severity likelihood could be as low as  $1/(n + 1)$  or as high as  $n_{p=0}/(n + 1)$ . Therefore, when these Weibull non-exceedance probabilities are averaged, the resulting centre of probability mass for multiple zeros becomes:

$$\bar{p}_0 = \frac{n_{p=0} + 1}{2(n + 1)} \quad (3)$$

where  $\bar{p}_0$  represents the mean probability of multiple zeros based on the Weibull plotting position function. This value

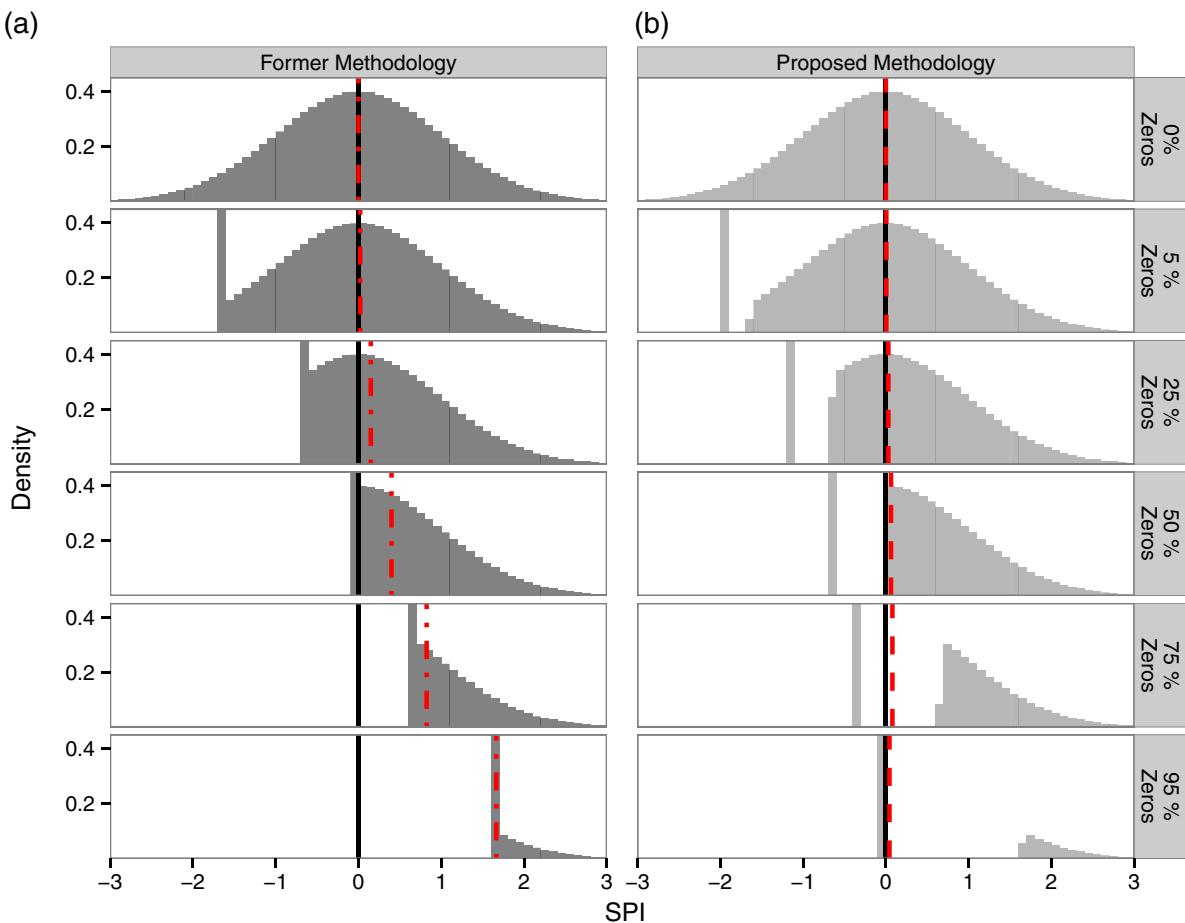


Figure 1. Simulated SPI distributions with increasing proportions of zero precipitation using the former methodology (a) and the proposed methodology (b).

is then used to calculate the SPI for the centre of probability mass for zero precipitation. The resulting piecewise probability distribution is then:

$$p(x) = \begin{cases} p_0 + (1-p_0) F(x_{p>0}, \lambda), & x > 0 \\ \frac{n_{p=0}+1}{2(n+1)}, & x = 0 \end{cases} \quad (4)$$

where  $p$  is the probability distribution and  $F(x, \lambda)$  represents the parametric univariate probability distribution for samples with detectable accumulated precipitation, fit with parameters  $\lambda$ . In this manner, SPI values for periods with detectable precipitation are identical to those calculated by Equation (1), but the likelihood of zero precipitation has been adjusted using the proposed methodology based on the mean Weibull plotting position,  $\bar{p}_0$ . As shown on right column of Figure 1, the proposed methodology maintains a mean SPI value of 0, retaining statistical interpretability through the entire range of  $p_0$ . SPI variance falls below 1 as  $p_0$  increases, which is an unavoidable consequence of any methodology. This also means that the minimum SPI value is limited by the proportion of zeros. However, the value presented is the true most likely value, with larger uncertainty bounds due to the inability to differentiate drought severity between periods with equal precipitation. At the most extreme ( $p_0 = 1$ ) conditions, the proposed solution collapses to apply an SPI of 0 to all periods with a large

uncertainty. While not informative for comparison purposes, this solution can be interpreted that when all periods have zero precipitation, zero precipitation is the typical climate condition ( $SPI = 0$ ). This makes statistical sense, whereas the former methodology approaches an SPI of infinity, which has no meaning. This proposed methodology, using the centre of probability mass, has been implemented in the SCI R package (Gudmundsson and Stagge, 2014). Alternative zero handling methods for calculating SPI have been proposed (Wilks, 1990), but have not gained widespread acceptance.

For the WFD, the threshold for zero precipitation is 0.01 mm. In this study, the most arid land cell in the driest month of the year (August) produced 23 months with zero precipitation out of the 30-year reference period (76%), when calculated for the single month SPI1. The remaining cells, months, and accumulation periods had significantly fewer periods with zero precipitation. At the SPI6 level, there were no land cells with zero precipitation. While in practice, one might consider removing cells with an excess of zero precipitation from the data set, this study did not censor the data because, for comparison purposes, all distributions were fit to the same underlying precipitation data with an equal number of zero values. Because climatic water balance can take on both positive and negative

values, creating a piecewise distribution is not required for the SPEI, making SPEI distribution fitting much simpler.

## 2.7. Evaluation criteria

Candidate distributions were tested and compared using several metrics. First, goodness of fit was calculated by the K–S test (Massey, 1951; Stephens, 1974), similar to a majority of SPI and SPEI studies. The K–S test calculates the maximum difference between the empirical cumulative distribution of sampled points and the theoretical cumulative distribution of the candidate distribution function. The A–D test (Anderson and Darling, 1954), a modified version of the K–S test with greater emphasis on the distribution tails, was also used to test goodness of fit. While the K–S test is often considered distribution-free, this is not true when applied to distributions estimated from the data (Lilliefors, 1967; Crutcher, 1975; Steinskog *et al.*, 2007), as is the case for this study and the majority of goodness-of-fit applications. The A–D test is also distribution specific, requiring calculation of critical values when values are not previously tabulated. Critical values for the K–S and A–D tests were therefore calculated using Monte Carlo bootstrap simulations. Where zero precipitation occurred in SPI testing, goodness-of-fit tests were applied only to the positive samples (i.e. the distribution that has been fit).

The S–W test (Shapiro and Wilk, 1965; Stephens, 1974) was used to test whether the calculated SPI and SPEI values are normally distributed. This test is based on the assumption that the resultant drought indices should be normally distributed ( $\mu = 0, \sigma = 1$ ) and independently sampled, as each distribution is fit based on a given day in different years. This test is based on a method proposed in Wu *et al.* (2007) and used subsequently (Kumar *et al.*, 2009) to test the suitability of SPI values in arid regions with zero precipitation. This study is the first use of this methodology as an alternative goodness-of-fit comparison for SPI/SPEI candidate distributions. The benefit of this test is that it directly tests the final index values, is independent of the candidate distribution, and has well-reviewed and sensitive critical values. Index values used for the S–W test were found to have no significant temporal autocorrelation, satisfying the requirement for this test that values should be independent and identically distributed. In the case of zero precipitation, where a single SPI value is applied for all zero precipitation samples, normally distributed index values are artificially generated and substituted for samples with zero precipitation prior to S–W normality testing. This ensures that the rejection of acceptance of normality is based only on the fit of years with detectable precipitation. While this decreases the likelihood of rejecting periods with many zeros, this is reasonable, given the high uncertainty of the true underlying probability distribution.

Distributions were further compared relative to one another based on the AIC (Akaike, 1974; Burnham and Anderson, 2004). AIC is commonly used to make relative comparisons, but is not a statistical test, and therefore cannot provide information about the absolute goodness of

fit. AIC is based on maximized likelihood, with an additional penalty related to the number of model parameters, included to test the balance between goodness of fit and model complexity. Except for the normal distribution in SPEI, all distributions have the same number of model parameters (two for SPI, three for SPEI), which reduces the AIC to a measure of log likelihood.

## 3. Results and discussion

### 3.1. SPI distribution fitting

Candidate distributions for SPI form two distinct groups with respect to the S–W test for normality, used to test goodness of fit. Figure 2 shows the S–W rejection frequency ( $\alpha = 5\%$ ), defined as the ratio of rejected distributions across all land surface cells in the domain and 365 daily-fitted distributions for each tested accumulation period (in total 1 440 000 distributions per accumulation period). The two groups can be classified as (1) long-accumulation distributions, of which the gamma distribution dominates, and (2) short-accumulation distributions, Weibull and Gumbel, of which the Weibull is consistently the best. When the two dominant distributions are directly compared, the gamma distribution performs best for all accumulation periods (3.71–9.24% rejection), except for the SPI-1, where it remains the second best distribution (13.36% rejection). A nearly identical pattern is shown when  $\Delta\text{AIC}$ , the difference in AIC between a distribution and the distribution with the best fit, is plotted (Figure 2). The K–S and A–D tests show identical patterns and so are only presented as Supporting Information.

The gamma distribution's general success is attributed to its relatively flexible shape parameter, which is clearly suited to the range of accumulated precipitation distributions in Europe. This is consistent with the findings by Lloyd-Hughes and Saunders (2002), which also recommend the two-parameter gamma distribution for use in normalizing SPI values in Europe, and by McKee *et al.* (1993), which uses the gamma distribution in their US example. While Guttman (1999) recommends the three-parameter gamma distribution (Pearson Type III) for 1035 sites in the contiguous United States, Giddings *et al.* (2005) notes that its improvement relative to the two-parameter gamma distribution is minor. While the authors are unaware of any studies providing a physical basis for gamma-distributed rainfall at the monthly scale, many studies (Mooley, 1973; Vlček and Huth, 2009) use a gamma distribution to model monthly, daily, and sub-daily precipitation. It follows that when daily gamma-distributed rainfall amounts are summed over 30–365 days, the resulting accumulated values should follow a similarly right-skewed gamma distribution, albeit with a less extreme tail. With longer accumulation periods, the Central Limit Theorem states that this distribution should eventually approach a normal distribution, as seen using this data.

Rejection frequencies for the gamma distribution in this study approach the random Type I error (5%, shown as a

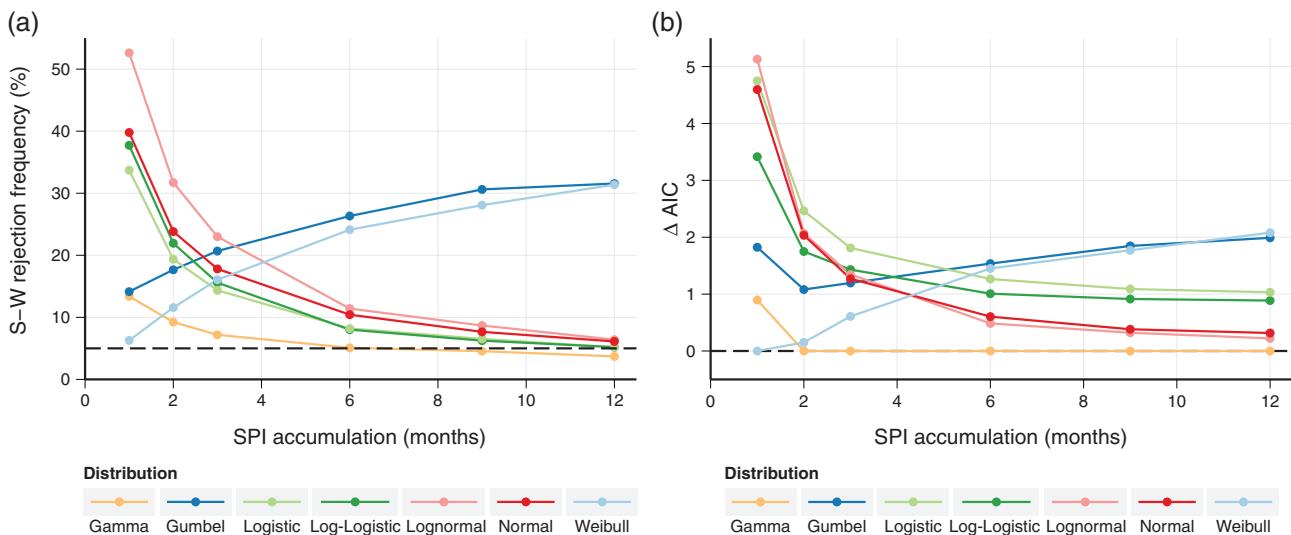


Figure 2. Shapiro–Wilk rejection frequency (%), a) and  $\Delta\text{AIC}$  (b) for SPI candidate distributions across all tested accumulation periods. Expected Type I error (5%) is shown by a dashed line.

dotted line in Figure 2) for accumulation periods longer than 2 months. A Type I error, often termed a ‘false positive’, occurs when one incorrectly rejects the null hypothesis and the rate at which this occurs is set by the alpha ( $\alpha$ ) level. In this example, samples randomly selected from a true gamma distribution would be rejected by the S–W test as being non-gamma distributed in 5% of the samples, on average. The Weibull distribution is only preferred over the gamma distribution at the 1-month accumulation period, but produces unacceptable failures ( $>25\%$ ) for accumulation periods equal to or greater than 6 months (Figure 2). This very specific success is attributed to its ability to model highly skewed distributions, with a sharp, asymptotic density near the zero bound; conditions that typically occur only during accumulation periods of 1 or 2 months. There was no seasonal pattern among rejection rates for any accumulation period, which allows a single distribution to model the entire year’s precipitation deficit. When precipitation is summed for 1 or more months, the differences in rainfall generation processes between frontal and convective rain events do not affect the SPI distributions.

It is important to note that the K–S test, commonly used as validation for distribution selection, is the least sensitive of the three goodness-of-fit tests (Table 1). K–S rejection rates are, on average, 7% lower than S–W rejection rates. This is particularly noticeable for highly skewed distributions (SPI-1 and 2), likely related to the widely recognized insensitivity of the K–S for differences in the tails of the distribution (Stephens, 1974; Mason and Schuenemeyer, 1983). This insensitivity in the tails occurs because, by definition, the empirical cumulative distribution and candidate distributions converge to 0 and 1 in the tails. This is especially problematic when using the K–S test to compare distributions for use in drought indices, where the distribution’s tails are of the most concern. The more sensitive A–D test, a variation of the K–S which gives more weight to the distribution tails, produced the

same general pattern with rejection rates that generally fell between the other tests (Table 1). The consistent pattern across all three goodness-of-fit tests and the  $\Delta\text{AIC}$  lends confidence to distribution recommendations and demonstrates that the S–W test can accurately discriminate between distributions, without the need for bootstrapping critical values, as in the K–S and A–D tests.

While the gamma distribution fits best for all averaging intervals apart from the SPI-1, it is important to investigate regional patterns, identifying regions and times of the year where or when it fails to adequately fit the accumulated precipitation data, potentially introducing a bias to the SPI. When S–W rejection frequencies for the gamma and Weibull distributions are plotted spatially (Figure 3), it is apparent that the gamma distribution has broad effectiveness across most of Europe and that the two distributions (gamma and Weibull) are directly opposed in some cases/regions. In isolated coastal regions of Denmark, northern France, and Greece, the gamma distribution is less capable of fitting the data (15–40% rejection), whereas the Weibull distribution produces rejection frequencies between 0 and 20% in these regions.

Relative distribution rankings (Figure 4) support the conclusions drawn from Figure 2, that the better fit of short-accumulation distributions (Weibull and Gumbel) is related to their ability to model highly skewed precipitation distributions where most values reside near the zero bound. This is confirmed by their effectiveness at low accumulation periods (SPI-1 and 2) for large regions of Europe and continued utility in isolated semi-arid portions of the Mediterranean (Turkey, Greece, and Spain) also for SPI-6 and 9.

The progression of relative AIC rankings shows that the long-accumulation distributions (gamma, normal, and lognormal) begin to dominate in more temperate climates at accumulation periods above 3 or 6 months. Again, the gamma distribution is the only distribution capable

Table 1. Kolmogorov–Smirnov (K–S), Anderson–Darling (A–D), and Shapiro–Wilk (S–W) rejection frequencies (%) for all tested distributions and SPI accumulation periods.

Distribution	SPI-1			SPI-2			SPI-3			SPI-6			SPI-9			SPI-12		
	K–S	A–D	S–W	K–S	A–D	S–W												
Gamma	8.76	10.29	13.36	7.08	8.03	9.24	6.64	7.28	7.18	5.32	5.60	5.09	5.53	5.50	4.54	5.14	5.18	3.71
Gumbel	10.43	11.00	14.14	11.48	13.95	17.66	12.87	15.12	20.69	13.76	17.73	26.33	16.29	20.24	30.62	17.21	20.82	31.56
Log-logistic	11.99	22.57	37.72	8.54	13.80	21.94	7.77	11.14	15.58	6.09	6.39	8.02	6.08	5.69	0.35	6.12	5.54	5.20
Lognormal	26.08	35.61	52.62	15.94	21.35	31.72	12.79	16.24	23.00	7.61	8.87	11.41	6.95	7.61	8.68	6.20	6.85	6.40
Logistic	10.00	19.15	33.72	7.90	12.21	19.35	7.39	10.62	14.35	6.07	6.76	8.20	6.40	6.08	6.53	6.45	5.65	5.16
Normal	17.96	24.28	39.79	12.41	16.38	23.83	12.41	12.95	17.81	7.89	8.88	10.43	7.89	7.71	7.66	6.54	6.91	6.11
Weibull	6.04	6.41	6.28	7.96	9.46	11.58	9.55	11.77	16.06	11.80	15.69	24.12	13.88	18.34	28.08	15.01	20.24	31.37

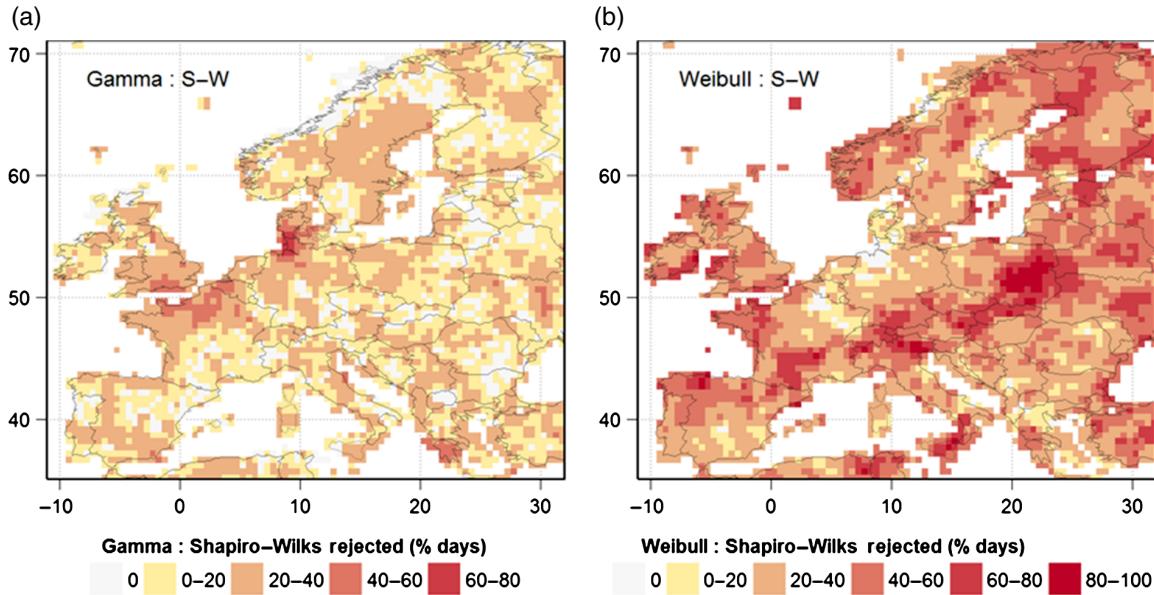


Figure 3. Spatial comparison of SPI-6 Shapiro–Wilk rejection frequencies (%) for the gamma (a) and Weibull (b) distributions.

of producing good relative fit across all accumulation periods and regions, showing its general effectiveness for producing accurate SPI values in Europe.

### 3.2. SPEI distribution fitting

Candidate distribution comparisons for SPEI show that the GEV distribution consistently produces the best goodness of fit across all accumulation periods (rejection frequency 2.48–2.92%), followed by the generalized logistic and Pearson Type III, which both have similar rejection frequencies (4.42–6.83%, Figure 5). All candidate SPEI distributions have lower rejection frequencies than those for SPI, suggesting that the distribution of accumulated climatic water balance is easier to fit with the available parametric distributions. This is partially due to the additional location parameter, but more likely related to the unbounded nature of climatic water balance, which avoids the problem of fitting a piecewise distribution bounded by zero. This assumption is supported by the relatively minor increase at low accumulation (1 and 2 months) for SPEI (Figure 5) compared with large increases in distribution fitting failures for the SPI (Figure 2) and also explains why no low-accumulation/high-accumulation groupings form within the SPEI rejection frequencies.

Selection of the GEV distribution as the preferred SPEI distribution differs from the previous analysis of SPEI candidate distributions (Vicente-Serrano *et al.*, 2010), which recommended the generalized logistic distribution, found here to be the second best. However, the recommendation in Vicente-Serrano *et al.* (2010) was based solely on L-moment diagrams because the K–S test was unable to distinguish between the candidate distributions. In this study, the K–S and A–D tests were both able to distinguish among the distributions, producing relative rankings similar to the S–W test (Table 2). The GEV distribution is again the recommended distribution, with rejection rates near  $\alpha = 0.05$ . Rejection rates for the GEV distribution using the S–W test were below the theoretical rejection rate, which may show some loss of power, but the overall ranking is confirmed by the K–S and A–D tests (Table 2). As with SPI testing, the S–W test was strictest, producing the highest rejection rates and successfully differentiating between the distributions (rejection frequencies of 2.48–6.78%).

Spatially, the GEV distribution provides the best fit for the majority of Europe across all accumulation periods (Figure 6). S–W rejection frequencies follow similar patterns as the AIC maps, although patterns are less

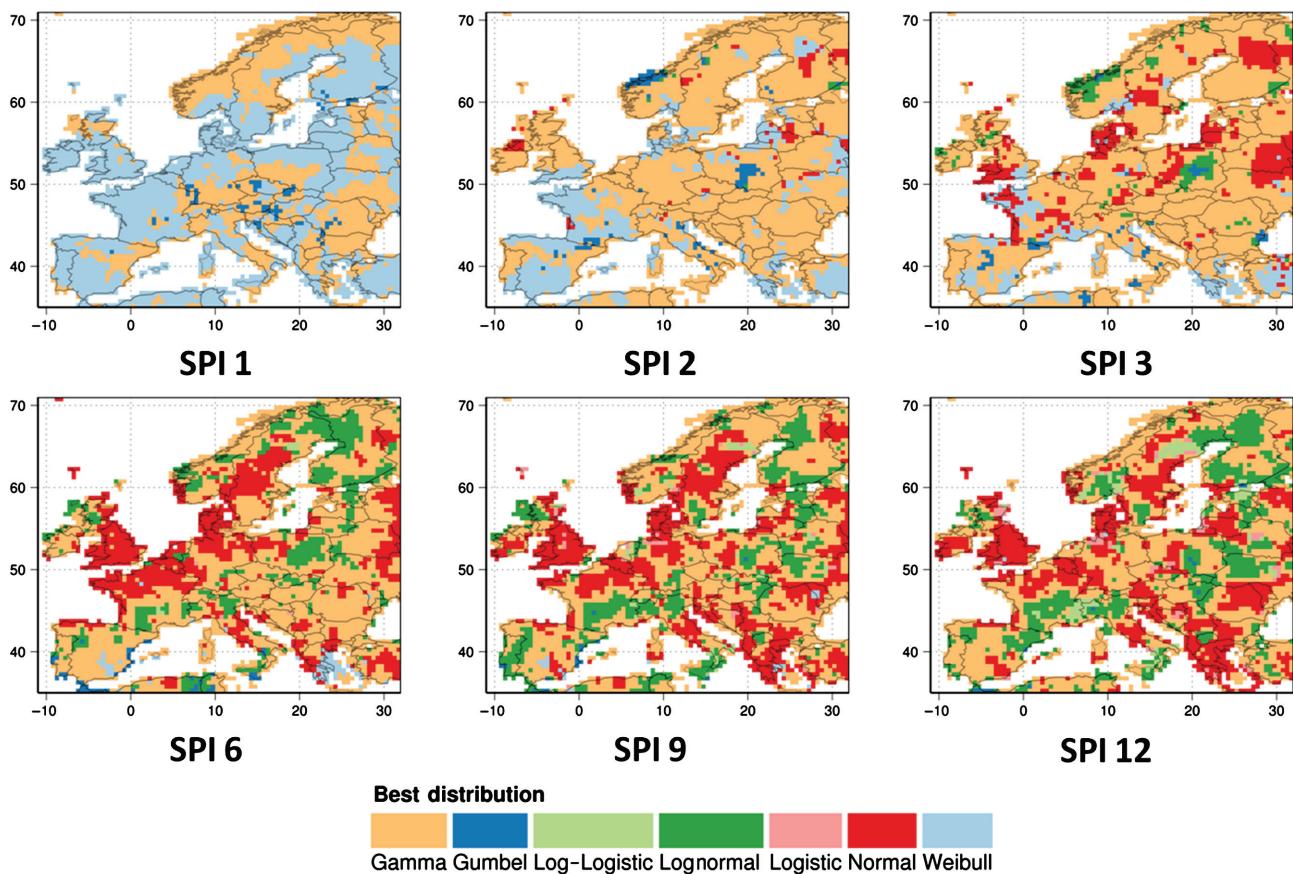


Figure 4. Spatial comparison of the SPI distribution with best relative fit (calculated by AIC) for the majority of tested time steps.

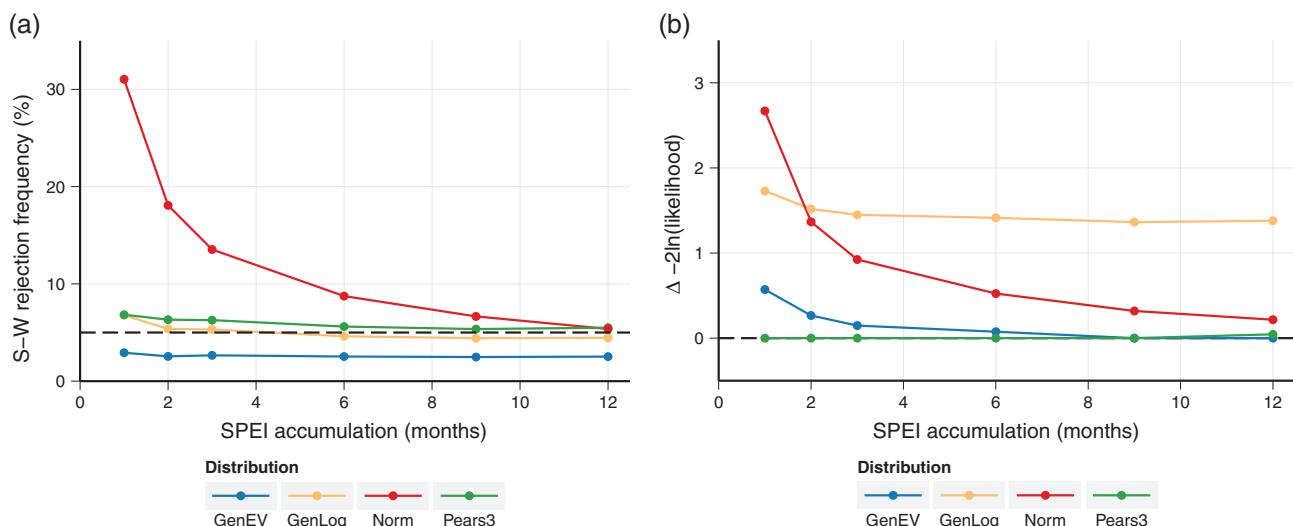


Figure 5. Shapiro–Wilk rejection frequency (%), a) and  $\Delta \log$  (likelihood) (b) for SPEI candidate distributions across all tested accumulation periods. Format is similar to Figure 2 with expected Type I error (5%) is shown by a dashed line; however,  $\log(\text{likelihood})$  (b) is presented instead of AIC to show the best possible model fit, regardless of the number of parameters. An equivalent plot showing AIC is included in Supporting Information.

pronounced and are therefore not presented. The Pearson Type III distribution may have some utility for low accumulation periods (<3 months) in semi-arid regions, as shown by its relative performance around the Mediterranean region (Figure 6) and at low accumulation periods with respect to AIC (Figure 5); however, this benefit

is extremely localized and this distribution should not be applied generally to Europe because of its poor fit elsewhere. Similarly, at high accumulation periods the normal distribution has some value for temperate zones, but its inflexibility and inability to capture short accumulation periods (Figure 5) render it useless as a general

Table 2. Kolmogorov–Smirnov (K–S), Anderson–Darling (A–D), and Shapiro–Wilk (S–W) rejection frequencies (%) for all tested distributions and SPEI accumulation periods.

Distribution	SPEI-1			SPEI-2			SPEI-3			SPEI-6			SPEI-9			SPEI-12		
	K–S	A–D	S–W	K–S	A–D	S–W	K–S	A–D	S–W	K–S	A–D	S–W	K–S	A–D	S–W	K–S	A–D	S–W
Gen Logistic	5.80	6.27	6.78	5.44	5.58	5.37	5.33	5.24	5.32	5.33	5.22	4.61	4.94	4.74	4.42	5.52	4.44	4.45
GEV	5.13	4.47	2.92	5.16	4.61	2.55	5.05	4.79	2.66	4.85	4.77	2.53	4.63	3.44	2.48	4.75	3.66	2.52
Pearson III	6.43	6.67	6.83	6.72	7.00	6.32	6.78	7.14	6.27	6.61	7.16	5.61	6.44	7.07	5.36	6.86	7.36	5.50
Normal	15.76	20.67	31.04	11.32	15.09	18.08	8.82	10.71	13.54	7.16	7.55	8.75	6.33	6.63	6.66	5.30	5.83	5.37

recommendation. Interestingly, despite having the second best performance with respect to the S–W test, the generalized logistic distribution does not dominate in any particular region (Figure 6). This may be caused by its lack of unique properties, which causes it to be masked by the similar, but better performing, GEV distribution. This phenomenon may also explain why the relative rankings by AIC differ from all three goodness-of-fit tests. These tests penalize fitting failures but do not measure differences between successful fits, while the AIC measures the average goodness of fit. Mean AIC values are so similar that they are essentially indistinguishable, suggesting the need to focus on goodness-of-fit tests and the spatial AIC comparisons, which all support the GEV distribution.

#### 4. Conclusions

The SPI is already a well-established and important index used to quantify and compare meteorological drought throughout the world and its newer climatic water balance variant, the SPEI, has promise for similar use provided it receives adequate testing. This study evaluated the selection of an appropriate univariate probability distribution for use in normalizing SPI and SPEI values in Europe. While these distribution recommendations are shown to be successful general recommendations for a continental scale analysis of Europe, it is important to use the proposed methodology to verify goodness of fit when working in a new region or with a new data set. Improper probability distributions have the potential to bias drought index values, exaggerating or minimizing the perceived severity of drought events.

Following extensive statistical testing and relative comparisons, for regional studies in Europe we recommend:

- 1 The two-parameter gamma distribution when calculating SPI, and
- 2 The GEV distribution when calculating SPEI.

Selection of the two-parameter gamma distribution for calculating SPI agrees with conclusions from previous studies, but the analysis here provides further support for this conclusion. The gamma distribution was found to have broad effectiveness in Europe, providing the best fit for accumulation periods from 2 to 12 months and the second best fit for 1 month, with consistently good

fit across the range of European climates. SPI distributions organize naturally into two groups, characterized as short-accumulation (1–3 months) and long-accumulation (>6 months) distributions. While the gamma distribution follows the long-accumulation pattern, it produces the most consistently good fit for all accumulation periods, except for 1 month, where the Weibull distribution, best among the short-accumulation distributions, slightly dominates. However, the Weibull distribution is only effective within a very narrow range of conditions, when accumulated precipitation is highly skewed for short (1–2 months) accumulation periods or in semi-arid regions, and is very poor for the remaining long-accumulation or temperate climates.

All candidate SPEI distributions produced better fits than the candidate SPI distributions because climatic water balance is not bounded by zero, removing difficulties in fitting piecewise distributions around this limit. The recommended distribution for use in calculating SPEI is the GEV distribution, which differs from the generalized logistic distribution recommended in the original SPEI analysis (Vicente-Serrano *et al.*, 2010). Short accumulation (1–2 months) periods remained the most difficult to fit, but this difference was less pronounced than in SPI goodness-of-fit tests.

While some have argued for a ‘multi-distribution’ approach (Sienz *et al.*, 2012), selecting different distributions for different regions and different times of the year, this approach adds a level of complexity in determining when to use new distributions and does not allow for comparisons across space and time. For instance, it would not be possible to examine how the scale parameter changes throughout the year for a site or how the shape parameter differs across Europe. In addition, while different distributions may appear similar in the central portion of the distribution, there are often larger differences in the tails, which could produce sudden changes in index values at the extremes when changing distributions. Given the statistical rejection levels near alpha for the recommended distributions, it is our recommendation to use a single, simple distribution and avoid interpretability issues and potential extrapolation issues with complex distributions that maybe overfit.

In addition to probability distribution recommendations for the SPI and SPEI in Europe, several methodological improvements are proposed, including use of the S–W test of normality, rather than the K–S test, to evaluate goodness of fit and use of the ‘centre of mass’ when calculating

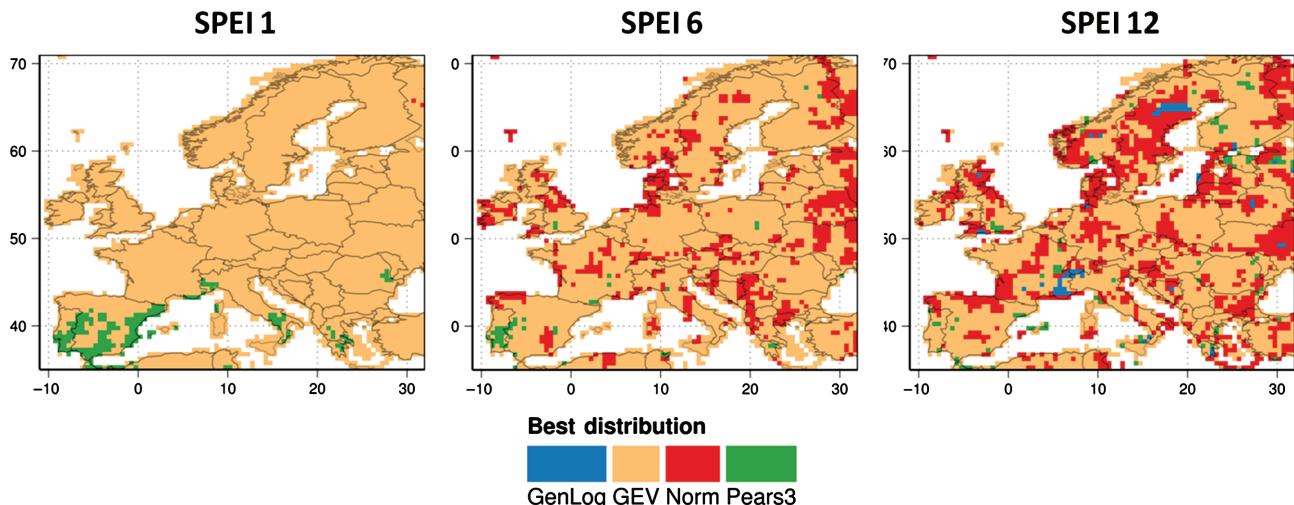


Figure 6. Spatial comparison of the SPEI distribution with best relative fit (calculated by AIC) for the majority of tested time steps, as in Figure 4.

SPI for periods with zero precipitation. The proposed S–W test for normality testing of the final calculated drought index was shown to effectively differentiate among the candidate distributions, producing similar results as the K–S and A–D tests with more sensitive results. The K–S test, used in previous studies, was found to be the least sensitive to proposed distributions. More importantly, the S–W test is simple and quick to implement, avoiding the need for Monte Carlo simulations of critical test values. As further evidence, the proposed S–W test produced rejection frequencies approaching the theoretical, random Type I error and generated spatial patterns similar to relative rankings calculated by AIC. The interpretability problem with applying the maximum exceedance probability to periods with zero precipitation was described in this paper and addressed by applying the ‘centre-of-mass’ adjustment.

Several general recommendations were demonstrated, though not tested, including the use of Penman–Monteith evapotranspiration when calculating SPEI, use of a daily timestep when data are available, and the use of limits on SPI/SPEI values based on the length of historical time series. Many of these methodological improvements have been implemented in the SCI R package (Gudmundsson and Stagge, 2014). Use of the FAO-56 Penman–Monteith equation with the Hargreaves modification was used as a more accurate, physically based measure of PET when calculating SPEI as recommended in Stagge *et al.* (2014). Setting limits on the SPI and SPEI index values ensures reasonable statistical values and removes the problem of extrapolation outside the historical range. Given the extensive testing provided herein, the SPEI warrants further research attention and expanded use as a reasonable and useful drought index.

## Acknowledgements

This study was financially supported by the EU FP7 project DROUGHT-R&SPI (contract no. 282769).

## Appendix A

### Modified Fao-56 Penman–Monteith equation

PET for this study is calculated by the modified FAO-56 Penman–Monteith equation. For a detailed explanation, see Allen *et al.* (1998). The modified Penman–Monteith equation reads:

$$\text{PET} = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T+273} u_2 (e_s - e_a)}{\Delta + \gamma (1 + 0.34u_2)} \quad (\text{A1})$$

where PET is measured in  $\text{mm day}^{-1}$ ,  $R_n$  represents net radiation ( $\text{MJ m}^{-2} \text{ day}^{-1}$ ),  $G$  is the soil heat flux density ( $\text{MJ m}^{-2} \text{ day}^{-1}$ ),  $\gamma$  is the psychrometric constant ( $\text{kPa}^\circ\text{C}^{-1}$ ),  $T$  is the mean daily temperature ( $^\circ\text{C}$ ),  $u_2$  is wind speed at 2 m ( $\text{m s}^{-1}$ ),  $\Delta$  is the slope of the saturation vapour pressure *versus* temperature curve at the given air temperature ( $\text{kPa}^\circ\text{C}^{-1}$ ), and  $e_s - e_a$  represents the saturation vapour pressure deficit (kPa).

The radiation term is based on the difference between net radiation at the crop surface,  $R_n$ , and soil heat flux density,  $G$ . Soil heat flux,  $G$ , may be ignored for 24-h time steps, as the magnitude is relatively small. Net radiation,  $R_n$ , is then calculated as the difference between incoming net shortwave ( $R_{ns}$ ) and net outgoing longwave ( $R_{nl}$ ) radiation:

$$R_n = R_{ns} - R_{nl} \quad (\text{A2})$$

Net incoming shortwave radiation is calculated as:

$$R_{ns} = (1 - 0.23) R_s \quad (\text{A3})$$

based on the assumed albedo of 0.23 of the reference crop, where shortwave radiation is calculated based on the Hargreaves and Samani (1985) approximation:

$$R_s = k_{R_s} \sqrt{(T_{\max} - T_{\min}) R_a} \quad (\text{A4})$$

In this equation,  $T_{\max}$  and  $T_{\min}$  represent maximum and minimum daily temperature ( $^\circ\text{C}$ ),  $R_a$  represents extraterrestrial radiation, and  $k_{R_s}$  is an adjustment coefficient. For

the purpose of this study,  $k_{R_s}$  is assumed to be 0.16 for the entire study area, though a value of 0.19 may be more applicable for ‘coastal’ locations. Atmospheric radiation,  $R_a$ , is calculated from latitude and Julian day by

$$\varphi_r = \frac{\pi}{180} \varphi_d \quad (A5)$$

$$d_r = 1 + 0.033 \cos\left(\frac{2\pi}{365}J\right) \quad (A6)$$

$$\delta = 0.409 \sin\left(\frac{2\pi}{365}J - 1.39\right) \quad (A7)$$

$$\omega_s = \cos^{-1}(-\tan[\varphi_r] \tan[\delta]) \quad (A8)$$

$$R_a = \frac{24 \times 60}{\pi} G_{sc} d_r [\omega_s \sin(\varphi_r) \sin(\delta) + \cos(\varphi_r) \cos(\delta) \sin(\omega_s)] \quad (A9)$$

where  $R_a$  is the extraterrestrial radiation ( $\text{MJ m}^{-2} \text{day}^{-1}$ ),  $\varphi_d$  the latitude (decimal degrees),  $\varphi_r$  the latitude (radians),  $J$  the Julian day,  $d_r$  the inverse relative distance of Earth–Sun,  $\delta$  the solar declination,  $\omega_s$  the sunset hour angle,  $G_{sc}$  the solar constant = 0.0820  $\text{MJ m}^{-2} \text{min}^{-1}$ .

Net outgoing longwave ( $R_{nl}$ ) radiation is calculated from temperature, vapour pressure, and relative shortwave radiation by:

$$R_{nl} = \sigma \left[ \frac{T_{\max,K} - T_{\min,K}}{2} \right] [0.34 - 0.14\sqrt{e_a}] \\ \left[ 1.35 \frac{R_s}{R_{so}} - 0.35 \right] \quad (A10)$$

where  $\sigma$  is the Stefan–Boltzmann constant ( $4.93 \times 10^{-9} \text{ MJ K}^{-4} \text{m}^{-2} \text{day}^{-1}$ ),  $T_{\max,K}$  and  $T_{\min,K}$  are daily maximum and minimum temperature (K),  $e_a$  is the actual vapour pressure (kPa), and  $R_s/R_{so}$  represents the relative shortwave radiation. Clear sky solar radiation,  $R_{so}$ , is calculated by:

$$R_{so} = (0.75 + 2 \times 10^{-5}z) R_a \quad (A11)$$

where  $z$  is the elevation above sea level (m). Saturation vapour pressure is related to air temperature and can be calculated using the following relationship:

$$e^o(T) = 0.611 e^{\left(\frac{17.27T}{T+237.3}\right)} \quad (A12)$$

where  $T$  is given in °C. In lieu of dewpoint temperature measurements, the daily minimum temperature,  $T_{\min}$ , may be used estimate the actual vapour pressure using this relationship:

$$e_a = e^o(T_{\min}) = 0.611 e^{\frac{17.27T_{\min}}{T_{\min}+237.3}} \quad (A13)$$

Similarly, mean daily saturation vapour pressure should be estimated as the mean between saturation vapour pressure at the daily maximum and minimum temperature:

$$e_s = \frac{e^o(T_{\max}) + e^o(T_{\min})}{2} \quad (A14)$$

Wind speed at 2 m is estimated from the WFD wind speed at 10 m using the power law:

$$\mu_2 = \mu_z \frac{4.87}{\ln(67.8z - 5.42)} \quad (A15)$$

where  $\mu_2$  is the wind speed at 2 m above ground level ( $\text{m s}^{-1}$ ) and  $\mu_z$  is the wind speed at  $z$  m above ground level ( $\text{m s}^{-1}$ ). In this case  $z = 10$  m. Finally, the saturation vapour pressure *versus* temperature curve at the given air temperature psychrometric constant,  $\Delta$ , is calculated based on mean daily temperature ( $T$ , °C):

$$\Delta = \frac{4098 \left[ 0.6108 e^{\frac{17.27T}{T+237.3}} \right]}{(T + 237.3)^2} \quad (A16)$$

and the psychrometric constant,  $\gamma$ , is calculated based on atmospheric pressure,  $P$ , where:

$$\gamma = \frac{1.013 \times 10^{-3}P}{0.622 \times 2.45} \quad (A17)$$

and

$$P = 101.3 \left( \frac{293 - 0.0065z}{293} \right)^{5.26} \quad (A18)$$

## Appendix B: SPI candidate distributions

### Gamma distribution

$$f(x) = \frac{1}{\alpha^\beta \Gamma(\beta)} x^{\beta-1} e^{-x/\alpha}, \quad x > 0 \quad (B1)$$

where

$$\Gamma(c) = \int_0^\infty e^{-x} x^{c-1} dx \quad (B2)$$

### Gumbel distribution

$$f(x) = \left( \frac{1}{\sigma} \right) e^{-z(x)-e^{-z(x)}}, \quad -\infty < x < \infty \quad (B3)$$

where

$$z(x) = \frac{x - \mu}{\sigma} \quad (B4)$$

### Log-logistic distribution

$$f(x) = \frac{\lambda \kappa (\lambda x)^{\kappa-1}}{\left[ 1 + (\lambda x)^\kappa \right]^2}, \quad x > 0 \quad (B5)$$

### Lognormal distribution

$$f(x) = \frac{1}{\beta x \sqrt{2\pi}} e^{\left[ -\frac{1}{2} \left( \log(x/\alpha)/\beta \right)^2 \right]}, \quad x > 0 \quad (B6)$$

### Logistic distribution

$$f(x) = \frac{\lambda^\kappa \kappa e^{\lambda x}}{\left[ 1 + (\lambda e^x)^\kappa \right]^2}, \quad -\infty < x < \infty \quad (B7)$$

### Normal distribution

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2 \right], \quad -\infty < x < \infty \quad (B8)$$

## Weibull distribution

$$f(x) = \left(\frac{\beta}{\alpha}\right)x^{\beta-1} \exp\left[-\left(\frac{1}{\alpha}\right)x^\beta\right], \quad x > 0 \quad (\text{B9})$$

## Appendix C: SPEI candidate distributions

### Generalized logistic distribution

$$f(x) = \begin{cases} \Delta \frac{(1+\xi z(x))^{-1-1/\xi}}{\sigma(1+(1+\xi z(x))^{-1/\xi})^2}, & \xi \neq 0, 1 + \xi z(x) > 0 \\ \Delta \frac{e^{-z(x)}}{\sigma(1+e^{-z(x)})^2}, & \xi = 0, -\infty < x < \infty \end{cases} \quad (\text{C1})$$

where

$$z(x) = \frac{x - \mu}{\sigma} \quad (\text{C2})$$

### GEV distribution

$$f(x) = \begin{cases} \left(\frac{1}{\sigma}\right)(1 + \xi z(x))^{-1/\xi-1} e^{-(1+\xi z(x))^{-1/\xi}}, & \xi \neq 0, 1 + \xi z(x) > 0 \\ \left(\frac{1}{\sigma}\right) e^{-z(x)-e^{-z(x)}}, & \xi = 0, \\ -\infty < x < \infty \end{cases} \quad (\text{C3})$$

where

$$z(x) = \frac{x - \mu}{\sigma} \quad (\text{C4})$$

### Pearson Type III distribution

$$f(x) = \frac{1}{a\Gamma(b+1)} \left(\frac{x-m}{a}\right)^b e^{-(x-m)/a}, \quad x > 0 \quad (\text{C5})$$

### Normal Distribution

See Appendix B.

### Supporting Information

The following supporting information is available as part of the online article:

**Figure S1:** Kolmogorov–Smirnov rejection frequency (%) for SPI candidate distributions across all tested accumulation periods. Expected Type I error (5%) is shown by a dashed line. Scale is identical to Figure 2.

**Figure S2:** Anderson–Darling rejection frequency (%) for SPI candidate distributions across all tested accumulation periods. Expected Type I error (5%) is shown by a dashed line. Scale is identical to Figure 2.

**Figure S3:** Kolmogorov–Smirnov rejection frequency (%) for SPEI candidate distributions across all tested accumulation periods. Expected Type I error (5%) is shown by a dashed line. Scale is identical to Figure 5.

**Figure S4:** Anderson–Darling rejection frequency (%) for SPEI candidate distributions across all tested accumulation periods. Expected Type I error (5%) is shown by a dashed line. Scale is identical to Figure 5.

**Figure S5:** ΔAIC for SPEI candidate distributions across all tested accumulation periods. Expected Type I error

(5%) is shown by a dashed line. Note that the normal distribution has one less parameter, which causes it to be the best model at long accumulation periods.

## References

- Akaike H. 1974. A new look at the statistical model identification. *IEEE Trans. Automatic Control* **19**(6): 716–723.
- Allen R, Pereira L, Raes D, Smith M. 1998. Crop evapotranspiration. FAO Irrigation and Drainage Paper 56, FAO, Rome.
- Amatya D, Skaggs R, Gregory J. 1995. Comparison of methods for estimating REF-ET. *J. Irrig. Drain. Eng.* **121**: 427–435.
- Anderson TW, Darling DA. 1954. A test of goodness of fit. *J. Am. Stat. Assoc.* **49**: 765–769.
- Beguería S, Vicente-Serrano SM, Angulo-Martínez M. 2010. A Multi-scalar Global Drought Dataset: The SPEIbase: A New Gridded Product for the Analysis of Drought Variability and Impacts. *Bull. Am. Meteorol. Soc.* **91**(10): 1351–1356, doi: 10.1175/2010BAMS2988.1.
- Beguería S, Vicente-Serrano SM, Reig F, Latorre B. 2013. Standardized precipitation evapotranspiration index (SPEI) revisited: parameter fitting, evapotranspiration models, tools, datasets and drought monitoring. *Int. J. Climatol.* **34**: 3001–3023.
- Burnham KP, Anderson DR. 2004. Multimodel inference: understanding AIC and BIC in model selection. *Sociol. Methods Res.* **33**: 261–304.
- Crutcher HL. 1975. A note on the possible misuse of the Kolmogorov–Smirnov test. *J. Appl. Meteorol.* **14**: 1600–1603.
- Delignette-Muller ML, Pouillot R, Denis J-B, Dutang C. 2013. fitdistrplus: help to fit of a parametric distribution to non-censored or censored data. R package version 1.0-3.
- Giddings L, Soto M, Rutherford BM, Maarouf A. 2005. Standardized Precipitation Index zones for México. *Atmosfera* **18**: 33–56.
- Gudmundsson L, Stagge JH. 2014. SCI: Standardized Climate Indices such as SPI, SRI or SPEI. R package version 1.0-1.
- Guttman NB. 1999. Accepting the Standardized Precipitation Index: a calculation algorithm. *J. Am. Water Resour. Assoc.* **35**: 311–322.
- Haddeland I, Heinke J, Voß F, Eisner S, Chen C, Hagemann S, Ludwig F. 2011. Effects of climate model radiation, humidity and wind estimates on hydrological simulations. *Hydrolog. Earth Syst. Sci. Discuss.* **8**: 7919–7945.
- Hargreaves GH, Samani ZA. 1985. Reference crop evapotranspiration from ambient air temperature. In *Proceedings of the Winter Meeting of American Society of Agricultural Engineers*, Chicago, IL, Paper No. 85-2517.
- Hayes M, Svoboda M, Wall N, Widhalm M. 2011. The Lincoln declaration on drought indices: universal meteorological drought index recommended. *Bull. Am. Meteorol. Soc.* **92**: 485–488.
- Hosking JRM, Wallis JR. 2005. *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press: Cambridge, UK.
- van Huijgevoort MHJ, van Loon AF, Hanel M, Haddeland I, Horvát O, Koutoulis A, Machlina A, Weedon G, Feneková M, Tsanis I, van Lanen HAJ. 2011. Simulation of low flows and drought events in WATCH test basins: impact of different climate forcing datasets. WATCH Technical Report 26, Wageningen University, Wageningen, The Netherlands.
- Jensen ME. 1973. *Consumptive Use of Water and Irrigation Water Requirements*. American Society of Civil Engineers: New York, NY.
- Kumar MN, Murthy CS, Sesha Sai MVR, Roy PS. 2009. On the use of Standardized Precipitation Index (SPI) for drought intensity assessment. *Meteorol. Appl.* **16**: 381–389, doi: 10.1002/met.136.
- Li L, Ngongondo CS, Xu C-Y, Gong L. 2013. Comparison of the global TRMM and WFD precipitation datasets in driving a large-scale hydrological model in southern Africa. *Hydrolog. Res.*, doi: 10.2166/nh.2012.175 (in press).
- Li L, Xu C-Y, Zhang Z, Jain SK. 2014. Validation of a new meteorological forcing data in analysis of spatial and temporal variability of precipitation in India. *Stochastic Environ. Res. Risk Assess.* **28**: 239–252.
- Lilliefors HW. 1967. On the Kolmogorov–Smirnov test for normality with mean and variance unknown. *J. Am. Stat. Assoc.* **62**: 399–402.
- Lloyd-Hughes B, Saunders MA. 2002. A drought climatology for Europe. *Int. J. Climatol.* **22**: 1571–1592.
- Makkonen L. 2006. Plotting positions in extreme value analysis. *J. Appl. Meteorol. Climatol.* **45**: 334–340.
- Makkonen L. 2008. Bringing closure to the plotting position controversy. *Commun. Stat. Theory Meth.* **37**: 460–467.

- Mason DM, Schuenemeyer JH. 1983. A modified Kolmogorov-Smirnov test sensitive to tail alternatives. *Ann. Stat.* **11**: 933–946.
- Massey FJ Jr. 1951. The Kolmogorov-Smirnov test for goodness of fit. *J. Am. Stat. Assoc.* **46**: 68–78.
- McKee TB, Doesken NJ, Kleist J. 1993. The relationship of drought frequency and duration to time scales. In *Proceedings of the 8th Conference on Applied Climatology*, Anaheim, CA, 17–22 January, 179–183.
- Mooley D. 1973. Gamma distribution probability model for Asian summer monsoon monthly rainfall. *Mon. Weather Rev.* **101**: 160–176.
- Shapiro SS, Wilk MB. 1965. An analysis of variance test for normality (complete samples). *Biometrika* **52**: 591–611.
- Sienz F, Bothe O, Fraedrich K. 2012. Monitoring and quantifying future climate projections of dryness and wetness extremes: SPI bias. *Hydrol. Earth Syst. Sci.* **16**: 2143–2157.
- Soláková T, De Michele C, Vezzoli R. 2014. Comparison between Parametric and Nonparametric Approaches for the Calculation of Two Drought Indices: SPI and SSI. *J. Hydrol. Eng.* **19**(9): 04014010. doi: 10.1061/(ASCE)HE.1943-5584.0000942.
- Stagge JH, Tallaksen LM, Xu C-Y, Van Lanen HAJ. 2014. Standardized Precipitation-Evapotranspiration Index (SPEI): sensitivity to potential evapotranspiration model and parameters. Hydrology in a Changing World: Environmental and Human Dimensions. In *Proceedings of FRIEND-Water 2014*, Daniell TM (ed). Montpellier, France, 7–10 October 2014. IAHS Publ. No. 363, IAHS Press, Centre for Ecology and Hydrology: Wallingford, UK, 367–373.
- Stahl K, Hisdal H. 2004. Hydroclimatology. In *Hydrological Drought – Processes and Estimation Methods for Streamflow and Groundwater. Developments in Water Sciences*, Tallaksen LM, van Lanen HAJ (eds). Elsevier Science BV: Amsterdam, 19–51.
- Steinskog DJ, Tjøstheim DB, Kvamstø NG. 2007. A cautionary note on the use of the Kolmogorov–Smirnov test for normality. *Mon. Weather Rev.* **135**: 1151–1157.
- Stephens MA. 1974. EDF statistics for goodness of fit and some comparisons. *J. Am. Stat. Assoc.* **69**: 730–737.
- Vicente-Serrano SM, Beguería S, López-Moreno JI. 2010. A multi-scalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index – SPEI. *J. Clim.* **23**: 1696–1718.
- Vicente-Serrano SM, López-Moreno JI, Gimeno L, Nieto R, Morán-Tejeda E, Lorenzo-Lacruz J, Beguería S, Azorin-Molina C. 2011. A multiscale global evaluation of the impact of ENSO on droughts. *J. Geophys. Res.* **116**: D20109, doi: 10.1029/2011JD016039.
- Vlček O, Huth R. 2009. Is daily precipitation gamma-distributed: adverse effects of an incorrect use of the Kolmogorov–Smirnov test. *Atmos. Res.* **93**: 759–766.
- Weedon GP, Gomes S, Viterbo P, Österle H, Adam JC, Bellouin N, Boucher O, Best M. 2010. The WATCH forcing data 1958–2001: a meteorological forcing dataset for land surface-and hydrological-models. WATCH Technical Report No. 22, Wageningen University, Wageningen, The Netherlands.
- Weedon GP, Gomes S, Viterbo P, Shuttleworth WJ, Blyth E, Österle H, Adam JC, Bellouin N, Boucher O, Best M. 2011. Creation of the WATCH forcing data and its use to assess global and regional reference crop evaporation over land during the twentieth century. *J. Hydrometeorol.* **12**: 823–848.
- Wilks DS. 1990. Maximum likelihood estimation for the gamma distribution using data containing zeros. *J. Clim.* **3**: 1495–1501.
- WMO. 2006. Drought monitoring and early warning: Concepts, progress and future challenges. WMO-No. 1006, World Meteorological Organization, Geneva, Switzerland.
- Wu H, Svoboda MD, Hayes MJ, Wilhite DA, Wen F. 2007. Appropriate application of the standardized precipitation index in arid locations and dry seasons. *Int. J. Climatol.* **27**: 65–79.