Chapter 16
Natural Disaster Risk Assessment Using Information Diffusion and Geographical Information System

Zhang Jiquan, Liu Xingpeng, and Tong Zhijun

Institute of Natural Disaster Research,
College of Urban and Environmental Sciences,
Northeast Normal University, Changchun, Jilin, 130024, P.R. China
`{zhangjq022,liuxp912,gis}@nenu.edu.cn`

Abstract. With the social and economic development, the losses caused by natural disasters were more and more serious. Natural disaster assessment, management and research are the important field, developing direction and hotspot issues on disaster science and geo-science in resent year. However, because most of the natural disasters are a small sample of events, and uncertainty of natural disasters, so natural disaster risk assessment is particularly difficult based on historical data. Information diffusion theory is useful method for natural disaster risk assessment based on small sample even; it is a fuzzy approach to quantitative analysis of the probability of natural disaster risk. Therefore, the information diffusion theory has unique advantages in natural disaster risk assessment and management. This chapter presents a Geographical Information Systems (GIS) and information diffusion theory-based methodology for spatio-temporal risk assessment of natural disasters, taking grassland fire disasters in the Northern China as the case study. Firstly, we discuss connotation and forming mechanism of natural disaster risk, basic theory and framework of natural disaster risk assessment and management. Secondly, we introduce information diffusion theories and Geographical Information Systems (GIS) in the form of definitions, theorems and applications comprehensively and systemically. Finally, we give the case study on application of information diffusion theory and Geographical Information Systems (GIS) on grassland fire disasters in the grassland area of the Northern China. We employed information matrix to analyze and to quantify fuzzy relationship between the number of annual severe grassland fire disasters and annual burned area. We also evaluated the consequences of grassland fire disaster between 1991 and 2006 based on historical data from 12 Northern China provinces. The results show that the probabilities of annual grassland fire disasters and annual damage rates on different levels increase gradually from southwest to northeast across the Northern China. The annual burned area can be predicted effectively using the number of annual severe grassland fire disasters. The result shows reliability as tested by two-tailed Pearson correlation coefficient. This study contributes as a reference in decision making for prevention of grassland fire disaster and for stockbreeding sustainable development planning. The fuzzy relationship could provide information to make compensation plan for the disaster affected area.

J. Lu, L.C. Jain, and G. Zhang: Handbook on Decision Making, ISRL 33, pp. 309–330.
springerlink.com

1 Introduction

A natural disaster is the effect of a natural hazard (e.g., flood, grassland fire disaster, hurricane, volcanic eruption, earthquake, or landslide) that affects the environment, and leads to financial, environmental and/or human losses. The resulting loss depends on the capacity of the population to support or resist the disaster, and their resilience [1]. This understanding is concentrated in the formulation: "disasters occur when hazards meet vulnerability"[2]. A natural hazard will hence never result in a natural disaster in areas without vulnerability, e.g. strong earthquakes in uninhabited areas.

Risk is the future safety. In quantitative terms, risk is often defined as the probability of an undesired outcome. Natural disaster risk is the appearance probability of some kind if disaster during a certain period. It is a function of disaster intensity and the appearance (probability). Risk assessment of the natural disaster is an important part of disaster reduction. Many international object of disaster reduction have involved this research. In many cases, however, it is practically impossible to precisely get the probability distribution we need. Sometimes the estimated values of the probabilities may be so imprecise as to be practically useless if we still regard them as crisp values.

For risk assessment of natural disaster, Completeness of the information is very important. However, in many cases, data are only a part of the facts, and information carried by them is incomplete. For example, a small sample which contains a few observations is incomplete data when we use them to study the natural disaster. Before there was fuzzy set theory, statisticians used to consider incomplete data in random uncertain viewpoint. For incomplete information, statistical methods are often ineffective. Fuzzy set theory provides a unifying framework for fuzzy information processing includes studying incomplete data. Information diffusion theory is an important method to process small sample using fuzzy set theory, and it was used in many areas such as risk analysis and disaster assessment.

Geographic Information Systems (GIS) are computer-based systems that store and process (e.g. manipulation, analysis, modeling, display, etc.) spatially referenced data at different points in time [3]. GIS may be used in archaeology, geography, cartography, remote sensing, land surveying, public utility management, natural resource management, precision agriculture, photogrammetry, urban planning, emergency management, military, navigation, aerial video, and localized search engines. The fuzzy processing and GIS were independent of each other in the past. At present, through the integration of information diffusion theory and GIS, fuzzy classification and identification in GIS, it can be widely used for many different problem domains including evaluating risk, risk zone, soil classification, crop-land suitability analysis, identifying and ranking wild fire risk.

Take grassland fire disaster as an example, this chapter presents the application of information diffusion theory in natural disaster risk analysis of small sample. Computations based on this analytical grassland fire disaster risk model can yield an estimated probability and burned areas value. This study indicates that the aforementioned model exhibits fairly stable analytical results, even when using a small set of sample data. The results also indicate that information diffusion theory and technology is highly capable of extracting useful information and therefore improves system recognition accuracy. This method can be easily applied and the analytical results produced are easy to understand. Results are accurate enough to act as a guide in disaster situations.

2 Basic Theory and Method of Information Diffusion

2.1 Information Diffusion

Information distribution [4], a method by which incomplete data is analyzed better, is a new way which allows us to divide an observation into two parts to belong to two subsets. The fundamental view of information diffusion is to cancel the restriction that an observation just belongs to two subsets in the domain. In information diffusion view, an observation can belong to all the subsets in the domain by different membership.

Let x_i ($i=1,2,3,\cdots,n$) be observations of natural disaster. Let X= $\{x_1,x_2,\ldots,x_n\}$, be a given sample and the universe of discourse be U. A mapping from X $\times$U to [0, 1], We call X a sample, and n its size.

$$\mu : \mathrm{X} \times \mathrm{U} \rightarrow [0,1]$$

$$(x,v) \mapsto \mu(x,v), \forall (x,v) \in X \times V$$

is called information diffusion of X on U, if it satisfies

(1) $\forall x_i \in$ X, if $u_o = x_i$, then $\mu(x_i,u_o)$ $=\sup_{u\in U}\{\mu(x_i,u)\}$

(2) $\forall x \in$ X, $\mu(x,u)$ is a convex function about u ;

$\mu(x,u)$ is called an information diffusion function of X on U. When U is discrete, the function also can be written as $\mu(x_i,u_j)$. μ is called a diffusion function and V is called a monitoring space. When $V=U$, we say that $\mu(x,v)$ is sufficient. The set $D(X)=\{\mu(x,u)|x\in X, u\in U\}$ is called the sample of fuzzy sets derived from X on U by diffusion μ .

Given an origin x_0 and a bin width h , the histogram is a frequency distribution on the intervals $A_j = [x_0+(j-1)h, x_0+jh], j=1,2,\ldots,m$. When X is incomplete, the corresponding histogram must be too rough. Information distribution method can improve it by using an allocation function instead of the associated characteristic function.

Let $\mu_j(j=1,2,3,\ldots,m)$ be the centers of the intervals $[x_0+(j-1)h, x_0+jh]$ of the histogram, the allocation function of X can be defined by

$$\mu_h(x_i,u_j)=\begin{cases} 1-|x_i-u_j|/h & if|x_i-u_j|\le h \\ 0 & if|x_i-u_j|>h \end{cases} \tag{1}$$

In fact, $\mu_h(x_i,u_j)$ is a membership function which indicates that x_i can belong to more than one interval.

A more reasonable estimation can be obtained by using Eq.(2).

$$\bar{f}(A_j)=\frac{1}{n}\sum_{i=1}^{n}\mu_h(x_i,u_j) \tag{2}$$

For $D(X)$, we have the principle of information diffusion to deal with small sample. This principle holds, at least, in the case of estimating a probability density function. For estimating a probability density function, in fact, the diffusion estimate is just a Parzen kernel estimate, but, there are many ways to do diffusion estimate. On the other hand, the whole of the kernel theory focuses on what properties the estimate possesses. Normal distribution, Poisson distribution and exponential distribution are most popular principle of information diffusion.

Take normal diffusion as example, we researched the similarities of information and molecules in diffusion action, and obtain a partial differential equation to represent the information diffusion. Solving the equation, we obtain the normal diffusion function

$$\mu(x,u)=\frac{1}{h\sqrt{2\pi}}\exp\left[-\frac{(x-u)^2}{2h^2}\right] \tag{3}$$

According to the relation between the membership and possibility, $\mu(x,u)$ would be regarded as some possibility that u will occur if x has occurred. The nearer the u is to x; the larger the possibility that u will occur. The largest possibility would be 1. For x, normalizing $\mu(x,u)$ along u, we have a membership function with respective to u,

$$\mu'(x,u)=\exp\left[-\frac{(x-u)^2}{2h^2}\right],u\in R. \tag{4}$$

Based on the two-point criterion and average distance assumption, we have Eq. (5) to calculate coefficient h.

$$h=\begin{cases} 0.6841(b-a) & for\quad n=5;\\ 0.5404(b-a) & for\quad n=6;\\ 0.4482(b-a) & for\quad n=7;\\ 0.3839(b-a) & for\quad n=8;\\ 2.6851(b-a)/(n-1) & for\quad n\geq 9, \end{cases} \tag{5}$$

where

$$b=\max_{1\leq i\leq n}\{x_i\},a=\min_{1\leq i\leq n}\{x_i\}.$$

Strictly speaking, the normalized distribution in Eq. (4) with the h calculated by Eq. (5) would be called simple normalized normal diffusion function. In short, we still call it normal diffusion function and write the corresponding fuzzy set A with x being centroid as

$$\mu_A(u) = \exp\left[-\frac{(x-u)^2}{2h^2}\right], u \in R. \tag{6}$$

The concept sketch of normal information diffusion can be shown as Fig.1.

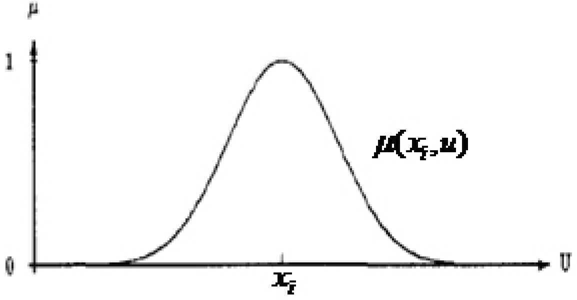

Fig. 1. Information diffusion function of x_i on a continuous universe of discourse U

2.2 Information Matrix

In the process of risk analysis of natural disaster, sometimes we should to determine the relationship between two variables, and in most cases, this relationship is not accurate, but a kind of fuzzy relations. Information matrix [5] is a useful tool to estimate the fuzzy relationships between natural disaster indicators.

Let $(x_i, y_i), i = 1,2,\ldots\ldots,n$, be observations of a given sample X with domain U of input and range V of output. Let $\mathrm{A}_\mathrm{j}, j = 1,2,\ldots\ldots,t$ and $B_k, k = 1,2,\ldots\ldots,l$ be fuzzy sets with membership functions $\mu_{A_j}(u), \mu_{B_k}(v),$ respectively. Let

$$\begin{cases} U = \{A_j \mid j = 1,2,\ldots\ldots,t\}; \\ V = \{B_k \mid k = 1,2,\ldots\ldots,t\}; \end{cases} \tag{7}$$

Their Cartesian product $U \times V$ is called an illustrating space. (A_j, B_k) is called an illustrating point.

Let $\circ$ be an operator defined on R (the set of real numbers).

$$q_{jk}(x_i, y_i) = \mu_{A_j}(x_i) \circ \mu_{B_k}(y_i) \tag{8}$$

is called information gain of (x_i, y_i) at (A_j, B_k) with respect to the operator.

Let

$$Q_{jk} = \sum_{i=1}^{n} q_{jk}(x_i, y_i). \tag{9}$$

Then

$$Q = \begin{array}{c c} & \begin{array}{cccc} B_1 & B_2 & \cdots & B_l \end{array} \\ \begin{array}{c} A_1 \\ A_2 \\ \vdots \\ A_t \end{array} & \begin{pmatrix} Q_{11} & Q_{12} & \cdots & Q_{1l} \\ Q_{21} & Q_{22} & \cdots & Q_{2l} \\ \vdots & \vdots & \vdots & \vdots \\ Q_{t1} & Q_{t2} & \cdots & Q_{tl} \end{pmatrix} \end{array} \tag{10}$$

is called an information matrix of X on $U \times V$.

3 The Application of GIS in Risk Assessment Natural Disaster

GIS (Geographic Information System) is an important tools, techniques and disciplines to acquisition, storage, analysis and management of geospatial data, in recent years has been widespread concerned and rapid development. Its powerful spatial analysis functions of geographic information are playing an increasingly important role in the GPS and route optimization. It is based on geospatial database, with the support of the computer's hardware and software, using the theory of systems engineering and information science, scientific management and comprehensive analysis of geographic data with spatial content, to provide information for management and decision-making. Simply, GIS is a technical system of comprehensive treatment and analyze geospatial data. Natural disaster management is a process that involves time, space and need to analysis of large amounts of data, so GIS has a wide range of application in natural disasters researches.

3.1 Components and Functions of GIS

GIS consists of hardware, software, data, personnel and methods. Hardware and software provide environment for GIS; Data is an important aspect to GIS; Methods provide solutions for the GIS; Personnel are the key and activity factors, it directly impact and coordinate other components.

Just for GIS itself, which mostly function is normally have five kinds of basic function, they are 1) The function of data collection and editing, 2) Attribute data editing and analysis, 3) The Function of Mapping, 4) The Function of Spatial database Management, and 5) The Function of Spatial Analysis.

3.2 Applications of GIS

Within the last 30 years, GIS has made remarkable development, it widely used in resources, environmental assessment, disaster prediction, land management, urban planning, telecommunications, transportation, military police, water electricity, utilities management, forestry, animal husbandry, statistics, business finance and almost all areas.

1) GIS and natural disasters risk information management

Because the factors influenced natural disasters is complexity, the natural disasters risk management data include real-time data monitoring data, historical disaster data, environmental data and regional socio-economic background data. These data have significant spatial geographical features. These regional and spatial characteristics for natural disasters management is a very important, such as natural background of different regions, population and pasture resources, the hut, and the spatial information distribution of other elements. GIS have the advantages to describe, manage, analyze and operate the spatial data, such as display the spatial data, inquire the correlative attribute data and orient goals. Particularly GIS powerful spatial analysis functions offer the effective tool for natural disasters risk management. Natural disasters risk assessment applying GIS mainly contains the following aspects:

(1) Natural disaster data maintenance

The natural disasters data input, and the output, updating and maintenance based on the electronic map. Maintain contents include regional basis geographical data (for example, administrative divisions, river and lake areas, settlements, population and contour lines), regional data on the natural environment (such as the average number of dry thunderstorms, drought index, wind speed, temperature, sunny days), natural disasters thematic data, natural disasters historical data (for example, when and where occurred fire, the yield loss, loss of livestock, the affected population and the impacted socio-economic, etc.). The data maintenance more intuitive and convenient based on the electronic map.

(2) Natural disasters data display and query

Using the function of positioning and layer management, the system could realize the display the all kinds of data from two-dimensional vector graphics applying electronic map, and even 3-D display if we have all kinds of needed data. Users could take zooming, roaming, and other operations, and could operate attribute data, such as integration operation.

(3) Natural disasters spatial analyze

Using specific spatial analysis of GIS, we could make a correlative analysis, such as buffer analysis for natural disasters influence area, Superposition analysis, and so on.

(4) Establish the risk management policy and emergency response plan

Based on upon operations, we could get adequate information on the natural disasters risk. It can offer provide for advancing measure and developing emergency response plans.

2) Natural disasters risk assessment is an important means and an important part of natural disasters risk management and a new perspective of the natural disasters risk Research. Natural disasters risk is a quantitative characteristic of natural disasters hazard and the possibility of the consequences become reality. According as the formation mechanism of the natural disaster risk, summarizes the formation mechanism of the natural disasters risk for the result of the interaction of hazard(H),exposure (E),vulnerability (V)and emergency response and recovery capability(R). Based on

the hazard (H), exposure (E), vulnerability (V) and emergency response and recovery capability(R) of the natural disasters risk, Choose the main influencing factors of the natural disasters that in including the natural factor and the social factor. Adopt to various methods such as linearity, Weighted Mathematical Model, the Analytic Hierarchy Process(AHP),the Weighted Comprehensive Analysis(WCA) , Information Diffusion Theory and Information Matrix, and so on , obtains the natural disasters risk index, and then to natural disasters risk degree in the study area for quantitative risk assessment. On this basis, supported by GIS, use its space analysis, study area is divided into a number of natural disasters risk areas, and conveys it in the form of digital maps, to verify the result.

3) GIS-based system of natural disasters risk management

The goal of this system constructs is the use advanced satellite remote sensing technology, the geographic information system, the database technology, the multimedia technology, the network technology , to geography space data comes from the different data pool, the different form, the different type carries on the unification processing. Integrating disaster risk assessment model, a powerful GIS spatial analysis functions in support, realize natural disasters hazards analysis, impact assessment and evaluation and prediction of loss, natural disasters vulnerability analysis, mitigation analysis, risk assessment and evaluation, for the fire risk integrated management and emergency response for the country and sub-regional providing the information and decision support.

The overall objective of system design is that friendly interface, easy operation and stable, reliable performance, a reasonable system structure and function, data safe and intelligent decision support for the natural disasters departments provide a practical, efficient, stable tool.

4 Case Analysis

Take grassland fire disaster as an example, this study presents a Geographical Information Systems (GIS) and information diffusion-based methodology for spatio-temporal risk analysis of grassland fire disaster to livestock production in the grassland area of the Northern China. Information matrix was employed to analyze and to quantify fuzzy relationship between the number of annual severe grassland fire disasters and annual burned area. We also evaluated the consequences of grassland fire disaster between 1991 and 2006 based on historical data from 12 Northern China provinces.

4.1 The Study Area and Statistical Analysis of Grassland Fire Disasters

Grassland fire disaster is a critical problem in China due to global warming and human activity. The northwestern and northeastern China face more challenges for mitigation of grassland fire disasters than other regions due to broad territory combined with the effects of complex physiognomy. According to statistical analysis of historical data of grassland fire disaster from 12 northern China provinces between 1991 and 2006, grassland fire disasters have been increasing gradually with economic development and population growth. The increased grassland fire disasters had significant impacts on the national stockbreeding economy. One of the main challenges is to establish the grassland fire disaster risk system so that the distribution of limited resources for disaster reduction and economic assistance can be made. Risk assessment is one of important means of natural disaster management. Risk assessment of natural disasters is defined as the assessment on both the probability of natural disaster occurrence and the degree of damage caused by natural disasters. In recent years, an increasing number of studies focus on natural disaster risk analysis and assessment of floodings, earthquakes and droughts among others. However less attention has been focused on grassland fire disasters. The occurrence of grassland fire disasters is due to both natural and human factors and their interactions. Traditional studies of grassland fire disaster are often limited to models in fire behavior, fire hazards and fire forecast. Disaster risk assessment has been used to manage wild fires. For example, Finney [6] calculated the wild fire risk by combining behavior probabilities and effects. Castro et al. [7] simulated the moisture content in shrubs to predict wild fire risk in Catalonia. Jaiswal et al. [8] applied a color composite image from the Indian Remote Sensing Satellite (IRS) and GIS for forest fire risk zoning in Madhya Pradesh, India. The grassland fire potentials depend on factors such as fuel type and density, topography, humidity, proximity to settlements and distances from roads. Thus grassland fire disasters occur randomly with uncertainties. In order to manage grassland fire disasters and compensate losses effectively, it is important to obtain probability and losses of grassland fire disasters on different risk levels.

In general, a disaster risk is defined as the outcome of probability multiplying potential losses. The main issues of risk assessment are implemented through estimating the probability distributions based on the historical data, which are usually substituted by frequencies. However, as grassland fire disasters are considered as the small sample event, it may not be unreasonable of using a frequency as a substitute instead of using an actual probability. Information diffusion was first put forward and proved with number theory. It is an effective method that can transform a traditional sample-point into a fuzzy set to partly fill the gap caused by incomplete data. Thus, information diffusion is suitable to deal with the small sample data. Yi et al. [9] estimated a frequency analysis method of flood disaster losses based on fuzzy mathematics theory of information diffusion with short time series of flood disaster samples. Huang et al. [10] discussed the benefit of the soft risk map calculated by the interior-outer-set model and proved that the soft risk map is better than the traditional risk map. Liu and Huang [11] established a fuzzy relationship between fire and surroundings based on the winter data of Shanghai. The purposes of the this part are to: (i) analyze the situation of grassland fire disasters, (ii) obtain the probabilities of annual grassland fire disasters and annual damage rates in different levels, and (iii) establish the fuzzy relationship between the number of annual severe grassland fire disasters and annual burned areas. The methodology in this study can be applied to study other natural disasters as well. The information from this study is potentially useful reference in decision making of grassland fire disaster prevention and stockbreeding sustainable development planning. The fuzzy relationship could provide information to make compensation plan for the disaster area.

Study Area

The main grassland areas in China lie in the arid and semi-arid northern 12 provinces of Inner Mongolia, Jilin, Heilongjiang, Liaoning, Xinjiang, Ningxia, Qinghai, Sichuan, Shaanxi, Shanxi, Hebei, and Gansu. Grassland fire disasters are one of the most concerned natural disasters there. According to historical data, 6,801 grassland fire disasters were reported over a period of 16 years between 1991 and 2006. About 6.22×10^8 ha of grasslands had been burned out in the northeastern and northwestern China, which represented an annual average of 3.89×10^7 ha.

Statistical Analysis of Grassland Fire Disasters

The grassland fire disaster is a natural disaster and a threat related to environment, people's livelihood, and socio-economic development, which is caused by the combination of both a natural hazard and a societal vulnerability. Spatial and temporal distributions of grassland fire disasters are determined by the interactions between natural condition and human activity. Historical data of grassland fire disaster in northern China from 1991 to 2006 illustrate the fluctuant trend of the number of grassland fire disasters and severe grassland fire disasters (Fig. 2). The trend indicates that the number of grassland fire disasters increased gradually, while the severe grassland fire disasters decreased (Fig.2). In recent years, although the provinces in northern China have strengthened management in fire prevention, grassland fire disaster risks are in increasing trend with economic growth and increased trend of human activities.

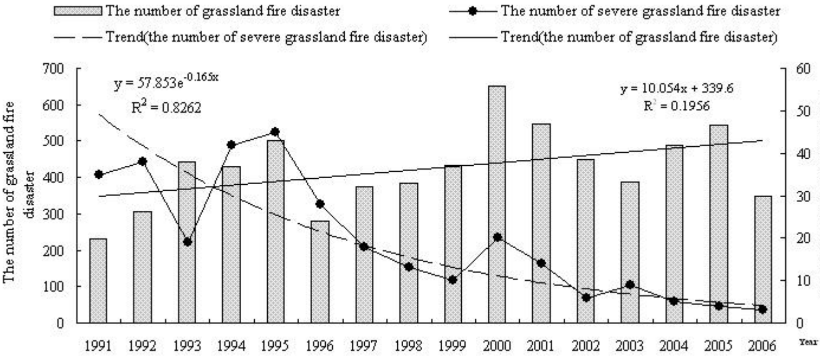

Fig. 2. Variation of grassland fire disasters and severe grassland fire disasters from 1991 to 2006 in northern China

Statistical data of fire seasons (from March to May and September to November) in northern China show a strong relationship between the number of grassland fire disasters and the burned area ($r = 0.98$, $p = 0.01$). The highest value emerged in April, followed by October, March, and May (Fig.3). The occurrence frequencies of April, October, March, and May account for about 41%, 16%, 13%, and 12%, respectively, while the burned area of April, October, March, and May account for about 47%, 14%, 8%, and 17%, respectively (Fig. 3).

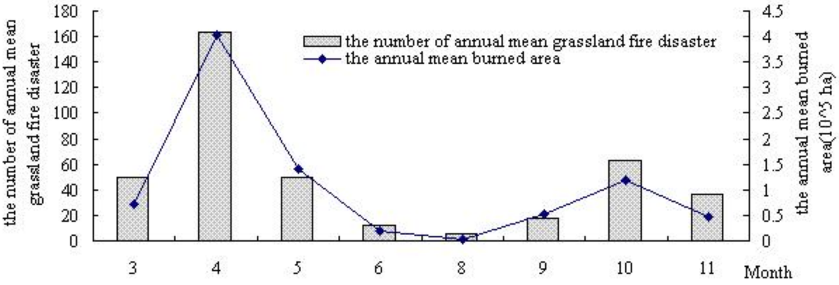

Fig. 3. Status of grassland fire disasters in fire seasons in northern China

For spatial distribution, grassland fire disasters in the provinces of Inner Mongolia, Hebei, Heilongjiang, and Xinjiang occurred more frequently than other provinces in northern China. The Inner Mongolia was the province with the most annual burned area (Fig.4). According to annual damage rates ($r = D_i / S_i$, r is annual damage rates, D_i is annual damage areas of province i, and S_i is grassland areas of province i), Heilongjiang and Inner Mongolia are the regions that are most likely affected by grassland fire disasters (Fig.5).

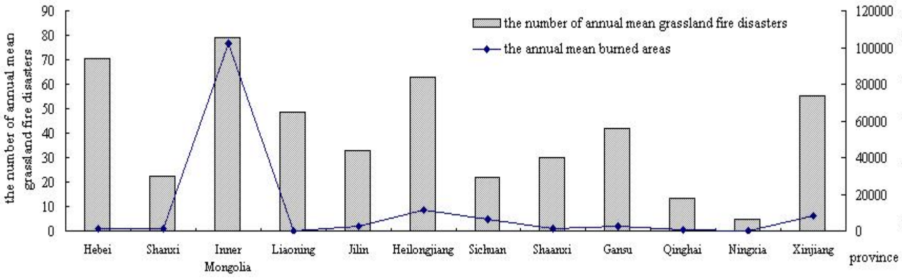

Fig. 4. Spatial distribution of grassland fire disasters in northern China provinces

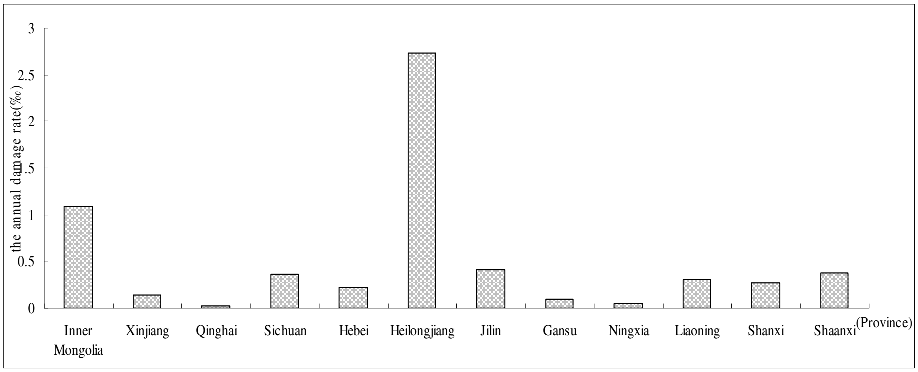

Fig. 5. Spatial distribution of annual damage rates of the grassland fire disaster in northern China provinces

In China, grassland fire disasters are classified into four grades with respect to its harmful level, i.e., grassland fire alarm, general grassland fire disaster, serious grassland fire disaster, and extraordinarily serious grassland fire disaster. Because serious grassland fire disaster and extraordinarily serious grassland fire disaster always have severe impacts on the environment, livelihood and socio-economic development we integrated those two grades into one category as the severe grassland fire disaster.

Data and Methods

Actual data sets are often incomplete due to various reasons in the study of grassland fire disaster [12]. Compared to accurate quantitative model, fuzzy mathematical model is a useful, objective technique to describe non-linear relationship in the grassland fire disaster. Information diffusion is an effective tool to deal with small sample event and establish the fuzzy nonlinear relationship. It has been widely used in disaster research such as earthquake and flooding among others. As a fuzzy mathematics method, information diffusion can offset the information deficiency in small sample event through transforming an observation sample into a fuzzy set.

Information Diffusion

Information diffusion is a fuzzy mathematics method that offset the lack of information in small sample event. The basic principle of information diffusion is to transform the crisp observations into fuzzy sets for partly filling up gaps caused by a scarcity of data so that the recognition of relationship between input and output could be improved. The detailed computational process is illustrated as follows.

Let $Y = \{y_1, y_2, y_3, \cdots y_n\}$ be a set of observations, called a given sample, and $U = \{u_1, u_2, u_3, \cdots u_m\}$ be the chosen framework space. If the observations cannot provide sufficient information to identify the precise relationship that is needed, then Y is called an incomplete data set. For any $y \in Y$, $u \in U$, the Eq.11 is called normal information diffusion.

$$\tilde{f}_j(u_i) = \frac{1}{h\sqrt{2\pi}} \exp[-\frac{(y_j - u_i)^2}{2h^2}] \tag{11}$$

Where y_j is a given sample, the universe U is the monitoring space, μ_i is the controlling points, h is the diffuse coefficient, and h is calculated using Table 1 [4].

Table 1. Relationship of the number of sample and diffuse coefficient

n	h	n	h
5	0.8146(b-a)	9	0.3362(b-a)
6	0.5690(b-a)	10	0.2986(b-a)
7	0.4560(b-a)	≥11	2.6851(b-a)/(n-1)
8	0.3860(b-a)		

h is diffuse coefficient, n is the number of sample, and b and a represents maximum and minimum value of sample, respectively.

It is assumes that

$$C = \sum_{i=1}^{n} \tilde{f}(u_i) \tag{12}$$

and the membership function of fuzzy subset is obtained as follows.

$$\mu_{y_j}(u_i) = \tilde{f}_j(u_i) \Big/ C \tag{13}$$

Where $\mu_{y_j}(u_i)$ is the normalizing information distribution of sample y_j.

Assumption

$$q(u_i) = \sum_{j=1}^{m} \mu_{y_j}(u_i) \tag{14}$$

The probability of u_i is calculated as follows.

$$p(u_i) = q(u_i) \Big/ \sum_{i=1}^{m} q(u_i) \tag{15}$$

In theory, $\sum_{i=1}^{m} q(u_i) = m$. The exceed probability ($p(u \ge u_i)$) of u_i is calculated as follows.

$$p(u \ge u_i) = \sum_{k=i}^{n} p(u_k) \tag{16}$$

Information Matrix

Let $H = \{(x_1, y_1), (x_2, y_2), \cdots (x_n, y_n)\}$ be a given sample, each sample has two elements of x as the input and y as the output. The input domain is U, $U = \{u_1, u_2, u_3, \cdots u_j\}$ j=1, 2, 3…, m, and output domain is V, $V = \{v_1, v_2, v_3, \cdots v_k\}$ k=1, 2, 3…, n. Eq.17 and Eq.18 are used to deal with information distribution.

$$q_{ijk} = \begin{cases} (1 - \dfrac{|u_j - x_i|}{\Delta x})(1 - \dfrac{|v_k - y_i|}{\Delta y}), & |u_j - x_i| < \Delta x \quad and \quad |v_k - y_i| < \Delta y \\ 0\,, & otherwise \end{cases} \tag{17}$$

$$Q_{kj} = \sum_{i}^{n} q_{ijk} \tag{18}$$

With Eq.17 and Eq.18, the simple information matrix ($Q = \{Q_{jk}\}_{m\times n}$) can be obtained on a $U \times V$ space. Based on the simple information matrix and applying Eq.19 and Eq.20, the fuzzy information matrix ($R = \{r_{jk}\}_{m\times n}$) can be obtained.

$$Q = \begin{array}{c} \\ u_1 \\ u_2 \\ \vdots \\ u_m \end{array} \begin{array}{c} \begin{array}{cccc} v_1 & v_2 & \cdots & v_n \end{array} \\ \begin{bmatrix} Q_{11} & Q_{12} & \cdots & Q_{1n} \\ Q_{21} & Q_{22} & \cdots & Q_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ Q_{m1} & Q_{m2} & \cdots & Q_{mn} \end{bmatrix} \end{array} \tag{19}$$

$$\begin{cases} S_k = \max\limits_{1\le j\le m} \{Q_{jk}\} \\ r_{jk} = Q_{jk} \big/ S_k \\ R = \{r_{jk}\}_{m\times n} \end{cases} \tag{20}$$

Fuzzy information matrix ($R = \{r_{jk}\}_{m\times n}$) establishes the fuzzy relationship of variables. It expresses the approximate relation of input-output. The basic element of fuzzy information is the fuzzy set described by a membership function. By max-min algorithm, the membership function of output variable could be described as Eq.21.

$$\tilde{I}_{x_i} = \bigvee_{x} [\mu(x_i, u_j) \wedge R] \tag{21}$$

Where, $\vee$ is symbol of max-product algorithm, $\wedge$ is symbol of min -product algorithm and the operator“ $\wedge$ ”is defined as expression, $\mu(e_i) = \min\{1, \Sigma \mu(x_i) r_{ij}\}$, R is fuzzy information matrix, and $\tilde{I}_{x_i}$ is the fuzzy output, and $\mu(x_i, u_j)$ is 1-dimension linear information distribution function. Let $X = \{x_1, x_2, x_3, \cdots x_m\}$ be a given sample, and $U = \{u_1, u_2, u_3, \cdots u_m\}$ be the chosen framework space with $\Delta = u_j - u_{j-1}$, j =2, 3…, m. For any x∈X, and any u∈U, the 1-dimension linear information distribution could be described as Eq.22.

$$\mu(x_i, u_j) = \begin{cases} (1 - \dfrac{|u_j - x_i|}{\Delta}), & |u_j - x_i| < \Delta \\ 0\,, & otherwise \end{cases} \tag{22}$$

Eq.21 is a fuzzy set about degree of membership calculated by fuzzy approximate reasoning. Because the information is too dispersive to apply in decision making, Eq.23 could calculate the fuzzy inference values and was employed to concentrate the information.

$$A_{x_i} = \sum_{k=1}^{n} \tilde{I}_{x_i}^{\alpha} \cdot v_k \Big/ \sum_{k=1}^{n} \tilde{I}_{x_i}^{\alpha} \tag{23}$$

Where, $\tilde{I}_{x_i}$ is fuzzy information distribution, v_k is the controlling points, α is constant, generally $\alpha = 2$, and A_{x_i} is the fuzzy inference value.

According to information diffusion theory, we calculated the probabilities of annual grassland fire disasters and annual damage rates in 12 northern China provinces based on the historical data from 1991 to 2006 collected by the Department of Grassland Fire Prevention, Supervision Center and Ministry of Agriculture of the People's Republic of China. Due to the significant importance of the annual burned area for policy makers and the strong correlations between the number of annual severe grassland fire disasters and annual burned area (r=0.728, p=0.01), we used the number of annual severe grassland fire disasters to predict the annual burned area in this study. The information matrix was employed during this process considering the efficiency of analyzing the fuzzy and nonlinear relationship between the annual several grassland fire disasters and annual burned area.

Results and Discussions

Probability Analysis of the Grassland Fire Disaster Risk

Traditionally, the probability of an event was calculated by frequency histogram. The frequency histogram is reasonable when the samples are abundant. However as a grassland fire disaster belongs to small sample event, we employed information diffusion to deal with the data of grassland fire disasters. By putting observation values of annual severe grassland fire disasters into the information diffusion function (Eq.11-Eq.16), Fig. 6 illustrates the exceed probabilities on different levels exceed probabilities.

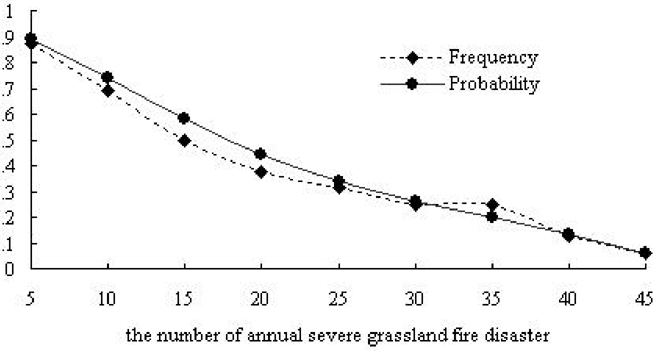

Fig. 6. Probabilities of annual severe grassland fire disasters based on information diffusion and frequencies

Fig.6 shows the scatter plot of measured (frequencies calculated by frequency histogram) and predicted (probabilities calculated by information diffusion) values of test samples in northern China. The predicted values agreed well with the measured values.

Fig.7 shows the correlation between measured and predicted annual probabilities of severe grassland fire disasters. The two-tailed Pearson correlation coefficient (r value) reached to 0.992 (p = 0.01). This shows that information diffusion is able to accurately predict the probabilities of grassland fire disasters on different levels.

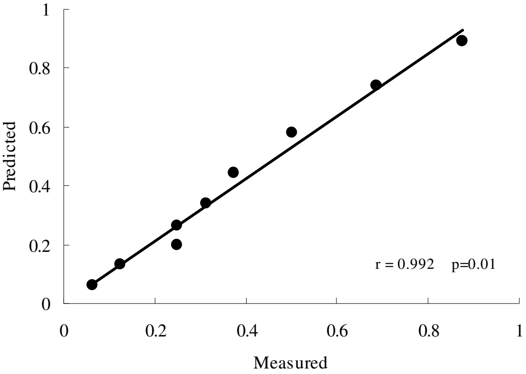

Fig. 7. Correlation between frequencies and probabilities of annual severe grassland fire disasters

The results in Fig.6 and Fig.7 indicate that probabilities obtained by normal information diffusion and frequency histogram are consistent. It means that normal information diffusion is useful to analyze probabilities of grassland fire disaster. If in the condition of having a complete dataset, the two methods should perform equally well. As the grassland fire disaster belong to the fuzzy events with incomplete data, therefore, the normal information diffusion is better than frequency histogram to analyze probabilities of the grassland fire disaster.

According to above method and the historical data of northern China, we calculated the probabilities of the number of annual grassland fire disasters and annual damage rates in 12 provinces using GIS. The results of spatial distributions are illustrated (Figs. 8, 9) show that the probabilities of annual grassland fire disasters and annual damage rates on different levels increased gradually from southwest to northeast across northern China. The Inner Mongolia and Heilongjiang are more easily affected by grassland fire disasters with high frequencies. The Qinghai and Shanxi provinces have less grassland fire disasters than other provinces in northern China.

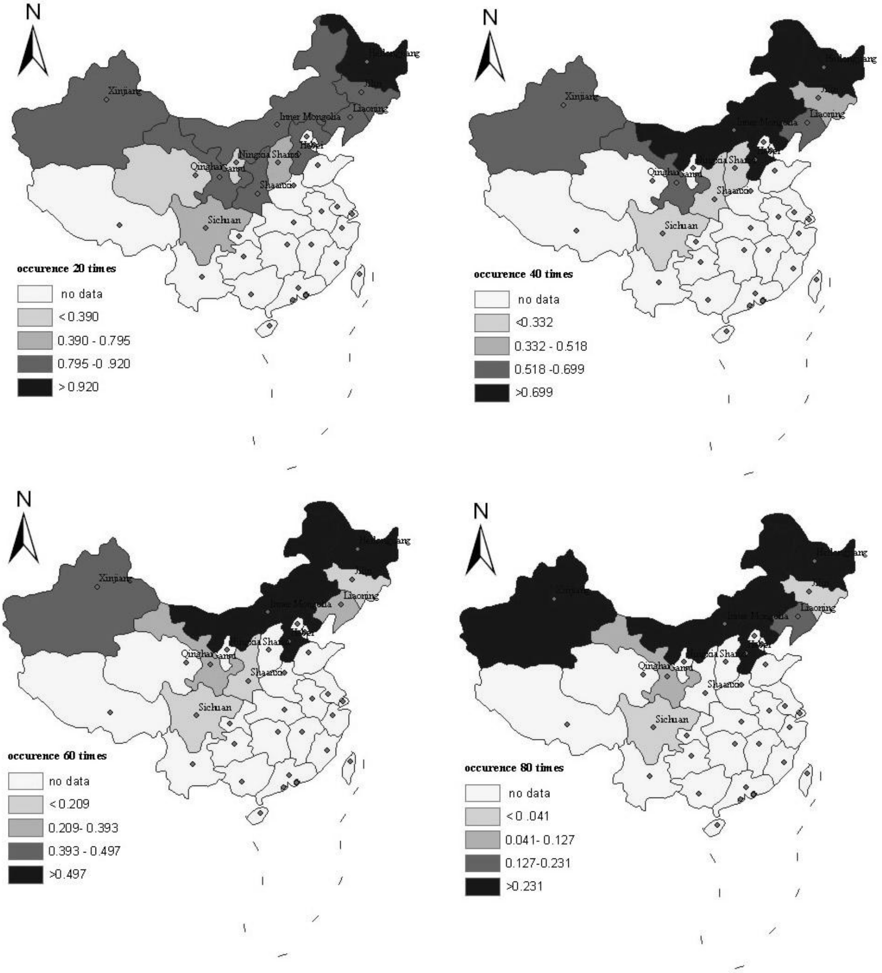

Fig. 8. Probabilities of annual grassland fire disasters on different levels in northern China

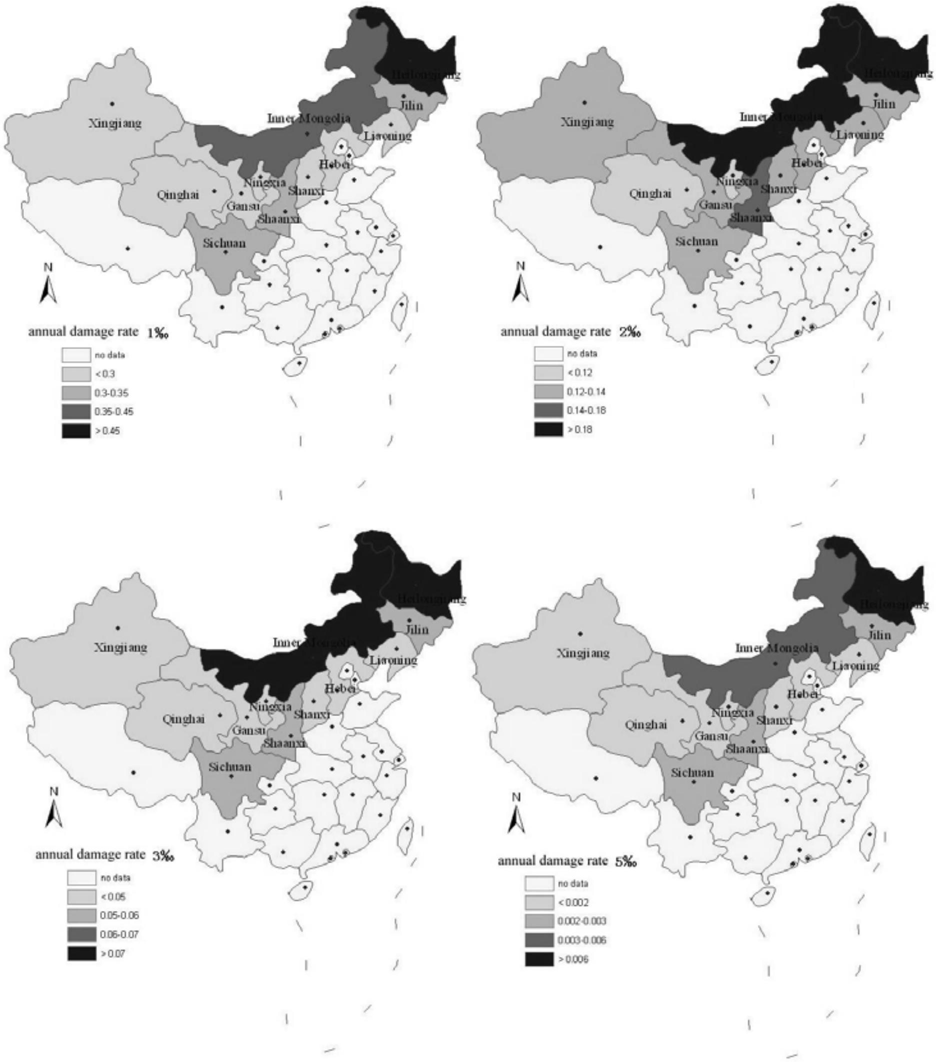

Fig. 9. Probabilities of annual damage rates on different levels in northern China

Prediction of Annual Burned Area

According to the method of information matrix, let X be the sample of annual severe grassland fire disasters, and Y be the sample of annual burned area.

$$X = \{35,38,19,42,45,28,18,13,10,20,14,6,9,5,4,3\}$$
$$Y = \{32.4,68.9,41.9,129.0,52.0,138.4,43.0,27.4,\cdots,4.74\}$$
$$H = \{X,Y\}$$

Correspondingly, we suppose that the chosen framework space of annual severe grassland fire disasters in input space is U, and the chosen framework space of annual burned area in output space is V. Annual severe grassland fire disasters range from 3 to 45, and nine controlling points are U= {5, 10, 15… 45}, step length Δ= 5. Chosen annual burned area V= {10, 20, 30… 140}, step length Δ=10.

Applying the observation values of two factors and information diffusion function (Eq.17-Eq.20), the simple information matrix and fuzzy matrix between two factors can be calculated as follows.

$$Q = \begin{array}{c|ccccc|} & 10.0 & 20.0 & 30.0 & \cdots & 140.0 \\ 5 & 1.63 & 0.00 & 0.30 & \cdots & 0.00 \\ 10 & 1.06 & 1.08 & 0.44 & \cdots & 0.00 \\ 15 & 0.54 & 0.41 & 0.89 & \cdots & 0.00 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 45 & 0.00 & 0.00 & 0.00 & \cdots & 0.00 \end{array} \qquad R = \begin{array}{c|ccccc|} & 10.0 & 20.0 & 30.0 & \cdots & 140.0 \\ 5 & 1.00 & 0.00 & 0.00 & \cdots & 0.00 \\ 10 & 0.65 & 1.00 & 0.33 & \cdots & 0.00 \\ 15 & 0.33 & 0.38 & 0.50 & \cdots & 0.00 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 45 & 0.00 & 0.00 & 0.00 & \cdots & 0.00 \end{array}$$

With the fuzzy relation matrix and Eq.21, we could obtain the corresponding annual burned area by a given observation of annual severe grassland fire disasters. For example, if we have an observation x_i =18, according to function Eq.22, the 1-dimension linear information distribution of observation x_i =18 can be obtained as follows (In all the formula, the signs "+" represent "and", the numerator represents subject degree, and the denominator represents discrete controlling point).

$$\mu(18, u_j) = \frac{0}{5} + \frac{0}{10} + \frac{0.4}{15} + \frac{0.6}{20} + \frac{0}{25} + \frac{0}{30} + \frac{0}{35} + \frac{0}{40} + \frac{0}{45}$$

According to the min -product algorithm, we obtain:

$$e = \frac{0.13}{10} + \frac{0.21}{20} + \frac{0.80}{30} + \frac{0.77}{40} + \frac{0.20}{50} + \frac{0}{60} + \frac{0}{70} + \frac{0}{80} + \frac{0}{90} + \frac{0}{100} + \frac{0}{110} + \frac{0}{120} + \frac{0}{130} + \frac{0}{140}$$

In order to concentrate the information, the max-product algorithm defined as follows.

$$\bar{e} = S_k \cdot e = (1.63 \times 0.13, 1.08 \times 0.21, 0.89 \times 0.80, \cdots, 0.5 \times 0) = (0.21, 0.23, 0.71, 0.82, \cdots, 0)$$

Normalize the $\bar{e}$, the final information distribution is obtained as follows.

$$\tilde{I}_{18} = \frac{0.27}{10} + \frac{0.28}{20} + \frac{0.87}{30} + \frac{1}{40} + \frac{0}{50} + \frac{0.19}{60} + \frac{0}{70} + \frac{0}{80} + \frac{0}{90} + \frac{0}{100} + \frac{0}{110} + \frac{0}{120} + \frac{0}{130} + \frac{0}{140}$$

In order to find the core of the information, the values of annual burned area can be calculated based on fuzzy inference (Eq.23).

$$A_{18} = \frac{10 * 0.27^{\wedge}2 + 20 * 0.28^{\wedge}2 + 30 * 0.87^{\wedge}2 + 40 * 1^{\wedge}2 + 50 * 0 + \cdots + 120 * 0 + 130 * 0 + 140 * 0}{0.27^{\wedge}2 + 0.28^{\wedge}2 + 0.87^{\wedge}2 + 1^{\wedge}2 + 0.19^{\wedge}2} = 34.40$$

According to above analysis, for one observation, $x_i = 18$, the annual occurrence probability is about 0.5 (from Fig.6); the maximum probable of annual burned area is 40; the secondary probable of annual burned area is 30; and the core of the information of annual burned area is 34.40. Compare with traditional probabilistic method, multi-valued risk result can provide more characteristics of risk system when we analyze the risk of system. By the methods described above, the results of annual burned area were calculated for the year of 1991 through 2006 (Fig.10).

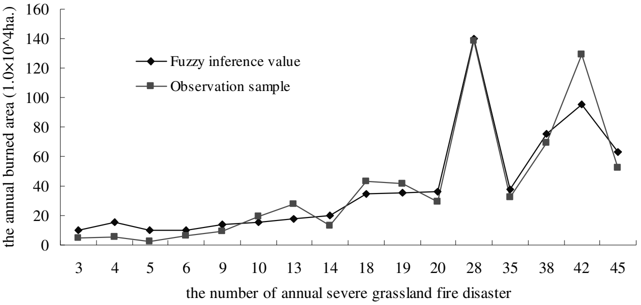

Fig. 10. A comparison between sample and fuzzy inference values of information diffusion matrix

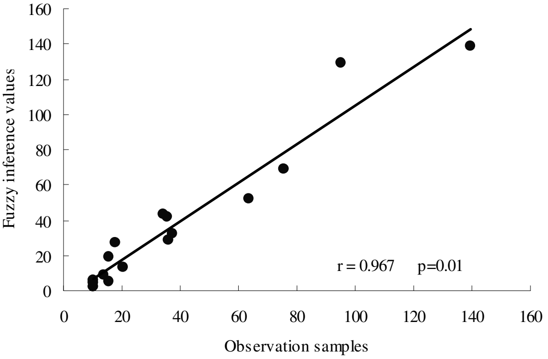

Fig. 11. Correlation between sample values and fuzzy inference values of annual burned areas

Fig.10 shows that the annual burned area simulated by information matrix matched well with the actual samples. The trends of annual burned areas between observation samples and fuzzy inference values are also consistent. From Fig. 10, it can be seen that annual burned area changes randomly, and it is impossible to be fitted by a linear function. It confirms the possibility of using the information diffusion theory for dealing with the fuzzy relationship of the grassland fire disaster. Fig.11 shows the correlation between observation samples and fuzzy inference values using information matrix. The two-tailed Pearson correlation coefficient (r value) reached to 0.967 (p = 0.01). This shows that information matrix is viable to accurately predict the annual burned area of grassland fire disasters on different levels. Because the relationship of between the number of annual severe grassland fire disasters and annual burned area is fuzzy, nonlinear, and could not be expressed by a linear relationship, the information matrix was employed and the annual burned area was predicted effectively using the number of annual severe grassland fire disasters with significant P values. The result could help in strategic decision makings to manage grassland fire disasters.

5 Conclusions

Grassland fires occur frequently in northern China provinces and cause significant property losses. In order to implement a compensation and disaster reduction plan, the spatio-temporal risk distribution of grassland fire disaster and the losses caused by grassland fires are among critically important information to grassland fire disaster managers. This study applied approximate reasoning of information diffusion to estimate probabilities and fuzzy relationship with scanty, incomplete data of the grassland fire disaster. Using information diffusion and information matrix, given observations, we can assess the spatio-temporal risk of grassland fire disaster and give an improved result to support risk management than traditional probability method.

Grassland fire disaster risk is an uncertain and complicated system. Considering the imprecise or incomplete information of grassland fire disaster, and the limits by the existing technique and experiment method, the analysis result could be inefficient and imprecise by employing only the traditional accurate model. If the risk values were estimated by classical statistics tool, we would lose information included in the given sample and could obtain wrong results. This study shows that information diffusion theory is effective to implement the impact and risk assessment associated with grassland fire hazards. The results could be used by the different level of authorities to develop grassland fire disaster risk protection plans. While the methods have certain advantages the accuracy in predicting could be improved by involving the impact factors of grassland fire disasters.

References

1. Bankoff, G., Frerks, G., Hilhorst, D.: Mapping Vulnerability: Disasters, Development and People (2003)
2. Wisner, B., Blaikie, P., Cannon, T., Davis, I.: At Risk - Natural hazards, people's vulnerability and disasters. Routledge, Wiltshire (2004)
3. Aronoff, S.: Geographic Information Systems: a Management Perspective. WDL Publications, Ottawa (1989)
4. Huang, C.F.: Principle of information diffusion. Fuzzy Sets and Systems 91, 69–90 (1997)
5. Huang, C.F., Moraga, C.: Extracting fuzzy if–rules by using the information matrix technique. Journal of Computer and System Sciences 70(1), 26–52 (2005)
6. Finney, M.A.: The challenge of quantitative risk analysis for wildland fire. Forest Ecology and Management 211, 97–108 (2005)
7. Castro, F.X., Tudela, A., Sebastià, M.T.: Modeling moisture content in shrubs to predict fire risk in Catalonia (Spain). Agricultural and Forest Meteorology 116, 49–59 (2003)

8. Jaiswal, R.K., Mukherjee, S., Raju, K.D., Saxena, R.: Forest fire risk zone mapping from satellite imagery and GIS. International Journal of Applied Earth Observation and Geoinformation 4, 1–10 (2002)
9. Yi, C., Huang, C.F., Pan, Y.Z.: Flood disaster risk analysis for Songhua river basin based on theory of information diffusion. In: 7th International Conference (ICCS 2007), part 3 (2007)
10. Huang, C.F., Inoue, H.: Benefit of soft risk map made by using information diffusion technique. In: International FLINS (Fuzzy Logic and Intelligent Technologies in Nuclear Science) Conference (2004)
11. Liu, J., Huang, C.F.: An information diffusion technique for fire risk analysis. Journal of Donghua University (Eng. Ed.) 21(3), 54–57 (2004)
12. Chen, J.N., Huang, H.K., Tian, F.Z., et al.: A selective Bayes classifier for classifying incomplete data based on gain ratio. Knowledge-Based Systems 21, 530–534 (2008)