Study on drought events in China based on time-varying nested Archimedean-copula function

Ziyang Zhaoa, Hongrui Wanga,*, Qiuyang Shiab and Cheng Wangb

a Beijing Key Laboratory of Urban Hydrological Cycle and Sponge City Technology, College of Water Sciences, Beijing Normal University, Beijing 100875, China
b Environment Science Division, Argonne National Laboratory, Lemont, IL 60439, USA

*Corresponding author. E-mail: henrywang@bnu.edu.cn

ABSTRACT

Drought forecasting, which can enable contingency actions to be implemented in advance of a drought, plays a significant role in reducing the risks and impacts of drought. In this study, a simulation framework of the occurrence probability of drought events based on a nested copula function and Gibbs sampling is proposed to effectively compensate for the high-dimensional problems and lack of initial data in traditional methods. Precipitation data of 718 meteorological stations from 1961 to 2018 in China was analyzed. The results showed that the occurrence location of drought events was mainly concentrated from 35° to 42° north latitude and 105° to 120° east longitude, with the occurrence time mainly concentrated from September to November. The Archimedean-copula function, constructed based on latitude, longitude, and occurrence time, could precisely determine the spatiotemporal joint probability distribution of drought events (RMSE: 0.01). The optimal time-varying nested Archimedean-copula functions were obtained from February to May (spring), June to September (summer) and October to January (autumn and winter). Compared with the nested Archimedean-copula function, the accuracy of Gibbs sampling and simulation based on the time-varying nested Archimedean-copula function was increased by 84.05% latitude and 69.76% longitude. The results provide an effective means for scientific drought forecasting, and water resource management departments can take preventive measures at an early stage.

Key words: Archimedean-copula function, drought, Gibbs sampling, probability distribution, run analysis theory

HIGHLIGHTS

• This is the first to attempt to propose the quantitative framework of drought forecast based on run analysis theory, Archimedean-copula function, and Gibbs sampling.
• The optimal time-varying nested Archimedean-copula functions are constructed.
• Drought prediction accuracy is improved.

1. INTRODUCTION

Water resources, as one of the most basic conditions for human survival and development, are essential natural and strategic resources that maintain the functions of natural ecosystems and support the evolution of socioeconomic systems (Lai et al. 2018; Yang et al. 2021). Currently, many countries and regions in the world are facing water resources issues, such as water resource scarcity (Zhao et al. 2020a, 2020b), water environment pollution (Zinatloo-Ajabshir et al. 2021a, 2021c) and water ecological destruction (Zinatloo-Ajabshir et al. 2021b). The problems of China are mainly manifested by fewer per capita water resources, uneven spatiotemporal distribution, serious pollution, low utilization efficiency, and inconsistency in regional economic development (Qian et al. 2014). To ensure the sustainable use of water resources in the future, China has adopted a series of safeguard measures in recent years, such as the South-to-North Water Diversion Project and the ‘three red lines’ of water resources (Zhang 2009; Zang et al. 2016). However, due to the scarcity, circulation, and uncertainty of water resources, the water scarcity risk always exists (Wang et al. 2018). Moreover, water scarcity in China has tended to increase with the rapid development of the economy and society (Hoekstra 2014). Reducing the population affected by water scarcity is one of the 17 Sustainable Development Goals identified by the United Nations, and is also a significant agenda for China (Sun et al. 2019).

It is generally believed that water scarcity is guided by the imbalance of water supply and water demand. The water demand is mainly determined by the level of socioeconomic development and the size of the population, and it is relatively stable over a period of time and is less likely to undergo dramatic changes (Qian et al. 2014). Thus, the water supply amount has become

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (http://creativecommons.org/licenses/by/4.0/).
the dominant factor in the occurrence of regional water scarcity. The total amount of regional water resources is represented by surface and underground water yields, that is, the sum of surface runoff and precipitation infiltration recharge (Wan et al. 2016). Furthermore, droughts can be divided into four categories: meteorological drought (Duan et al. 2014), hydrological drought (Li et al. 2020), agricultural drought, and socio-economic drought (Mishra & Singh 2010, 2011). Drought is a natural phenomenon where rainfall in a certain area has been reduced for a period of time, thus affecting society, economy, and environment. In addition, drought is a normal cyclical climatic event characterized by many complex factors, related not only to rainfall, but also to topography, soil properties, and water facilities (Chang 1991). Drought events, as represented by meteorological drought, are mainly focused on precipitation, and are interrelated hydrological events, such as drought duration, drought intensity and drought severity (Wong et al. 2013; Choi et al. 2018). Generally, drought has become a major problem in regional water scarcity research. At present, the research direction of drought events can be summarized into three categories: (a) the identification, classification, and real-time monitoring of drought events based on GIS and remote sensing (Liu & Kogan 1996; Gao et al. 2014); (b) the construction of drought event indicator systems and risk based on the theory of disaster systems (Mishra & Singh 2010, 2011); (c) the probability modeling and evaluation of drought events based on mathematical statistics and multiple interdisciplinary studies (Chung & Salas 2000; Shiau 2006). To reduce losses and avoid the adverse consequences caused by drought events, it is urgent to start research from the identification and the occurrence probability of drought events.

In recent years, with the development of statistical methods and machine learning technology, a variety of studies have employed a number of forecasting methods, such as Markov chain (Lohani & Loganathan 1997), time series analysis (Mishra & Desai 2005), support vector machine (Ganguli & Reddy 2014), and artificial neural network (Araghil et al. 2017) to establish drought forecast models. However, these methods are limited to univariate analysis of drought characteristics. For drought events that include multiple related variables, univariate analysis is limited in its ability to reflect the joint distribution of multiple variables. In general, dynamical forecasts based on the copula model are a promising tool and have been increasingly used for drought forecasting globally because they can flexibly perform joint probability analysis of multiple variables (Qian et al. 2018).

At present, the copula method has been widely used in drought studies due to its ability to address uncertainties in multifactors, conditional probability analysis, and recurrence period analysis (Dupuis 2007; Genest & Favre 2007; Qian et al. 2018). Copula theory, which originated from financial risk analysis, was first proposed as an important tool for constructing multivariate joint distributions (Genest et al. 2007). In fact, it is a function that connects the joint distribution with the edge distribution, which is also called a connection function (Ma et al. 2013). The copula method was first used to analyze hydrological problems caused by drought in 2006. In recent years, many scholars have used bivariate and trivariate copula functions to analyze precipitation in northern Iran (Mirabbsi et al. 2012), India (Ganguli & Reddy 2012), peninsular Malaysia (Yusof et al. 2013), and Texas (Reddy & Singh 2014); Chinese scholars have also established a bivariate copula function model of drought duration and drought intensity to study Dujiangyan (Li et al. 2016), southwest China (Zuo et al. 2014), and the Aksu River Basin (Zhang et al. 2014). Although the copula model has unique advantages and extensive applications in the field of drought forecasting, there are still two shortcomings (Qian et al. 2018). First, the traditional copula model is more complicated when dealing with high-dimensional problems. Second, drought forecasting based on the copula model cannot compensate for the lack of initial data, which affects forecast accuracy. This scenario requires a combination of the copula model and other approaches to improve the forecast effect. Nested function and Gibbs sampling provide an effective research tool for resolving the above problems.

In this paper, a simulation framework that can determine the occurrence probability of drought events was proposed. The core of the simulation framework is run analysis theory, the time-varying Archimedean-copula function model, and Gibbs sampling. As an example, China was selected to assess serious water resource situations based on precipitation data of 718 meteorological stations from 1961 to 2018. Generally, the main objectives of this study are as follows: (1) to identify drought events; (2) to obtain a spatially adaptive time-varying copula joint distribution; and (3) to study the nationwide regular patterns of drought events. This study will provide a reference for researchers and decision-makers to guide water resource management and agricultural development.

2. MATERIALS AND METHODS

2.1. Study area and data

China has significant regional differences in natural conditions, which determine the spatial characteristics of meteorological elements. Precipitation is a key factor affecting regional agricultural systems, and the features in China show an uneven...
distribution, where the north is drier and colder than the south (Wang et al. 2019). A total of 718 meteorological stations, with daily precipitation data for a period from 1961 to 2018, were provided by the National Meteorological Information Center of China (Figure 1) (http://data.cma.cn/). Precipitation data were homogenized, and missing values were interpolated from ten adjacent stations (Shahid et al. 2018; Shahid & Rahman 2021).

2.2. Run analysis theory

Run analysis theory, as a common method of identifying high and low precipitation levels, is a type of time series method (Figure 2). It refers to different events before and after the same type of events that occur continuously, such as alternate occurrences of wetness and dryness, and flow changes of river runoff. If the annual precipitation data are regarded as a discrete sequence, the multiyear average precipitation $Q_0$ is selected as the standard quantity. When $Q_i - Q_0 > 0$, the variation is positive and belongs to a wet year; when $Q_i - Q_0 < 0$, the variation is negative and belongs to a dry year. When $Q_i - Q_0 > 0$ ($Q_i - Q_0 < 0$) occurs continuously, this indicates a continuous wet year (continuous dry year), which is called a positive run (negative run), and the continuous time is the length of run. In practical research, the wet years and the dry years are analyzed according to the classification criteria, and the frequency of continuous wet years and continuous dry years are counted. Additionally, the probability of continuous high and low precipitation in each meteorological station is calculated based on run analysis theory. This is of great significance for taking preventive measures in advance to reduce the losses caused by risk events.

Generally, the application of run analysis theory to identify drought events is roughly divided into four steps:

1. Deciding on the nature of water scarcity: Water scarcity research relies on analytical factors, such as runoff, precipitation, and soil moisture. Considering the definition of water scarcity and the difficulty of data acquisition, this study selected the precipitation data of 718 meteorological stations in China from 1961 to 2018.
2. Determining the time step: The time step can be hours, days, months, seasons, years, etc. Considering that a drought event is a natural phenomenon with slow development and long duration, months were used as the time step in this study.
3. Establishing the truncation level: The truncation level is mainly used to distinguish different events in historical data, which reflects the demand for incoming water. It can be a constant, a function that changes over time, or a certain percentage of runoff and precipitation; 50% of monthly average precipitation was selected as the truncation level in this study.

![Figure 1](http://data.cma.cn/) | Location of 718 meteorological stations over China (the province information is just to show the location in detail). This figure was created by using ArcGIS 10.2 for maps (URL: http://www.esri.com).
Calculating the water scarcity degree: If $S$ is the water scarcity degree, $M$ is the water scarcity severity (run-severity), and $D$ is the water scarcity duration (run-length), then $S = M \times D$. This formula describes the relationship among the three basic characteristics of drought events: the greater the absolute value of $S$ (negative value), the more severe the drought.

2.3. Copula method

2.3.1. Nested Archimedean-copula function

The Archimedean-copula function is a class of copula functions with simple form and good combination. Its stable properties and simple construction have led to its extensive use (Salvadori & De Michele 2004; Serinaldi & Grimaldi 2007). The high-dimensional Archimedean-copula function has a variety of constructors, of which the nested method is widely used (Müller & Scarsini 2005; Serinaldi & Grimaldi 2007). For three random variables of the three dimensions $X, Y,$ and $Z$, the joint distribution can be expressed as $C(x, y, z) = C_2[C_1(x, y), z]$, where $C_1$ is the joint distribution copula function of $X$ and $Y$, and $C_2$ is the joint distribution copula function of $C_1$ and $Z$, that is, $C(x, y, z)$ is the three-dimensional joint distribution function of $X, Y,$ and $Z$.

In summary, the joint distribution function of drought events constructed by the Archimedean-copula nesting method is as follows:

$$C = C(Z, C_1)$$

where $Z$ is the occurrence time, $C_1 = C_1(X, Y)$, $X$ is latitude, and $Y$ is longitude. Therefore, the drought events can be expressed as:

$$C = C(X, Y, Z) = [Z, C_1(X, Y)]$$

According to the associativity of the Archimedean-copula function (Salvadori & De Michele 2007), the three-dimensional copula functions $C(X, Y, Z)$ and $C_2[Z, C_1(X, Y)]$ are equivalent. The commonly used typical Archimedean-copula functions are the Clayton copula, Frank copula, and Gumbel copula.

2.3.2. Selection of Archimedean-copula function

Copula parameter estimation generally uses the maximum likelihood estimate method. Let $F_x, F_y$ be differentiable and the copula function $C(u, v)$ be second-order differentiable, where $u = F_x(x), v = F_y(y)$. The two-dimensional copula density
function can be expressed as:
\[ f_{xy}(x, y) = f_x(x)f_y(y)C(u, v; \theta) \] (3)
and then
\[ C(u, v; \theta) = \frac{f_{xy}(x, y)}{f_x(x)f_y(y)} \] (4)

Finally, the maximum likelihood estimate for parameter \( \theta \) is
\[ L(\theta) = \arg\max \ln C(u, v; \theta) \] (5)

This study selected the best-fitting copula on the basis of the goodness of fit. To compare the goodness of fit of the five copula functions, the average error (AE), root mean square error (RMSE), Akaike information criterion (AIC), and Bayesian information criterion (BIC) were obtained.

2.4. Gibbs sampling
Gibbs sampling is a series of samplings generated by the joint distribution of two or more random variables, whose greatest advantage is that only the conditional distribution of a single variable is considered. Then, the dimensionality reduction of high-dimensional problems is solved. Compared with other sampling methods, Gibbs sampling has less autocorrelation and better convergence, and is suitable for samples with data loss. The main idea of Gibbs sampling is to set the edge distribution \( f_x, f_y \) of two-dimensional variables \( (x, y) \), examine the conditional distribution \( f(x|y), f(y|x) \), and simplify the calculation with a joint density function \( f(x, y) \). Generally, the application of Gibbs sampling is roughly divided into four steps: (1) initialize \( y_0 \) and calculate \( x_0 \) by the \( f(x|y = y_0) \); (2) update \( y_1 \) by \( f(x = x_0|y) \); (3) repeat \( k \) times to obtain a Gibbs sequence of length \( k \); (4) if the Gibbs sequence converges to a stationary distribution, then each value in the sequence is independent of the initial value. Then, this stationary distribution is the target distribution that needs to be simulated.

2.5. Technology roadmap
In summary, to determine the occurrence probability of drought events, a model-based simulation framework is proposed (Figure 3). This framework mainly includes drought identification based on run analysis, joint probability distribution of drought events based on the copula method, and sampling and simulation based on Gibbs sampling.

3. RESULTS

3.1. Drought event identification
Based on the precipitation data of 718 meteorological stations in China from 1961 to 2018, the distribution of drought events in latitude and longitude was identified by run analysis theory. According to the statistical results (Figure 4), the occurrence of drought events was mainly concentrated from 35° to 42° north latitude (frequency/class width >0.04) and 105° to 120° east longitude (frequency/class width >0.025). Additionally, the occurrence time of drought events was mainly concentrated from September to November (frequency/class width >0.15).

3.2. Three-dimensional nested Archimedean-copula function construction
To simulate the spatiotemporal distribution of drought events, a nested Archimedean-copula function of drought events was obtained as follows: (1) estimation of edge probability distribution; (2) Archimedean-copula function construction of location; (3) nested Archimedean-copula function construction of location and time.

Through the nonparametric estimation of the probability distribution of latitude, longitude and time, the cumulative probability distribution was procured (Figure 5). Moreover, the K-S method was used to test the rationality. The \( p \)-values of the K-S test are 0.2334 (latitude \( X \)), 0.3502 (longitude \( Y \)) and 0.1254 (time \( Z \)), respectively, which are all greater than 0.05. This shows that the three empirical distribution functions of latitude, longitude and time can be used to estimate the edge distribution of the original variable under a significance level of 95%.
Furthermore, the joint probability distribution of location $C_1$(latitude $X$, longitude $Y$) was constructed based on the Archimedean-copula function, which was fitted by the three Clayton, Frank and Gumbel functions. Finally, the Gumbel copula has the minimum AE, RMSE, AIC, and BIC, indicating that it had the best goodness of fit (Table 1), and the Gumbel copula with the minimum sum of squared deviation was selected as the joint function (Figure 6).
Additionally, the spatiotemporal joint probability distribution of $C_2(t, Z, \text{location})$ was constructed. Finally, the Gumbel copula has the minimum AE, RMSE, AIC, and BIC, indicating that it had the best goodness of fit (Table 2), and the Gumbel copula with the minimum sum of squared deviation was selected as the joint function (Figure 7).

**Table 1** | The parameter estimation and goodness of fit based on the Archimedean-copula function of location

| Functions       | Clayton | Frank | Gumbel |
|-----------------|---------|-------|--------|
| Parameter estimation $\theta$ | – 0.1928 | 0.4904 | 1.1258 |
| AE ($10^{-2}$) | 2.1509  | 2.1241 | 2.1035 |
| RMSE ($10^{-1}$) | 1.4962  | 1.4824 | 1.4231 |
| AIC             | – 1.0154 | – 1.0175 | – 1.0187 |
| BIC             | – 1.0142 | – 1.0021 | – 1.0154 |

**Table 2** | The parameter estimation and goodness of fit based on the Archimedean-copula function of location

| Functions       | Clayton | Frank | Gumbel |
|-----------------|---------|-------|--------|
| Parameter estimation $\theta$ | – 0.0774 | 0.4664 | 1.0002 |
| AE ($10^{-2}$) | 2.3268  | 2.4502 | 1.9365 |
| RMSE ($10^{-1}$) | 1.5674  | 1.3268 | 1.2982 |
| AIC             | – 1.3254 | – 1.3689 | – 1.3761 |
| BIC             | – 1.3268 | – 1.2645 | – 1.3892 |
Through the Monte Carlo GoF test, the test statistics $p (0.1883)$ were obtained. The actual joint empirical distribution and the established nested two-dimensional Archimedean-copula joint distribution had a very high degree of coincidence (RMSE: 0.01) (Figure 8). It can be concluded that the joint probability distribution of time and location by using the nested Archimedean copula is scientific and reasonable, and the Archimedean-copula function can accurately determine the spatiotemporal joint probability distribution of drought events.

### 3.3. Gibbs sampling based on nested Archimedean-copula function

Based on the Gumbel copula model constructed in 3.2, Gibbs sampling of the same sample size was performed in the observed samples. The longitude and latitude of future drought events were simulated (Figure 9).

Furthermore, based on the Gibbs sampling simulation results, the average longitude and latitude of the drought events in each month were calculated according to the different occurrence times of each drought event in the simulation results. Comparing the average latitude and longitude of the simulated sample and actual sample with the monthly trend (Figure 10), it is found that there is a great difference between the actual curve and the predicted curve, indicating that there is still a large gap between the fitted models and the real model (RMSE: 3.26 (latitude), 2.48 (longitude)). Therefore, the nested Archimedean-copula function established above still has large room for improvement. Finally, to improve the accuracy of fitting and
prediction, this study tried to establish a time-varying nested Archimedean-copula function model according to the different categories of drought events.

3.4. Time-varying nested Archimedean-copula function construction

According to the trend of the average longitude and latitude of drought events with the change of months (solid line in Figure 10), drought events can be divided into three categories:

1. Category A: In February, March, April and May (spring), the latitude of the drought events decreased, while the longitude decreased first and then increased.
2. Category B: In June, July, August and September (summer), the latitude and longitude of the drought events showed a trend of turbulence.
3. Category C: In October, November, December and January (autumn and winter), the latitude and longitude of drought events increased.

Figure 9 | Sample distribution scatters of Gibbs sampling. This figure was created by using MATLAB 2018b (URL: http://www.mathworks.com).

Figure 10 | Comparison of simulated and actual latitude and longitude of drought events based on the nested Archimedean-copula function.
The Archimedean-copula models are constructed for the three categories A, B and C, and the optimal functions can be found in the following (Table 3):

1. The optimal joint distribution function of latitude and longitude of category A is a Clayton copula, and the optimal joint distribution function of time and location is a Gumbel copula.
2. The optimal joint distribution function of latitude and longitude of category B is a Frank copula, and the optimal joint distribution function of time and location is a Gumbel copula.
3. The optimal joint distribution function of latitude and longitude of category C is a Gumbel copula, and the optimal joint distribution function of time and location is a Gumbel copula.

In addition, as shown in Table 4, the three categories of time-varying nested Archimedean copula all passed the Monte Carlo GoF test. Therefore, different time periods only affect the joint distribution function of latitude and longitude of drought events ($C_1$) but have no effect on the joint distribution function of time and location of drought events ($C_2$); that is, the joint distribution function of time and location of drought events ($C_2$) is fixed at different times of the year. Moreover, the parameters $\theta$ of the joint distribution function $C_2$ have not changed, suggesting that drought will occur at the same time for a certain region each year. Compared with the non-time-varying nested Archimedean-copula model, the fitting effect of the joint distribution of drought events obtained by the three categories of A, B and C has been greatly improved.

### Table 3 | The parameter estimation and goodness of fit based on the Archimedean-copula function of location

| Joint distribution function | Parameter estimation $\theta$ | Clayton | Frank | Gumbel |
|----------------------------|-------------------------------|---------|-------|--------|
| Category A Latitude and longitude ($C_1$) | AE ($10^{-2}$) | $1.45 \times 10^{-6}$ | 0.36727 | 1.0974351 |
| | RMSE ($10^{-1}$) | 2.3164 | 2.4562 | 2.8934 |
| | AIC | -1.5627 | -1.3694 | -1.2563 |
| | BIC | -1.3649 | -1.2694 | -1.3182 |
| Time and location ($C_2$) | AE ($10^{-2}$) | $1.45 \times 10^{-6}$ | -2.00139 | 1.0000014 |
| | RMSE ($10^{-1}$) | 2.2364 | 2.2694 | 1.6497 |
| | AIC | -1.55987 | -1.3529 | -1.5764 |
| | BIC | -1.2643 | -1.5329 | -1.5764 |
| Category B Latitude and longitude ($C_1$) | AE ($10^{-2}$) | $1.45 \times 10^{-6}$ | -0.56059 | 1.0754822 |
| | RMSE ($10^{-1}$) | 2.6435 | 2.3294 | 2.3497 |
| | AIC | -1.3649 | -1.4697 | -1.3964 |
| | BIC | -1.2346 | -1.3092 | -1.2396 |
| Time and location ($C_2$) | AE ($10^{-2}$) | $1.45 \times 10^{-6}$ | -1.55987 | 1.0000014 |
| | RMSE ($10^{-1}$) | 2.5673 | 2.5634 | 1.2534 |
| | AIC | -1.8637 | -1.2694 | -1.9346 |
| | BIC | -1.2039 | -1.1647 | -1.5237 |
| Category C Latitude and longitude ($C_1$) | AE ($10^{-2}$) | $1.45 \times 10^{-6}$ | 1.32744 | 1.1904942 |
| | RMSE ($10^{-1}$) | 1.9643 | 1.8643 | 1.5697 |
| | AIC | -1.3465 | -1.8643 | -2.0349 |
| | BIC | -1.6349 | -1.2349 | -1.9403 |
| Time and location ($C_2$) | AE ($10^{-2}$) | $1.45 \times 10^{-6}$ | -1.08826 | 1.0000014 |
| | RMSE ($10^{-1}$) | 2.2356 | 2.4967 | 1.8649 |
| | AIC | -2.0619 | -2.9631 | 1.9873 |
| | BIC | -1.2346 | -1.3790 | -1.5037 |

3.5. Gibbs sampling based on the time-varying nested Archimedean-copula function

Based on the principle of Gibbs sampling and the Archimedean-copula function of three time periods, this study simulated the spatiotemporal distribution of future drought events according to the occurrence times of the drought events from...
February to May (spring), from June to September (summer), and from October to January (autumn and winter). Finally, the simulated sample distribution scatters of the three time periods were obtained with the Gibbs sampling simulation results (Figure 11).

Comparing the average latitude and longitude of the simulated sample and actual sample with the monthly trend (Figure 12), it is found that the coincidence degree is high (RMSE: 0.52 (latitude), 0.75 (longitude)). This also proved the reliability of the established nested Archimedean-copula function as a predictive model. Additionally, it is verified that the time-varying nested Archimedean-copula function model constructed by observed samples at different time periods is much better than the whole nested Archimedean-copula function model. That is, it is reasonable and necessary to carry out classification modeling analysis according to the different time periods of drought events.

4. DISCUSSION

Since the first publication on hydrological problems caused by drought based on the copula method, numerous studies have confirmed that the copula model has a unique advantage both in China and abroad on multivariate distributions of drought characteristics, such as the intensity, duration, frequency, and recurrence. China is a vast area with different climatic regions, and it is not surprising that the spatiotemporal patterns of drought events can have regional characteristics.

4.1. The relationship between drought and precipitation

Generally, drought refers to a long-term meteorological condition of minimal rain or absence of rain, which makes soil water insufficient and induces crop water imbalance. The definition of drought usually depends on the selected influencing factors; thus, droughts can be divided into four categories: meteorological drought (Duan et al. 2014), hydrological drought (Dracup et al. 1980), agricultural drought, and socioeconomic drought (Figure 13) (Mishra & Singh 2010, 2011). Different types of drought have different research priorities. Meteorological drought refers to a water scarcity caused by the imbalance of evaporation and precipitation, and water expenditure is greater than water income. Agricultural drought refers to the water deficit caused by the continuous lack of soil moisture during the growing period of crops. Hydrological drought refers to the imbalance of surface water or groundwater in a certain period due to the long-term shortage of precipitation, which reduces river flows, lake water levels, reservoir water storage and so on. Socioeconomic drought is an abnormal water deficit caused by the imbalance between water supply and water demand in the natural system and the human socioeconomic system. For these four types of droughts, meteorological drought is a natural phenomenon characterized by the reduction of precipitation, while

| Time-varying nested Archimedean copula | p     |
|----------------------------------------|-------|
| From February to May (spring)          | 0.2479|
| From June to September (summer)        | 0.2175|
| From October to January (autumn and winter) | 0.3721|

Table 4 | Monte Carlo GoF test results of time-varying nested Archimedean-copula function

Figure 11 | Sample distribution scatters of three Gibbs samplings. This figure is created by using MATLAB 2018b (URL: http://www.mathworks.com).
agricultural drought, hydrological drought and socioeconomic drought are more concerned with human and social aspects (Adarsh et al. 2019). In general, meteorological drought is the basis for the three other types of drought (Figure 13). Therefore, this study mainly uses precipitation to study meteorological drought. This is consistent with Fan et al. (2017), who noted the similarity coefficient between drought probability and low precipitation probability in China.
4.2. Identification of drought events

Run analysis theory can directly start from the simple statistical run phenomenon and reveal the probability of the occurrence of the run phenomenon, which is an effective method to study drought. Therefore, the application of run analysis theory in the field of drought has become increasingly popular in recent years, and many scholars have used the theory to analyze drought events in the northern regions of south Africa (Herbst et al. 1966), Beijing (Fan et al. 2017) and the Zhangweinan Canal Basin (Bai et al. 2018). Although the direct analysis of precipitation data has been presented as research has progressed, the Standardized Precipitation Index (SPI), Palmer Drought Severity Index (PDSI), and Standardized Precipitation Evaporation Index (SPEI) are the most commonly used and effective drought indices (Figure 13) (Heim 2002; Bothe et al. 2010; Hernandez & Uddameri 2014). The SPI is an index used to measure the degree of drought and flooding, and it is mainly based on the distribution of precipitation, which is obtained by calculating the cumulative frequency of precipitation after normalization (Nikravesh et al. 2020). The PDSI is based on a simple two-layer soil moisture budget model that considers the current soil moisture supply (precipitation) and demand (evaporation). The SPEI is mainly used to detect drought distribution based on the SPI and can reflect the duration and accumulation of drought. The SPI calculation is convenient because only precipitation is considered, and the PDSI is usually used because of its wide perspective, whereas the SPEI emphasizes the effect of warming on drought (Hayes et al. 1999; Dai et al. 2004; Vicente-Serrano et al. 2010). In general, the precipitation index of the SPI, PDSI and SPEI is used to identify drought events based on run analysis theory and is also a possible research direction at present. Moreover, there are still some shortcomings in the study of drought events. For the law of drought change, many scholars have extracted the variables of drought duration and drought severity. They have paid less attention to the study of drought areas and determined the probability of extreme drought events. This study proposed a simulation framework that can analyze the joint probability distribution of drought events. However, the study of drought areas has not been considered.

4.3. Thoughts on the study of drought events in China

There are two main geographical features in China (Fan et al. 2016). First, the land area is vast, and the topography and physiognomy are complex. The east and south of China are close to the ocean with more precipitation, while the north and west of China are deep inside the continent and drier. Second, the latitude and longitude spans are large and include the tropics, subtropics, warm temperate zones, middle temperate zones, cold temperate zones and the Qinghai-Tibet Plateau. The difference in thermal properties between land and sea causes the precipitation to decrease from the southeast coast to the northwest inlands (Wang et al. 2019). In general, precipitation and temperature are the key factors affecting regional drought events and the features in China’s regional system show an uneven distribution. To explore the influences of precipitation factors on drought events over China and offer suggestions for drought prediction, this study used precipitation data based on run analysis theory, the copula method and Gibbs sampling to analyze the three-dimensional spatiotemporal probability distribution of drought events in China. The significant regional differences in China’s natural conditions determine the spatial characteristics of Chinese meteorological factors. The study area can be divided into seven regions in accordance with the properties of geography, land use type, climate and farming system: northeast China (NE), northwest China (NW), central China (CC), northern China (NC), southern China (SC), eastern China (EC) and southwest China (SW) (Figure 13) (Song et al. 2011; Wang et al. 2019). The optimal joint distribution Archimedean-copula function for each of the seven regions may not be consistent with China as a whole, but the overall accuracy is bound to increase. Separate research on China’s seven regions will be our future research direction, which will fully consider geographical characteristics, while offering more effective guiding significance for the prevention of local drought events.

4.4. Other perspectives of drought research

Solar radiation is an important source of energy for atmospheric motion and a direct driver of surface temperature, precipitation and heat flux. The radiation budget at the top of the atmosphere maintains an approximate balance, while the radiation transmission process at the earth’s surface is very complicated (Bisht et al. 2005). Moreover, the annual average radiation income from satellite observations indicates that nearly 50% of solar radiation energy is absorbed by the earth’s surface. This part of the energy exchanges with the surface and the atmosphere through sensible heat flux, latent heat flux and soil heat flux, which affects the surface energy balance and land surface process changes and provides power for near-surface atmospheric movement (Yang et al. 2008). At the same time, the balance of the surface radiation budget is also closely related to regional evapotranspiration and hydrological processes (Mecikalski et al. 2018). In general, solar radiation has great
significance for the study of drought. This study summarizes the factors affecting solar radiation based on a literature review and data query. The influencing factors are divided into two categories. One is the periodic factors, such as season, geography and meteorology, and the other is nonperiodic factors, such as natural disasters (Figure 13).

In detail, the solar incident angle, the duration of illumination and the intensity of sunshine are different in each season, which is the main manifestation of the impact of seasonal change on drought. Therefore, the solar elevation angle and the duration of sunshine can be considered seasonal factors in the study. The influence of geographical factors on solar radiation should not be underestimated; thus, latitude, longitude and altitude can be considered in the study. The latitude, longitude and season are mainly considered in this paper. Meteorological conditions mainly affect the scattering and absorption of solar radiation, and weather conditions and cloud conditions can be considered in the study. Although natural disasters (such as rain and snow disasters, wind anomalies, sand dust and typhoons) can also be meteorological factors to some extent, they have a strong impact on drought events. Finally, other perspectives of drought research are shown in Figure 13.

4.5. Uncertainties

This study presents a simulation framework that can construct a joint probability distribution of drought events. However, it should be noted that some methods related to the framework are not perfect; thus, uncertainties are potentially introduced into our results. Although drought event identification based on run analysis theory is the foundation of this study, the recognition process has deficiencies such as lag and rigidity. Generally, the accumulation and recession effects of drought are regarded as a physical process. The process description can be divided into four stages of accumulation stage, outbreak stage, disposal stage and regression stage, which can realize the dynamic identification of drought events and eliminate uncertainties. Moreover, the joint probability distribution of drought events based on the Archimedean copula is the focus of this study. Despite this, to improve the rationality and interpretability of drought events, a variety of methods (e.g., the meta-elliptic copula and Plackett copula) should be used to explore the joint probability distribution. In addition, this study only determined the trend of average latitude and longitude of drought events in China as a whole, which is not sufficient. Thus, different regional assessments of the joint probability distribution should be conducted with the geographical features of China for applicability.

5. CONCLUSIONS

Although drought is the main problem in the field of water scarcity, studies of the spatiotemporal probability distribution of drought events are relatively rare. In this study, a simulation framework of the occurrence probability of drought events is proposed to effectively compensate for the high-dimensional problems and lack of initial data in traditional methods. The core of the simulation framework is run analysis theory, a time-varying nested Archimedean-copula function model, and Gibbs sampling. The run analysis theory mainly identifies the drought events, and the copula function determines the spatiotemporal probability distribution of drought events, while the Gibbs model performs sampling and simulation. And this study demonstrated the application of the proposed simulation framework in China based on precipitation data of 718 meteorological stations from 1961 to 2018.

The occurrence location of drought events is mainly concentrated from 35° to 42° north latitude and 105° to 120° east longitude, while the occurrence time is mainly concentrated from September to November. The three-dimensional nested copula function can accurately determine the spatiotemporal joint probability distribution of drought events (RMSE: 0.01). Gibbs sampling and simulation of the time-varying nested Archimedean-copula function (RMSE: 0.52 (latitude), 0.75 (longitude)) is more accurate than the nested Archimedean-copula function (RMSE: 3.26 (latitude), 2.48 (longitude)). Compared with the nested Archimedean-copula function, the accuracy of Gibbs sampling and simulation based on the time-varying nested Archimedean-copula function was increased by 84.05% latitude and 69.76% longitude. And the drought will occur at the same time every year. The proposed simulation framework established the spatiotemporal joint probability distribution of drought events, which provided an effective means for predicting drought events scientifically. In the future, it should pay more attention to supplementary data and multidisciplinary methods to achieve more accurate predictions of drought events.

AUTHOR CONTRIBUTIONS

The first author contributed substantially to conceptualization, methodology, validation, data curation, data interpretation, and writing. All authors participate in drafting the article or revising it critically, and give final approval of the version to be submitted.
FUNDING
This research was supported by the National Key Research and Development Program of China (2019YFC0408902), the National Natural Science Foundation of China (Grant No. 51879010 and 51479003), and the Graduate Innovation Fund in Beijing Key Laboratory of Urban Hydrological Cycle and Sponge City Technology (HYD2020IFDC03).

CONFLICTS OF INTEREST
The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT
Data cannot be made publicly available; readers should contact the corresponding author for details.

REFERENCES
Adarsh, S., Kumar, D. N., Deepthi, B., Gayathri, G., Aswathy, S. S. & Bhagyasree, S. 2019 Multifractal characterization of meteorological drought in India using detrended fluctuation analysis. International Journal of Climatology 39 (11), 4234–4255.

Araghi, A., Mousavi-Baygi, M., Adamowski, J., Martinez, C. & van der Ploeg, M. 2017 Forecasting soil temperature based on surface air temperature using a wavelet artificial neural network. Meteorological Applications 24 (4), 603–611.

Bai, Q., Song, Z., Wang, H., Deng, C. & Zhao, Y. 2018 Quantitative analysis of the impact of natural factors and human factors on hydrological system using the SWAT model: the Zhanweinan Canal Basin case. Journal of Natural Resources 33 (9), 1575–1587.

Bisht, G., Venturini, V., Islam, S. & Jiang, L. 2005 Estimation of the net radiation using MODIS (Moderate Resolution Imaging Spectroradiometer) data for clear sky days. Remote Sensing of Environment 97 (1), 52–67.

Bothe, O., Fraedrich, K. & Zhu, X. 2010 The large-scale circulations and summer drought and wetness on the Tibetan plateau. International Journal of Climatology 30 (6), 844–855.

Chang, T. J. 1991 Investigation of precipitation droughts by use of kriging method. Journal of Irrigation and Drainage Engineering 117 (6), 935–943.

Choi, W., Byun, H.-R., Cassardo, C. & Choi, J. 2018 Meteorological and streamflow droughts: characteristics, trends, and propagation in the Milwaukee River basin. Professional Geographer 70 (3), 463–475.

Chung, C. H. & Salas, J. D. 2000 Drought occurrence probabilities and risks of dependent hydrologic processes. Journal of Hydrologic Engineering 5 (3), 259–268.

Dai, A., Trenberth, K. E. & Qian, T. 2004 A global dataset of Palmer Drought Severity Index for 1870–2002: relationship with soil moisture and effects of surface warming. Journal of Hydrometeorology 5 (6), 1117–1130.

Dracup, J. A., Lee, K. S. & Paulson Jr., E. G. 1980 On the statistical characteristics of drought events. Water Resources Research 16 (2), 289–296.

Dupuis, D. J. 2007 Using copulas in hydrology: benefits, cautions, and issues. Journal of Hydrologic Engineering 12 (4), 381–393.

Duan, K., Xiao, W., Wei, Y. & Liu, D. 2014 Multi-scale analysis of meteorological drought risks based on a Bayesian interpolation approach in Huai River basin, China. Stochastic Environmental Research and Risk Assessment 28 (8), 1985–1998.

Fan, J., Wu, L., Zhang, F., Xiang, Y. & Zheng, J. 2016 Climate change effects on reference crop evapotranspiration across different climatic zones of China during 1956–2015. Journal of Hydrology 542, 923–937.

Fan, L., Wang, H., Wang, C., Lai, W. & Zhao, Y. 2017 Exploration of use of copulas in analysing the relationship between precipitation and meteorological drought in Beijing, China. Advances in Meteorology 2017, 4650284.

Ganguli, P. & Reddy, M. J. 2012 Risk assessment of droughts in Gujarat using bivariate copulas. Water Resources Management 26 (11), 3301–3327.

Ganguli, P. & Reddy, M. J. 2014 Ensemble prediction of regional droughts using climate inputs and the SVM–copula approach. Hydrological Processes 28 (19), 4989–5009.

Gao, Z., Wang, Q., Cao, X. & Gao, W. 2014 The responses of vegetation water content (EWT) and assessment of drought monitoring along a coastal region using remote sensing. GIScience & Remote Sensing 51 (1), 1–16.

Genest, C. & Favre, A.-C. 2007 Everything you always wanted to know about copula modeling but were afraid to ask. Journal of Hydrologic Engineering 12 (4), 347–368.

Genest, C., Favre, A.-C., Béliveau, J. & Jacques, C. 2007 Metaelliptical copulas and their use in frequency analysis of multivariate hydrological data. Water Resources Research 43 (9), W09401.

Hayes, M. J., Svoboda, M. D., Willhite, D. A. & Vanyarkho, O. V. 1999 Monitoring the 1996 drought using the Standardized Precipitation Index. Bulletin of the American Meteorological Society 80 (3), 429–438.

Heim, R. R. 2002 A review of twentieth–century drought indices used in the United States. Bulletin of the American Meteorological Society 83 (8), 1149–1166.

Herbst, P. H., Bredenkamp, D. B. & Barker, H. M. G. 1966 A technique for the evaluation of drought from rainfall data. Journal of Hydrology 4, 264–272.
Hernandez, E. A. & Uddameri, V. 2014 Standardized precipitation evaporation index (SPEI)-based drought assessment in semi-arid south Texas. *Environmental Earth Sciences* 71 (6), 2491–2501.

Hoekstra, A. Y. 2014 Water scarcity challenges to business. *Nature Climate Change* 4 (5), 318–320.

Lai, W. L., Wang, H. R. & Zhang, J. 2018 Comprehensive assessment of drought from 1960 to 2013 in China based on different perspectives. *Theoretical and Applied Climatology* 134 (1–2), 585–594.

Li, T., Wang, S., Zhuang, W. & Liu, T. 2016 Application of the theory of run and copula function to the joint distribution of two-dimension drought variables. *Journal of Arid Land Resources and Environment* 30 (6), 77–82.

Li, Q., He, P., He, Y., Han, X., Zeng, T., Lu, G. & Wang, H. 2020 Investigation to the relation between meteorological drought and hydrological drought in the upper Shaying River Basin using wavelet analysis. *Atmospheric Research* 234, 104743.

Liu, W. T. & Kogan, F. N. 1996 Monitoring regional drought using the Vegetation Condition Index. *International Journal of Remote Sensing* 17 (14), 2761–2782.

Lohani, V. K. & Loganathan, G. V. 1997 An early warning system for drought management using the Palmer Drought Index. *JAWRA Journal of the American Water Resources Association* 33 (6), 1375–1386.

Ma, M. W., Song, S. B., Ren, L. L., Jiang, S. H. & Song, J. L. 2013 Multivariate drought characteristics using trivariate Gaussian and Student t copulas. *Hydrological Processes* 27 (8), 1175–1190.

Medikalı, J. R., Shoemaker, W. B., Wu, Q., Holmes, M. A., Paech, S. J. & Sumner, D. M. 2018 High-resolution GOES insolation–evapotranspiration index data for water resource management in Florida: 1995–2015. *Journal of Irrigation and Drainage Engineering* 144 (9), 04018025.

Mirabbasi, R., Fakheri-Fard, A. & Dinapashoh, Y. 2012 Bivariate drought frequency analysis using the copula method. *Theoretical and Applied Climatology* 108 (1–2), 191–206.

Mishra, A. K. & Desai, V. R. 2005 Drought forecasting using stochastic models. *Stochastic Environmental Research and Risk Assessment* 19 (5), 326–339.

Mishra, A. K. & Singh, V. P. 2010 A review of drought concepts. *Journal of Hydrology* 391 (1–2), 202–216.

Mishra, A. K. & Singh, V. P. 2011 Drought modeling – a review. *Journal of Hydrology* 403 (1–2), 157–175.

Müller, A. & Scarsini, M. 2005 Archimedean copulae and positive dependence. *Journal of Multivariate Analysis* 95 (2), 434–445.

Nikravesh, G., Aghababaei, M., Nazari-Sharabian, M. & Karakouzian, M. 2020 Drought frequency analysis based on the development of a two-variate standardized index (rainfall–runoff). *Water* 12 (9), 2599.

Qian, L., Wang, H. & Zhang, K. 2014 Evaluation criteria and model for risk between water supply and water demand and its application in Beijing. *Water Resources Management* 28 (13), 4433–4447.

Qian, L. X., Wang, H. R., Dang, S. Z., Wang, C., Jiao, Z. Q. & Zhao, Y. 2018 Modelling bivariate extreme precipitation distribution for data-scarce regions using Gumbel–Hougaard copula with maximum entropy estimation. *Hydrological Processes* 32 (2), 212–227.

Reddy, M. J. & Singh, V. P. 2014 Multivariate modeling of droughts using copulas and meta-heuristic methods. *Stochastic Environmental Research and Risk Assessment* 28 (3), 475–489.

Salvadori, G. & De Michele, C. 2004 Frequency analysis via copulas: theoretical aspects and applications to hydrological events. *Water Resources Research* 40 (12), W12511.

Salvadori, G. & De Michele, C. 2007 On the use of copulas in hydrology: theory and practice. *Journal of Hydrologic Engineering* 12 (4), 369–380.

Serinaldi, F. & Grimaldi, S. 2007 Fully nested 3-copula: procedure and application on hydrological data. *Journal of Hydrologic Engineering* 12 (4), 420–430.

Shahid, M., Cong, Z. & Zhang, D. 2018 Understanding the impacts of climate change and human activities on streamflow: a case study of the Soan River basin, Pakistan. *Theoretical and Applied Climatology* 134 (1–2), 205–219.

Shahid, M. & Rahman, K. U. 2021 Identifying the annual and seasonal trends of hydrological and climatic variables in the Indus Basin Pakistan. *Asia-Pacific Journal of Atmosphere Sciences* 57 (2), 191–205.

Shiau, J. T. 2006 Fitting drought duration and severity with two-dimensional copulas. *Water Resources Management* 20 (5), 795–815.

Song, Y., Achberger, C. & Lindermholm, H. W. 2011 Rain-season trends in precipitation and their effect in different climate regions of China during 1961–2008. *Environmental Research Letters* 6 (3), 034025.

Sun, S., Fu, G. T., Bao, C. & Fang, C. L. 2019 Identifying hydro-climatic and socioeconomic forces of water scarcity through structural decomposition analysis: a case study of Beijing city. *Science of the Total Environment* 687, 590–600.

Vicente-Serrano, S. M., Beguería, S. & López-Moreno, J. I. 2010 A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *Journal of Climate* 23 (7), 1696–1718.

Wan, L., Zhou, J. X., Guo, H. Y., Cui, M. & Liu, Y. G. 2016 Trend of water resource amount, drought frequency, and agricultural exposure to water stresses in the karst regions of South China. *Natural Hazards* 80 (1), 23–42.

Wang, H., Fan, L., Liang, Y. & Wang, C. 2018 An integrated approach for water scarcity evaluation – a case study of Yunnan, China. *Environment Development and Sustainability* 20 (1), 109–127.

Wang, Z., Ye, A., Wang, L., Liu, K. & Cheng, L. 2019 Spatial and temporal characteristics of reference evapotranspiration and its climatic driving factors over China from 1979–2015. *Agricultural Water Management* 213, 1096–1108.

Wong, G., van Lanen, H. A. J. & Torfs, P. J. F. 2013 Probabilistic analysis of hydrological drought characteristics using meteorological drought. *Hydrological Sciences Journal* 58 (2), 253–270.
Yang, K., Pinker, R. T., Ma, Y., Koike, T., Wonsick, M. M., Cox, S. J., Zhang, Y. & Stackhouse, P. 2008 Evaluation of satellite estimates of downward shortwave radiation over the Tibetan Plateau. *Journal of Geophysical Research: Atmospheres* **113**, D17204.

Yang, S., Li, C., Lou, H., Wang, P., Wu, X., Zhang, Y., Zhang, J. & Li, X. 2021 Role of the countryside landscapes for sustaining biodiversity in karst areas at a semi centennial scale. *Ecological Indicators* **123**, 107315.

Yusof, F., Hui-Mean, F., Suhaila, J. & Yusof, Z. 2013 Characterisation of drought properties with bivariate copula analysis. *Water Resources Management* **27** (12), 4183–4207.

Zang, Z., Zou, X., Xi, X., Zhang, Y., Zheng, D. & Sun, C. 2016 Quantitative characterization and comprehensive evaluation of regional water resources using the Three Red Lines method. *Journal of Geographical Sciences* **26** (4), 397–414.

Zhang, Q. 2009 The South-to-North Water Transfer Project of China: environmental implications and monitoring strategy. *Journal of the American Water Resources Association* **45** (5), 1238–1247.

Zhao, Z., Wang, H., Wang, C., Li, W., Chen, H. & Deng, C. 2020a Changes in reference evapotranspiration over Northwest China from 1957 to 2018: variation characteristics, cause analysis and relationships with atmospheric circulation. *Agricultural Water Management* **231**, 105958.

Zhao, Z., Wang, H., Yu, C., Deng, C., Liu, C., Wu, Y., Yan, J. & Wang, C. 2020b Changes in spatiotemporal drought characteristics over northeast China from 1960 to 2018 based on the modified nested Copula model. *Science of the Total Environment* **739**, 140528.

Zinatloo-Ajabshir, S., Heidari-Asil, S. A. & Salavati-Niasari, M. 2021b Recyclable magnetic ZnCo2O4-based ceramic nanostructure materials fabricated by simple sonochemical route for effective sunlight-driven photocatalytic degradation of organic contaminants. *Ceramics International* **47** (7), 8959–8972.

Zinatloo-Ajabshir, S., Heidari-Asil, S. A. & Salavati-Niasari, M. 2021a Enhanced visible-light-driven photocatalytic performance for degradation of organic contaminants using PbWO4 nanostructure fabricated by a new, simple and green sonochemical approach. *Ultrasonics Sonochemistry* **72**, 105420.

Zuo, D. D., Hou, W., Yan, P. C. & Feng, T. C. 2014 Research on drought in southwest China based on the theory of run and two-dimensional joint distribution theory. *Acta Physica Sinica* **63** (23), 250204.