



# Comprehensive security risk factor identification for small reservoirs with heterogeneous data based on grey relational analysis model

Jing-chun Feng<sup>a,b,c</sup>, Hua-ai Huang<sup>a,d</sup>, Yao Yin<sup>a,b</sup>, Ke Zhang<sup>a,b,\*</sup>

<sup>a</sup> Business School, Hohai University, Nanjing 211100, China

<sup>b</sup> Institute of Project Management, Hohai University, Nanjing 211100, China

<sup>c</sup> Jiangsu Provincial Collaborative Innovation Center of World Water Valley and Water Ecological Civilization, Nanjing 211100, China

<sup>d</sup> Guangxi Flood Control and Drought Relief Headquarters, Nanning 530023, China

Received 7 June 2019; accepted 17 October 2019

Available online 19 December 2019

## Abstract

Identification of security risk factors for small reservoirs is the basis for implementation of early warning systems. The manner of identification of the factors for small reservoirs is of practical significance when data are incomplete. The existing grey relational models have some disadvantages in measuring the correlation between categorical data sequences. To this end, this paper introduces a new grey relational model to analyze heterogeneous data. In this study, a set of security risk factors for small reservoirs was first constructed based on theoretical analysis, and heterogeneous data of these factors were recorded as sequences. The sequences were regarded as random variables, and the information entropy and conditional entropy between sequences were measured to analyze the relational degree between risk factors. Then, a new grey relational analysis model for heterogeneous data was constructed, and a comprehensive security risk factor identification method was developed. A case study of small reservoirs in Guangxi Zhuang Autonomous Region in China shows that the model constructed in this study is applicable to security risk factor identification for small reservoirs with heterogeneous and sparse data.

© 2019 Hohai University. Production and hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

**Keywords:** Security risk factor identification; Heterogeneous data; Grey relational analysis model; Relational degree; Information entropy; Conditional entropy; Small reservoir; Guangxi

## 1. Introduction

Small reservoirs are indispensable components of China's water conservancy projects and flood control systems. The latest Chinese statistics from 2011 indicated the existence of 93308 small reservoirs out of a total of 98002 reservoirs (MWRPRC, 2013). Reservoirs are not supposed to be disaster-bearing bodies, but they can be prone to disasters. The regional

governments and their dam authorities take an administrative leadership responsibility for reservoirs with a capacity of more than 1 million m<sup>3</sup> or a dam height of more than 15 m. Although small reservoirs are defined legally in China, management specifications are lacking. Small reservoirs in remote areas are usually managed by township water conservation stations or villagers. Due to the lack of management personnel, funds, and monitoring, it is difficult to obtain comprehensive and accurate reservoir operation data.

Most of China's small reservoirs were built from the 1950s to the 1970s. Due to the constraints of economic conditions, technical expertise, and management systems at that time, most small reservoirs have problems, such as poor construction quality, low construction standards, and various engineering defects (MWRPRC, 2010), which lead to reservoir

This work was supported by the National Nature Science Foundation of China (Grant No. 71401052), the National Social Science Foundation of China (Grant No. 17BGL156), and the Key Project of the National Social Science Foundation of China (Grant No. 14AZD024).

\* Corresponding author.

E-mail address: [zhangke@nuaa.edu.cn](mailto:zhangke@nuaa.edu.cn) (Ke Zhang).

Peer review under responsibility of Hohai University.

operation risks (Duckstein and Plate, 1987). In some extreme cases, sudden flooding can cause excessive water level rise in the reservoir area, which is liable to result in dam leakage, excessive deformation, and even dam collapse. These factors transform reservoirs into disaster-causing bodies (Georgakakos, 2006), and cause incalculable losses to inhabitants at lower reaches of reservoirs. Therefore, identifying the impact factors of security risks for small reservoirs is of a practical importance.

There is a complex and non-linear relationship between the security risk of small reservoirs and their design specifications, construction quality, operation and maintenance, and natural environment. Due to the lack of samples and data, most of the previous studies on the security risk of small reservoirs have adopted expert survey methods. Guided by years of basic reservoir management experience, Fu et al. (2011) discussed the risk management countermeasures for small reservoirs from three aspects, i.e., water resources management, water quality management, and dam failure risk management, and explored some security management measures to reduce the risk for small reservoirs. Yang (2014) focused on security risks for small reservoirs in the operation process from the perspective of the uncertainty of risks. In view of the lack of both basic data and funds for investigation of small reservoirs in China, Sheng et al. (2008) proposed a risk analysis method suitable for small reservoirs by means of expert experience. These studies identify security risk factors for small reservoirs based on expert experience. It is a challenge to find unique and potential models for security problems and related factors for small reservoirs due to their inherent uncertainty characteristics, including data deficiency and data inaccuracy.

The three most recognized theories for investigation of uncertain systems are probability and statistics, fuzzy mathematics, and the grey system theory. The theory of probability and statistics views the stochastic uncertainty with an emphasis on identifying probability distributions to explain site-specific data using statistical tools (Runnenburg, 1985). Fuzzy logic emerged in the context of the fuzzy set theory, introduced by Zadeh (1965). A fuzzy set assigns a degree of membership, typically a real number in the interval  $[0,1]$ , to elements of a universe. Fuzzy logic arises by assigning truth degrees from a standard set of  $[0,1]$  to propositions, where 0 represents “totally false”, 1 represents “totally true”, and other numbers refer to “partially true” (Mendel, 2014; Liu et al., 2018). The grey system theory focuses on uncertainty problems with small samples and inadequate information that cannot easily be treated with probability (Liu and Forrest, 2013). Therefore, the grey system theory is more suitable for analyzing small reservoir security risks with incomplete data.

The grey relational analysis is an important branch of the grey system theory, and it is usually applied to identifying the factors of a system with deficient data. Since Deng (1989) introduced the grey relational axiom, scholars have introduced several grey relational analysis models (Hipel, 2011; Luo et al., 2015), such as the grey dynamic trend incidence model (Wang et al., 2017), grey comprehensive relational model (Wang et al., 2018), trapezoid grey relational degree

model (Yan et al., 2016), multivariable grey relational analysis model (Zhang et al., 2014; Wang et al., 2019), and so forth. These models focus on relational analysis of numerical data, and there has been little research on the relational analysis of categorical data, or the relational analysis of both categorical data and numerical data. However, in real applications, it is necessary to measure the relational degree between categorical data, as well as that between categorical data and numerical data. Therefore, this paper introduces a new grey relational analysis model for heterogeneous source data, and discusses its applicability to identifying the security risk factors for small reservoirs.

In this study, a new grey relational analysis model was established to objectively identify comprehensive security risks for small reservoirs with heterogeneous and deficient data. Then, it was used to cluster theoretical factors of comprehensive security risks for small reservoirs, and identify key factors according to the results of cluster analysis. The model developed in this study was validated through a case study of a small reservoir database of Guangxi Zhuang Autonomous Region.

## 2. Construction of security risk factor set for small reservoirs

### 2.1. Theoretical analysis of dam security risks

The comprehensive security risk of small reservoirs is the possibility of their failure and the losses to downstream regions because of design, construction, operation, and maintenance defects. As opposed to the resilience of the project, security risk analysis pays more attention to the economic and social hazards to downstream residents. The construction of a security risk factor set for small reservoirs is the basis for identification of their security risks. With the comprehensive function, a reservoir dam is a complex system involving climate, geography, the ecological environment, socio-economics, and engineering technology. There are many factors affecting the security, closely related to many disciplines such as mechanics and geology, involving the coupling of various media, including water bodies and rocks (Li et al., 2016; Cai et al., 2018). Many practitioners have evaluated the security risk of reservoir dams from different perspectives. Saedi et al. (2014) created an HIRARC model for the evaluation of environmental safety and health of a hydroelectric power generation plant. Li et al. (2010) established a multi-level fuzzy comprehensive risk assessment model for dam security from seven aspects: engineering quality, dam operation management, flood control standard, structural security, seepage security, metal structure, and seismic performance. Based on the entropy weight method and the normal cloud model, Feng (2015) developed a dam security risk evaluation system from three main aspects: the environment, seepage, and dam deformation. Zhang et al. (2017) quantitatively evaluated the security risk of actual projects in torrential flood-prone areas based on nine major categories, including meteorological conditions, geological conditions, flood control

security, seepage security, structural security, seismic security, metal structure security, engineering quality, and operational management.

Existing studies generally describe security risks from two aspects: internal factors and external factors of reservoir dams. Internal factors are mainly related to the construction technology of reservoir dams, such as the engineering quality, seepage flow, and seismic performance, while external factors are mainly related to the meteorological environment, geological environment, and operational management. There are many factors related to both internal factors and external factors. When internal and external adverse factors work together, they will trigger security risks, leading to catastrophic reservoir dam accidents. Based on the research mentioned above, this study selected relevant factors from three aspects, i.e., the environment, technology, and management, to construct a reservoir dam security risk indicator system.

## 2.2. Security risk factors of reservoir dams

### 2.2.1. Environmental security risks

Environmental security risks refer to the security risk triggered by the reservoir dam when it is stressed by natural factors, such as geography and climate. Reservoir dams depend heavily on the geographical environment and geological conditions. Therefore, this study selected five indicators to evaluate the environmental security risk for reservoir dams: the geological hazard index, engineering geological conditions in hub areas, precipitation distribution, impact of human activities, and vegetation status. The corresponding factors include the basic seismic intensity, geological conditions of the dam base, average annual precipitation, affected population, and major landslide bodies in the reservoir area. These factors of some reservoirs are measured and recorded in the command system for the flood control and drought relief system, engineering management system of the regional water administration department, and information systems of hydrological and meteorological departments. By integrating these data sources, the measurement of the factors can be obtained.

### 2.2.2. Technical security risks

Technical security risks refer to the security risk caused by multiple factors, mainly resulting from design technology, construction technology, or operation technology. A reservoir dam has a long construction period, a large investment scale, and a wide range of specialties. Therefore, the technical security risk depends on technology to a large degree. During the development and construction of a reservoir dam, the technical security risk should be minimized. Technical security risks of a reservoir dam exist throughout a life cycle that can be divided into three stages: engineering design planning, construction, and operational management. The security risk triggered in any stage may potentially threaten the next stage, resulting in an accumulation of the security risk of the entire project.

This study selected the dam type, form of anti-seepage body, anti-seepage measures of the dam base, and seismic

precautionary intensity as the impact factors in the stage of engineering design planning; the year of completion as the impact factor in the stage of construction; and the construction status of automatic hydrological monitoring systems, hydrological monitoring mode, engineering condition monitoring mode, construction status of automatic engineering condition monitoring system, and maximum seepage flow as the impact factors in the stage of operational management of the project.

### 2.2.3. Management security risks

Reservoir dam management refers to using legal, administrative, technical, economic, and other means to scientifically and reasonably organize the construction and operation of reservoir dams, to ensure the security of reservoir dams, to promote benefits, and to meet the needs of social and economic development for the comprehensive benefits of reservoir dams. There are a large number of small reservoir dams in China, a considerable number of which were built in the 1950s and 1960s. Due to the low technical level in those years, they have more engineering quality problems, and the long-term negligence of management has further increased the security risk of those projects. Especially in extreme cases, the problem of engineering quality is more prominent, and the security risk of projects is greatly stimulated, resulting in an unprecedented safety threat to projects. The role of human beings always runs through the management and production processes of the entire water conservancy system, and to a large extent, human beings play a dominant and controlling role. Therefore, inappropriate human behaviors can become an important source of vulnerability to the security risk for reservoir dams. According to relevant data, due to negligence and misconduct of human activities (such as improper construction methods, inadequate security measures, and management omissions), dam accidents occur often. Therefore, assessing the staff of a reservoir dam is critical to the effective improvement of the security and stability of dams and reduction of security risks. This paper describes the management security risk of small reservoirs from three aspects: the reservoir security management mechanism, business capability, and competence. Their corresponding factors are the registration status, the number of mid- and above-level engineers in the management unit, and the number of senior and above-level engineers in the management unit, respectively. The security risk indicator system and quantitative factors are shown in Table 1.

## 3. Clustering method based on grey relational analysis model for heterogeneous data

### 3.1. Construction of grey relational analysis model for heterogeneous data

Most data fall into one of two groups: numerical or categorical. Numerical data such as the dam height and average annual precipitation are obtained from measurement. Mathematical operations can be applied to them. Categorical data represent characteristics of objects, such as the hydrological

Table 1  
Security risk indicator system and quantitative factors.

Primary indicator	Secondary indicator	Factor	Variable	Best practice recommendation
Environmental	Geological hazard index	Basic seismic intensity	$X_1$	8
	Engineering geology	Dam base geological conditions	$X_2$	Fine
	Precipitation distribution	Average annual precipitation	$X_3$	1000–1500 mm
	Impact of human activities	Affected population	$X_4$	<1000
	Vegetation status	Major landslide bodies in the reservoir area	$X_5$	None
Technical	Design technology	Dam type	$X_6$	Concrete dam
		Form of anti-seepage body	$X_7$	Clay anti-seepage body or concrete impervious wall
		Dam base anti-seepage measures	$X_8$	Curtain grouting
		Seismic precautionary intensity	$X_9$	>6
	Construction technology	Year of completion	$X_{10}$	The 1960s to present
	Operation technology	Construction status of automatic hydrological monitoring system	$X_{11}$	Complete
		Hydrological monitoring mode	$X_{12}$	Auto
		Engineering condition monitoring mode	$X_{13}$	Manual monitoring
		Construction status of automatic engineering condition monitoring system	$X_{14}$	Complete
		Maximum seepage flow	$X_{15}$	5 m <sup>3</sup> /h
		Management	Security management mechanism	Registration status
	Business capability	Number of mid- and above-level engineers in the management unit	$X_{17}$	1–5
	Competence	Number of senior and above-level engineers in the management unit	$X_{18}$	1–2

monitoring mode and dam base anti-seepage measures of a reservoir. Categorical data can take numerical values, but those values do not have mathematical meaning. Categorical data also include qualitative data or yes/no data.

Existing studies on grey relational analysis models focus on the relational degree between numerical data, and they cannot deal with the relational degree between categorical data, as well as the relational degree between categorical data and numerical data.

Because the data of risk factors are heterogeneous and sparse, the Euclidian distance and cosine similarity are not suitable for measuring the similarity among heterogeneous factors. Therefore, this paper introduces a new grey relational analysis model based on the information entropy of heterogeneous source data. The concept of information entropy was introduced by Shannon (1951). Information is a relatively complex and abstract concept, which makes it difficult to quantify and measure. Shannon transferred the concept of entropy into physics to describe the uncertainty of information.

In information theory, information entropy and conditional entropy are two measures for the uncertainty of the information content of a message. They can be exploited to identify the risk indicators of small reservoirs.

(1) Information entropy: Suppose that the original sequence  $X_0 = \{x_{01}, x_{02}, \dots, x_{0n}\}$  is the categorical data, and the comparison sequence  $Y_0 = \{y_{01}, y_{02}, \dots, y_{0n}\}$  is the numerical data. In order to measure the relational degree between  $X_0$  and  $Y_0$ , the numerical data sequence  $Y_0$  is first discretized with the

equal-width method or the equal-frequency method, and converted into a categorical data sequence. For sequences  $X_0$  and  $Y_0$ , information entropies are calculated separately with the following formula:

$$H(X) = - \sum_{i=1}^n p(x_i) \ln p(x_i) \tag{1}$$

where  $H(X)$  is the information entropy of a random sequence  $X$ ; and  $p(x_i)$  is the probability of each status, with  $p(x_i) \geq 0$ , and  $\sum_{i=1}^n p(x_i) = 1$ .  $p(x_i)$  denotes the ratio of the number of samples of category  $i$  to all samples. A greater entropy  $H(X)$  means a greater number of variants of the random variable  $X$  and a greater amount of information carried by  $X$ .

(2) Conditional entropy: Conditional entropy refers to the probability that a particular information condition appears under certain conditions (Malings and Pozzi, 2016). Assuming that there are two random sequences  $X$  and  $Y$ , and the numbers of possible values  $x_i$  and  $y_j$  of the random sequences  $X$  and  $Y$  are  $n$  and  $m$ , respectively, the conditional entropy can be defined as follows:

$$H(X|Y) = - \sum_{i=1}^n \sum_{j=1}^m [p(y_j) p(x_i|y_j) \ln p(x_i|y_j)] \tag{2}$$

where  $p(x_i|y_j)$  denotes the probability of a particular value  $x_i$  at a certain value  $y_j$ , and  $H(X|Y)$  is the conditional entropy, indicating the uncertainty of the random sequence  $X$  at the known sequence  $Y$ .

(3) Relational degree:  $V(X_1|X_2)$  denotes the relational degree between the pair of sequences  $X_1$  and  $X_2$ , and is calculated as follows:

$$V(X_1, X_2) = 1 - \frac{HI(X_1, X_2)}{\max HI(X_I, X_J)} \quad (3)$$

$$HI(X_I, X_J) = \alpha |H(X_I) - H(X_J)| + (1 - \alpha) \frac{H(X_I|X_J) + H(X_J|X_I)}{2} \quad (4)$$

$|H(X_I) - H(X_J)|$  measures the absolute proximity of the information entropy of  $X_I$  and  $X_J$ . With a smaller value of  $|H(X_I) - H(X_J)|$ , the absolute proximity of their information entropies becomes higher, and their absolute distributions are closer.  $H(X_I|X_J)$  measures the relative proximity of the information entropy of  $X_I$  and  $X_J$ . With a smaller value of  $H(X_I|X_J)$ , the uncertainty of  $X_I$  under the known condition of  $X_J$  is lower, and the relative distributions of  $X_I$  and  $X_J$  are closer. In Eq. (3),  $HI(X_I, X_J)$  is the comprehensive proximity measure of the information entropy between  $X_I$  and  $X_J$  from both absolute and relative perspectives. Hence,  $V(X_1, X_2)$  can be used to measure the similarity of  $X_1$  and  $X_2$ , with a range of 0–1. A greater value of  $V(X_1, X_2)$  indicates a greater similarity of the two sequences. In Eq. (4),  $\alpha \in [0, 1]$ . In general,  $\alpha$  is set to 0.5, and it will be greater than 0.5 when more emphasis is put on the absolute proximity.

### 3.2. Clustering procedure of grey relational analysis model based on heterogeneous data

Assume that there are  $n$  observational objects, with  $M$  risk factors for each object, constituting  $M$  sequences, denoted as  $X_1, X_2, \dots, X_M$ . According to grey relational clustering, the grey relational degree  $V(X_I, X_J)$  between any two sequences  $X_I$  and  $X_J$  can be calculated, and a grey relational degree matrix  $A$  can be obtained.

Assume that the  $K$ th security risk factor for small reservoirs can be expressed as  $X_K = \{x_{K1}, x_{K2}, \dots, x_{Kn}\}$ , where  $x_{Ki}$  ( $i = 1, 2, \dots, n$ ) denotes the observed value of the  $K$ th risk factor for the  $i$ th small reservoir. For small reservoirs, the security risk factors can be divided into discrete factors and continuous factors. For the former, such as the basic seismic intensity and the dam base geological conditions, the information entropy can be directly obtained, while for the latter, such as the average annual precipitation and the year of completion, they need to be discretized to obtain information entropy. Then, according to the grey relational analysis model for heterogeneous data, the relational degrees between factors are obtained, and the relational degree matrix is generated. In the application process, the relational degree threshold  $\gamma$  of the cluster analysis is set according to the actual situation. When the relational degree between factors  $X_I$  and  $X_J$  satisfies  $V(X_I, X_J) \geq \gamma$ ,  $X_I$  and  $X_J$  are regarded as similar indicators, and the clustering result can be obtained by means of the grey relational degree matrix  $A$ .

In summary, grey relational cluster analysis is conducted of security risk factors for small reservoirs with heterogeneous data. The steps are as follows:

Step 1: Collecting data of risk factors listed in Table 1 from the command system for the flood control and drought relief system, engineering management system, and information systems of hydrological and meteorological departments.

Step 2: Extracting security risk factor data from raw data, transforming the extracted data, and constructing the sample data set for risk factor identification.

Step 3: Calculating the relational degree between any two factors according to Eq. (3), and constructing the relational degree matrix.

Step 4: Setting the relational degree threshold  $\gamma$ , and conducting cluster analysis of the factors according to the relational degree matrix.

Step 5: Choosing the factors with sufficient data as typical risk factors in each category.

## 4. Case study

### 4.1. Background

#### 4.1.1. Database tables

The samples selected in this study were from the Office of the Guangxi Flood Control and Drought Relief Headquarters. The relevant data were extracted from the reservoir database. The reservoir database contains data from super-large reservoirs (with a storage capacity greater than  $10^{10} \text{ m}^3$ ), large reservoirs (with a storage capacity between  $10^9 \text{ m}^3$  and  $10^{10} \text{ m}^3$ ), medium-sized reservoirs (with a storage capacity between  $10^8 \text{ m}^3$  and  $10^9 \text{ m}^3$ ), small reservoirs (with a storage capacity between  $10^7 \text{ m}^3$  and  $10^8 \text{ m}^3$ ), and mini-reservoirs (with a storage capacity between  $10^6 \text{ m}^3$  and  $10^7 \text{ m}^3$ ). The model selected the data of the small reservoirs and mini-reservoirs, and then extracted non-empty data of each factor.

Database tables included the basic information table of reservoirs, the table of reservoir hydrological characteristics, the dam table, the table of the downstream impact, the table of the reservoir management system, and the table of the reservoir operation management. All kinds of tables were formulated according to the *Structures and Identifiers of Database for Construction and Management of Water Projects* (SL 700–2015) issued by the Ministry of Water Resources of the People's Republic of China (MWRPRC). These tables can be described as follows:

(1) The basic information table of reservoirs describes the basic information, such as the reservoir code and reservoir name. It has 4467 records in total.

(2) The table of reservoir hydrological characteristics has 29 indicators, comprising the reservoir code, control basin area, and river length, with 4466 records in total.

(3) The dam table describes dam information, having 15 indicators, with 4466 records in total. It includes the dam base geological conditions, dam type, form of anti-seepage body, and dam base anti-seepage measures.

(4) The table of the downstream impact describes the area, population, and towns that may be affected by a dam break, having 10 indicators, with 4421 records in total.

(5) The table of the reservoir management system has 18 indicators, with 4428 records in total. It includes the numbers of mid- and above-level engineers as well as senior and above-level engineers in the management unit.

(6) The table of reservoir operation management describes operation management information, having 29 indicators, with 1010 records in total. It includes major landslide bodies in the reservoir area, the construction status of the automatic hydrological monitoring system, the hydrological monitoring mode, the engineering condition monitoring mode, the construction status of the automatic engineering condition monitoring system, and the maximum seepage flow.

4.1.2. Data acquisition and cleanup

Data from small reservoirs were scarce. Based on the reservoir code of the basic information table of reservoirs, a multi-table joint query of data of all the tables was conducted, and a total of 941 records were obtained.

After extracting and cleaning up the data, relatively complete data of 10 samples were obtained (Table 2), involving seven small reservoirs (including the Shimai, Jiaoe, Maqiao-Raojiang, Yangda, Hebao, Nanchang, and Damiao reservoirs) and three mini-reservoirs (including the Chitou, Changjiangkou, and Fenghuang reservoirs). Dam base geological conditions and the maximum seepage flow of these reservoirs had some empty values, and those values were supplemented using the mean imputation method. For dam base geological conditions ( $X_2$ ), empty values were modified to be sandy loam. For the maximum seepage flow ( $X_{15}$ ), empty values were modified to be 0. Considering that nine values out

of the 10 values of the maximum seepage flow were 0, which was inconsistent with the actual situation, the factor was discarded in the case study. Therefore, 17 factors ( $X_1$  through  $X_{14}$  and  $X_{16}$  through  $X_{18}$ ) in Table 1 were selected for cluster analysis.

4.2. Cluster analysis of factors

The relational degree matrix of these 17 factors after calculation is shown in Table 3. According to grey relational clustering, the hierarchical diagram can be obtained, with the distinctive threshold  $\gamma$  in the region of 0–1, as shown in Fig. 1.

The relational degree threshold  $\gamma$  of the cluster analysis was set at 0.7 in this study. According to the relational degree matrix in Table 3, security risk factors were divided into four categories with distinctive background colors. Category 1 is  $\{X_1, X_2, X_5, X_6, X_8, X_9, X_{16}\}$ , Category 2 is  $\{X_3, X_{10}, X_{13}, X_{14}, X_{17}, X_{18}\}$ , Category 3 is  $\{X_7, X_{11}, X_{12}\}$ , and Category 4 is  $\{X_4\}$ .

4.3. Result analysis

Based on the theoretical analysis of security risk factors for reservoir dams described in this paper, the corresponding factors were divided into the environmental, technical, and management factors. After cluster analysis, these factors were divided into four categories.

(1) Category 1 includes the security risk factors closely related to the geological conditions of the reservoir dam. In hydraulic engineering construction, geological work is basic and vital work, running throughout the construction process. Relevant studies show that major landslide bodies in the

Table 2 Security risk factor data of 10 reservoirs.

Reservoir	$X_1$	$X_2$	$X_3$ (mm)	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$
Chitou	<6		1930	1000	None	HED	CFS	Curtain grouting	Unfortified
Changjiangkou	<6		1930	3000	None	HED	CDW	Curtain grouting	Unfortified
Shimai	<6	Sandy loam	1930	15000	None	HED	CFS	Curtain grouting	Unfortified
Jiaoe	<6	Sandy loam	1930	8000	None	HED	CFS	Curtain grouting	Unfortified
Maqiao-Raojiang	<6	Sandy loam	1930	8000	None	HED	CFS	Curtain grouting	Unfortified
Yangda	<6		1086	1000	None	HED	CFS	Curtain grouting	Unfortified
Fenghuang	<6		2100	1000	None	HED	CFS	Curtain grouting	Unfortified
Hebao	<6	Sandy loam	1933	15000	None	HED	CFS	Curtain grouting	Unfortified
Nanchang	<6		2100	45000	None	HED	CFS	Curtain grouting	Unfortified
Damiao	<6		1933	20000	None	HED	CASB	Curtain grouting	Unfortified

Reservoir	$X_{10}$	$X_{11}$	$X_{12}$	$X_{13}$	$X_{14}$	$X_{15}$ (L/s)	$X_{16}$	$X_{17}$	$X_{18}$
Chitou	1953	UC	Auto-manual	Manual	No monitoring	0	Registered	24	6
Changjiangkou	1953	WC	Manual	Manual	No monitoring	0	Registered	24	6
Shimai	1958	UC	Auto-manual	Manual	No monitoring	0	Registered	24	6
Jiaoe	1958	UC	Auto-manual	Manual	No monitoring	0	Registered	24	6
Maqiao-Raojiang	1958	UC	Auto-manual	Manual	No monitoring	0	Registered	24	6
Yangda	2011	UC	Auto-manual	Automatic	Automatic	0.8	Registered	1	0
Fenghuang	1956	UC	Auto-manual	Manual	No monitoring		Registered	24	6
Hebao	1968	UC	Auto-manual	Manual	No monitoring		Registered	24	6
Nanchang	1956	UC	Auto-manual	Manual	No monitoring		Registered	24	6
Damiao	1958	UC	Auto-manual	Manual	No monitoring		Registered	24	6

Note: HED is the homogeneous earth dam, CFS is the concrete face slab, CDW is the concrete diaphragm wall, UC means under construction, WC means without construction, and CASB is the clay anti-seepage body.

Table 3  
Relational degree matrix of factors.

Factor	Relational degree																	
	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>	X <sub>8</sub>	X <sub>9</sub>	X <sub>10</sub>	X <sub>11</sub>	X <sub>12</sub>	X <sub>13</sub>	X <sub>14</sub>	X <sub>16</sub>	X <sub>17</sub>	X <sub>18</sub>	
X <sub>1</sub>	1.00	1.00	0.64	0.00	1.00	1.00	0.29	1.00	1.00	0.64	0.29	0.11	0.64	0.64	1.00	0.64	0.64	
X <sub>2</sub>	1.00	1.00	0.64	0.00	1.00	1.00	0.29	1.00	1.00	0.64	0.29	0.11	0.64	0.64	1.00	0.64	0.64	
X <sub>3</sub>	0.64	0.64	1.00	0.16	0.64	0.64	0.43	0.64	0.64	1.00	0.48	0.47	1.00	1.00	0.64	1.00	1.00	
X <sub>4</sub>	0.00	0.00	0.16	1.00	0.00	0.00	0.37	0.00	0.00	0.16	0.33	0.42	0.16	0.16	0.00	0.16	0.16	
X <sub>5</sub>	1.00	1.00	0.64	0.00	1.00	1.00	0.29	1.00	1.00	0.64	0.29	0.11	0.64	0.64	1.00	0.64	0.64	
X <sub>6</sub>	1.00	1.00	0.64	0.00	1.00	1.00	0.29	1.00	1.00	0.64	0.29	0.11	0.64	0.64	1.00	0.64	0.64	
X <sub>7</sub>	0.29	0.29	0.43	0.37	0.29	0.29	1.00	0.29	0.29	0.43	0.77	0.75	0.43	0.43	0.29	0.43	0.43	
X <sub>8</sub>	1.00	1.00	0.64	0.00	1.00	1.00	0.29	1.00	1.00	0.64	0.29	0.11	0.64	0.64	1.00	0.64	0.64	
X <sub>9</sub>	1.00	1.00	0.64	0.00	1.00	1.00	0.29	1.00	1.00	0.64	0.29	0.11	0.64	0.64	1.00	0.64	0.64	
X <sub>10</sub>	0.64	0.64	1.00	0.16	0.64	0.64	0.43	0.64	0.64	1.00	0.48	0.47	1.00	1.00	0.64	1.00	1.00	
X <sub>11</sub>	0.29	0.29	0.48	0.33	0.29	0.29	0.77	0.29	0.29	0.48	1.00	0.72	0.48	0.48	0.29	0.48	0.48	
X <sub>12</sub>	0.11	0.11	0.47	0.42	0.11	0.11	0.75	0.11	0.11	0.47	0.72	1.00	0.47	0.47	0.11	0.47	0.47	
X <sub>13</sub>	0.64	0.64	1.00	0.16	0.64	0.64	0.43	0.64	0.64	1.00	0.48	0.47	1.00	1.00	0.64	1.00	1.00	
X <sub>14</sub>	0.64	0.64	1.00	0.16	0.64	0.64	0.43	0.64	0.64	1.00	0.48	0.47	1.00	1.00	0.64	1.00	1.00	
X <sub>16</sub>	1.00	1.00	0.64	0.00	1.00	1.00	0.29	1.00	1.00	0.64	0.29	0.11	0.64	0.64	1.00	0.64	0.64	
X <sub>17</sub>	0.64	0.64	1.00	0.16	0.64	0.64	0.43	0.64	0.64	1.00	0.48	0.47	1.00	1.00	0.64	1.00	1.00	
X <sub>18</sub>	0.64	0.64	1.00	0.16	0.64	0.64	0.43	0.64	0.64	1.00	0.48	0.47	1.00	1.00	0.64	1.00	1.00	

reservoir area (reservoir topographic conditions), the dam base geological conditions, the basic seismic intensity, and the seismic precautionary intensity are important factors for the determination of the dam type (Gu et al., 2014), while dam base anti-seepage measures are generally determined according to the dam type.

(2) Category 2 includes the security risk factors related to construction and operation management of the reservoir dam. First, the automatic engineering condition monitoring system is an important part of reservoir information construction. There is no doubt that the engineering condition monitoring mode is closely related to the construction status of the system. There are generally four engineering condition monitoring modes: automatic monitoring, manual monitoring, the combination of automatic and manual monitoring, and no monitoring. For completed systems, the mode of automatic monitoring or the combination of automatic and manual

monitoring is always adopted. For the systems under construction, being introduced, or temporarily unplanned, the mode of manual monitoring or no monitoring is generally adopted. Second, most of China's reservoir projects are located in mountainous areas, where working and living conditions are relatively tough, and welfare benefits are poor. Technical experts that have been cultivated and exercised through long-term training often have high turnover, leading to various long-running problems in some reservoir management units, such as unreasonable personnel structure, low technical quality, and weak management and responsibility. Third, two factors, namely, the number of mid- and above-level engineers and the number of senior and above-level engineers in the management unit, are used to describe the personnel security risk of small reservoirs. Large numbers indicate a high quality of management personnel, a high level of technical competence, a high level of security management awareness, and a relatively high construction level of the automatic engineering condition monitoring system. Most of China's small reservoirs were built from the 1950s to the 1970s, and due to the constraints of economic conditions, the technical level, and the management systems at that time, most small reservoirs have problems, such as the poor construction quality, low construction standards, and many underlying dangers in engineering. The factor of the year of completion largely reflects the construction quality of the reservoir dam.

(3) Category 3 includes the security risk factors related to hydrological monitoring of the reservoir dam. The hydrological monitoring system is the basis of reservoir operation management and flood control dispatching. The system collects and processes real-time hydrological data such as rainfall and water level in the monitoring area through data acquisition, transmission, storage, and processing (Ma et al., 2009). In recent years, with the frequent occurrence of various natural

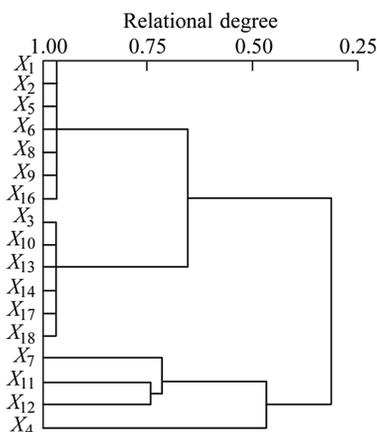


Fig. 1. Hierarchical diagram of factors.

disasters and the increasing investment in water conservancy projects in China, some reservoirs have upgraded their hydrological monitoring systems from the security perspective. However, the manual monitoring mode is generally used for hydrological monitoring. For the systems under construction, being introduced, or temporarily unplanned, manual monitoring mode or no monitoring is generally adopted.

(4) Category 4 only includes the impact factor of the affected population. Since the beginning of the 21st century, more attention has been paid to human life, and the loss of lives has become the focus of public and social attention. At present, research on small reservoir risk-induced life loss is still underdeveloped in China. Unlike other impact factors, the affected population is a cost-type factor of risk. Once a reservoir dam breaks, it will cause immeasurable losses of lives and property to people living downstream. By integrating construction and management data from water conservancy projects, this study has identified the affected population as a major security risk factor from the viewpoint of data analysis, providing an important theoretical basis for water conservancy safety regulators.

## 5. Conclusions

(1) The grey relational analysis model for heterogeneous data introduced in this paper was effectively used for the correlation analysis of categorical data and numerical data. It is applicable to identification of security risk factors of small reservoirs with heterogeneous and scarce data.

(2) The case study of comprehensive security risk factor identification of Guangxi small reservoirs shows that the geological conditions, construction and operation management, hydrological monitoring, and affected population are four risk clusters extracted from heterogeneous data of 17 factors. Because the data of these factors can be obtained directly from the information system, the results are conducive to selecting risk factors and constructing risk assessment models of small reservoirs with insufficient and heterogeneous data.

Due to the lack of data, this study only selected some factors affecting the security risk of small reservoirs in Guanxi Zhuang Autonomous Region in China, and the results obtained with this method may have some limitations. With the development of the operational management informationization of small reservoirs and the supplementation of various data, more factors can be analyzed in the future to provide a better guarantee for the prediction and early warning of security risks for small reservoirs.

## Acknowledgements

The authors would like to thank the Department of Water Resources and the Office of Guangxi Flood Control and Drought Relief Headquarters, the Nanjing Hydraulic Research Institute, and the Beijing Guoxinhuayuan Technology Co., Ltd. (BGT) for their great support in providing the reservoir data for the present study.

## References

- Cai, Q., Wang, Z.S., Zhu, S.Y., Ding, J.L., 2018. A real-time risk assessment and synthetical early warning model of earth and rockfill dams. *Adv. Sci. Technol. Water Resour.* 38(4), 57–63 (in Chinese). <https://doi.org/10.3880/j.issn.1006-7647.2018.04.010>.
- Deng, J.L., 1989. Introduction to grey system theory. *J. Grey Syst.* 1(1), 1–24.
- Duckstein, L., Plate, E.J., 1987. *Engineering Reliability and Risk in Water Resources*. Springer, Dordrecht. <https://doi.org/10.1007/978-94-009-3577-8>.
- Feng, X.H., 2015. Evaluation of dam safety based on cloud model and entropy weight method. *Water Resour. Power* 33(11), 57–60 (in Chinese).
- Fu, H.H., Li, Z.W., Hu, J.M., 2011. Risk management and response countermeasures for small reservoirs. *China Rural Water Hydropower* (4), 150–151 (in Chinese).
- Georgakakos, K.P., 2006. Analytical results for operational flash flood guidance. *J. Hydrol.* 317(1–2), 81–103. <https://doi.org/10.1016/j.jhydrol.2005.05.009>.
- Gu, H., Huang, X.F., Ding, J., 2014. Influence factors of high dam reservoir basin deformation based on finite element analysis and weights analysis. *Adv. Mater. Res.* 919–921, 1354–1360. <https://doi.org/10.4028/www.scientific.net/AMR.919-921.1354>.
- Hipel, K.W., 2011. Grey systems: Theory and applications. *Grey Syst. Theor. Appl.* 1(3), 274–275. <https://doi.org/10.1108/gs.2011.1.3.274.1>.
- Li, H.E., He, Y.J., Liu, X.Q., 2016. *Key Techniques for Early Warning and Disaster Mitigation of Reservoir Disasters in Torrential Flood-Prone Areas*. China Water&Power Press, Beijing (in Chinese).
- Li, X.L., Li, C.L., Li, D.Y., Wang, H.B., 2010. Evaluation of dam safety based on AHP combining with fuzzy math. *Yangtze River* 41(17), 92–95 (in Chinese). <https://doi.org/10.3969/j.issn.1001-4179.2010.17.024>.
- Liu, C., Shen, Z.Z., Gan, L., Xiong, S.F., Lu, J., 2018. Comprehensive evaluation of rehabilitation degree of dangerous reservoirs based on fuzzy gray clustering and combination weighting method. *Adv. Sci. Technol. Water Resour.* 38(3), 36–41 (in Chinese). <https://doi.org/10.3880/j.issn.1006-7647.2018.03.007>.
- Liu, S., Forrest, Y.L., 2013. Advances in grey systems research. *Underst. Complex. Syst.* 25(2), 1–18. <https://doi.org/10.1007/978-3-642-13938-3>.
- Luo, D., Wei, B.L., Lin, Y.W., 2015. The optimization of several grey incidence analysis models. *J. Grey Syst.* 27(4), 1–11. <https://doi.org/10.1109/GSIS.2015.7301849>.
- Ma, F.P., Zhang, X.Y., Li, L.Z., 2009. Construction of water quality monitoring and security system in Douhe Reservoir water head area of Tangshan. *Environ. Sci. Technol.* 32(3), 119–121 (in Chinese).
- Malings, C., Pozzi, M., 2016. Conditional entropy and value of information metrics for optimal sensing in infrastructure systems. *Struct. Saf.* 60, 77–90. <https://doi.org/10.1016/j.strusafe.2015.10.003>.
- Mendel, J.M., 2014. General type-2 fuzzy logic systems made simple: A tutorial. *IEEE Trans. Fuzzy Syst.* 22(5), 1162–1182. <https://doi.org/10.1109/TFUZZ.2013.2286414>.
- Ministry of Water Resources of the People's Republic of China (MWRPRC), 2010. Security management measures for small reservoirs. *China Water Resour.* 30(12), 5–6 (in Chinese).
- Ministry of Water Resources of the People's Republic of China (MWRPRC), 2013. *Bulletin of First National Census for Water*. China Water&Power Press, Beijing (in Chinese).
- Runnenburg, J.T., 1985. Probability theory and its application. *Stat. Neerl.* 39(2), 181–189. <https://doi.org/10.1111/j.1467-9574.1985.tb01137.x>.
- Saedi, A.M., Thambirajah, J.J., Pariatamby, A., 2014. A HIRARC model for safety and risk evaluation at a hydroelectric power generation plant. *Saf. Sci.* 70, 308–315. <https://doi.org/10.1016/j.ssci.2014.05.013>.
- Shannon, C.E., 1951. Prediction and entropy of printed English. *Bell Syst. Tech. J.* 30(1), 50–64. <https://doi.org/10.1002/j.1538-7305.1951.tb01366.x>.
- Sheng, J.B., Feng, J.Y., Peng, X.H., 2008. Research on risk analysis of small reservoir dams. *Hydro-Sci. Eng.* (1), 28–35 (in Chinese). <https://doi.org/10.3969/j.issn.1009-640X.2008.01.005>.
- Wang, J., Pan, X., Wang, L.G., Wei, W., 2018. Method of spare parts prediction models evaluation based on grey comprehensive correlation degree

- and association rules mining: A case study in aviation. *Math. Probl. Eng.*. <https://doi.org/10.1155/2018/2643405>.
- Wang, J.J., Hipel, K.W., Dang, Y.G., 2017. An improved grey dynamic trend incidence model with application to factors causing smog weather. *Expert Syst. Appl.* 87, 240–251. <https://doi.org/10.1016/j.eswa.2017.06.012>.
- Wang, L.H., Yin, K.D., Cao, Y., Li, X.M., 2019. A new grey relational analysis model based on the characteristic of inscribed core (IC-GRA) and its application on seven-pilot carbon trading markets of China. *Int. J. Environ. Res. Public Health* 16(1). <https://doi.org/10.3390/ijerph16010099>.
- Yan, F., Qiao, D.Y., Qian, B., Ma, L., Xing, X.G., Zhang, Y., Wang, X.G., 2016. Improvement of CCME WQI using grey relational method. *J. Hydrol.* 543(B), 316–323. <https://doi.org/10.1016/j.jhydrol.2016.10.007>.
- Yang, X.G., 2014. Analysis of uncertainty factors in security operation risks of small reservoirs. *Sci. Technol. Innov.* (32), 235 (in Chinese). <https://doi.org/10.3969/j.issn.1673-1328.2014.32.222>.
- Zadeh, L.A., 1965. Fuzzy sets. *Inf. Control* 8(3), 338–353. [https://doi.org/10.1016/S0019-9958\(65\)90241-X](https://doi.org/10.1016/S0019-9958(65)90241-X).
- Zhang, H.Y., Fu, C.S., Huang, H.Y., Yuan, K., 2017. Comprehensive evaluation on safety of reservoir dam in susceptible area of mountain torrent. *Water Conserv. Sci. Technol. Econ.* 23(9), 63–68. <https://doi.org/10.3969/j.issn.1006-7175.2017.09.014> (in Chinese).
- Zhang, K., Zhang, Y.T., Qu, P.P., 2014. Comprehensive multivariate grey incidence degree based on principal component analysis. *J. Syst. Eng. Electron.* 25(5), 840–847. <https://doi.org/10.1109/JSEE.2014.00097>.