Applied Mathematical Modelling 36 (2012) 626–637

Contents lists available at ScienceDirect

Applied Mathematical Modelling

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

Analysis and prediction of hazard risks caused by tropical cyclones in Southern China with fuzzy mathematical and grey models

Hexiang Liu [a], Da-Lin Zhang [b,*]

[a] School of Mathematical Sciences, Guangxi Teachers Education University, Nanning, Guangxi 530023, PR China
[b] Department of Atmospheric and Oceanic Science, University of Maryland, College Park, Maryland 20742, USA

ARTICLE INFO

Article history:
Received 2 January 2011
Received in revised form 23 May 2011
Accepted 1 July 2011
Available online 12 July 2011

Keywords:
Combined weights
Fuzzy mathematical models
Hazard risk analysis
Exceeded probability
Tropical cyclones
Grey prediction model

ABSTRACT

A hazard-risk assessment model and a grey hazard-year prediction model (GHYPM) are constructed by integrating recent advances in the fuzzy mathematics, grey theory and information spread technique, and then applied to 17-year tropical cyclones (TCs) hazards in Southern China. In constructing the models, a genetic fuzzy mathematical algorithm is first developed to calculate the categorical and ranking weights of TC hazard impact and cause indicators, from which their combined weights are obtained after optimization. The hazard impact and cause index series are then found by coupling the combined weights with their corresponding down-scaled indicators. A two-dimensional normal-spread technique is employed to create a primitive information matrix and a fuzzy relation matrix in order to make fuzzy rough inference of hazard risks with the factorial space theory. An exceeded probability model is developed to assess the possibility of exceeding any given hazard-year category. Results from the GHYPM show that the simulated hazard risk values are more or less consistent with the hazard-impact index series, with more than 60% probability of exceeding a moderate hazard year in Southern China. Results also show small relative errors of the GHYPM, indicating its applicability to the prediction of TC hazard-years up to 20 years.

1. Introduction

Skillful analysis and accurate prediction of natural hazards are of extreme importance for hazard mitigation. However, the fuzzy and stochastic nature of hazards and the pertinent data collection usually cause data incompleteness for statistical analysis and prediction. In addition, the duration of useful data samples for natural disasters is often too short, for instance, only 15–25 years in China. Clearly, such a small sample size can hardly be analyzed, using conventional statistical or parameter-estimation methods, to reach scientifically meaningful conclusions. Thus, numerous fuzzy mathematics, grey models (GMs) and information spread techniques have been developed since 1960s [1] to aid in the risk assessment of natural disasters and sporadically occurring events.

For example, Sun and Ma [2] showed the important roles of fuzzy rough set models, based on certain fuzzy compatible relations, in decision-making when they were applied to clinical diagnoses. Hundecha et al. [3] developed a fuzzy-logic based rainfall-runoff model to analyze the flood-related hazard conditions and predict possible hazard risks. Iliadis and Spartalis [4] used the fuzzy logics to study the risks of forest fires in Greece. Yui et al. [5] applied a fuzzy multi-objective scheme to the short-term (<24 h) prediction of rainfall associated with tropical cyclones (TCs); TCs are defined herein as having

* Corresponding author. Tel.: +1 301 405 2018; fax: +1 301 314 9482.
E-mail address: dalin@atmos.umd.edu (D.-L. Zhang).

0307-904X/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.apm.2011.07.024

the maximum surface wind speed of at least 17.2 m s^{-1}. Cheng and Wang [6] constructed a prototype hazard index model for estimating TC hazards with a grey analytical hierarchy process (GAHP) – a grey correlation algorithm. Zhang and Liu [7] developed a fire-hazard predictive grey system model [GM (1, 1)] for fire accidents by taking advantage of the powerful predictive skill for uncertainties with GM (1, 1) theory.

According to Huang [8], we may view hazard risk analyses as a process to identify a functional relationship between the probability distributions of hazard causes (e.g., rainfall amounts or surface wind strengths) and hazard impacts or hazard carriers (e.g., on human-beings, buildings, crops) using information matrices (i.e., inputs vs. outputs). In this regard, Zhang et al. [9] analyzed the geographical characteristics of storm-surge hazards caused by TCs using the hazards data collected along the coastal region of Guangdong Province during the period of 1949–2005. In their study, a risk assessment model with the use of an information spread technique included was applied to TC storm-surge hazards by calculating the maximum water accumulation index at six tidal stations. Their model results were shown to compare favorably to those from a numerical storm-surge model and observations.

Since many natural hazards occur randomly and irregularly, Feng and Hong [10] introduced the concept of hazard entropy that reaches a maximum value under some given constraints. They found that the hazard damage and loss series follows the P-III distribution when the hazard entropy is maximized, and then estimated the recurrence time interval of future disaster losses using historical hazard data and the properties of the P-III distribution. Andrade et al. [11] proposed a stochastic model, based on the Bayesian and Monte Carlo Markov Chain algorithms, to improve the estimated allocation of water volumes in a reservoir for the control of flood waves. With conventional statistical analysis methods, Lu [12] calculated the hazard-risk index series (HRIS) of the TC-caused human casualties, inundations of agricultural area, and damaged houses over the Shanghai metropolitan region during the years of 1949–1990. Although Chen et al. [13] have later improved the HRIS, it is far from perfect and rigor in terms of either the selected indicators or the relative importance of various index series in hazard risk assessment.

While the previous studies have demonstrated the successful applications of fuzzy mathematics, GMs, and information spread techniques to hazard risk analyses, little attention has been paid to the development of an integrated hazard-risk assessment and prediction system by combining various weights, index series and models with fuzzy algorithms in order to extract valuable information from limited samples. In particular, few studies have been conducted to optimize hazard-impact index series (HIIS) and hazard-cause index series (HCIS) for the risk assessment of natural hazards. Thus, it is our intention to fill in this gap by developing a fuzzy GAHP-based, genetic projection pursuit algorithm, following closely the work of Jin et al. [14], and a GM (1, 1) model, following Liu et al. [15] and Liu and Zhi [16], for the prediction of TC hazard risk years, respectively. Then, they will be tested with the TC hazard data collected for the period of 1990–2007 over Southern China consisting of the provinces of Guangdong, Guangxi, and Hainan. Note that we assess the TC hazard risks herein with fuzzy mathematical models, based on the given hazard-cause intensity and hazard-carrier's vulnerability. Clearly, this risk assessment approach would contain many uncertainties due to the fuzzy and random nature of natural hazards, changes in hazard-carrier's vulnerability and societal preparedness, as well as the reliability of hazard indicators used in the risk assessment model.

The next section describes the database used for the present study, and shows how to construct a genetic projection pursuit algorithm, and then use it to determine the categorical and ranking weights of hazard-impact and hazard-cause indicators, optimize their combined weights, and eventually obtain the HIIS and HCIS by coupling the combined weights with the corresponding indicators. Section 3 discusses the construction of an exceeded probability model for assessing the probability of exceeding any TC hazard-year category for the given database. Section 4 presents the derivation of a grey hazard-year prediction model (GHYPM), based on the GM (1, 1) theory. A summary and conclusions are given in the final section.

2. Construction of a hazard risk assessment model

2.1. Data base

The TC hazard data used for this study are from the ''Dictionary of Hazards in China'' for the years of 1990–2000 [17], and some statistics hazard data for the years of 2001–2007 (even up to 2011) provided by the National Meteorological Data and Hazard Information Internet (http://www.laxf.gov.cn/qbzq/qbswebsite/default.asp). The TC-related rainfall and wind data are from ''Annual Report of Tropical Cyclones'' [18]. However, the year of 2004 is excluded in this study because of its anomalous fewer landfalling TCs, less damages and economic losses, and its relatively higher rate of missing data.

2.2. Construction of the hazard-impact index series

To obtain an index series that could characterize the variability of TC hazards, we must consider the following two factors: (a) data representativeness, and (b) data continuity. For these two reasons, we select the 17-year (i.e., 1990–2007) database that records relatively more complete TC hazards over Southern China than earlier years. Because TC hazard risks are mainly associated with human casualties, property damages, and recovery of agricultural and industrial production, we define the following five parameters as the TC *hazard-impact indicators*: the number of human mortalities (RM), the affected population (RP), the affected agricultural area (RA), damaged buildings (RB), and direct economic losses (RL). Similarly, we consider the following three factors as the TC *hazard-cause indicators* over Southern China: the annual-total number (CN), the annual-mean daily maximum rainfall amount (CR), and the annual-mean daily maximum wind (CW) of TCs. Keep in mind that the above two groups of hazard indicators will be listed herein in the same order as listed above in the HIIS and HCIS, respectively, unless otherwise stated. The same is also held for various weight series to be derived. It should be mentioned that some indicators are similar to those used by Lu [12] and Chen et al. [13]. Appendices A and B tabulate the original data, after some minor data quality control and needed calculations, in accordance with the above selected indicators.

As we know, the magnitudes of assigned weights will reflect the rankings and roles of individual indicators in the integrated fuzzy assessment of hazard risks. They also signify to some degree the accuracy of a multi-indicator assessment system and affect the quality of the final assessment results. To take full use of the categorical and ranking information of various hazard indicators in the original high-dimensional space, Jin et al. [14] showed the use of an accelerated generic algorithm (AGA) to calculate the combined weights from the categorical and ranking weights. Thus, we construct a fuzzy GAHP-based, genetic projection pursuit algorithm with the inclusion of some AGA algorithm in order to calculate the categorical and ranking weights, and their combined weights are then obtained after optimization. Steps to attain these weights are described below.

(a) Standardize the hazard-impact indicator matrix, x_{ij} ($i = 1, 2, \ldots, 5$; $j = 1, 2, \ldots, 17$), from the 17-year sample data, as tabulated in Appendix A. Normally, the larger the magnitudes of the selected indicators, the more meaningful are the final risk assessment results, which is the case herein. Thus, it is desirable to define the membership weight or down-scaled matrix r_{ij} ($i = 1, 2, \ldots, 5$; $j = 1, 2, \ldots, 17$) as

$$r_{ij} = \frac{x_{ij}}{\max\limits_{1\le j\le 17}\{x_{ij}\} + \min\limits_{1\le j\le 17}\{x_{ij}\}}, \quad i = 1, 2, \ldots, 5, \tag{1}$$

where $\max\limits_{1\le j\le 17}\{x_{ij}\}$ and $\min\limits_{1\le j\le 17}\{x_{ij}\}$ are the maximum and minimum values of the jth indicator, respectively. Calculation of Eq. (1) gives the following dimensionless, fuzzy weighted hazard-impact indicator matrix,

$$\mathbf{R} = (r_{ij})_{5\times 17} = 10^{-4} \times \begin{bmatrix} 876 & 4745 & 73 & 3029 & 5365 & 3942 & 10000 & 1989 & 201 & 401 & 1022 & 1314 & 1259 & 1460 & 0 & 4562 & 310 \\ 1444 & 4575 & 1016 & 9784 & 6255 & 8176 & 4816 & 3995 & 1216 & 1718 & 2381 & 7205 & 3129 & 6352 & 216 & 7380 & 1273 \\ 1265 & 2778 & 802 & 9716 & 4935 & 5465 & 3079 & 3027 & 647 & 1528 & 2988 & 6025 & 1721 & 3692 & 284 & 5414 & 890 \\ 763 & 1738 & 1535 & 5247 & 9958 & 3813 & 9026 & 5007 & 119 & 274 & 2582 & 3663 & 447 & 1943 & 42 & 4468 & 179 \\ 243 & 1760 & 515 & 5313 & 6371 & 4714 & 9851 & 2589 & 311 & 998 & 2406 & 8102 & 1082 & 2780 & 149 & 8071 & 897 \end{bmatrix}.$$

(b) Determine the categorical weights of the hazard-impact indicators, $A_1 = a_{i1}(i = 1, 2, \ldots, 5)$, using a genetic projection pursuit algorithm. This algorithm is useful for analyzing the statistical properties of multi-dimensional datasets with non-normal distribution and nonlinearity [14]. In particular, it can help determine the projection direction of data characteristics in a multi-dimensional space such that they can be examined in reduced dimensions. In this study, we will use this algorithm to project the multi-dimensional information (i.e., multiple parameters involved in hazard-impact and cause indicators) to a one-dimensional (1D) line embedded in the multi-dimensional space. This is done by first synthesizing the membership weight matrix r_{ij} ($i = 1, 2, \ldots, 5, j = 1, 2, \ldots, 17$) into a linear base function: $\boldsymbol{e} = (e_1, e_2, e_3, e_4, e_5)$, where $\boldsymbol{e}$ is a unit vector,

$$z_j = \sum_{i=1}^{5} e_i r_{ij} \quad (j = 1, 2, \ldots, 17). \tag{2}$$

To obtain more realistically the aforementioned weights, it is necessary to ensure the one-dimensional distribution of the projected points locally as dense as possible, better condensed in clusters, and globally as well spread as possible. For this purpose, we calculate the standard deviation (S_Z) of z_j,

$$S_Z = \sqrt{\frac{1}{16}\sum_{j=1}^{17}(z_j - \bar{z})^2}, \tag{3}$$

which measure the spread of the data. The local density (D_Z) of the projected points can be determined by,

$$D_Z = \sum_{i=1}^{5}\sum_{j=1}^{17}(K - d_{ij})u(t)(K - d_{ij}), \tag{4}$$

where the window radius of the local density is defined as $K = 0.1S_z$, distance $d_{ij} = |z_i - z_j|$, $t = K - d_{ij}$, and the unit step function $u(t)$ is 0 for $t < 0$ and 1 for $t \geqslant 0$. The associated objective function after projection [19] is then written as

$$Q_e = S_z D_z. \tag{5}$$

Since different projection directions reflect different characteristics of hazard-impact and cause indicators, an optimally projected direction should capture more important characteristics of the multi-dimensional information. In this study, the maximum value and optimized projection direction of Q_e are obtained by maximizing the projected objective function with a genetic algorithm, i.e.,

$$\begin{aligned} &\max Q_e = S_Z D_Z, \\ &\text{s.t.} \quad \sum_{i-1}^{5} e_i^2 = 1, \quad \text{and} \quad e_i \geq 0. \end{aligned} \tag{6}$$

This gives the optimized projection direction as $\boldsymbol{e}^*$ = (0.8075 0.3076 0.1511 0.2096 0.4319). Normalization of $\boldsymbol{e}^*$ yields the categorical weights of the hazard-impact indicators: $A_1 = (a_{i1};\ i = 1, 2, \ldots, 5)$ = (0.4233 0.1612 0.0792 0.1099 0.2264). The categorical weight series so obtained reveals that the RM indicator is much more significant than the other indicators.

(c) Determine the ranking weights of the hazard-impact indicators, $A_2 = (a_{i2};\ i = 1, 2, \ldots, 5)$, using the GAHP method. As we know, systematic assessment of irregular data is essentially an optimized ranking process. Clearly, the greater an element, r_{ij}, in the fuzzy indicator matrix, **R**, the more important role it has in determining the ranking weight of the indicators being concerned. This implies that for any i-th indicator the greater its algebraic addition, the more significant influence it may have on the optimized ranking weights. Let

$$s_i = \sum_{j=1}^{17} r_{ij}, \tag{7a}$$

and from

$$c_{ij} = \begin{cases} \frac{s_i - s_j}{max\{s_i\} - \min\{s_i\}}(c_m - 1) + 1, & s_i \geq s_j, \\ \left(\frac{s_j - s_i}{\max\{s_i\} - \min\{s_i\}}(c_m - 1) + 1\right)^{-1}, & s_i < s_j, \end{cases} \tag{7b}$$

where $\max\{s_i\}$ and $\min\{s_i\}$ are the maximum and minimum values of s_i, respectively, and $c_m = \min\{9, \int[\max\{s_i\}/\min\{s_i\} + 0.5]\}$ is a constant, we obtain a level 1-9 decision matrix **C** to be used for calculating the ranking weights,

$$\mathbf{C} = (c_{ii}) = 10^{-4} \times \begin{bmatrix} 10000 & 11137 & 6457 & 14512 & 9413 \\ 8979 & 10000 & 6015 & 13376 & 8503 \\ 15488 & 16624 & 10000 & 20000 & 14864 \\ 6891 & 7476 & 5000 & 10000 & 6607 \\ 10624 & 11760 & 6728 & 15136 & 10000 \end{bmatrix}.$$

According to fuzzy mathematics, the decision matrix $\mathbf{C} = (c_{ij}) = a_{i2}/a_{j2})\ i, j = 1, 2, \ldots, n)$ has the reflective (i.e., $c_{ii} = 1$ when $i = j$), reciprocal (i.e., $c_{ij} = 1/c_{ji}$), and consistency (i.e., $c_{ij}\ c_{jk} = c_{jk}$) properties. The consistency property implies that the quantitative relationship among different elements can be transformed, whereas a positive reciprocal matrix has a single eigenvalue, and it must be a positive real number corresponding to a positive eigenvector. Moreover, it can be shown that when an n-order positive reciprocal matrix **C** has the maximum eigenvalue of $\lambda_{\max} \geq n$, **C** has the consistency property [20]. Thus, calculation of the eigenvalues of **C** (i.e., **CIC**) gives the consistency indicator: CIC = 4.0647×10^{-4} that is much less than 0.10; see [20] for its rationality. Note that if the matrix **C** does not meet the consistency requirement, a modified decision matrix can be derived using the optimization algorithms in Matlab [14]. The ranking weights of the hazard-impact indicators so obtained are: $A_2 = (a_{i2};\ i = 1, 2, \ldots, 5)$ = (0.1361 0.2914 0.1932 0.1759 0.2034), which show the significant contribution of the RP indicator, and then the RL indicator. Since this result differs from that of the categorical weights, we have to calculate the combined weights for being used in our risk assessment model.

(d) Determine the combined weights between the categorical (a_{i1}) and ranking (a_{i2}) weights $A = (a_i;\ i = 1, 2, \ldots, 5)$. Solving the following optimization equation,

$$\begin{aligned} &\min\ F = \sum_{j=1}^{5}\sum_{i=1}^{5}\left(\mu|a_{i1} - a_i|r_{ij} + (1-\mu)|a_{i2} - a_i|r_{ij}\right), \\ &\text{s.t.} \quad \sum_{i=1}^{5} a_i = 1, \quad \text{and } a_i \geq 0; \quad i = 1, 2, \ldots, 5, \end{aligned} \tag{8}$$

gives the combined weights of the TC hazard-impact indicators: $A = (a_i;\ i = 1, 2, \ldots, 5)$ = (0.3362 0.1788 0.1261 0.1345 0.2244). Note that in Eq. (8) parameter μ is set to 0.5, implying that all categorical weights are considered equal contributions. The combined weights show that the RM indicator is relatively more significant than the other indicators, which is the same as that from the categorical weights.

Because the original hazard impact indicators range in several orders of magnitude, it is necessary to minimize the influences of the large different magnitudes on the ranking of HIIS. After some experimentation, we found it is more reasonable to down-scale the magnitudes of RM, RP, RA, RB, and RL to the range of [0, 1], and represented by $I_{d_j}, I_{f_j} I_{h_j}, I_{e_j}, I_{s_j}$, respectively, i.e.,

$$I_{d_j} = \begin{cases} \log d_j - 1, & d_j \geqslant 100 \text{ persons} \\ \frac{d_j}{100}, & d_j < 100 \text{ persons} \end{cases}, \quad I_{f_j} = \begin{cases} \log f_j - 2, & f_j \geqslant 10^6 \text{ persons} \\ \frac{f_j}{10000}, & f_j < 10^6 \text{ persons} \end{cases},$$

$$I_{h_j} = \begin{cases} \log h_j - 2, & h_j \geqslant 6667 \text{ hm}^2 \\ \frac{h_j}{100000}, & h_j < 6667 \text{ hm}^2 \end{cases}, \quad I_{e_j} = \begin{cases} \log e_j - 2, & e_j \geqslant 10^4 \text{ houses} \\ \frac{e_j}{10000}, & e_j < 10^4 \text{ houses} \end{cases}, \text{ and}$$

$$I_{s_j} = \begin{cases} \log s_j - 2, & s_j \geqslant 10^9 \text{ RMB} \\ \frac{s_j}{10}, & s_j < 10^9 \text{ RMB} \end{cases}.$$

Applying the combined weight series $A = (a_i; i = 1, 2, \ldots, 5)$ to the above down-scaled hazard-impact indicators leads to the following HIIS,

$$y_j = a_1 I_{d_j} + a_2 I_{f_j} + a_3 I_{h_j} + a_4 I_{e_j} + a_5 I_{s_j}. \tag{9}$$

Table 1 lists the values of y_j for the present study.

2.3. Construction of the hazard-cause index series

The HCIS can be constructed, following the same procedures as those associated with the HIIS. Specifically, the genetic projection pursuit algorithm is used to obtain the optimized projection direction (0.5377 0.0914 0.8382) for the hazard-cause indicators (i.e., CN, CR, and CW) and their dimensionless, normalized categorical weights (0.3665 0.0623 0.5713). This result indicates the more significant contribution of the CW indicator. Applying the GAHP method gives the decision matrix of ranking weights,

$$\mathbf{C} = (c_{ii}) = 10^{-4} \times \begin{bmatrix} 10000 & 5000 & 9352 \\ 20000 & 10000 & 19307 \\ 10693 & 5179 & 10000 \end{bmatrix},$$

and calculation of its eigenvalues (i.e., **CIC**) yields the consistency indicator: CIC = 9.6370×10^{-5}, which is much less than 0.10, indicating that the above decision matrix is satisfied with the matrix consistency requirements [20]. Then, we obtain the ranking weights (0.2451 0.4955 0.2594) showing more pronounced contribution of the CR indicator, and the combined weights $a'_k = (0.3750\ 0.4415\ 0.1835)$, ($k = 1, 2, 3$), after performing optimization. The combined weights show more important contributions of the CR indicator, followed closely by the CN indicator.

Table 1
The hazard-impact index (y_j), hazard-cause index (x_j), and the assessed hazard-risk index ($\tilde{y}_j$) caused by TCs over Southern China in Years 1990–2007.

No.	Year	Impact index	Cause index	Risk index
1	1990	0.60	54.39	0.84
2	1991	0.87	48.89	0.66
3	1992	0.09	45.47	0.48
4	1993	1.10	53.32	0.78
5	1994	1.17	50.97	0.72
6	1995	1.07	52.69	0.78
7	1996	1.25	54.52	0.84
8	1997	0.83	51.85	0.74
9	1998	0.67	36.27	0.45
10	1999	0.19	40.72	0.37
11	2000	0.59	42.91	0.42
12	2001	0.90	53.86	0.78
13	2002	0.44	44.08	0.41
14	2003	0.75	51.19	0.73
15	2005	0.34	47.85	0.57
16	2006	1.14	55.23	0.84
17	2007	0.37	50.31	0.67

Unlike the hazard-impact indicators, the original data associated with hazard-cause indicators do not exhibit large differences in magnitude. Thus, we may construct the HCIS by summing the weighted average of the original data directly with the combined weights obtained above. Let the CN, CR, and CW index series be x_{1j}, x_{2j}, x_{3j} ($j = 1, 2, \ldots, 17$), we attain the HCIS as

$$x_j = \sum_{k=1}^{3} a'_k x_{kj}, \quad j = 1, 2, \ldots, 17. \tag{10}$$

See Table 1 for the values of x_j.

2.4. Assess hazard risks using the two-dimensional normal spread and factorial space theory

As mentioned before, assessing natural hazard risks deals mainly with the identification of a functional relationship between the probability distribution of hazard-cause factors (as inputs) and a hazard-impact system (as outputs) [8]. Because of the limited data samples, we adopt the information matrix algorithm to build an assessment model in order to objectively assess TC hazard risks over Southern China.

First, we need to construct a primitive information matrix. For this purpose, the HCIS x_j and the HIIS y_j are treated as an input and output sample, respectively, namely, $X = \{(x_1, y_1), (x_2, y_2), \ldots, (x_{17}, y_{17})\}$. Next, we define U and V as a discrete input and output field, respectively, in a monitoring space, where $U = \{u_1, u_2, \ldots, u_{12}\}$ and $V = \{v_1, v_2, \ldots, v_{14}\}$. To ensure the monitoring accuracy, we use small spatial steps, i.e., $\Delta_x = 1.8956$ and $\Delta_y = 0.0965$, in U and V, respectively. Thus, the selected monitoring space of X constitutes its input monitoring space, i.e.,

$$\begin{aligned} U &= \{u_1, u_2, \ldots, u_{12}\} \\ &= \{35.32 \quad 37.22 \quad 39.12 \quad 41.01 \quad 42.91 \quad 44.80 \quad 46.70 \quad 48.60 \quad 50.49 \quad 52.39 \quad 54.28 \quad 56.18\}, \end{aligned}$$

and the selected monitoring space of Y constitutes the output monitoring space of X, i.e.,

$$V = \{v_1, v_2, \ldots, v_{14}\} = \{0.05 \quad 0.14 \quad 0.24 \quad 0.33 \quad 0.43 \quad 0.53 \quad 0.62 \quad 0.72 \quad 0.82 \quad 0.91 \quad 1.01 \quad 1.11 \quad 1.20 \quad 1.30\}$$

With the use of the following two-dimensional (2D) normal spread algorithm

$$\mu((x_j y_j), (u_h, v_k)) = \left[\frac{1}{h_x\sqrt{2\pi}} \exp\left(-\frac{(u_h - x_j)^2}{2h_x^2}\right)\right]\left[\frac{1}{h_y\sqrt{2\pi}} \exp\left(-\frac{(v_k - y_j)^2}{2h_y^2}\right)\right], \tag{11}$$

where subscripts $j = 1, 2, \ldots, 17$; $h = 1, 2, \ldots, 12$; and $k = 1, 2, \ldots, 14$; $h_x = \frac{2.6851(b_x - a_x)}{n-1}$ and $h_y = \frac{2.6851(b_y - a_y)}{n-1}$ are the spread parameters in (x_j, y_j) domain, $n = 17$, $a_x = \min_{1 \leqslant j \leqslant 17}\{x_j\} a_x = \max_{1 \leqslant j \leqslant 17}\{x_j\}$, $b_y = \min_{1 \leqslant j \leqslant 17}\{y_j\}, b_y = \max_{1 \leqslant j \leqslant 17}\{y_j\}$, we can spread the sample information in (x_j, y_j) space into the monitoring space $U \times V$, and then obtain a fuzzy set $\{g_{jhk}\}$ that contains the spread information at each point in the monitoring space.

Summation of the fuzzy set $\{g_{jhk}\}$, i.e.,

$$G_{hk} = \sum_{j=1}^{17} G^{(j)} = \sum_{j=1}^{17} (g_{jhk})_{12\times14}, \tag{12}$$

gives the primitive information matrix of X in the monitoring space $U \times V$, namely,

$$G = (G_{hk})_{12\times14} = 10^{-4} \times \begin{bmatrix} 496 & 697 & 887 & 1161 & 1637 & 2239 & 2643 & 2531 & 1921 & 1146 & 536 & 196 & 56 & 13 \\ 1193 & 1626 & 1912 & 2143 & 2502 & 2964 & 3207 & 2918 & 2143 & 1251 & 576 & 209 & 60 & 13 \\ 2178 & 2930 & 3376 & 3572 & 3719 & 3830 & 3661 & 3029 & 2075 & 1153 & 516 & 186 & 56 & 14 \\ 3180 & 4242 & 4919 & 5229 & 5294 & 5074 & 4413 & 3334 & 2131 & 1147 & 528 & 217 & 87 & 36 \\ 3951 & 5226 & 6130 & 6630 & 6710 & 6260 & 5239 & 3870 & 2537 & 1516 & 862 & 486 & 276 & 153 \\ 4286 & 5699 & 6787 & 7409 & 7437 & 6825 & 5754 & 4559 & 3497 & 2634 & 1937 & 1374 & 925 & 571 \\ 3984 & 5501 & 6788 & 7535 & 7547 & 6986 & 6304 & 5827 & 5486 & 5018 & 4314 & 3450 & 2532 & 1654 \\ 3095 & 4613 & 6078 & 7056 & 7333 & 7238 & 7377 & 7956 & 8571 & 8692 & 8133 & 6994 & 5432 & 3686 \\ 2003 & 3308 & 4755 & 5955 & 6730 & 7412 & 8512 & 10081 & 11587 & 12430 & 12325 & 11191 & 9071 & 6340 \\ 1087 & 2004 & 3179 & 4405 & 5588 & 6906 & 8609 & 10650 & 12637 & 14113 & 14686 & 13956 & 11721 & 8411 \\ 495 & 1018 & 1806 & 2829 & 4053 & 5485 & 7130 & 8936 & 10781 & 12450 & 13525 & 13380 & 11601 & 8533 \\ 188 & 437 & 880 & 1563 & 2477 & 3540 & 4654 & 5803 & 7055 & 8385 & 9473 & 9720 & 8677 & 6529 \end{bmatrix}.$$

After finding the primitive information matrix, we can now construct a fuzzy-relation matrix in order to make fuzzy rough inference. Since factorial relations can be used to synthesize the essence of concepts or identities, and since a fuzzy set can be used to quantify these variables, and even extrapolate them [21], we may consider $U \times V$ as a factorial space displaying a knowledge structure of the information matrix [22,23]. To this end, let

$$s_k = \max_{1\leqslant h\leqslant 12}\{G_{hk}\}(k = 1, 2, \ldots, 14), \quad \text{and} \quad r_{hk} = \frac{G_{hk}}{s_k}, \quad h = 1, 2, \ldots, 12,$$

we obtain the following causal fuzzy relation matrix (with their magnitudes scaled to the range of [0, 1]) from the primitive information matrix G:

$$R_f = \{r_{hk}\}_{12\times 14} = 10^{-4} \times \begin{bmatrix} 1157 & 1224 & 1307 & 1541 & 2169 & 3021 & 3070 & 2376 & 1520 & 812 & 365 & 140 & 48 & 15 \\ 2783 & 2853 & 2816 & 2844 & 3315 & 3999 & 3725 & 2740 & 1696 & 886 & 392 & 149 & 51 & 16 \\ 5080 & 5142 & 4973 & 4741 & 4928 & 5167 & 4252 & 2844 & 1642 & 817 & 351 & 134 & 47 & 17 \\ 7419 & 7444 & 7247 & 6940 & 7015 & 6846 & 5126 & 3130 & 1687 & 813 & 359 & 156 & 74 & 42 \\ 9217 & 9169 & 9030 & 8799 & 8891 & 8445 & 6085 & 3634 & 2007 & 1074 & 587 & 348 & 235 & 179 \\ 10000 & 10000 & 9997 & 9833 & 9853 & 9208 & 6683 & 4281 & 2767 & 1867 & 1319 & 985 & 789 & 669 \\ 9295 & 9653 & 10000 & 10000 & 10000 & 9425 & 7322 & 5472 & 4341 & 3556 & 2937 & 2472 & 2160 & 1938 \\ 7220 & 8094 & 8954 & 9365 & 9716 & 9766 & 8569 & 7471 & 6783 & 6159 & 5538 & 5011 & 4635 & 4320 \\ 4672 & 5804 & 7005 & 7903 & 8917 & 10000 & 9887 & 9466 & 9169 & 8807 & 8392 & 8019 & 7739 & 7430 \\ 2536 & 3516 & 4683 & 5847 & 7404 & 9317 & 10000 & 10000 & 10000 & 10000 & 10000 & 10000 & 10000 & 9856 \\ 1154 & 1787 & 2661 & 3755 & 5370 & 7400 & 8282 & 8391 & 8531 & 8822 & 9210 & 9588 & 9898 & 10000 \\ 439 & 766 & 1296 & 2075 & 3283 & 4776 & 5406 & 5449 & 5583 & 5942 & 6450 & 6965 & 7403 & 7652 \end{bmatrix}.$$

Based on the fuzzy set in the given U field, $\widetilde{A} = \{\mu_{\underset{\sim}{A}}(u_1), \mu_{\underset{\sim}{A}}(u_2), \ldots, \mu_{\underset{\sim}{A}}(u_{12})\}$, and the fuzzy relation matrix R_f in the monitoring space of $U \times V$, we can make the following fuzzy rough inference:

$$\mu_{\underset{\sim}{B}}(v_k) = \sup_{u_h\in U}\{\mu_A(u_h) \wedge r_{hk}\} = \max_{u_h\in U}\{\min\{\mu_{\underset{\sim}{A}}(u_h), r_{hk}\}\}, \quad v_k \in V, \tag{13}$$

and attain the output fuzzy set in V

$$\widetilde{B} = \left\{\mu_{\underset{\sim}{B}}(v_1), \mu_{\underset{\sim}{B}}(v_2), \ldots, \mu_{\underset{\sim}{B}}(v_{14})\right\}. \tag{14}$$

Through the calculation of the gravitational center of the above fuzzy set

$$\tilde{y}_j = \sum_{k=1}^{14} \mu_{\underset{\sim}{B}}(v_k) \cdot v_k \bigg/ \sum_{k=1}^{14} \mu_{\underset{\sim}{B}}(v_k), \tag{15}$$

we obtain the estimated values of the output samples, i.e., the hazard risk index values or HRIS as roughly determined by the hazard-cause indictors of TCs (see the last column in Table 1). We find that the standard deviation of assessment errors so obtained is 0.0628, with a mean error of 0.2583. The correlation coefficient between the HRIS and HIIS given in Table 1 is 0.7461. Because they do not follow the normal distribution when chronologically ordered, they are not appropriate for t-test. However, the HRIS and HIIS have similar distributions, based on the Kolmogorov–Smirnov test, with a 91.87% probability and k = 0.4118.

It is apparent from Table 1 that the highest hazard-risk value of 0.84 corresponds to the years of 1990, 1996 and 2006, whereas the lowest risk value of 0.37 corresponds to the year of 1999. These results are consistent with those obtained in accordance with the HIIS (y_j), especially for 1996 and 2006 in which Herb (1996), Sally (1996), Bilis (2006), and Saomai (2006), hitting Southern China, were listed as 4 of the top 10 costliest (and deadliest) TCs from 1983 to 2006 [24]. However, there is one exception for 1990, particularly when it is compared to 1994. We may attribute this exceptional result to the two different ways to calculate HIIS and HRIS. Specifically, the year of 1990 has a hazard-impact index value (i.e., 0.60) that is smaller than that of 1994 (i.e., 1.17) due to the smaller contributions of all the *hazard-impact indicators* in the combined weights, whereas the hazard-risk index value of 1990 (i.e., 0.84) is larger than that of 1994 (i.e., 0.72) due to the dominant RP and RW *hazard-cause indicators* (see Appendices A and B).

Of course, the above comparison also indicates the likely importance of including the total rainfall amounts of individual TCs instead of the daily mean rainfall amount as one of the hazard-cause indicators. For example, Typhoon Abe (1990) produced a total rainfall amount of 468 mm in two days, whereas Typhoon Fred (1994) left behind 678 mm in four days. One can clearly understand why Fred could produce more hazard problems, based on the RP indicator alone. Moreover, because the HRIS is assessed after spreading the HCIS information and some fuzzy mathematical manipulations, it should not be considered as accurate as the HIIS. Nevertheless, the above issues will be examined in our future studies.

3. An exceeded-probability model for the hazard risk index series

Because of the randomness and irregularity of natural hazards, it is often desirable for decision-makers to know the possibility of reaching or exceeding certain hazard levels, as compared to previous hazard events. Thus, in this section, we construct an exceeded probability model, following Huang [8] and Huang and Shi [25], to estimate the possibility of exceeding various categories of TC hazards over Southern China, based on the results obtained in the preceding section.

Given the hazard-impact index field $Z = \{z_1, z_2, \ldots, z_s\} = \{z_1, z_2, \ldots, z_{14}\} = \{0, 0.1, 0.2, \ldots, 1.3\}$, we can use the following one-dimensional normal spread algorithm [8,23] to spread the HIIS information carried by y_j to every point in Z:

$$f_j(z_k) = \frac{1}{h_y\sqrt{2\pi}} \exp\left[-\frac{(y_j - z_k)^2}{2h_y^2}\right]. \quad (16)$$

Its corresponding member function of the fuzzy subset $\tilde{y}$ is

$$\mu_{y_j}(z_k) = \frac{f_j(z_k)}{\sum_{k=1}^{s} f_j(z_k)} = \frac{f_j(z_k)}{\sum_{k=1}^{14} f_j(z_k)}. \quad (17)$$

Let

$$g(z_k) = \sum_{j=1}^{17} \mu_{y_j}(z_k), \quad (18)$$

and define the frequency of y_j falling onto z_k as the estimated probability, i.e.,

$$\begin{aligned} p(z_k) &= \frac{g(z_k)}{\sum_{k=1}^{s} g(z_k)} = \frac{g(z_k)}{\sum_{k=1}^{14} g(z_k)} \\ &= 10^{-4} \times \{278 \quad 416 \quad 543 \quad 644 \quad 718 \quad 775 \quad 822 \quad 860 \quad 889 \quad 913 \quad 924 \quad 889 \quad 766 \quad 562\}, \end{aligned}$$

we obtain the probability of the assessed risk value that exceeds z_k by taking the accumulative summation of $p(z_k)$, i.e.,

$$\begin{aligned} P(z_k) &= \sum_{l=k}^{14} p(z_k) \\ &= 10^{-4}\{10000 \quad 9722 \quad 9306 \quad 8762 \quad 8118 \quad 7400 \quad 6624 \quad 5802 \quad 4942 \quad 4053 \quad 3141 \quad 2217 \quad 1328 \quad 562\}. \end{aligned}$$

Clearly, the larger the value of the hazard-risk index, the better is the assessed parameter. Therefore, applying the above exceeded probability result to the HIIS given in Table 1 yields the following categorical order:

(i) the catastrophic hazard year ($y_j > 1.2$) of 1996;
(ii) the major hazard years ($0.9 < y_j \leqslant 1.2$) of 1994, 2006, 1993, 1995, and 2001;
(iii) the moderate hazard years ($0.6 < y_j \leqslant 0.9$) of 1991, 1997, 2003, and 1998;
(iv) the light hazard years ($0.3 < y_j \leqslant 0.6$) of 1990, 2000, 2002, 2007, and 2005; and
(v) the diminutive hazard years ($0 < y_j \leqslant 0.3$) of 1999, and 1992.

That is, the probabilities of exceeding moderate, major and catastrophic hazards are 0.6624, 0.4053, and 0.1328, respectively. Obviously, the results conform the reality of the TC-caused hazard levels during the years of 1990 to 2007 in Southern China. They could also be roughly estimated from the original data given in Appendices A and B.

4. An improved grey hazard-year prediction model

The grey estimation system theory is built upon the concepts of correlation space, and discrete but smooth functions. It takes any stochastic process as the temporal and spatial variations of a grey process in a given four-dimensional domain such that any stochastic variable can be treated as a grey variable in this system. It can also consider a randomly varying (in time and space) discrete series as a manifestation of a potentially ranking series so that the former can be transformed into a ranking series using appropriate transformation algorithms [15]. In this regard, predicting the future years of TC hazard risks involves examining the hazard year series and HIIS and then finding their inherent characteristics. This can be achieved by constructing a GM (1, 1) model, based on the previous TC hazard-year series, to estimate the recurrence time intervals of future TC hazard risks, which is referred to as the GHYPM.

From the results obtained in Section 3, the hazard years exceeding the moderate hazard category are: $\{Y^{(0)}(k)\}(k = 1, 2, 3, \ldots, 10) = \{1991, 1993, 1994, 1995, 1996, 1997, 1998, 2001, 2003, 2006\}$. The corresponding HIIS is $10^{-4} \times (8681, 11025, 11692, 10692, 12507, 8349, 6676, 9045, 7536, 11397)$ and the corresponding hazard-year series is $\delta^{(0)} = (2, 4, 5, 6, 7, 8, 9, 12, 14, 16)$.

To ensure the predictive accuracy, we use the fifth order roots to improve the smoothness of the original discrete data and obtain the following summed series:

$$\left\{\delta^{(1)}(k)=\sum_{i=1}^{k}\delta^{(0)}(i)\right\} \quad (k=1,2,\ldots,10)$$
$$=10^{-4}\times(11487 \quad 24682 \quad 38479 \quad 52789 \quad 67547 \quad 82704 \quad 98222 \quad 114660 \quad 131612 \quad 149023). \tag{19}$$

Obviously,

$$\{k\delta^{(0)}(k)/\sum_{i=1}^{k-1}\delta^{(0)}(i)\}, \quad (k=2,3,\ldots,10)=10^{-4}\times(11487 \quad 5590 \quad 3719 \quad 2796 \quad 2244 \quad 1876 \quad 1673 \quad 1478 \quad 1323)$$

is a series in a deceased ranking, so it is indeed a smooth series. Averaging the hazard-year series between neighboring points, i.e., $z^{(1)}(k)=\frac{1}{2}[\delta^{(1)}(k)+\delta^{(1)}(k+1)]$ $k=1,2,\ldots,9$, leads to the neighboring-mean series of $\delta^{(1)}$, i.e.,

$$Z^{(1)}=10^{-4}\times(18085 \quad 31581 \quad 45634 \quad 60168 \quad 75125 \quad 90463 \quad 106441 \quad 123136 \quad 140318).$$

A GHYPM may be derived from the following ordinary differential equation:

$$\frac{d\delta^{(1)}}{dt}+a\delta^{(1)}=b, \tag{20}$$

where the parameters a and b are obtained by applying the least-square algorithm to (20), $\hat{a}=\begin{bmatrix}a\\b\end{bmatrix}=10^{-4}\times\begin{bmatrix}-342\\12653\end{bmatrix}$. The solution to Eq. (20) is then

$$\delta^{1}(t)=q_m\exp(-at)+\frac{b}{a}, \tag{21}$$

where q_m is calculated, following Liu and Zhi [16], from

$$q=\sum_{i=2}^{10}\frac{\exp[-a(i-1)]}{\delta^{(0)}(i)}\Bigg/\sum_{i=2}^{10}\left(\frac{\exp[-a(i-1)]}{\delta^{(0)}(i)}\right)^2=1.2826.$$

By setting $q = q_m[1-\exp(a)]$, we obtain q_m = 38.1040. Therefore, the solution to (20) gives the GM (1, 1) form of the temporal response series for TC hazard-risk years,

$$\hat{\delta}^{(1)}(k+1)=q(1-\exp(a))^{-1}\exp(-ak)+\frac{b}{a}=q_m\exp(-ak)+\frac{b}{a}, \quad k=0,1,\ldots,9. \tag{22}$$

After performing recovery transformations in temporal response through accumulative subtractions between $\hat{\delta}^{(1)}(k)$ in (19) and $\hat{\delta}^{(1)}(k+1)$ in (22), we obtain the following solution to the GHYPM,

$$\hat{\delta}^{(0)}(k+1)=\hat{\delta}^{(1)}(k+1)-\hat{\delta}^{(1)}(k)=q_m(1-\exp(a))\exp(-ak)=\left(\hat{\delta}^{(1)}(1)-\frac{b}{a}\right)(1-\exp(a))\exp(-ak)$$
$$=1.1069(1-\exp(-0.0342))\exp(0.0342k)$$
$$=10^{-4}\times(12826 \quad 13272 \quad 13735 \quad 14213 \quad 14708 \quad 15220 \quad 15751 \quad 16299 \quad 16867 \quad 17455)$$
$$(k=0,1,\ldots,9). \tag{23}$$

Transforming the above predicted results through accumulative subtractions, the simulated $\delta^{(0)}$series is recovered as,

$$\hat{\delta}^{(0)}=10^{-4}\times(34705 \quad 41185 \quad 48876 \quad 58002 \quad 68833 \quad 81685 \quad 96938 \quad 115039 \quad 136520 \quad 162011). \tag{24}$$

We can estimate the error series between the simulated $\delta^{(0)}$ and predicted $\hat{\delta}^{(0)}$ by defining

$$\varepsilon(k)=\{\delta^{(0)}-\hat{\delta}^{(0)}\} \quad (k=1,2,\ldots,10), i.e.,$$

$$\varepsilon=10^{-4}\times(-14705 \quad -1185 \quad 1124 \quad 1998 \quad 1167 \quad -1685 \quad -6938 \quad 4961 \quad 3480 \quad -2011),$$

and the relative error series, $\Delta_k=\left\{\left|\frac{\varepsilon(k)}{\delta^{(0)}}\right|\right\}$, ($k$ = 1, 2, ..., 10), i.e.,

$$\Delta=10^{-4}\times(7353 \quad 296 \quad 225 \quad 333 \quad 167 \quad 211 \quad 771 \quad 413 \quad 249 \quad 126). \tag{25}$$

It is apparent that the mean relative error from the above series (25) is $\bar{\Delta}=0.1014$, so the relative accuracy is $1-\bar{\Delta}=0.8986$ and the simulated accuracy is $1-\Delta_7$ = 0.9874. These results are all better than the respective mean relative accuracy and simulated accuracy of 0.8620 and 0.9028 as obtained by Liu and Zhi [16]. In particular, our improved algorithm is more rigorous than theirs from the theoretical perspective.

With the GHYPM solution, we may predict the hazard-year series after the year of 2006 as $\hat{\delta}(11)=19.2263$, $\hat{\delta}(12)=22.8163$, $(\hat{\delta}(13)=27.0767$, $(\hat{\delta}(14)=32.1326$, $(\hat{\delta}(15)=38.1325$. After performing recovery transformations, these hazard-year series corresponds to the years of 2009, 2012, 2017, 2022, 2028, respectively, namely, after 3, 7, 11, 16, and 22 years from the most recent year of the exceeded probability of moderate TC hazards. Indeed, there were 9 landfalling TCs in China in 2009, most of which took place in the coastal area of Southern China (see http://www.laxf.gov.cn/qbzq/qbswebsite/default.asp). This verification is encouraging and it indicates that this GM (1, 1) model could be useful for predicting the future hazard years, assuming the recurrence of TC hazards, with appropriate transformations and adjustments due to changes in other parameters.

5. Summary and concluding remarks

In this study, a fuzzy mathematical model and a modified GHYPM are constructed to estimate and predict TC hazard-year risks in Southern China, respectively. In constructing these models, a genetic projection pursuit algorithm is developed to determine the categorical and ranking weights of the hazard-impact and hazard-cause indicators, and their combined weights, obtained through optimization, are then employed to construct the HIIS and HCIS after data down-scaling, based on different magnitudes of the hazard-impact indicators. Fig. 1 provides a flow chart of assessing the TC-hazard risks and predicting the associated risk-recurrence years.

To identify the input–output functional relation between the probability distributions of the hazard-cause and hazard-impact indicators, a two-dimensional normal spread algorithm is utilized to construct the primitive information and fuzzy relation matrix, which allows us to make fuzzy rough inference of hazard risks from the HCIS with the factorial space theory (see Fig. 1). For the TC hazards in Southern China, our calculation shows that the highest HRIS value corresponds to the years of 1990, 1996, and 2006 and the lowest HRIS value corresponds to the year of 1999, which are consistent with the results from the HIIS, except for the year of 1990. The standard deviation in estimation error is 0.0628, with a mean error of 0.2583. We have attributed the inconsistent result for 1990 to the use of the hazard-impact and hazard-cause indicators for calculating the HIIS and assessing the HRIS values, respectively.

An exceeded probability model, based on the TC hazard-impact indicators, has been constructed to estimate the probability of exceeding a certain hazard-impact category (Fig. 1). We group the TC hazard impacts during the past 20 years in Southern China into five categories. Results show that the probability of exceeding a moderate, major, and catastrophic hazard year is 0.6624, 0.4053, and 0.1328, respectively.

A dynamic GHYPM, expressed by a time-dependent differential equation, has been constructed through the GM (1, 1) theory using the original discrete hazard-year series, and its solution is represented in terms of a temporal response function. In constructing the GHYPM, the original HIIS is first summed accumulatively to make it in a decreased ranking, and the fifth root transformation is then performed to ensure its smoothness and predictive accuracy. Results show that the dynamical model has a mean relative error, a mean relative accuracy, and a simulation accuracy of 0.1014, 0.8986, and 0.9874, respectively. Because the response function $\hat{\delta}(k+1)$ grows at an exponential rate, i.e., $e^{0.0342k} \to \infty$ as $k \to \infty$, the model has a relatively better performance for the first 5 recurrence time intervals, and its predictive accuracy drops rapidly after the year of 2028. Therefore, this model is more suitable for the hazard-year prediction within the first 20 years from 2007.

It should be mentioned that because of the incompleteness and limited size of the data samples and unavoidable errors in data collection, we have noted that in some cases the hazard-impact and hazard-cause indicators could not be easily syn-

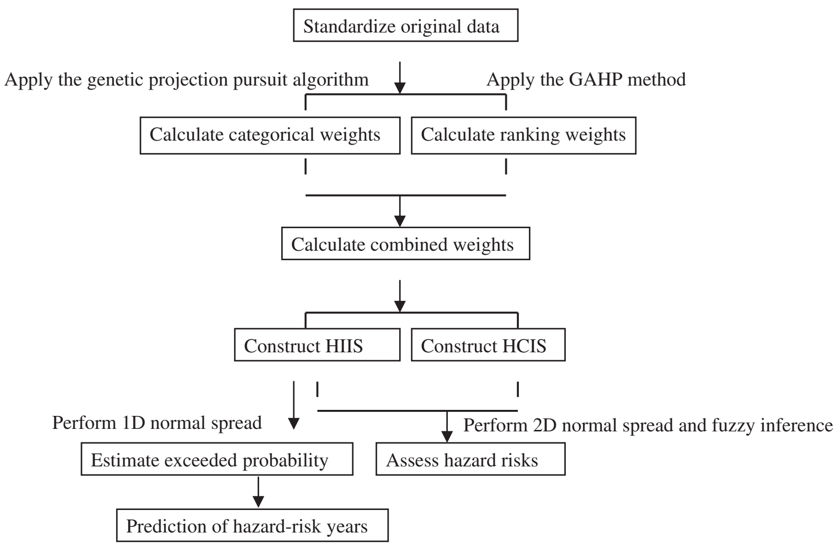

Fig. 1. A flow chart of the fuzzy assessment of TC-hazard risks and the prediction of TC-hazard-risk years.

chronized. In our future work, we will attempt to examine whether or not such a data sample could be considered as a piece of fuzzy information with certain fuzzy relations in order to take full use of all data samples for the analyses and prediction of TC hazard risks. Nevertheless, the above-mentioned problems may impose some limitations on the application of the fuzzy mathematical model and GHYPM constructed in this study to the analysis and prediction of TC hazard risks in the other coastal regions around the world. Thus, more future work is needed to gain insight into the inherent characteristics of TC hazard-impact data, and construct appropriate hazard-cause indicators in order to improve the quality and accuracy of these fuzzy mathematical models.

Acknowledgements

We would like to thank Prof. Zhaohua Wu of the Florida State University for his initial editorial assistance. This work was funded by China's NSFC Grants 41065002 and 11061008, the Key Scientific Research Grant 0993002-4 of Guangxi Province, and the research Grant 200911MS151 of the Department of Education of Guangxi Province. The second author was supported by the NSF Grant ATM0758609.

Appendix A

The impact of TC hazards occurring during the period of 1990–2007 in Southern China in terms of the number of human mortalities (RM), the affected population (RP), the affected agricultural area (RA), damaged houses (RB), and direct economic losses (RL).

Year	RM	RP (10^4 people)	RA (10^4 acres)	RB (10^4)	RL (10^8 RMB)
1990	48	675.75	36.1325	3.4303	7.3085
1991	260	2140.587	79.3374	7.8168	52.8285
1992	4	475.4359	22.9047	6.9037	15.4521
1993	166	4578.1435	277.5331	23.6005	159.5211
1994	294	2926.96	140.9629	44.7883	191.2678
1995	216	3825.8	156.1027	17.151	141.5335
1996	548	2253.46	87.96	40.5974	295.757
1997	109	1869.29	86.4736	22.522	77.74
1998	11	568.95	18.491	0.5355	9.3474
1999	22	804.12	43.641	1.234	29.973
2000	56	1114.08	85.357	11.6114	72.232
2001	72	3371.592	172.0839	16.4745	243.261
2002	69	1463.91	49.1659	2.011	32.4885
2003	80	2972.21	105.4598	8.7408	83.4772
2005	0	101.0595	8.0984	0.1897	4.4737
2006	250	3453.4	154.6494	20.0966	242.3109
2007	17	595.82	25.4351	0.8072	26.9266

Appendix B

The characteristics of TC hazards occurring during the period of 1990–2007 in Southern China in terms of the annual-total landed number (CN), the annual-mean daily maximum rainfall (CR), and the annual-mean daily maximum wind (CW) of TCs.

Year	CN	CR (mm)	CW ($m\ s^{-1}$)
1990	4	107.74	29
1991	8	87.94	38.5
1992	6	86.19	28.17
1993	8	100.62	32.125
1994	7	98.12	27.375
1995	11	98.32	28.1
1996	6	104.26	34
1997	2	104.73	26.5
1998	3	69.43	24.5
1999	8	73.13	29.6

Appendix B (*continued*)

Year	CN	CR (mm)	CW ($m\ s^{-1}$)
2000	4	80.5	32
2001	8	101.73	32.4
2002	7	82.39	27.67
2003	8	92.53	40
2005	3	93.99	28.5
2006	6	108.65	27.3
2007	5	95.44	34.3

References

[1] L.A. Zadeh, Fuzzy sets, Inform. Control 8 (1965) 338–353.
[2] B.-S. Sun, W.-M. Ma, Fuzzy rough set model on two different universes and its application, Appl. Math. Model. 35 (2011) 1798–1809.
[3] Y. Hundecha, A. Bardossy, H.-W. Theisen, Development of a fuzzy logic based rainfall-runoff model, Hydrol. Sci. J. 46 (2001) 363–377.
[4] L.S. Iliadis, S.I. Spartalis, Fundamental fuzzy relation concepts of a D.S.S. for the estimation of natural disasters' risk (the case of a trapezoidal membership function), Math. Comput. Model. 42 (2005) 747–758.
[5] P.-S. Yui, S.-T. Chen, T.-C. Yang, The potential of fuzzy multi-objective model for rainfall forecasting from typhoons, Nat. Hazards 34 (2005) 131–150.
[6] S.-P. Cheng, R.-Y. Wang, Analyzing hazard potential of typhoon damage by applying grey analytic hierarchy process, Nat. Hazards 33 (2004) 77–103.
[7] Y.L. Zhang, M. Liu, Application of grey prediction model GM (1, 1) in fire accident, Acac Scientiarum Naturalium Universitais Nankaiensis (Natural Science Edition) 42 (2009) 11–15.
[8] C.-F. Huang, Information matrix method for risk analysis of natural disaster, J. Natural Dis. 15 (2006) 1–10.
[9] J.-X. Zhang, C.-F. Huang, X.-L. Liu, Geographical distribution and risk assessment of typhoon storm surge disasters in coastal regions of Guangdong Province in 1949–2005, J. Basic Sci. Eng. 16 (2008) 393–401.
[10] L.-H. Feng, W.-H. Hong, On the principle of maximum entropy and the risk analysis of disaster loss, Appl. Math. Model. 33 (2009) 2934–2938.
[11] G.M. Andrade, M.D. Fragoso, A.A.F.M. Carneiro, A stochastic approach to the flood control problem, Appl. Math. Model. 25 (2001) 499–511.
[12] W.-F. Lu, The Shanghai area tropical cyclone disaster situation's appraisal and the famine year forecast, J. Natural Dis. 4 (1995) 40–45.
[13] P.-Y. Chen, Y.-H. Yang, X.-T. Lei, Y.-Z. Qian, Cause analysis and preliminary hazard estimate of typhoon disaster in China, J. Natural Dis. 18 (2009) 64–73.
[14] J.-L. Jin, Y.-M. Wei, J. Ding, System evaluation model based on combined weights, Math. Pract. Theory 33 (2003) 51–58.
[15] S.-F. Liu, Y.-G. Dang, Z.-G. Fang, Grey System Theory and its Application, Science Press, Beijing, 2005. pp. 172–175.
[16] F.-Q. Liu, C.-J. Zhi, Improvement of grey prediction model GM (1, 1), Math. Pract. Theory 35 (2005) 11–15.
[17] K.-G. Wen, China Meteorological Disasters Dictionary, Meteorological Press, Beijing, 2007. pp. 199–250.
[18] China Meteorological Administration (1991–2008). Tropical Cyclone Yearbook (1990–2007).
[19] J.H. Friedman, J.W. Turkey, A projection pursuit algorithm for exploratory data analysis, IEEE Trans. Comput. 23 (1974) 881–890.
[20] T.L. Saaty, The Analytic Hierarchy Process, McGraw-Hill, New York, 1980.
[21] X.-T. Peng, P.-Z. Wang, A. Kandel, Knowledge acquisition by random sets, Int. J. Intell. Syst. 11 (1996) 113–147.
[22] P.-Z. Wang, A factor space approach to knowledge representation, Fuzzy Sets Syst. 36 (1990) 113–124.
[23] L.-X. Wang, J.M. Mendel, Generating fuzzy rules by learning through examples, IEEE Trans. Syst. Man Cybernet. SMC-22 (1992) 1414–1427.
[24] Q. Zhang, L.-G. Wu, Q.-F. Liu, Tropical cyclone damages in China (1983–2006), Bull. Am. Meteor. Soc. 90 (2009) 489–495.
[25] C.-F. Huang, Y. Shi, Towards Efficient Fuzzy Information Processing Using the Principle of Information Diffusion, Physica-Verlag, Springer, Heidelberg, 2002. pp. 45–110.