Journal of
Water & Climate Change

Journal of Water and Climate Change Vol 13 No 12, 4302 doi: 10.2166/wcc.2022.345

A new method for assessment of regional drought risk: information diffusion and interval mapping adjustment based on k-means cluster points

Wentao Xu [a], Ming Tang [a,b,*] and Yanlei Li [a]

[a] School of Hydraulic & Ecological Engineering, Nanchang Institute of Technology, Nanchang, Jiangxi 330099, China
[b] Jiangxi Key Laboratory of Hydrology-Water Resources and Water Environment, Nanchang, Jiangxi 330099, China
*Corresponding author. E-mail: frankt@nit.edu.cn

WX, 0000-0003-2544-2879

ABSTRACT

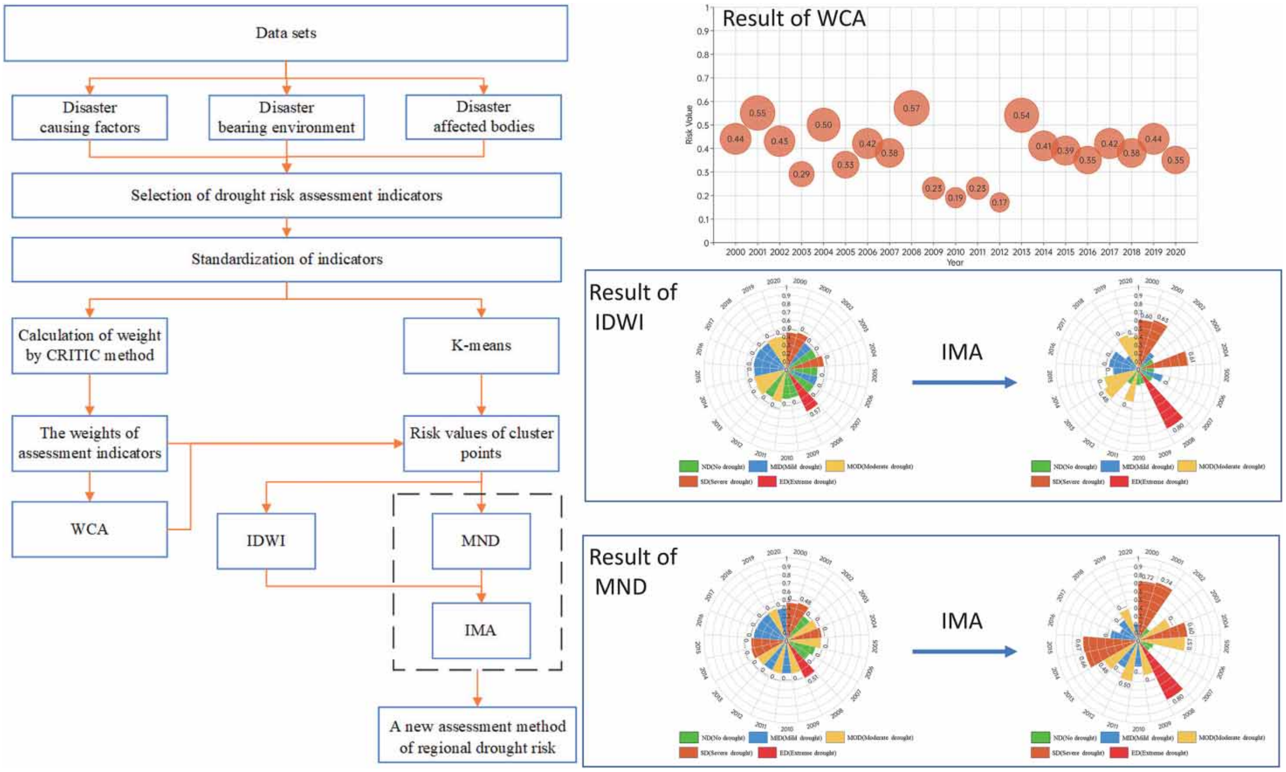
Aiming at the defect that there is no ability for the conventional weighted comprehensive assessment method (WCA) to grade drought risk directly, a method based on k-means cluster points to realize the classification of drought risk is proposed in this paper. On the basis of calculating the drought risk value of cluster points, the inverse distance weight interpolation method (IDWI) and multidimensional normal diffusion method (MND) were used to quantify the drought risk value, and the discrimination between the risk value and grade was improved by interval mapping adjustment (IMA). In this paper, the drought risk of Anhui Province from 2000 to 2020 was calculated to verify the above method. The results show that: (1) The drought risk quantification method based on information redistribution of k-means cluster point can not only realize automatic risk classification, but also re-quantify the risk value of the assessment object in the same risk grade, which makes up for the defects that the conventional WCA cannot carry out grade division and the conventional clustering method cannot assign the risk value of the assessment object. (2) The result of information redistribution based on MND is closer to the actual drought situation and more reasonable than IDWI. (3) The dispersion effect of risk value obtained by information redistribution based on k-means cluster point can be improved by the IMA of drought risk. It improves the discrimination degree of risk value, so that the grades can be displayed more intuitively. The defect of the WCA is overcome by the new method proposed in this paper, the follow-up utilization space is widened, and the thinking of risk quantification in drought risk assessment is broadened.

Key words: drought risk assessment, information diffusion, interval mapping adjustment, k-means clustering, cluster points

HIGHLIGHTS

- The drought risk was graded based on k-means.
- The drought risk is quantified by inverse distance weight interpolation method and multidimensional normal diffusion method.
- The discrimination of drought risk can be improved by mapping the adjustment on interval.

GRAPHICAL ABSTRACT



ABBREVIATIONS

CRITIC	Crtiteria Importance Through Intercriteria Correlation
ED	extreme drought
IPCC	Intergovernmental Panel on Climate Change
IDWI	inverse distance weight interpolation
IMA	interval mapping adjustment
MND	multidimensional normal diffusion
ND	no drought
MID	mild drought
MOD	moderate drought
SD	severe drought
WCA	weighted comprehensive assessment

1. INTRODUCTION

In the Sixth IPCC (Intergovernmental Panel on Climate Change) Assessment Report, it is pointed out that global warming has caused the increase of atmospheric evaporation demand and drought events (IPCC 2021). Drought events tend to have a serious impact on production, life and ecology due to its long latency and duration. The losses caused by drought can be effectively reduced and the ability to cope with drought can be enhanced by drought risk management. In recent years, drought risk management has been transformed from single drought resistance to comprehensive drought resistance (Xu *et al.* 2013), in which region-oriented assessment of drought risk is the core content of drought risk management, as well as the main means of regional sustainable development management and combating global climate change (Jin *et al.* 2016). At present, there are three assessment theoretical modes of regional drought (Jin *et al.* 2016): the first is the curve assessment method based on the analysis of historical drought loss frequency, the second is the curve assessment method based on the physical genetic process of drought loss risk and the third is the assessment of drought risk based on the components of drought risk loss (hereinafter referred to as weighted comprehensive assessment method (WCA)). The WCA can build assessment indicator system of drought risk according to the difference of different regional differentiation, and the weight calculation results can reflect the influence of various factors contributing to drought risk. The WCA helps to analyse the overall regional drought risk, which has been widely used with its low data requirements, strong operability and flexibility (Ai *et al.* 2020; Wang *et al.* 2021a).

At present, there are many studies on the assessment of regional drought risk using conventional WCA. Chou *et al.* (2019) used the analytic hierarchy process (AHP) to calculate the risk values of hazard, vulnerability and exposure of vulnerable eco-regions in China during 1988–2017, and predicted the change of drought risk in the future. Zhao *et al.* (2020) made a quantitative analysis of drought risk in China from three aspects: hazard, vulnerability and exposure, and found out the main influencing factors. Huang *et al.* (2015) constructed an integrated drought index combining meteorological, hydrological and agricultural factors based on the variable fuzzy set theory and calculated the spatial distribution characteristics of the annual average and seasonal average integrated drought index in the Yellow River Basin. Through the analysis of the above studies, the drought risk of the assessment objects in the region can be compared by the conventional WCA, i.e., the relative quantification of drought risk (Jin *et al.* 2016). However, it is difficult to carry out classification of risk grade directly. The cluster analysis had been introduced into the assessment of drought risk to classify research objects in some studies. Shim *et al.* (2021) used k-means clustering (clustering based on Euclidean distance between sample data) to divide South Korean island areas into three clusters for the assessment of drought vulnerability and pointed out the main vulnerability factors affecting each cluster. Moura *et al.* (2021) identify homogenous zones over Paraiba State in relation to the state, duration and severity of droughts that have occurred during 1998–2017 using hierarchical clustering based on rainfall data. The risk grades of the research areas according to the drought risk values were only classified but the risk values of the assessment objects were not assigned in the above two studies, and the results could not reflect the distinction between the drought risk values of different assessment objects in the same grade. According to the above defects, some researchers refined risk value. Tang *et al.* (2009) classified the grade on the basis of comprehensive drought risk value of each city in Anhui Province calculated by the conventional mutation evaluation method and mapped the comprehensive drought risk values to 10 uniform interval range 0–1 by interval mapping adjustment (IMA). Thus, the differentiation degree of risk value of each city can be increased, and the grade and size of evaluation value can be shown more intuitively. Qu *et al.* (2015) constructed the evaluation index system from the perspectives of hazard causing factors, hazard bearing bodies and hazard-pregnant environment, and calculated the agricultural drought risk degree of all provinces in China based on the WCA and AHP. After determining the grade division of a single index by using the normal distribution, the fuzzy comprehensive evaluation method was used to classify and assign the grade of hazard, vulnerability and exposure. Sahana *et al.* (2021) divided drought index ranging from 0 to 1 into five grades based on calculating the drought hazard index and the drought vulnerability index by multivariate standardized drought index and determined the number of each category by k-means clustering. To sum up, it can be found that there is the defect for the conventional WCA in the division of risk grade. At the same time, the combination of machine learning and drought research is one of the mainstream methods. Machine learning is a reliable choice for both assessing current drought risk and predicting future drought trends (Shamshirband *et al.* 2020; Band *et al.* 2022).

This paper aimed to solve the defect that conventional WCA cannot grade drought risk and referred to the research idea that k-means clustering was used to classify the drought area in other literature works, and k-means clustering was introduced to grade drought risk. In order to improve the discrimination degree of risk value of the assessment objects, the drought risks of the assessment objects were quantified by using IDWI and MND, respectively, on the basis of the calculation for the risk value of cluster points. Finally, the risk grade was shown more accurately through IMA, which provided a new idea for the assessment of regional drought risk.

2. DATA AND METHODS

2.1. Study area and data source

Anhui Province is taken as the research object in this study (Figure 1). Anhui Province is located between 114°54′–119 °37′ E and 29°41′–34 °38′ N. It straddled the Yangtze River and Huaihe River Basin with different topography and features; its terrain is high in the southwest. Its south is located in the subtropical monsoon climate zone, and its north is located in the temperate monsoon climate zone; therefore, it is hot and rainy in summer, but the distribution of rainfall is uneven. The rainfall varies greatly between seasons, as well as in space, which is easy to cause drought disasters (Tang 2008). Figure 2 shows the annual variation of the Standardized Precipitation Index (SPI) in Anhui Province, indicating that drought events occurred frequently in Anhui Province from 2000 to 2020.

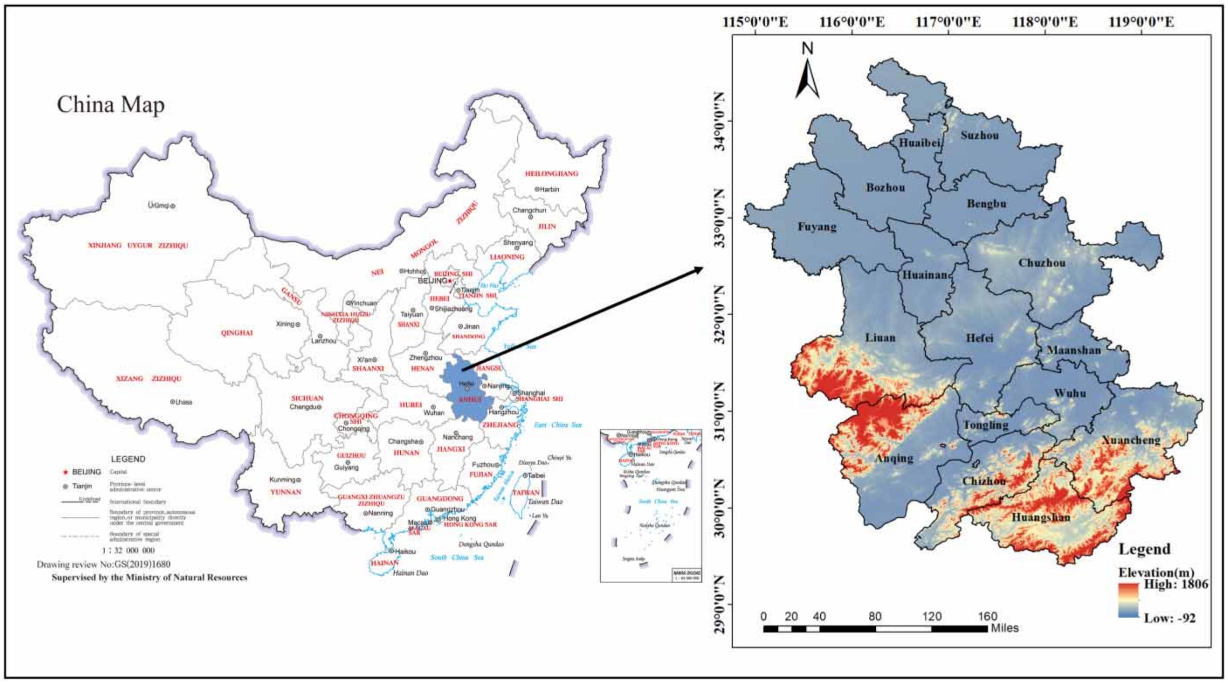


Figure 1 | A topographic map of Anhui Province.

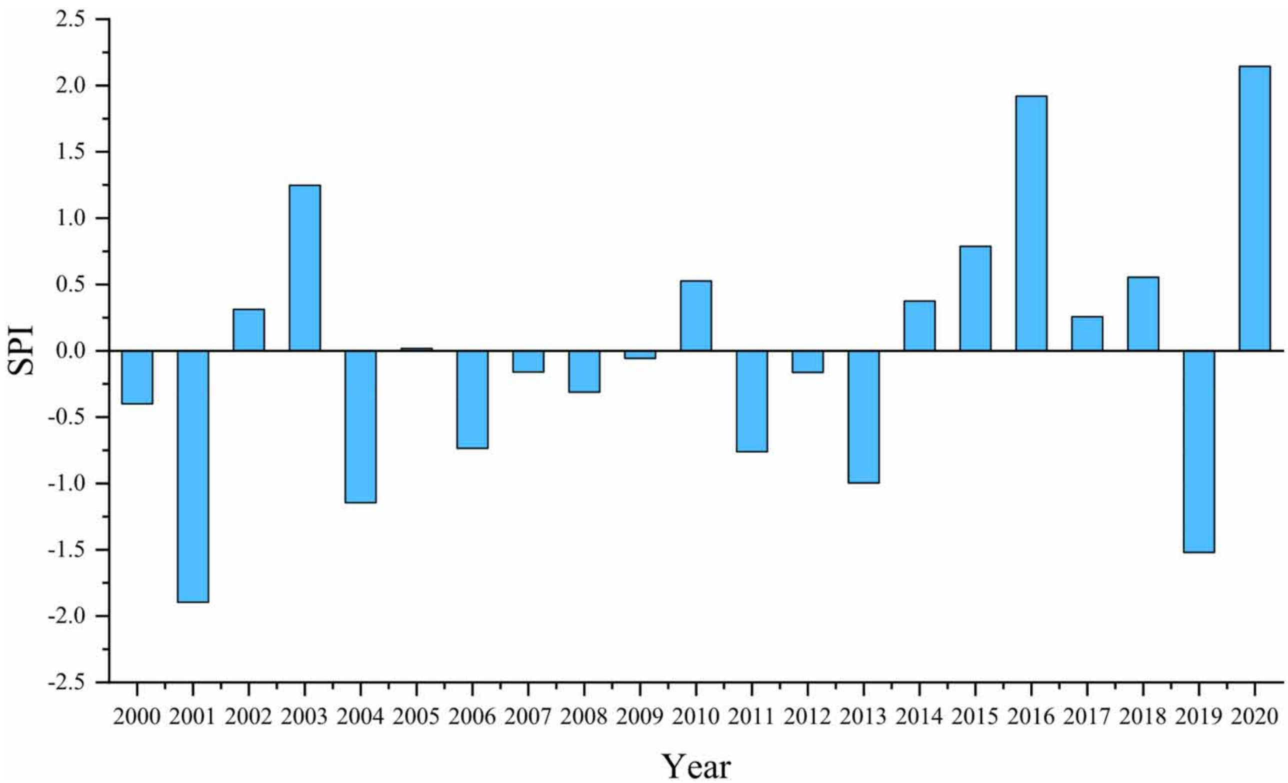


Figure 2 | The annual variation of SPI in Anhui Province from 2000 to 2020.

Regional drought system is a nonlinear dynamical system coupled by regional natural, economic and social factors, which is characterized by non-stationarity (Luo *et al.* 2019). In this study, assessment indicators are selected from three aspects which make up the nature disaster system, including hazard of disaster causing factors, vulnerability of disaster bearing environment and vulnerability of disaster affected bodies, to jointly construct a drought risk assessment indicator system in the Anhui Province. The data used from 2000 to 2020 are from the Anhui Statistical Yearbook (http://tjj.ah.gov.cn/ssah/qwfbjd/tjnj/index.html), 65 Years of Flood Control and Drought Relief in Anhui Province (Compiled by Anhui Flood Control and Drought Relief Headquarters Office), Flood and Drought Disasters in Anhui Province (Compiled by Department of Water Resources in Anhui Province), Drought Resistance Manual of Anhui Province (Compiled by Department of Water Resources in Anhui Province), etc. The above data are the annual scale data of the Anhui Province from 2000 to 2020, reflecting the overall level of Anhui Province in each year.

Based on previous research results (Tang 2008; Tang *et al.* 2009) and related studies (Fang *et al.* 2019; Kyatengerwa *et al.* 2020), and considering the accessibility of index data, eight indicators including precipitation anomaly percentage, temperature anomaly percentage, water resources per unit area, proportion of primary output value, proportion of effective irrigated area in cultivated land area, proportion of cultivated land, proportion of animal husbandry and forestry output value and proportion of agricultural population were selected as drought risk assessment indicators in this study, and they can respond to the driving mechanism of drought system.

2.2. Methods

In this study, a new method for assessment of regional drought risk is developed by combining the k-means clustering, WCA, MND and IMA into a general framework. In addition, we compare the conventional WCA and IDWI with the new method (Figure 3). In detail, data of eight indicators are standardized. On the one hand, drought risk value is calculated by the conventional WCA. On the other hand, the drought risk is graded by k-means clustering and the risk value of each cluster center

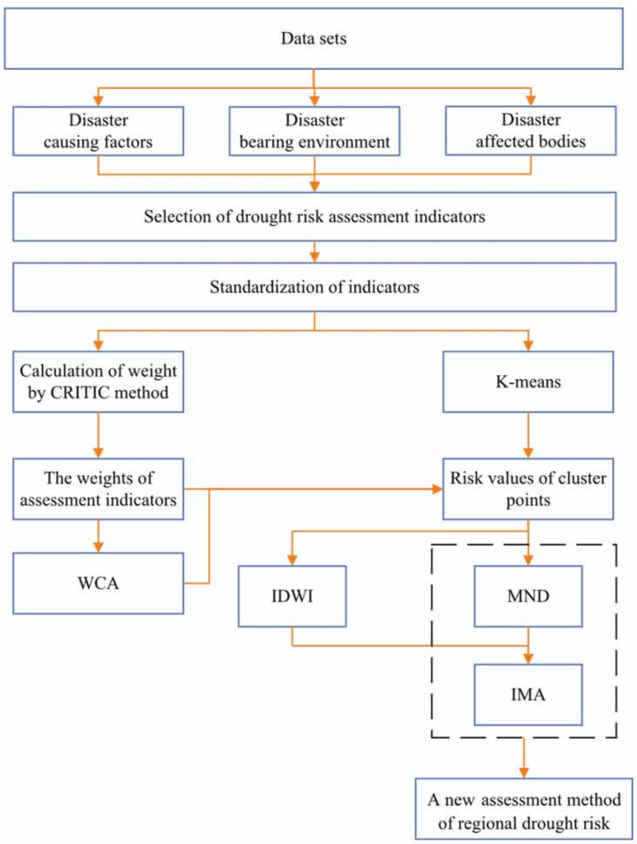


Figure 3 | The methodological framework used in this study.

is calculated based on WCA, and then the risk values of assessment objects are calculated by IDWI and MND to compare the risk quantification ability of the two methods. Finally, IMA is used to improve the risk discrimination between assessment objects.

2.2.1. Comprehensive assessment of drought risk based on the CRITIC weighting method

2.2.1.1. Calculation of weight. The Criteria Importance Through Intercriteria Correlation (CRITIC) method was proposed by Diakoulaki in 1995, which measured the objective weight of indicators based on the contrast intensity and conflict between indicators and its weight was more reasonable and stable than independence weighting method, entropy weighting method, coefficient of variation method (Diakoulaki *et al.* 1995; Wang *et al.* 2022). The weight is calculated as follows:

First, eight indicators are divided into positive indicators and reverse indicators according to their promoting or inhibiting effects on drought risk, and the data of eight indicators are standardized by formulas (1) and (2) to eliminate the influence of dimensions among indicators. Then, the variance S_j and conflict R_j among eight indicators are, respectively, calculated by formulas (3) and (4). Based on the results of S_j and R_j, the information amount C_j is calculated based on formula (5) and the weight of each indicator W_j is calculated based on formula (6). The formulas are as follows (Wang *et al.* 2021b):

$$\text{Positive indicator: } y_{ij} = \frac{y_{ij} - y_j^{\min}}{y_j^{\max} - y_j^{\min}} \tag{1}$$

$$\text{Reverse indicator: } y_{ij} = \frac{y_j^{\max} - y_{ij}}{y_j^{\max} - y_j^{\min}} \tag{2}$$

$$\begin{cases} \bar{y}_j = \dfrac{1}{m}\sum_{i=1}^{m} y_{ij} \\ S_j = \sqrt{\dfrac{\sum_{i=1}^{m}(y_{ij} - \bar{y}_j)^2}{m-1}} \end{cases} \tag{3}$$

$$R_j = \sum_{j'=1}^{n}(1 - r_{jj'})(j \neq j') \tag{4}$$

$$C_j = S_j \times R_j \tag{5}$$

$$W_j = \frac{C_j}{\sum_{j=1}^{n} C_j} \tag{6}$$

where $y_j^{\min}$ and $y_j^{\max}$ are the minimum and maximum of the jth indicator, y_{ij} is the standardized value of the jth indicator, $\bar{y}_j$ and S_j are the mean and standard deviation of the jth indicator and $r_{jj'}$ is the correlation coefficient between the jth and the j'th indicator.

2.2.1.2. WCA method. With reference to the index evaluation method, the WCA is used in this paper to calculate the comprehensive assessment value of drought risk based on standardized treatment and weight calculation of indicators. The WCA is a method to convert multiple indicator values into a single value according to the weight of each indicator. The calculation formula is as follows (Zhao *et al.* 2020):

$$R_i = \sum_{\mathrm{j}}^{n} W_j \times y_{ij} \tag{7}$$

where R_i is the drought risk value of the ith year, W_j is the weight of the jth indicator obtained by CRITIC method and y_{ij} is the normalized value of the jth indicator in the ith year.

2.2.2. Classification based on k-means clustering and calculation of cluster point risk value

2.2.2.1. The classification based on k-means clustering. K-means clustering is an unsupervised clustering algorithm, which has good applicability in drought assessment (Leščešen *et al.* 2019). k-means clustering algorithm is characterized by simple implementation, efficient and concise, and its clustering of drought risk grade is more objective than artificial classification. Data are usually divided into k clusters according to the Euclidean distance and the initial cluster results are optimized through iteration. The main steps are as follows (Yan *et al.* 2022):

- k samples were selected from the standardized samples in Section 2.2.1.1 as the initial cluster centers $a = a_1, a_2, \cdots, a_k$.
- Calculate the distance from each sample x_i to cluster centers in the dataset and divide it into the cluster whose cluster center with the smallest distance to x_i.
- According to the cluster result, recalculate the new cluster center a_j:

$$\mathrm{a}_j = \frac{1}{|c_j|}\sum_{x_i \in c_j} x_i \tag{8}$$

where c_j is the jth cluster and $|c_j|$ is the number of members in the jth cluster.

- Repeat Steps (2) and (3) until the termination iteration condition (usually the number of iterations, the minimum error) is reached. So far, the samples have been clustered into k classes and the clustering center of each class can be obtained.

2.2.2.2. The calculation of risk value of cluster centers. Based on the above cluster results, k clusters and their corresponding cluster centers can be obtained. According to the idea of WCA, the drought risk value of each clustering center can be calculated. The formula is as follows (Zhao *et al.* 2020):

$$r_i = \sum_{j=1}^{n} W_j \times a_{ij} \tag{9}$$

where r_i is the risk value of ith cluster center, a_{ij} is the value of the ith cluster center at the jth indicator and W_j is the weight of the jth indicator obtained by CRITIC method.

2.2.3. The information redistribution of cluster point based on IDWI

The information redistribution of cluster point based on IDWI is to combine k-means clustering and IDWI to requantify the risk value of the assessment objects in the same risk grade. After k clusters and their cluster centers are obtained by k-means clustering, the risk value of each cluster center is calculated based on the weight obtained by CRITIC weighting method and the risk value of each cluster member is obtained by IDWI. The main steps are as follows (Imen *et al.* 2021):

- Calculate the square of the distance from cluster members to each cluster center D (Ohlert *et al.* 2022)

$$D_{ij} = \sum_{j=1}^{\mathrm{k}} (x_i - a_j)^2 \tag{10}$$

where x_i is the ith cluster member and a_j is the jth cluster center.

- Calculate the weight of cluster members to each cluster center W_{ij}

$$W_{ij} = \frac{1/D_{ij}}{\sum_{j=1}^{k} 1/D_{ij}} \tag{11}$$

- Calculate the risk value at each cluster member R_i'

$$R_i' = \sum_{j=1}^{k} W_{ij} \times r_j \tag{12}$$

where r_j is the risk value of the jth cluster center.

2.2.4. The information redistribution of cluster point based on MND

The information redistribution of cluster point based on MND is similar to the information redistribution based on IDWI. The difference is that the information of cluster points is diffused to each cluster member through the MND after calculating the risk value of each cluster center, namely get drought risk value of each year in Anhui Province. The main processes of MND are as follows (Huang & Shi 2002):

- P is a sample set containing m sample points, and each sample point contains n components. The sample is denoted by $P = \{(x_{11}, x_{12}, \cdots, x_{1n}), (x_{21}, x_{22}, \cdots, x_{2n}), \cdots, (x_{m1}, x_{m2}, \cdots, x_{mn})\}$, and its domain is denoted by $U_1, U_2, \cdots, U_m$ and $U_i \subset R$, which is denoted by $U_i = \{u_{1i}, u_{2i}, \cdots, u_{ti}\}$. In the case of this study, P is the cluster center sample set, and U is the standardized data set of the assessment indicator.
- According to the principle of information diffusion, the information carried by the sample points $(x_{1n}, x_{2n}, \cdots, x_{mn})$ in P is allocated to the points in U_i according to μ by the MND, the formula is as follows (Yu *et al.* 2020; Zhang *et al.* 2022):

$$\mu_{jk} = \frac{1}{2\pi h_1 h_2 \cdots h_n} \exp\left[-\frac{(u_{k1} - x_{j1})^2}{2h_1^2} - \frac{(u_{k2} - x_{j2})^2}{2h_2^2} - \cdots - \frac{(u_{kn} - x_{jn})^2}{2h_n^2}\right] \tag{13}$$

The diffusion coefficient h_n is calculated by the following formula (Liu *et al.* 2018):

$$h_n = \begin{cases} 0.8146(b-a), n=5 \\ 0.5690(b-a), n=6 \\ 0.4560(b-a), n=7 \\ 0.3860(b-a), n=8 \\ 0.3362(b-a), n=9 \\ 0.2986(b-a), n=10 \\ 2.6851(b-a)/(n-1), n \geq 11 \end{cases} \tag{14}$$

where $b = \max_{1 \leq i \leq n} \{x_i\}$, $a = \min_{1 \leq i \leq n} \{x_i\}$.

- The information matrix Q of P on U is calculated according to Step (2), and the weight W of information assigned to each point in the domain is calculated as follows (Lu *et al.* 2014):

$$Q_j = \sum_{k=1}^{t} \mu_{jk} \tag{15}$$

$$W_{jk} = \frac{\mu_{jk}}{Q_j} \tag{16}$$

- Calculate the risk value of each point in the domain:

$$R'_k = \sum_{j=1}^{m} W_{jk} \times \mathrm{r_j} \tag{17}$$

where r_j is the risk value of the jth cluster center.

2.2.5. The IMA on drought risk

Considering that the risk quantification method based on cluster points is closely related to the risk value of cluster centers, and the risk value of cluster centers is close to each other, which cause the little inter-annual variation between the risk value of the assessment objects after quantification, and the discrimination of drought risk value of different years between different risk grades will be low. Therefore, it is considered to map the quantified risk to the range of [0,1], which can not only affect the classification result of drought risk grade, but also make the discrimination degree of risk values in the same grade or different grades more obvious. The mapping process is as follows:

- According to the risk value of cluster centers, the k-means cluster results were set into five grades: No drought (ND) – mild drought (MID) – moderate drought (MOD) – severe drought (SD) – extreme drought (ED). R_1, R_2, R_3, R_4 were, respectively, set as the minimum value of the four grades range from MID to ED after the quantification to risk of the assessment objects, and they were defined as the classification thresholds of the grades range from ND to ED, namely $[0, R_1]$, $[R_1, R_2]$, $[R_2, R_3]$, $[R_3, R_4]$ and $[R_4, 1]$.

- The intervals in Step (1) are, respectively, mapped to [0, 0.2], [0.2, 0.4], [0.4, 0.6], [0.6, 0.8] and [0.8, 1], and the mapping formula is as follows:

$$r' = \frac{r - R_{\min}}{R_{\max} - R_{\min}} \times (t_{\max} - t_{\min}) + t_{\min} \tag{18}$$

where r and r' are the risk values before and after mapping; $R_{\min}$ and $R_{\max}$ are the upper and lower limits of each subinterval before adjustment; $t_{\min}$ and $t_{\max}$ are the upper and lower limits of the target subinterval, respectively.

3. RESULTS AND DISCUSSION

3.1. Comprehensive assessment results of drought risk based on CRITIC weighting method

The drought risk in Anhui Province from 2000 to 2020 was evaluated in this paper. According to formulas (1) – (6), the weight of each evaluation indicator could be calculated. Table 1 shows the weights calculated based on the CRITIC method.

On the basis of objective weighting of the evaluation indicators, the drought risk value of the Anhui Province from 2000 to 2020 was calculated according to the WCA. Figure 4 shows the comprehensive assessment results of drought risk based on CRITIC weighting method. It can be found from the figure that the drought risk value is between 0.15 and 0.6 in the Anhui Province, with obvious differentiation between high value and low value and large inter-annual variation. However, this method cannot determine the drought risk grade of each year and can only compare the relative magnitude of drought risk between years.

3.2. Assessment results of drought risk based on IDWI

Figure 5 shows the assessment results of drought risk based on the IDWI. After the classification of the k-means cluster points, the risk values of the five cluster points are 0.3056, 0.3827, 0.4024, 0.4966 and 0.5705, respectively. It can be found from Figure 5(a) that the drought risk can be classified into five risk grades after classification, which has a better ability of classification compared with the traditional WCA. At the same time, the drought risk of each year can be quantified by IDWI. However, the drought risk values of the years in ND, MID and MOD are close, and the discrimination degree is weak. After interval mapping, the variation range of inter-annual risk value was increased and the discrimination degree of drought risk value between different grades was significantly improved under the condition of maintaining the original change relationship of drought risk value in Figure 5(b), thus achieving the purpose of drought risk refinement.

3.3. Drought risk assessment results based on MND

Figure 6 shows the assessment results of drought risk based on MND. The drought risk value of each year is quantified by MND based on the risk values of five cluster points. Compared with Figure 5(a), the inter-annual variation in drought risk value has increased relatively in Figure 6(a). However, the difference in risk value for MID and MOD needs to be improved. After the interval mapping, the discrimination degree of drought risk value between different grades was significantly

Table 1 | Weights of evaluation indicators based on the CRITIC weighting method

Weighting method	Precipitation anomaly percentage	Temperature anomaly percentage	Water resources per unit area	Proportion of primary output value	Proportion of effective irrigated area in cultivated land area	Proportion of cultivated land	Proportion of animal husbandry and forestry output value	Proportion of agricultural population
Weights based on the CRITIC weighting method	0.110	0.126	0.097	0.096	0.092	0.259	0.105	0.115

Figure 4 | Drought risk value based on the CRITIC weighting method.

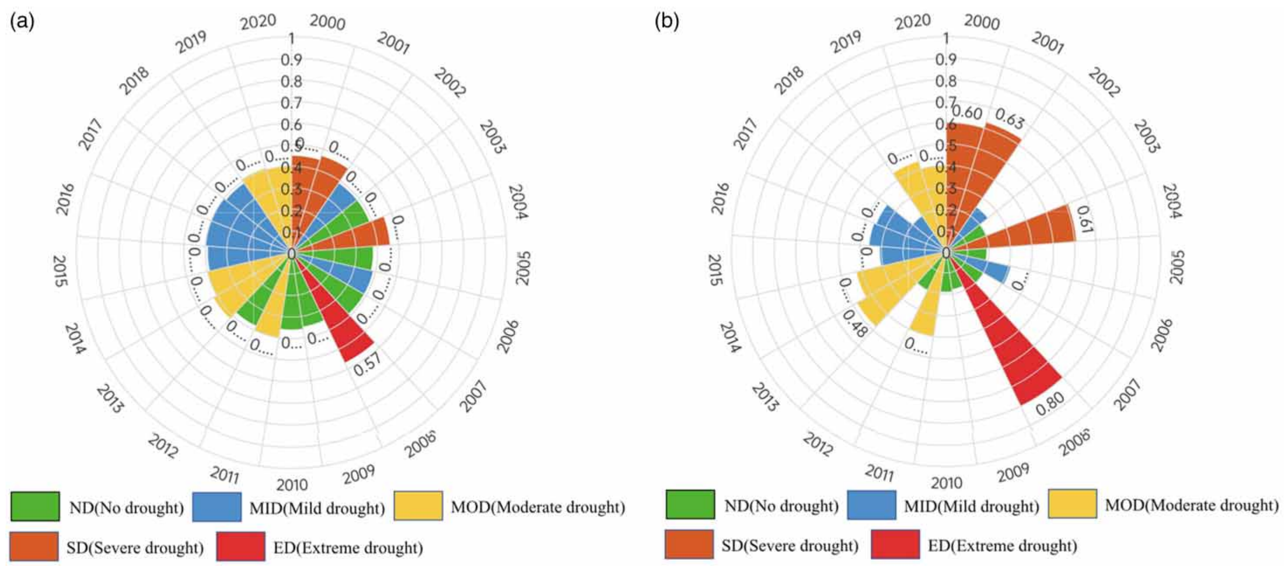


Figure 5 | Drought risk value and grade based on IDWI: (a) before mapping and (b) after mapping.

improved under the condition of maintaining the original change relationship of drought risk value in Figure 6(b), which shows that the method is reasonable in this paper.

3.4. Analysis of the results of three drought risk assessment methods

From Table 2, it can be found that the inter-annual variation trends of the drought risk values calculated by the three methods are consistent basically, and the differences only exist in individual years. The two drought risk quantification methods based on k-means cluster points proposed in this paper not only have the ability of WCA to quantify drought risk, but also make up for the defect that it cannot classify drought risk and can effectively divide drought risk into five grades.

By comparing Figure 5(a) and 5(b) with Figure 6(a) and 6(b), it can be found that although the quantification and grade division of drought risk have been realized without introducing interval mapping, there are only minor differences in the risk values between different grades and within the same grade. However, after the introduction of interval mapping, the difference in drought risk values was obvious visually, especially the discrimination degree of risk values in different years

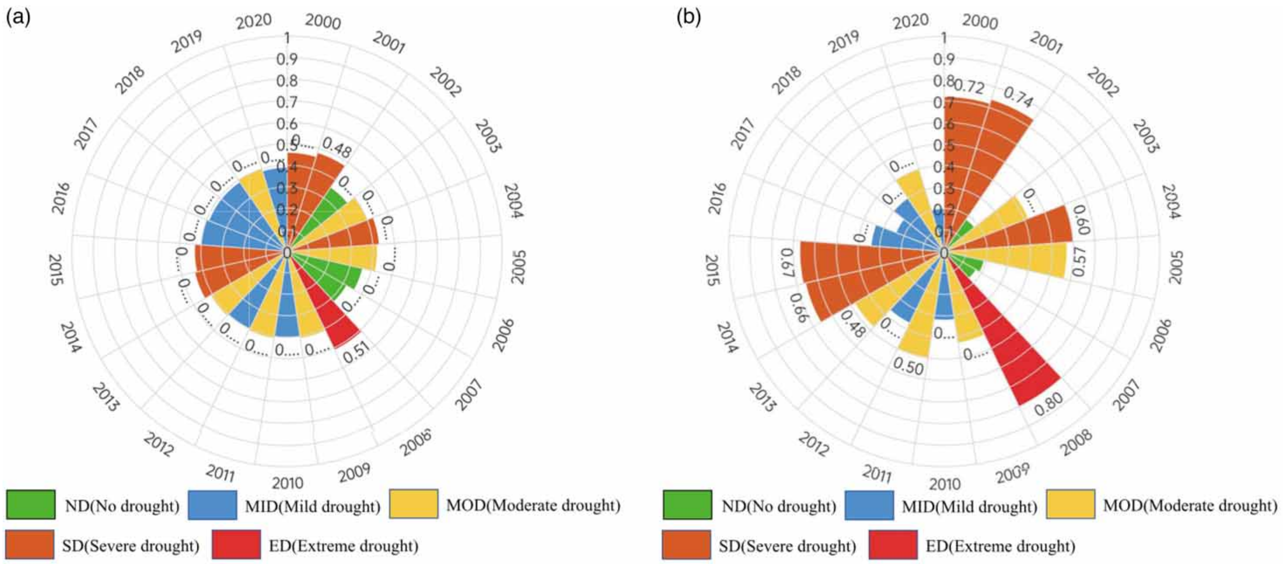


Figure 6 | Drought risk value and risk grade based on MND: (a) before mapping and (b) after mapping.

between different grades was effectively separated. Therefore, it is effective and necessary to introduce interval mapping in this case.

3.5. Analysis of two risk quantification methods based on k-means cluster points

Tables 3 and 4 are the comparison of the results obtained by two risk quantification methods and the drought record in statistical material. It can be found from the tables that the results obtained by two risk quantification methods are not in line with the records in the statistical material in some years.

For the one phenomenon that the assessment grade in this paper is lower than the record in the statistical material, the main reason is that Anhui Province is located in the monsoon region, and the uneven distribution of rainfall within the year leads to obvious wet and dry seasons, which easily leads to drought. However, the annual scale data selected in this paper cannot fully reflect the contradiction between the supply and demand of water and the imbalance of water resources within the year.

For the other phenomenon that the assessment grade in this paper is higher than the record in the statistical material, the main reason is that the risk grade in the yearbook represents the overall drought disaster situation in the Anhui Province which is the actual disaster after human intervention. However, the drought risk grade in this paper is comprehensively assessed from the perspectives of hazard of disaster causing factors, vulnerability of disaster bearing environment and vulnerability of disaster affected bodies, which represents the theoretical grade of drought risk that may occur. It is important to note that drought risk is not the same as drought disaster, and only when the possibility of impact and harm caused by drought becomes real, will risk translate into disaster (Qu *et al.* 2015). Thus, the results in this paper may be higher than the historical records.

It can be found that only 50% of the years in which the assessment results obtained by information redistribution based on IDWI matched the historical record exactly or differed by one grade. However, the result obtained by information redistribution based on MND can reach 71.43%, indicating that the use of information diffusion method to quantify drought risk can improve the rationality of the results and make it closer to the actual situation. This advantage may be caused by the following reasons: first, the regional drought system is characterized by high dimensionality, nonlinearity, correlation complexity and uncertainty (Zhang 2022). Second, the calculation principle of MND is also nonlinear, while the principle of IDWI is linear conversely (Huang *et al.* 2021). Third, IDWI is susceptible to the influence of points with a short distance (Ohlert *et al.* 2022).

3.6. Limitations and prospects

In addition to the two factors mentioned in Section 3.5., the results of this paper are also influenced by the data availability. In fact, there are many indicators that can be used for drought risk assessment research. However, some data are not collected in

Table 2 | Drought risk values in the Anhui Province based on different risk assessment methods

State	Methods	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Before mapping	WCA	0.44	0.55	0.43	0.29	0.50	0.33	0.42	0.38	0.57	0.23	0.19	0.23	0.17	0.54	0.41	0.39	0.35	0.42	0.38	0.44	0.35
	IDWI	0.45	0.47	0.39	0.39	0.46	0.38	0.39	0.39	0.57	0.35	0.36	0.40	0.38	0.42	0.40	0.39	0.40	0.40	0.39	0.41	0.40
	MND	0.46	0.48	0.36	0.40	0.43	0.42	0.36	0.32	0.51	0.41	0.40	0.41	0.40	0.41	0.44	0.43	0.40	0.39	0.39	0.40	0.39
After mapping	IDWI	0.60	0.63	0.25	0.20	0.61	0.19	0.31	0.20	0.80	0.18	0.19	0.40	0.20	0.48	0.43	0.31	0.36	0.35	0.20	0.44	0.40
	MND	0.72	0.74	0.18	0.42	0.60	0.57	0.19	0.17	0.80	0.43	0.32	0.50	0.37	0.48	0.66	0.67	0.34	0.24	0.30	0.40	0.20

Table 3 | Comparison of the drought grade obtained by the two risk quantification methods and recorded in the Statistical Yearbook

Year	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Statistical yearbook	★	★	●	◆	●	◆	■	▲	▲	■	●	■	■	■
IDWI	■	■	●	▲	■	▲	●	▲	★	▲	▲	◆	▲	◆
MND	■	■	▲	◆	■	◆	▲	▲	★	◆	●	◆	●	◆

Note: ▲ indicates ND (No drought); ● indicates MID (Mild drought); ◆ indicates MOD (Moderate drought); ■ indicates SD (Severe drought); and ★ indicates ED (Extreme drought).

Table 4 | The statistical table of deviations between the drought grade obtained by the two risk quantification methods and recorded in the Statistical Yearbook

Grade differences	Equal	One-grade differences	Two-grade differences	Three-grade differences	Four-grade differences
IDWI	14.29%	35.71%	28.57%	14.29%	7.14%
MND	28.57%	42.86%	14.29%	7.14%	7.14%

the dataset built in this paper (some optional indicators lack systematic statistical data), so the selection of assessment indicators is limited. In addition, there is no short-term statistical data for some indicators to choose from (such as proportion of cultivated land and agricultural population), so the annual scale data selected in this paper will have a certain impact on the quantification accuracy of risk, which will result in deviations between the results in this paper and the records in statistical material.

However, it is foreseeable that with the improvement of the spatiotemporal attributes of statistical data, we can choose a shorter time scale and build a more refined risk assessment indicator system. Therefore, on the one hand, we will strengthen cooperation with the official departments to obtain more detailed statistical data while narrowing the scope of the research area. On the other hand, we will try to find the relationship between the hazard of disaster causing factors, vulnerability of disaster bearing environment, vulnerability of disaster affected bodies and the drought disaster, so as to keep the theoretical disaster situation consistent with the actual disaster situation. Thus, the above problems can be effectively solved, and the quantification of risk grades will be more accurate.

4. CONCLUSION

Anhui Province was taken as the research object and the annual scale assessment of drought risk was carried out in this paper. Aiming at the defects of conventional WCA and according to the classification method based on k-means cluster points, the drought risk value in the Anhui Province during 2000–2020 was quantified by IDWI and MND, and the drought risk value was refined by IMA, which can expand the way of risk assessment. The main conclusions are as follows:

- The risk of drought can be quantified by the conventional WCA, but the risk is difficult to be graded directly, which needs to be graded by other methods.
- On the basis of k-means clustering and WCA to calculate the risk value of cluster centers, IDWI and MND were used to calculate the drought risk value of each clustering member, which realized the combination of drought risk quantification and grade division.
- Although the quantification of drought risk can be realized by information redistribution based on IDWI and MND, the drought risk value obtained by MND method is more reasonable and closer to the actual situation, which benefits from its nonlinear diffusion principle.
- By introducing IMA of drought risk, the dispersion effect of risk value obtained by redistribution of information based on k-means clustering points is greatly improved. The discrimination between the risk values of the assessment object is improved, especially the risk values of different grades, so that the grades can be displayed more intuitively. Therefore, the follow-up utilization space is widened by this method, and the thinking of risk quantification in drought risk assessment is broadened.

In the future, we will try to promote the statistics of small-scale data and the establishment of relevant databases for drought risk assessment. At the same time, we will apply the method proposed in this paper to other drought-prone regions in China. For other regions, we will adjust the assessment indicators according to the data availability in the region under the premise that assessment indicators still respond to the driving mechanism of drought system.

ACKNOWLEDGEMENTS

This research was funded by the Key Science and Technology Project of Education Department of Jiangxi Province (No. GJJ201901). The authors thank Water Resources Department of Anhui Provincial and Statistics Bureau of Anhui Province, for providing the data for this study. The authors would like to thank all the reviewers who participated in the review.

AUTHOR CONTRIBUTIONS

W.X. did data curation, formal analysis, investigation, methodology, software, visualization, writing of the original draft. M.T. conceptualized the study, acquisition of funds, did investigation, methodology, project administration, resources, supervision, writing, review and editing of the article. Y.L. wrote the original draft.

DATA AVAILABILITY STATEMENT

All relevant data are available from an online repository or repositories: http://tjj.ah.gov.cn/ssah/qwfbjd/tjnj/index.html.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Ai, P., Chen, B., Yuan, D., Hong, M. & Liu, H. 2020 Dynamic risk assessment of drought disaster: a case study of Jiangxi Province, China. *Journal of Water and Climate Change* **12** (5), 1761–1777. https://doi.org/10.2166/wcc.2020.141.

Band, S. S., Karami, H., Jeong, Y.-W., Moslemzadeh, M., Farzin, S., Chau, K.-W., Bateni, S. M. & Mosavi, A. 2022 Evaluation of time series models in simulating different monthly scales of drought index for improving their forecast accuracy. *Frontiers in Earth Science* **10**. https://doi.org/10.3389/feart.2022.839527.

Chou, J., Xian, T., Zhao, R., Xu, Y., Yang, F. & Sun, M. 2019 Drought risk assessment and estimation in vulnerable eco-regions of China: under the background of climate change. *Sustainability* **11** (16), 4463. https://doi.org/10.3390/su11164463.

Diakoulaki, D., Mavrotas, G. & Papayannakis, L. 1995 Determining objective weights in multiple criteria problems: the critic method. *Computers & Operations Research* **22** (7), 763–770. https://doi.org/10.1016/0305-0548(94)00059-H.

Fang, G., Tu, Y., Wen, X., Yan, M. & Tan, Q. 2019 Study on the development process and evolution characteristics of meteorological drought in the Huaihe River Basin from 1961 to 2015. *Journal of Hydraulic Engineering* **50** (05), 598–611. https://doi.org/10.13243/j.cnki.slxb.20180986.

Huang, C. & Shi, Y. 2002 *Towards Efficient Fuzzy Information Processing: Using the Principle of Information Diffusion*. Springer Science & Business Media, Heidelberg.

Huang, S., Chang, J., Leng, G. & Huang, Q. 2015 Integrated index for drought assessment based on variable fuzzy set theory: a case study in the Yellow River basin, China. *Journal of Hydrology* **527**, 608–618. https://doi.org/10.1016/j.jhydrol.2015.05.032.

Huang, H., Yin, K. & Jin, G. 2021 Information diffusion interpolation model based on the PSO-CV algorithm and its application on source regions of the Yellow River. *Pearl River* **42** (09), 21–27. https://doi.org/10.3969/j.issn.1001-9235.2021.09.004.

Imen, K., Ghofrane, L. & Ahmed, G. 2021 Use of GIS based inverse distance weighted interpolation to assess surface water quality: case of Wadi El Bey, Tunisia. *Environmental Technology & Innovation* **24**, 101892. https://doi.org/10.1016/J.ETI.2021.101892.

IPCC 2021 *Climate Change 2021: the Physical Science Basis*. Cambridge University Press, Cambridge.

Jin, J., Song, Z., Cui, Y., Zhou, Y., Jiang, S. & He, J. 2016 Research progress on the key technologies of drought risk assessment and control. *Journal of Hydraulic Engineering* **47** (03), 398–412. https://doi.org/10.13243/j.cnki.slxb.20150945.

Kyatengerwa, C., Kim, D. & Choi, M. 2020 A national-scale drought assessment in Uganda based on evapotranspiration deficits from the Bouchet hypothesis. *Journal of Hydrology* **580** (C), 124348. https://doi.org/10.1016/j.jhydrol.2019.124348.

Leščešen, I., Dolinaj, D., Pantelić, M. & Popov, S. 2019 Drought assessment in Vojvodina (Serbia) using k-means cluster analysis. *Journal of the Geographical Institute 'Jovan Cvijić' SASA* **69** (1), 17–27. https://doi.org/10.2298/IJGI1901017L.

Liu, J., Li, S., Wu, J., Liu, X. & Zhang, J. 2018 Research of influence of sample size on Normal information diffusion based on the Monte Carlo method: risk assessment for natural disasters. *Environmental Earth Sciences* **77** (13), 1–8. https://doi.org/10.1007/s12665-018-7612-2.

Lu, H., Yang, L.-Z. & Gao, J.-M. 2014 The application of information diffusion technique in probabilistic analysis to grassland biological disasters risk. *Ecological Modelling* **272**, 264–270. https://doi.org/10.1016/j.ecolmodel.2013.10.014.

Luo, D., Wang, X. & Li, H. 2019 Prediction of regional agricultural drought loss rate based on multi-scale combined model. *Journal of North China University of Water Resources and Electric Power (Natural Science Edition)* **40** (03), 1–6. https://doi.org/10.19760/j.ncwu.zk.2019029.

Moura, B. N. R., Guimarães, S. C. A., da, S. R. M., dos, S. C. A. C., Zhong, L. & Wyndham, Q. N. 2021 Geospatial cluster analysis of the state, duration and severity of drought over Paraíba State, northeastern Brazil. *Science of the Total Environment* **799**, 149492. https://doi.org/10.1016/J.SCITOTENV.2021.149492.

Ohlert, P. L., Bach, M. & Breuer, L. 2022 Accuracy assessment of inverse distance weighting interpolation of groundwater nitrate concentrations in Bavaria (Germany). *Environmental Science and Pollution Research International*. https://doi.org/10.1007/S11356-022-22670-0.

Qu, Y., Gao, H., Lv, J., Su, Z., Cheng, X. & Sun, H. 2015 Agricultural drought disaster risk assessment in China based on the regional disaster system theory. *Journal of Hydraulic Engineering* **46** (08), 908–917. https://doi.org/10.13243/j.cnki.slxb.20150049.

Sahana, V., Mondal, A. & Sreekumar, P. 2021 Drought vulnerability and risk assessment in India: sensitivity analysis and comparison of aggregation techniques. *Journal of Environmental Management* **299**, 113689. https://doi.org/10.1016/J.JENVMAN.2021.113689.

Shamshirband, S., Hashemi, S., Salimi, H., Samadianfard, S., Asadi, E., Shadkani, S., Kargar, K., Mosavi, A., Nabipour, N. & Chau, K.-W. 2020 Predicting standardized streamflow index for hydrological drought using machine learning models. *Engineering Applications of Computational Fluid Mechanics* **14** (1), 339–350. https://doi.org/10.1080/19942060.2020.1715844.

Shim, I., Kim, H., Hong, B., An, J. & Hwang, T. 2021 Drought vulnerability assessment and cluster analysis of island areas taking Korean island areas at Eup (Town) and Myeon (Subcounty) levels as study targets. *Water* **13** (24), 3657. https://doi.org/10.3390/W13243657.

Tang, M. 2008 *Risk Management on the Drought Disaster: Theory, Method and Applications*. Wuhan University, Wuhan.

Tang, M., Shao, D., Yao, C. & Huang, X. 2009 Improved catastrophe theory based evaluation method and its application to drought disaster risk evaluation. *Journal of Hydraulic Engineering* **40** (07), 858–862 + 869. https://doi.org/10.13243/j.cnki.slxb.2009.07.017.

Wang, L., Zhang, J., Elmahdi, A., Shu, Z., Wu, Y. & Wang, G. 2021a Evolution characteristics and relationship of meteorological and hydrological droughts from 1961 to 2018 in Hanjiang River Basin, China. *Journal of Water and Climate Change* **13** (1), 224–246. https://doi.org/10.2166/wcc.2021.267.

Wang, X., Zhou, Z., Sun, L., Xie, G. & Lou, Q. 2021b Research on the evaluation index system of 'new energy cloud' operation mode based on CRITIC weighting method and AHP method. *IOP Conference Series: Earth and Environmental Science* **831** (1). https://doi.org/10.1088/1755-1315/831/1/012017.

Wang, D., Li, S., Toktarbek, S., Jiakula, N., Ma, P. & Feng, Y. 2022 Research on the coordination between agricultural production and environmental protection in Kazakhstan based on the rationality of the objective weighting method. *Sustainability* **14** (6), 3700. https://doi.org/10.3390/SU14063700.

Xu, L., Wang, H., Ma, X., Chen, J. & Ma, J. 2013 Study on framework and adaptive strategies of drought disaster risk management in Yunnan Province. *Areal Research and Development* **32** (02), 103–108.

Yan, S., Liu, W., Yang, P., Wu, F. & Yan, Z. 2022 Multi group sparrow search algorithm based on K-means clustering. *Journal of Beijing University of Aeronautics and Astronautics*, 1–13. https://doi.org/10.13700/j.bh.1001-5965.2022.0328.

Yu, X., Yu, X., Li, C. & Ji, Z. 2020 Information diffusion-based risk assessment of natural disasters along the Silk Road Economic Belt in China. *Journal of Cleaner Production* **244** (C), 118744. https://doi.org/10.1016/j.jclepro.2019.118744.

Zhang, G. 2022 *Grey Modeling Method for Regional Agricultural Drought Diaster Prediction and Assessment*. North China University of Water Resources and Electric Power.

Zhang, J., Wang, J., Chen, S., Tang, S. & Zhao, W. 2022 Multi-hazard meteorological disaster risk assessment for agriculture based on historical disaster data in Jilin Province, China. *Sustainability* **14** (12), 7482. https://doi.org/10.3390/SU14127482.

Zhao, J., Zhang, Q., Zhu, X., Shen, Z. & Yu, H. 2020 Drought risk assessment in China: evaluation framework and influencing factors. *Geography and Sustainability* **1** (3), 220–228. https://doi.org/10.1016/j.geosus.2020.06.005.

First received 19 September 2022; accepted in revised form 9 November 2022. Available online 17 November 2022