

Analysis and modelling of flood risk assessment using information diffusion and artificial neural network

Qiong Li^{1*}, Xingwen Jiang² and Donghan Liu³

¹*School of Mathematics and Physics, Hubei Polytechnic University, Huangshi, Hubei, China*

²*College of Hydropower and Information Engineering, Huazhong University of Science and Technology, Wuhan, Hubei, China*

³*Hubei Polytechnic University, Huangshi, Hubei, China*

ABSTRACT

Floods are a serious hazard to life and property. The traditional probability statistical method is acceptable in analysing the flood risk but requires a large sample size of hydrological data. This paper puts forward a composite method based on artificial neural network (ANN) and information diffusion method (IDM) for flood analysis. Information diffusion theory helps to extract as much useful information as possible from the sample and thus improves the accuracy of system recognition. Meanwhile, an artificial neural network model, back-propagation (BP) neural network, is used to map the multi-dimensional space of a disaster situation to a one-dimensional disaster space and to enable resolution of the grade of flood disaster loss. These techniques all contribute to a reasonable prediction of natural disaster risk. As an example, application of the method is verified in a flood risk analysis in China, and the risks of different flood grades are determined. Our model yielded very good results and suggests that the methodology is effective and practical, with the potentiality to be used to forecast flood risk for use in flood risk management. It is also hoped that by conducting such analyses lessons can be learned so that the impact of natural disasters such as floods can be mitigated in the future.

Keywords: artificial neural network, information diffusion, flood, risk analysis, assessment

INTRODUCTION

Natural disasters are increasing alarmingly worldwide. Flooding is a common natural disaster which often causes loss of property and human life. Recent flooding disasters have shown the vulnerability of the so-called developed and developing countries to such events. In China, flood disasters occur frequently, and about two-thirds of the country's area is facing the threat of different types and degrees of floods, as a result of both natural and unnatural influences, such as social and economic factors (Chen, 2010). Natural disasters present a great challenge to society today. Flood risk assessment for an area is important for flood disaster managers to be able to implement a compensation and disaster-reduction plan. As severe floods occur frequently, flood risk assessment and management play an important role in guiding governments in making timely and effective decisions for disaster rescue and relief.

Risk management for the operation of an existing flood protection system is the sum of actions for a rational approach to flood disaster mitigation. Its purpose is the control of flood disasters, in the sense of being prepared for a flood, and acting to minimise its impact. It includes the process of risk analysis, which forms the basis for decisions on maintaining and improving the system.

Risk analysis is a challenging task at the present. Assessing flood risk is difficult because of the lack of objective measures of acceptable risk, scarcity of data, and an abundance of unknown probability distributions. Flood risk analysis methods progressed from the direct integral method, Monte Carlo method,

and mean first-order-second-moment method, to advanced first-order-second-moment method, second-order-method and JC method (design point in first-order second-moment method). Theories and methods of flood risk analysis have been established based on the work of several authors: e.g. Ang and Tang (1984), Ashkar and Rousselle (1981), Diaz-Granados et al. (1984), Kuczera (1982), Stedinger and Taylor (1982), Todorovic and Rousselle (1971), Todorovic and Zelenhasic (1970), Wood and Rodríguez-Iturbe (1975). Recently, many risk analysis approaches have been based on using linguistic assessments instead of numerical values. Using fuzzy sets theory (Zadeh 1965), data may be defined in vague, linguistic terms, such as low probability, serious impact, or high risk. These terms cannot be defined meaningfully with a precise single value, but fuzzy sets theory provides the means by which these terms may be formally defined in mathematical logic.

In traditional flood risk analysis, the probability statistics method is usually used to estimate hydrological variables' exceedance probability, because it is based on a mature basic theory and has easy application. However, problems exist in the feasibility and reliability of the method's outcomes, without considering its fuzzy uncertainty. Especially in the case of small samples, results based on the classical statistical methods are usually unreliable. It is also rather difficult to collect long sequence flood data and the sample is usually small. Information diffusion theory can be applied enabling as much useful underlying data as possible to be extracted from the sample and thus improving the accuracy of system recognition (Huang, 2002; Palm, 2007).

Information diffusion is a fuzzy mathematic set-value method for samples, which considers optimising the use of fuzzy information of samples in order to offset the information deficiency. In order to map the multi-dimensional space of a disaster situation onto a one-dimensional disaster space

* To whom all correspondence should be addressed.

☎ +086-13339903642; e-mail: liqiong070108@163.com

Received 13 January 2012; accepted in revised form 1 October 2013.

nonlinearly, and to test the grade criteria for flood disaster loss, resolving the non-uniformity problem of evaluating results expressed as disaster loss indexes, an artificial neural network model, back-propagation (BP) neural network, is suggested for evaluating the degree of flood disaster, where the disaster's degree of loss is a continuous real number.

In this study we propose a composite method, based on information diffusion and artificial neural networks, to establish a flood risk assessment model that can be applied with a small number of measured samples. This method is then successfully applied to flood risk analysis in China. The principle of the modelling framework is briefly described in the next section. This is a new attempt at applying information diffusion theory and artificial neural networks to flood risk analysis. Computations based on this flood risk model can yield an estimated flood damage value that is relatively accurate. The aforementioned model exhibits fairly stable results, even when using a small set of sample data. This also indicates that information diffusion technology is highly capable of extracting useful information and can therefore improve system recognition accuracy. This method can also be easily applied and understood, as illustrated by the example given.

MATERIALS AND METHODS

The essence of flood disaster risk analysis is to estimate the probability distribution of an index. Because of the incomplete data that is usually available, the application of traditional statistical methods cannot guarantee high precision in the results; fuzzy mathematical treatments are therefore necessary for a small sample size. This paper uses artificial neural network techniques and obtains continuous degree index values for the samples; the degree values for observed samples are then turned into fuzzy sets by information diffusion method, to finally produce the risk values. This method is tested in a case study, showing that it is superior to traditional statistical models, and offers a means improve on the results of traditional estimation.

Basis of artificial neural network methods

Artificial neural networks (ANN) are massive parallel interconnected networks of simple (usually adaptive) nodes which are intended to interact with objects of the real world in the same way that biological nervous systems do (Simon, 2009). ANN was proposed based on modern biology research concerning human brain tissue, and can be used to simulate neural activity in the human brain (Markopoulos et al., 2008). ANN has the topological structures of information processing, distributed in parallel. The mapping of input and output estimation responses is obtained via combinations of non-linear functions (Srivaree-Ratana et al., 2002).

In terms of their structures, neural networks can be divided into two types: feedforward networks and recurrent networks. In a feedforward network, the neurons are generally grouped into layers. Signals flow from the input layer through to the output layer via unidirectional connections, the neurons being connected from one layer to the next, but not within the same layer. The multi-layer perceptron (MLP) is perhaps the best known type of feedforward network. For the typical multi-layer perceptron of the feed-forward mode neural network there is an input layer, output layer, and hidden layer. Neurons in the input layer only act as buffers for distributing the input signals x_j to neurons in the hidden layer. Each neuron j in the hidden

layer sums up its input signals x_j after weighting them with the strengths of the respective connections w_{ji} from the input layer and computes its output y_j as a function f of the sum:

$$y_j = f\left(\sum w_{ji}x_i\right) \quad (1)$$

in which f can be a simple threshold function or sigmoidal, hyperbolic tangent or radial basis function. The output of neurons in the output layer is computed similarly.

The back-propagation (BP) algorithm, a gradient descent algorithm, is the most commonly adopted MLP training algorithm. It gives the change Δw_{ji} , the weight of a connection between neurons i and j , as follows:

$$\Delta w_{ji} = \eta \delta_j x_i \quad (2)$$

where:

η is a parameter called the learning rate and
 δ_j is a factor depending on whether neuron j is an output neuron or a hidden neuron.

For output neurons:

$$\delta_j = \left(\frac{\partial f}{\partial net_j}\right)(y_j^{(t)} - y_j) \quad (3)$$

and for hidden neurons:

$$\delta_j = \left(\frac{\partial f}{\partial net_j}\right) \sum_q w_{jq} \delta_q \quad (4)$$

In Eq. (3), net_j is the total weighted sum of input signals to neuron j and $y_j^{(t)}$ is the target output for neuron j .

The neural cell of each layer only affects the status of the next neural cell. If the expected output signals cannot be obtained in the output layer, the weight values of each layer of the neural cells must be modified. Erroneous output signals will be backward from the source. Finally, the signal error will arrive in certain areas with repeated propagation. After the neural networks' training procedure is complete, the forecast information can be analysed with weight values and thresholds.

Information diffusion method

Information diffusion is a fuzzy mathematic set-value method for samples, which optimises the use of fuzzy information of samples in order to offset the information deficiency (Huang and Shi, 2002). The method can turn an observed sample into a fuzzy set, that is, turn a single point sample into a set-value sample. The simplest model of information diffusion is the normal diffusion model.

Information diffusion: Let X be a set of samples, and V be a subset of the universe, $\mu: X \times V \rightarrow [0,1]$ is a mapping from $X \times V$ to $[0,1]$. $\forall (x, v) \in X \times V$ is called a kind of information diffusion of X on V , and satisfies 3 conditions as follows (Huang and Shi, 2002):

- It is decreasing. $\forall x \in X, \forall v', v'' \in V$, if $\|v' - x\| \leq \|v'' - x\|$, then $\mu(x, v') \geq \mu(x, v'')$, where μ is the diffusion function.
- $\forall x \in X$; let v^* be observed value of x , which satisfies $\mu(x, v^*) = \max_{v \in V} \mu(x, v)$.
- $\mu(x, v)$ is conservative. If and only if $\forall x \in X$, its integral value on the universe is 1: $\int_U \mu(x, u) du = 1$.

In particular, if the random variable's domain is discrete, suppose it is $U = \{u_1, u_2, \dots, u_m\}$, the conservation condition is $\sum_{i=1}^m \mu(x, u_i) = 1 \quad (\forall x \in X)$.

Let $X = \{x_1, x_2, \dots, x_m\}$ be a sample, and $U = \{u_1, u_2, \dots, u_m\}$ be the discrete universe of X . x_i and u_j are called the sample point and the monitoring point, respectively. $\forall x_i \in X, \forall u_j \in U$, we diffuse the information carried by x_i to u_j at gain $f_i(u_j)$ by using the normal information diffusion shown in Eq. (5).

$$f_i(u_j) = \exp\left[-\frac{(x_i - u_j)^2}{2h^2}\right], u_j \in U \quad (5)$$

where:

h is called the normal diffusion coefficient, calculated by Eq. (6) (Huang and Shi, 2002; Huang, 2005).

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

where $b = \max\{x_i\}$; $a = \min\{x_i\}$

$$\text{Let } C_i = \sum_{j=1}^m f_i(u_j) \quad (7)$$

We obtain a normalised information distribution on U determined by x_p , shown in Eq. (8).

$$\mu_{x_i}(u_j) = \frac{f_i(u_j)}{C_i} \quad (8)$$

For each monitoring point u_j , summing all normalised information, we obtain the information gain at u_j , which came from the given sample X . The information gain is shown in Eq. (9).

$$q(u_j) = \sum_{i=1}^n \mu_{x_i}(u_j) \quad (9)$$

$q(u_j)$ means that, with the information diffusion technique we infer that there are $q(u_j)$ (generally is not an integer) sample points in terms of statistic averaging at the monitoring point u_j . Obviously $q(u_j)$ is not usually a positive integer, but is certainly a number not less than zero. It is assumed that:

$$Q = \sum_{j=1}^m q(u_j) \quad (10)$$

where: Q is the sum of the sample size of all $q(u_j)$.

Theoretically, $Q = n$, but due to the numerical calculation error, there is a slight difference between Q and n . Therefore, we can employ Eq. (11) to estimate the frequency value of a sample falling at u_j :

$$p(u_j) = \frac{q(u_j)}{Q} \quad (11)$$

The frequency value can be taken as the estimation value of its probability. The probability value of transcending u_j is calculated by:

$$P(u_j) = \sum_{k=j}^m p(u_k) \quad (12)$$

where: $P(u_j)$ is the required risk estimation value.

Flood disaster risk assessment

According to the above theory, we can calculate the probabilities of each degree of flood disaster in China, based on the

historical data from 1950 to 2009 collected by the Ministry of Water Resources of the People's Republic of China (see Table 1).

Year	Disaster area (1 000 ha)	Inundated area (1 000 ha)	Dead Population (persons)	Collapsed houses (x 10 000)
1950	6 559.00	4 710.00	1 982	130.50
1951	4 173.00	1 476.00	7 819	31.80
1952	2 794.00	1 547.00	4 162	14.50
1953	7 187.00	3 285.00	3 308	322.00
1954	16 131.00	11 305.00	42 447	900.90
1955	5 247.00	3 067.00	2 718	49.20
1956	14 377.00	10 905.00	10 676	465.90
1957	8 083.00	6 032.00	4 415	371.20
1958	4 279.00	1 441.00	3 642	77.10
1959	4 813.00	1 817.00	4 540	42.10
1960	10 155.00	4 975.00	6 033	74.70
1961	8 910.00	5 356.00	5 074	146.30
1962	9 810.00	6 318.00	4 350	247.70
1963	14 071.00	10 479.00	10 441	1 435.30
1964	14 933.00	10 038.00	4 288	246.50
1965	5 587.00	2 813.00	1 906	95.60
1966	2 508.00	950.00	1 901	26.80
1967	2 599.00	1 407.00	1 095	10.80
1968	2 670.00	1 659.00	1 159	63.00
1969	5 443.00	3 265.00	4 667	164.60
1970	3 129.00	1 234.00	2 444	25.20
1971	3 989.00	1 481.00	2 323	30.20
1972	4 083.00	1 259.00	1 910	22.80
1973	6 235.00	2 577.00	3 413	72.30
1974	6 431.00	2 737.00	1 849	120.00
1975	6 817.00	3 467.00	29 653	754.30
1976	4 197.00	1 329.00	1 817	81.90
1977	9 095.00	4 989.00	3 163	50.60
1978	2 820.00	924.00	1 796	28.00
1979	6 775.00	2 870.00	3 446	48.80
1980	9 146.00	5 025.00	3 705	138.30
1981	8 625.00	3 973.00	5 832	155.10
1982	8 361.00	4 463.00	5 323	341.50
1983	12 162.00	5 747.00	7 238	218.90
1984	10 632.00	5 361.00	3 941	112.10
1985	14 197.00	8 949.00	3 578	142.00
1986	9 155.00	5 601.00	2 761	150.90
1987	8 686.00	4 104.00	3 749	92.10
1988	11 949.00	6 128.00	4 094	91.00
1989	11 328.00	5 917.00	3 270	100.10
1990	11 804.00	5 605.00	3 589	96.60
1991	24 596.00	14 614.00	5 113	497.90
1992	9 423.30	4 464.00	3 012	98.95
1993	16 387.30	8 610.40	3 499	148.91
1994	18 858.90	11 489.50	5 340	349.37
1995	14 366.70	8 000.80	3 852	245.58
1996	20 388.10	11 823.30	5 840	547.70
1997	13 134.80	6 514.60	2 799	101.06
1998	22 291.80	13 785.00	4 150	685.03

TABLE 1
Value of flood indexes for China over a 60-year period

Year	Disaster area (1 000 ha)	Inundated area (1 000 ha)	Dead Population (persons)	Collapsed houses (x 10 000)
1999	9 605.20	5 389.12	1 896	160.50
2000	9 045.01	5 396.03	1 942	112.61
2001	7 137.78	4 253.39	1 605	63.49
2002	12 384.21	7 439.01	1 819	146.23
2003	20 365.70	12 999.80	1 551	245.42
2004	7 781.90	4 017.10	1 282	93.31
2005	14 967.48	8 216.68	1 660	153.29
2006	10 521.86	5 592.42	2 276	105.82
2007	12 548.92	5 969.02	1 230	102.97
2008	8 867.82	4 537.58	633	44.70
2009	8 748.16	3 795.79	538	55.59

TABLE 2
Flood disaster rating standard

Disaster level	Damage area (1 000 ha)	Inundated area (1 000 ha)	Dead population (persons)	Collapsed houses (x 10 000)	Recurrence interval (years)	Grade number
Small flood	0~9 045	0~4 989	0~3 446	0~112.1	<2	1
Medium flood	9 045~14 197	4 989~8216.7	3 446~5113	112.1~247.7	2~5	2
Large flood	14 197~20 388	8 216.7~13 000	5 113~10 676	247.7~754.3	5~20	3
Extreme flood	20 388~80 000	13 000~50 000	10 676~10 0000	754.3~5 000	>20	4

TABLE 3
Disaster degree estimations based on the BP network evaluation

Sample	Degree value	Sample	Degree value
1	0.4968	31	1.1020
2	3.4372	32	2.6992
3	1.9236	33	1.6720
4	1.6588	34	3.4072
5	3.1968	35	2.0740
6	1.1336	36	0.6060
7	0.2592	37	0.3456
8	0.3244	38	1.9396
9	2.3124	39	2.3460
10	2.4576	40	1.7848
11	2.7404	41	2.7104
12	0.9520	42	3.2744
13	0.3892	43	1.8760
14	0.2496	44	3.0172
15	0.1996	45	2.2672
16	1.0908	46	2.100
17	1.7048	47	3.1532
18	1.4108	48	2.5816
19	1.4584	49	2.1920
20	1.4224	50	0.5588
21	1.8180	51	0.2980
22	1.7900	52	0.4520
23	1.8504	53	0.4144
24	1.9704	54	0.6748
25	1.3876	55	0.6804
26	4.0000	56	1.7912
27	1.7960	57	1.1428
28	0.9656	58	2.3168
29	1.7068	59	0.6428
30	1.9700	60	1.3928

We select the set of 60 records as the large sample, and then 30 records are randomly chosen to form a small sample in order to provide a comparison of the results obtained using the method for a small sample. Damage area, inundated area, dead population, and collapsed houses were chosen as the disaster indicator for flood risk analysis. By frequency analysis the disaster level is classified into 4 levels: small, medium, large and extreme (see Table 2).

In order to map the multi-dimensional space of the disaster situation onto a one-dimensional disaster situation, a relationship between the disaster degree and the disaster indexes is needed. Because it is impossible to describe the relationship using a related function, we adopt the 'simulation' and 'memory' of the neural networks in flood degree evaluation. This is because the advantages of neural networks can be used to simulate and record the relationship of the input and output variables in the complex 'function', through training and learning and without the use of any mathematical models.

We used damage area, inundated area, dead population, and collapsed houses as input variables and disaster grading value as an output variable, and then set the nodes of the input as 4 and of the output layers as 1. This follows from Kolmogorov's theorem (Hecht-Nielsen, 1987) that the number of nodes in the hidden layer is at least $2n + 1$, where n is the number of nodes in the input layer. Since $n = 4$, the number of nodes in the hidden layer is at least 9. Considering accuracy, we determined that the number of nodes in the hidden layer is 10. Thus, we obtained the topology structure (4, 10, and 1) of the neural networks for flood degree forecasting.

The four flood grades are small, medium, large and extreme flood, with degree values in the interval [0,1], [1,2], [2,3], [3,4]; we use the disaster grading standard boundary values (Table 2) as 5 four-dimensional training samples for training and learning in the BP neural network. Meanwhile initial parameters of BP model weights and biases are randomly assigned before the commencement of training. With 100 000 cycles of training and learning in the training samples, the global error of the networks is set $E = 10^{-6}$. Learning rate and impulse parameter of the network are changed adaptively, and function trainlm is used for fast training.

The calculated output values are compared with the expected values where the mean square error is 5.498909×10^{-8} , indicating a good fit. Thus the BP neural network has completed the training procedure. So we can use the BP network to forecast disaster degrees of all the samples with the weighting coefficients and the thresholds modified. The flood degree estimations are listed in Table 3.

Based on the BP neural network, the disaster degree values of the 60 samples can be calculated (see Table 3), that is the sample points set $X = \{x_1, x_2, \dots, x_n\}$. We selected the set of 60 records as the large sample whose degree value ranges from 0 to 4. Then 30 records are randomly chosen to form a small sample to compare with the large sample.

The universe of discourse, namely the monitoring points set, is taken as $U = \{u_1, u_2, u_3, \dots, u_{41}\} = \{0, 0.1, 0.2, \dots, 4.0\}$. The normalised information distribution of each x_i , that is, $\mu_{x_i}(u_j)$, can be obtained according to Eqs. (5)–(8), then, based on Eqs. (9)–(12), disaster risk estimation, namely probability

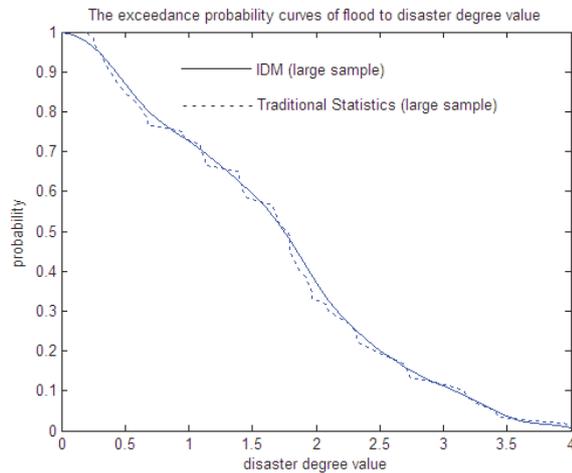


Figure 1

Comparisons of the risks by BP-IDM and the traditional statistical method

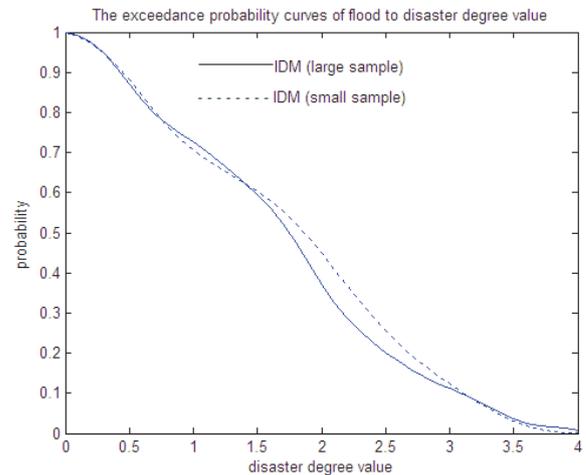


Figure 2

Comparison of the risks by BP-IDM with small sample and large sample

risk value, is calculated. The relationship between the recurrence interval N (years) and the probability p can be expressed as $N = 1/p$. The exceedance probability curve of flood disaster degree value is shown as Fig. 1, with a comparison to that obtained by the traditional statistical method.

RESULTS AND DISCUSSION

By information diffusion method (IDM), we obtain the exceedance probabilities on the different disaster degree values shown in Fig. 1.

In Fig. 1, the results reflect that the risk of the flood decreases smoothly with disaster degree value, using the BP-IDM model. The curve of the BP-IDM model is smoother and more accurate than that obtained using the traditional statistical method.

Thirty records were randomly chosen to form a small sample, and analysed in the same way for comparison with results from the large sample. These results are compared in Figs. 2 and 3.

Figure 2 provides a comparison of the two curves for estimated risk using the BP-IDM model with the small sample and large sample. The two curves match well, which indicates that the result is largely unchanged when the sample size changes; the method results are thus stable and not greatly affected by the size of the sample. If the analysis results for a very large sample are used as the standard, the BP-IDM method results are considered to be closer to the standard than the results of the statistical method, as proven by some experiments (Huang and Shi, 2002).

In Fig. 3, we compare two curves for the estimated risk with a small sample and large sample by frequency statistics. The mean error between the results for a large vs. small sample by frequency statistics is 0.0428, which is larger than that obtained for the BP-IDM model.

Figures 2 and 3 indicate that the results for the small sample analysed by the BP-IDM model are satisfactory. The results reflect the fact that the risk of flood decreases smoothly with the increase in disaster degree value, and that the BP-IDM model gives more satisfactory results than the statistical method for practical problems. The results obtained with the BP-IDM model are closer to the standard. Table 4 presents a comparison of the mean errors between the results with large vs. small samples

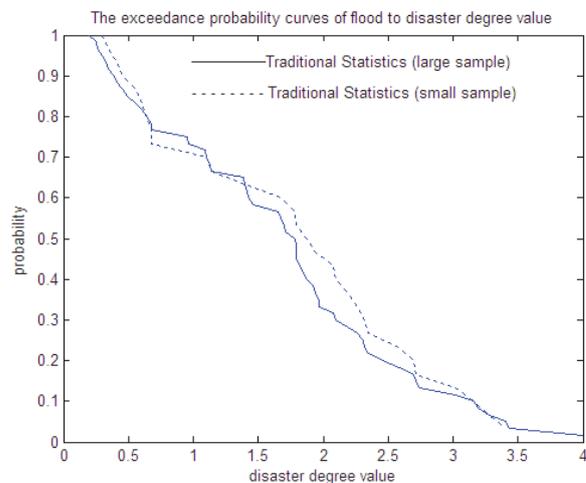


Figure 3

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

Method	BP-IDM	Statistics
Mean error	0.0227	0.0428

using the BP-IDM model and traditional statistics. Table 4 also shows that the mean error given by the BP-IDM model is much smaller than that given by the statistical method.

Flood disaster risk evaluation for China

The following 4 categories of flood degree are used (Chen, 2009):

- If $1.0 \leq H \leq 1.5$, then flood degree is small (1st grade).
- If $1.5 < H \leq 2.5$, then flood degree is medium (2nd grade).
- If $2.5 < H \leq 3.5$, then flood degree is large (3rd grade).
- If $3.5 < H \leq 4$, then flood degree is extreme (4th grade).

The result in Fig. 2 also illustrates the risk estimation, i.e., the exceedance probability of the disaster degree value. From this information, we know the risk estimation is 0.0356 when the

Disaster level	Small flood	Medium flood	Large flood	Extreme flood
Exceedance probability risk	0.7258	0.5955	0.2011	0.0356
Recurrence interval (years)	1.3778	1.6793	4.9727	28.0899

disaster indicator is 3.5. In other words, floods exceeding the 3.5 degree value (extreme floods) occur every 28.0899 years. Similarly, the probability of floods exceeding 2.5 degrees (large floods) is 0.2011, which means that floods exceeding that intensity occur every 4.9727 years. These findings indicate the serious situation relating to floods in China. The frequency and recurrence interval of the floods of the four grades are shown in Table 5.

Limitations

A study of some simulation experiments by Huang reveals that the superiority of the information diffusion method is dependent on whether we are blind to the population and whether the size of a given sample is small (Huang and Shi, 2002). 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 the 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 in order 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 the information diffusion method as little benefit can be obtained from this.

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. Flood risk analysis of an area is important for flood disaster managers to be able to implement a compensation and disaster-reduction plan. But results from use of traditional statistics for flood risk analysis are frequently inaccurate, especially in the case of small samples. In the present study, a comprehensive fuzzy BP-IDM method for flood disaster risk assessment is developed. This method provides an enhanced implementation of the information diffusion process which better corresponds to the actual situation.

In fact, 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 BP neural network and information diffusion technique to improve probability estimation, and test this method using an example. The proposed method can be generalised as an integration of techniques and has been tested as stable and reliable. In view of the theoretic system of flood risk assessment developed thus far and the fact that observed time-series of flood disasters are quite short or even unavailable, the method adopted in the paper is indisputably an effective and practical method.

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

ACKNOWLEDGMENTS

This work was 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).

REFERENCES

- ANG AHS and TANG WH (1984) *Probability Concepts in Engineering Planning and Design. Volume 2, Decision, Risk, and Reliability*. Wiley, New York.
- ASHKAR F and ROUSSELLE J (1981) Design discharge as a random variable: A risk study. *Water Resour. Res.* **17** 577–591.
- SRIVAREE-RATANA C, KONAK A and SMITH AE (2002) Estimation of all-terminal network reliability using an artificial neural network. *Comput. Oper. Res.* **29** 849–868.
- CHEN S (2009) *Theory and Model of Variable Fuzzy Sets and its Application* (1st edn.). Dalian University of Technology Press, Dalian.
- CHEN G (2010) *China's Floods in 2010: A Stress Test for State Capacity*, Singapore. East Asian Institute, National University of Singapore.
- DIAZ-GRANADOS M, VALDES J and BRAS R (1984) A physically based flood frequency distribution. *Water Resour. Res.* **20** 995–1002.
- HECHT-NIELSEN R (1987) Kolmogorov's mapping neural network existence theorem. *Proc. of the IEEE First Int. Conf. on Neural Networks*, San Diego, CA, Volume III. IEEE, Piscataway, NJ.
- HUANG C (2002) Information diffusion techniques and small-sample problem. *Int. J. Inf. Technol. Decis. Making* **1** 229–250.
- HUANG CF (2005) *Risk Assessment of Natural Disaster: Theory and Practice* (in Chinese). Science Press, Beijing.
- HUANG C and SHI Y (2002) *Towards Efficient Fuzzy Information Processing: Using The Principle of Information Diffusion*. Physica Verlag, Heidelberg.
- KUCZERA G (1982) On the relationship between the reliability of parameter estimates and hydrologic time series data used in calibration. *Water Resour. Res.* **18** 146–154.
- MARKOPOULOS AP, MANOLAKOS DE and VAXEVANIDIS NM (2008) Artificial neural network models for the prediction of surface roughness in electrical discharge machining. *J. Intell. Manuf.* **19** 283–292.
- PALM R (2007) Multiple-step-ahead prediction in control systems with Gaussian process models and TS-fuzzy models. *Eng. Appl. Artif. Intell.* **20** 1023–1035.
- SIMON H (2009) *Neural Networks and Learning Machines* (3rd edn.). China Machine Press, Beijing.
- STEDINGER JR and TAYLOR MR (1982) Synthetic streamflow generation: 1. Model verification and validation. *Water Resour. Res.* **18** 909–918.
- TODOROVIC P and ROUSSELLE J (1971) Some problems of flood analysis. *Water Resour. Res.* **7** 1144–1150.
- TODOROVIC P, ZELENHASIC E (1970) A stochastic model for flood analysis. *Water Resour. Res.* **6** 1641–1648.
- WOOD EF and RODRÍGUEZ-ITURBE I (1975) A Bayesian approach to analyzing uncertainty among flood frequency models. *Water Resour. Res.* **11** 839–843.
- ZADEH LA (1965) Fuzzy sets. *Inf. Control* **8** 338–353.