Improved grey clustering method in risk zonation of mountain flash flood disaster

Qiong Li[1], Jianzhong Zhou[2*], Jiaming Cai[3], Huaiwei Sun[4]

[1]School of Mathematics and Physics, Hubei Polytechnic University, Huangshi 435003, China
[1,2,3,4]College of Hydropower and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China

Abstract**—** Flash floods are considered one of the worst weather-related natural disasters. Flash floods are dangerous because they are sudden and highly unpredictable. Identification of the locations of high-risk areas has a major effect on the improvement of flash flood disaster control and prevention. Earlier work conducted on flood disaster risk zonation was commonly based on Digital Elevation Mode (DEM) data and statistical yearbook data and used an index, such as rainfall, topography, slope, or river distribution, with the analytic hierarchy process (AHP) method to determine the weighting. In this method, the final regional risk map was created by using ArcGIS map algebra superposition. In the present study, an improved gray clustering method is put forward to improve the comprehensive evaluation of the risk of mountain flash flood disasters by constructing the exponential whitening function and by using the information entropy weight method, which produces results that are more accurate and more reliable than those of the traditional method. This improved method can make full use of the limited information available, improving not only the resolution but also the influence of the subjective method, and produces more objective and accurate evaluation results. We obtain the risk degree by combining the information entropy weight and improved whitening function approaches in a gray clustering methodology. Additionally, a method is applied to develop models for mapping the risk grade in zones of 1436 towns and counties in Hubei Province with remotely sensed (RS) data and the ArcGIS platform. The results show that the improved approach is useful for rapidly assessing flash flood hazard and vulnerability and for completing risk assessments in mountain areas.

Keywords— ***improved gray clustering method, whitening function, information entropy, flood disaster risk, evaluation, zonation.***

I. INTRODUCTION

Flash flooding is one of the major natural disasters that may hamper human development in flash flood areas. A mountain flash flood disaster is one of the most serious natural disasters. Mountain flash flood disasters occur suddenly, are considerably destructive, have short durations and cause serious harmed in the form of many casualties and considerable property loss. China is one of the countries that seriously suffer from mountain flash floods. Jonkman (2005) studied flash flood data from 1975 to 2002 and found that flash flood mortality is higher than that for other natural hazards. The potential for flash flood casualties and damages is also increasing in many regions due to the social and economic development which imply pressure on land use. Consequently, the flash flood hazard is expected to increase in frequency and severity because of the impacts of global change on climate, severe weather in the form of heavy rains and river discharge conditions. Therefore, the management of flash flood risks is a critical component of public safety and quality of life.

As one of the important and fundamental steps in flood regionalization, flood risk evaluation has general public concern. Many achievements in flood risk assessment research have been realized. Currently, the main methods of evaluating and locating flood disaster risk include geomorphologic methods (Haruyama et al. 1996), hydrology-hydraulic models, system simulation methods (Solaimani et al. 2005; Elawad et al. 2004; Smemoe et al. 2004),methods based on historical disaster data(Liu and Shi 2001;Huang et.al) and ancient flood data(Bonito et al. 1998; Bonito et al. 2003; Bonito et al. 2004), methods based on remote sensing and GIS (Sanyal and Lu 2006), and machine system analysis methods (Tang and Zhu 2005;Li et al. 2005). However, fairly few research papers have focused on mountain flash flood zonation. Even rarer are integrated analyses of flood risk that comprehensively consider flood formation mechanisms, climate, geomorphology, river water systems, and historical flood data. The research conducted in this paper is focused on comprehensive risk assessment and zonation, considering the important factors affecting flash floods: precipitation, topography, water system, vegetation, GDP, population, cultivated land and flood control capacity.

In fact, the main characteristics of natural disaster systems are the uncertainty and complexity of the system. Determination of the weights of the indicators is a problem because of the wide range of both natural uncertainty and approaches. Some in myriad sample can be solved by probability and statistics ways, and some in kenning uncertainty can be dealt with by fuzzy mathematics. However, there also exists another category on uncertainty in less data and little sample, incomplete information and devoid of experience, which is suitable to be dealt with only by gray system theory (Deng,2005). In general terms, the uncertainty in less data and incomplete information is designated grayness. Thus, systems possessing grayness are said to be gray systems.

The gray theory provides the applications of clustering analysis, relational analysis, predication, and decision for the gray system (Deng, 1989). The so-called ‘‘gray” means that system information is incomplete, unclear, and uncertain. It is a useful method to address the problems of limited, deficient, and no rules available for data processing. Its analysis makes use of minor data and does not demand strict statistical procedures and inference rules. Recent studies have emphasized the importance of a comprehensive assessment of the flood risk using the gray system method (Liu, 2010).

To address the problem of nonadjacent domain weighted superposition failure caused by the traditional gray clustering whitening function, this paper proposes a whitening function construction method based on exponential distribution, avoiding the condition of a zero weighting, and discusses on the steps of flood risk assessment in detail.

In view of the complexity of the causes and the randomness of the occurring process of the flood disaster, we proposed a comprehensive assessment by introducing the concept of information entropy into the improved gray method and constructing a typical exponential whitening weight function. Based on the above characteristic, this method can effectively solve the zero-weight problem, make full use of the simple data and largely reserve the information implied in the clustering weight by modifying the clustering weight with the values reflected by the entropy. Finally, with data from Hubei Province, we graded the risk evaluation of 1436 towns in Hubei Province, which verifies the validity and objectivity of the method described in the paper. This illustrative example verifies that this method is simple and reasonable and can extend application range of the gray clustering in flash flood zonation.

II. GRAY CLUSTERING MODELING AND IMPROVEMENTS

2.1 Traditional gray clustering modeling and limitations

The concept of gray clustering is based on samples of every object, in accordance with the given indicators by means of the whitening function to abstract the actual sample into scales, and objects are sorted into their corresponding gray league to synthesize the priorities for selecting an appropriate alternative, which is said to be gray clustering.

As in the fuzzy approach, in this paper, different ranges of risk assessment are described as very low, low, medium, high, and very high by the so-called ‘‘whiten-weight function”. The common ones are triangular and trapezoidal functions. The trapezoidal functions are usually used and can be represented as

$$f_{\tilde{A}}(x)=\begin{cases} 0 & x<a \\ \dfrac{x-a}{b-a} & a\le x\le b \\ 1 & b\le x\le c \\ \dfrac{x-d}{c-d} & c\le x\le d \\ 0 & x>d \end{cases} \tag{1}$$

where $x\in R$, $a\le b\le c\le d$, a and b are the lower limit and upper limit of $\tilde{A}$.Specifically, $\tilde{A}$ is the triangle fuzzy number when $b=c$. $\tilde{A}$ is a real number when $a=b=c=d$, as shown in Fig. 1

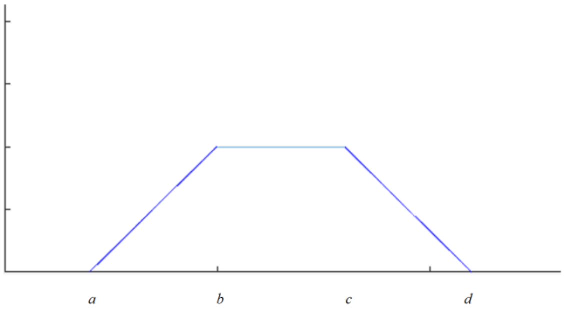

FIG. 1.TRADITIONAL LINEAR WHITENING FUNCTION

Fig. 1 shows the construction of the whitening function in the traditional gray clustering model. The whitening function is piecewise linear, and the whitening function of each rank is non-zero when it is adjacent to the interval the data belongs to, meaning that there is a relationship between only two adjacent levels in the coverage of the whitening function, and useful information is lost. In addition, the whitening function of the rank that is not adjacent to the level is zero, possibly causing a problem of zero weight.

The shortcoming of existing gray clustering methods is that the range of the linear whitening function is too narrow, especially when the distribution of the monitoring value is discrete, causing the loss of useful information. The improved gray clustering method makes better use of characteristic information in research data and calculates risk evaluation levels of the areas.

2.2 Improved gray clustering

2.2.1 Exponential whitening function

In this paper, we consider the issue that the linear whitening function in a conventional gray clustering model considers the relationship between only adjacent levels, and we construct f_{jk}, which can effectively expand the coverage of a whitening function and greatly improve the utilization of information, as shown in Fig. 2.

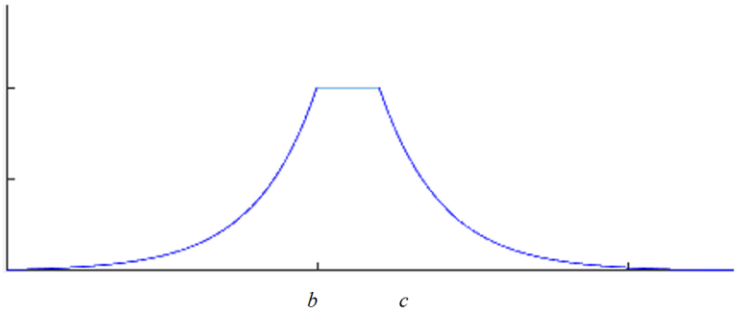

FIG. 2. EXPONENTIAL WHITENING FUNCTION

Let

$$f_{\tilde{A}}(x)=\begin{cases} e^{\frac{x-b}{b}} & x\in(0,b) \\ 1 & x\in[b,c] \\ e^{\frac{c-x}{c}} & x\in(c,+\infty) \end{cases} \tag{2}$$

If $x\in(0,b)$, the function $f_{\tilde{A}}(x)$ monotonically increases in the interval $(0,b)$, and $\lim\limits_{x\to 0^+} f_{\tilde{A}}(x)=0$. If $x\in(c,+\infty)$, the function $f_{\tilde{A}}(x)$ monotonically decreases in the interval $(c,+\infty)$, and $\lim\limits_{x\to+\infty} f_{\tilde{A}}(x)=0$.

The exponential whitening function can effectively overcome the limitation of the linear whitening function in the traditional gray clustering method, considering the defects between adjacent grades only, broaden the coverage of the whitening function and greatly improve the utilization of the available information.

2.2.2 Clustering weights based on entropy

It is essential to analyze the relative importance of each index in the flood risk evaluation index system. In previous research, however, flood risk evaluation index systems lacked effective methods to calculate the weights of indexes, making flood risk assessment difficult. The most commonly used method to determine index weighting is the analytic hierarchy process (AHP). However, the analytic hierarchy process is a highly subjective method that compares the indexes to each other and subjectively determines the relative importance by the judgment of individuals. The information entropy method can overcome the limitations of the subjective imprecision of the traditional AHP method, use the limited data in the weighting process and better reflect the natural conditions in the evaluation results.

The concept of entropy is derived from thermodynamics, which is used to describe the irreversible phenomena of the movement of ions or molecules. Later, entropy was introduced into information theory by Shannon (Shannon, 1948) to measure the uncertainty, stability, and quantity of information in the system. In the actual evaluation or decision-making process, the role of the indicators and the amount of information transmitted vary. When the value of an object in a certain index is greater, the entropy value is lesser, which shows that more effective information is provided by this index; the greater the role of the comprehensive evaluation is, the greater the weight should be, and vice versa.

Anentropy-based clustering weight can avoid subjective influence, produce more reasonable and objective weights, and ensure standardization and accuracy of the flash flood evaluation and zonation. The process of information entropy is as follows.

Suppose the sample set is $\{d_1, \ d_2, \cdots, d_m\}$ and every sample has n indicators. The sample indicator matrix is D

$$D = \begin{pmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ d_{m1} & d_{m2} & \cdots & d_{mn} \end{pmatrix} \tag{3}$$

where x_{ij} is the jth indicator of sample i, i=1,2,…m, and j=1,2,…n.

Because the entropy variables range from 0 to 1, we need to address the original sample value.In this paper, the normalization method used is as follows:

$$u_{ij} = \frac{d_{ij}}{\sum_{i=1}^{m} d_{ij}} \tag{4}$$

For a j-th index in the system, the information entropy is as follows:

$$e_j = -k \sum_{i=1}^{m} u_{ij} \ln u_{ij} \tag{5}$$

where $k = 1/\ln n$ and $0 < e < 1$.

We can define the clustering weights of the j-th index as follows:

$$w_j = (1 - e_j)(m - \sum_{j=1}^{n} e_j) \tag{6}$$

Combined with the entropy weight method, the Eq. to calculate the gray clustering coefficient can be rewritten as follows:

$$\sigma_{ik} = \sum_{j=1}^{n} f_{jk}(d_{ij}) w_j \tag{7}$$

where $f_{jk}(d_{ij})$ is the whitening function of the j-th index and belongs to grade k.

III. FLASH FLOOD RISK EVALUATION MODELING FRAMEWORK

Based on abovementioned comprehensive method that includes entropy weighting and an exponential whitening function, the flash flood risk evaluation methods are conducted in the following steps. First, a reasonable evaluation index system is established. Second, the comprehensive weights of different indicators are calculated by the entropy weighting method. Finally, gray clustering analysis is carried out with the improved whitening function to achieve reasonable risk evaluations of the regions that have potential flash flood risks.

In this research, flood risk, hazard and vulnerability are determined by the so-called ‘‘whiten-weight function” and described by the following categories: very low, low, medium, high, and very high. Here, the exponential whitening functions are used, which can be represented by Eq. (2).

Detailed steps are described below.

Step 1: We calculate the improved whitenization weight function $x_{ie}^{(s)}$ for the indexes $x_i (\mathrm{i} = 1, 2, 3, ..., n)$ belonging to the gray cluster e $(e = 1, 2, 3, \cdots, k)$. According to Table 1, whitenization weight functions are defined by boundary parameters, as shown in Eq. (2).

Then, let

$$x_i^{(s)} = \sum_{e=1}^{k} x_{ie}^{(s)} \tag{8}$$

For the data s , its evaluation indexes are $x_i (\mathrm{i} = 1, 2, 3, ..., n)$, and the gray evaluation coefficient of the data s belonging to the gray clusters e ($e = 1, 2, \cdots, k$) is recorded as $r_{ie}^{(s)}$. Then,

$$r_{ie}^{(s)} = \frac{x_{ie}^{(s)}}{x_i^{(s)}} \tag{9}$$

The evaluation weight matrix of the data s for different evaluation gray clusters can be determined by combining the gray evaluation weight vectors of all indexes of data s :

$$R^{(s)} = \begin{pmatrix} r_{1,1}^{(s)} & r_{1,2}^{(s)} & \ldots & r_{1,k}^{(s)} \\ r_{2,1}^{(s)} & r_{2,2}^{(s)} & \ldots & r_{2,k}^{(s)} \\ \ldots & \ldots & \ldots & \ldots \\ r_{n,1}^{(s)} & r_{n,2}^{(s)} & \ldots & r_{n,k}^{(s)} \end{pmatrix} \tag{10}$$

Step 2: We calculate the weight vector of the indicators U_R with the information entropy method with Eq. (6).

Step 3: To conduct gray clustering, the comprehensive assessment model of the data s is calculated as $B^{(s)} = U_R \cdot R^{(s)} = (\mathrm{b}_1^{(s)}, \mathrm{b}_2^{(s)}, ..., \mathrm{b}_k^{(s)})$ (11), or Eq.(7), where U_R is the importance weight vector of the indicators. If $b_k = \max\{\mathrm{b}_1^{(s)}, \mathrm{b}_2^{(s)}, ..., \mathrm{b}_k^{(s)}\}$ (12), then the data s belongs to the cluster k , and all data classifications are obtained.

IV. CASE STUDY

4.1 Study area

Hubei Province is located in the middle reaches of the Yangtze River. Hubei Province has a diverse range of land forms including mountains, hills and plains. Mountainous areas account for approximately 55.5% of the total area of the province; hillocks, 24.5%; and the plains and lake areas, 20%. The province covers an area of 185,900 km^{2}and has a population of 58,160 thousand people. The province is surrounded by mountains to the east, west, and north, and the middle area is primarily a flat, incomplete basin with a southern opening. The altitude of this area is generally 20~100 meters. The average precipitation is 800~1,600 millimeters with significant seasonal differences; most precipitation occurs in summer, and less precipitation occurs in winter. In Hubei Province, flash floods were recorded in 15 of the last 100 years. In the 1960s, 1980s and 1990s, this area had several flood currents and debris flows, resulting in heavy casualties and property loss. Mountain flash floods seriously threaten the safety of the people in mountainous areas across Hubei Province.

4.2 Data and indexes

Risk reflects the expected number of deaths and injuries as well as the severity of property damage and economic activity disruption due to a particular natural phenomenon, generally defined as the product of the hazard probability and its consequences. Risk can be viewed as a function of hazard and vulnerability (Maskrey, 1989).

Thus,

$$\text{RISK=Vulnerability + Hazard} \tag{13}$$

where vulnerability is the probability of any physical, structural or socioeconomic element to a natural hazard being damaged, destroyed or lost. Vulnerability is not static but must be considered as a dynamic process, integrating changes and developments that alter and affect the probability of loss and damage of all the exposed elements.

Hazard is the probability that in a given period in a given area, an extreme potentially damaging natural phenomena occurs that induces air, earth or water movements, which affect a given zone.

Risk is the estimate of the total expected losses for a given area, and specific risk is the expected degree of loss to a given category of elements at risk as a function of hazard and vulnerability. Risk can be related directly to the concept of disaster, given that it includes the total losses and damages that can be suffered after a natural hazard: dead and injured people, damage to property and interruption of activities. Risk implies a future potential condition, a function of the magnitude of the natural hazard and of the vulnerability of all the exposed elements in a determined moment.

From the above risk model in Eq. (13), the factors that affect the flood risk can be classified into the following 2 categories: hazard factors and vulnerability factors.

Although a significant database was generated on hazard and vulnerability parameters of Hubei Province, only 15 of the parameters were used by the statistical correlation test, given that we cannot use two indicators that are strongly correlated in the same model.

A. Hazard factors

Selection of the flood hazard evaluation index is based on the flash flood analysis, considering disaster environment and disaster bearing body, combined with the index characteristics and the existing data. He BY (2002) analyzed 4 factors: precipitation, topography, river network density and historic flood frequency. Tang C and Zhu J (2005) considered 6 main factors: terrain, river network density, rainstorm occurrence, modulus of flood peak, debris flow density and comprehensive disaster degree. Jiang WG (2008) selected 7 factors as the hazard evaluation indexes: maximum 3-day rainfall, rainstorm frequency, vegetation coverage, rivernet density, standard deviation of the elevation per unit area, and old and young populations per unit area.

There are many factors that affect flooding. However, the two main factors are precipitation and underlying surface. The precipitation is determined by meteorological conditions. The precipitation characteristics including total precipitation, rainfall duration and intensity are the major factors causing hazard disaster; underlying factors including location, landform, vegetation, other environmental conditions and human activities are indirect hazard factors (Fan et al., 2008).

The topographic factor is also a parameter or index that is meaningful for the study and expression of geomorphic features. Topography is closely related to the degree of flood risk. The influence of topography on flood formation is mainly manifested in two aspects: relative elevation and terrain slope. Therefore, this study uses relative elevation and slope as topographic indicators to reflect the severity of the mountain floods, and this study uses ArcGIS software to extract the distribution of geomorphic features in Hubei Province from DEM data.

In the present study, based on the obtained data, the following parameters are selected as mountain flash flood hazard assessment indexes: precipitation, critical precipitation, water level, average gradient, relative elevation difference, river network density, and vegetation coverage.

B. Vulnerability factors.

The concept of vulnerability has been continuously widened and broadened towards a more comprehensive approach encompassing susceptibility, exposure, coping capacity and adaptive capacity, such as physical, social, economic, environmental and institutional vulnerability. Although a significant database was generated for the vulnerability parameters, only 8 of them were used.

In this paper, the vulnerability assessment indexes are divided into two categories; one category is related to natural disasters, and the other is related to socio-economic indicators. Disaster-related indicators include the current flood prevention capability, direct economic loss caused by earlier floods, and number of deaths and missing people caused by earlier floods. Socio-economic indicators include the proportion of the agricultural areas, population density, and residential housing property values in flood-controlled areas, total agricultural output and local financial revenue.

After obtaining the hazard and vulnerability, risk assessment is an easy task. According to the risk definition, risk can be determined by the following Eq., which is derived from Eq. (13).

$$R=0.8H+0.2\ V \tag{14}$$

where R is the risk assessment of a mountain flood disaster, H is the hazard assessment of a mountain flood disaster, and V is the vulnerability assessment of a mountain flood disaster.

4.3 Example

Based on this database, we selected 10 towns in Hubei Province to describe the proposed method. First, the weight coefficient of the risk evaluation indexes was obtained by the entropy method with the data of hazard indexes and vulnerability indexes from 10 towns in Hubei Province, as shown in Tables 1 and 2; these towns were selected as the research object of the flash flood disaster regions.

TABLE 1
HAZARD INDEX DATA OF 10 TOWNS

Town	Precipitation (mm)	Critical rainfall (mm)	Water level(m)	Average gradient(°)	Relative height difference(m)	River network density (km/km^2)	Vegetation coverage (%)
Longgang Town	84.66	126.90	1.84	18.18	431	0.88	80.41
Xiangkou County	61.01	72.90	0.88	28.56	610	0.86	95.28
Fengxi County	122.44	75.94	0.41	39.22	1619	0.62	93.31
Leizu Town	204.39	193.60	1.36	32.43	679	0.54	95.22
Wantan Town	105.63	178.10	34.19	20.80	1312	0.58	98.43
Chebu Town	105.88	217.25	-0.29	3.32	33.00	0.51	76.97
Sanli County	84.33	125.75	1.29	19.17	719.00	0.54	95.25
Qingtaiping Town	98.46	124.08	-3.57	23.23	982.00	0.49	95.89
Shadaogou Town	115.84	132.21	-0.05	32.23	1193.00	0.59	94.84
Rongmei Town	136.18	150.00	1.64	30.33	1205.00	0.58	97.39

TABLE 2
VULNERABILITY INDEX DATA OF 10 TOWNS

Town	Current flood prevention capability (year)	Proportion of agricultural area (%)	Population density (people/km2)	Direct economic loss by historical flash floods (10,000 RMB)	Deaths and missing in historical flash floods (people)	Residential housing property in prevention area (10,000 RMB)	Total agricultural output (tons)	Local financial revenue (10,000 RMB)
Longgang Town	1.02	0.18	301.61	120000	7	12.59	31069.86	4890.50
Xiangkou County	2.89	0.02	53.28	39966	11	21.78	14990.80	1374.00
Fengxi County	2.91	0.04	29.48	2000	16	11.24	18062.50	2375.00
Leizu Town	7.15	0.07	70.81	3100	5	28.56	17153.29	16714.29
Wantan Town	3.99	0.09	53.82	2497	45	180.63	17441.71	4157.14
Chebu Town	3.14	0.24	299.34	4950.00	0	11.79	27450.33	17193.50
Sanli County	5.80	0.26	254.61	159.00	2	43.21	38012.70	15926.00
Qingtaiping Town	31.99	0.14	144.64	8000.00	8	62.42	26683.83	4519.92
Shadaogou Town	3.04	0.07	101.33	303.00	0	25.89	24048.00	3232.00
Rongmei Town	16.32	0.08	133.95	112.00	0	40.36	19733.57	30286.29

In the process of gray clustering, the hazard and vulnerability assessments were divided into five classifications of standard risk level: I-grade, II-grade, III-grade, IV-grade and V-grade. These grades represents very high risk, high risk, moderate risk, low risk and very low risk levels, respectively, as shown in Table 3.

TABLE 3
RISK LEVEL STANDARDS

Evaluation indexes		I-grade	II-grade	III-grade	IV-grade	V-grade
Hazard	Precipitation (mm)	<10	10~70	70~00	100~150	>150
	Critical precipitation (mm)	>171.2	149.8~171.2	128.5~149.8	107.1~128.5	<107.1
	Water level (m)	<-1.4	-1.4~-0.5	-0.5~0.1	0.1~1.4	>1.4
	Average gradient (°)	<5	5~10	10~30	30~50	>50
	Relative height difference (m)	<100	100~200	200~500	500~1000	>2000
	River network density (km/km^2)	<0.5	0.5~0.7	0.7~0.9	0.9~2	>2
	Vegetation coverage (%)	>95	80~95	60~80	40~60	<40
Vulnerability	Current flood prevention capability (year)	>100	10~100	5~10	2~5	<2
	Proportion of agricultural area (%)	<0.2	0.2~0.3	0.3~0.5	0.5~0.8	>0.8
	Population density (people/km2)	0~150	150~300	300~420	420~500	>500
	Direct economic loss by historical flash floods (10,000 RMB)	<100	100~200	200~300	300~1000	>1000
	Deaths and missing in historical flash floods (people)	<2	2~5	5~10	10~30	>30
	Residential housing property in prevention area (10,000 RMB)	<10.2	10.2~16.1	16.1~51.2	51.2~80.46	>80.46
	Total agricultural output (tons)	<10000	10000~20000	20000~30000	30000~50000	>50000
	Local financial revenue (10,000 RMB)	<3000	3000~5000	5000~10000	10000~20000	>20000

Based on the categorization above (Table 3), the exponential whitening function for every index was determined. For example, the exponential whitening functions for precipitation belonging to grades I, II, III, IV and V are established as follows:

$$f_1 = \begin{cases} 1 & x \in [0,10] \\ e^{\frac{10-x}{10}} & x \in (10,+\infty) \end{cases} \qquad f_2 = \begin{cases} e^{\frac{x-10}{10}} & x \in [0,10] \\ 1 & x \in (10,70] \\ e^{\frac{70-x}{70}} & x \in (70,+\infty) \end{cases} \qquad f_3 = \begin{cases} e^{\frac{x-70}{70}} & x \in [0,70] \\ 1 & x \in (70,100] \\ e^{\frac{100-x}{100}} & x \in (100,+\infty) \end{cases} \qquad f_4 = \begin{cases} e^{\frac{x-100}{100}} & x \in [0,100] \\ 1 & x \in (100,150] \\ e^{\frac{150-x}{150}} & x \in (150,+\infty) \end{cases}$$

$$f_5 = \begin{cases} 1 & x \in (150,+\infty) \\ \frac{x-125}{125} & x \in [0,150] \end{cases}$$

Based on the Hubei Province data in Tables 1 and 2, flash flood disaster risks in Hubei Province were evaluated according to Eq.s (8)-(12), as shown in Table 4.

TABLE 4
RISK EVALUATION RESULTS

Town	Entropy-improved gray clustering
Longgang Town	III
Xiangkou County	IV
Fengxi County	V
Leizu Town	IV
Wantan Town	II
Chebu Town	I
Sanli County	IV
Qingtaiping Town	III
Shadaogou Town	V
Rongmei Town	IV

V. RESULTS AND DISCUSSION

As an example, we calculated the degrees of risk for 10 towns in Hubei Province to compare the improved gray clustering method with the traditional gray clustering method (Liu, 2010) and fuzzy number-gray clustering method, as shown in Table 5.

TABLE 5
COMPARISON OF THE RESULTS FROM THREE METHODS

Town	Improved gray clustering	Traditional gray clustering	Fuzzynumber-gray clustering
Longgang Town	III	III	III
Xiangkou County	IV	IV	IV
Fengxi County	V	V	V
Leizu Town	IV	V	IV
Wantan Town	II	I	II
Chebu Town	I	I	I
Sanli County	IV	IV	IV
Qingtaiping Town	III	IV	III
Shadaogou Town	V	IV	V
Rongmei Town	IV	IV	IV

Using these three methods, the risk evaluation of mountain flood disaster areas in 10 towns across 10 counties is provided in Table 5. It can be seen from Table 5 that the results of the evaluation of four towns are slightly different by using the traditional gray clustering method, compared to the other two methods. This is because the range of the linear whitening function is too narrow in the traditional gray clustering method, and when the distribution of the monitoring value is discrete, it loses a considerable amount of useful information; for example, this is true for Leizu Town and Wantan Town. For all these towns, the entropy-improved gray clustering method produces risk levels that are more consistent with the actual counts of deaths and missing people as well as the direct economic loss in historical flash floods. It can be seen that the improved gray clustering method makes full use of the limited information available, more effectively reflecting the risk level of the region. The results show that the improved evaluation method produces results that are more objective and accurate.

Additionally, the Fuzzy number-gray clustering analysis and evaluation method is also used to calculate the results. The results of these two methods are analyzed, as shown in Table 5. The results in Table 5 show that the results of the two evaluation methods are both consistent with the actual disaster loss documented. The results show that the calculation results based on the improved gray clustering method can reflect the degree of the mountain flood risk more effectively and have high reliability in the field of flood disaster risk assessment and zonation.

It can be seen from Table 5 that the two methods produce identical risk levels for 10 towns in flash flood regions. The resultant risk levels of all towns basically agree with the actual number of deaths and missing people and the direct economic loss in historical flash floods. Although the number of deaths and missing people in historical flash floods in Shadaogou Town and Rongmei Town is 9, the precipitation, critical precipitation and current flood prevention capacity are weaker than other towns, significantly influencing their risk levels. In other words, the above evaluation results are reasonable and reliable for risk zonation.

In this paper, a grid-based flash flood risk zonation of Hubei Province was accomplished by using hydrometeorology, landform and socioeconomic characteristics data and the spatial data extraction, sampling, interpolation and analysis functions in ArcGIS. The distribution of risk grading was obtained from the risk grades of 1436 towns and counties in Hubei Province. Furthermore, a flash flood risk zonation map of Hubei Province was compiled (Fig. 3)

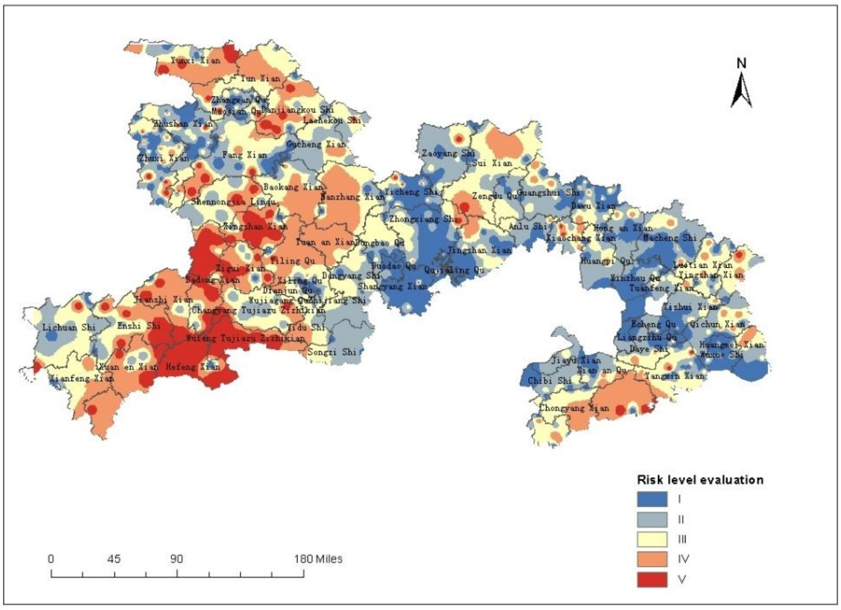

FIG. 3. FLASH FLOOD RISK ZONATION OF HUBEI PROVINCE

Furthermore, with the same method, we produced the maps of hazard and vulnerability across Hubei Province to investigate their spatial distribution, as shown in Fig. 4 and Fig. 5, respectively. The flash flood risk map identified the areas of the highest risk and the areas highly vulnerable to flash flood.

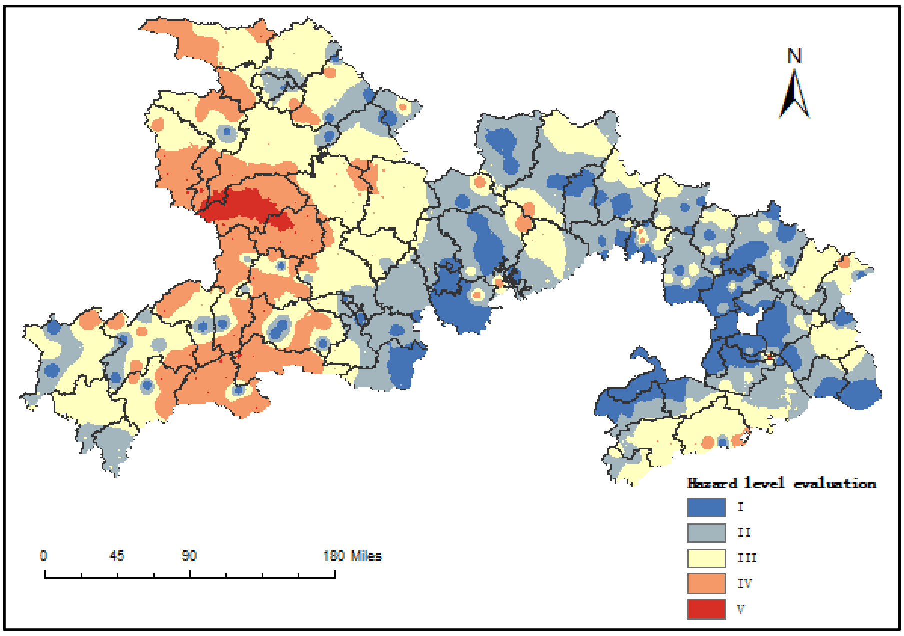

FIG. 4. FLASH FLOOD HAZARD ZONATION OF HUBEI PROVINCE

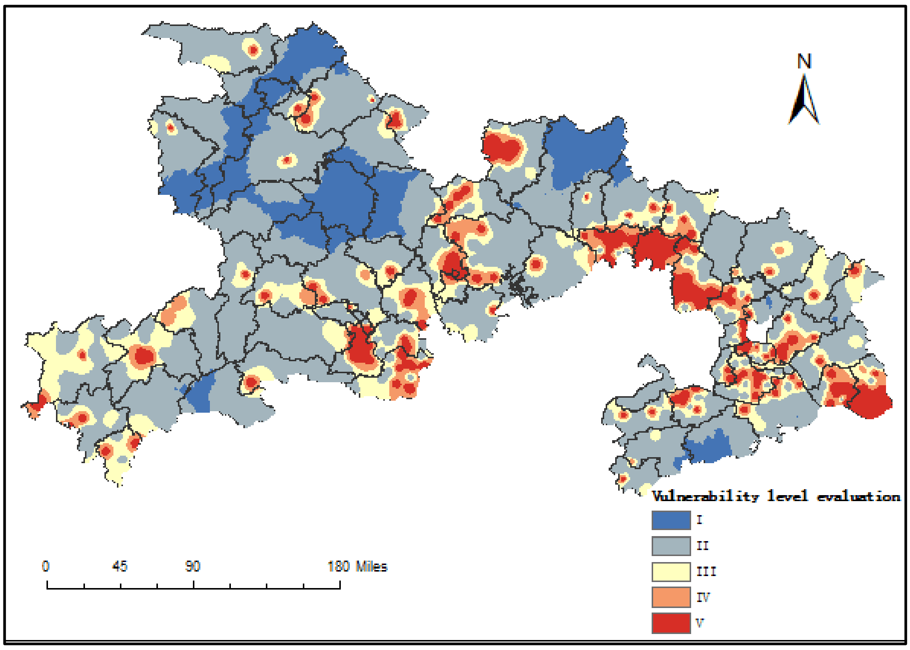

FIG. 5. FLASH FLOOD VULNERABILITY ZONATION OF HUBEI PROVINCE

VI. CONCLUSIONS

In the risk assessment of mountain floods in Hubei Province, the villages and towns are taken as the evaluation units in the study area, and the geometric calculation function is used to calculate the area proportion of different risk grades in ArcGIS, as shown in Table 6.

TABLE 6
AREA PROPORTION OF DIFFERENT RISK GRADES

risk grade	area（km^2）	area proportion（%）
Very low	43880.75	23.23%
low	16089.42	8.52%
medium	60416.60	31.98%
high	36401.06	19.27%
Very high	32123.65	17.00%

Based on the hazard map in Fig. 4, some of the southeastern regions (such as Luotian County, Xishui County, Tuanfeng County, Daye City, Tongcheng County, and Tongshan County), western regions (such as Yunxi County, Yun County, Danjiangkou City, Gucheng County, Enshi and other areas) and northern regions (such as Zaoyang City and Suizhou County) of Hubei Province pose the highest degree of flash flood hazard.

However, based on the hazard map in Fig. 5, some of the southern regions (such as Zhijiang City, Yidu City, Songzi City and Wujiagang District), northern regions (such as Laohekou City and Danjiangkou City) and eastern regions (such as Anlu City, Huangpi District, Xinzhou District, Huarong District and Jingshan County) of Hubei Province are most vulnerable to flash flood.

Based on the risk map that considers both the hazard and the vulnerability, shown in Fig. 3, Zouma Town, Sha Road Town, Five Peak Town, Xintang County, Flower County and 209 other towns and counties are exposed to the highest flash flood risk level in Hubei Province. Among the highest risk areas, Ma Zhen occupies the largest area, 1417.61 km^2, and accounts for 4.5% of the total very high-risk area.

Among 1436 towns and counties in Hubei Province, 196 towns are exposed to the low and very low flash flood risk. Juwan Town has the highest proportion of area in the low and very low flash flood risk. There are 407 towns, such as Tianjia Town, with a moderate flash flood risk, 624 towns with a high flash flood risk, and 209 towns with a very high flash flood risk.

The method presented in this paper improves the traditional gray clustering method with the following changes: constructing an exponential whitening function, effectively overcoming the shortcomings of the traditional gray clustering method that considers the defects between only adjacent grades in the linear whitening function, broadening the coverage of the whitening function and greatly improving the utilization of the available information. Additionally, the entropy-based method for determining the clustering weight can avoid subjective influences, such as the AHP method, and produce more reasonable and objective weights and evaluations. Lastly, this research is useful for identifying the regions that are threatened by the highest risk and can be easily applied to flash flood zonation for disaster assessments.

In this paper, the issues of the existing gray clustering method are analyzed. To overcome the shortcoming of the existing gray clustering method and the method for determining weights, an improved gray clustering method that includes entropy is proposed. This improved method is used in a case study. The improved gray clustering method makes better use of characteristic information from a research database, and, compared to other methods, the improved gray clustering calculates risk levels of evaluation units more accurately and quickly. The results demonstrate that this method is simple and feasible, and the result is reasonable and accurate. It is reasonable to apply this method to the risk assessment and zonation of mountain flash floods and other disasters.

The results show that the improved approach is useful in rapidly assessing flash flood hazard and vulnerability as well as the risk assessment in mountain areas and could be adopted, with appropriate modification, elsewhere in areas with flash flooding.

Flood disaster risk assessment is very important for disaster prevention, decision-making and management, since reasonable planning and management in flood-prone areas not only reduces the flash floods loss and guarantees the safety of human lives in hilly areas but also provides disaster risk precaution information for local residents and promotes the sustained and stable development of a social economy. Flood disaster risk assessment helps to quickly determine the prevention level and complete reinforcement measured in dangerous areas, to greatly reduce the workload and to improve work efficiency, which is important to promote flash flood relief work.

ACKNOWLEDGMENTS

This work is supported by a Grant from Major research program of the National Natural Science Foundation of China (Project No. 91547208), a Grant from the scientific research project of Hubei Provincial Department of Education (Project No. B2017255), a Grant from Postdoctoral Science Foundation of China (2017M612460), a Grant from the National Natural Science Foundation of China (Project No. 51579107), and a Grant from the Key projects of the National Natural Science Foundation of China (Project No. 51239004).

REFERENCES

[1]. Benito G, Baker VR, Gregory KJ．Palaeohydrology and environmental change，John Wiley & Sons Inc，Chichester,1998,317-333.

[2]. Benito G, Lang M, et.al. Use of systematic, palaeoflood and historical data for the improvement of flood risk estimation. Review of scientific methods, Natural Hazards,2004.31:623-643.

[3]. Benito G, Sánchez-Moya Y, Sopeña A．Sedimentology of high-stage flood deposits of the Tagus River, Central Spain, Sedimentary Geology, 2003,157:107-132.

[4]. Borga M, Anagnostou EN, Blöschl G, Creutin JD. Flash flood forecasting, warning and risk management: the HYDRATE project. Environmental Science & Policy 14 (7), 834-844.

[5]. Chang D Y .Applications of the Extent Analysis Method on Fuzzy AHP[J] .European journal of operational research ,1996,(3):649-655.

[6]. Chen HL, Chen G, Ding GP. Evaluation of regional flood disaster risk based on GIS. Yangtze River, 2003, 34 (6):49 - 51.

[7]. Chen, S. Y. and Fu, G. T.: Combining fuzzy iteration model with dynamic programming to solve multiobjective multistage decision making problems, Fuzzy Sets Syst.2005, 152: 499-512.

[8]. Chen, S. Y. and Guo, Y.: Variable fuzzy sets and its application in comprehensive risk evaluation for flood-control engineering system, Fuzzy Optimization Decision Making, 2006,5:153-162.

[9]. Chen YN, Yang SQ. Application of grey clustering in flood disaster classification [J]. Arid region geography (Chinese version), 1999, 22 (3): 37-42.

[10]. Cheng XT, Yang L, Chen XJ, et al. Numerical model of flood routing in flood diversion and storage areas. Journal of natural disasters, 1996,5 (1):34-40.

[11]. Chubey MS, Hathout S. Integration of RADARSAT and GIS modelling for estimating future Red River flood risk,2004,59:237-246.

[12]. Deng JL, The Primary Methods of Grey System Theory, Huazhong University of Science and Technology Press, 2005.

[13]. Fan SX, Cheng YC, Gao Y. Flood design and control. Chemical Industry Press, 2008.

[14]. Fan YX, Luo Y, Chen QS. Determination of weight of comprehensive assessment index for vulnerability of regional disaster bearing body. Disaster science, 2001, 16 (1):85-87.

[15]. Haruyama, Shigeko, Ohokura et al. Geomorphological zoning for flood inundation using satellite data. GeoJournal,1996,38(3):73-278

[16]. He BY, Zhang HL, Zhang S, et.al..GIS-Based Risk Evaluation for Flood Disasterin Hubei Province. Journal of Natural Disasters, 2002,11(4):84-89 (inChinesewithEnglishabstract).

[17]. He BY, Zhang HL. Zhang Sui. Hubei province flood disaster risk assessment based on GIS. Journal of natural disasters.2002,14(4):84-89.

[18]. Huang CF. Theory and practice of natural disaster risk assessment. Beijing: Science Press, 2005.

[19]. Huang SF, Xu M, Chen DQ. Extraction of river network density supported by GIS and its application to flood risk analysis. Journal of natural disasters, 2001,10(4):129 – 132.

[20]. Huang SF. Study on the theory and method of flood risk analysis. Doctoral dissertation, Institute of geography, Chinese Academy of Sciences.1999.

[21]. Islam MM, Sado K, Development of flood hazard maps of Bangladesh using NOAA-AVHRR images with GIS, Hydrological Sciences Journal,2000,45:337-355.

[22]. Islam MM, Sado K, Development priority map for flood countermeasures by remote sensing data with geographic information system, Journal of Hydrologic Engineering,2002,7:346-355.

[23]. Islam MM, Sado K. Flood hazard assessment in Bangladesh using NOAA AVHRR data with geographical information system, Hydrological Processes,2000,14:605-620.

[24]. Jiang T, Xu PZ. Social vulnerability assessment of natural disasters in the study. China proceedings of the National Academy of Sciences, 1996,3:186-191.

[25]. Jiang WG, LJ, Li ZW, et.al. Fuzzy Assessment of the Population Risk of Flood Disaster. Journal of Hunan University (Natural Sciences), 2008, 35(9):84-87 (in Chinese with English abstract).

[26]. Jiang WG, LJ, Wu JJ,et.al. Risk assessment System for Regional Flood Disaster(Ⅱ): Model and Application. Journal of Natural Disasters, 2008,17(6):105-109(in Chinese with English abstract)

[27]. Jin JL, Wei YM, Fu Q, et al. Theoretical framework of flood disaster risk management. Water resources and hydropower technology, 2002, 33 (9):40-42.

[28]. JonkmanSN. Global perspectives on loss of human life caused by floods. Natural hazards 34 (2), 151-175.

[29]. Li BN. Weibull model for flood disaster assessment. Journal of natural disasters, 2005,14 (6):32- 36.

[30]. Li JQ, Zhang YS, Wang LP, et al. Analysis of structural characteristics of integrated flood risk. Management and environment of Changhong basin.2005,14 (6):805-809.

[31]. Li JQ, Zhang YS, Wang LP, et al. Risk structure and comprehensive evaluation method of flood consolidation platform (1) - macro aspect. Journal of Wuhan University (Engineering Science), 2005,38 (5):19-23.

[32]. Li JR, Ding ZX, Huang SF. The study on evaluation model based on spatial distribution social economic loss of flood disaster database. Institute of water conservancy and Hydropower of China.2003,1(2):104-110.

[33]. Liu L, Zhou JZ, Yang L, et al. Application of grey clustering based on entropy weight in flood evaluation [J]. Journal of natural disasters, 2010 (4): 213-218.

[34]. Liu XL, Shi PJ. Theory and Practice Research on the model of regional flood risk assessment. Journal of natural disasters.2001, (2):66-72.

[35]. Liu YH. Application Research of index weight determination method based on trapezoidal fuzzy number. Journal of Nanjing University of Information Science and Technology, 2009,1(4):369-372.

[36]. Ma JM, Zhou KY, Lu JK. Historical data quantification and analysis of Flood flood disaster risk. China Institute of water conservancy and Hydropower.1997,1(2).

[37]. Ma ZW, Xu YP, Li JJ. Fractal dimension of river morphology and its relation to flood taking the middle and lower reaches of the Yangtze River as an example. Advances in Water Sciences, 2005, 16, (4):530 – 534

[38]. Maskrey A. Disaster Mitigation: A Community Based Approach. Front Cover. Oxfam, 1989.

[39]. Mu JB, Hou KF. Application of grey clustering method in water quality assessment [J]. environmental science, 1991 (2): 86-89.

[40]. Qin NX, Jiang T. Division and evaluation in the lower reaches of the Yangtze River flood risk based on GIS. Journal of natural disasters,2005,4(5):1-7.

[41]. Sanyal J, Lu XX. Remote sensing and GIS-based flood vulnerability assessment of human settlements: a case study of Gangetic West Bengal, India. Hydrological Processes. 2005,19:3699-3716.

[42]. Sanyal J, Lu XX. GIS-based flood hazard mapping at different administrative scales: A case study in Gangetic West Bengal, India. Singapore Journal of Tropical Geography.2006,27:207-220.

[43]. Satty T L .The Analytic Hierarchy Process[M] .New York:McGraw-Hill ,1980.

[44]. Shannon C E. A m athem atical theory of comm un icat ion[J] . B ell System T echn ical Jou rn a,l 1948, 27 (3) : 379 - 423.

[45]. Shi PJ. Atlas of Chinese natural disaster. Beijing: Science Press, 2003.

[46]. Smemoe CM.Floodplain risk analysis using flood probability and annual exceedance probability maps.Brigham Young University,2004.

[47]. Solaimani K, Mohammadi H, Ahmadi M Z.Habibnejad M. Flood occurrence hazard forecasting based on geographical information system． Int J Environ Sci Tech,2005,2(3):253-258．

[48]. Su BD, Jiang T, Guo YY, et al. A flood risk dynamic simulation model based on GIS raster data and its application. Journal of Hohai University,2005,33(4):370-374.

[49]. Tang C, Shi YE, A study on multi target evaluation method for urban flash flood. Advances in geographical science,2006,25 (4):13-21.

[50]. Tang C, Zhu J. A GIS Based Regional Torrent Risk Zonation. Acta Geographica Sinica, 2005, 60(1): 87-94 (in Chinese with English abstract).

[51]. Tang C, Zhu J. Risk zoning of mountain torrents disaster based on GIS. Journal of geography. 2005, 60 (1):87-94.

[52]. Wei YM, Zhang LP, Fan Y. Simulation study of flood disaster evolution based on Swarm. Journal of management science,2002,5 (6):39-46.

[53]. Xu JT. PZ. XG. Flood vulnerability concept model. China mitigation.1997,7(2):24-29.

[54]. Xu ZK, Huang HP, Wei BW, et al. A comprehensive evaluation method for flood risk based on systematic multilevel grey model -- a case study of Poyang Lake River Basin [J]. south to North Water Transfer and water conservancy science and technology, 2015, 13 (1): 20-23.

[55]. Yang PG, Dai EF, Wu SH, et al. Spatial pattern of flood risk in the downstream levee reserve of the Yellow River. Science Bulletin (Supplement). 2006,51:148-154.

[56]. Yang XC, Zhu XH. Study on fractal dimension and relation of flood and flood in the seven watersheds of China. Disaster science, 2002,17(3):9-13.

[57]. Zadeh, L. A.: Fuzzy sets, Inf. Control, 8, 338–353, 1965.

[58]. Zhan XG. Zhu GR, Wen YY. Study on flood risk assessment of plain area using GIS method: 6 counties and cities in Jingzhou are inverted. Resources and environment of the Yangtze River Basin, 2003.12 (4):388-392.

[59]. Zhang C, Wan Q, Zhang JQ, et al. Flood risk assessment method based on grid data-a case study of new flood in Japan. Geo information science, 2003,(4):69-73.

[60]. Zhou CH, Wan Q. Huang SF, et al. Study on Flood Disaster Risk Regionalization Based on GIS. Journal of geographical science.2000,55(1):15 - 24.