Flood Disaster Loss Evaluation Model Based on Projection Pursuit

Juliang Jin, Yu Zhang, Libing Zhang

College of Civil Engineering, Hefei University of Technology, Hefei, Anhui 230009, China, jinjl66@xinhuanet.com

Abstract: In order to resolve the non-uniformity problem of evaluation results of disaster loss indexes, and to raise the evaluation result precision of flood disaster loss, a projection pursuit（PP）based flood disaster loss model is suggested, where the results of flood disaster loss evaluation are real numbers. A scheme of PP modeling is also presented to reduce the computational amount, and a new function of projection indexes is given. It is suggested that both the function of projection indexes and the parameters of PP model can be optimized by using a real coding based genetic algorithm. The calculation example shows that the model is effective and general, which can be applied to evaluating other natural disaster loss. [Nature and Science 2003;1(1):82-85].

Key words：flood disaster loss; evaluation; projection pursuit (PP); genetic algorithm

1. Introduction

Flood disaster loss evaluation is to evaluate the damage degree caused by flood disaster according to flood disaster loss evaluation criterions, existing flood disaster loss evaluation index values and disaster loss evaluation model (Jin, 2002). The result of disaster loss evaluation, named as disaster loss grade, disaster grade or disaster degree (Yu, 1993; Zhao, 1993) is of important instructional significance to the flood disaster management. Flood disaster loss is involved with many factors including the natural environment and social economy etc. There are still no uniform evaluation index systems and grade criterions of flood disaster loss internationally. So evaluation problem of flood disaster loss is still one of difficulties and hotspots of researches on flood disaster.

Due to the non-uniformity of the flood disaster loss evaluation results of all indexes, it is difficult to evaluate disaster loss by directly using the disaster loss index criterions determined by historical experience of flood disaster (Jiang, 1996). So some models for evaluating disaster loss, including disaster loss degree measurement method, fuzzy comprehensive evaluation, matter element analysis, neural network and so on, have been presented one after the other (Jin, 2002;Yu, 1993; Zhao, 1997; Li, 1994). Their evaluation results are always discrete disaster grades, and the disaster grade precision is coarse. While the disaster loss indexes of actual flood disaster are usually continuous real numbers, in other words, even though the flood disasters are of the same disaster grade according to the actual models for evaluating disaster loss grade, their corresponding disaster loss index values are often different remarkably. So it is inconvenient to guide the flood disaster management.

How to synthesize a multi-index problem into a single-index problem scientifically and objectively is still the focus of researches on flood disaster loss evaluation, because only being in one dimension space makes it possible to flood disaster loss evaluate (Ren, 1998). Therefore, as an exploratory data analysis method, projection pursuit (PP) model for evaluating flood disaster loss, where the parameters are optimized by real coding based accelerating genetic algorithm (RAGA) (Jin, 2000), is suggested and applied to a case in this paper.

2. Flood Disaster Loss Evaluation Model Based on Projection Pursuit

PP is a kind of exploratory statistical method to analyze and process non-normal and high dimension data (Friedman, 1974; Li, 1997). Its basic idea is to project high dimension space data to projective values in low dimension space, to describe some structure by using a projective index function, to search the optimal projective index function, and to analyze the structure characters of the high dimension space data by the projective values, or to construct mathematical model according to the scattering dot figure formed by the projective values and the researched system output values. The problem to construct and optimize projective index function is the key to successfully applying PP method. The problem is very complex, and the computation of traditional PP methods is large (Friedman, 1974; Li, 1997), which restricts the wide application of PP technique and the deep study on it.

Here a simple scheme is presented, where a PP model for evaluating flood disaster loss is founded by using RAGA, which includes three steps as follows：

Step 1: to construct projective index function. Let $y(i)$

be the experiential grade of a certain flood disaster, which is produced according to the grade criterion table of flood disaster loss evaluation, and let $\{x^*(j,i)|j=1\sim p, i=1\sim n\}$ be the sample set of the disaster loss indexes, where n is the number of the flood disaster and p is the number of disaster loss indexes, respectively. The more the disaster loss is, the greater the disaster loss indexes are, and the higher the grades of the flood disaster loss are. Let the lowest grade of disaster loss be 1, and let the highest grade of disaster loss be N. Founding the model for evaluating flood disaster loss means constructing mathematical relation between $\{x^*(j,i)|j=1\sim p\}$ and $y(i)$. Here the aim of PP method is to synthesize the p dimension data $\{x^*(j,i)|j=1\sim p\}$ to one dimension $z(i)$ named projective value with the projective direction $\alpha=(\alpha(1), \alpha(2), \cdots, \alpha(p))$ by the following formula

$$z(i)=\sum_{j=1}^{p}\alpha(j)x(j,i) \tag{1}$$

where α is an unit length vector, then we can construct mathematical relation according to the scattering dot figure of $z(i)$ and $y(i)$. Equation (2) can be used to standardize the indexes both to eliminate the dimensional effect and to make the PP model be of generality:

$$x(j,i)=[x^*(j,i)-E_x(j)]/S_x(j) \tag{2}$$

where $\{x(j,i)|j=1\sim p\}$ are the standardized values of $\{x^*(j,i)|j=1\sim p\}$, $E_x(j)$ and $S_x(j)$ are mean value and standard deviation of the disaster loss index series $\{x^*(j,i)|i=1\sim n\}$.

When synthesizing projective value, the projective values should contain as much variation information of $\{x(j,i)\}$ as possible, in other words, the standard deviation S_z of $z(i)$ is as great as possible. Meanwhile, absolute value $|R_{zy}|$ of the related coefficient of $z(i)$ and $y(i)$ should be as great as possible. So the synthesized projective values can contain as much variation information of independent variable system $\{x(j,i)|j=1\sim p\}$ as possible, and can guarantee that the projective value is of good interpretability to attributive variable $y(i)$ (Ren, 1998). Based on the above demands, a projective index function can be designed as follows:

$$Q(\alpha)=S_z+|R_{zy}| \tag{3}$$

where | | is to calculate absolute value, S_z is standard deviation of projective value $z(i)$, namely

$$S_z=\left[\sum_{i=1}^{n}(z(i)-E_z)^2/(n-1)\right]^{0.5} \tag{4}$$

and R_{zy} is the related coefficient of $z(i)$ and $y(i)$, namely

$$R_{zy}=\frac{\sum_{i=1}^{n}(z(i)-E_z)(y(i)-E_y(i))}{\left[\sum_{i=1}^{n}(z(i)-E_z)^2\sum_{i=1}^{n}(y(i)-E_y(i))^2\right]^{0.5}} \tag{5}$$

In equations (4) and (5), E_z and E_y are the mean values of the series $\{z(i)\}$ and $\{y(i)\}$, respectively.

Step 2: to optimize the projective index function. The value of the projective index function $Q(\alpha)$ is changed only according to the variation of the projective direction α when the grade sample set of flood disaster loss $\{y(i), i=1\sim n\}$ and the disaster index set $\{x^*(j,i)|j=1\sim p, i=1\sim n\}$ have been determined. Different projective directions reflect different data structure character, and the optimal projective direction is the direction that best discovers some structure character of the high dimension sample data. The optimal projective direction can be estimated by resolving the following optimal problem:

$$\max\ Q(\alpha)=S_z+|R_{xy}| \tag{6}$$

$$\text{s.t.}\quad \sum_{j=1}^{p}\alpha^2(j)=1 \tag{7}$$

It is a complex and nonlinear optimization problem, where the optimized variables are $\{\alpha(j)|j=1\sim p\}$, and it is difficult to resolve the problem by using the traditional methods (Friedman, 1974; Li, 1997). As a kind of general optimization methods based on the mechanics of natural selection and natural genetics, RAGA can be applied to deal with the optimization problem easily and effectively (Jin, 2000).

Step 3: to found the PP model of evaluating flood disaster loss. The projective value $z^*(i)$ of the ith flood disaster can be gained by substituting the equation (1) with the optimal projective direction α^* according to Step 2. Then we can found the corresponding mathematical model according to the scattering dot figure of $z^*(i)\sim y(i)$. Present writers find what the scattering dot figure of $z^*(i)\sim y(i)$ reflecting here is monotonically increasing relation between $z^*(i)$ and $y(i)$: when $z^*(i)$ is greater than a certain threshold, it is determined to be the highest grade of flood disaster (Grade N); when $z^*(i)$ is less than another certain threshold, it is determined to be the lowest grade of flood disaster (Grade 1); when $z^*(i)$ is between the two thresholds, it is determined to be the medium grade of flood disaster. This is a relation that both the upper segment and the lower segment have limits, and the middle segment varies and increases rapidly and progressively. So it is appropriate to take Logistic Curve as the model of evaluating flood disaster loss, namely (Jin, 1997).

$$y^*(i)=\frac{N}{1+e^{c(1)-c(2)z^*(i)}} \tag{8}$$

where $y^*(i)$ is the calculated value of the grade of the ith flood disaster; the highest grade N is the upper limit of the Logistic Curve; $c(1)$ and c(2), being undetermined parameters, are integral constant and increase rate, and they can be determined by resolving the following minimization problem with using RAGA (Jin, 2000; Jin, 1997):

$$\min F(c(1), c(2)) = \sum_{i=1}^{n} [y^*(i) - y(i)]^2 \quad (9)$$

3. Case Study

Flood disaster area $x^*(1,i)$ and direct economic loss $x^*(2,i)$ are taken as the grade evaluation indexes of flood disaster loss. The frequency analysis was done of the actual series data of Henan Province of China between 1950 and 1990, and then the grade criterions of flood disaster loss of Henan Province are gained (Table 1) (Jiang, 1996).

Table 1. Evaluation Criterions of Flood Disaster Loss of Henan Province of China (Jiang, 1996)

evaluation indexes	ordinary disaster	fairly great disaster	great disaster	super disaster
flood disaster area (km^2)	<46.7	46.7~136.7	136.7~283.3	>283.3
direct economic loss (10^8 yuan)	<9.5	9.5~31.0	31.0~85.0	>85.0

Random evaluation indexes values and corresponding sample series of experiential grades of flood disaster loss can be gained by using the steps as follows: 1) The values of experiential disaster loss grades 1,2,3 and 4 are for the four values of ordinary disaster, fairly great disaster, great disaster, and super disaster, respectively. 2) The index value of the left extreme point of ordinary disaster can be determined as 0.5 times the index value of the right extreme point of ordinary disaster, and the index value of the right extreme point of the super disaster can be determined as 3 times the index value of the left extreme point of the super disaster. So every disaster loss grade has its index value range. 3) Five values can be gained by using uniform random number in the index value range of each disaster loss grade. The direct economic loss and the flood disaster area should have the same uniform random number in the same sample dot, considering that direct economic loss and flood disaster area are generally of positive relativity. 4) Every boundary value is chosen once from Table 1, and the corresponding disaster loss grade value is chosen as the arithmetic mean value of two disaster loss grade values related to the boundary value. So 23 sample dots are gained from No.1 to No.23 in Table 2.

Table 2. Comparison between Experiential Grade Values and Calculated Grade Values of PP Model

No.	evaluation indexes		$z^*(i)$	grades of flood disaster loss		No.	evaluation indexes		$z^*(i)$	grades of flood disaster loss	
	$x^*(1,i)$	$x^*(2,i)$		experiential	PP model		$x^*(1,i)$	$x^*(2,i)$		experiential	PP model
1	38.70	7.900	-1.179	1.0	1.375	17	157.30	38.600	-0.472	3.0	2.486
2	38.50	7.800	-1.180	1.0	1.374	18	283.30	85.000	0.424	3.5	3.499
3	32.10	6.500	-1.215	1.0	1.323	19	556.90	67.100	2.174	4.0	3.966
4	24.20	4.900	-1.257	1.0	1.264	20	649.50	194.900	2.766	4.0	3.987
5	36.40	7.400	-1.191	1.0	1.358	21	602.30	180.700	2.464	4.0	3.979
6	46.70	9.500	-1.136	1.5	1.438	22	446.50	134.000	1.468	4.0	3.897
7	97.60	21.700	-0.843	2.0	1.896	23	694.90	208.500	3.056	4.0	3.992
8	60.40	12.800	-1.057	2.0	1.558	1950	72.92	9.900	-1.047	2.0	1.573
9	112.60	25.200	-0.757	2.0	2.035	1954	148.13	20.656	-0.690	2.0	2.143
10	56.20	11.800	-1.081	2.0	1.521	1956	203.92	27.521	-0.437	3.0	2.538
11	80.60	17.600	-0.941	2.0	1.739	1957	179.10	24.858	-0.545	3.0	2.373
12	136.70	31.000	-0.618	2.5	2.258	1963	375.46	94.927	0.827	4.0	3.722
13	259.10	76.100	0.252	3.0	3.364	1964	301.24	47.836	0.092	3.0	3.213
14	200.10	54.400	-0.167	3.0	2.915	1975	141.97	116.439	0.295	3.0	3.400
15	280.10	83.800	0.401	3.0	3.482	1982	279.84	121.127	0.792	4.0	3.707
16	236.10	67.600	0.088	3.0	3.209	1984	172.06	51.619	-0.287	3.0	2.755

The disaster grade index values $\{x^*(j,i)|j=1\sim2, i=1\sim23\}$ in Table 2 are transformed into standardized series $\{x(j,i)|j=1\sim2, i=1\sim23\}$. The standardized series and disaster loss grade value series $\{y(i)| i=1\sim23\}$ substitute the equations (1), (4), (5) and (6) in turn, and then the projective index function of the example is gained. After the function is optimized by using RAGA, its maximum value is 2.34, and the optimal projective direction is α^*= (0.7066, 0.7076). Making α^* substitute the equation (1), and then the projective values $z^*(i)$ is gained (Table 2). The scattering dot figure of $z^*(i)\sim y(i)$ shows that equation (8) can be used to describe the relation of $z^*(i)$ and $y(i)$, where N is 4, $c(1)$ and $c(2)$ in equation (8) can be estimated by using RAGA to optimize equation (9). Then PP model for evaluating flood disaster loss grade in Henan Province is

$$y^*(i) = \frac{4}{1+e^{-1.2578-1.6152z^*(i)}} \quad (10)$$

where $y^*(i)$ is the calculated disaster loss grade value of the ith flood disaster. The calculated grade value of each flood disaster of PP model is in Table 2, and the result of error analysis between $y^*(i)$ and $y(i)$ is listed in Table 3.

Table 3. Error Analysis between Experiential Grade Values and Calculated Grade Values of PP Model

percent of absolute error falling the following range (%)						mean absolute error (disaster grade)	mean relative error (%)
[0,0.1]	[0,0.2]	[0,0.3]	[0,0.4]	[0,0.5]	[0, 0.6]		
34.78	43.48	60.87	82.61	95.65	100.00	0.22	13.42

Comparing the experiential grade values and the calculated grade values of PP model in Table 2 with the grade criterions in Table 1, the calculated values are reasonable, which more exactly describe the influence of the quantity difference of disaster loss index values on determining the disaster loss grade.

Table 3 shows that PP model can be used to describe the relationship between the projective values of flood disaster loss and the disaster loss grades. The 9 great disaster loss samples of 41 years data, which took place from 1950 to 1990 in Henan Province (Jiang, 1996), have been evaluated by using PP model, and their evaluation results can be seen from No.1950 to No.1984 in Table 2. Their experiential grade values are chosen as the evaluation results of the neural network model (Jin, 2002). The evaluation results of the two models are consistent basically, and the grade precision of disaster loss of PP model is higher. Now most of the calculated disaster loss grade values of other evaluation models are discrete, and they lack necessary transition range between the adjacent disaster loss grades. Take the flood disaster of 1950 for example, the disaster loss indexes of flood disaster area and direct economic loss are all near the bounders of the ordinary disaster and fairly great disaster, so it is reasonable that this year's disaster loss grade is evaluated as 1.573 by using PP model.

4. Conclusion

At the present time, the calculation results of the presented models for evaluating flood disaster loss are mostly discrete grades, and the grade precision of the calculation results are also coarse. For the sake of raising evaluation precision, a new model – PP model, where disaster loss grades are continuous real numbers, is suggested for evaluating flood disaster loss. A scheme of PP modeling is presented to reduce the computational amount and a new projection index function is given. It is suggested that both the function and the parameters of PP model can be optimized by using a real coding based genetic algorithm developed by the authors. The calculation example shows that PP model is effective and general, which can be applied to evaluating other natural disasters loss.

Foundation Item: The Excellent Young Teachers Program in Higher Education Institute of MOE, P.R.C. [Department of Education and Person (2002)350], Anhui Provincial Science and Technology Fund of Excellence Youth, Anhui Provincial Natural Science Foundation (No. 01045102), Chinese National Natural Science Fund Project (No. 40271024).

Correspondence to:
Juliang Jin
College of Civil Engineering
Hefei University of Technology
Hefei, Anhui 230009, Chnia
E-mail: jinjl66@xinhuanet.com

References

Friedman J H, Turkey J W. A projection pursuit algorithm for exploratory data analysis. IEEE Trans. on Computer 1974;23(9):881-890.

Jiang J, Ji X, Liu L, et al. Preliminary analysis of flood and drought disasters between 1950 and 1990 in Henan province. Journal of Catastrophology 1996;11(4):69-73 (in Chinese).

Jin J, Yang X, Ding J. Real coding based accelerating genetic algorithm. Journal of Sichuan University (Engineering Science Edition) 2000;32(4):20-24 (in Chinese).

Jin J, Yang X, Liu Y, et al. Application of genetic algorithm for estimating logistic curve parameters. System Sciences and Comprehensive Studies in Agriculture 1997;13(3):186-190 (in Chinese).

Jin J, Zhang X, Ding J. Projection pursuit model for evaluation grade of flood disaster loss. Systems Engineering—Theory & Practice 2002;22(2):140-144 (in Chinese).

Li Z, Deng X. Preliminary exploration of natural disaster loss assessment model based on matter element analysis. Journal of Natural Disasters 1994;3(2):28-33 (in Chinese).

Li Z. Projection pursuit technology and its progress of application. Journal of Nature in China 1997;19(4):224-227 (in Chinese).

Ren R, Wang H. Theory, methods and examples of data analysis for multi-variable statistics. Chinese National Defence Industry Press. 1998:109-113, 149-163 (in Chinese).

Yu Q. Localization and its improvement of disaster degree measurement method. Journal of Natural Disasters 1993;2(2):8-10 (in Chinese).

Zhao A, Ma Z. Research on assessment index system of natural disaster loss. Journal of Natural Disasters 1993;2(3):1-7 (in Chinese).

Zhao L, Wang K, Qiu P. Synthetic evaluation of disaster. Systems Engineering—Theory & Practice 1997;17(3):63-69 (in Chinese).