The Impact of Two Sources of Uncertainty on the Drought Index SPEI and Drought Assessment (Postprint)

Authors: Han Rucun

Date: 2022-10-21T00:00:00+00:00

Abstract

Against the backdrop of global climate warming, drought events are becoming more frequent and persistent, with their impact scope gradually expanding. When employing drought indices for drought assessment and research, the evaluation results often exhibit certain uncertainties due to the influence of multiple factors. Taking the Heihe River Basin as the study area, this study analyzes the impacts of two uncertainty sources—probability distribution models and parameter estimation errors—on the Standard ized Precipitation Evapotranspiration Index (SPEI) and drought characteristic variables (drought intensity, drought peak, and drought duration). The results indicate that both sources affect SPEI and drought characteristic variables, with the impact becoming more pronounced as SPEI becomes more extreme. Their influence on extreme and severe droughts is substantially greater than that on mild and moderate droughts. For extreme and severe droughts, probability distribution models contribute greater uncertainty to drought intensity and drought peak, while parameter estimation errors contribute greater uncertainty to drought duration. These research findings can provide support for accurate drought assessment and offer theoretical backing for formulating more precise and effective drought mitigation decisions in drought prevention and mitigation efforts, thereby avoiding potential inadequacies in disaster reduction capacity or waste of drought mitigation resources.

Full Text

Abstract

Drought issues are becoming more frequent and persistent against the backdrop of global warming, with their impact range gradually expanding. When employing drought indices for drought assessment and research, the results often exhibit certain uncertainties due to multiple influencing factors. This study analyzes the effects of two uncertainty sources—probability distribution models and parameter estimation errors—on the Standardized Precipitation Evapotranspiration Index (SPEI) and drought assessment in the Heihe River Basin. The results demonstrate that both sources affect SPEI and drought characteristic variables (drought intensity, drought peak, and drought duration), with their influence becoming more pronounced as the index becomes more extreme. Their impact on extreme and severe drought far exceeds that on mild and moderate drought. For extreme and severe drought, probability distribution models introduce greater uncertainty in drought intensity and drought peak, while parameter estimation errors cause greater uncertainty in drought duration. These findings can support accurate drought assessment and provide theoretical support for formulating more precise and effective drought mitigation decisions, thereby avoiding potential inadequacies in disaster reduction capacity or waste of drought-resistant resources.

Keywords: probability distribution model; parameter estimation; uncertainty; Standardized Precipitation Evapotranspiration Index (SPEI); Heihe River

1 Introduction

Global warming represents one of the most serious environmental challenges facing humanity, attracting widespread attention from governments and the public worldwide. Against this backdrop, drought has become a focal issue in ecology, meteorology, and hydrology. Drought indices serve as important tools for drought research. However, due to influences from samples, models, and parameters, results from drought monitoring and assessment using these indices typically contain uncertainties.

Previous studies have extensively investigated sample uncertainty in drought indices. For instance, Hong et al. [?] verified the sampling uncertainty of drought index calculations using Monte Carlo simulation methods, while Zhang et al. [?] evaluated the impact of sample uncertainty on SPI-based drought assessment. Regarding model selection, different probability distribution models are required to fit relevant data when calculating drought indices. Due to varying climatic conditions across regions, the optimal theoretical probability distribution models differ. Studies indicate that the Log-Logistic distribution is optimal for the United States and China' s Poyang Lake Basin [?], the Generalized Normal distribution better suits Brazil [?], and the Generalized Extreme Value distribution is recommended for Europe [?]. Concerning parameters, previous research has focused more on different parameter estimation methods. Begueria et al. [?] compared two parameter estimation methods—Maximum Likelihood Estimation and Unbiased Probability Weighted Moments—for SPEI calculation. However, the uncertainty arising from model parameter estimation errors themselves and the uncertainty introduced by different theoretical probability distribution models in drought assessment have received less attention in existing research.

Given this context, this paper examines the influence of different probability distribution models and parameter estimation errors on the drought index SPEI and drought assessment. Since SPEI considers both precipitation and evapotranspiration effects on drought, it was selected as the target analysis index. The Heihe River Basin, the second largest inland river basin in Northwest China, serves as the study area. Drought has always been the primary natural disaster in this basin [?], with 21 drought and flood disasters occurring cumulatively over recent years, of which droughts accounted for 71% [?]. Under global warming, drought conditions in the basin' s middle and lower reaches have become increasingly severe, with the most severe autumn-winter-spring consecutive drought since 2000 occurring in 2001 [?]. Frequent drought disasters not only make the local ecological environment exceptionally fragile but also constrain healthy economic and social development. The middle reaches' corridor plain represents a favorable region for agricultural development. However, controlled by both the northwest airflow from the east coast of Asia and the downdraft at the edge of large terrain, this area is one of the driest regions with human activity at the same latitude globally. The peak water demand period for crops occurs from May to July, which coincides with the basin' s dry season, creating exceptionally prominent "bottleneck" drought conflicts during summer irrigation [?].

This study evaluates the impact of six probability distribution models and model parameter estimation errors on SPEI and characteristic variables including drought intensity, drought peak, and drought duration. The objective is to provide support for accurate drought assessment and theoretical backing for more precise and effective drought mitigation decisions to avoid potential inadequacies in disaster reduction capacity or waste of drought-resistant resources.

2 Study Area and Data

2.1 Study Area Overview

The Heihe River Basin is located in the middle of the Hexi Corridor, between 97°37 ~102°06 E and 37°44 ~42°40 N, covering an area of approximately 14.29$×10^{4}$ km^2 [Figure 1: see original paper]. Influenced by both westerly circulation and polar cold air masses, the basin' s climate exhibits distinct north-south and east-west differences, spanning three climate zones from south to north: semi-arid and semi-humid, semi-arid, and extremely arid desert regions [?]. The southern mountainous area in the upper reaches receives 400–500 mm of mean annual precipitation with a mean annual temperature of 2.8–7.6°C and a cool climate. The middle reaches receive 100–150 mm of precipitation with a mean annual temperature of 6.0–8.0°C. The lower receives less than 50 mm of precipitation, with the Ejin Plain in the lower reaches having a mean annual temperature of approximately 2.0°C. Overall, precipitation across the basin is scarce but concentrated, with long sunshine hours, strong evaporation, and large diurnal temperature variations. In recent years, the basin' s mean temperature has shown a significant upward trend, with the Ejin Banner station showing an increase rate of 0.42°C · $(10a)^{-1}$ [?], exacerbating drought issues in the basin.

2.2 Data Sources

Nine national meteorological stations within the basin were selected as research stations . After examining data completeness and continuity, daily precipitation and temperature datasets from 1960–2017 were used for SPEI calculation. Data were obtained from the National Meteorological Science Data Center (http://data.cma.cn/), with good data quality.

provides basic information for the meteorological stations, including station names, latitude, longitude, mean annual precipitation, and mean annual temperature.

3 Methods

3.1 Standardized Precipitation Evapotranspiration Index (SPEI) and Drought Characteristic Variables

Based on Vicente-Serrano et al. [?], SPEI considers temperature effects on drought and is commonly used for drought characteristic analysis in temperature-sensitive regions. The three-parameter Log-Logistic distribution model is recommended to describe the difference between precipitation and potential evapotranspiration. The calculation process follows these steps:

First, probability distribution models are used to fit the cumulative difference series between monthly precipitation (P) and potential evapotranspiration (PET) for each month j, obtaining optimal parameter sets and corresponding variance-covariance matrices, where j represents the month and n represents sample size.

Drought events are defined when the drought index falls below a set threshold. For example, when SPEI is between -1.5 and -2.0, a severe drought event occurs. Based on run theory [?], drought duration is defined as the time period when the drought index remains between threshold levels. Drought intensity is defined as the ratio of the cumulative drought index value during a drought event to drought duration. Drought peak is defined as the minimum drought index value during a drought event.

shows drought classification based on SPEI values, ranging from extremely wet (SPEI $>$ 2.0) to extremely dry (SPEI $\leq$ -2.0).

3.2 Probability Distribution Models and Goodness-of-Fit Tests

To evaluate the impact of different probability distribution models on SPEI estimation, six distributions were selected for calculation: three-parameter distributions (Log-Logistic, Generalized Extreme Value, Pearson Type III) and two-parameter distributions (Normal, Weibull, Logistic). Probability density functions for each distribution are provided in the references [?].

Maximum Likelihood Estimation was used for parameter estimation of different distribution models, with Anderson-Darling (AD) and Kolmogorov-Smirnov

(KS) tests employed for goodness-of-fit evaluation. The KS test calculates the maximum difference between given and empirical cumulative distribution functions without assuming a specific distribution. The AD test is an improvement on KS, being more sensitive to outliers at both distribution ends. Detailed calculation methods are provided in Fadhilah et al. [**?**].

3.3 Multivariate Normal Method

The multivariate normal method [**?**] was used to quantify uncertainty in SPEI estimates caused by parameter estimation errors. Based on the asymptotic normal assumption of optimal parameter sets and variance-covariance matrices, N groups of parameter sets were randomly generated (N = 2000 in this study). SPEI values were calculated for each parameter group, and the 95% confidence interval was defined using the 2.5th and 97.5th percentiles. This confidence interval represents the uncertainty range considering parameter estimation error effects.

Potential evapotranspiration was calculated using the Thornthwaite method [**?**] as the basis for fitting the difference series between precipitation and potential evapotranspiration using different probability distribution models.

4 Results

4.1 Effects of Probability Distribution Models

4.1.1 Goodness-of-Fit Test Results [Figure 2: see original paper] shows test statistics for different probability distribution models for January data series. All stations' statistics are below critical values at the 0.05 significance level, indicating these distributions can adequately fit the series. The black dashed boxes indicate minimum statistics values. Log-Logistic distribution shows the best fit among three-parameter models, with the smallest AD statistics at most stations (66.7%). Among two-parameter models, Logistic distribution shows the best fit with the smallest statistics at 55.6% of stations, while Cauchy distribution shows the poorest fit.

Results for other months show similar patterns. Log-Logistic and Logistic distributions are the best-fitting models among three-parameter and two-parameter models, respectively. While all selected distributions are suitable for fitting precipitation and potential evapotranspiration difference series in the study area, further investigation is needed to determine how different distribution choices affect SPEI uncertainty.

4.1.2 Impact on SPEI Using all distributions to calculate SPEI for each station, results were compared with those from the optimal distribution (Log-Logistic). Taking Ejin Banner station in January as an example, Log-Logistic distribution calculated SPEI values of -2.31 and 2.13, indicating extreme drought and extreme wet conditions, respectively. In contrast, Cauchy distribution calculated values of -0.89 and 0.91, indicating moderate drought and moderate wet conditions. Apart from Cauchy distribution, Logistic distribution shows the largest differences.

[Figure 3: see original paper] presents residuals between SPEI values from alternative distributions and the optimal Log-Logistic distribution. The pattern shows that as |SPEI| increases (wetter or drier conditions), differences increase, particularly for moderate drought/wetness and above. When -1 < SPEI < 1 (no or mild drought), residual differences range from -0.24 to 0.31. For moderate drought/wetness, the range is [-0.52, 0.58]. For severe drought/wetness, it expands to [-0.89, 0.91], and for extreme drought/wetness, it reaches [-1.31, 1.35]. The more extreme the value, the greater the uncertainty introduced by distribution model selection.

4.1.3 Impact on Drought Characteristic Variables Based on run theory, drought duration, intensity, and peak were extracted for all stations and classified according to drought categories. [Figure 5: see original paper] shows statistics for drought characteristics across all stations under different drought categories. The 95% confidence intervals for mild, moderate, severe, and extreme drought intensity are [0.5, 2.1], [2.1, 4.3], [4.3, 8.9], and [8.9, 16.2], respectively. Uncertainty increases with drought category, with severe and extreme drought intensity uncertainty being 11–20 times greater than mild drought uncertainty. Similar patterns appear for drought peak, with 95% confidence intervals of [0.5, 1.2], [1.2, 1.8], [1.8, 2.4], and [2.4, 3.1] for the four categories. In contrast, drought duration uncertainty shows much smaller variation across categories, with mild drought duration showing the smallest uncertainty.

4.2 Effects of Parameter Estimation Errors

4.2.1 Parameter Estimation and Errors Maximum Likelihood Estimation was used to calculate distribution parameters (shape, scale, location). shows optimal values and 95% confidence intervals for parameters of the three-parameter Log-Logistic distribution at Yeniugou, Jiuquan, and Ejin Banner stations for January. The 95% confidence intervals for shape, scale, and location parameters are [10.2, 18.7], [1.2, 1.8], and [-43.2, -28.7] for Yeniugou; [8.9, 15.4], [1.1, 1.6], and [-38.7, -25.4] for Jiuquan; and [7.3, 12.8], [0.9, 1.4], and [-32.1, -21.5] for Ejin Banner.

4.2.2 Impact on SPEI [Figure 4: see original paper] and [Figure 6: see original paper] show SPEI estimates and 95% confidence intervals considering parameter estimation uncertainty. The confidence interval width increases as SPEI becomes more extreme. For no or mild drought/wetness, the 95% confidence interval is [-0.8, 0.8]. For moderate drought/wetness, it expands to [-1.5, 1.5]. For severe drought/wetness, it increases to [-2.3, 2.3], and for extreme drought/wetness, it reaches [-3.1, 3.1]. The more extreme the drought or wetness condition, the greater the uncertainty caused by parameter estimation errors.

4.2.3 Impact on Drought Characteristic Variables [Figure 5: see original paper] also shows drought characteristics under different drought categories resulting from parameter estimation uncertainty. The 95% confidence intervals for mild, moderate, severe, and extreme drought intensity are [0.8, 2.4], [2.4, 4.8], [4.8, 10.2], and [10.2, 18.5], respectively. For drought peak, the intervals are [0.6, 1.3], [1.3, 1.9], [1.9, 2.6], and [2.6, 3.4]. For drought duration, the intervals are [2.1, 4.2], [4.2, 6.8], [6.8, 9.5], and [9.5, 12.3] months.

Uncertainty increases with drought category for intensity and peak, with severe and extreme drought showing significantly greater uncertainty ranges than mild and moderate drought. For drought duration, uncertainty differences across categories are relatively small, with mild drought duration showing the smallest uncertainty, followed by severe and extreme drought, and moderate drought showing intermediate uncertainty.

5 Discussion

Goodness-of-fit test results indicate that all six selected distributions are suitable for fitting precipitation and potential evapotranspiration difference series in the study area. Among three-parameter models, Log-Logistic shows the strongest applicability, while among two-parameter models, Logistic performs best [Figure 2: see original paper]. If only fitting effectiveness is considered, Logistic distribution could serve as a candidate for fitting difference series due to its simplicity with only two parameters. However, comparing SPEI values reveals substantial differences between Logistic and Log-Logistic distributions at extreme values. Although two-parameter models have fewer parameters and simpler calculations, they introduce greater uncertainty in SPEI calculations [Figure 3: see original paper].

Calculating the relative contribution of different probability distribution models to SPEI uncertainty reveals that distribution models are a major source of SPEI uncertainty. Vergni et al. [?] also found that, besides data series length, probability distribution models are the primary source of SPEI uncertainty. Laimighofer and Laaha [?] concluded that distribution models contribute significantly to SPEI uncertainty. This study further reveals that higher drought (or wetness) categories show greater uncertainty from different distribution models, particularly for severe and extreme drought, where uncertainty increases significantly and can even affect drought category determination [Figure 3: see original paper]. Additionally, distribution models introduce uncertainty in assessing drought intensity, peak, and duration. As drought category increases, uncertainty in drought intensity and peak caused by distribution models increases, with severe and extreme drought showing uncertainty ranges 7–10 times greater than mild and moderate drought [Figure 5: see original paper]. This indicates that uncertainty in severe and extreme drought intensity and peak caused by probability distribution models is substantial and cannot be ignored. However, uncertainty in drought duration does not change significantly across categories, with relatively similar uncertainty ranges [Figure 5: see original paper].

Parameter estimation errors also cause SPEI uncertainty. Zhang et al. [?] found that both Log-Logistic and Generalized Extreme Value distributions adequately fit the Heihe River Basin. Three-parameter models show better fitting performance than two-parameter models. Although two-parameter models with fewer parameters and simpler calculations can serve as alternative distributions for fitting precipitation and potential evapotranspiration difference series, they can cause considerable uncertainty in SPEI calculations. As drought categories become more extreme, uncertainty in SPEI caused by model parameter estimation errors increases. This study, using Log-Logistic distribution, reaches similar conclusions: more extreme conditions lead to greater SPEI uncertainty from parameter estimation errors [Figure 4: see original paper]. Additionally, higher drought categories show greater uncertainty in drought intensity and peak caused by parameter estimation errors, with severe and extreme drought showing uncertainty ranges significantly larger than mild and moderate drought [Figure 5: see original paper]. Uncertainty ranges for drought duration across categories are relatively similar [Figure 5: see original paper].

Compared with probability distribution models, parameter estimation errors produce wider 95% confidence intervals for mild, moderate, and severe drought SPEI, indicating greater uncertainty. Parameter estimation errors also cause greater uncertainty in severe drought duration, while uncertainty ranges for other drought characteristic variables are similar. This demonstrates that both uncertainty sources affect drought characteristic assessment, particularly impacting severe and extreme drought intensity and peak. Therefore, drought assessment requires careful selection of optimal probability distribution models and parameter estimation methods to reduce uncertainty and provide support for accurate drought assessment and forecasting.

It should be noted that besides the two uncertainty sources discussed, other sources exist in drought assessment, such as different potential evapotranspiration estimation methods. Aadhar and Mishra [?] pointed out that potential evapotranspiration estimation methods contribute significantly to uncertainty in final drought assessment results. Due to data limitations, this study only used the Thornthwaite method to calculate potential evapotranspiration, which also introduces some uncertainty into assessment results. This aspect will be comprehensively considered in future research.

6 Conclusions

This study examines the Heihe River Basin and SPEI to investigate the effects of two uncertainty sources—probability distribution models and parameter estimation errors—on drought index and characteristic variables including drought intensity, drought peak, and drought duration. The main conclusions are:

1. **Probability distribution model selection is crucial for drought assessment reliability.** Overall, the three-parameter Log-Logistic distribution demonstrates the strongest applicability, while the two-parameter Logistic distribution performs best among two-parameter models. However, two-parameter models, despite having fewer parameters and simpler calculations, introduce considerable uncertainty in SPEI calculations. As drought categories become more extreme, uncertainty from different distribution models increases, significantly affecting severe and extreme drought assessment. Uncertainty in drought intensity and peak caused by distribution models increases with drought category, with severe and extreme drought uncertainty ranges reaching over 7 times those of mild and moderate drought. Uncertainty in drought duration shows little variation across categories.

2. **Parameter estimation errors also contribute to SPEI uncertainty.** The more extreme the SPEI value, the greater the uncertainty caused by parameter estimation errors. Higher drought categories show greater uncertainty in drought intensity and peak, with severe and extreme drought uncertainty ranges significantly exceeding those of mild and moderate drought. Uncertainty ranges for drought duration across categories are relatively similar. Compared with distribution models, parameter estimation errors produce greater uncertainty for mild, moderate, and severe drought SPEI, as well as for severe drought duration.

3. **Both uncertainty sources affect drought assessment**, particularly for severe and extreme drought intensity and peak. Compared with distribution models, parameter estimation errors cause greater uncertainty for mild, moderate, and severe drought SPEI, and greater uncertainty for severe drought duration. This indicates that both uncertainty sources must be considered for precise drought assessment. Selecting optimal probability distribution models and parameter estimation methods can reduce uncertainty in drought assessment results.

References

[1] Jin Ling, Wang Yongfang, Guo Enliang, et al. Evaluation of drought hazards in Inner Mongolia based on SPEIbase v.2.6 dataset[J]. Arid Land Geography, 2022, 45(3): 695-705.

[2] Hong Xingjun, Guo Shenglian. Hydrological drought assessment and uncertainty analysis for the upper Yangtze River[J]. China Flood & Drought Management, 2018, 28(10): 14-20.

[3] Zhang Ruihan, Wang Yimin, Guo Aijun. Impact of sample uncertainty on SPI index based drought evaluation[J]. Journal of Northwest A & F University (Natural Science Edition), 2019, 47(11): 134-142, 154.

[4] Guttman N B. Accepting the standardized precipitation index: A calculation algorithm[J]. Journal of the American Water Resources Association, 1999, 35(2): 311-322.

[5] Hong Xingjun, Guo Shenglian, Zhou Yanlai. Applicability of standardized precipitation index with alternative distribution functions[J]. Journal of Water Resources Research, 2013, 2(1): 33-41.

[6] Liu Liu, Liu Lili, Suo Ying. Spatiotemporal evolution of hydro-meteorological variables in the Heihe River Basin in recent 53 years[J]. Arid Zone Research, 2017, 34(3): 465-478.

[7] Sienz F, Bothe O, Fraedrich K. Monitoring and quantifying future climate projections of dryness and wetness extremes: SPI bias[J]. Hydrology and Earth System Sciences, 2011, 8(6): 10635-10677.

[8] Gabriel C B, Monica C M. Inadequacy of the gamma distribution to calculate the standardized precipitation index[J]. Revista Brasileira de Engenharia Agrícola e Ambiental, 2015, 19(12): 1129-1135.

[9] Stagge J H, Tallaksen L M, Gudmundsson L, et al. Candidate distributions for climatological drought indices (SPI and SPEI)[J]. International Journal of Climatology, 2015, 35(13): 4027-4040.

[10] Begueria S, Vicente-Serrano S M, Reig F, et al. Standardized precipitation evapotranspiration index (SPEI) revisited: Parameter fitting, evapotranspiration models, tools, datasets and drought monitoring[J]. International Journal of Climatology, 2014, 34(10): 3001-3023.

[11] Vicente-Serrano S M, Begueria S, Lopez-Moreno J I. A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index[J]. Journal of Climate, 2010, 23(7): 1696-1718.

[12] Yevjevich V. An objective approach to definitions and investigations of continental hydrologic droughts[J]. Journal of Hydrology, 1969, 7(3): 353, doi: 10.1016/0022-1694(69)90110-3.

[13] Fadhilah Y, Foo H M, Jamaludin S. Rainfall characterisation by application of standardised precipitation index (SPI) in Peninsular Malaysia[J]. Theoretical and Applied Climatology, 2014, 115(3-4): 503-516.

[14] Nixon R M, Wonderling D, Grieve R D. Non-parametric methods for cost effectiveness analysis: The central limit theorem and the bootstrap compared[J]. Health Economics, 2010, 19(3): 316-333.

[15] Degeling K, Ijzerman M J, Koopman M, et al. Accounting for parameter uncertainty in the definition of parametric distributions used to describe individual patient variation in health economic models[J]. BMC Medical Research Methodology, 2017, 17: 170, doi: 10.1186/s12874-017-0437-y.

[16] Zhang Jingtian, Xi Haiyang. Spatiotemporal dynamics of groundwater levels in a desert riparian forest and its response to surface runoff[J]. Arid Land

Geography, 2020, 43(2): 388-397.

[17] Ren Chaoxia, Lu Yuqi, Yang Dayuan. Study on drought and flood changes in recent 2000a in Heihe River Basin[J]. Journal of Arid Land Resources and Environment, 2009, 23(4): 90-93.

[18] Cao Ling, Dou Yongxiang, Zhang Deyu. Effect of climate change on ecological environment of Heihe Field[J]. Journal of Arid Meteorology, 2003, 21(4): 45-49.

[19] Zhang Ying. Uncertainty analysis of two typical meteorological drought indexes: A case study in Heihe River Basin[D]. Beijing: China University of Geosciences (Beijing), 2020.

[20] Feng Q, Cheng G D, Endo K N. Towards sustainable development of the environmentally degraded River Heihe Basin, China[J]. Hydrological Sciences Journal, 2001, 46(5): 647-658.

[21] Vergni L, Di L B, Todisco F, et al. Uncertainty in drought monitoring by the standardized precipitation index: The case study of the Abruzzo region (central Italy)[J]. Theoretical and Applied Climatology, 2017, 128: 13-26.

[22] Laimighofer J, Laaha G. How standard are standardized drought indices? Uncertainty contributions for the SPI & SPEI case[J]. EGU General Assembly, 2020: EGU2020-4716, doi: 10.5194/egusphere-egu2020-4716.

[23] Zhang Y, Li Z. Uncertainty analysis of standardized precipitation index due to the effects of probability distributions and parameter errors[J]. Frontiers in Earth Science, 2020, 8: 76, doi: 10.3389/feart.2020.00076.

[24] Aadhar S, Mishra V. Increased drought risk in South Asia under warming climate: Implications of uncertainty in potential evapotranspiration estimates[J]. Journal of Hydrometeorology, 2020, 21(12): 2979-2996.

Note: Figure translations are in progress. See original paper for figures.