Nat. Hazards Earth Syst. Sci., 13, 239–249, 2013
www.nat-hazards-earth-syst-sci.net/13/239/2013/
doi:10.5194/nhess-13-239-2013

Natural Hazards and Earth System Sciences
Open Access

Fuzzy approach to analysis of flood risk based on variable fuzzy sets and improved information diffusion methods

Q. Li

School of Mathematics and Physics, Hubei Polytechnic University, Huangshi, Hubei, China

Correspondence to: Q. Li (liqiong070108@163.com)

Received: 10 January 2012 – Published in Nat. Hazards Earth Syst. Sci. Discuss.: –
Revised: 24 May 2012 – Accepted: 10 November 2012 – Published: 8 February 2013

Abstract. The predictive analysis of natural disasters and their consequences is challenging because of uncertainties and incomplete data. The present article studies the use of variable fuzzy sets (VFS) and improved information diffusion method (IIDM) to construct a composite method. The proposed method aims to integrate multiple factors and quantification of uncertainties within a consistent system for catastrophic risk assessment. The fuzzy methodology is proposed in the area of flood disaster risk assessment to improve probability estimation. The purpose of the current study is to establish a fuzzy model to evaluate flood risk with incomplete data sets. The results of the example indicate that the methodology is effective and practical; thus, it has the potential to forecast the flood risk in flood risk management.

1 Introduction

Floods are rare natural mutation phenomenon. They occur frequently in China, where approximately two-thirds of its area is threatened by different types and degrees of floods (Chen, 2010). These phenomena are the results of natural and unnatural causes, such as social and economic factors. In order to reduce the so severe losses that caused by natural disasters, a large number of engineering and non-engineering relief efforts were carried out. One important work of which is carrying on flood disaster risk analysis.

In the case of natural hazards, risk is most meaningful when expressed in terms of potential human suffering and/or economic losses (Wilson and Crouch, 1987; Maskrey, 1989; Smith, 1996). Besides the probability of a hazard occurring, risk must include the potential adverse consequences that can result from the hazard event. To summarize all definitions of "risk", would be to say the term refers to "the possibility of some adverse events."

Risk analysis incorporates the likelihood of a specific event and the severity of the outcome. This process combines both the severity and the probability of all relevant hazard loss scenarios. Risk managers use two different evaluative methods in risk and hazard analysis: deterministic and probabilistic.

Deterministic analysis relies on the laws of physics and chemistry, or on correlations developed through experience or testing, to predict the outcome of a particular hazard scenario. In the deterministic approach, one or more possible designs can be developed that represent the worst possible credible events. In this approach, the frequency of possible occurrences need not be evaluated.

Probabilistic analysis evaluates the statistical likelihood that a specific event will occur and what losses and consequences will result. In addition to using analysis techniques and experimental findings, the approach uses considerable statistics including the incorporation of historical information. But the conventional hydrological frequency analysis method often becomes invalid because of the shortage of historical measured data in degree and frequency (Efstratiadis et al., 2010).

People often use "expectation," instead of "probability distribution" to compare risks or mathematically combine relevant quantities using one or some operators. The risk of quantitative expression also becomes the "risk degree". In some fields, an essentially multi-dimensional risk problem can be simplified into a one-dimensional (1-D) problem using the proper expression. However, common sense dictates that we must be very careful in using the simplified results in the field of general risk analysis.

In general, to obtain a way to control or manage a system, we use a mathematical model that closely represents the system. The mathematical model is solved and its solution is applied to the system. Models are idealized representation but the nature of disasters is changing, and disasters are becoming more complex.

In fact, the main characteristics of natural disaster systems are the uncertainty and complexity of the system. A subsystem of natural disaster systems is the human society system. Disasters have a broad social impact. They cause deaths, injuries and monetary losses. However, they can also redirect the character of social institutions, result in new and costly regulations imposed on future generations, alter ecosystems and disturb the stability of political regimes. So when judging the consequences of future natural disasters, it is necessary to analyze the human society structure (land use, property distribution, structures, etc). Human society structure is very complex with many uncertain factors, which we must fully consider in analyzing natural disaster systems. In fact, even on natural disaster phenomena, uncertainty and complexity still cannot be ignored. Uncertainty, ambiguity, constant change, and surprise are problems that disaster management has to overcome. Unfortunately, most management strategies are designed for a predictable world and a static view of natural disasters.

One very common claim states that mathematical models fail when applied in management sciences due to system complexity or computational difficulty. A risk system may only be studied by certain state equations, on the condition that these state equations could be found. However, in many cases, to obtain the state equation and all data is very difficult. Going a step further, if we employ other methods to simplify the system analysis, obtaining the precise relations we need is also difficult. In other words, the relations we obtain are usually imprecise.

Probability method is another kind of simplified research work, but it is not appropriate to replace risk analysis with the probability analysis. Altay and Green (2006) show that probability theory and statistics are very frequently utilized methods in the research literature (about 20 % of research effort included in their review), but much less frequently in practice. Because feasibility and reliability problems exist in terms of practical issues without considering the fuzzy uncertainty. Sometimes, results based on the classical statistical methods are sometimes very unreliable and unstable for cases with limited or small sample issues. In fact, the collection of long sequence disaster data is quite a challenging task, especially when the sample is small. In this case, only records of the same age or of similar ages may be used for flood risk analysis.

Information diffusion is just a fuzzy mathematical set-value method for samples, considering that the use of fuzzy information of samples is optimized to offset information deficiency. The information diffusion theory helps extract the useful underlying information from the sample as much as possible, improving system recognition accuracy (Huang, 2002; Palm, 2007).

2 Fuzzy risk

When we study a risk system using a traditional probabilistic method, it is usually difficult to ascertain if a hypothesis of probability distribution is suitable, and sometimes we meet the problem of small samples, wherein the data is too scanty to make a decision. The problem has shown that empirical Bayesian methods (Carlin and Louis, 1997) and kernel methods (Breiman et al., 1977; Chen, 1989; Devroye and Gyorfi, 1985; Hand, 1982; Parzen, 1962; Silverman, 1986; Wertz, 1978) need further development. Obtaining a precise relation between events and probabilities of occurrence is difficult. To represent the imprecision, the best way is to employ fuzzy sets to represent the relations.

Uncertainty is another factor to consider. It is defined in simple language as the lack of definiteness and has important implications in what integrated disaster management could achieve. Thus, uncertainty should be considered in all disaster management decisions.

Fuzzy set theory, which deals with uncertainties and allows the incorporation of the opinions of decision makers, may provide an appropriate tool for establishing disaster risk management systems, such as fuzzy rule-based techniques and the combination of the fuzzy approach with other techniques. Risk is expressed in terms of fuzzy risk only when we study it by a fuzzy method. Some early related applications can be found in the literature (Brown, 1979; Clements, 1977; Dong et al., 1985; Esogbue et al., 1992; Hadipriono and Ross, 1991; Hoffman et al., 1978).

The concept of the fuzzy set was proposed by Zadah (Zadeh, 1965), who bestowed media and fuzziness scientific description and great significance in the academic world. However, the fuzzy set is static if the relativity and variability are not considered. Therefore, the theory is in conflict with the variability of the interim form. Some defects of traditional fuzzy sets are due to approaching the media, variable fuzzy phenomenon, and variable fuzzy objects by static concepts, theory, and method of traditional fuzzy sets.

In light of the foregoing, the theory and method of variable fuzzy sets (VFS) was proposed by Chen based on opposite fuzzy sets and the definitions of a relative difference function (Guo and Chen, 2006; Wu et al., 2006). The method of Chen is the innovation and extension of the static fuzzy set theory established by Zadeh (Zadeh, 1965), which is very important in theory and applications. The VFS theory was the extension of fuzzy sets theory which was established by Chen (Chen and Guo, 2006; Guo and Chen, 2006; Wu et al., 2006). The comprehensive evaluation of VFS effectively eliminates the border effect on assessment result and monitors estimation standard error. This method can determine relative membership functions and membership degrees of

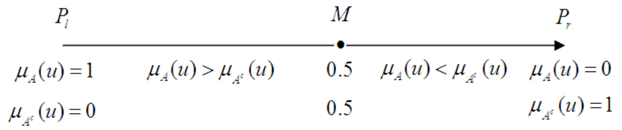

Fig. 1. The relation of the relative membership degree function $\mu_A(u)$; $\mu_{A^c}(u)$ and the relative difference function $D(u)$.

disquisitive objectives (or indices) scientifically and reasonably. This technique can also make full use of one's knowledge and experience, as well as quantitative and qualitative information of index systems, to obtain weights of objectives (or indices) for comprehensive disasters evaluations (Wang et al., 2011; Zhang et al., 2011).

In the present study, VFS was combined with information diffusion as an integration of techniques. The method proposed in the current research uses fuzzy multiple indicators of comprehensive evaluation of VFS and converts the multi-dimensional indicators of the samples into one-dimensional degree values. Then, the method turns the degree values of the observed sample into fuzzy sets by information diffusion method, finally obtaining the risk values. The method is then tested by a case showing that the method is superior to the traditional statistical model and improves the result of traditional estimation.

3 Modeling framework of variable fuzzy sets and information diffusion

3.1 Variable fuzzy sets

A is a fuzzy concept in the domain of U, and u is any element of U ($u \in U$). A and A_c are a pair of opposite fuzzy concepts (or two basic fuzzy attributes). The relative membership degree RMD of A_c is any element u of U to A is $\mu_A(u)$, which ranges from 0 to 1 continuously. The RMD to A^c is μ_A^c, which ranges from 1 to 0. RMD satisfies $\mu_A(u)+\mu_{A^c}(u)=1$ (Chen and Guo, 2006; Guo and Chen, 2006).

Hence, for any element u ($u \in U$), a pair of numbers $\mu_A(u)$ and $\mu_{A^c}(u)$ and in the continuum $[0,1]$ are assigned to u.

$\mu_A(u)$ and $\mu_{A^c}(u)$ are the RMD functions that express the levels of acceptability and repellency, respectively. The mapping $\mu_A(u),\ \mu_{A^c}(u): u \mapsto \mu_A(u), \mu_{A^c}(u) \in [0,1]$ is shown in Fig. 1.

When two basic fuzzy attributes are equal, the dynamic equilibrium can be established (Fig. 1). When the RMD $\mu_A(u)$ is larger than $\mu_{A^c}(u)$, the major property of u is acceptability, and the minor property is repellency. When it changes from $\mu_A(u) > \mu_{A^c}(u)$ to $\mu_A(u) < \mu_{A^c}(u)$, the conclusion is the exact opposite.

Definition: Let $D(u) = \mu_A(u) - \mu_{A^c}(u)$, where $D(u)$ is defined as the relative difference function of u to A as

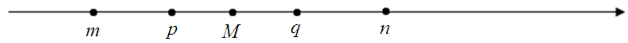

Fig. 2. Relationship between point M, and internals $[p,q]$, $[m,n]$.

the mapping $D: u \to D(u) \in [-1,1]$ (Chen and Guo, 2006; Guo and Chen 2006; Wu et al., 2006). Hence, from the Eq. $\mu_A(u)+\mu_{A^c}(u)=1$ we can infer that $D(u)=2\mu_A(u)-1$ or $\mu_A(u)=1+D(u)/2$. Let

$$\begin{aligned} A_+ &= \{u\,|u \in U, 0 < D(u) < 1\} \\ A_- &= \{u\,|u \in U, -1 < D(u) < 0\} \\ A_0 &= \{u\,|u \in U, D(u) = 0\} \\ V_0 &= \{(u,D)\,|u \in U, D(u) \\ &= \mu_A(u) - \mu_{A^c}(u), D \in [-1,1]\} \end{aligned}$$

A_+, A_- and A_0 are the attracting sets, repelling sets, and balance boundaries of VFS V_0, respectively. For any element u, the set V_0 is defined as Variable Fuzzy Sets (VFS).

For example, the interval $[p,q]$ is considered as the attracting sets W_0 of W, implying that $0 < D_A(u) \leq 1$ and $\mu_A(u) < \mu_{A^C}(u)$, and W is the interval $[m,n]$ which includes W_0 ($W_0 \subset W$). $[m,p]$ and $[q,n]$ are the repelling sets of W implying that $-1 \leq D_A(u) < 0$ and $\mu_A(u) < \mu_{A^C}(u)$ in the repelling sets $[m,p]$ and $[q,n]$. Then, M is a value which satisfies $D_A(u)=1$, implying $\mu_A(u)=\mu_{A^c}(u)=0.5$, where M is usually different from the mid-value of interval $[p,q]$ (see Fig. 2).

If x is placed at right side of M, the relative difference function $(\forall u \in W)\, D_A(u)$ can be expressed as

$$\begin{cases} D(u) = (\frac{u-p}{M-p})^\beta & u \in [p,M] \\ D(u) = -(\frac{u-p}{m-p})^\beta & u \in [m,p] \end{cases} \tag{1}$$

According to the formula $\mu_A(u)=1+D(u)/2$, Eq. (1) can be transformed to Eq. (2):

$$\begin{cases} \mu_A(u) = 0.5[1+(\frac{u-p}{M-p})^\beta] & u \in [p,M] \\ \mu_A(u) = 0.5[1-(\frac{u-p}{m-p})^\beta] & u \in [m,p] \end{cases} \tag{2}$$

and if x is placed at left side of M, the difference function is

$$\begin{cases} D(u) = (\frac{u-q}{M-q})^\beta & u \in [M,q] \\ D(u) = -(\frac{u-q}{n-q})^\beta & u \in [q,n] \end{cases} \tag{3}$$

and Eq. (3) can also be transformed to Eq. (4):

$$\text{Or} \begin{cases} \mu_A(u) = 0.5[1+(\frac{u-q}{M-q})^\beta] & u \in [M,q] \\ \mu_A(u) = 0.5[1-(\frac{u-q}{n-q})^\beta] & u \in [q,n]. \end{cases} \tag{4}$$

In which β is an index larger than 0, usually $\beta=1$. Equations (1) to (4) are linear functions. They satisfy the following conditions: (i) $u=p$ and $u=q$, $D(u)=0$ and $\mu_A(u)=\mu_{A^c}(u)=0.5$; (ii) $u=M$, $D(u)=1$ and $\mu_A(u)=1$; and (iii) $u=m$ and $u=n$, $D(u)=-1$ and $\mu_A(u)=0$. Based on Eqs. (2) and (4), the values of relative membership degree function $\mu_A(u)$ of inquisitive indices are obtained.

3.2 Information diffusion

In the current research, we first use fuzzy multiple indicators of comprehensive evaluation of VFS and convert the multi-dimensional indicators of the samples into one-dimensional degree values. Then, we turn the degree values of the observed sample into fuzzy sets by information diffusion method, finally obtaining the risk values.

Information diffusion is a fuzzy mathematic set-value method for samples. The use of fuzzy information of samples is optimized to offset information deficiency. This method can transform an observed data into a fuzzy set, turning a single point sample into a set-value sample.

In this study some improvements are made to improve the information diffusion method:

1. Diffusion velocity k is replaced with variable $k(u)$, and it is used to obtain the disaster risk estimation. This method provides an improved information diffusion process implementation and higher accuracy.

2. An optimal discretization algorithm is designed. According to the diffusion function assumed in the IIDM equation, the MacCormack scheme is used during the process to obtain the numerical solutions. Finally, the method is compared with NIDM and traditional statistical method by examining the mean error.

3.2.1 Definition of information diffusion

Information diffusion: suppose X is the sample set and V is the universal field, and information diffusion is a mapping $\mu : X \times V \rightarrow [0,1]$ that satisfies the following (Huang and Shi, 2002):

1. It is decreasing. $\forall x \in X, \forall \upsilon', \upsilon'' \in V$, if $\left\|\upsilon' - x\right\| \leq \left\|\upsilon'' - x\right\|$, then $\mu(x,\upsilon') \geq \mu(x,\upsilon'')$. μ is the diffusion function.

2. $\forall x \in X$. Let υ^* be the observed value of x, which satisfies $\mu(x,\upsilon^*) = \max_{\upsilon \in V} \mu(x,\upsilon)$.

3. $\mu(x,\upsilon)$ is conservative. If and only if $\forall x \in X$, its integral value on the universe is 1, viz. $\int_U \mu(x,u)du = 1$.

3.2.2 Improved information diffusion method (IIDM)

In Shang and Jin (2002), the diffusion function of information diffusion method is described as follows:

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\left(k\frac{\partial u}{\partial x}\right). \tag{5}$$

In which k is a constant. So Eq. (4) can be written as

$$\frac{\partial u}{\partial t} = k\frac{\partial^2 u}{\partial x^2}. \tag{6}$$

But the information carried by the samples moves from a higher level of consistency to a lower one because of inherent uneven distributions (Gu and Li, 2002). So a function of u, i.e. $k(u)$, instead of the constant k in the current paper. $k(u)$ can be selected freely if this function satisfies some conditions. The simplest quadratic function in which $k(u) = (u+1)^2$ can be used in the current paper, satisfies the conditions as following:

1. The function is non-negative increasing.

2. The diffusion velocity is irregular which means the diffusion velocity becomes faster as the consistency increases.

3. There is little diffusion even without consistency.

So the diffusion function of improved information diffusion method (IIDM) can be written as

$$\begin{cases} \frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\left(k(u)\frac{\partial u}{\partial x}\right) \\ u|_{t=0} = \delta(x) \end{cases}, \tag{7}$$

it can be transformed as

$$\frac{\partial u}{\partial t} = k(u)\frac{\partial^2 u}{\partial x^2} + \frac{\partial k(u)}{\partial u}\left(\frac{\partial u}{\partial x}\right)^2. \tag{8}$$

The MacCormack difference scheme can be selected to get the solutions. As a special form of the Lax-Wendroff difference scheme, the MacCormack difference scheme follows a two-step predictor-corrector mechanism. This technique is a second-order-accurate method because it is used with forward differences on the predictor and with rearward differences on the corrector (Anderson and Wendt, 1995).

So the diffusion function of the problem can be described as

$$v_t(w_n, x) = u(w_n - x, t) \tag{9}$$

in which $W = \{w_1, w_2, \cdots, w_n\}$ is the sample degree variable, and $u(w_n - x, t)$ is the solution of Eq. (7). The information function carried by the sample $\{w_1, w_2, \cdots, w_n\}$ which diffuses to the point x field can be described as

$$f_t(x) = \frac{\sum_{i=1}^{n} v_t(w_i, x)}{n}. \tag{10}$$

3.2.3 Principle of NIDM

The information diffusion method is called normal information diffusion when its diffusion function is normal distribution. Let $W = \{w_1, w_2, \cdots, w_n\}$ be a sample, and $U = \{u_1, u_2, \cdots, u_m\}$ be the discrete universe. w_i and u_j are called a sample point and a monitoring point, respectively. $\forall w_i \in W \forall u_j \in U$ we diffuse the information carried by w_i to u_j

Table 1. Flood disaster rating standard.

Disaster level	Damage area (thousand hectares)	Inundated area (thousand hectares)	Dead population (persons)	Collapsed houses (ten thousand)	Recurrence interval (years)	Grade number
Small flood	0 ~ 9045	0 ~ 4989	0 ~ 3446	0 ~ 112.1	<2	1
Medium flood	9045 ~ 14197	4989 ~ 8216.7	3446 ~ 5113	112.1 ~ 247.7	2 ~ 5	2
Large flood	14 197 ~ 20 388	8216.7 ~ 13 000	5113 ~ 10 676	247.7 ~ 754.3	5 ~ 20	3
Extreme flood	20 388 ~ 80 000	13 000 ~ 50 000	10 676 ~ 1 00 000	754.3 ~ 5000	>20	4

at gain $f_i(u_j)$ by using the normal information diffusion method (NIDM) shown in Eq. (11).

$$f_i(u_j) = \exp\left[-\frac{(w_i - u_j)^2}{2h^2}\right], u_j \in U \tag{11}$$

where h is called normal diffusion coefficient, calculated by Eq. (12) (Huang and Shi, 2002).

$$h = \begin{cases} 0.8146(b-a), & n=5; \\ 0.5690(b-a), & n=6; \\ 0.4560(b-a), & n=7; \\ 0.3860(b-a), & n=8; \\ 0.3362(b-a), & n=9; \\ 0.2986(b-a), & n=10; \\ 0.6851(b-a)/(n-1), & n \geq 11. \end{cases} \tag{12}$$

where $b = \max\limits_{1\leq i\leq n}\{x_i\}; a = \min\limits_{1\leq i\leq n}\{x_i\}$.

4 Flood disaster risk analysis based on VFS and IIDM

The disastrous loss data collected by the Ministry of Water Resources of the People's Republic of China cover the period from 1950 to 2009. The damaged and inundated areas, dead population, and collapsed houses were chosen as the disaster indicators in the flood risk analysis. Using actual historical data from 1950 to 2009, the flood frequency analysis was used to determine the flood disaster rating standard (Table 1). The flood cases were divided into four groups, namely, small, medium, large, and extreme. Floods that occurred every 2 yr were classified as small floods, those that occurred every 2 to 5 yr were medium floods, those that occurred every 5 to 20 yr were large floods, and those that occurred after more than 20 yr were considered extreme floods. The range of the floods in each group was decided using the flood disaster rating standard (Table 1).

4.1 VFS for comprehensive assessment of the flood degree

Based on the study of Chen and Guo (2006) and Table 1, for each degree $h(h = 1,2,3,4)$, the following matrices of parameters are established to calculate RMD of VFS:

Table 2. Scale preferences used in the pairwise comparison process.

Range	Category	Score
Superior	Absolutely superior	9
	Very strongly superior	7
	Strongly superior	5
	Moderately superior	3
Equal	Equal	1
Inferior	Absolutely inferior	1/9
	Very strongly inferior	1/7
	Strongly inferior	1/5
	Moderately inferior	1/3

$$I_{[p,q]} = \begin{bmatrix} [0,9045] & [9045,14197] & [14197,20388] & [20388,80000] \\ [0,4989] & [4989,8216.7] & [8216.7,13000] & [13000,50000] \\ [0,3446] & [3446,5113] & [5113,10676] & [10676,100000] \\ [0,112.1] & [112.1,247.7] & [247.7,754.3] & [754.3,5000] \end{bmatrix}$$

$$I_{[m,n]} = \begin{bmatrix} [0,14197] & [0,20388] & [9045,80000] & [14197,80000] \\ [0,8216.7] & [0,13000] & [4989,50000] & [8216.7,50000] \\ [0,5113] & [0,10676] & [3446,100000] & [5113,100000] \\ [0,247.7] & [0,754.3] & [112.1,5000] & [247.7,5000] \end{bmatrix}$$

$$M = \begin{bmatrix} 0 & 10762 & 18324 & 80000 \\ 0 & 6064 & 11406 & 50000 \\ 0 & 4002 & 8822 & 100000 \\ 0 & 157 & 585 & 5000 \end{bmatrix}.$$

Index x locates at either the left side or right side of point M, then Eq. (2) or Eq. (4) is selected for calculating RMD $\mu_h(u_{ij})$ of indices based on matrices $I_{[p,q]}$, $I_{[m,n]}$, and M, where h is the grade number ($h = 1,2,3$, and 4), i is the index number ($i = 1,2,3$, and 4), and j is the sample number ($j = 1,2\cdots 32,\ldots 60$).

Then a two-level hierarchy is constructed to obtain the weights of the evaluation indicator. The goal is to ascertain "the weights of the evaluation indicators". The evaluation indicators (attributes) are damage area, inundated area, dead population, and collapsed houses.

The pairwise comparison is conducted using a scale based on the proposal of Saaty (Saaty, 1980) detailed in Table 2. To illustrate the kind of results obtained, Table 3 presents a pairwise comparison matrix drawn from the information provided from the expert for the evaluation of the importance of

Table 3. Pairwise comparison of the alternatives with respect to flood disasters.

	Damage area	Inundated area	Dead population	Collapsed houses
Damage area	1	1/2	1/9	1/3
Inundated area	2	1	1/5	1/2
Dead population	9	5	1	3
Collapsed houses	3	2	1/3	1

Table 4. Vector of weights of the alternatives with respect to flood disasters.

	Flood impact
Damage area	0.0655
Inundated area	0.1189
Dead population	0.6043
Collapsed houses	0.2113
CR	0.0030

the factors. Then, the consistency of the comparison matrix was tested and the relative weights of the elements are computed along with the consistency ratio (CR) as presented in Table 4. If the CR is below 10 %, the judgments are considered consistent.

According to AHP, we obtain the normalized weights of the evaluation indicators as

$$W=[0.06550.11890.60430.2113]=w_i. \tag{13}$$

With the following variable fuzzy recognition model Eq. (14) proposed by Wu et al. (2006), the synthetic disaster degree H of each index can be obtained.

$$u'_h(x_j)=\left\{1+\left[\frac{\sum\limits_{i=1}^{m}[w_i(1-\mu(x_{ij})_h]^p}{\sum\limits_{i=1}^{m}[w_i\mu(x_{ij})_h]^p}\right]^{\frac{\alpha}{p}}\right\}^{-1} \tag{14}$$

where $h=1$, 2, 3, 4, and H is synthetic disaster degree. x_j represent sample j, and x_{ij} is the i-th index value of the sample. w_i is the disaster index weight, m is the number of indexes ($m=4$). α is the rule parameter of model optimization and p is the distance parameter, usually $\alpha=1$ or 2; $p=1$ or 2. Using Eq. (14), the synthetic relative membership degree of each index $u'_h(x_j)$ for floods can be obtained. Then, after normalizing these indices, the normalized synthetic relative membership degrees $u_h(x_j)$ of each index are obtained.

Using Eq. (15), the synthetic disaster degree of the sample is obtained as shown in Table 5.

$$H=(1,2,3,4)\times u_h(x_j) \tag{15}$$

Table 5. The disaster degree values during the 60 yr in China.

Sample	Degree value	Sample	Degree value
1	1.2975	31	1.7058
2	2.4769	32	2.2728
3	1.5831	33	2.4727
4	1.5564	34	2.6061
5	3.2447	35	1.777
6	1.2101	36	1.7948
7	3.2537	37	1.5432
8	2.131	38	1.628
9	1.4648	39	1.8295
10	1.7051	40	1.5528
11	2.2814	41	1.6377
12	2.1606	42	2.7769
13	2.0829	43	1.3983
14	3.449	44	1.7725
15	2.2234	45	2.7208
16	1.1692	46	2.0294
17	1.0678	47	2.8727
18	1.0227	48	1.4821
19	1.0524	49	2.3079
20	1.9909	50	1.4299
21	1.1141	51	1.2857
22	1.1142	52	1.1281
23	1.0727	53	1.4456
24	1.3453	54	1.3957
25	1.2092	55	1.1273
26	3.9117	56	1.4349
27	1.1177	57	1.3782
28	1.3402	58	1.2435
29	1.0624	59	1.0483
30	1.3200	60	1.0439

4.2 Flood risk evaluation using improved information diffusion

Based on VFS, the disaster degree values of the 60 samples are calculated (Table 5), that is, the sample point set $W=\{w_1,w_{2,}\cdots,w_n\}$. The information function carried by the sample $\{w_1,w_{2,}\cdots,w_n\}$ which diffuses to the point x can be obtained according to Eqs. (5) to (8). Then, based on Eqs. (9) to (10), disaster probability risk estimation is calculated.

The IIDM is used in the form as follows:

$$\begin{cases}\frac{\partial u}{\partial t}=\frac{\partial}{\partial x}(K(u)\frac{\partial u}{\partial x})\\ u\,|_{t=0}=\delta(x)\end{cases} \tag{16}$$

where the initial value $\delta(x)$ is assumed as normal distribution. The variance of this value should be a fairly small one that can facilitate information diffusions.

Assume $u|_{t=0}=\sqrt{\frac{3}{\pi}}\exp(-3x^2)$, and the MacCormack technique is applied. The interval of x is set as $[-4,4]$ which suits the disaster degree value interval and satisfies

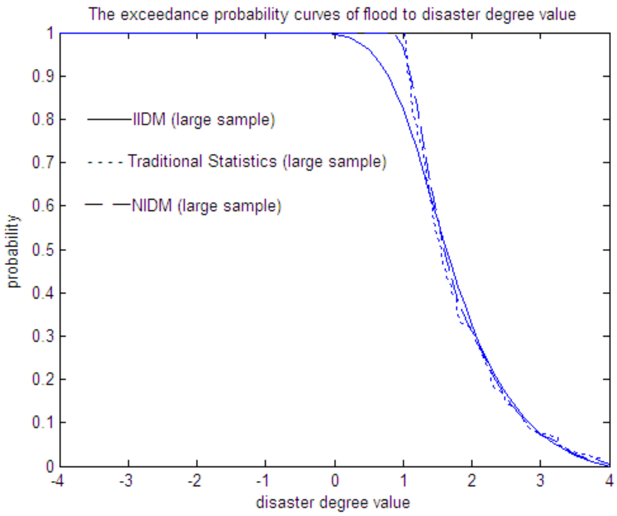

Fig. 3. Comparisons of the risks by IIDM, the traditional statistical method, and NIDM.

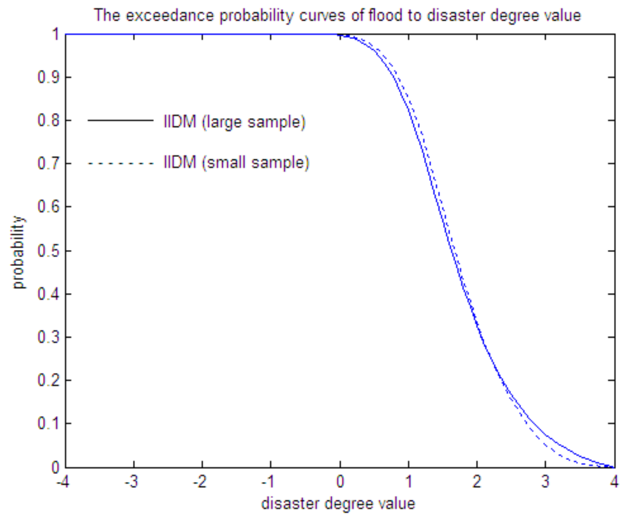

Fig. 4. Comparison of the risks by IIDM with the small sample and the large sample.

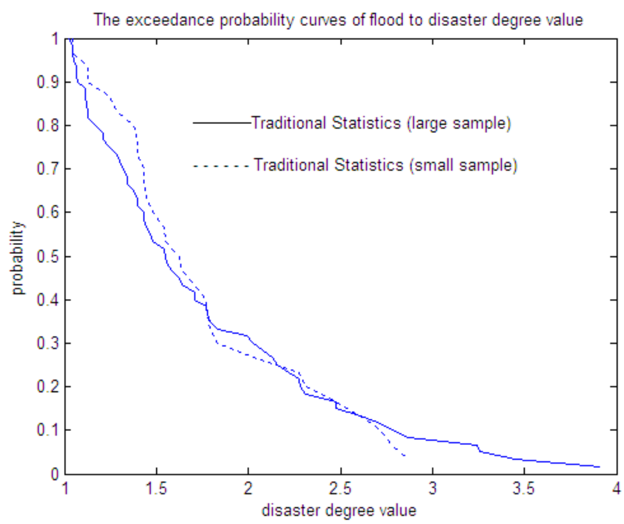

Fig. 5. Comparison of the risks by traditional statistics with the small sample and the large sample.

the symmetry requirement in normal distribution. The interval $[-4, 4]$ is divided into 500 subintervals, and the time t is equal to 1000. With the MacCormack technique the numerical solutions for the problem $u_1(x), u_2(x), \cdots, u_{1000}(x)$ can then be obtained.

As described above, the information diffusion function is $v_t(w_n, x) = u(w_n - x, t)$. $W = \{w_1, w_{2,} \cdots, w_n\}$ will be the values of the sample by VFS, whereas $u(w_n - x, t)$ is the solution of Eq. (8). The information function for disaster degree can be written as $f_t(x) = \frac{\sum_{i=1}^{n} v_t(w_n,x)}{n}$. Then, the disaster risk estimate or the probability risk value is calculated. The relationship between the recurrence interval N (years) and the probability p can be expressed as $N = 1/p$. The flood exceedance probability curve to the disaster degree value with the comparison of that by the traditional statistical method and NIDM is shown in Fig. 3.

5 Discussion

5.1 Results and discussion

In order to testify this method, all sixty records are selected to comprise the large sample, and the estimated risk of this set is calculated by the IIDM and other two methods. Then, 30 records are randomly selected to comprise a small sample set and are analyzed in the same way. Comparisons of these sample sets using IIDM, NIDM and the statistical method are demonstrated in Figs. 4 and 5, respectively.

Figure 4 shows the difference between the two curves of the estimated risk with small sample and large sample by VFS-IIDM model. From Fig. 4, two curves match well, which indicates that the result barely changes when the sample size changes, and that the method is stable and barely affected by the size of the sample. The analysis results for a very large sample can be used as the standard, so the VFS-IIDM method is considered closer to the standard than the statistical method, as proven by the following Monte Carlo experiments. In Fig. 5, we compare two curves of the estimated risk with small sample and large sample by frequency statistics. The mean errors between the results are much bigger and reach the value of 0.0522, which indicates that it is unstable and changeable with the sample size.

Figures 4 and 5 indicate that the results of the small sample analyzed by VFS-IIDM model are satisfactory. The results

Table 6. Comparison of three methods.

Method	IIDM	NIDM	Statistics
Mean error	0.0081	0.0237	0.0522

reflect the fact that the risks of the floods decrease smoothly with the increase in degree value, and that the VFS-IIDM model works better for practical problems. Table 6 presents a comparison of the mean errors between the results with large sample and small sample by VFS-IIDM model, NIDM model and traditional statistics. Comparing with those calculated by the other two methods (see Table 6), IIDM approach is much better because the mean error of VFS-IDM model is smallest.

The results also illustrate the risk assessment values and recurrence interval values of different disaster levels in China.

The disaster degree value obtained by VFS in Table 5 shows the effect of each disaster. The result in Fig. 4 illustrates the risk estimation, i.e., the exceedance probability of the disaster degree value, which indicates that the degrees of the flood impacts occur with a corresponding frequency probability. For example, according to the results calculated in Fig. 4, the exceedance probability of the flood whose impact is 2.8 degrees is 0.1057. In other words, floods exceeding the 2.8 degree value (extreme floods) occur every 9.4607 ($N = 1/p$) years. In fact, the probability risks of the floods with each flood impact degree may be obtained (see Table 7).

Based on the standard of four grades (Chen, 2009), the following categories are used:

a. if $1.0 \leq H \leq 1.5$, then the flood degree is small (1st grade);

b. if $1.5 < H \leq 2.5$, then the flood degree is medium (2nd grade);

c. if $2.5 < H \leq 3.5$, then the flood degree is large (3rd grade); and

d. if $3.5 < H \leq 4$, then the flood degree is extreme (4th grade).

Table 7 shows that the exceedance probability risk estimation is 0.0252 when the disaster impact degree is 3.5. In other words, floods exceeding the 3.5 degree value (extreme floods) occur every 39.6825 yr. Similarly, the probability of floods exceeding 2.5 degrees (large floods) is 0.1672, indicating that floods exceeding that intensity occur every 5.9809 yr. These findings demonstrate the serious situation of floods in China. The frequency and the recurrence interval of the four grades of floods are shown in Table 8.

5.2 Method evaluated using Monte Carlo method

To further evaluate our method, we compared it with other methods using some simulation experiments by Monte Carlo method. As flood frequency always obeys some kind of distribution. Some experiments are conducted using the normal distribution, exponential distribution and log-normal distribution.

Table 7. The disaster risk values.

Degree value	Risk	Degree value	Risk
0.1	0.9929	2.1	0.2885
0.2	0.9885	2.2	0.2534
0.3	0.9821	2.3	0.2216
0.4	0.9730	2.4	0.1929
0.5	0.9604	2.5	0.1672
0.6	0.9436	2.6	0.1443
0.7	0.9219	2.7	0.1239
0.8	0.8948	2.8	0.1057
0.9	0.8617	2.9	0.0896
1.0	0.8228	3.0	0.0753
1.1	0.7785	3.1	0.0627
1.2	0.7297	3.2	0.0515
1.3	0.6776	3.3	0.0416
1.4	0.6236	3.4	0.0328
1.5	0.5692	3.5	0.0252
1.6	0.5156	3.6	0.0185
1.7	0.4639	3.7	0.0127
1.8	0.4150	3.8	0.0077
1.9	0.3693	3.9	0.0035
2.0	0.3272	4.0	0.0000

An experiment is conducted using the normal distribution $N(0, 1)$. We obtain 10 numbers randomly from the standard normal distribution $N(0, 1)$. The average divergence is obtained as $\rho = 0.0594$ after 50 simulation experiments. Then, let $n = 12, \ldots, 22$ and respectively simulate 50 experiments. We then obtain Table 8, which shows the average divergence-ρ of the normal information diffusion estimate compared with the average divergence ρ' of the histogram estimate. The relative error of the histogram estimate and information diffusion method is calculated as $e = \frac{\rho' - \rho}{\rho'}$. The results show that information diffusion method is better than the histogram estimate. Roughly speaking, for a small sample, the information diffusion method can improve a histogram estimator to reduce the mean error by about 15.63 %.

For the two other experiments in exponential distribution and log-normal distribution, we refer to Tables 10 and 11 to show the average and divergences-p, p' of the histogram estimate and information diffusion method, respectively. Tables 10 and 11 show that, when n is small, the information diffusion method is better than the histogram estimate in relation to the exponential distribution and log-normal distribution.

When n is small, the method of information diffusion method is superior with respect to almost any distribution. Furthermore, for a given sample whose size is very large and which is drawn from a distribution, the new method is not the

Table 8. Flood disaster risk evaluation values.

Disasters level	Small flood	Medium flood	Large flood	Extreme flood
Exceedance probability risk	0.8228	0.5692	0.1672	0.0252
Return period (years)	1.2154	1.7569	5.9809	39.6825

Table 9. Average divergence and the relative error for $N(0,1)$.

n	10	12	14	16	18	20	22
	0.0594	0.0472	0.0446	0.0443	0.0432	0.0417	0.0403
	0.0637	0.0579	0.0570	0.0548	0.0540	0.0476	0.0453
	0.0664	0.1847	0.2190	0.1909	0.1994	0.1231	0.1109

Table 10. Average divergence and the relative error for $E(1.5)$.

n	10	12	14	16	18	20	22
	0.0876	0.0881	0.0854	0.0822	0.0821	0.0724	0.0708
	0.1225	0.1094	0.1022	0.0916	0.0885	0.0835	0.0793
	0.2845	0.1946	0.1646	0.1027	0.0729	0.1333	0.1072

Table 11. Average divergence and the relative error for log-normal distribution.

n	10	12	14	16	18	20	22
	0.0391	0.0262	0.0274	0.0388	0.0364	0.0373	0.0264
	0.0849	0.0647	0.0670	0.0544	0.0606	0.0545	0.0463
	0.5395	0.5951	0.5910	0.2868	0.3993	0.3156	0.4298

best because the large sample can provide abundant statistical information.

A study of the above simulation experiments reveals that the superiority of information diffusion method is dependent on whether we are blind to the population and whether the size of a given sample is small. In the experiments, the given sample is considered fuzzy due to its small size, so some benefits can be obtained by information diffusion method. The work efficiency of information diffusion method is about 35 % higher than that of the histogram estimate. That is, if no knowledge is available about the population from which the given sample is drawn, and if the sample size is small, we have to obtain more observations, adding about 35 %, to guarantee that the estimation is as good as the one given by the fuzzy method.

However, if we have a lot of knowledge about the population to confirm an assumption, the statistical object with respect to a given sample is clearer. So if the size of a given sample is large, there is an abundance of statistical information in the sample. In this case, it is unnecessary to replace the statistics with information diffusion method as little benefit can be obtained using it.

6 Conclusions

Disaster risk analysis is a complex multi-criteria problem crucial to the success of strategic decision making in disasters. In China floods occur frequently and cause significant property losses and casualties, and flood risk analysis of an area is important for flood disaster managers so they could implement a compensation and disaster-reduction plan. But Traditional statistics are frequently inaccurate, especially in small sample problems. In the present study, a comprehensive fuzzy method for flood disaster risk assessment is developed. This method provides an enhanced implementation of information diffusion process and better corresponds to the actual situation.

Disaster risk, as a natural or societal phenomenon, is neither precise nor certain. In the current paper, we use a fuzzy method of flood risk assessment based on VFS theory and improved information diffusion technique to improve probability estimation. In the improved information diffusion method, the diffusion velocity k is replaced with variable $k(u)$. The method is based upon VFS and the IIDM which has been tested in the example. The proposed method can be generalized as an integration of techniques and has been tested as stable and reliable. The results are consistent with practical problems.

In view of the theoretic system of flood risk assessment developed thus far and the fact that observed series of disasters are quite short or even unavailable, the method adopted in the paper is indisputably an effective and practical method.

As fuzzy risk analysis involves more imprecision, uncertainty, and partial truth in natural and societal phenomena, this research in flood risk assessment must promote the study of the foundations of fuzzy risks. Neither the classical models nor this proposed model govern the nature of the physical processes. They are introduced as a compensation for their own limitations in the understanding of the processes concerned. It is based upon an improved modeling framework, and it can be extended to some other disasters.

We hope that further technological developments in flood control and many new effective methods of flood risk analysis can be used to obtain prediction accuracy. And by conducting such analysis, lessons can be learned so that the impact of natural disasters, such as the floods in China, can be prevented or mitigated in the future.

Acknowledgements. This work is supported by a grant from the National Basic Research Program of China (Project No. 2007CB714107), a grant from the Key Projects in the National Science and Technology Pillar Program (Project No. 2008BAB29B08), and a grant from the Special Research Foundation for the Public Welfare Industry of the Ministry of Science and Technology and the Ministry of Water Resources (Project No. 201001080).

Edited by: T. Glade
Reviewed by: A. N. Morote and two anonymous referees

References

Altay, N. and Green, W. G.: OR/MS research in disaster operations management, European J. Operational Res., 175, 475–493, 2006.

Anderson, J. D. and Wendt, J. F.: Computational fluid dynamics, McGraw-Hill, 1995.

Breiman, L., Meisel, W., and Purcll, M.: Variable kernel estimates of multivariate densities, Technometrics, 19, 135–144, 1977.

Brown, C. B.: A fuzzy safety measure, J. Eng. Mechanics Division, 105, 855–872, 1979.

Carlin, B. P. and Louis, T. A.: Bayes and empirical Bayes methods for data analysis, Stat. Comput., 7, 153—154, 1997.

Chen, S.: Theory and model of variable fuzzy sets and its application, 1st Edn., Dalian Univercity of Technology Press, Dalian, 2009.

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

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

Chen, X. R.: Non-parametric Statistics, Shanghai Science and Technology Press, Shanghai, 1989.

Clements, D. P.: Fuzzy ratings for computer security evaluation, 1977.

Devroye, L. and Gyorfi, L.: Nonparametric Density Estimation, Wiley, 1985.

Dong, W., Shah, H., and Wongt, F.: Fuzzy computations in risk and decision analysis, Civil Eng. Syst., 2, 201–208, 1985.

Efstratiadis, A., Vasiliades, L., and Loukas, A.: Review of existing statistical methods for flood frequency estimation in Greece, EU COST Action ES0901: European Procedures for Flood Frequency Estimation (FloodFreq) – 3rd Management Committee Meeting, Prague, 2010.

Esogbue, A. O., Theologidu, M., and Guo, K.: On the application of fuzzy sets theory to the optimal flood control problem arising in water resources systems, Fuzzy Sets Syst., 48, 155–172, 1992.

Gu, C. and Li, D: Mathematical and Physical Equations, Higher Education Press, Beijing, 2002.

Guo, Y. and Chen, S. Y.: Application of Variable Fuzzy Sets in Classified Prediction of Rockburst, ASCE, Proc. Sess. GeoShanghai, 112–118, 2006

Hadipriono, F. C. and Ross, T. J.: A rule-based fuzzy logic deduction technique for damage assessment of protective structures, Fuzzy Sets Syst., 44, 459–468, 1991.

Hand, D. J.: Kernel discriminant analysis, Research Studies Press, Chichester, UK, 1982.

Hoffman, L. J., Michelman, E. H., and Clements, D.: SECURATE-Security evaluation and analysis using fuzzy metrics, IEEE Computer Society, 531 pp., 1978.

Huang, C.: Information diffusion techniques and small-sample problem, Int. J. Inf. Tech. Decision Making, 1, 229–250, 2002.

Huang, C. and Shi, Y.: Towards efficient fuzzy information processing: using the principle of information diffusion, Physica Verlag, 2002.

Huang, D., Chen, T., and Wang, M. J. J.: A fuzzy set approach for event tree analysis, Fuzzy Sets Syst., 118, 153–165, 2001.

Maskrey, A.: Disaster mitigation: A Community Based Approach, Oxford, Oxfam, 1989.

Palm, R.: Multiple-step-ahead prediction in control systems with Gaussian process models and TS-fuzzy models, Eng. Appl. Artif. Intell., 20, 1023–1035, 2007.

Parzen, E.: On estimation of a probability density function and mode, Ann. Math. Stat., 33, 1065–1076, 1962.

Saaty, T.: The Analytic Hierarchy Process, McGraw-Hill, New York, 1980.

Shang, H. and Jin, P.: Information Diffusion Method in Risk Analysis, Proceedings of the 5th International FLINS Conference, Computational Intelligent Systems for Applied Research, 189–197, 2002.

Slobodan, P. S.: System Approach to Management of Disasters, New Jersey, John Wiley & Sons Inc., 2011.

Smith, K.: Environmental Hazards: Assessing risk and Reducing Disaster, 2nd Edn., New York, Routledge, 1996.

Silverman, B. W.: Density estimation for statistics and data analysis, Chapman & Hall/CRC, 1986.

Wang, Y., Wang, D., and Wu, J.: A variable fuzzy set assessment model for water shortage risk: Two case studies from China, Human Ecol. Risk Assess., 17, 631–645, 2011.

Wertz, W.: Statistical density estimation: a survey, Vandenhoeck & Ruprecht, 1978.

Wilson, R. and Crouch, E. A. C.: Risk assessment and comparision: an introductin, Science, 236, 267–270, 1987.

Wu, L., Guo, Y., Chen, S., and Zhou, H.: Use of Variable Fuzzy Sets Methods for Desertification Evaluation, Comput. Int. Theory Appl., 38, 721–731, 2006.

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

Zhang, D., Wang, G., and Zhou, H.: Assessment on agricultural drought risk based on variable fuzzy sets model, Chinese Geograph. Sci., 21, 167–175, 2011.