

## Sensitivity Analysis of Mountain Flash Flood Risk: Case of Licheng County in China

Haofang Wang<sup>1</sup> Jincun Zhang<sup>1</sup> Ruan Yun<sup>1</sup> Ranhang Zhao<sup>1\*</sup>

<sup>1</sup> School of Civil Engineering, Shandong University, Jinan 250061, P.R., China;  
whf29@sdu.edu.cn; zhangjincun@sdu.edu.cn;  
yunruan@sdu.edu.cn; ranhang-z@sdu.edu.cn

\* Correspondence: ranhang-z@sdu.edu.cn; Tel.: +8613705316167

**Abstract:** Flash flood is one of the most significant natural disasters in China, particularly in mountainous area, causing heavy economic damage and casualties of life. Accurate risk assessment is critical to an efficient flash flood management. There are more than 530,000 small watersheds in 2058 counties in China where flash flood should be prevented. In practice, with limited fund and different risk levels, the priorities of each small watershed for flash flood prevention and control are also needed for an efficient flash flood management. This paper, take Licheng county in China as an example, aims to give out these priorities for management. First, sensitive indexes are identified among index system, which includes 9 indexes based on underlying surface characteristics of small watershed in hilly region. Second, the range of each index and the rank division of each index for evaluation are determined. Based on the rank divisions for evaluation, the flash flood risk grade eigenvalue (H) is calculated by Variable Fuzzy Method (VFM) using 1000 samplings generated by Latin hypercube sampling method. Third, the key sensitivity factors that affect flash flood risk grade eigenvalue (H) are assessed by two different global sensitivity analysis methods -- stepwise regression analysis and mutual entropy. Both results indicate that watershed slope (S) is the most sensitive factor; the second is antecedent precipitation index (CN); while other factors are slightly different sensitive in sequence. This study shows that stepwise regression analysis and mutual information analysis are appropriate for the sensitivity analysis of mountain flash flood risk. Finally, based on watershed slope (S), the priorities of flash flood prevention and control of 119 small watersheds in Licheng county are given out.

**Keywords:** sensitive analysis; variable fuzzy method; mutual entropy; stepwise regression analysis; mountain flash flood risk

## 1. Introduction

Mountain flash flood can cause extensive disruptions to property, infrastructure and human lives. The nature of the mountain flash flood-producing is the interaction of atmosphere, land geology and geomorphology, vegetation and soils and the activities of people [1-4]. The atmosphere affects the rainfall characteristics whether it is short-duration highly intense rainfall or long-duration low intensity rainfall. Other factors such as land geology and geomorphology, vegetation and soils and the activities of people (hereinafter refers as underlying surface) mainly affect flood runoff production and flow routing. In the same climate conditions, even the same rainfall characteristics, different underlying surfaces will cause different runoff productions and flow routings, which will lead to different flood risk. In China, there are more than 530,000 small watersheds in 2058 counties where flash flood should be prevented in China [5]. For many small watersheds, especially ungauged or sparse gauged, flash floods occur often and cause extensive damage to property, infrastructure and human lives. Each small watershed has different underlying surface characteristics. Because the underlying surfaces are complex giant systems, the impact of each underlying surface factor on flash flood risk may have a relationship of linear or nonlinear with randomness and fuzziness. In order to reveal the most key influencing factors, proper sensitivity analysis will be needed. There are two kinds of sensitivity analysis methods which are local sensitivity analysis and global sensitivity analysis [6-7]. Local sensitivity analysis is by evaluating how the outputs change by varying one input variable at one time. However, the global sensitivity analysis ( GSA ) is by evaluating all variable's contribution to the outputs change by varying all input variables simultaneously[8]. In this study, GSA is used. There are many GSA methods, such as scatter plots analysis applied to a microparasite and macroparasite transmission model[9-10], stepwise regression analysis applied to groundwater model[11,12,15,18], Latin hypercube sampling with partial rank correlation coefficient index (LHS-PRCC) used to urban drainage modeling or other models [8,9,13], standardized regression coefficients and Morris screening method applied to sewer flow and water quality modelling and others[14,18], mutual information method widely applied to the probability distribution of groundwater level series[12], Storm Water Management Model[8], ground water models[15], neural network models[16] , flood inundation models[17], and others[19-20 ]. Although many GSA

methods have been proposed in many fields of science and engineering [21-25], there are no reports in mountain flash flood risk evaluation as we know. In this study, mutual entropy is applied to sensitivity analysis of flash flood risk evaluation. Mutual entropy is a good method to delineate nonlinear and complex multivariable relationship. In order to verify the rationality of mutual entropy, stepwise regression analysis is selected to compare with.

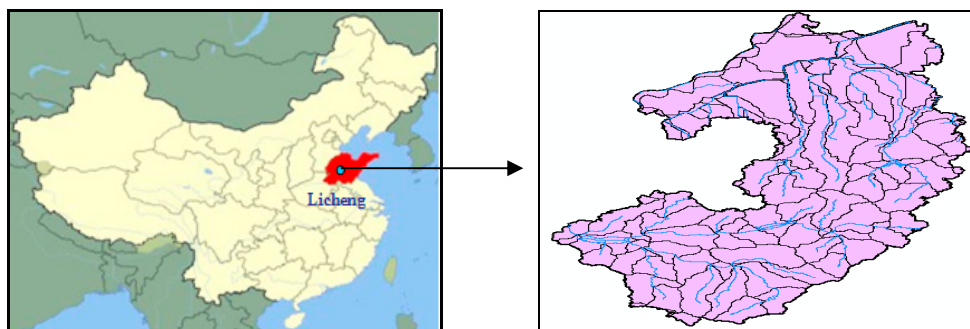
First, an index system for flash flood risk evaluation is constructed. Owing to the impact of each index on flash flood risk with a linear or nonlinear relationship which has randomness and fuzziness, variable fuzzy method (VFM) is adopted to evaluate flash flood risk. VFM, developed by Chen Shouyu in 2005 [26-27], is an effective evaluation method for fuzzy system with linear or nonlinear relationship, and is widely used in water resources evaluation [27-28], water environment evaluation [29-31], flood control operations [32], land suitability evaluation [37], water quality assessment [38-40] and flood risk assessment [41]. The purpose of this study is to find out the sensitivity factors for flash flood risk evaluation and determine the priorities of flash flood prevention and control of numerous small watersheds in hilly regions.

## 2 Materials and Methods

### 2.1 Study Site

Licheng county, located at the transition zone between mountainous areas in the middle of Shandong province, China, has a area  $1298 \text{ km}^2$  with  $E116^{\circ}49' \sim 117^{\circ}29'$  and  $N36^{\circ}20' \sim 36^{\circ}62'$ . As shown in Fig1. Licheng county belongs to the warm-temperate zone, monsoon climate with four different seasons. With an uneven precipitation distribution in space, the average annual precipitation is 665.7 mm, which take place mainly in 6 ~ 9 month accounting for 66% of the annual rainfall. The geomorphic types are mainly three parts from south to north: mountains, hills and plains. The elevation is between 20 m and 975 m. The mountain and hill area in south accounts for 67% of total area. Licheng county belongs to flash flood prone region and the one of the pilot counties of flash flood prevention in China [57]. With the increase of extreme weather events, especially in recent years, flash floods often lead to serious disasters. According to statistics, big flash flood disasters happened in the year of 1963, 1964, 1985, 1987, 1994, 1996, 2000 and 2010, with seriously destroyed

bridges and roads, inundated farmland and buildings, and a direct economic loss of about 99.1 million RMB[42-43]. Flash flood disaster has become a prominent problem and a restraining factor to Licheng's sustainable development. According to the survey of the flash flood disaster, there are 119 small watersheds listed as flash flood disaster prevention in Licheng county shown in Fig1. In order to prevent and control flash flood for 119 small watersheds in an orderly way, sensitivity analysis is needed based on GSA to find sensitive factors that influencing flash flood risk.



**Figure 1.** Location and small watersheds distribution in Licheng county

## 2.2 Index selection

As we known, mountain flash flood is caused by short-duration high-intense rainfall. For same rainfall conditions, mountain flash flood is mainly affected by watershed's underlying surface conditions. In this study, watershed's underlying surface conditions affecting flash flood are considered only.

In small hilly watersheds, because of small watershed area, low stream density, steeper watershed slope, steeper channel slopes and short flow routing time, flash floods are characterized by fast flood rise and fall, large peak flow, and small runoff volume. Peak flow is the main consideration in determining the scale of water conservancy projects and flash flood warning projects. So the index selection mainly considers the impact on the flood peak.

The main runoff production factors are watershed area and antecedent precipitation index.

(1) Watershed area (F). For the same rainfall and underlying surface conditions, the bigger the watershed area is, the larger the watershed runoff.

(2) Antecedent precipitation index (API). API symbolizes the soil moisture before a rainfall event begins. With different API, same rainfall pattern will lead to different runoff volume. The higher API is, the bigger runoff volume. API is mainly affected by

soil water content and evaporation. For same evaporation condition, the soil water content is affected by land use type, soil type, hydrological management level and so on. In this paper, API is described by CN of the SCS model. CN is a dimensionless value influenced by soil types, land use types and hydrological management level.

The main factors influencing flow routing are watershed shape factor, watershed slope, longest flow path length, slope of the longest flow path, main stream length and slope of main stream, and width of main stream.

(1) Watershed shape factor ( $K_e$ ).  $K_e$  is the ratio of the average width to the length of watershed. For the same watershed area and same storm rain, different watershed shapes can lead to different discharge curve on watershed outlet. For fan-shaped watershed, it is easy to produce sharp and thin discharge curve. For long and narrow shape watershed, it is easy to produce short and fat discharge curve.

(2) Watershed slope ( $S$ ).  $S$  influences overland flow. For small watersheds in hilly regions, overland flow is the main part of total flow. The bigger the watershed slope is, the faster the overland flow velocity, then the shorter routing time and the shorter peak time is formed.

(3) Longest flow path length ( $L_1$ ).  $L_1$  refers to the longest distance from watershed outlet along the water flow to watershed divide, by which the flow routing time is determined. The shorter the longest flow path length is, the shorter flow concentration time, then the flood peak time is earlier.

(4) Slope of the longest flow path ( $J_1$ ).  $J_1$  is the average slope of the longest flow path. The greater the gradient is, the greater the flow velocity. So the flow routing time becomes shorter. Then it leads to a faster flash flood rising and pushes a greater pressure on flash flood warning.

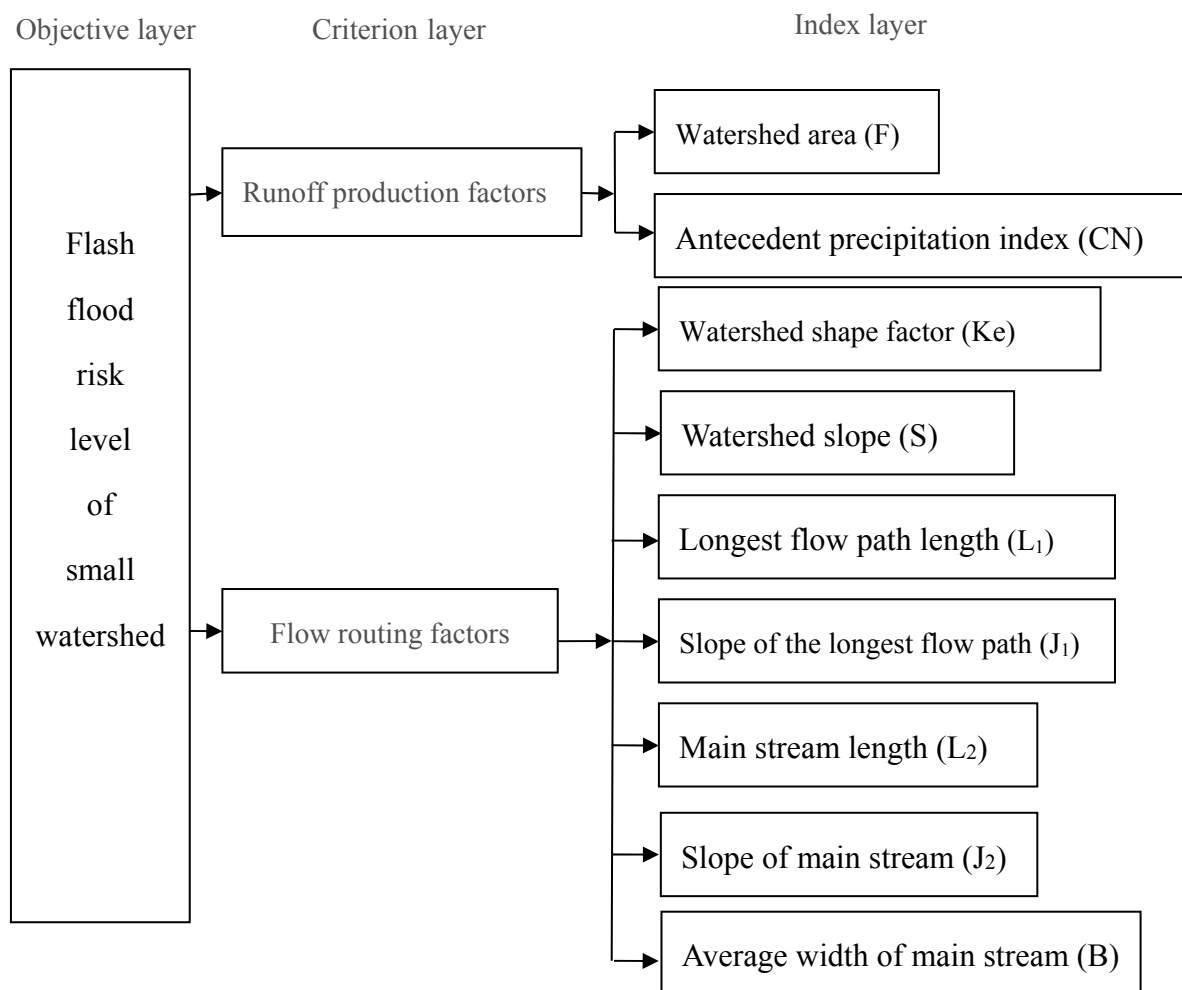
(5) Main stream length ( $L_2$ ).  $L_2$  influences stream flow routing. The longer the stream length is, the longer the stream flow routing time.

(6) Slope of main stream ( $J_2$ ).  $J_2$  influences stream flow routing. The bigger the slope is, the shorter the stream flow routing time.

(7) Average width of main stream ( $B$ ):  $B$  influences river flow capacity, the wider river, the stronger flow ability, then the peak flow is small. The opportunities of overbank flow become less. The risk of flash flood disasters also reduce. Therefore, the average width of main stream is one of the main evaluation indicators of flash flood.

From what has been discussed above, the index system for flash flood risk

evaluation is built and shown in Fig.2. The evaluation index system includes objective layer, criterion layer and index layer. The objective layer represents flash flood risk level of small watershed. The criterion layer is the factors of runoff production and flow routing. The index layer is the main factors influencing runoff production and flow routing of small watershed. The range of each index is ascertained based on the field measurement data of small watersheds in flash flood disaster prevention and control regions of Shandong Province showed in Table 1.



**Figure 2.** Evaluation index system of flash flood risk in small watershed

**Table 1** The range of each index

No.	Index	Range
x1	F(km <sup>2</sup> )	0.2-50
x2	Ke	0.01-0.58
x3	S	0.0002-0.645
x4	CN	30-90
x5	L1(km)	0.096-31.9
x6	J1(‰)	0-107
x7	L2(km)	0.4-26
x8	J2(‰)	0-46
x9	B(m)	2-72

### 2.3 Flash Flood Risk Evaluation Method

Based on the concept of fuzzy sets depicting imprecision or vagueness of Zadeh, Chen Shouyu established a fuzzy set theory and a engineering fuzzy set theory [33], and Variable fuzzy method (VFM) developed in 2005 [26-27]. VFM can scientifically and reasonably determine membership degrees and relative membership functions of objectives or indexes at level interval relating to fuzzy system [35-36]. So in this paper, VFM is applied to evaluate the flash flood risk of small watersheds in hilly regions. The detail calculation process is as follows[30,33,35,36]:

Suppose the object  $j$  ( $j = 1, 2, \dots, n$ ) has  $i$  ( $i = 1, 2, \dots, m$ ) indices, and index eigenvalues  $x_{ij}$  as follows:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} = (x_{ij}) \quad (1)$$

The indices can be evaluated by  $c$  grading levels, and the criteria interval matrices of  $m$  indices and  $c$  levels can be expressed as follows :

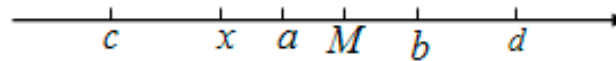
$$I_{ab} = \begin{bmatrix} [a_{11}, b_{11}] & [a_{12}, b_{12}] & \cdots & [a_{1c}, b_{1c}] \\ [a_{21}, b_{21}] & [a_{22}, b_{22}] & \cdots & [a_{2c}, b_{2c}] \\ \vdots & \vdots & \vdots & \vdots \\ [a_{m1}, b_{m1}] & [a_{m2}, b_{m2}] & \cdots & [a_{mc}, b_{mc}] \end{bmatrix} = ([a_{ih}, b_{ih}]) \quad (2)$$

Where  $i = 1, 2, \dots, m; h = 1, 2, \dots, c$ . Level 1 means the superior level and level  $c$  means the inferior level. If the index is the larger the better, then  $a > b$ . If the index is the smaller the better, then  $a < b$ . For every  $[a_{ih}, b_{ih}]$ , the range of interval  $[c_{ih}, d_{ih}]$  can

be determined according to the upper and lower bound of its adjacent intervals [30,33,35,36].

$$I_{cd} = \begin{bmatrix} [c_{11}, d_{11}] & [c_{12}, d_{12}] & \cdots & [c_{1c}, d_{1c}] \\ [c_{21}, d_{21}] & [c_{22}, d_{22}] & \cdots & [c_{2c}, d_{2c}] \\ \vdots & \vdots & \vdots & \vdots \\ [c_{m1}, d_{m1}] & [c_{m2}, d_{m2}] & \cdots & [c_{mc}, d_{mc}] \end{bmatrix} = ([c_{ih}, d_{ih}]) \quad (3)$$

Suppose  $X_0 \in [a, b]$  belongs to attracting sets, i.e.  $0 < D_A(u) \leq 1$ . Then  $X \in [c, d]$  belongs to the upper bound and lower bound of a certain range which includes  $X_0$  ( $X_0 \subset X$ ). As shown in Fig.3.



**Figure 3.** The position relationship between variable  $x$ ,  $M$  and interval  $X_0, X$

According to the definition of  $V$ , it is known that  $[c, a]$  and  $[b, d]$  are all repelling sets, i.e.  $-1 \leq D_A(u) < 0$ . Suppose that  $M$  is a point value belonging to the interval  $[a, b]$  and it is relative difference degree  $D_A(u) = 1$ . So it is relative membership degree  $\mu_A(u) = 1$ . Then  $M$  is the matrix of all the point value belonging to  $\mu_A(x_{ij})_h = 1$ .

$M_{ih}$  is an important parameter and can be obtained according to the criteria eigenvalue  $a_{ih}$  and  $b_{ih}$  as follows:

$$M = \begin{bmatrix} M_{11} & M_{12} & \cdots & M_{1c} \\ M_{21} & M_{22} & \cdots & M_{2c} \\ \vdots & \vdots & \vdots & \vdots \\ M_{m1} & M_{m2} & \cdots & M_{mc} \end{bmatrix} = (M_{ih}) \quad (4)$$

$$M_{ih} = \frac{c-h}{c-1} a_{ih} + \frac{h-1}{c-1} b_{ih} \quad (5)$$

Eq. (1) satisfies the following three special conditions:

- (1) When  $h = 1$ , then  $M_{i1} = a_{i1}$ ; (2) when  $h = c$ , then  $M_{ic} = b_{ic}$ ; (3) when  $h = \frac{c+1}{2}$



then  $M_{ih} = \frac{a_{ih} + b_{ih}}{2}$ . If  $x_{ij}$  belongs to the interval  $[a_{ih}, b_{ih}]$ , and  $x_{ij}$  is located at the left of  $M_{ih}$ , the relative membership degree of  $x_{ij}$  to level  $h$  can be calculated as follows :

$$\mu_A(x_{ij})_h = \begin{cases} 0.5(1 + \frac{x_{ij} - a_{ih}}{M_{ih} - a_{ih}}) & x_{ij} \in [a_{ih}, M_{ih}] \\ 0.5(1 - \frac{x_{ij} - a_{ih}}{c_{ih} - a_{ih}}) & x_{ij} \in [c_{ih}, a_{ih}] \end{cases} \quad (6)$$

On the other hand, if  $x_{ij}$  belongs to the interval  $[a_{ih}, b_{ih}]$ , and  $x_{ij}$  is located at the right of  $M_{ih}$ , the relative membership degree of  $x_{ij}$  to level  $h$  can be calculated as follows [33,35,36]:

$$\mu_A(x_{ij})_h = \begin{cases} 0.5(1 + \frac{x_{ij} - b_{ih}}{M_{ih} - b_{ih}}) & x_{ij} \in [M_{ih}, b_{ih}] \\ 0.5(1 - \frac{x_{ij} - b_{ih}}{d_{ih} - b_{ih}}) & x_{ij} \in [b_{ih}, d_{ih}] \end{cases} \quad (7)$$

$$\mu_A(x_{ij})_h = 0 \quad x_{ij} \notin [c_{ih}, d_{ih}] \quad (8)$$

Finally, if  $x_{ij}$  does belong to the interval  $[c_{ih}, d_{ih}]$ , then  $D_A(u) = -1$ ,  $\mu_A(x_{ij})_h = 0$ .

Eq. (6) and Eq. (7) satisfy the following special conditions:

- (1) When  $x_{ij} = a_{ih}$  or  $x_{ij} = b_{ih}$ , then  $D_A(u) = 0$ ,  $\mu_A(x_{ij})_h = 0.5$ ;
- (2) When  $x_{ij} = M_{ih}$ , then  $D_A(u) = 1$ ,  $\mu_A(x_{ij})_h = 1$ ;
- (3) When  $x_{ij} = c_{ih}$  or  $x_{ij} = d_{ih}$  then  $D_A(u) = -1$ ,  $\mu_A(x_{ij})_h = 0$ .

The matrix of the relative membership degree can be computed by equations (6–8).

$$U_j = [\mu_A(x_{ij})_h] \quad (9)$$

As such, a synthetic relative membership degree vector can be obtained by the variable fuzzy set theory.

$${}_j u'_h = \left\{ 1 + \left[ \frac{\sum_{i=1}^m [\omega_i (1 - \mu_A(x_{ij})_h)]^p}{\sum_{i=1}^m [\omega_i \mu_A(x_{ij})_h]^p} \right]^{\alpha/p} \right\}^{-1} \quad (10)$$

Where  $\omega$  is the weight of indices (the weight of each index represents its contribution to flash flood risk, it has a profound effect on the assessment results. In the sensitivity analysis process, we assume the weight of each index is same in order to objectively evaluate the sensitivity of each index to flash flood risk);  $\alpha$  is the rule parameter of optimization model ( $\alpha = 1$  for least single method and  $\alpha = 2$  for least square method); and  $p$  is distance parameter ( $p = 1$  denotes hamming distance and  $p = 2$  denotes Euclidean distance). There are four kinds of combinations for different  $\alpha$  and  $p$ .

By Eq. (10), the synthetic relative membership degree of each index about the grade level  $h$  is obtained:

$$U'_j = [{}_j u'_h] \quad (11)$$

After normalizing  ${}_j u'_h$ , the normalized synthetic relative membership degree of each index about the grade level  $h$  is obtained:

$$U_j = [{}_j u_h] \quad (12)$$

Where 
$${}_j u_h = \frac{{}_j u'_h}{\sum_{h=1}^c {}_j u'_h}$$

The grading level eigenvalues can be computed by Eq. (13), and the evaluation level can be determined according to the value of  $H$ .

$$H_j = [1 \quad 2 \quad \dots \quad c] U_j \quad (13)$$

In variable fuzzy evaluation process, evaluation rank division is the key step. In this study, owing to the different nature of each index, some indicators, such as CN, watershed slope, positively correlate with flash flood risk. However, some indicators such as longest flow path length negatively correlate with flash flood risk. In order to facilitate the flash flood risk evaluation, statistical analysis is carried on the following characteristic values of standard deviation, mean, median, the maximum and minimum values of each index according to field measurements data. Based on this statistical analysis and reference literatures [44-46], 3 risk levels are divided and they are high, medium and low. The characteristic values of each index for different risk levels are determined and shown in Table2.

**Table 2.** The characteristic value of each index in different risk levels

Risk level	Ch	F (km <sup>2</sup> )	Ke	S	CN	L <sub>1</sub> (km)	J <sub>1</sub> (‰)	L <sub>2</sub> (km)	J <sub>2</sub> (‰)	B (m)
high	1	30-50	0.29-0.58	0.359-0.645	70-90	0.096-8.52	45-107	0-6.5	19-46	2-15
medium	2	20-30	0.15-0.29	0.139-0.359	50-70	8.52-17.6	15-45	6.5-19	6.5-19	15-35
low	3	0-20	0.01-0.15	0.0002-0.139	30-50	17.6-31.9	0-15	19-26	0-6.5	35-72

Note: Ch= characteristic value

## 2.4 Sensitivity Analysis Methods

We applied Latin hypercube sampling, stepwise regression analysis and mutual entropy analysis in this case study.

### 2.4.1 Latin Hypercube Sampling

There are many random sampling methods such as Monte Carlo sampling [12,14,15], Latin hypercube sampling [8,10,22], quasi-random sequence sampling [47,49], cluster sampling [48,49] and so on. In this study, Latin hypercube sampling (LHS) was used to sample. The LHS was described by McKay in 1979 [50]. LHS can be applied to multiple variables and achieve a reasonably random distribution [51]. This method is simple and avoids the repeated sampling which makes the indexes uniformly distribute with the same probability trend. In this study, we used LHS to generate 1000 samples for each index based on the scope of each index in Table 1.

### 2.4.2 Stepwise regression analysis

Stepwise regression analysis is a commonly used sensitivity method which is simple, convenient and capable of describing the effect of individual input factors on outputs [15, 52]. Stepwise regression analysis is multiple regressions which includes three main steps:

Firstly, construct a simple regression model. Perform a correlation analysis between each of the input factors and an output variable, the variable with the most significant partial correlation coefficient is selected to build regression model, and then other variables are imported into the model one by one.

Secondly, perform significant test for all the variables in the model. Every time each variable is imported into model, a significant test of all the variables in the model is performed, When the previously imported variables become insignificant due to the

subsequent imported variable, the previously variables are removed from the model. And repeating this process until no further variables can be deleted.

Finally, construct multiple regressions model. The regression model should contain all the variables with the significant influence to output variable, and the variables outside of the model are insignificant.

The stepwise regression equation is:

$$y = b_0 + b_1x_1 + \dots + b_kx_k \quad (14)$$

Where  $y$  is fitted output,  $x_i$  is input variables,  $b_i$  is estimated coefficient which determined by stepwise regression model.

### 2.4.3 Mutual Entropy

Mutual entropy analysis is a global sensitivity analysis method which is based on the concept of entropy, solves possible non-monotonicity in input-output relationships by mutual information [15,53,54]. The mutual entropy of two variables is regarded as the correlative extent between these two variables and the distribution character of two variables can be expressed by contingency tables[55,56].The entropies of  $x$  and  $y$  can be defined as follows[8,15]:

$$H(x) = -\sum_{i=1}^n p(x_i) \ln p(x_i) \quad (15)$$

$$H(y) = -\sum_{j=1}^m p(y_j) \ln p(y_j) \quad (16)$$

The values of discrete random variable  $x$  take  $x_i (i = 1, 2, \dots, n)$ , and the values of discrete random variable  $y$  take  $y_j (j = 1, 2, \dots, m)$ .

The joint entropy  $H(x, y)$  of  $x$  and  $y$  denotes the average total information gained by observing both  $x$  and  $y$ . The joint entropy  $H(x, y)$  of  $x$  and  $y$  is defined as:

$$H(x, y) = -\sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \ln p(x_i, y_j) \quad (17)$$

The mutual information  $I(x, y)$  of  $x$  and  $y$  is the amount of information shared by  $x$  and  $y$ . The mutual information is zero if and only if  $x$  and  $y$  are independent. So the mutual information between the random variables  $x$  and  $y$  can be considered a

measure of dependence between these variables, or better yet, the statistical correlation of  $x$  and  $y$ .

$$I(x, y) = H(x) + H(y) - H(x, y) \quad (18)$$

In mutual entropy analysis, the sensitivity of the output to input variables is estimated by two indicators of uncertainty coefficient and R statistic [8,12].

#### 1. Uncertainty coefficient

Uncertainty coefficient is as a quantitative measure of association, defined by the two limits identified above.

$$U(x, y) = 2 \left[ \frac{I(x, y)}{H(x) + H(y)} \right] \quad (19)$$

Which varies the range [0,1]. 0 indicates  $x$  is independent of  $y$ , and 1 indicates  $x$  and  $y$  are completely dependent.

#### 2. R-statistic

R-statistic is association defined on the basis of mutual information. R can be used to provide a computational representation of variable importance.

$$R(x, y) = [1 - \exp\{-2I(x, y)\}]^{1/2} \quad (20)$$

Which R varies [0,1]. R is 0 if  $x$  and  $y$  are independent, and is 1 if  $x$  is completely related.

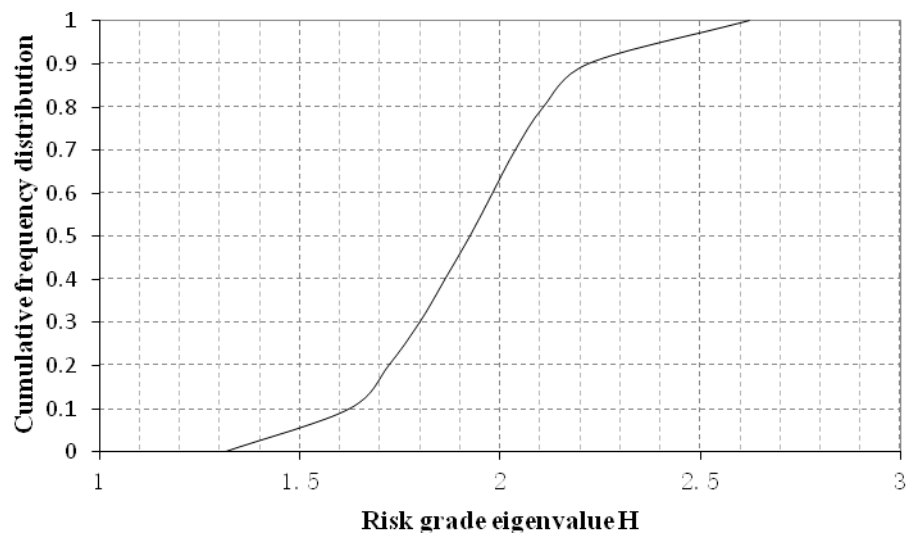
### 3 Results and Discussion

#### 3.1 Statistics Analysis of the Evaluation Results from Variable Fuzzy Method

Using 1000 samplings generated by Latin hypercube sampling, according to the evaluation rank division of each index, the flash flood risk grade eigenvalue (H) are calculated, as shown in Table 3. According to Table 3, the cumulative frequency distribution of H is calculated and presented in Figure 4. The statistical parameters for the distribution are given in Table 4. Figure 4 shows that high H has high empirical frequency for all indexes. Because these statistical parameters cannot be used to identify which index is the most contributive to the H, stepwise regression analysis and mutual entropy analysis are adopted here to identify key sensitive indexes.

**Table 3.** The risk grade eigenvalue (H) of 1000 samplings calculated by VFM

F (km <sup>2</sup> )	Ke	S	CN	L <sub>1</sub> (km)	J <sub>1</sub> (‰)	L <sub>2</sub> (km)	J <sub>2</sub> (‰)	B(m)	H
44.054	0.085	0.186	85.742	1.180	2.9	23.332	22.8	16.613	1.874
7.4067	0.235	0.318	50.848	24.582	13.2	11.566	38.7	42.738	2.195
38.013	0.470	0.619	56.181	7.750	2.8	3.154	7.9	54.838	1.788
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
5.583	0.292	0.615	82.466	30.332	106.7	20.758	29.9	9.067	1.702
25.324	0.190	0.342	32.078	5.837	46	1.273	0.3	28.498	1.987
34.15	0.167	0.137	69.131	6.651	91	13.388	11.9	34.917	1.878

**Figure 4.** Cumulative frequency distribution of risk grade eigenvalue (H)**Table 4.** Statistical parameters of risk grade eigenvalue H of 1000 samples

Statistical parameter	Risk grade eigenvalue H
Maximum	2.623
Minimum	1.313
Mean	1.920
Median	1.924
Variance	0.051
5th percentile	1.547
95th percentile	2.305

### 3.2 Scatter plots

Scatter plots can illustrate the relationship between input variable and output outcomes [9-10]. Because random sampling exit uncertainty and the evaluation results from Variable fuzzy method (VFM) also exit uncertainty, the relationship between each sampling and evaluation results which is linear, monotonic, and nonmonotonic

problems are needed to illustrate by scatter plots. Appropriate sensitivity analysis method can be selected according to the relationship between each sampling and evaluation results. In this study, the 1000 samples of each index are used as inputs; and the values of variable fuzzy evaluation risk grade eigenvalue (H) are calculated by Variable fuzzy method (VFM). Scatter plot of each index is mapped. According to the scatter plots illustrating the relationship between the grade eigenvalue (H) and each index, the influence degree of each index to result can be analyzed.

Scatter plots (Figure5) show the variation of each index with changes of risk grade eigenvalue (H). From Figure 5 we find that the relationships between each index and the risk grade eigenvalue (H) are all linear. Some are positive correlation such as the width of main stream channel, the longest length of flow path and length of main stream channel. The others are negative correlation. According to the scatter plots, we find the sensitivity of all index are not evident. Other sensitivity analysis methods need to be used.

### 3.3 Stepwise regression analysis

From the scatter plots we know the relationship between each index and the risk grade eigenvalue (H) are all linear. Stepwise regression analysis is a convenient linear regression method [52]. In this study, the significant test value  $p$  is 0.05. The stepwise criteria is followed that: when the probability of significant test value  $p$  is less than 0.05, the variable is to enter; when  $p$  is more than 0.1, the variable is to remove. The regression model is constructed by importing regression variable one by one in Table5. Different regression model has different importing variables and partial correlation coefficients (R), adjusted coefficients ( $R^2$ ) and standard error of the estimate. As we know the bigger  $R^2$ , the smaller the standard error of the estimate, then the regression model has good fitting. But it does not mean that  $R^2$  is bigger, the regression model fitting is better. Because when the importing variables increase in model, the partial correlation coefficient will increase, while the model fitting is not necessarily better. Therefore, considering the above factors, the regression model 6 is selected. Then the standardized partial correlation coefficient of each importing variable of model 6 are calculated and shown in Table 6. Key sensitive factor analysis based on standardized partial correlation coefficient shows that the key sensitive factor is watershed slope (S), the next in turn are antecedent precipitation index (CN), longest flow path length ( $J_1$ ), watershed shape factor ( $K_e$ ), average width of

main stream (B) and watershed area(F).

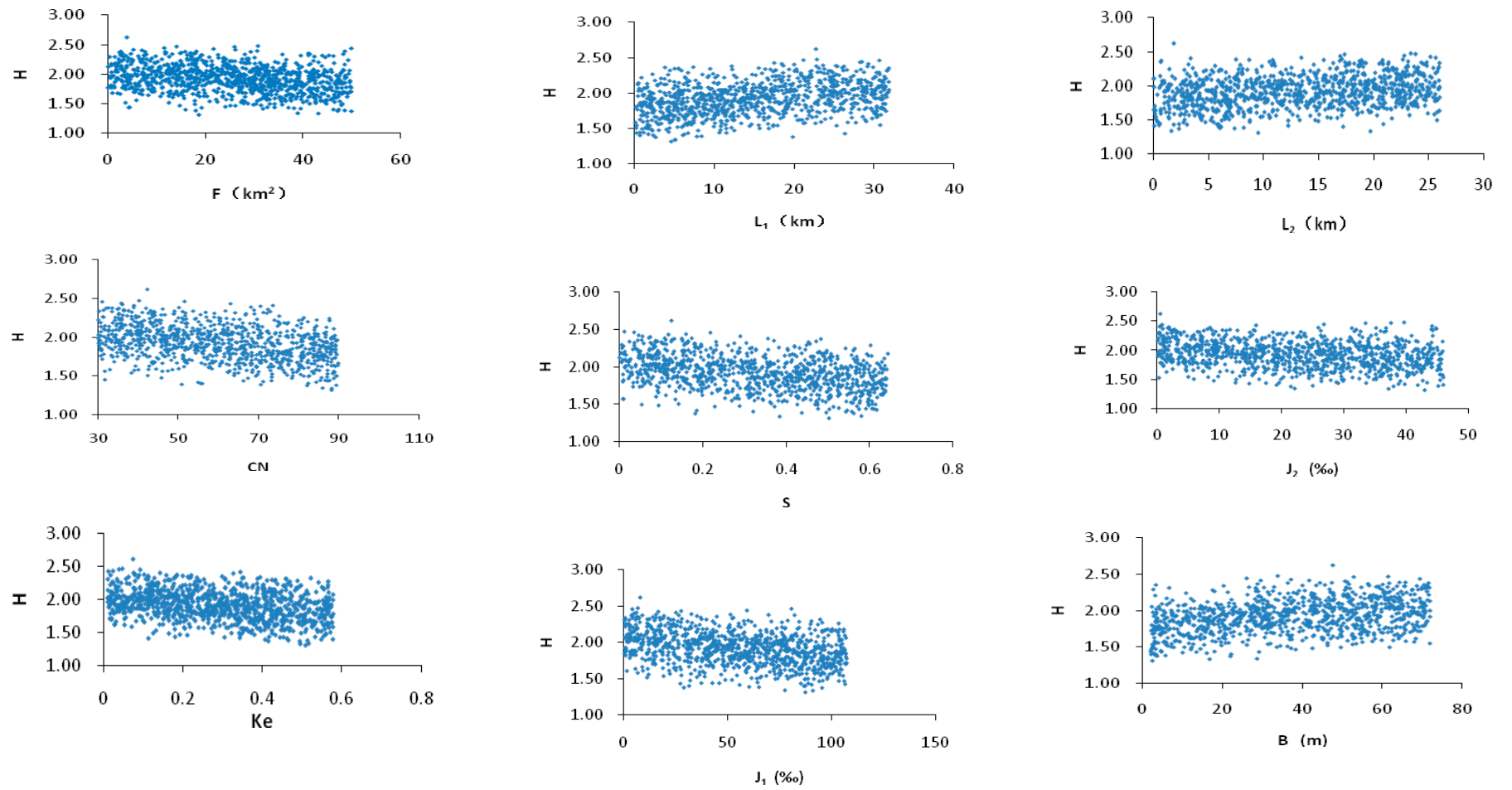
**Table 5.** Regression model summary

Model	Importing variable	R	R <sup>2</sup>	Adjusted R <sup>2</sup>	Standard error of the estimate
1	X3	0.388	0.150	0.150	0.208
2	X9	0.523	0.273	0.272	0.192
3	X4	0.630	0.397	0.396	0.175
4	X2	0.703	0.494	0.492	0.161
5	X5	0.769	0.591	0.589	0.144
6	X1	0.832	0.692	0.691	0.125

**Table 6.** Coefficients of the regression model

The regression model	Variable	Unstandardized partial correlation coefficient	Standardized partial correlation coefficient	P-value
6	constant	2.348		0
	X3	-0.451	-0.370	0
	X9	0.004	0.321	0
	X4	-0.005	-0.349	0
	X2	-0.464	-0.332	0
	X5	0.008	0.337	0
	X1	-0.005	-0.320	0





**Figure 5.** Scatter plots between each index and results of risk grade eigenvalue  $H$

### 3.4 Mutual Information Analysis

Using Equations(17), (18), (19) and (20), the joint entropy  $H(x,y)$ , the mutual entropy  $I(x,y)$ , the uncertainty coefficient  $U(x,y)$  and R-statistic are calculated based on the risk evaluation results of 1000 samples, respectively. R statistic to determine the sensitive ranking of indexes is shown in Table 7. The first column and the second column show the indexes in a lesser important order. According to R statistic, watershed slope (S) is the most sensitive index to flash flood risk in this study area. And in turn, antecedent precipitation index (CN), average width of main stream (B) and watershed shape factor (Ke) are more sensitive indexes. Although there exist slightly differences between the results from mutual entropy and stepwise regression analysis, importance ranking of index is similar for both mutual entropy analysis and stepwise regression analysis.

**Table 7.** Results of mutual information analysis

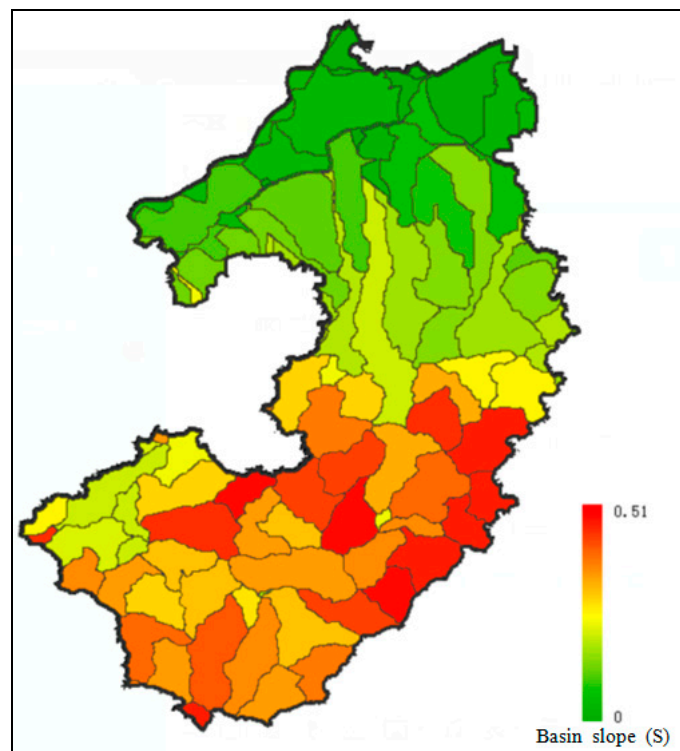
Rank	Index	H	I	U	R
1	S	4.140	0.121	0.057	0.327
2	CN	4.148	0.117	0.055	0.322
3	B	4.149	0.114	0.054	0.319
4	Ke	4.166	0.098	0.046	0.296
5	J <sub>1</sub>	4.177	0.089	0.042	0.282
6	L <sub>1</sub>	4.170	0.086	0.040	0.279
7	J <sub>2</sub>	4.185	0.079	0.037	0.268
8	F	4.182	0.075	0.035	0.261
9	L <sub>2</sub>	4.195	0.064	0.030	0.241

### 3.5 The priorities for management in Licheng county

According to the key factor---watershed slope (S), the priorities of each small watershed for management are determined. If the watershed has the same watershed slope, then the high antecedent precipitation index (CN) has the priorities for management. The priorities are shown in Table 8 and Figure 6. These properties provide a certain decision support to prevent and control flood disaster for small watersheds and will be helpful for the management of these small watersheds.

**Table 8** The priorities of 119 small watersheds For management in Licheng county

Priorities	Watershed code	Watershed slope
1	WDA84111C0000000	0.51
2	WDA84111F0000000	0.508
3	WDA8500121A00000	0.474
⋮	⋮	⋮
117	WDD1140600000000	0.00001
118	WDD110012G000000	0.00001
119	WDD110012E000000	0.00001

**Figure 6.** The priorities diagram of 119 small watersheds based on watershed slope in Licheng county

#### 4. Conclusions

After the construction of an index system for flash flood risk evaluation, the flash flood risk from underlying surface characteristics of small watersheds in hilly region is analyzed. We discover a linear relationship between each index and its risk grade eigenvalue (H).

Based on stepwise regression analysis and mutual information analysis, we find the key sensitivity factors of flash flood risk are watershed slope (S) and antecedent precipitation index (CN). It also shows that stepwise regression analysis and mutual

entropy analysis are two capable methods to identify key sensitivity factors of flash flood risk.

According to the key sensitivity factors discovered above, the priorities of each small watershed for flash flood prevent and control are determined, which will effectively help the management of these 119 small watersheds in Licheng county.

In this study, 119 small watersheds in Licheng county in China are taken as an example. Although the result of priorities of 119 small watersheds can not be extrapolated to other small watersheds, the analysis procedure proposed in this paper and the two sensitivity analysis methods----stepwise regression analysis and mutual information analysis ----can be used for small watersheds in other regions.

**Acknowledgements:** This research was supported by (1) the project “ Flash Flood Disaster Prevention and Control Technology Application Research in Shandong Province” (No. SZJSYY-YJ201501 and No.SZJSYY-YJ201502), and (2) the project “Shandong Provincial Water Conservancy Scientific Research and Technology Promotion” (No. SDSLKY201402).

**Author Contributions:** Haofang Wang conceived and designed the survey; Jincun Zhang contributed analysis tools; Ruan Yun analyzed the data; and Haofang Wang and Ranhang Zhao wrote the paper.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Reference

1. Maruša Špitalar a,b; Jonathan J; Gourley et al. Analysis of flash flood parameters and human impacts in the US from 2006 to 2012. *Journal of Hydrology*. **2014**,519, 863-870.
2. Creutin, J.D.; Borga, M.; Grunfest, E.; Lutoff, C.; Zoccatelli, D. A space and time framework for analyzing human anticipation of flash floods. *Journal of Hydrology*.**2013**,482, 14-24.
3. Daniele Norbiato;Marco Borga.et al. Flash flood warning based on rainfall thresholds and soil moisture conditions: An assessment for gauged and ungauged basins. *Journal of Hydrology*. **2008**,362, 274-290.
4. M. Borga E.N. Anagnostou; G. Bloßchl; J.-D.Creutin. Flash flood forecasting, warning and risk management: the HYDRATE project. *Environmental Science & Policy*.**2011**, 14, 834-844.
5. Chen Jin. The current situation, problems and countermeasures of mountain flash flood disaster control, *CHINA WATER RESOURCES*, **2016**,14,9-11.
6. Rojas R; Feyen L; Dassargues A. Sensitivity analysis of prior model probabilities

- and the value of prior knowledge in the assessment of conceptual model uncertainty in groundwater modelling. *Hydrol Process.* **2009**, 23,1131-1146.
7. Singh VP. The use of entropy in hydrology and water resources. *Hydrol Process.* **1997**, 11,587-626
  8. Chuanqi Li; Wei Wang; Jianzhi Xiong; Pengyu Chen. Sensitivity Analysis for Urban Drainage Modeling Using Mutual Information, *Entropy.* **2014**,16, 5738-5752.
  9. Blower, S.M.; Dowlatabadi, H. Sensitivity and uncertainty analysis of complex-models of disease transmission—An HIV model, as an example. *Int. Stat. Rev.* **1994**, 62, 229-243.
  10. Jianyong Wu; Radhika Dhingra elta. Sensitivity analysis of infectious disease models: methods, advances and their application. *Journal of the Royal Society Interface.* **2013**, June,1-15.
  11. Wang, H.,Du, P.; Zhao, D. Global sensitivity analysis for urban rainfall-runoff model. *China Environ. Sci.* **2008**, 28, 725-729.
  12. Zeng, X., Wang, D.,Wu, J. Sensitivity analysis of the probability distribution of groundwater level series based on entropy. *Stoch Environ. Res. Risk Assess.* **2012**, 26, 345-356.
  13. Helton, J.C., Davis, F.J. Illustration of sampling-based methods for uncertainty and sensitivity analysis. *Risk Anal.* **2002**, 22, 591-622.
  14. Gamerith, V., Neumann, M.B., Muschalla, D. Applying global sensitivity analysis to the modelling of flow and water quality in sewers. *Water Res.* **2013**, 47, 4600-4611.
  15. Mishra, S.; Deeds, N.; Ruskauff, G. Global sensitivity analysis techniques for probabilistic ground water modeling. *Ground Water.* **2009**, 47, 727-744.
  16. Bonnländer, B.V.; Weigend, A.S. Selecting input variables using mutual information and nonparametric density estimation. *In Proceedings of the International Symposium on Artificial Neural Networks (ISANN'94)*, Tainan, Taiwan. **1994**, 42-50.
  17. Pappenberger F; Beven KJ;Ratto M; Matgen P. Multi-method global sensitivity analysis of flood inundation models. *Adv Water Resour.* 2008, 31, 1-14.
  18. J. C. Helton; F. J. Davis. Illustration of sampling-based methods for uncertainty and sensitivity analysis. *Risk Analysis.* **2002**, 22, 591-622.
  19. Harrold, T.I.; Sharma, A.; Sheather, S. Selection of a kernel bandwidth for measuring dependence in hydrologic time series using the mutual information criterion. *Stoch Environ. Res. Risk Assess.* **2001**, 15, 310-324.
  20. Kumar PTK; Sekimoto H. Reduction of systematic uncertainty in the transmission measurement of iron using entropy based mutual information. *Radiat Meas.* **2009**, 44,168-172
  21. Vezzaro, L.; Mikkelsen, P.S. Application of global sensitivity analysis and uncertainty quantification in dynamic modeling of micropollutants in stormwater runoff. *Environ. Model. Softw.* **2012** , 27, 40-41.

22. Muleta, M.K.; Nicklow, J.W. Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model. *J. Hydrol.* **2005**, *306*, 127-145.
23. Sobol, I.M. On sensitivity estimates for nonlinear mathematical models. *Matematicheskoe Modelirov.* **1990**, *2*, 112-118.
24. McCarthy M. A.; Burgman;M. A.; Ferson, S. Sensitivity analysis for models of population viability. *Biological Conservation.***1995**,*73*, 93-100.
25. Patil, S. R.; Frey, H. C. Comparison of sensitivity analysis methods based upon applications to a food safety risk model. *Risk Analysis.* **2004**, *24*,573-585.
26. Chen Shouyu. Theory and model of engineering variable fuzzy set—mathematical basis for fuzzy hydrology and water resources. *Journal of Dalian University of Technology.***2005**, *45*, 308-312.
27. Chen Shouyu; Hu Jimin. Variable fuzzy assessment method and its application in assessing water resources carrying capacity.*SHUILI XUEBAO.***2006**,*37*, 264-271.
28. Xinyu Wan; Pingan Zhong,elta. Variable sets and fuzzy rating interval for water allocation options assessment.*Water Resources Management.***2014**, *28*, 2833-2849.
29. Jianzhu Li;SenmingTan,elta. A new method of change point detection using variable fuzzy sets under environmental change. *Water Resources Management.* **2014**, *28*, 5125-5138.
30. Wenchuan Wang; Dongmei Xu, elta. Assessment of river water quality based on theory of variable fuzzy sets and fuzzy binary comparison method. *Water Resources Management.* **2014**, *28*, 4183-4200.
31. Yuankun Wang; Dong Sheng,elta. Variable fuzzy set theory to assess water quality of the Meiliang bay in Taihu lake basin.*Water Resources Management.***2014**, *28*, 867-880.
32. X. J. Wang; R. H. Zhao; Y. W. Hao. Flood control operations based on the theory of variable fuzzy sets. *Water Resources Management.***2011**,*25*,777-792.
33. Chen SY. The theory and application of engineering fuzzy sets. *National Defence Industry Press*, Beijing, **1998**.
34. Chang N; Chen HW; King SK. Identification of river water quality using the fuzzy synthetic evaluation approach. *Journal of Environmental Management.* **2001**, *63*, 293-305.
35. Chen SY; Guo Y. Variable fuzzy sets and its application in comprehensive risk evaluation for flood-control engineering system. *Fuzzy Optim Decis Ma.* **2006**, *5*, 153-162.
36. Chen SY. Theory and model of variable fuzzy sets and its application. *Dalian university of technology press*, Dalian, **2009**.
37. Chen S; Chai H; Su Y. Variable fuzzy sets methods and application on land suitability evaluation. *Transactions of the CSAE.***2007**, *23*, 95-97.
38. Charkhabi AH; Sakizadeh M. Assessment of spatial variation of water quality parameters in the most polluted branch of the Anzali Wetland, Northern Iran. *Pol*

- J Environ Stud.* **2006**, 15, 395-403.
39. Chen FZ; Song XL; Hu YH; Liu ZW; Qin BQ. Water quality improvement and phytoplankton response in the drinking water source in Meiliang Bay of Lake Taihu, China. *Ecol Eng.***2009**,35,1637-1645.
  40. Dahiya S; Singh B; Gaur S; Garg VK; Kushwaha HS. Analysis of groundwater quality using fuzzy synthetic evaluation. *J Hazard M.* **2007**,147,938-946.
  41. Zou Q; Zhou JZ; Zhou C; Song LX; Guo J. Comprehensive flood risk assessment based on set pair analysis-variable fuzzy sets model and fuzzy AHP. *Stoch Env Res Risk A.* **2013**, 27,525-546.
  42. Qingbin Meng; Shengle Cao; Chengzhong Zhao. Jinan flood control practice and exploration. *China Water & Power Press, Beijing*,**2008**.
  43. Xia Xue. Research on small Watershed classification and early-warning indicators of flash flood disasters in prevention areas [D]. *Shandong University*.**2015**.
  44. Huang Minshen; Huang Chengchen. Research on grade model offlood risk assessment. *Journal of Catastrophology.* **2007**, 22,1-5.
  45. Luo Pei; Zhang Tairu; Du Jun. GIS and hazy evaluation method based on hazard evaluation for flood disaster, taking Chongqing as an example. *Journal of China West Normal Univer8ity(Natural Sciences)*.**2007**,28,165-171.
  46. Liu Hexiang; Xu Qingjuan. Fuzzy valuation and forecast of regional flood and water logging risk. *Journal of Catastrophology.* **2007**, 22,38-42.
  47. MF Huang; H Jing ;YR Zhong ; FY Liu. Sampling Method on 3D Model Surface Based on Quasi Random Sequence. *Computer Engineering*.**2008**, 34,263-265.
  48. DR Smith; RF Villeda; DP Lemarié. Application of adaptive cluster sampling to low-density populations of freshwater mussels. *Environmental & Ecological Statistics*.**2003**, 10, 7-15.
  49. FOLken; D Rotem. Random sampling from databases: a survey. *Statistics & Computing.* **1995**, 5,25-42.
  50. McKay, M.D.; Beckman, R.J.; Conover, W.J. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics.* **1979**, 21,239-245.
  51. Taylor & Francis. Controlling correlations in Latin Hypercube Samples. *Journal of the American Statistical Association.* **1994**, 89,1517-1522.
  52. Bergante S; Facciotto G; Minotta G. Identification of the main site factors and management intensity affecting the establishment of Short-Rotation-Coppices (SRC) in Northern Italy through stepwise regression analysis. *Cent Eur J Biol.***2010**, 5,522-530.
  53. Moddemeijer R. On estimation of entropy and mutual information of continuous distributions. *Signal Process.* **1989**,16, 233-248.
  54. Haussler D; Oppen M. Mutual, information, metric entropy and cumulative relative entropy risk. *Ann Stat.* **1997**, 25,2451-2492.

55. Mishra S.; Knowlton, R.G. Testing for input-output dependence in performance assessment models. In *Proceedings of the 10th International High-Level Radioactive Waste Management Conference*, LasVegas, Nevada, IL: American Nuclear Society, 30 March-2 April **2003**; pp. 882-887.
56. Mogheir, Y.; de Lima, J.; Singh, V.P. Characterizing the spatialvariability of groundwater quality using the entropy theory: I. Synthetic data. *Hydrol. Process.* **2004**, 18, 2165-2179.
57. Dongya Sun; Dawei Zhang; Xiaotao Cheng. Framework of national non-structural measures for flash flood Disaster prevention in China. *Water.* **2012**, 4, 272-282.