# Research on intelligent prediction and zonation of basin-scale flood risk based on LSTM method

**Minzhi Yang · Ping-an Zhong · Jieyu Li · Weifeng Liu · Yinghui Li · Kun Yan · Yinyang Yuan · Yihui Gao**

**Abstract** Global climate change and human activities aggravate the frequency of flood disasters. Flood risk includes natural flood risk and risk of economic and social disasters, which is displayed intuitively by flood risk zonation maps. In this paper, we take the disaster-causing factors, the disaster environment, the disaster-bearing body, and the disaster prevention and mitigation capability into consideration comprehensively. Eleven influencing indexes including annual maximum 3-day rainfall and rainfall in flood season are selected, and the virtual sown area of crops is innovated. Taking the Huaihe River Basin (HRB) as the research area, the flood risk prediction of the basin is explored by using the long short-term memory (LSTM). The results show that LSTM can be successfully applied to flood risk prediction. The short-term prediction results of the model are good, and the area where the risk is seriously underestimated (the high and very high risk are identified as the very low risk) accounts for only 0.98% of the total basin on average. The prediction results can be used as a reference for watershed management organizations, so as to guide future flood disaster prevention.

M. Yang · P.-a. Zhong · J. Li · W. Liu
College of Hydrology and Water Resources, Hohai University, No. 1, Xikang Road, Nanjing 210098, China

P.-a. Zhong (✉)
National Engineering Research Center of Water Resources Efficient Utilization and Engineering Safety, Hohai University, No. 1, Xikang Road, Nanjing 210098, China
e-mail: zpa_hhu@163.com

Y. Li
College of Agriculture and Water Conservancy Engineering, Suihua University, Yellowriver Road, Suihua 152061, China

K. Yan · Y. Gao
Zhejiang Institute of Hydraulics & Estuary, No. 50, Fengqi East Road, Hangzhou 310020, China

Y. Yuan
Chengdu Water Authority, No. 68, Shujin Road, Chengdu 610042, China

## Introduction

Flood disaster is a type of natural disaster with sudden occurrence, huge destructiveness, and high frequency (Hall et al. 2005; Budiyono et al. 2015; Foudi et al. 2015; Yin et al. 2015; Khosravi et al. 2016; Lyu et al. 2018). In recent years, under the dual influence of climate change and human activities, its occurrence frequency, damage range, and disaster loss have increased in varying degrees (Apel et al. 2004; Zhou et al. 2012; Chen et al. 2013). Flood risk reflects the combination of disaster possibility and consequences in a certain region or basin, including not only natural aspects but also economic and social factors. Understanding the temporal and spatial characteristics of flood risk is an important prerequisite to guide current and future flood control and disaster reduction work. Flood risk assessment is an effective means to reveal the

complex relationship between risk influencing factors and risk levels. Flood risk zonation is the quantification of the flood risk, which can intuitively display the risk distribution, provide accurate temporal and spatial information of the risk, and is widely used in the field of the flood risk. It is also the development trend of visual monitoring and management of flood control in the future (Li et al. 2013; Zou et al. 2013; Castillo-Rodriguez et al. 2014).

Because the flood disaster is affected by many factors including nature, economy and society, and engineering, it is involved in the complex process of causing, gestating, and bearing disaster. Flood risk assessment has become one of the hotspots and difficulties in international water conservancy and disaster circles. In the assessment process, a large amount of real-time data is needed for flood inundation simulation and the instant flood risk assessment based on the water dynamics (Bonn and Dixon 2005; Badrzadeh et al. 2015). The one-dimensional or two-dimensional water dynamic model used for modeling is not a macro consideration for a large river basin and is generally applicable to a small watershed on the contrary, so is therefore not among the ranges discussed herein. Only the macro-scale annual flood risk is discussed in this paper. At present, there are two kinds of mainstream evaluation methods, among which the construction of an evaluation index system based on risk components from the perspective of causes is relatively mature, and the research achievements are quite fruitful. For example, the risk evaluation model is constructed from the angle of hazard and vulnerability (Wu et al. 2015; Wu et al. 2017; Peng 2018), or from three aspects of disaster-causing factors, disaster environment, and the disaster-bearing body (Liu et al. 2008; Liu et al. 2009). The weight of index also changes from using a single subjective or objective weight to constructing a comprehensive weight, thus balancing the multiple effects of the subjective and objective factors, and the spatial heterogeneity of decision makers' attitudes towards risk preference is also considered (Xiao et al. 2017). There are many factors involved in this method which have strong regional differences, and most of them call for expert decision-making. For example, four hydrologists, four engineers, and eight end-users are required to score repeatedly in a paper (Ouma and Tateishi 2014), which consumes a lot of manpower and material resources and makes the calculation more complex.

With the rapid development of artificial intelligence and machine learning techniques, flood risk evaluation based on intelligent algorithms is gradually entering the field of hydrology and risk management and provides a new perspective and a new methodology for flood risk evaluation. Some intelligent algorithms have been applied in specific watersheds or regions. Among them, random forests showed strong ability in flood risk assessment of Dongjiang River Basin (Wang et al. 2015), and the effect of risk classification proved to be reliable. Deng (2013) successfully applied feature weighted support vector machine based on improved parameter optimization for flood risk assessment. In order to realize the function of a comprehensive evaluation of the multi-dimensional flood disaster index in one-dimensional continuous space, Huang et al. (2010) established an optimized support vector machine model suitable for flood disaster level evaluation. The results showed that the model had strong generalization ability. In addition, decision tree (Kubal et al. 2009; Tingsanchali and Karim 2010), Bayesian network (Li et al. 2010), BP neural network (Lai et al. 2011), and other methods (Wang et al. 2013; Lai et al. 2015) have been used to evaluate flooding. However, most of these intelligent algorithms do not consider the time correlation of dynamic time series as well as revealing the evolution process of flood risk from the mechanism scale, and it is difficult to deeply excavate the inherent laws contained in dynamic time series. Most studies only evaluate the risk that has occurred in the past and mostly simulate the multi-year average flood risk of a basin or a region, and fail to give guidance for the future flood risk and its trend.

Flood risk prediction can be regarded as a typical time-series prediction problem with both nonlinear and nonstationary characteristics, which increases the difficulty of predicting flood risk. The LSTM proposed in 1997 has the concept of time series (Hochreiter and Schmidhuber 1997). The long-term memory ability enables it to study deeply and mine the potential laws between the data more deeply, which improves the accuracy and reliability of the prediction results. It has been applied to the fields of finance, acoustics, and medicine (Graves and Schmidhuber 2005; Fischer and Krauss 2018), among many other fields.

In view of the complex nonlinear relationship between flood risk evaluation index and flood risk level, the temporal correlation of flood risk itself, the large number of indexes, and the large amount of data, LSTM is introduced to carry out empirical research in the HRB

(schematic in Fig. 1). An intelligent identification and prediction model of flood risk of the basin based on LSTM is constructed, and the intelligent prediction and zonation of the future flood risk of the basin is studied, which provides an efficient, convenient, and accurate way for flood risk evaluation, and the prediction results can provide valuable reference for watershed management institutions as well as decision makers.

## Study area

The Huaihe River Basin (HRB, 30° 55′~36° 36′ N, 111° 55′~121 °25′ E) straddles four provinces of Henan, Jiangsu, Shandong, and Anhui (Fig. 2). It is located in the overlapping regions of three transitional climates, covering mid to low latitudes and sea to land and is the third largest watershed in China. The river basin is densely populated; the average population density is the highest of the seven major Chinese basins. The cultivated land is vast and composed of grain, cotton, and oil base in China.

Due to its unique geographical position, the typical drought and waterlogging characteristics (drought will occur without rainfall, waterlogging will occur together with rainfall, and flood will occur together with heavy rainfall) of this basin are formed. The rainfall from June to August accounts for more than half of the annual rainfall of the basin, and once floods occurred, the middle reaches are prone to waterlogging due to the flat terrain and the poor flood discharge. The floods can be compound because of the numerous tributaries and few entrances to the sea. According to the historical records of 522 years from 1470 to 1991, the flood occurred every 3 years on average in the HRB. With the rapid development of global climate change and urbanization, it has brought more arduous challenges to the current and future flood control and disaster reduction in HRB (Yang and Li 2003; Yang et al. 2012).

## Data

### Index selection

We consider the disaster-causing factors, disaster environment, disaster-bearing body, and disaster prevention and mitigation capability. We end up selecting 11 evaluation indexes. The index system is shown in Fig. 3.

A comprehensive flood risk assessment should not only take into account the risk caused by flood transit in flood season and untimely drainage of local heavy rainfall but also include damage to people's lives and property. Therefore, in order to comprehensively describe the flood risk of the basin, we selected in the disaster-causing factors the rainfall in the flood season and the annual
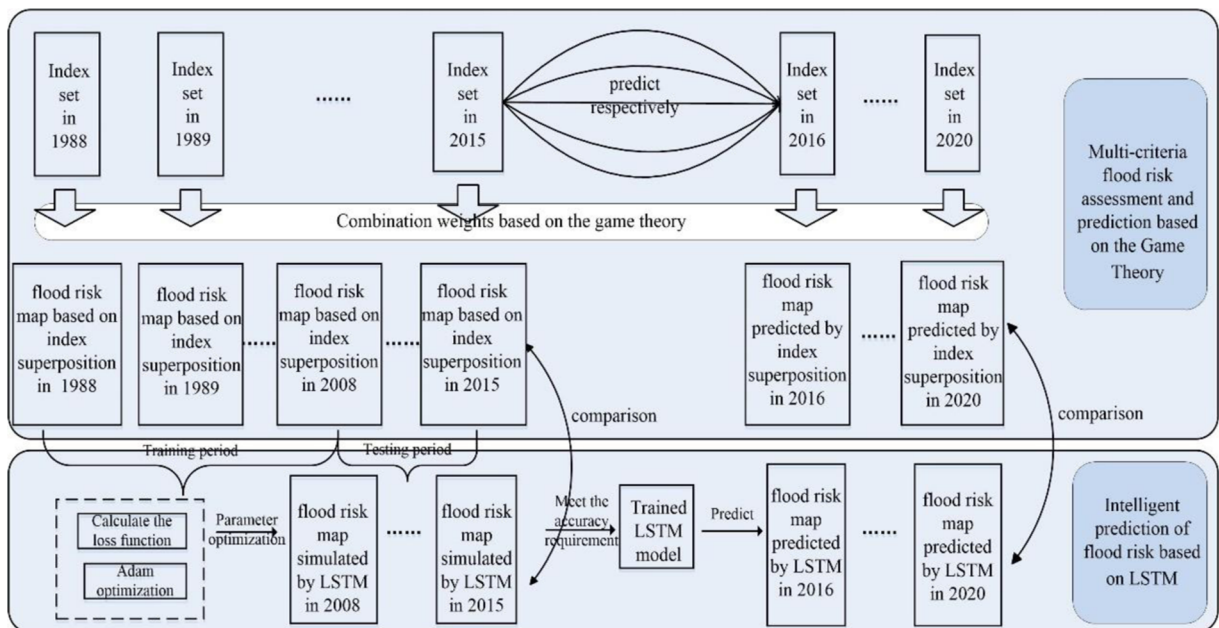


**Fig. 1** Flow chart of intelligent prediction and zonation of the flood risk of HRB based on LSTM
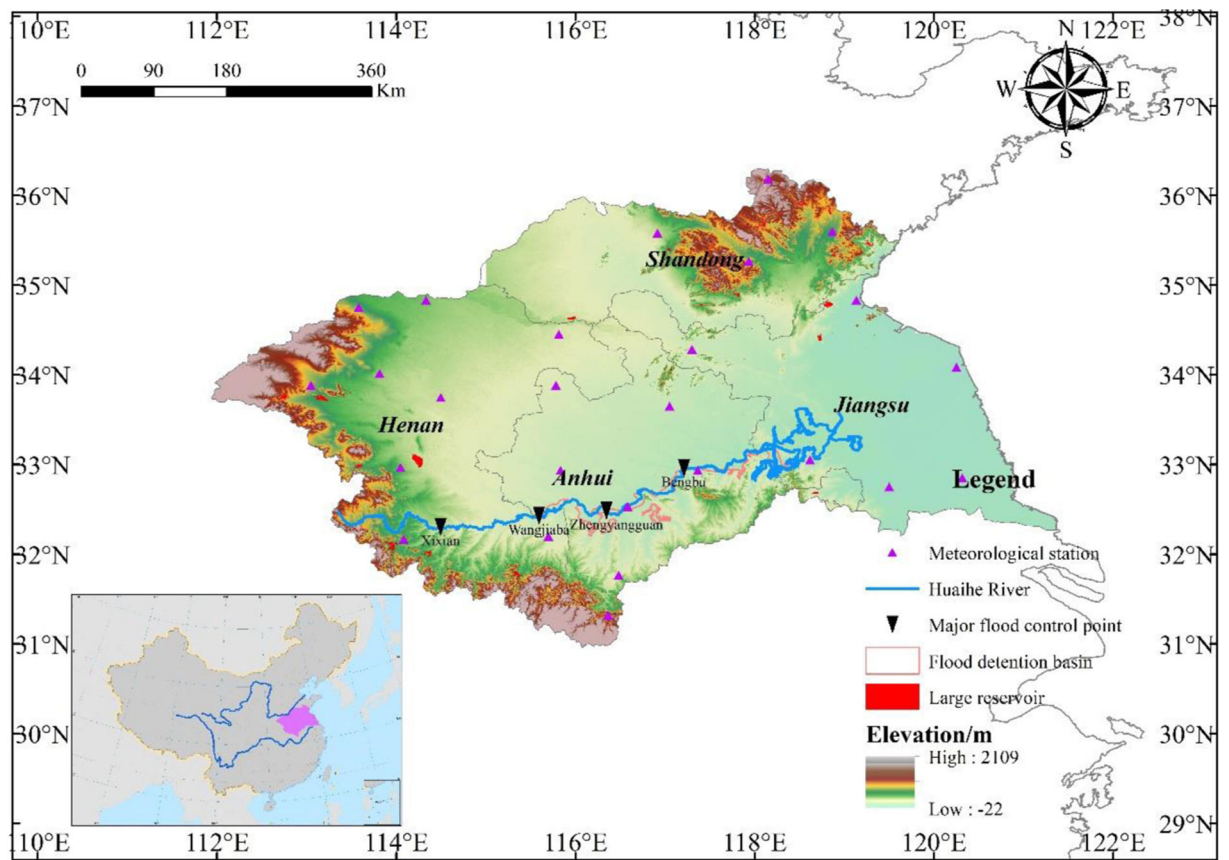
**Fig. 2** Sketch map of the HRB

maximum 3-day rainfall. The disaster environment includes elevation, slope, river buffer, and drainage density. The disaster-bearing body includes population density, GDP density, and virtual sown area of crops. The disaster

prevention and mitigation capability take reservoir storage modulus and flood detention basin modulus into account.

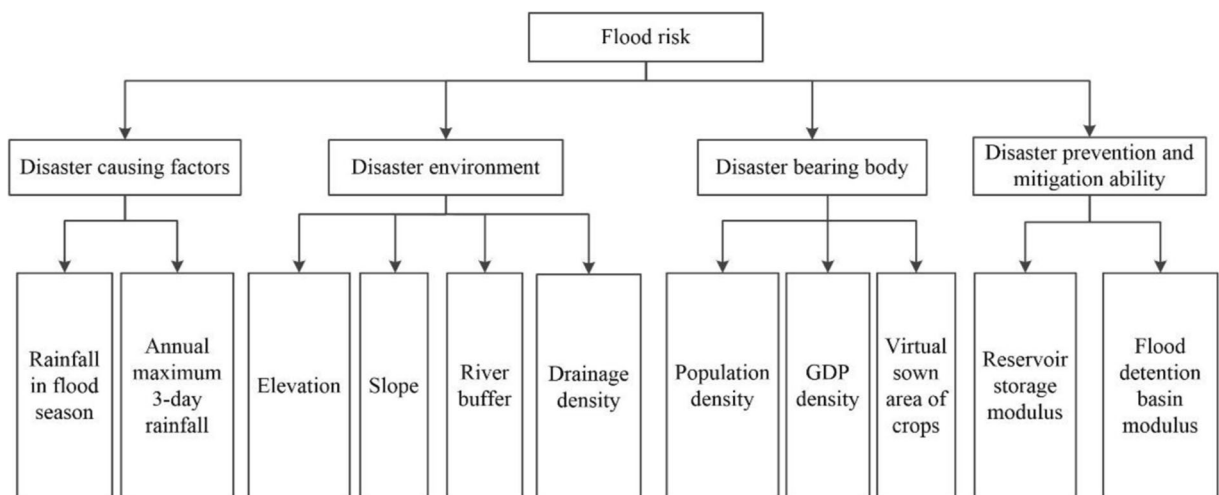Among them, the positive indexes (the larger the index value, the greater the flood risk value) are rainfall



**Fig. 3** Structure diagram of flood risk assessment index system

in flood season, annual maximum 3-day rainfall, river buffer, drainage density, population density, GDP density, and virtual sown area of crops. Negative indexes include elevation, slope, reservoir storage modulus, and flood detention basin modulus.

1. Rainfall in flood season (*XP*): Large-scale, high-intensity, and long-duration rainfall is the direct cause of flood disaster, and the rainfall in flood season accounts for the vast majority of the annual rainfall. Therefore, flood disaster is closely related to the rainfall in flood season. The larger the rainfall in flood season, the more vulnerable it is to the flood disaster.

2. Annual maximum 3-day rainfall (*P3*): Rainfall is the direct factor in the formation of floods. According to the difference of climate, area, and underlying surface conditions, this index is selected. The greater the annual maximum 3-day rainfall, the more vulnerable it is to the flood disaster.

3. Elevation (*H*): Most of the flood disasters occur in low-lying areas, and the higher terrain far away from the river can effectively avoid the infringement of flood events. Therefore, the higher the elevation, the less likely it is to be affected by the flood disaster.

4. Slope (*S*): Slope is an important factor affecting the speed of water flow. In the upper reaches of the river where the slope is large, the flood velocity is fast, and the flood recedes rapidly, and the probability of a flood disaster is relatively small, while the terrain of the lower reaches of the river is flat, and the flood is stranded and superimposed for a long time, which is more prone to a flood disaster. Therefore, the greater the slope of the terrain, the less likely it is to be affected by the flood disaster.

5. River buffer (*RB*): The surrounding area of the middle and lower reaches of the river and the lake is called a buffer. The buffer, in addition to the threat of local rainfall, is also threatened by the formation of infiltration, levee, and collapse of the flood. Therefore, the flood risk in this part of the area is higher than that of other places, and the risks of high-grade rivers are relatively high due to the large runoff.

6. Drainage density (*D*): refers to the total length of the river within a unit area. The higher the drainage density, the more channels of flood discharge, the wider the adjacent river area, and the more vulnerable it is to the flood disaster.

7. Population density (*PD*): refers to the number of permanent residents per unit land area (km$^2$), which can reflect the number and distribution of disaster-bearing bodies in the basin. The greater the population density, the more concentrated the population distribution, the more vulnerable it is to the flood disaster.

8. GDP density (*GD*): refers to the gross domestic product (GDP) created by unit land area (km$^2$), which can reflect the efficiency of land use, the density of output value and the level of economic development. The larger the GDP density, the more concentrated the distribution of GDP, the more vulnerable it is to the flood disaster.

9. Virtual sown area of crops (*A*): It reflects the virtual sown area under the situation of rice as the standard crop planting in the basin. The larger the virtual sown area of the crops, the more rice is planted, so the greater the flood disaster tendency is.

10. Reservoir storage modulus (*SD*): Reservoir is an important flood control project, which plays the role of stagnant flood storage and flood peak abatement. The larger the flood control capacity above the flood control section, the greater the regulation effect on the flood, the more difficult to be affected by the flood disaster.

11. Flood detention basin modulus (*FD*): Flood detention basin is a common emergency flood control project, which plays a role in sacrificing the part to protect the whole and taking the initiative to exchange smaller controllable losses for greater flood control benefits. Therefore, the greater the flood detention basin modulus, the less likely it is to be affected by the flood disaster.

Data source

The data sources are shown in Table 1.

Index processing

The grid layers with a resolution of 90 m × 90 m of the indexes above are generated by ArcGIS, and the spatial distribution method of each index is shown as follows.

**Table 1** Index sources

| Category | Index | Year | Data sources |
|---|---|---|---|
| Disaster-causing factors | XP | 1988–2015 | China Meteorological Data Network |
| | P3 | 1988–2015 | |
| Disaster environment | H | – | DEM |
| | S | – | From elevation |
| | RB | – | From elevation |
| | D | – | National Geomatics Center of China |
| Disaster-bearing body | PD | 1988–2015 | Yearbook of the Jiangsu, Anhui, Shandong, |
| | GD | 1988–2015 | and Henan province |
| | A | 1988–2015 | |
| Disaster prevention and mitigation capability | FD | 1988–2015 | Hydrographic bureau of HRB |
| | SD | 1988–2015 | |

Note: (1) Data have already been subjected to strict homogenization quality control; (2) Limited by the yearbook information, the evaluation year begins in 1988. There are 39 large-scale reservoirs in the HRB, and the reservoir built after 1988 includes the Banqiao Reservoir (1993), the Shimantan Reservoir (1996), the Bailianya Reservoir (2008), and the Yanshan Reservoir (2009)

### XP and P3

The inverse distance weight (IDW) method (Lu and Wong 2008) is used for interpolation. This method is a kind of spatial distribution method which fully considers the regional relationship between various factors. Because of its simple principle and accurate result, it has been widely used in spatial analysis of discrete points. It is assumed that each discrete point has a local influence on the interpolation point, which weakens with the increase of distance. The corresponding tool bar in ArcGIS is IDW interpolation.

### H and S

Elevation is from the DEM data; slope is extracted by DEM data and the "slope" function in ArcGIS.

Suppose that the probability of flood disaster is close to 0 above the elevation $h$ m, and the probability of flood disaster is 1 from the lowest point of the basin to $x$ m, indicating that the probability of flood risk is the same. From $x$ m to $h$ m, the linear assignment from 1 to 0 indicates that with the increase of elevation, the risk of flood disaster decreases gradually. $x$ and $h$ are determined according to the general situation of the basin. $h$ can be determined according to the maximum inundation height of the historical flood, and $x$ is generally 0.

### RB

The distance from a river or reservoir means that the risk of flood disaster is different, and on the premise of equal distance, the higher the grade of the river or the larger the reservoir capacity, the greater the flood risk, so the buffers are divided into different levels according to the distance from the river or reservoir of different levels, and the scope of the buffer is properly modified according to the terrain at the same time. The correction principle is that the buffer is wider in the reach with large flow and gentle terrain, and narrower in the reach with small flow and steeper terrain. The specific assignment is shown in Table 2.

### D

The grid distribution map is generated by ArcGIS, and the river system distribution map and grid distribution map are superimposed. Count the river length in each grid, divide

**Table 2** River buffer value

| Distance | River level | | | Reservoir level | |
|---|---|---|---|---|---|
| | 1–3 level | 4 level | 5 level | 1 level | 2 level |
| 0–5 km | 1 | 0.7 | 0.6 | 1 | 0.9 |
| 5–6 km | 0.9 | 0.6 | 0.5 | 0.9 | 0.8 |
| 6–7 km | 0.8 | 0.5 | 0.4 | 0.8 | 0.7 |
| 7–8 km | 0.7 | 0.4 | 0.3 | 0.7 | 0.6 |
| 8–9 km | 0.6 | 0.3 | 0.2 | 0.6 | 0.5 |
| 9–10 km | 0.5 | 0.2 | 0.1 | 0.5 | 0.4 |
| > 10 km | 0 | 0 | 0 | 0 | 0 |
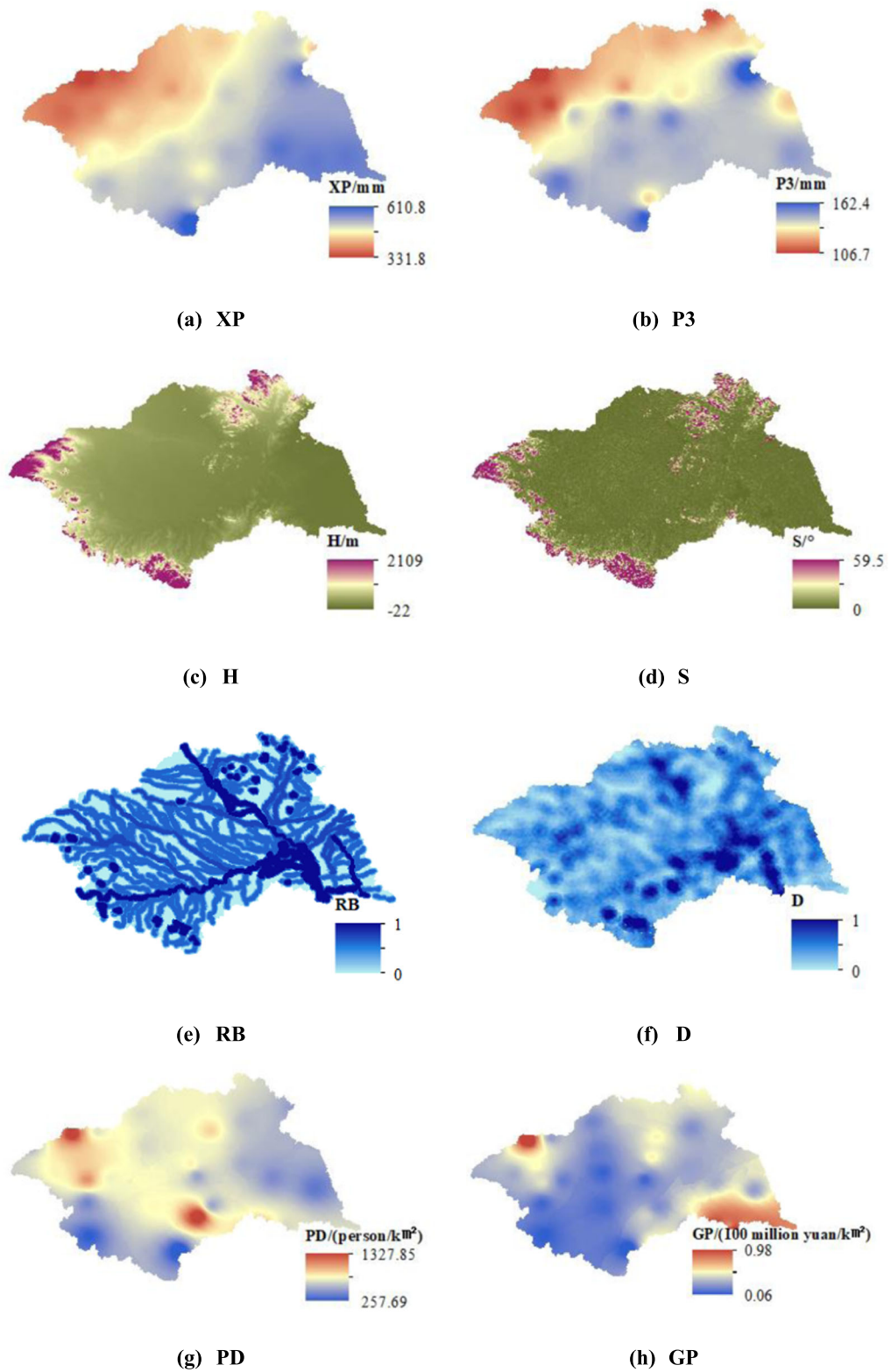
**Fig. 4** Distribution of indexes (a) XP, (b) P3, (c) H, (d) S, (e) RB, (f) D, (g) PD, (h) GP, (i) A, (j) SD, and (k) FD. Note: the XP, P3, H, S, RB, and D are the multi-year average value; PD, GDP, A, SD, and FD are the latest data in 2015
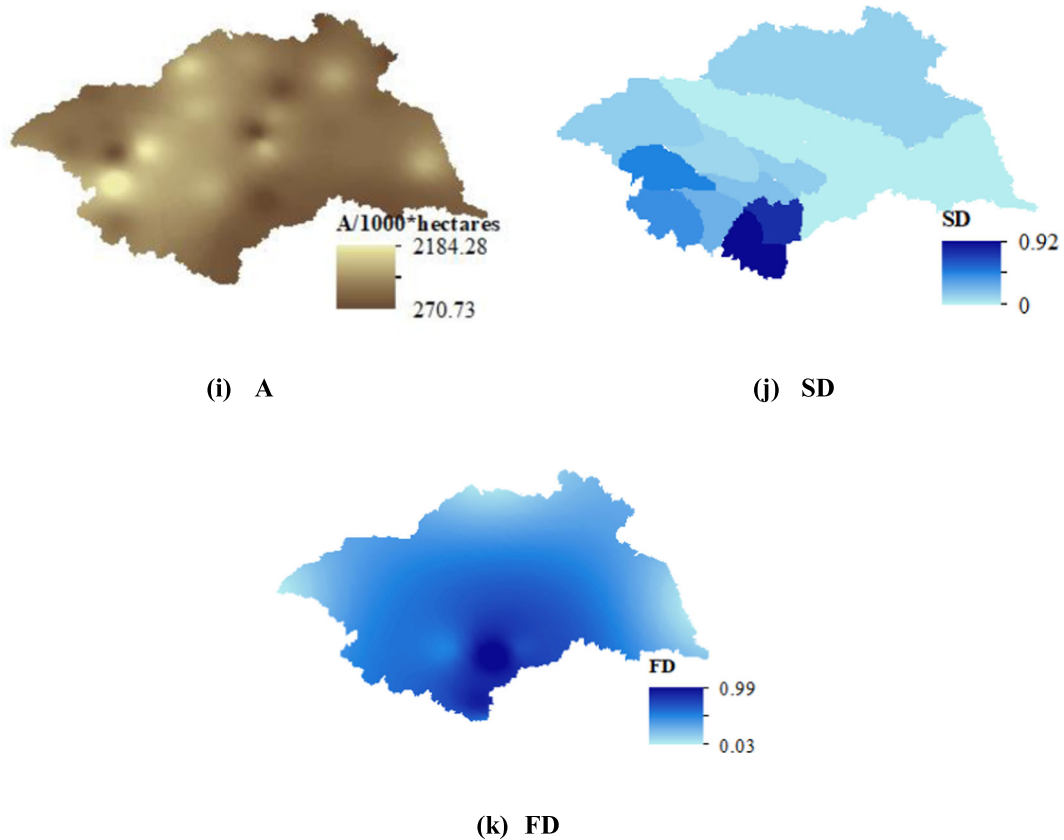
(i)  A



(j)  SD



(k)  FD

**Fig. 4**  continued.

by the grid area, and the drainage density of the grid is obtained. The drainage density value is assigned to the corresponding grid. The formula is as follows:

$$D_i = \frac{l_i}{a_i} \quad (i = 1, 2, \cdots, n) \tag{1}$$

where $n$ is the grid number, $D_i$ is the drainage density of the $i$th grid, $l_i$ is the river length of the $i$th grid, and $a_i$ is the area of the $i$th grid.

*PD and GD*

Because the population and GDP are not evenly distributed, the spatial distribution of population density and GDP density adopts the spatial kernel density estimation method (Sheather and Jones 1991). The corresponding tool bar in ArcGIS is the kernel density.

*A*

In the same area, it is assumed that the planting density of the same crop is fixed, the larger the sown area is, the more severe the flood disaster will be. Because of the different flood tolerance time of the different crops, the flood risk of them is different from the same sown area. Therefore, in order to fully reflect the difference of flood risk in areas with different planting structures with the same sown area, we comprehensively consider the degree of flood tolerance of different crops from the point of view of disaster-bearing body, unify the total sown area of crops containing all kinds of crops in different regions into the virtual sown area of rice (the standard crop), so that the agricultural losses of different regions are comparable. Thus, the index of "virtual sown area of crops" is constructed. The larger the index value is, the greater the flood risk will be, and this index can better reflect the flood risk. The specific description is as follows.

First, select the main crops planted in the area, and screen crops that are still growing in the main flood

season of each year (June to August), which are crop $A_j$ ($j = 1, 2, \ldots, m$), respectively.

Assume that crop $A$ will experience $n$ growth periods during the flood season, namely $X_i$ ($i = 1, 2, L, n$). The duration of each growth period is $T_i$ ($i = 1, 2, L, n$) $d$, $\sum_{i=1}^{n} T_i = 92$, respectively. The flood tolerance duration corresponding to different growth period (the allowable flooding time for crops without significant yield reductions) is $t_i$ ($i = 1, 2, L, n$) $d$, respectively, so the comprehensive flood tolerance duration is $\sum_{i=1}^{n} \frac{t_i \cdot T_i}{92} d$. From this, it is calculated that the comprehensive flood tolerance duration of the reference standard crop rice is 5 $d$.

The crop loss coefficient is defined as the reciprocal after the standardization of the comprehensive flood tolerance duration of crops, which reflects the disaster tendency of crops with different comprehensive flood tolerance duration compared to rice, so the loss coefficient of crop $A_j$ is $\frac{5}{T_{Aj}}$. The virtual sown area of crops is defined as the product of loss coefficient and actual sown area, so the virtual sown area of crop $A_j$ is $\frac{5}{T_{Aj}} \cdot S_{Aj}$, where $S_{Aj}$ is the actual sown area of the crop $A_j$, then the total virtual sown area of crops growing in the flood season of the region is $\sum_{j=1}^{m} \frac{5 \cdot S_{Aj}}{T_{Aj}}, j = 1, 2, \cdots, m$.

### SD and FD

The existence of watershed makes the flood risk caused by different tributaries have regional characteristics. The tributaries continue to meet, the more downstream, the greater the flood volume. We adopt the method proposed by Wu et al. (2015, 2017) to avoid the spatial cross-border transfer of flood control capacity in the process of reservoir storage modulus distribution.

The flood detention basin refers to low-lying areas and lakes that store flood temporarily outside the river embankment. It is an important part of a flood control system and an effective measure to ensure flood control safety and reduce flood disasters in key areas. The calculation principle and distribution method of it are the same as the reservoir storage modulus.

Elevation, slope, drainage density, river buffer, and flood detention basin modulus do not change much with time, so it is considered that these five indexes remain unchanged in a short period of time. The characteristic distribution of each index is shown in Fig. 4.

According to the correlation analysis of indexes, it is found that the correlation coefficients between indexes are less than 0.9, so it is considered that the interaction between indexes is within the reasonable range.

### Standardization rules

Because the numerical range and unit of each index are different, which is not conducive to unified calculation, the indexes need to be standardized from their original values to the value ranging from 0 to 1 in order to eliminate the influence of measurement scales and make different indexes comparable both in space and time. To the positive indexes, the formula is as follows:

$$X_{is} = \frac{X_i - X_{i\min}}{X_{i\max} - X_{i\min}} \tag{2}$$

where $X$ is the standardized value of the index, $X_i$ is the initial value of the index, and $X_{max}$ and $X_{min}$ are the space-time maximum and the minimum value of the index, respectively.

To the negative indexes, the formula is as follows:

$$X_{is} = 1 - \frac{X_i - X_{i\min}}{X_{i\max} - X_{i\min}} \tag{3}$$

The meaning of variables in the formula is the same as above.

### Methodology

Multi-criteria flood risk assessment based on the game theory

Determining a suitable weight is a critical step during the evaluation process. In this paper, the gray correlation–binary comparison analysis method is used to calculate the subjective empirical weight of the index; the improved CRITIC method is used to calculate the objective weight of the index. The final optimization weight of each index is given by using the game theory coupling two kinds of weights, so as to overcome the defects caused by the subjective influence of the two different kinds of weights.

The evaluation process is as follows. Firstly, determine the appropriate factors, and this part has been introduced in the "Data" section. Secondly, determine the optimal comprehensive weights, which will be introduced in the "Weight definition" section. Then the

**Table 3**  Mood operator, scale, and fuzzy membership degree

| Mood operator | I | II | III | IV | V | VI | VII | VIII | IX | X |
|---|---|---|---|---|---|---|---|---|---|---|
| Scale | 0.50 | 0.55 | 0.60 | 0.65 | 0.70 | 0.80 | 0.85 | 0.90 | 0.95 | 1.00 |
| Fuzzy membership degree | 1.000 | 0.818 | 0.667 | 0.538 | 0.429 | 0.250 | 0.176 | 0.111 | 0.053 | 0.000 |

flood risk index is obtained by using the weights to stack the index layers, and then the risk level is classified by the natural classification method, from which the annual flood risk zonation map is obtained, and this part will be introduced in the "Flood risk index calculation" and "Flood risk classification" sections. The applications implemented are ArcGIS and Excel.

*Weight definition*

*Gray correlation–binary comparison analysis method* In view of the lack of superiority comparison between elements in the AHP method, Chen and Guo (2006) put forward the fuzzy set theory of nonstructural decision-making. In this paper, the binary comparison method is selected and the traditional gray correlation method is coupled. It does not need to carry out the consistency test, and the calculation is simple, which can overcome the disadvantage that the traditional AHP method neglects comparison of the superiority of the elements. The gray correlation method and binary comparison analysis method are introduced below.

1.  The gray correlation method

    The gray correlation method can mine partly known information to measure the correlation degree between indexes. The main steps are divided into calculating the correlation coefficient and calculating the correlation degree (Zhang and Zhang 2007).

2.  Gray correlation–binary comparison analysis method

a)  Rank the index importance

    After using the gray correlation method to determine the rank of the index importance, the expert opinions are introduced. The expert opinions are consulted many times and adjust appropriately, and the reasonable ranking which can best reflect the actual situation is finally determined.

b)  Select the scale and calculate the weight

    According to the ranking of each index, the appropriate mood operator and fuzzy membership degree are selected for each index from Table 3. When selecting mood operator, the ranking adjusted by the experts must be strictly followed, so that the mood operator satisfies the weak consistency, i.e., each index is compared with the most important index in turn, and the relative importance of each index can be described in Table 3.

    The relationship between scale and fuzzy membership degree in Table 3 can be expressed by the relationship

**Fig. 5** The relationship between the scale and fuzzy membership degree
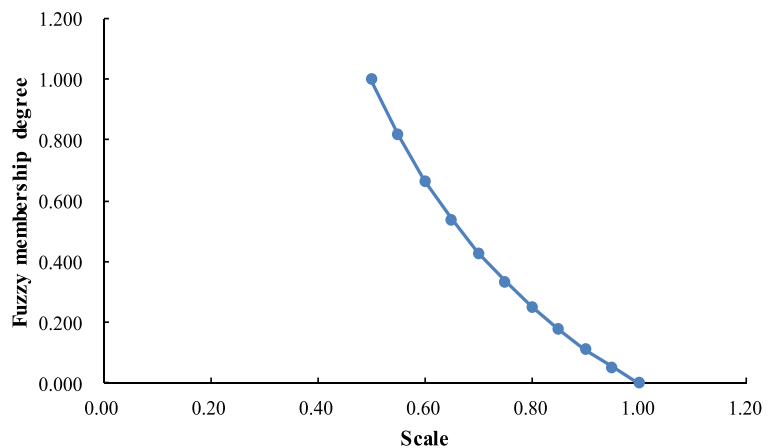
curve shown in Fig. 5. When the mood operator is described between the two mood operators in Table 3, the fuzzy membership degree value is obtained from Fig. 5 by the interpolation method. The fuzzy membership degree of each index is used as the weight coefficient, and the final weight is obtained after standardization.

*The improved CRITIC method* The CRITIC (Criteria Importance Through Intercriteria Correlation) method is an objective weighting method proposed by Diakoulaki et al. (1995), whose basic idea is to comprehensively measure the objective weight of each index by comparison intensity and conflict.

Because of the different dimensions of each index, the standard deviation could not tell the difference between them directly. In this paper, we use the improved CRITIC method. The contrast intensity is reflected by the standard deviation coefficient (the ratio of standard deviation to mean value) instead of the standard deviation.

The conflict quantitative index $\sum_{i=1}^{n}\left(1-r_{ij}\right)$ between the $j$ index and the other indexes is defined by the CRITIC method, where $r_{ij}$ represents the correlation coefficient between the $i$ index and the $j$ index, $n$ is the total number of the indexes, and considering the negative correlation coefficient, the improved CRITIC method measures the conflict between the indexes by calculating $\sum_{i=1}^{n}\left(1-\left|r_{ij}\right|\right)$.

The information amount $G_j$ contained in the $j$ index is expressed as $G_j = \delta_j \cdot \sum_{i=1}^{n}\left(1-\left|r_{ij}\right|\right)$, where $\delta_j$ represents the inter-class standard deviation coefficient of the $j$ index, and the improved CRITIC weight is expressed as $w_j = \frac{G_j}{\sum_{j=1}^{n} G_j}$.

*Combination weighting based on the game theory* Game theory is the theory of rational behavior and decision equilibrium when the behavior of decision-making subject affects each other (Myerson 2013). The basic idea of combination weighting based on the game theory is to find compromise or balance between different weights, so that the deviation between each basic weight and possible weight is minimized and a relatively coordinated comprehensive weight can be obtained. The concrete performance is finding the optimal weight coefficient $\alpha_k^*$ to minimize the deviation between the optimal combination weight $w_k^*$ and the $w_k$ calculated by different methods, which is $\min\left\|\sum_{k=1}^{n}\alpha_k \cdot w_k^T - w_i^T\right\|_2$, $i = 1, 2, \cdots, n$. $\alpha_k$ is then calculated and $\alpha_k^* = \frac{\alpha_k}{\sum_{k=1}^{n}\alpha_k}, k = 1, 2, \cdots, n$ is obtained after standardization, and the optimal combination weight is represented as $w^* = \sum_{k=1}^{n}\alpha_k^* \cdot w_k^T$, $n = 2$.

*Flood risk index calculation*

After the standardization of the 11 indices mentioned above, the evaluation matrix $B$ shown in formula (4) is formed, and $n$ is the total number of points of the 90 m × 90 m grid layer. The flood risk index *Risk* is expressed as the product of the evaluation matrix and the optimal comprehensive weight, and each element of the result vector shown by formula (5) represents the risk value of the corresponding point on the actual grid layer.

$$B = \begin{bmatrix} P_1 & R_1 & H_1 & S_1 & RB_1 & D_1 & PD_1 & GD_1 & A_1 & SD_1 & FD_1 \\ P_2 & R_2 & H_2 & S_2 & RB_2 & D_2 & PD_2 & GD_2 & A_2 & SD_2 & FD_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ P_i & R_i & H_i & S_i & RB_i & D_i & PD_i & GD_i & A_i & SD_i & FD_i \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ P_n & R_n & H_n & S_n & RB_n & D_n & PD_n & GD_n & A_n & SD_n & FD_n \end{bmatrix} \quad (4)$$

$$(i = 1, 2, \cdots, n)$$

$$Risk = f(F, E, U, V) = B \times W = \begin{bmatrix} P_1 & R_1 & H_1 & S_1 & RB_1 & D_1 & PD_1 & GD_1 & A_1 & SD_1 & FD_1 \\ P_2 & R_2 & H_2 & S_2 & RB_2 & D_2 & PD_2 & GD_2 & A_2 & SD_2 & FD_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ P_i & R_i & H_i & S_i & RB_i & D_i & PD_i & GD_i & A_i & SD_i & FD_i \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ P_n & R_n & H_n & S_n & RB_n & D_n & PD_n & GD_n & A_n & SD_n & FD_n \end{bmatrix} \times \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_i \\ \vdots \\ W_{11} \end{bmatrix} = \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_i \\ \vdots \\ R_n \end{bmatrix} \quad (5)$$

where $F$, $E$, $U$, and $V$ are the flood risk described by the four aspects of indexes which are disaster-causing factors, disaster environment, disaster-bearing body and disaster prevention and mitigation capability.

*Flood risk classification*

We classify the flood risk by Jenks natural breaks classification method (Jiang 2013). The basic idea of this method is to minimize the variance within each level and maximize the variance between the different levels. Output the annual flood risk index value and classify them into five levels, i.e., very low, low, moderate, high and very high, and then the annual flood risk zonation map is obtained.

Intelligent prediction model of flood risk based on LSTM

The intelligent algorithm can simplify the operation process without setting the index weight and classification standard. LSTM has natural advantages in processing time series data, so LSTM is selected to predict the flood risk intelligently.

*Model principle of LSTM*

The LSTM improves the structure of traditional recurrent neural network (RNN) by introducing specific memory unit into the hidden layer to control the addition, deletion, modification, and provision of information to other neurons in the model. The working mechanism of this memory unit mainly comes from three "gates": memory gate, input gate, and output gate. The internal structure of the standard LSTM module is shown in Fig. 6. The specific

working principle can be found in (Hochreiter and Schmidhuber 1997). Through the subtle operation of these three "gates" units, LSTM completes the connection of the time relationship before and after the time series data, which makes it always maintain the feature extraction of the time series data and complete the data influence on the next or more time steps in the current time step state.

*Modeling steps*

The LSTM flood risk prediction model includes the following steps. (1) Select samples and divide training set and testing set. (2) Use backpropagation through time (BPTT) algorithm for model training, set the minimum loss function as the optimization objective, and select adaptive moment estimation (Adam) algorithm to update the weights. (3) Particle swarm optimization algorithm is used to optimize the model parameters to minimize the fitting root mean square error of the testing set. (4) The future flood risk prediction will be carried out after the training and testing model both meet the accuracy requirements. (5) The flood risk prediction zonation map is generated based on ArcGIS.

## Results

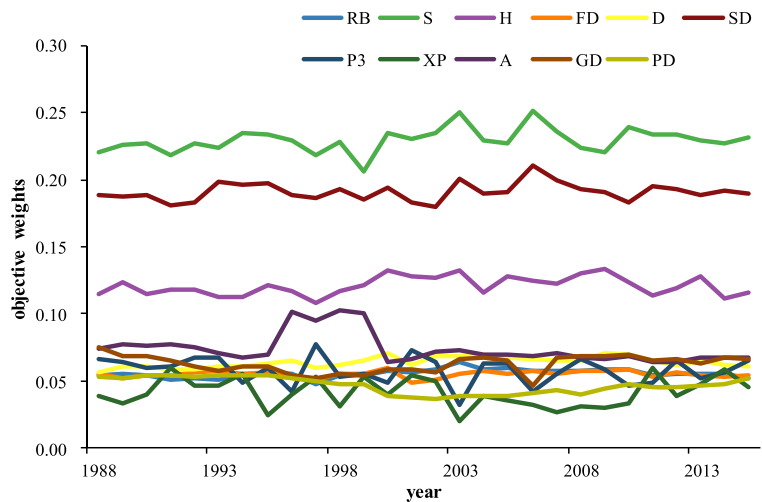Flood risk assessment and prediction of the HRB based on the game theory

*Weight determination*

The subjective weights of the indexes remain unchanged and the objective weights change yearly.
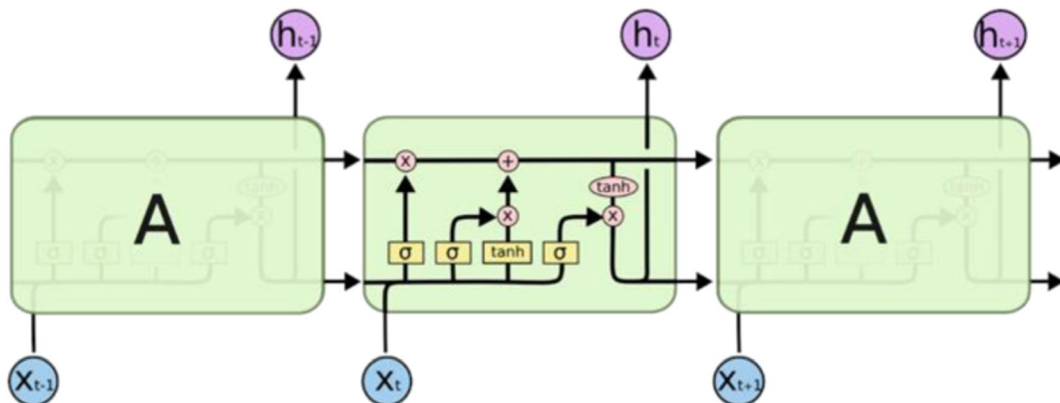


**Fig. 6** The internal structure of the LSTM module

**Fig. 7** The trends of the objective weights



Firstly, the gray correlation-binary comparison weight is determined, and then the objective weight of each index from 1988 to 2015 is calculated by using the improved CRITIC method. We use the game theory to couple the two kinds of weights to get the optimal combination weight of each index from 1988 to 2015 finally.

The subjective weights are determined according to the order of RB, S, H, FD, D, SD, P3, XP, A, GD, and PD as follows: 0.0008, 0.1575, 0.1635, 0.0537, 0.0988, 0.0625, 0.0823, 0.1168, 0.2106, 0.0165, and 0.037, respectively. The first three most important indexes are virtual sown area of crops, elevation and slope, which is slightly different from the rankings in reference (Sheng et al. 2010; Zhang et al. 2014). This is because although rainfall is the direct cause of the flood disaster and when the flood comes, people generally carry out organized flood rescue, the amount of rainfall may not reflect the flood disaster degree comprehensively and objectively. The rainfall has not been selected as the most important influencing index, which is more in line with reality.

The changing trends of the objective and combination weights of 11 indexes are shown in Fig. 7 and Fig. 8, respectively. It can be seen from Fig. 7 that the first three most important indexes of the objective weights are slope, reservoir storage modulus and elevation, and the weight values of the indexes do not fluctuate with time.

The results of the combination weights show that the slope is the most important index, the virtual sown area of crops and elevation follow after. The combination weight of each factor does not fluctuate with time which is similar to the objective weight. Therefore, game

**Fig. 8** The trends of the combination weights
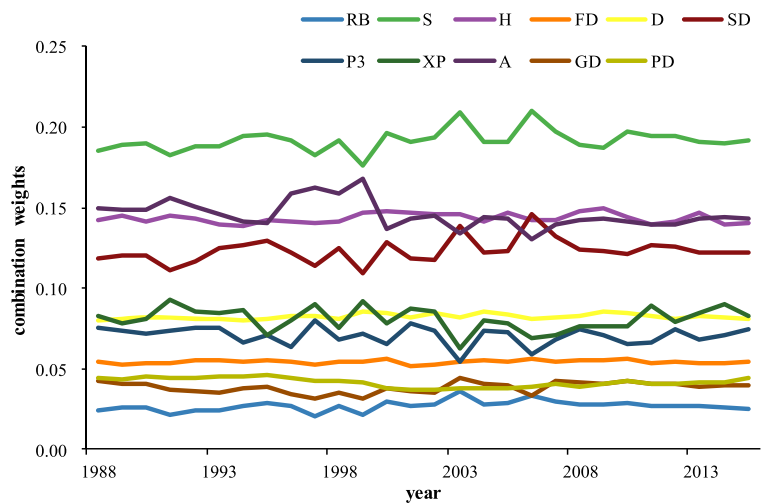
theory can effectively couple the two kinds of weights, thus giving more realistic weights value.

*Historical flood risk zonation*

The annual flood risk indexes of HRB from 1988 to 2015 are obtained by combination weights and risk influencing indexes, and the corresponding evolution zonation maps of flood risk of HRB are obtained according to the classification method of natural breaks.

We divided the value of the flood risk index into five levels. The flood risk is very low when the value ranges in 0.193157–0.49376, low in 0.49376–0.57989, moderate in 0.57989–0.63063, high in 0.63063–0.67579, and very high in 0.67579–0.822399.

The flood risk zonation maps of HRB in 2015 are given in four aspects: disaster-causing factors, disaster environment, disaster-bearing body, and disaster prevention and mitigation capability (Fig. 9) and the general annual flood risk zonation map is also shown in Fig. 9.
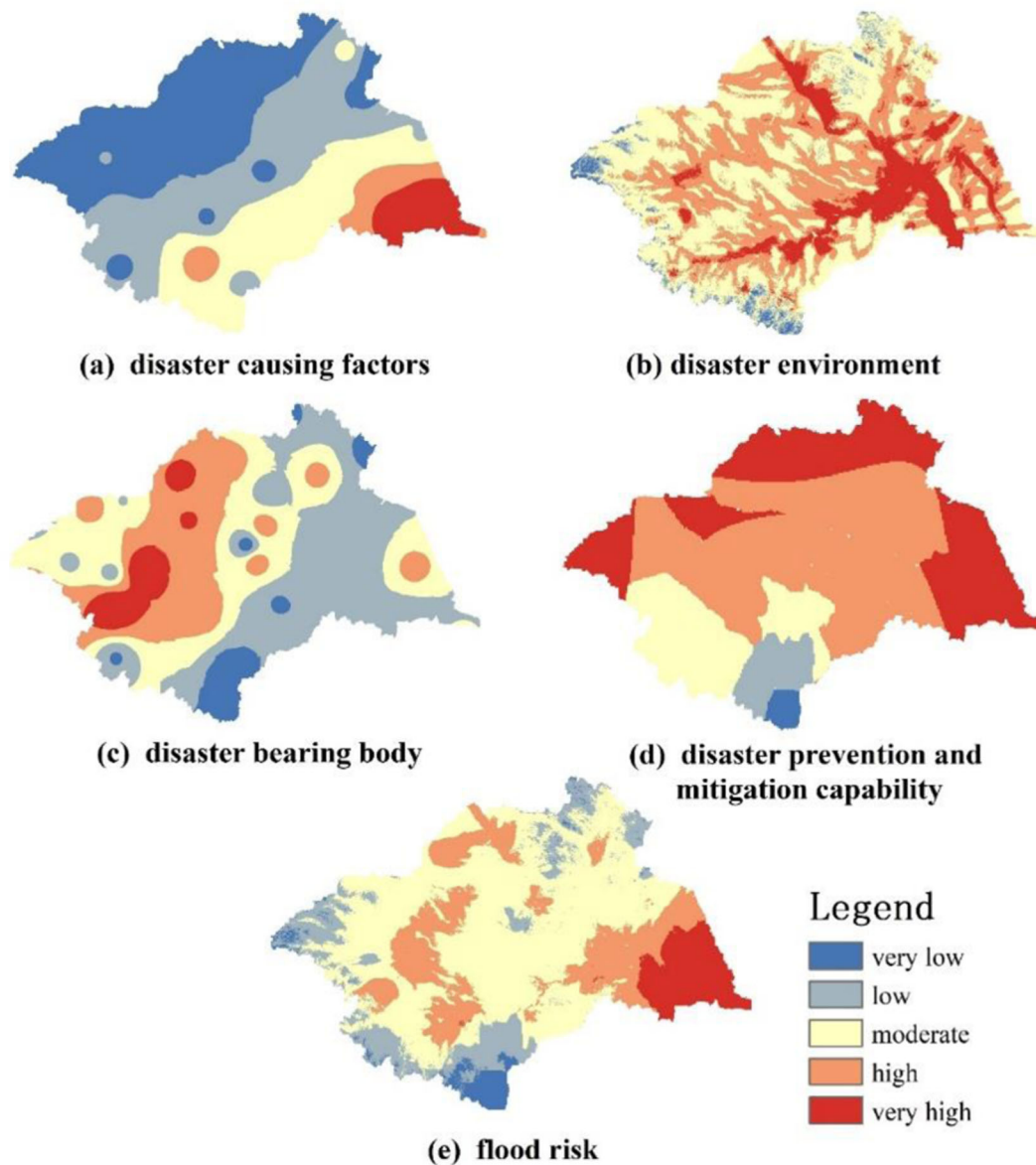


**Fig. 9** Flood risk zonation maps of HRB in 2015

It can be seen from the figure that the disaster-causing factors are all rainfall-related factors, so the flood risk caused by disaster-causing factors is gradually reduced from the southeast to northwest and from coastal area to the inland.

The high flood risk area of the disaster environment is concentrated on both sides of the main stream of the Huaihe River. With the distance from the river becomes farther and farther away, the risk decreases gradually. The flood risk of the Dabieshan mountain in the southernmost part of the basin where elevation is 300 m–1774 m and the Funiushan Mountain and Tongbaishan Mountain in the west corner of the basin where elevation is 200 m–500 m are the lowest because of the obvious vertical distribution of the terrain, which is consistent with the actual situation.

The risk of disaster-bearing body reflects the development degree of economy, population, and agriculture. The area with the highest flood risk is located near Zhumadian, Zhoukou, Shangqiu, and Heze cities. In these areas, the proportion of farmland vulnerable to the flood is high, the economy is underdeveloped, the population is dense, the general awareness of flood control is weak, the speed of flood evacuation is slow, and the comprehensive effect makes the risk concentration the highest. The area with the lowest risk is located in Lu'an city and the southern mountain area, which is remote, sparsely populated, backward in economic development and not suitable for the development of agriculture.

In view of the flood risk of disaster prevention and mitigation capability, the reservoir storage modulus is the highest in Wangjiaba, Jiangjiaji, and Runheji area, so the risk is the lowest, while the reservoir storage modulus above Xixian county, Yishusi area, and Lixiahe area is the lowest, so the risk is the highest on the opposite.

In 2015, the very high flood risk area in HRB is located in Lixiahe area, accounting for 7.8% of the whole basin area, which is basically consistent with the very high risk area of disaster-causing factors, disaster environment, and disaster prevention and mitigation capability. Lixiahe area is located in the middle of Jiangsu province, belonging to the coastal plain, with humid climate, low-lying terrain, large population, and developed economy, resulting in the greatest flood risk. The high risk area is located in the Mengwa flood detention area and the surrounding low-lying zone, the southeast of the basin, the junction of Henan and Anhui

province, and the west of Shandong province, accounting for 20.9% of the basin area. The very low risk area is located in Dabieshan mountain area, accounting for 3.3% of the basin area. It is characterized by high terrain and backward economy and is also the area with the largest reservoir storage modulus, which is consistent with the actual situation, indicating that the evaluation results are reasonable.

*Future flood risk prediction*

In order to predict the future flood risk of HRB by using a combination weighting method based on game theory, the prediction value of each index must be obtained first.

*Future indexes prediction* It is considered that the elevation, slope, river buffer, and drainage density will remain unchanged in the short term (5 years) in the future. It is assumed that there is no new reservoir constructed in the basin, so the latest data are used for reservoir storage modulus and flood detention basin modulus. The prediction methods of other indexes are as follows.

*XP and P3* The rainfall in flood season and the annual maximum 3-day rainfall are both time series factors. LSTM is selected to construct the prediction model, and 1988–2010 is selected as the training period and 2011–2015 as the testing period. The results show that the *RMSE* of the model during the training period is 81.1 and 25.9, and the *RMSE* of the testing period is 179.9 and 59.1, respectively, which both meet the accuracy requirements. The trained LSTM model is used to predict the rainfall in flood season and the annual maximum 3-day rainfall in the next 5 years.

*PD* In order to facilitate the study, the following four assumptions are put forward. (1) The statistical data consulted and cited are accurate and effective. (2) The main influencing parameters of population change are birth, death, and net migration, and other secondary factors are not taken into account. (3) Large-scale natural and man-made disasters are not taken into account. (4) There will be no major population migration in the country within the projected period of time.

We adopt the population-development-environment (PDE) model (Holmes et al. 1994) and take the universal two-child policy on the fertility rate of the population into consideration. The medium fertility data and
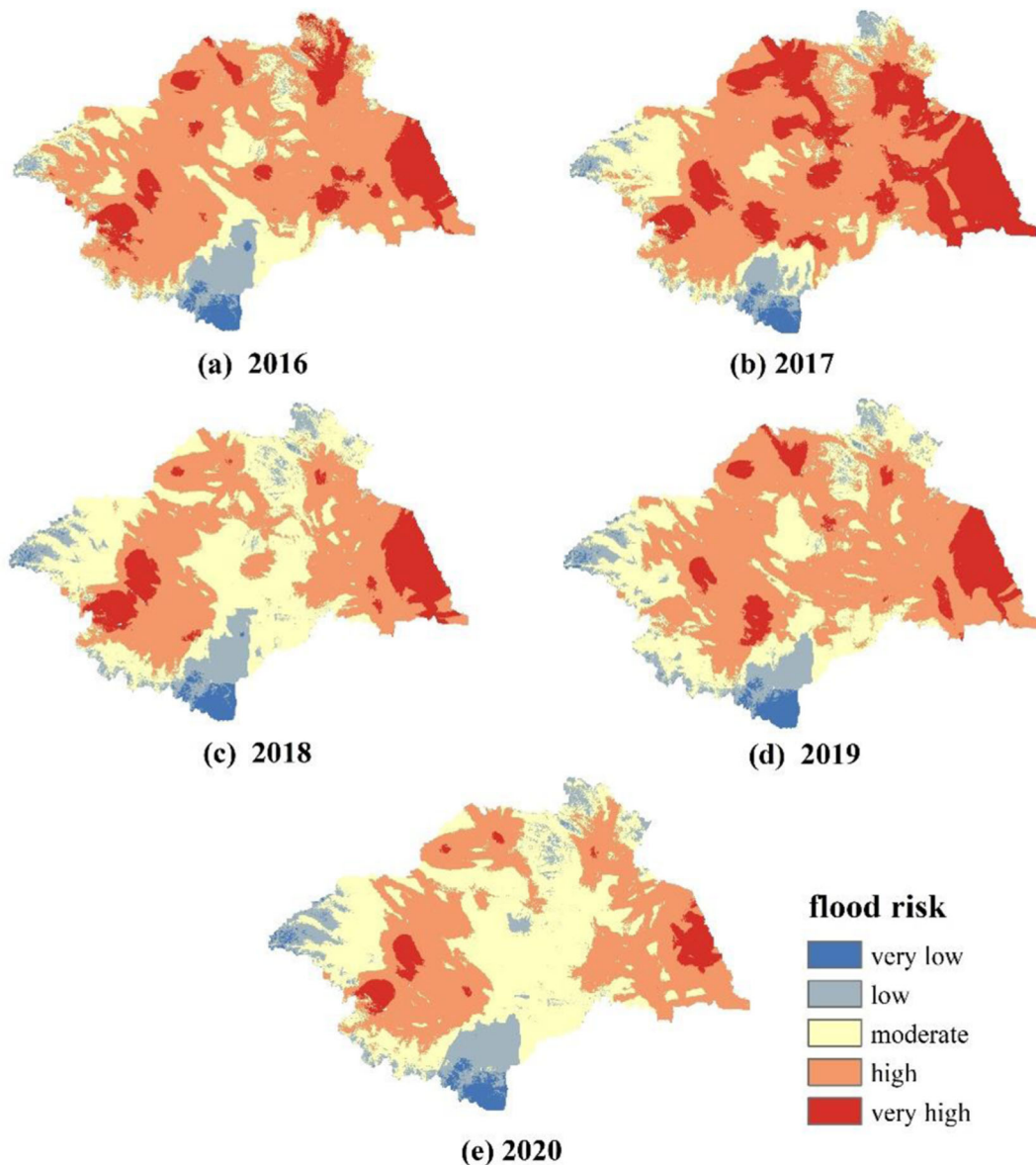
**Fig. 10** Flood risk zonation maps of HRB during 2016–2020

mortality data are substituted into the model to calculate the total population of each area of the basin in the future, and the population density is obtained by dividing the total population by the area.

*GD* Autoregressive integrated moving average (ARIMA) model is a time series modeling method proposed by Box and Jenkins (1976) in 1970 and is widely used in GDP prediction. ARIMA modeling is divided into the following four steps: (1) sequence

stabilization, (2) model order determination, (3) parameter estimation and model diagnosis, and (4) substitute the selected parameters for prediction.

The GDP data of the prefecture-level cities in HRB from 1988 to 2015 (GDP which has been converted to constant price in 1988) are selected as samples, in which the training period is from 1988 to 2008 and the testing period from 2009 to 2015. After meeting the accuracy requirements, the GDP density is obtained by dividing the predicted GDP by the area.

**Table 4** The basin area ratio of different flood risk levels during 2016–2020 (%)

| Year | Very low risk | Low risk | Moderate risk | High risk | Very high risk |
|------|------|------|------|------|------|
| 2016 | 1.9 | 6.7 | 17.9 | 60.1 | 13.4 |
| 2017 | 1.7 | 5.9 | 16.9 | 48.6 | 26.9 |
| 2018 | 2.2 | 9.6 | 37.4 | 41.0 | 9.8 |
| 2019 | 2.2 | 8.1 | 26.6 | 53.7 | 9.3 |
| 2020 | 2.1 | 10.3 | 43.1 | 40.0 | 4.5 |
| 5-year average | 2.0 | 8.1 | 28.4 | 48.7 | 12.8 |

A It is considered that the planting structure of crops will not change greatly in a short period of time, and the virtual sown area of crops in each year in the future will be taken as the average of the virtual sown area of crops in the previous 5 years.

*Future flood risk zonation map display* The flood risk zonation maps of the HRB during 2016–2020 are shown in Fig. 10, and the statistical table of the basin area ratio of different flood risk levels is shown in Table 4. It can be seen from the figure and the table that in the next 5 years, Lixiahe area, Xixian county, and Bantai-Wangjiaba interval are always in the very high flood risk areas, which need to be focused on. Dabieshan mountain area, Funiushan mountain area, and Tongbaishan mountain area are always in the very low risk area, so the prevention efforts can be appropriately reduced.

The area proportions of very low risk, low risk, and very high risk are relatively stable. The moderate risk coverage tends to increase, with the area percentage increasing from 17.9 to 43.1%. The flood risk around Bengbu city is changed from high to moderate, and the probability of flood disaster with high level is low, but it is possible to have local flood or urban waterlogging, and the high-risk area shows a decreasing trend, with the proportion of area decreasing from 60.1 to 40.0%. However, Xihua city surrounding, Mengwa flood storage area, Yanzhou city surrounding, Hongzehu Lake, and the eastern coast of Jiangsu are always in the high risk area, which still need to be paid attention to. The 5-year average results show that the high flood risk area accounts for the largest proportion while the very low-risk area accounts for the smallest proportion.

## Flood risk intelligent prediction model of HRB based on LSTM

An intelligent prediction model of the flood risk in HRB based on LSTM is constructed. Taking the flood risk value of HRB from 1988 to 2015 as the reference value, the model is trained to predict the future flood risk of the basin intelligently, and the zonation maps of the flood risk are therefore obtained, which are compared with the flood risk zonation maps obtained by the method based on index weighting during the same period.

### Modeling steps

In order to verify that LSTM can be used in flood risk assessment and zonation, an intelligent prediction model of flood risk in HRB based on LSTM is constructed through the following steps in accordance with the flow chart (Fig. 1).

1. Sample selection. There are 28 flood risk zonation maps in chronological order during 1988–2015. After resampling operation, the grid layers are output as numerical matrices with the resolution of 2000 m × 2000 m, each matrix represents the flood risk grid layer of each year, and each element of the matrix has both geographical coordinate position information and flood risk value, and each matrix has 66,793 elements (sample points).
2. Model construction. The intelligent prediction models of flood risk of HRB are constructed by LSTM. The flood risk value at each sample point from 1988 to 2010 is taken as the model training set, and the flood risk value during 2011–2015 is used as the model testing set, so a total of 66,793 models need to be established.
3. Model training. Given that the number of initial random seed is 1 and the number of training steps is 500, 66,793 intelligent flood risk prediction models constructed in step (2) are trained respectively. Particle swarm optimization algorithm is used to find the optimal parameters (segmentation window length, state vector size and learning rate), so that the root mean square error of the model testing set $RMSE = \sqrt{\sum_{i=1}^{n} \frac{(y_i - y_i')^2}{n}}$ is minimized, where $y_i$ and $y_i'$ are the true values of the sample and the outputs of the model respectively, $n$ is the

**(a)  2016**                                    **(b)  2017**
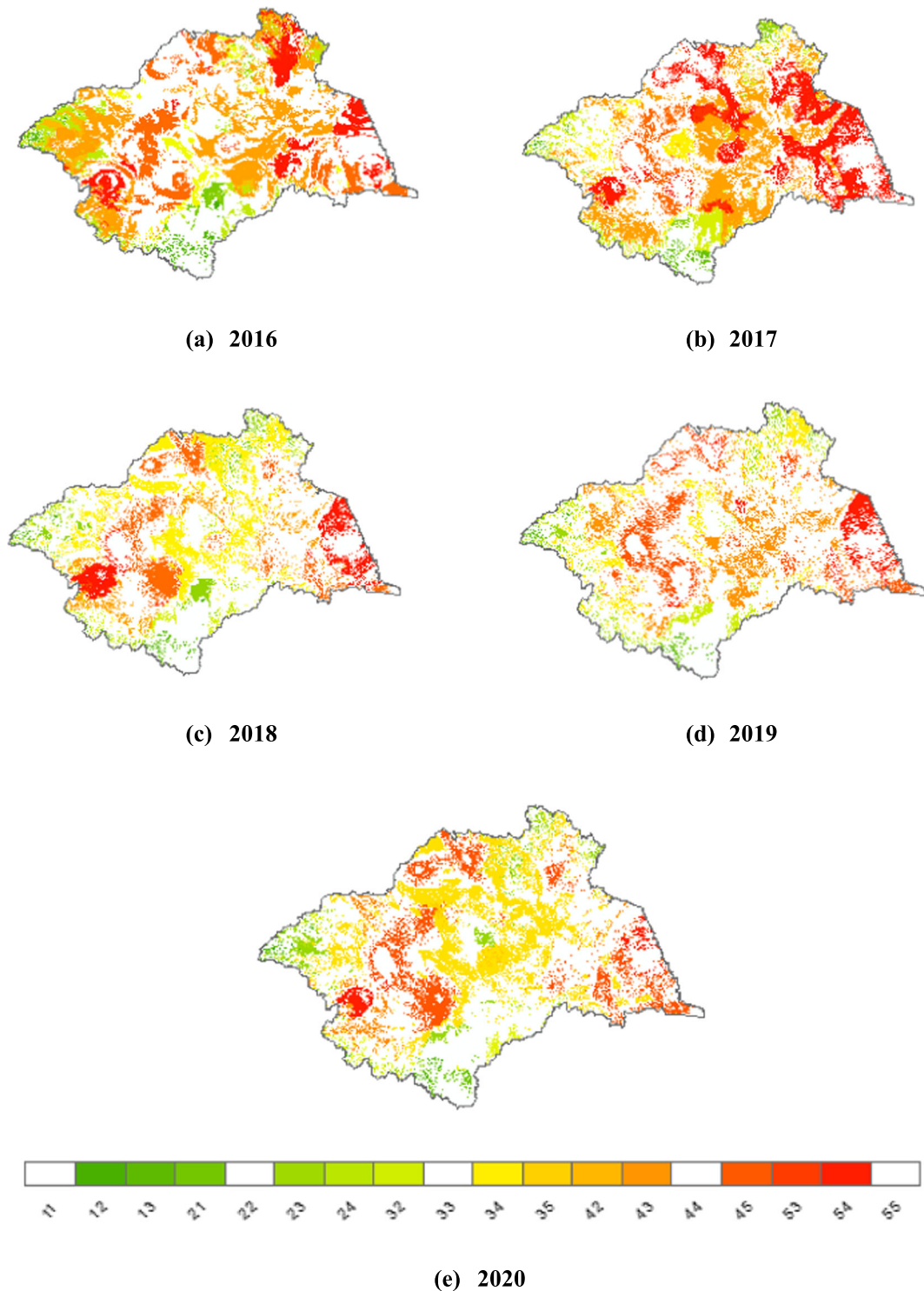
**(c)  2018**                                    **(d)  2019**

**(e)  2020**

**Fig. 11** Distribution of difference points between index weighting method and LSTM method: (a) 2016, (b) 2017, (c) 2018, (d) 2019, (e) 2020

**Table 5** Statistical table of difference points

| Year | Ratio of the difference points (%) | Ratio of first level difference points to the total (%) | Ratio of overestimation of risk to underestimation of risk by LSTM |
|------|------|------|------|
| 2016 | 46.33 | 93.73 | 0.40 |
| 2017 | 46.62 | 95.56 | 0.17 |
| 2018 | 34.40 | 98.29 | 2.05 |
| 2019 | 27.50 | 99.71 | 0.92 |
| 2020 | 35.74 | 98.68 | 5.34 |

Note: first level difference points are 12, 21, 23, 32, 34, 43, 45, and 54; second level and above are 24, 31, 35, 41, 42, 52, and 53; the overestimated risk by LSTM refers to 12, 13, 23, 24, 34, 35, and 45; underestimated risk refers to 21, 31, 32, 42, 42, 42, 52, 53, and 54

length of the testing set, and the model training is completed after meeting the accuracy requirements.

4. Risk prediction. The trained model is used to predict the flood risk value of each point in the future, the risk matrix is therefore formed, and then it is converted to the grid layer. Finally, the future flood risk zonation map of the basin is obtained by using the natural breaks classification method.

### Results display and comparison between prediction results

The average *RMSE* of the 66,793 models during the training period is 0.000964, and the average *RMSE* during the testing period is 0.004601, which both meet the accuracy requirements. Therefore, LSTM is used to predict the flood risk of the HRB in the short term (2016–2020), and the difference between the two kinds of risk maps is compared and analyzed based on the risk map predicted by the index weighting method at the same time in the "Flood risk assessment and prediction of the HRB based on the game theory" section. The steps are as below. (1) The minimum to maximum flood risks of the two flood risk maps are marked as 1 to 5, respectively. (2) The flood risk map based on index weighting is multiplied by 10 by using the Raster Calculator tool in ArcGIS. (3) The flood risk zonation map predicted by LSTM is superimposed on the result of step (2), and the distribution of difference points between index superposition and LSTM

classification results during 2016–2020 is obtained (see Fig. 11), and the difference points are counted in Table 5.

In Fig. 11 and Table 5, the first number of each difference point is index weighting flood risk level, and the second number is LSTM flood risk level. As shown in Fig. 11, "12" represents "index weighting method as very low flood risk, but LSTM method as low flood risk", "35" represents "index weighting method as moderate flood risk, but LSTM method as very high flood risk", and so on.

It can be seen from the table that the ratio of the difference points of 5 years are 38% on average, and the difference points are mainly concentrated on the first level, and the ratio of the points that are risk-overestimated to the points that are risk-underestimated by LSTM has no obvious rule.

As can be seen from Fig. 11, the most dangerous situations for which the risk is underestimated are 41, 42, 52, and 53, i.e., the high and very high flood risk are classified as very low risk, which account for only 0.98% of the area (5-year average). The overestimation of the risk is a conservative scheme to some extent, and considering the flood control safety of the basin, the scheme is more often used in practice. However, for 24 and 35, the low level is classified as the high level, and for this false early warning, the two situations are more wasteful of human and material resources, which account for only 0.22% of the area (5-year average).

In summary, LSTM has good prediction ability and the classification results are reasonable. The model establishment and the making of risk zonation map have good application value for flood disaster decision and analysis, so that the corresponding flood control and early warning scheme can be worked out in time for areas with high flood risk level.

### Discussion

Using LSTM to simulate flood risk in the HRB has obvious advantages, and there is no need to collect index information; the calculation is simple and the effect is good, so is the extension effect. The flood risk zonation map predicted by risk value training is basically consistent with the predicted map based on index superposition, which can be used for short-term prediction of flood risk of the basin in the

future. The error may be produced for the reasons: (1) Because of the short series of 28 years, we adopt the method of replacing time with space, which may lead to insufficient mining of data features so that the recognition accuracy is not high and may not give full play to the real advantages of LSTM. (2) The LSTM parameters are not optimal. (3) The accuracy of the predicted value of risk influencing index is not good enough.

## Conclusion

According to the theory of the natural disaster system of the river basin, 11 evaluation indexes are selected on the basis of comprehensive consideration of disaster-causing factors, disaster environment, disaster-bearing body, and disaster prevention and mitigation capability. LSTM is introduced into the field of flood risk assessment. Taking the HRB, a typical area of climate transition, as the research area, a flood risk assessment, and prediction model based on LSTM, is constructed. The main conclusions are as follows:

(1) Slope, virtual sown area of crops, and elevation are important influencing factors of flood risk in HRB.
(2) The LSTM model does not need to set the index weight in advance, and the data mining ability is strong, which greatly simplifies the operation process. The results show that the prediction effect of the model is good and the area with two or more levels of risk deviation only accounts for 1.2%, so LSTM can be used for short-term prediction of flood risk in the HRB.
(3) The Lixiahe area, Xixian county, and Bantai-Wangjiaba interval will be in the high flood risk area in the short term in the future, and the Dabieshan mountain area, Funiushan mountain area, and Tongbaishan mountain area will be in the low flood risk area.

This paper verifies the applicability of LSTM in the field of flood risk prediction and extends the application of deep learning technology. Based on the present work, further research can be carried out, such as seeking more effective parameter optimization methods from many LSTM model parameters, selecting more research areas with different climate

types, carrying out correction research of the flood risk level prediction and so on.

Based on the historical data, the prediction model is established by using the data-driving technology in reverse. In the next step, the reliability prediction method can be studied by using the extracted key features and elements from the relevant domain knowledge. In addition, because of the wide coverage area of the HRB, the hydrological stations are few and distribute unevenly, and the runoff data is difficult to obtain, the runoff factors are not taken into account when the evaluation index is selected. The hydrological model can be coupled to simulate the runoff data, so as to reduce the uncertainty of flood risk assessment caused by the incomplete index system.

## References

Apel, H., Thieken, A. H., Merz, B., & Blöschl, G. (2004). Flood risk assessment and associated uncertainty. *Natural Hazards and Earth System Science, 4*(2), 295–308.

Badrzadeh, H., Sarukkalige, R., & Jayawardena, A. W. (2015). Hourly runoff forecasting for flood risk management: application of various computational intelligence models. *Journal of Hydrology, 529*, 1633–1643.

Bonn, F., & Dixon, R. (2005). Monitoring flood extent and forecasting excess runoff risk with RADARSAT-1 data. *Natural Hazards, 35*(3), 377–393.

Box, G. E., & Jenkins, G. M. (1976). *Time series analysis: forecasting and control San Francisco*. Calif: Holden-Day.

Budiyono, Y., Aerts, J., Brinkman, J., Marfai, M. A., & Ward, P. (2015). Flood risk assessment for delta mega-cities: a case study of Jakarta. *Natural Hazards, 75*(1), 389–413.

Castillo-Rodriguez, J. T., Escuder-Bueno, I., Altarejos-Garcia, L., & Serrano-Lombillo, A. J. (2014). The value of integrating information from multiple hazards for flood risk analysis and

management. *Natural Hazards and Earth System Sciences, 14*, 379–400.

Chen, S. Y., & Guo, Y. (2006). Variable fuzzy sets and its application in comprehensive risk evaluation for flood-control engineering system. *Fuzzy Optimization and Decision Making, 5*(2), 153–162.

Chen, S., Luo, Z. K., & Pan, X. B. (2013). Natural disasters in China: 1900-2011. *Natural Hazards, 69*(3), 1597–1605.

Deng, W. P. (2013). *Research on flood disaster comprehensive evaluation model based on intelligent algorithm*. Huazhong University of Science and Technology (in Chinese).

Diakoulaki, D., Mavrotas, G., & Papayannakis, L. (1995). Determining objective weights in multiple criteria problems: the critic method. *Computers & Operations Research, 22*(7), 763–770.

Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research, 270*(2), 654–669.

Foudi, S., Osés-Eraso, N., & Tamayo, I. (2015). Integrated spatial flood risk assessment: the case of Zaragoza. *Land Use Policy, 42*, 278–292.

Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks, 18*(5–6), 602–610.

Hall, J. W., Sayers, P. B., & Dawson, R. J. (2005). National-scale assessment of current and future flood risk in England and Wales. *Natural Hazards, 36*(1–2), 147–164.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780.

Holmes, E. E., Lewis, M. A., Banks, J. E., & Veit, R. R. (1994). Partial differential equations in ecology: spatial interactions and population dynamics. *Ecology, 75*(1), 17–29.

Huang, Z. W., Zhou, J. Z., Song, L. X., Lu, Y. L., & Zhang, Y. C. (2010). Flood disaster loss comprehensive evaluation model based on optimization support vector machine. *Expert Systems with Applications, 37*(5), 3810–3814.

Jiang, B. (2013). Head/tail breaks: a new classification scheme for data with a heavy-tailed distribution. *The Professional Geographer, 65*(3), 482–494.

Khosravi, K., Pourghasemi, H. R., Chapi, K., & Bahri, M. (2016). Flash flood susceptibility analysis and its mapping using different bivariate models in Iran: a comparison between Shannon's entropy, statistical index, and weighting factor models. *Environmental Monitoring and Assessment, 188*(656), 656.

Kubal, C., Haase, D., Meyer, V., & Scheuer, S. (2009). Integrated urban flood risk assessment - adapting a multicriteria approach to a city. *Natural Hazards & Earth System Sciences, 9*(6), 1881–1895.

Lai, C. G., Wang, Z. L., & Song, H. J. (2011). Flood risk assessment of Beijiang River basin based on BP neural network. *Hydropower Energy Science, 29*(3), 57–59 (in Chinese).

Lai, C. G., Chen, X. H., Chen, X. Y., Wang, Z. L., Wu, X. S., & Zhao, S. W. (2015). A fuzzy comprehensive evaluation model for flood risk based on the combination weight of game theory. *Natural Hazards, 77*(2), 1243–1259.

Li, L. F., Wang, J. F., Leung, H. T., & Jiang, C. S. (2010). Assessment of catastrophic risk using Bayesian network constructed from domain knowledge and spatial data. *Risk Analysis, 30*(7), 1157–1175.

Li, G. F., Xiang, X. Y., Tong, Y. Y., & Wang, H. M. (2013). Impact assessment of urbanization on flood risk in the Yangtze River Delta. *Stochastic Environmental Research and Risk Assessment, 27*(7), 1683–1693.

Liu, J. F., Li, J., Liu, J., & Cao, R. Y. (2008). Integrated GIS/AHP-based flood risk assessment: a case study of Huaihe River Basin in China. *Journal of Natural Disasters, 17*(6), 110–114 (in Chinese).

Liu, J., Jiang, W. G., Du, P. J., Sheng, S. X., Liu, J. F., & Cao, R. Y. (2009). Rainstorm risk assessment of Huaihe River based on correlation analysis. *Journal of China University of Mining and Technology, 38*(5), 735–740 (in Chinese).

Lu, G. Y., & Wong, D. W. (2008). An adaptive inverse-distance weighting spatial interpolation technique. *Computers & Geosciences, 34*(9), 1044–1055.

Lyu, H. M., Sun, W. J., Shen, S. L., & Arulrajah, A. (2018). Flood risk assessment in metro systems of mega-cities using a GIS-based modeling approach. *Science of the Total Environment, 626*, 1012–1025.

Myerson, R. B. (2013). *Game theory*. Harvard University Press.

Ouma, Y. O., & Tateishi, R. (2014). Urban flood vulnerability and risk mapping using integrated multi-parametric AHP and GIS: methodological overview and case study assessment. *Water, 6*(6), 1515–1545.

Peng, S. H. (2018). Preparation of a flood-risk environmental index: case study of eight townships in Changhua County, Taiwan. *Environmental Monitoring and Assessment, 190*(174), 174.

Sheather, S. J., & Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society: Series B (Methodological), 53*(3), 683–690.

Sheng, S. X., Shi, L., Liu, J. F., Ye, J. Y., & Liu, J. (2010). Risk assessment of heavy rain and flood in the depression area along the Huaihe river. *Geographical Research, 29*(03), 416–422 (in Chinese).

Tingsanchali, T., & Karim, F. (2010). Flood-hazard assessment and risk-based zoning of a tropical flood plain: case study of the Yom River, Thailand. *International Association of Scientific Hydrology Bulletin, 55*(2), 145–161.

Wang, Z. L., Chen, X. H., Lai, C. G., & Zhao, S. W. (2013). Flood risk assessment model based on particle swarm rule mining algorithm. *System Engineering Theory and Practice, 33*(6), 1615–1621 (in Chinese).

Wang, Z. L., Lai, C. G., Chen, X. H., Yang, B., Zhao, S. W., & Bai, X. Y. (2015). Flood hazard risk assessment model based on random forest. *Journal of Hydrology, 527*, 1130–1141.

Wu, Y. N., Zhong, P. A., Zhang, Y., Xu, B., Ma, B., & Yan, K. (2015). Integrated flood risk assessment and zonation method: a case study in Huaihe River basin, China. *Natural Hazards, 78*(1), 635–651.

Wu, Y. N., Zhong, P. A., Xu, B., Zhu, F. L., & Ma, B. (2017). Changing of flood risk due to climate and development in Huaihe River basin, China. *Stochastic Environmental Research and Risk Assessment, 31*(4), 935–948.

Xiao, Y. F., Yi, S. Z., & Tang, Z. Q. (2017). Integrated flood hazard assessment based on spatial ordered weighted averaging method considering spatial heterogeneity of risk preference. *Science of the Total Environment, 599-600*, 1034–1046.

Yang, H., & Li, C. Y. (2003). The relation between atmospheric intraseasonal oscillation and summer severe flood and drought in the Changjiang-Huaihe River basin. *Advances in Atmospheric Sciences, 20*(4), 540–553.

Yang, C. G., Yu, Z. B., Hao, Z. C., Zhang, J. Y., & Zhu, J. T. (2012). Impact of climate change on flood and drought events in Huaihe River Basin, China. *Hydrology Research, 43*(1–2), 14–22.

Yin, J., Ye, M. W., Yin, Z., & Xu, S. Y. (2015). A review of advances in urban flood risk analysis over China. *Stochastic Environmental Research and Risk Assessment, 29*(3), 1063–1070.

Zhang, Y. J., & Zhang, X. (2007). Grey correlation analysis between strength of slag cement and particle fractions of slag powder. *Cement and Concrete Composites, 29*(6), 498–504.

Zhang, Z. T., Gao, C., Liu, Q., Zhai, J. Q., Wang, Y. J., Su, B. D., & Tian, H. (2014). Risk assessment of storm floods and floods in the Huaihe river basin in different return periods. *Geographical Research, 33*(07), 1361–1372 (in Chinese).

Zhou, Q., Mikkelsen, P. S., Halsnæs, K., & Arnbjerg-Nielsen, K. (2012). Framework for economic pluvial flood risk assessment considering climate change effects and adaptation benefits. *Journal of Hydrology, 414-415*, 539–549.

Zou, Q., Zhou, J. Z., Zhou, C., Song, L. X., & Guo, J. (2013). Comprehensive flood risk assessment based on set pair analysis-variable fuzzy sets model and fuzzy AHP. *Stochastic Environmental Research and Risk Assessment, 27*(2), 525–546.