Uncertainty in SPI Calculation and Its Impact on Drought Assessment in Different Climate Regions over China

WEN WANG,a JINGSHU WANG,a AND RENATA ROMANOWICZb

a State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering Sciences, Hohai University, Nanjing, China
b Institute of Geophysics, Polish Academy of Sciences, Warsaw, Poland

(Manuscript received 26 October 2020, in final form 15 March 2021)

ABSTRACT: Uncertainty in the calculation of a standardized precipitation index (SPI) attracted growing concerns in the hydrometeorology research community in the last decade. This issue is addressed in the present study from the perspective of candidate probability distributions, the data record length, the cumulative time scale, and the selection of a reference period with the bootstrap and Monte Carlo methods using daily precipitation data observed in four climate regions across China. The impacts of the uncertainty in an SPI calculation on drought assessment are also investigated. Results show that the gamma distribution is optimal in describing the cumulative precipitation in China; among the four time scales investigated in the present study (i.e., 10, 20, 30, and 90 days), the minimal time scale appropriate for SPI calculation is 20 days for the humid region, 30 days for the semihumid/semiarid region and Tibetan Plateau (mostly its eastern part), and 90 days for the arid region. The uncertainty in SPI calculation decreases with the increase of time scale and record length, essentially as a consequence of the decrease of the confidence interval width of gamma distribution parameters with the increase of time scale and record length. But there is little improvement for the parameter estimation with record length longer than 70 years. There is greater uncertainty for high absolute SPI values than for small ones; consequently, there is greater uncertainty in assessing extreme droughts than moderate droughts. Reference period selection has large impacts on drought assessment, especially in the context of climate change. The uncertainty of the SPI calculation has large impacts on categorizing droughts, but no impact on assessing the temporal features of drought variation.

SIGNIFICANCE STATEMENT: The standardized precipitation index (SPI) is the most commonly used drought index over the world for drought assessment, but there are several issues not well recognized in its application, such as the uncertainty resulting from the choice of time scales, the record length, and the reference period for its calculation. A comprehensive evaluation of these issues will be helpful for better interpreting SPI values for drought assessments, especially in the context of climate change.

KEYWORDS: Drought; Indices; Uncertainty; Probability forecasts/models/distribution; Climate change

1. Introduction

Due to its flexibility, spatial–temporal comparability, and simple calculation, the standardized precipitation index (SPI) (McKee et al. 1993) has been applied to drought assessment globally (Guttman 1999; Sienz et al. 2012; W. Wang et al. 2015; Stagge et al. 2015; Okpara et al. 2017; Blain et al. 2018; Guenang et al. 2019). SPI is recommended for meteorological drought assessment by the World Meteorological Organization (2012), stating that SPI allows the user to confidently compare historical and current droughts between different climatic and geographic locations. However, the SPI is a relative measure whose calculation depends on the probability density function adopted, on the method used for parameter estimation and on the reference time period used in the estimation (Paulo et al. 2016). There are several related issues concerning the drought research community in the application of SPI in different places, including the appropriate probability distribution to fit the cumulative precipitation, the appropriate time scale, the record length, and the nonstationarity of pluviometric series.

An appropriate choice of probability distribution for precipitation is the premise of reliable SPI calculation. Different distributions (including probability distribution types and probability distribution parameters) would lead to different SPI values (Guttman 1999). The original formula of the SPI calculation by McKee et al. (1993) assumes that cumulative precipitation follows a gamma distribution (GAM). The gamma distribution has been confirmed to be the optimal choice for SPI calculation in most regions of the world. For example, Stagge et al. (2015) recommended the use of the gamma distribution for general use when calculating SPI across all accumulation periods and regions within Europe; Okpara et al. (2017) showed that gamma distribution performs the best for monthly rainfall in West Africa; Blain et al. (2018) recommended the use of the gamma distributions for calculating SPI with 1–12-month time scales in a tropical–subtropical region of Brazil; and Zhao et al. (2020) showed that gamma distribution exhibits greatest stability among several candidate one-parameter and two-parameter distributions at different time scales in China. Meanwhile, some other distributions are also suggested by different researchers for different climate regions, different accumulation periods and different datasets. Guttman (1999) suggested using the Pearson type III (PE3)
distribution for calculating SPI at different time scales in the contiguous United States; Angelidis et al. (2012) showed that for SPI of 12 or 24 months, the lognormal or the normal probability distribution can be used for simplicity, instead of gamma, producing almost the same results; Sienz et al. (2012) showed that the Weibull type distribution (WEI) gives distinctly improved fits compared to the gamma distribution for monthly precipitation in Europe and the contiguous United States; Vergni et al. (2017) showed that PE3 and generalized normal (GNO) perform slightly better than other distributions in central Italy for SPI calculation with 3-, 6-, and 12-month time scales; Svensson et al. (2017) found that a three-parameter Tweedie (TWE) distribution fits precipitation in the United Kingdom nearly as well as the traditionally used three-parameter distributions; Khanmohammadi et al. (2018) found that some other distribution types fit cumulative precipitation better than gamma distribution in their study stations in Iran; Guenang et al. (2019) found that the Weibull and gamma functions are concurrently the best fits for SPI calculation at time scales not more than 9 months in most areas over Central Africa; Pieper et al. (2020) advocate the employment of the exponential Weibull distribution as the basis for SPI calculation over the global land area; and Raziei (2021) recommended the use of PE3 for fitting precipitation aggregated at all considered time scales (from 1 to 24 months) throughout Iran.

SPI is flexible and can be calculated for multiple time scales. The time scales used in the literature mostly vary from 1 to 24 months, and different time scales reflect the impact of drought on the availability of different water resources. For example, one may want to look at a 1- or 2-month SPI for meteorological drought, anywhere from 1- to 6-month SPI for agricultural drought, and something like 6-month up to 24-month SPI or more for hydrological drought analyses and applications (WMO 2012). However, in regions where rainfall is low, caution should be taken when analyzing SPI at short time scales. Wu et al. (2007) found that SPI values at short time scales such as 1 week can be used in any season in the eastern part of the United States, but in the western part short time scales of SPI can be unreliable due to the high frequency of no precipitation cases. The difficulty of interpreting SPI in arid areas due to statistical issues linked to inaccuracies in the estimation of the gamma distribution has been recognized by many authors (e.g., Spinoni et al. 2014). Nevertheless, the quantitative analysis of which time scale is appropriate in different climate regions is not well recognized.

An accurate estimation of the distribution parameter is essential in the SPI calculation, whereas a certain amount of data is commonly required to achieve reliable parameter estimation. Usually at least 30 years of a pluviometric series are needed to calculate SPI (McKee et al. 1993) and it is believed that the longer the length of record used in the SPI calculation, the more reliable the SPI values will be (Wu et al. 2005). Guttman (1994) investigated the effect of sample size on the precipitation distribution parameter estimation by L-moments, and found that about 40–60 years of record are needed for parameter estimation stability in the central part of the distribution, and about 70–80 years of record for stability in the tails. Carbone et al. (2018) found that SPI estimates for 30-yr reference periods have considerably more uncertainty than those made from 60-yr records, and extreme events have significant influence on SPI estimates, even for records exceeding 60 years.

While an increase in record length improves the stability of the SPI calculation, long records may lead to the aggravation of nonstationarity in the precipitation records. Drought is a situation of water deficit compared to the normal climate status in a reference period. Paulo et al. (2016) compared SPI values calculated from long records of data and its subperiods and found that the impacts of the reference period on the computed SPI were large due to the precipitation change in the context of climate change. Kwon and Sung (2019) showed that the conclusion about the direction of future drought change may be significantly different if different reference periods are used for SPI calculation. Therefore, while discussing the influence of record length on the reliability of parameter estimation, we need to take the nonstationarity of the precipitation record into account. The uncertainty that stems from the selection of the baseline period has been considered as one source of uncertainty in climate change assessment (Mohammed and Scholz 2019). It is necessary to analyze the impact of baseline (or reference) period selection on SPI calculation.

How the uncertainty in the calculation of SPI affects the drought assessment is another issue not well recognized. For instance, Guttman (1999) stated that there is very little difference in the number, duration, intensity, or regional variation of dry events resulting from the SPI computations using different distributions. Many others take the view that an improper probability distribution might impart bias to the index values, exaggerating or minimizing drought severity (Sienz et al., 2012; Stagge et al. 2015; Guerreiro et al. 2017; Zhang and Li 2020; Zhao et al. 2020).

SPI is one of the recommended indices in the Code for the Classification of Meteorological Drought in China (GB/T20481-2006). A comprehensive evaluation of the uncertainties resulting from different sources in the calculation of SPI in different climate regions is necessary. In this paper, using daily precipitation data at stations in four climate regions (humid region, semihumid/semi-arid region, arid region, and Tibetan Plateau) across China, the goodness-of-fit of two two-parameter distributions (i.e., gamma distribution and Weibull distribution) and three three-parameter distributions [i.e., generalized extreme value distribution (GEV), Pearson type III distribution, and Tweedie] are compared. Moreover, the uncertainties in SPI calculation resulting from time scale, record length and reference period selection, are quantitatively analyzed in terms of the confidence interval and coefficient of variation using the bootstrap and Monte Carlo methods. The impacts of the uncertainty in SPI calculation on drought assessment are also investigated.

2. Data and methods

The overall methodology in the present study includes the following main procedures: 1) select nonstationary daily precipitation data observed at different sites in four climate zones in China; 2) investigated the goodness-of-fit of five candidate
distribution models for cumulative precipitation data at different time scales based on the Kolmogorov–Smirnov (KS) test and Akaike information criterion (AIC); 3) determine the minimal time scale of cumulative precipitation for different climate zones based on KS test, and then investigate how the uncertainty in SPI calculation decrease with the increase of time scale with the Monte Carlo method; and 5) investigate the effects of uncertainty on drought classification and trend analysis considering the distorted values of SPI and different choices of reference periods with both the bootstrap and the Monte Carlo method.

a. Data

Annual rainfall over China generally decreases gradually from the humid coasts of southeast China to the arid inland areas of northwest China (Fig. 1a). Daily precipitation observations at more than 600 sites are provided by the national meteorological information center of China Meteorological Administration (http://data.cma.cn/). To mitigate the influence of nonstationarity, we selected 64 stations (Fig. 1b) in four climate regions (i.e., humid region, semihumid/semiarid region, arid region, and Tibetan Plateau), whose annual pluviometric series during 1960–2016 exhibited no trend at the 0.1 significance level by the Mann–Kendall (MK) test (Mann 1945). Only three stations in the Tibetan Plateau have missing data during the 1960–2016 period, and the missing data are no more than 1% of the daily series. Because of the scarcity of gauging stations in the Tibetan Plateau for reliable spatial interpolation, the missing data in a day are replaced by their long-term average value. We also check the homogeneity of variance of the precipitation time series at the monthly time scale by Fligner–Killeen test (Fligner and Killeen 1976), which shows only the time series at six stations in arid regions exhibit nonhomogeneity at the 0.05 significance level. But further investigation by MK test shows that there is no significant temporal trend in variance present in those series.

b. Standardized precipitation index

The SPI calculation is composed of a transformation of one probability distribution (e.g., gamma) to a normal distribution based on the long-term cumulative precipitation record at a given time scale. Table 1 presents the classification of droughts in terms of SPI values and the occurrence probability of each drought theoretically. The calculation procedure is as follows:

1) Calculate cumulative precipitation series over a specific time with a “moving window” approach on a daily basis. In our study, we calculate SPI with eight time scales, i.e., 10, 20, 30, 60, 90, 120, 180, and 365 days, so as to get eight daily SPI series with different scales.

2) Select optimal probability density function (PDF). According to the results of Zhao et al. (2020), WEI and GAM are the best distributions among one-parameter and two-parameter distributions for SPI calculation in China. In the present study, we focus on the comparison of GAM and WEI with three three-parameter PDFs, i.e., GEV, PE3, and TWE for fitting the cumulative precipitation series. The formulae of the five PDFs are listed in Table 2. The optimal PDF is determined based on the KS test and AIC (section 3c).

3) The inverse normal (Gaussian) function is then applied to the cumulative probability, resulting in the SPI. This procedure is an equiprobability transformation, consisting of calculating the mixed distribution function of zero precipitation and cumulative distribution function of nonzero precipitation and transforming it to a standard normal distribution (McKee et al. 1993).

c. Evaluation of the goodness-of-fit of probability distributions

The goodness-of-fit of the PDF for the sample data is evaluated based on the nonparametric KS test and AIC.

<table>
<thead>
<tr>
<th>SPI</th>
<th>Drought categories</th>
<th>Theoretical occurrence probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; −1</td>
<td>No drought</td>
<td>84.1%</td>
</tr>
<tr>
<td>(−1.5, −1]</td>
<td>Moderate drought</td>
<td>9.2%</td>
</tr>
<tr>
<td>(−2, −1.5]</td>
<td>Severe drought</td>
<td>4.4%</td>
</tr>
<tr>
<td>≤−2</td>
<td>Extreme drought</td>
<td>2.3%</td>
</tr>
</tbody>
</table>

TABLE 1. Classification of drought categories.
The test statistic \( D \) of the KS test measures the largest distance between the empirical cumulative distribution function \( F(x) \) fitted by sample series and the theoretical cumulative distribution function \( G(x) \), that is,

\[
D = \max|F(x) - G(x)|. \tag{1}
\]

The hypothesis regarding the distributional form is rejected if the test statistic \( D \) is greater than the critical value at a given significance level. The goodness-of-fit is indicated by the \( p \) value. If the \( p \) value is greater than a given significance level (0.05 here), the null hypothesis is accepted, and the larger the \( p \) value, the better the candidate distribution is.

The Shapiro–Wilk test is one of the most powerful and most frequently used normality tests, but it does not specify the parameters of normal distribution (i.e., the values of mean and standard deviation). When it is applied to SPI series, additional parameters of normal distribution (i.e., the values of mean and standard deviation) are estimated by the empirical function and the fitted distribution function, respectively. The distribution function with the smallest \( AIC \) is considered the most suitable distribution.

\[ AIC = n \log(SSE/n) + 2m, \tag{2} \]

where \( n \) is the length of the observation series and \( m \) is the number of parameters. \( SSE \) is the sum squared residual of the fitted model and is estimated as follows:

\[
SSE = \sum_{i=1}^{n} (O_{i} - P_{i})^2, \tag{3}
\]

where \( O_{i} \) and \( P_{i} \) are the probabilities of each observation estimated by the empirical function and the fitted distribution function, respectively. The distribution function with the smallest \( AIC \) is considered the most suitable distribution.

d. **Bootstrap method for uncertainty evaluation**

Bootstrap is a resampling technique proposed by Efron (1979) to make inferences about a sampling distribution by resampling the sample itself with replacement. The bootstrap method has been commonly used in evaluating the uncertainty of an SPI calculation (Liu et al. 2014; Hu et al. 2015; Vergni et al. 2017).

Given a cumulative precipitation series \( X = (x_{1}, x_{2}, \ldots, x_{n}) \), the Monte Carlo procedure for uncertainty evaluation is as follows:

1) Resample \( X \) \( N \) times (we take \( N = 1000 \)) with replacement from the precipitation series, and obtain \( N \) bootstrapped samples of length \( m \), \( X_j = (x_{j1}, x_{j2}, \ldots, x_{jm}) \), \( j = 1, 2, 3, \ldots, N \), \( m \leq n \).
2) For the assumed probability distribution (here, gamma distribution), the maximum likelihood method is used to estimate distribution parameters of each sample \( X_j \) and \( N \) groups of parameters (\( \alpha \) and \( \beta \)) are obtained.
3) Calculate the 90% (difference between 5th and 95th percentiles) confidence interval width (CIW) and coefficient of variation (CV) of \( \alpha \) and \( \beta \) to quantify the uncertainty of distribution parameter estimation.
4) Using \( N \) groups of parameters, calculate the cumulative probabilities \( P(x_i) \) for each value \( (x_i) \) of the cumulative precipitation \( X \). Then \( N \) values of \( P(x_i) \) are transformed to \( N \) values of SPI through an equiprobability transformation.
5) Calculate the 90% CIW and CV of \( N \) SPI values at \( x_i \) to quantify the uncertainty of the SPI calculation. The smaller CIW and CV, the smaller the uncertainty.

e. **Monte Carlo method for uncertainty evaluation**

The bootstrap method cannot satisfy the need for evaluating the uncertainty of parameter estimation with different lengths (longer than the length of observation data) because the size of the bootstrap sample should not exceed the original sample size. Therefore, the Monte Carlo method (or parametric bootstrap method, with assumptions made about the distributions of the data) is used to generate random samples from a given probability distribution to analyze the influence of data size on parameter estimation.

Given a cumulative precipitation series \( X = (x_{1}, x_{2}, \ldots, x_{n}) \), the Monte Carlo procedure for uncertainty evaluation is as follows:

1) Using the maximum likelihood method to estimate the parameters of a given distribution (here, \( \alpha \) and \( \beta \) of the gamma distribution) of the precipitation series \( X \).
2) Generate \( N \) groups of random samples of length \( m \), \( X_j = (x_{j1}, x_{j2}, \ldots, x_{jm}), j = 1, 2, 3, \ldots, N \). Here, both \( N \) and \( m \) can be set of arbitrary size depending on the experiment.
3) Repeat steps 2–5 with the bootstrap method to get the 90% CIW and CV of \( \alpha \), \( \beta \), and SPI.

All the above computations, including the data quality examination, SPI calculation, the evaluation of the goodness-of-fit of probability distributions, the bootstrap procedure, and

---

**Table 2. Probability distribution formulas.**

<table>
<thead>
<tr>
<th>Distribution name</th>
<th>Probability density function</th>
<th>Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gamma</td>
<td>( f(x) = \frac{1}{\beta \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} )</td>
<td>( \alpha ) and ( \beta )</td>
</tr>
<tr>
<td>Weibull</td>
<td>( f(x) = \beta / \alpha (x/\alpha)^{\frac{\beta}{\alpha}-1} \exp \left( -x/\alpha \right)^{\beta} )</td>
<td>( \alpha ) and ( \beta )</td>
</tr>
<tr>
<td>Generalized extreme value</td>
<td>( f(x) = \frac{1}{\eta \Gamma(\gamma)} \left( 1 - k \left( \frac{x - \delta}{\eta} \right) \right)^{\gamma-1} \exp \left( - \left[ 1 - k \left( \frac{x - \delta}{\eta} \right) \right]^{\gamma} \right) )</td>
<td>( \alpha ), ( \beta ), and ( k )</td>
</tr>
<tr>
<td>Pearson III</td>
<td>( f(x) = \frac{1}{\beta} \left( \frac{x - \mu}{\beta} \right)^{\alpha-1} e^{-(x-\mu)^{2}/\beta} )</td>
<td>( \alpha ), ( \beta ), and ( \mu )</td>
</tr>
<tr>
<td>Tweedie</td>
<td>( f(x) = a(x, \phi) \exp \left( \frac{1}{\phi} [x - \kappa(\theta)] \right) )</td>
<td>( \mu ), ( \theta ), and ( \xi )</td>
</tr>
</tbody>
</table>
Monte Carlo procedure for uncertainty analysis, are implemented with R software (R Core Team 2019).

3. Results

a. Optimal probability distribution for calculating SPI

Five PDFs, i.e., GAM, WEI, GEV, PE3, and TWE, are compared in terms of their goodness-of-fit to observed precipitation of different accumulation period in different regions. Figure 2 shows the $p$ values of the KS test for the five PDFs. It is shown that at the 10-day time scale in each climate region considerable number of cases fail to follow any candidate distribution, indicating that SPI values at short time scales like 10 days cannot be used anywhere in China; all distributions are acceptable with time scale longer than 10 days according to a KS test with $p$ values greater than 0.05.

AIC is further applied to comparing the goodness-of-fit of PDFs. We define the win rate of a PDF as the percentage of a PDF taking the smallest AIC value among five PDFs for all cumulative precipitation series with a given time scale at all calendar days over the year at all stations in each climate region (i.e., in total $365 \times$ the number of stations in each climate region) according to AIC. Table 3 presents mean $p$ values and the win rate of each PDF for the precipitation series with different time scales in different climate regions. No PDF wins in all four regions and over all time scales. However, PE3 performs the poorest as its win rate is almost 0 in four regions. As the time scale increases, the win rate of WEI decreases from 11% to 40% to zero, which means it is unsuitable for precipitation series at a longer time scale. Comparing all the distribution types, GAM, GEV, and TWE have higher win rates than the others, with win rates of 23%–70%, 15%–40%, and 13%–33%, respectively, and GAM has the highest win rate at time scales $\geq 30$ days in all regions. Therefore, GAM is confirmed to be the optimal choice for calculating SPI in China, and it is chosen to calculate the SPI value in the following investigation.

b. Effects of time scale on SPI calculation

The KS test is used to test if the computed SPI series at different time scales follows the standard normal distribution based on bootstrap samples of precipitation observations. In the present study, we consider a time scale to be suitable for SPI calculation at a site if the percentage of the SPI series following the standard normal distribution at the given time scale over the 365 days of the year at a given site is greater than 90%. The percentage of standard normality in SPI series at 10-, 20-, 30-, 90-, and 365-day time scales in different seasons for all sites in four regions is shown in Figs. 3a–d. In the humid region, the percentage of standard normality is higher than 0.9 in spring, summer, and whole year at a 20-day time scale, and close to 100% at time scales $\geq 30$ days. In the semihumid/semiarid region...
In the humid region, the percentage of normality is higher than 0.9 for all seasons at time scales \( \geq 30 \) days. In the Tibetan Plateau, the percentage of normality is higher than 0.9 for all seasons except for winter at a 30-day time scale and higher than 0.9 for all seasons at time scales \( \geq 90 \) days. In the arid region, the percentage of normality is higher than 0.9 for all seasons at time scales \( \geq 90 \) days. As a result, from the perspective of the standard normality rate, among the time scales we investigated here (i.e., 10, 20, 30, and 90 days), a minimal time scale of 20 days can be applied in humid regions, a 30-day time scale in semi-humid/semi-arid regions, a 90-day time scale in arid regions, and a 365-day time scale in the Tibetan Plateau.

### Table 3. The mean \( p \) values and the win rates of candidate PDFs according to AIC in four climate regions. The bold numbers indicate the best performance among the five PDFs.

<table>
<thead>
<tr>
<th>Region</th>
<th>Distribution</th>
<th>10-day scale</th>
<th>30-day scale</th>
<th>90-day scale</th>
<th>365-day scale</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean ( p ) Rate (%)</td>
<td>Mean ( p ) Rate (%)</td>
<td>Mean ( p ) Rate (%)</td>
<td>Mean ( p ) Rate (%)</td>
<td></td>
</tr>
<tr>
<td>Humid</td>
<td>GAM 0.62 23.69</td>
<td>0.92 55.81</td>
<td>0.94 57.70</td>
<td>0.95 50.49</td>
<td></td>
</tr>
<tr>
<td></td>
<td>WEI 0.62 11.48</td>
<td>0.94 4.57</td>
<td>0.94 0</td>
<td>0.95 0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GEV 0.54 40.12</td>
<td>0.93 24.36</td>
<td>0.95 21.21</td>
<td>0.96 30.13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PE3 0.59 0</td>
<td>0.92 0</td>
<td>0.95 0</td>
<td>0.96 0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TWE 0.6 24.71</td>
<td>0.93 15.26</td>
<td>0.94 21.09</td>
<td>0.96 19.38</td>
<td></td>
</tr>
<tr>
<td>Semi humid/semi-arid</td>
<td>GAM 0.7 26.01</td>
<td>0.93 39.12</td>
<td>0.95 43.58</td>
<td>0.97 46.84</td>
<td></td>
</tr>
<tr>
<td></td>
<td>WEI 0.69 23.71</td>
<td>0.95 9.11</td>
<td>0.96 0.05</td>
<td>0.97 0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GEV 0.68 23.81</td>
<td>0.94 23.71</td>
<td>0.96 24.120</td>
<td>0.96 23.25</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PE3 0.69 0</td>
<td>0.94 0</td>
<td>0.96 0</td>
<td>0.96 0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TWE 0.72 28.06</td>
<td>0.94 28.06</td>
<td>0.95 32.25</td>
<td>0.98 29.92</td>
<td></td>
</tr>
<tr>
<td>Arid</td>
<td>GAM 0.3 23.50</td>
<td>0.72 38.91</td>
<td>0.93 42.90</td>
<td>0.94 57.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>WEI 0.3 29.29</td>
<td>0.73 10.22</td>
<td>0.96 28.58</td>
<td>0.97 27.92</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GEV 0.27 15.85</td>
<td>0.7 19.84</td>
<td>0.96 28.58</td>
<td>0.97 27.92</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PE3 0.28 0</td>
<td>0.71 0</td>
<td>0.96 0</td>
<td>0.96 0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TWE 0.62 31.36</td>
<td>0.72 31.04</td>
<td>0.96 28.25</td>
<td>0.97 15.03</td>
<td></td>
</tr>
<tr>
<td>Tibetan Plateau</td>
<td>GAM 0.59 47.81</td>
<td>0.91 56.72</td>
<td>0.93 66.18</td>
<td>0.93 68.25</td>
<td></td>
</tr>
<tr>
<td></td>
<td>WEI 0.55 15.19</td>
<td>0.94 6.39</td>
<td>0.95 0</td>
<td>0.95 0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GEV 0.57 16.45</td>
<td>0.92 21.42</td>
<td>0.93 20.71</td>
<td>0.95 28.20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PE3 0.56 0</td>
<td>0.93 0</td>
<td>0.95 0</td>
<td>0.96 0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TWE 0.63 20.55</td>
<td>0.93 15.47</td>
<td>0.96 13.11</td>
<td>0.94 3.55</td>
<td></td>
</tr>
</tbody>
</table>

**FIG. 3.** The percentage of normality of SPI of all studied stations at 10-, 20-, 30-, 90-, and 365-day scales in four climate regions based on 1960–2016 precipitation series.
is suitable for the semihumid/semiarid regions and the Tibetan Plateau, and a 90-day time scale is recommended for the arid region.

The results for the Tibetan Plateau should be treated with caution, because most stations investigated there are located in the comparatively wet eastern part of the plateau, and for its dry western parts the 30-day time scale may be not enough long for an SPI calculation. In addition, the time scales between 30 and 90 days (e.g., 60 days) are not investigated in the present study, but could be possible choices for the arid region. However, considering the popularity of the time scales of 90 days, we suggest using 90 days (or 3 months) as the minimal time scale for SPI calculation when assessing droughts over entire China.

We further choose Fuzhou, Zhengzhou, Turpan, and Lasa as representative stations (shown in Fig. 1) for a humid region, semihumid/semiarid region, arid region, and the Tibetan Plateau, respectively, to evaluate the uncertainty resulting from different time scale options in terms of CV and CIW of SPI using cumulative precipitation series at the four representative stations with different time scales (i.e., 30, 90, 180, and 365 days) at the first and 180th day of a year with the bootstrap method. The average CV and CIW values for these SPI series and gamma distribution parameters ($\alpha$ and $\beta$) are all presented in Table 4. It is shown that in general CIW and CV of SPI both decrease as time scale increases, which indicates that the uncertainty of the SPI calculation decreases as the time scale increases. Although the CIW of $\alpha$ increases with the increase of time scale (because of the increases of cumulative precipitation with the increase of time scale) and the CIW of $\beta$ shows no clear changing pattern, CVs of both $\alpha$ and $\beta$ decrease with the increase of time scale (except for the case of the cumulative precipitation at the first day in Fuzhou), indicating that the uncertainty in the estimation of gamma distribution parameters decreases with the increase of time scale. That leads to the decrease of the uncertainty in SPI calculation as the time scale increases.

c. Uncertainty resulting from record length

Gamma distribution parameters are estimated with the maximum likelihood method for each of the cumulative precipitation series at the 180th day of the year for four representative stations with different time scales (20 days for Fuzhou, 30 days for Zhengzhou and Lasa, and 90 days for Turpan). The Monte Carlo method is then applied to simulate 10,000 samples of 100 points (representing 100 years) for each set of parameters. Subsequently, we take 30–100 points (representing 30–100 years) from each sample to estimate the shape parameter $\alpha$ and scale parameter $\beta$ of the gamma distribution, and calculate the 90% CIW of $\alpha$ and $\beta$. The relative CIW, which is the ratio of CIW with $n$ = 30–100 to the CIW with $n$ = 30, is shown in Fig. 4. As the data size increases, the relative CIW decreases, which indicates a decrease of the uncertainty in parameter estimation. The effect of data size on the shape parameter $\alpha$ is larger than that on the scale parameter $\beta$. The decrease of relative CIW tends to get slower with an increase of data size, and the relative CIW is close to being stable after the data size reaches 70–80 points, or, 70–80 years record for the SPI calculation.

The uncertainties of parameter $\alpha$ and $\beta$ estimates under different data sizes are also evaluated in terms of the CVs of

<table>
<thead>
<tr>
<th>Station</th>
<th>Day of year</th>
<th>Value</th>
<th>20 days</th>
<th>30 days</th>
<th>90 days</th>
<th>180 days</th>
<th>365 days</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fuzhou</td>
<td>1st SPI</td>
<td>0.54 (0.17)</td>
<td>0.53 (0.16)</td>
<td>0.51 (0.15)</td>
<td>0.51 (0.15)</td>
<td>0.52 (0.14)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\alpha$</td>
<td>0.58 (0.17)</td>
<td>0.60 (0.17)</td>
<td>1.47 (0.14)</td>
<td>5.44 (0.20)</td>
<td>20.02 (0.22)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>12.91 (0.17)</td>
<td>20.41 (0.18)</td>
<td>16.67 (0.14)</td>
<td>45.90 (0.20)</td>
<td>33.17 (0.19)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>180th SPI</td>
<td>0.60 (0.20)</td>
<td>0.59 (0.18)</td>
<td>0.55 (0.17)</td>
<td>0.54 (0.16)</td>
<td>0.53 (0.16)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\alpha$</td>
<td>2.38 (0.28)</td>
<td>4.36 (0.23)</td>
<td>12.69 (0.22)</td>
<td>16.09 (0.20)</td>
<td>19.80 (0.19)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>39.17 (0.26)</td>
<td>39.50 (0.22)</td>
<td>20.18 (0.20)</td>
<td>17.65 (0.20)</td>
<td>31.68 (0.17)</td>
<td></td>
</tr>
<tr>
<td>Zhengzhou</td>
<td>1st SPI</td>
<td>—</td>
<td>0.55 (0.19)</td>
<td>0.54 (0.18)</td>
<td>0.54 (0.17)</td>
<td>0.54 (0.17)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\alpha$</td>
<td>—</td>
<td>0.52 (0.19)</td>
<td>1.69 (0.19)</td>
<td>7.13 (0.18)</td>
<td>11.48 (0.17)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>—</td>
<td>8.08 (0.19)</td>
<td>15.88 (0.18)</td>
<td>22.78 (0.16)</td>
<td>22.23 (0.16)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>180th SPI</td>
<td>—</td>
<td>0.51 (0.17)</td>
<td>0.50 (0.15)</td>
<td>0.49 (0.15)</td>
<td>0.49 (0.14)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\alpha$</td>
<td>—</td>
<td>0.67 (0.19)</td>
<td>1.83 (0.16)</td>
<td>2.83 (0.16)</td>
<td>15.43 (0.13)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>—</td>
<td>26.95 (0.19)</td>
<td>17.29 (0.17)</td>
<td>20.45 (0.17)</td>
<td>17.03 (0.14)</td>
<td></td>
</tr>
<tr>
<td>Turpan</td>
<td>1st SPI</td>
<td>—</td>
<td>—</td>
<td>0.52 (0.16)</td>
<td>0.52 (0.16)</td>
<td>0.50 (0.15)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\alpha$</td>
<td>—</td>
<td>—</td>
<td>0.47 (0.19)</td>
<td>1.08 (0.18)</td>
<td>1.48 (0.14)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>—</td>
<td>—</td>
<td>3.23 (0.24)</td>
<td>2.42 (0.17)</td>
<td>2.49 (0.15)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>180th SPI</td>
<td>—</td>
<td>—</td>
<td>0.54 (0.17)</td>
<td>0.50 (0.15)</td>
<td>0.48 (0.15)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\alpha$</td>
<td>—</td>
<td>—</td>
<td>0.39 (0.19)</td>
<td>0.55 (0.14)</td>
<td>2.33 (0.13)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>—</td>
<td>—</td>
<td>3.12 (0.19)</td>
<td>3.86 (0.19)</td>
<td>2.72 (0.17)</td>
<td></td>
</tr>
<tr>
<td>Lasa</td>
<td>1st SPI</td>
<td>—</td>
<td>0.59 (0.19)</td>
<td>0.57 (0.17)</td>
<td>0.53 (0.16)</td>
<td>0.52 (0.15)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\alpha$</td>
<td>—</td>
<td>0.95 (0.22)</td>
<td>0.90 (0.22)</td>
<td>0.53 (0.15)</td>
<td>14.14 (0.14)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>—</td>
<td>14.06 (0.24)</td>
<td>25.52 (0.25)</td>
<td>16.99 (0.27)</td>
<td>18.57 (0.22)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>180th SPI</td>
<td>—</td>
<td>0.57 (0.19)</td>
<td>0.55 (0.17)</td>
<td>0.54 (0.17)</td>
<td>0.52 (0.16)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\alpha$</td>
<td>—</td>
<td>0.16 (0.41)</td>
<td>2.83 (0.40)</td>
<td>4.57 (0.31)</td>
<td>14.20 (0.20)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>—</td>
<td>55.30 (0.41)</td>
<td>32.81 (0.30)</td>
<td>61.93 (0.24)</td>
<td>11.71 (0.18)</td>
<td></td>
</tr>
</tbody>
</table>
\(\alpha\) and \(\beta\), shown in Fig. 5. The CV of \(\beta\) is slightly greater than that of \(\alpha\) for the same data size, so the uncertainty of scale parameter \(\beta\) is slightly higher than that of the shape parameter \(\alpha\). The CVs of \(\alpha\) and \(\beta\) decrease with an increase of data size, as for the variation of CIW. The CV of \(\alpha\) becomes stable after the data size reaches 70–80, while the CV of \(\beta\) keeps going down even after the data size reaches 70–80. Therefore, from the perspective of minimizing the uncertainty of parameter estimations with minimal record length, the optimal time series length for the SPI calculation is 70–80 years.

d. Uncertainty in drought assessment resulting from uncertainty in SPI calculation

1) IMPACT OF SPI UNCERTAINTY ON DROUGHT CLASSIFICATION

Taking the gamma distribution parameters \((\alpha = 4.47, \beta = 49.05)\) fitted to the 30-day cumulative precipitation series at the 180th day of Fuzhou Station as an example, we investigate the impact of SPI calculation uncertainty on drought classification. We first generate 5000 series of 30-point data as the original cumulative precipitation series with parameter \(\alpha = 4.47\) and \(\beta = 49.05\), and calculate the original SPI (SPI\(_O\)) series for each of them using the true parameters (i.e., \(\alpha = 4.47\) and \(\beta = 49.05\)). Then for each cumulative precipitation series, we calculate SPI for 1000 bootstrap samples (referred to as SPI\(_B\)) to get a 90% CIW. Using the same procedure, the CIW for 60- and 90-point series can also be derived.

Figures 6a–c show the CIW for 5000 data series of 30, 60, and 90 points, respectively. First, it is shown that the CIW decreases as data size increases, which indicates that the uncertainty decreases with an increase of data size. Second, the CIW of SPI is low in the middle part near SPI equal 0, and increases non-linearly from the middle to the two sides, indicating much larger uncertainty under drier or wetter conditions. In other words, the larger anomaly of precipitation, the less reliable the SPI. That means there is larger uncertainty in assessing extreme droughts than in assessing moderate ones.

Then we investigate how drought may be falsely categorized because of the uncertainty in SPI calculation. To do this, for each point of a 30-point precipitation series, we determine its true drought category based on its SPI\(_O\) value, and then count the number of drought categories correctly identified, underestimated and overestimated based on its 1000 SPI\(_B\) values for the same point. In this way we can calculate the proportion of correct identifications, underestimated identification, and overestimated identification for all the points of 5000 30-point precipitation series based on SPI\(_B\) in comparison to the SPI\(_O\). The results are shown in Fig. 7a. Similarly, we evaluate the drought identification for 5000 60-point and 5000 90-point precipitation series, respectively (Figs. 7b,c).

Comparing Figs. 7a–c, we find that, a small percentage of nondroughts (2.8%, 1.9%, and 1.5% for 30-, 60-, and 90-point series, respectively) have been overcategorized as moderate droughts, but a considerable percentage (18.9%, 14.2%, and 11.7% for 30-, 60-, and 90-point series, respectively) of moderate droughts have been undercategorized as nondroughts. Similarly, a high percentage (21.5%, 13.8%, and 11.6% for 30-, 60-, and 90-point series, respectively) of extreme droughts have been undercategorized as severe droughts; meanwhile similar
amounts (18.9%, 12.9%, and 10.3% for 30-, 60-, and 90-point series, respectively) of severe droughts have been overcategorized. The probability of false categorization decreases with an increase of data size.

Based on the theoretical occurrence probability of each drought class in Table 1, and the rate of miscategorization in Fig. 7, we can quantitatively define how confident we are when categorizing a drought in terms of its SPI value. Considering that class A drought is the lowest class (i.e., nondrought), and assuming that the theoretical occurrence probability of class A drought and class B drought is $P_A$ and $P_B$, respectively, and the probability of correctly categorizing class A drought is $P_{A|A}$, the probability of undercategorizing class B for class A is $P_{A|B}$, then the proportion of all identified class A droughts being true class A droughts ($P_{\text{correct}}$) is

$$P_{\text{correct}} = P_A P_{A|A} / (P_A P_{A|A} + P_B P_{A|B}).$$

(4)

and the proportion of all identified class A droughts being undercategorized class B droughts ($P_{\text{under}}$) is

$$P_{\text{under}} = P_B P_{A|B} / (P_A P_{A|A} + P_B P_{A|B}).$$

(5)

Similarly, the proportion of overcategorized droughts ($P_{\text{over}}$) can be calculated if class A is the top class and class B could be overcategorized for class A.

In the case when class A is the middle class, and class C can be falsely categorized as class A with a rate of $P_{A|C}$, the proportion of correctly categorizing class A droughts ($P_{\text{correct}}$) is given by

$$P_{\text{correct}} = P_A P_{A|A} / (P_A P_{A|A} + P_B P_{A|B} + P_C P_{A|C}).$$

(6)

Parameters $P_{\text{correct}}$, $P_{\text{under}}$, and $P_{\text{over}}$ for each drought category identified in terms of a given SPI series are shown in Table 5. Taking the extreme drought as an example, when the data size is 30 points, only 68.5% extreme droughts are true, whereas 31.5% are actually severe droughts; when the data size reaches 90 points, there are only 81.8% extreme droughts are true, whereas 18.2% are actually severe droughts. With an increase of data size, we have greater confidence in the drought category.

2) IMPACT OF SPI CALCULATION UNCERTAINTY ON DROUGHT TREND ANALYSIS

To investigate the impact of the SPI calculation uncertainty on drought trend analysis, one station is selected in each climate region where the 365-day cumulative precipitation series exhibit a significant trend. At each station an original SPI series for the cumulative precipitation series is derived first. Then, we generate 1000 SPI$_B$ series based on 1000 bootstrap samples. Finally, a test of the trend of the original SPI series as well as 1000 SPI$_B$ series is performed using Mann–Kendall trend test.

Fig. 6. The 90% CIW of SPI derived from bootstrap samples for the 5000 simulated (a) 30-, (b) 60-, and (c) 90-point series with parameter $\alpha = 4.47$ and $\beta = 49.05$.

Fig. 7. The proportion of drought category indicated by SPI based on bootstrapped samples vs the original SPI (SPI$_O$) at lengths of (a) 30, (b) 60, and (c) 90.
3) UNCERTAINTY IN DROUGHT ASSESSMENT RESULTING FROM UNCERTAINTY IN REFERENCE PERIOD SELECTION

Taking the gamma distribution parameters \((\alpha = 4.47, \beta = 49.05)\) of the 30-day cumulative precipitation series on the 180th day of Fuzhou Station as an example, the uncertainty of drought assessment for a simulated 200-yr precipitation series is analyzed. The procedure is as follows: 1) generate 10 000 series of 30 points and 1 series of 200 points which follow the gamma distribution with \(\alpha = 4.47\) and \(\beta = 49.05\); 2) estimate \(\alpha\) and \(\beta\) for the 10 000 30-point series, which are assumed to be different reference periods of 30 years; 3) calculate the SPI for the 200-point series with the 10 000 estimates of \(\alpha\) and \(\beta\), obtaining 10 000 SPI series of 200 points; 3) summarize the number of droughts of different classes in the 10 000 SPI series. The results are presented as boxplots in Fig. 8. Theoretically, the number of extreme droughts, severe droughts and moderate droughts calculated based on the 200-point series standardized with correct parameters \((\alpha = 4.47, \beta = 49.05)\) should be 4.6, 8.8, and 18.4. But the drought assessment results based on simulated reference period data differ from the theoretical result greatly: the number of extreme droughts varies from 0 to 20, the variation range of severe droughts is 5–21, and for moderate droughts is 10–26. It is obvious that the uncertainty of the reference period selection brings large uncertainty in drought assessment.

The uncertainty in drought assessment resulting from reference period selection is essentially an issue of the representativeness of the reference period. For the 10 000 simulated series of 30 points, boxplots of their mean values and standard deviations are presented in Fig. 9. It is shown that, while the theoretical mean value is \(\alpha\beta = 219.3\) and standard deviation is \(\sqrt{\alpha\beta^2} = 103.7\) for a gamma distribution with \(\alpha = 4.47\) and \(\beta = 49.05\), there is a large variation in the simulated series. The mean values of simulated series vary from 152.5 to 289.6, and standard deviations vary from 49.4 to 190. The result indicates that a reference period of 30 years cannot unusually be representative of the climate normal.

4. Discussion

a. How the number of zeros affects the minimal time scale in different regions?

SPI has flexibility in the choice of time scales, but there is no widespread agreement on which time scale is appropriate and less uncertain for different climate regions when applying SPI to evaluating drought. WMO (2012) suggests caution when applying short time scale SPI in areas with less precipitation. Wu et al. (2007) found that when applying SPI in the western United States, short time scale SPI is not reliable due to the high frequency of zero values in precipitation series. An SPI series must be capable of meeting the normality assumption, but the high frequency of zero leads to truncating distribution of SPI at zero precipitation and SPI values that do not follow a standard normal distribution (Stagge et al. 2015; Blain et al. 2018). Figure 10 shows the percentage of zeros \(P_0\) in the cumulative precipitation series of different time scales at all stations in four climate regions of China. It is shown that due to the impacts of the monsoon climate in China, there are more zero values in winter and spring precipitation than in summer and autumn precipitation in all four regions. The percentage \(P_0\) is mostly less

![Fig. 8. The number of extreme, severe, and moderate droughts for a 200-point series based on different 30-point reference series coming from the gamma distribution with same parameters (\(\alpha = 4.47\) and \(\beta = 49.05\)).](image-url)
than 20% for the 20-day cumulative precipitation series in humid region, 30-day series in semihumid/semiarid region and the Tibetan Plateau, and fully less than 10% for 90-day series in arid region in different seasons. Although Stagge et al. (2015) proposed a method of handling zeros based on the “center of mass” of the zero distribution, but the minimum SPI value is still limited by the proportion of zeros and does not follow a standard normal distribution. Blain (2012) also explained that for periods in which the probability associated with the zero precipitation value is close to 0.5, the SPI may erroneously indicate the end of an existing drought (or a decrease in its severity). However, the effect of zeros on the standardization of precipitations attached with time scale is not quantified yet. Here we make a quantitative analysis about how the amount of zeros may affect the standardization process.

We first fit gamma distributions to the 30-day cumulative precipitation at the 1st day and the 180th day of the year for those four representative stations, and simulate 10,000 precipitation series of different length \((L = 30–100)\) with the gamma distribution parameters. Then, we replace the last 1 to \(L/2\) values of each series with zeros, and calculate SPI values for those series with different number of zeros. Finally, we test if those SPI series follow a standard normal distribution or not with the KS test at the 0.05 significance level. The simulation results show that there is a threshold of \(P_0\) for each precipitation series, above which the standardized precipitation would not follow a standard normal distribution. The threshold of \(P_0\) is only related to the data size, whatever the gamma distribution parameters we applied to the simulations. Such a relationship between data size and \(P_0\) threshold is shown in Fig. 11.

The longer data size, the less zeros can be contained in a precipitation series to make its standardized series follow the standard normal distribution.

According to such a relationship, for a cumulative precipitation series of 57 years long (1960–2016) as in the present study, the maximum percentage of zeros is about 0.17, that is, a maximum 10 zeros can be contained in the series to be standardized following standard normal distribution. A close inspection of Fig. 10 shows that cumulative precipitation series with time scale of 20 days for humid region, 30 days for semihumid/semi-arid region and the Tibetan Plateau, and 90 days for arid region in different seasons meet such a requirement mostly. That is essentially why the minimal time scales for SPI calculation are 20 days for humid region, 30 days for semihumid/semi-arid region and the Tibetan Plateau, and 90 days for arid region.

**b. Impacts of nonstationarity of precipitation series on drought assessment**

Although we show that a record length of 70–80 years is optimal for an SPI calculation, which is in agreement with the result of Guttmann (1994) who found that about 70–80 years of record is needed for parameter estimation stability in the tails, whereas DeGaetano et al. (2015) and Carbone et al. (2018) who find record lengths of 60–70 years typically result in stable parameters and representative SPI values, it is very common that significant nonstationarity exhibits in a precipitation time series longer than 30 years. Growing concern about the nonstationary analyses of hydrological processes has stimulated research on a nonstationary approach to drought assessment by taking either time or some climate indices as a covariate. A number of variants of SPI which tack nonstationarity in droughts has been developed in the last decade, which can be grouped into three categories: 1) calculate SPI using adjusted precipitation series considering the varying local-time means (Türkeş and Tatli 2009); 2) varying probability distribution parameters with time linearly (Russo et al. 2013) or nonlinearly (Y. Wang et al. 2015); 3) varying probability distribution parameters with climate indices linearly (Li et al. 2015) or in a flexible way with the Generalized Additive Model for Location, Scale, and Shape (GAMLSS) modeling framework (Rashid and Beecham 2019).

However, drought is a phenomenon involving a water deficit compared to the climate normal, and temporal variation of drought characteristics (such as the trend in drought severity) is a major issue in drought assessment (W. Wang et al. 2015), whereas the time-varying index makes the “normal” keep varying and may deviate far from its long-term climate normal, therefore, the application of a nonstationary SPI to drought assessment may lead to contradictory results in comparison with classical stationary SPI. For instance, Park et al. (2019) showed that the nonstationary method underestimated the drought severity than the stationary SPI for two severe drought cases during the last 10 years in South Korea. Shiau (2020) showed that because of a clearly increasing rainfall trend at Taipei over the last four decades, the stationary SPI series indicated less frequent and less severe droughts than the nonstationary SPI series did. It is therefore concerning if such a nonstationary practice may reduce the capability of detecting the temporal variability of droughts or detecting multiyear mega-droughts which affected many regions during the last decades, such as the 2012–14 drought in California (e.g., Griffin and Anchukaitis 2014) and the central Chile mega-drought during 2010–18 (Garreaud et al. 2020). How to comprehensively assess droughts by capturing the nonstationary in drought...
features on one hand and not losing the long-term temporal features on the other hand is a challenge for future research.

Another challenge in drought assessment is how to deal with the nonstationarity in the assessment of future droughts. There are several approaches to assessing future droughts using standardized drought indices in the literature: (i) calculate standardized drought indices using data of the entire simulation period including historic period and future period (Duffy et al. 2015; Guerreiro et al. 2017; Yao et al. 2020); (ii) calculate nonstationarity SPIs using data of the entire simulation period (Russo et al. 2013); and (iii) calculate drought indices in the future using probability distribution parameters fitted to precipitation data in the baseline period (Touma et al. 2015; Kwon and Sung 2019; Wan et al. 2018; Taylor et al. 2013), or by mapping future precipitation to the standardized empirical probability in the baseline period (Sun et al. 2019). All the three approaches have their pros and cons. The first approach fits probability model to long-term nonstationary precipitation series directly. It can be implemented easily with all data used for classical SPI calculation, but SPI is computed under the stationary assumption (Türkeş and Tatlı 2009). Nonstationary distributions are more and more advocated for nonstationary situations, and the nonstationary modeling of the drought index often outperforms the stationary modeling (Russo et al. 2013; Y. Wang et al. 2015; Das et al. 2020). The second approach takes the nonstationarity into account directly by fitting time-varying distributions, and the drought index calculation is more robust because it uses data over the whole period rather than a
subperiod (Russo et al. 2013), but it may lead to the loss of important temporal characteristics of droughts as discussed earlier. The third approach is most commonly used in the future drought assessment in the literature. It is known that SPI is a relative drought index which measures the deviation of precipitation from the normal of a specific period (McKee et al. 1993), but there is no consensus about which period could be specified for computing the normal precipitation, whereas we have shown that the selection of the reference period brings large uncertainty in drought assessment. On the other hand, climate normal changes over time. For instance, the World Meteorological Organization (2017) defines the climatological standard normal as the most-recent 30-yr period finishing in a year ending with 0 (1981–2010, 1991–2020, and so forth). Whatever the option of the reference period by different authors with the third approach, one thing is sure: climate normals in the past are used for assessing drought in the future with the third approach, sometimes in the far future 100 years ahead. That is equivalent to assessing the current drought situations using precipitation data in the early half of the twentieth century, which was conducted by nobody in the literature, because it is not meaningful. Thus, it is questionable to assess droughts in the future with the climate normal in a baseline period far away from the local time of the future. How to effectively conduct future drought assessment is worth further investigation.

c. How reliable are the bootstrap method and Monte Carlo method for uncertainty analysis?

The bootstrap method is often used as a major tool for evaluating the uncertainty in not only SPI calculation (e.g., Hu et al. 2015; Vergni et al. 2017) but also flood frequency analysis (e.g., Bomers et al. 2019). Here we would like to investigate the reliability of the bootstrap method and Monte Carlo method for uncertainty analysis. The procedure is as follows: 1) using a set of gamma distribution parameters ($\alpha = 4.47, \beta = 49.05$) that are fitted to the 30-day cumulative precipitation series at the 180th day at Fuzhou, generate 10,000 sets of samples with lengths of 30, 60, and 90, respectively; 2) use the bootstrap and Monte Carlo methods to estimate 90% CI for $\alpha$, $\beta$, and SPI for those 10,000 simulations of different length, respectively; and 3) calculate the percentage of 10,000 simulations that the true values of the parameters fall into the 90% CI of bootstrap and Monte Carlo methods, respectively. The results are presented in Table 6, which show that 1) the rates of parameters falling in their 90% CIs by bootstrap method and Monte Carlo method are always smaller than 90%, which indicates that the uncertainties are underestimated by both methods; 2) with the increase of data size, the rates of parameters falling in the their 90% CIs get closer to 90%, indicating that the underestimation of uncertainty is smaller with the increase of data size; and 3) the covering rates by the bootstrap method are always smaller than those by the Monte Carlo method, which means that the Monte Carlo method gives a better estimation of CIW than the bootstrap method for uncertainty analysis.

Our results are in agreement with those of Kyselý (2010), who showed that both bootstrap and Monte Carlo method (referred to as parametric bootstrap in their work) underestimate the width of the confidence intervals in cases of GEV distribution and the generalized Pareto distribution, and the parametric bootstrap is superior to the nonparametric one. Basing on our results for the gamma distribution and Kyselý’s results for the GEV distribution and the generalized Pareto distribution, we suggest using Monte Carlo method as the first choice for the uncertainty analysis in the hydrometeorological probability modeling fitting and frequency analysis.

5. Conclusions

Using the daily precipitation data of 64 meteorological stations in four climate regions in China from 1960 to 2016, we investigated the performance of two two-parameter distributions (i.e., gamma distribution and Weibull distribution) and three three-parameter distributions (i.e., generalized extreme value distribution, Pearson type III distribution, and Tweedie distribution) in fitting the cumulative precipitation of different time scales, and the uncertainties in the SPI calculation with different time scales, record lengths and reference periods with the bootstrap and Monte Carlo methods. The influence of uncertainty of the SPI calculation on drought assessment is also investigated. The main outcomes from the present research are as follows:

1) While all five candidate distributions can describe cumulative precipitation series well according to KS test, the optimal choice for calculating SPI with time scales from 30 to 365 days in China is the gamma distribution. Among the time scales investigated in the present study (i.e., 10, 20, 30, and 90 days), the recommended minimal temporal scale appropriate for SPI calculation is 20 days for the humid region, 30 days for the semihumid/semiarid region and Tibetan Plateau (mostly its eastern part), and 90 days for the arid region. That is, from the perspective of assessing droughts over all of China, 90 days (or 3 months) is the recommended time scale. The major factor affecting the time scale option is the percentage of zero values in cumulative precipitation series. The longer the data recorder length, the less zeros can be contained in a precipitation series to make it follow the standard normal distribution after standardization.

2) The uncertainty in SPI calculation decreases with the increase of time scale and record length, essentially as a
consequence of the decrease of the confidence interval width of gamma distribution parameters with the increase of time scale and record length. However, there is little improvement for the parameter estimation with data size longer than ~70 years. Therefore, from the perspective of reducing uncertainty, the optimal record length for calculating SPI is about 70 years.

3) The uncertainty in SPI calculation affect drought classification greatly, and there are larger uncertainties in assessing extreme droughts than assessing moderate ones because the uncertainty in SPI calculation increases with the increase of absolute values of SPI. But the uncertainty of a drought calculation has no impact on the drought temporal characteristics when analyzing the trend with a nonparametric trend test method.

4) The uncertainty resulting from the selection of the reference period will bring large uncertainties in drought assessment. It is very common to assess future droughts by calculating drought indices in the future based on probability distribution parameters or empirical distributions of precipitation data in the baseline period. On the one hand there is considerable uncertainty in the drought assessment resulting from the choice of reference period, on the other hand cautions must be taken when assessing droughts in the future with the climate normal in a baseline period far away from its local time of the future.

Acknowledgments. The financial support from the National Natural Science Foundation of China (41961134003) and the Polish National Science Centre (contract 2018/30/Q/ ST10/00654) is gratefully acknowledged.

Data availability statement. Precipitation data associated with the paper are provided by the national meteorological information center of China Meteorological Administration (available at http://data.cma.cn/).

REFERENCES


