Ecological Modelling 272 (2014) 264–270

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

The application of information diffusion technique in probabilistic analysis to grassland biological disasters risk

Lu Hao [a,b,*], Li-Zhe Yang [a,b], Jing-Min Gao [c,d]

[a] International Center for Ecology, Meteorology, and Environment, Jiangsu Key Laboratory of Agricultural Meteorology, Nanjing University of Information Science & Technology, Nanjing 210044, China
[b] Department of Applied Meteorological Science, Nanjing University of Information Science & Technology, Nanjing 210044, China
[c] Department of Climate Change and Public Policy, Nanjing University of Information Science & Technology, Nanjing 210044, China
[d] Binjiang College, Nanjing University of Information Science & Technology, Nanjing 210044, China

ARTICLE INFO

Article history:
Received 7 December 2012
Received in revised form 9 October 2013
Accepted 13 October 2013
Available online 7 November 2013

Keywords:
Normal information diffusion
Small sample
Probability
Risk assessment
Pest and rodent disaster

ABSTRACT

Biological disaster risk analysis is a complicated system. The incompleteness (gray areas), the non-clarity (fuzziness) and the uncertainty (randomness) of the data cause many difficulties that must be addressed with the risk assessment. In China, grasshopper and rodent disasters often occur in remote pastoral regions. This causes the monitored data of biological disaster to have a short series and span a large spatial and temporal scale. As available data are small sample in size, the use of risk assessment is often limited. The grassland biological disaster is a complex non-linear system. For the complex non-linear problems, effective conclusion can not be obtained from the accurate probability theory and mathematical statistics theory, but the fuzziness method may be a better method. In this paper, the one-dimension information diffusion technology adopted in evaluating the grassland biological disaster risk for the small statistical sample. The results show that: The information diffusion technology can make up for the information blank caused by the incompleteness of data, can change the single-valued samples into set-valued samples and excavate the internal law contained in the incomplete sample so as to achieve the aim of making full use of the information. It also can be seen that the diffusion results obtained under different starting control points or different interval step sizes have relatively good consistency and continuity. Based on such stability, a biological disaster risk forecast method can be derived, and the risk map using the reciprocals of different transcending probability values to demonstrate the regional differences on the same disaster level was also made by combining with GIS technology. Compared to other mature theories and technologies, the theory and method of fuzzy information optimization processing has its shortcomings especially in the selection of information diffusion function and information diffusion coefficient, and many improvements are needed.

1. Introduction

The essence of risk is uncertainty, and randomness and fuzziness are the most basic connotation of uncertainty. The uncertain problems involved in the disaster risk assessment are extremely complex because they do not only include the internal random uncertainty but also include the fuzzy uncertainty arising from the inaccuracy, fuzziness and incompleteness of information and data set. How to quantize and express the uncertainty in disaster risk assessment is difficult. At present, the most widely and maturely applied method is the probability description based on probability and mathematical statistics. However, this traditional method is only effective for the statistical risk. Because of the extreme complexity of natural disaster system, many input-output relations (including the relation between events and probability) can only be expressed as fuzzy relations. Therefore, the mathematical foundation of risk analysis should be fuzzy set theory (Zadeh, 1965).

Fuzzy set theory (Zadeh, 1965) and rough set theory (Pawlak, 1982) are two good methods to deal with uncertainty. Applying the fuzzy set theory to solve the uncertainty in the risks can better cope with the inconsistency and non-uniformity of data, the insufficiency of sample size and other problems.

In many circumstances, it is hard to find large enough sample in size. What a small sample reflects is incomplete information and it has fuzzy uncertainty. For a long time, people have always been seeking for the methods of analyzing the experimental results of small sample size. At present, there are mainly the

* Corresponding author at: College of Applied Meteorology, Nanjing University of Information Science & Technology, No. 219, Ningliu Road, Nanjing 210044, China. Tel.: +86 2558699957; fax: +86 2558699957; mobile: +86 18952067076.
E-mail address: hl_haolu@163.com (L. Hao).

0304-3800/$ – see front matter
http://dx.doi.org/10.1016/j.ecolmodel.2013.10.014

following methods: bootstrap, Bayes, Bayes bootstrap and Monte Carlo simulation. The bootstrap method (Efron, 1979) is a re-sampling process, i.e., using existing data to imitate an unknown distribution. This method can be applied to conduct the estimation of confidence intervals or the statistical hypothesis testing of parameters. The Bayes method (Berger, 1985) is to use a prior distribution; whether the prior distribution is reasonable or not can directly affect the assessment result. The key to using Bayes method is how to determine the prior distribution according to prior information. In practical application, however, there usually is little or no prior information. Bayes bootstrap method (Donald and Rubin, 1981) is a statistical processing method for the error and use a random weighted statistic to imitate the distribution of error of estimation; this method has a better effect than the Bootstrap method under mean square. Monte Carlo Simulation (Anderson, 1986) takes probability and statistics theory as main theory, and random sampling as main means. The method and its program structure are simple and it can solve the complex random problems without requirement for the simplification and hypothesis of mathematical model; but it has the shortcomings of slow convergence rate and large error.

The technology of fuzzy information optimization processing with "information distribution" and "information diffusion" as the core is an emerging data processing technology presented and developed by Chinese scholar Huang C.F. (2005). The object of fuzzy information optimization processing is the incomplete information, especially the fuzzy information of small sample. Simply speaking, the information diffusion can make up for the deficiency of sample information and change a traditional data set into a fuzzy set by optimizing the use of the sample.

The purpose of information diffusion is to excavate as much useful information as possible from the incomplete sample and enhance the accuracy of system identification. The most successful application of the technology is in the aspects of artificial neural network, geodetic data processing, risk analysis and evaluation system. In the aspect of risk analysis, the information diffusion has been applied in such disasters as earthquake, fire, drought, snowstorm and typhoon on (Li et al., 2010; Liu et al., 2010; Huang and Claudio, 2005; Hao et al., 2012).

With global warming (IPCC, 2007a,b) as well as severe degeneration and desertification due to overgrazing (Zhang et al., 2003), the grassland in China has an increasing trend for the incidence of biological disasters (Cheung, 2009; Shi et al., 2002; Yu et al., 2009). The pest- and rodent-induced disasters have become one of the three major natural disasters (drought, blizzard and pest/rodent) that affect the sustainable development of animal husbandry and cause the deterioration of ecological environment in the grassland of China (Zhang, 2003).

The implementation of risk analysis for grassland biological disasters is an important task aimed at reducing grassland losses. Many studies on pest risk analysis have generated a significant number of data (Yemshanov et al., 2010; Koch et al., 2009; Sutherst and Maywald, 1991; Strauss, 2010).

Due to the dual nature of uncertainty and randomness of the biological disasters, the risk analysis method uses mathematical methods such as probability and statistics to analyze the historical data of disasters and does not need an in-depth understanding on the causes and dynamic processes of disasters, so it is simple and operable. The key of risk statistical assessment is to estimate the conditional probability $P(h|t, s)$ occurred with intensity of h in a certain period of t, and in a given area of s (Huang, 2005). As the data series are short and the spatial and temporal scales are large for grassland pest/rodent monitoring, the pest/rodent disasters are considered as the small sample events. The implementation of risk assessment in China is usually limited due to the constraints of obtaining data.

In this paper, we applied information diffusion of the fuzzy mathematical approach in risk analysis of grassland biological disasters. We selected pests/rodents as the study subjects for grassland biological disasters and attempted to obtain relatively stable risk assessment results.

This paper mainly answers the following three questions:

(1) How to deal with the small sample size problem in grassland biological disaster risk analysis and the fuzzy uncertainty of sample data?
(2) How about the stability of information diffusion technology in grassland biological disaster risk analysis?
(3) How to use the information diffusion technology to realize probabilistic risk assessment to grassland biological disaster?

2. Materials and methods

2.1. Data resources

In our study, data were obtained from basic reports of grassland construction and utilization in The Chinese Animal Husbandry Yearbook (1999–2008) and prairie disaster statistics in the National Grassland Monitoring Report and Chinese Agriculture Yearbook (1997–1998). In addition, some indicators, such as the disaster index, were obtained through calculations with relevant formulae. The northern area encompasses Hebei, Inner Mongolia, Liaoning, Jilin, Heilongjiang, Shaanxi, Gansu, Qinghai, Ningxia and Xinjiang, a total of 10 provinces (Autonomous regions) (Fig. 1).

2.2. Information diffusion of the fuzzy mathematical approach

2.2.1. The information diffusion theory

Fuzzy mathematics is a mathematical method used to study and process unclear phenomena. The information diffusion is a set-valued fuzzy mathematical processing method that can change single-valued samples into set-valued samples. The purpose of information diffusion is to identify the maximum amount of useful information to improve the accuracy of risk recognition when the sample information is insufficient.

The information diffusion principle is as follows (Huang, 2005):

Let $X = \{x_1, x_2, \ldots, x_n\}$ be a given sample to estimate the relationship R of the universe U.

Assume γ is a reasonable operator, $\chi(x_i, u)$ is associate characteristic function, then the non-diffusion estimate is

$$\hat{R}(\gamma, X) = \left\{ \gamma(\chi(x_i, u)) \middle| x_i \in X, u \in U \right\} \quad (1)$$

If and only if X is not complete, then there must exist a diffusion function $\mu\,(x_i, u)$ and a corresponding operator γ'. Replace $\chi(x_i, u)$ with $\mu\,(x_i, u)$, and replace γ with γ', then the diffusion estimate is

$$\tilde{R}\left[\gamma', D(X)\right] = \left\{ \gamma'(\mu(x_i, u)) \middle| x_i \in X, u \in U \right\} \quad (2)$$

that satisfies

$$\left\| R - \tilde{R} \right\| < \left\| R - \hat{R} \right\| \quad (3)$$

where $||.||$ is the deviation between the estimated relationship and the true relationship.

Information diffusion principle ensures there must be a reasonable spread function that can improve non-proliferation estimate when a given sample is not complete, i.e., when X is not complete, there must be some way to be able to pick up fuzzy information of X, in order to more accurately estimate the function approximation of a relation R. However, this principle did not provide the method of how to find this spread function.

Fig. 2 gives the explanation of the information diffusion principle.