The triggering threshold estimation of Majiagou landslide based on data mining

Lei Zhang\textsuperscript{1,2}, Yifei Cui\textsuperscript{1*}, Bin Shi\textsuperscript{2}

\textsuperscript{1} State Key Laboratory of Hydrosience and Engineering, Tsinghua University, Beijing 100084, China
\textsuperscript{2} School of Earth Sciences and Engineering, Nanjing University, Nanjing 210023, China

* Corresponding authors’ e-mail: yifeicui@mail.tsinghua.edu.cn (Y.C.)

Abstract: The deformation of the reservoir landslide is mainly governed by the combined action of rainfall and the fluctuation of the reservoir water level. The determination of the threshold of triggering factors is of great significance in the stability analysis and evaluation of the potential landslide. Existing empirical threshold model is mainly based on statistical analysis to fit the explicit function between triggering factors and displacement, which is widely used in rainfall-triggered landslides. However, for reservoir landslides, the relationship among rainfall intensity, reservoir water level, fluctuation rate and landslide displacement is highly nonlinear, which hindered the application of existing empirical threshold model. To tackle the scientific challenge, a novel data mining-based threshold estimation method is proposed in this study. The Majiagou landslide, located in the Three Gorges Reservoir (TGR) region, is selected as the study site. Firstly, the Distributed Fiber Optical Sensing (DFOS) technology has been adopted to record the rainfall, reservoir water level fluctuations, and deformation information for two years in real time; Then, the evolution pattern of Majiagou landslide was analyzed in depth; Finally, the cluster analysis and decision tree algorithm are used to determine the threshold value of the rainfall and the fluctuation of the reservoir water level. Among which, 80% of the data set is used for training model, and the remaining 20% is used for validation. The study here provides a new and effective method to estimate the triggering threshold and contribute to the prediction and early warning of reservoir landslides.

Keywords: Reservoir landslide; Triggering factors; Threshold estimation; Data mining

1. Introduction
An estimated of 5386 landslides have occurred since the impoundment of the Three Gorges Reservoir (TGR) in 2003 [1]. Compared with other landslides, reservoir landslides are more widespread, diversified and intense. Infrastructure such as roads, bridges, houses have been destroyed in patterns that can threaten people's lives [2-3]. The estimation of triggering factors’ threshold is an economical and effective way to achieve stability evaluation and early-warning of reservoir landslide.

The landslide threshold estimation methods can be mainly divided into physical threshold method and empirical threshold method in the existing research. The physical threshold of landslide is determined based on rainfall infiltration, hydrological conditions and landslide failure mechanism.
Whereas the empirical threshold of landslide is determined by statistical analysis of the monitoring data. As there are no complex mechanical models and assumptions involved in empirical threshold model, which makes it widely adopted in landslide threshold estimation. Caine [4] reviewed 73 rainfall induced shallow landslides, and proposed the rainfall intensity duration threshold. Thereafter, many scholars have focused on the study of triggering threshold estimation of landslide. The accumulated rainfall duration empirical threshold model is proposed by Cannon [5]. For specific, Jibson [6] proposed an empirical threshold model of cumulative rainfall intensity. However, reservoir landslides in the Three Gorges Reservoir (TGR) region are governed by rainfall and reservoir water level fluctuation. At present, most studies only focus on the rainfall threshold and ignores the influence of reservoir water level.

In this regard, data mining as an effective technique to establish nonlinear relationship among different parameters has been introduced in landslide recently. A few landslide-related analysis methods have integrated data mining algorithms in different forms, including decision tree [7-8], Bayesian network [9-10], artificial neural network [11-12], support vector machine (SVM) [13]. However, data mining methods are generally used in risk assessment and early-warning and the triggering factors’ threshold of reservoir landslide is rarely investigated.

In this study, Majiagou landslide, a traditional reservoir landslide located in Three Gorges Reservoir (TGR) Region was selected as the study site. The real-time monitoring system has been installed and the associated rainfall, reservoir water level and displacement were recorded. For the sake of estimating the threshold of triggering factors, data mining method including the cluster analysis and decision tree algorithm was proposed. By learning the characteristics of training dataset, the nonlinear relationship between rainfall, reservoir water level and displacement can be established. Thus, the threshold of rainfall and reservoir water level can be determined. This study provides a scientific method to determine the threshold of triggering factors, which is of great significant in early-warning of reservoir landslide.

2. Study site
2.1 General setting of Majiagou landslide
The Majiagou landslide, initiated when the reservoir was first impounded in 2003, is 560 m long, 180-210 m wide, and covers an area of 9.8×104 m2. The geographic location of the Majiagou landslide is 31°01’08″~31°01’17″ North latitude and 110°41’48″~110°42’10″ East longitude. It is situated on the left bank of the Zhaxi River, a tributary river of the Yangtze River in Zigui County, Hubei Province (Fig.1). The front of the landslide is a fluvial alluvial terrace and forms multi-stage gentle platform and steep sided ridges. Two East-West trending gullies, with an undercutting depth of 50-60 m, shapes the boundary of Majiagou landslide (Fig.1). Its average gradient is approximately 15° within the range of 12° to 20° according to the relief elevation of the ground surface.

The Majiagou landslide are mainly constituted of surficial deposits and sedimentary bedrock (Fig.2). Between them, the soft stratum exists. The surficial deposits are mainly composited of residual silty clay with gravels interbedded with little sandstone, whose permeability is around 6.5×10^-4 m/s. The bedrock, with dip direction of 270-290° and dip angle of 25-30°, comprises quartz sandstones interbedded with thin silty mudstones of the Jurassic Suining Formation with a permeability of around 1.0×10^-7 m/s. The annual rainfall and reservoir water level fluctuate seasonally. The annual rainfall of Majiagou landslide is abundant, about 1066 mm. Among them, the rainfall from May to September is the most concentrated, accounting for more than two-thirds of the annual rainfall. The water level in the reservoir fluctuates annually from 145 m to 175 m.
2.2 Monitoring system based on DFOS

The remote real-time monitoring system was installed on Majiagou landslide. The system consists of the real-time monitoring unit, remote processing unit (with data storage and analysis capacity) and decision-making unit (Fig.3).

2.2.1 Real-time monitoring unit

(1) Rainfall monitoring

Daily rainfall data were collected by a weather station that was installed on the landslide. The monitoring results can be transmitted to the control center via the GPRS module.

(2) Reservoir water level monitoring
The daily reservoir water level was measured by the Hydrology Bureau of Changjiang Water Resources Commission and presented at the website (http://www.cjsyw.com).

![Figure 3. Real-time monitoring system](image)

(3) Displacement monitoring
Two in-place inclinometers were installed just in the deep sliding zone of inclinometers B8 and B9 [14], which were introduced into the landslide monitoring station through optical cables and connected to a 16-channel FBG demodulator in January 2016.

2.2.2 Remote processing unit
When the real-time monitoring unit is in operation, singularity and noise inevitably exist because of the characteristic of the signal, the limitations of the electronic instrument, and the disturbance of the ambience. To minimize this problem, the singular values generated in the monitoring process were eliminated using the singular value test, and the ellipse filter was used to filter the noise signals that occurred during the test process.

3. Triggering factors’ threshold estimation method
3.1 Estimation method

3.1.1 Two-step clustering method
The two-step clustering method realizes the data clustering process by pre-clustering and clustering (Fig.4). Pre-clustering uses “sequential” method to roughly divide samples into several sub categories. Based on pre-clustering, the clustering process also judges whether the sub-clusters generated in the pre-clustering can be merged according to the “degree of affinity” of the samples, and finally the sample data are divided into L categories.

![Fig. 4 Two-step clustering](image)
The establishment of decision tree model can be divided into the following two steps: (1) Decision tree model training stage: according to the specified training data set, selecting the appropriate decision tree model. (2) Decision tree model prediction stage: Clarify each type of data in the data set based on the trained decision tree model, forming classification rules, and then predict the unknown data set.

3.2 Modeling process

As shown in Fig.5, the process of threshold estimation includes model establishment, evaluation model and model implementation.

Model establishment: Firstly, the rainfall index, reservoir water level index and landslide deformation index are expressed by two-step clustering method. Then, the landslide deformation evolution state is set as the output term and the external hydrological inducing factor as the input term. The dataset is divided into training dataset and test dataset. Among which, 80% of the data set is used for training model, and the remaining 20% is used for validation. Thus, the decision tree model is established.

Evaluation model: The decision tree model is evaluated according to the evaluation standard of data mining. It is noted that the decision tree model should have high accuracy and not conflict with the current experience and knowledge.

Model implementation: Generating classification and prediction rules of input variables and output variables. The decision tree model can predict output variables according to the new input variables. According to the decision tree model, the rule criterion between the landslide deformation evolution and the external triggering factors can be established to realize the accuracy estimation of landslide threshold.

Figure 5. The general steps of threshold estimation of Majiagou landslide through data mining

4. Result and analysis

Fig. 6 depicts the relationship between horizontal cumulative displacement and the hydrological factors. The cumulative horizontal displacement didn’t increase with time, but showed a stepped increase tendency. In early May 2016 and July 2017, intense rainfall occurred, the landslide displacement increased correspondingly. Nevertheless, the fluctuation of reservoir water level
indicated a strong seasonal influence on the deformation of Majiagou landslide. This indicated that rainfall and reservoir water level fluctuation are closely related to the deformation of the landslide.

According to the previous research results, the quantitative indexes of rainfall are Monthly maximum daily rainfall and Total monthly rainfall intensity, the quantitative indexes of reservoir water level are monthly average water level and monthly change of water level, and the monthly velocity is selected as the quantitative index of landslide deformation. The rainfall index, reservoir water level index and landslide deformation index are expressed by two-step clustering method, as illustrated in Tables 1, 2 and 3.

**Table 1 Qualitative value of rainfall**

| Monthly maximum daily rainfall intensity (mm) | Qualitative value | Total monthly rainfall intensity $q_{\text{month}}$ (mm/month) | Qualitative value |
|---------------------------------------------|-------------------|-------------------------------------------------------------|-------------------|
| (70-109)                                    | Heavy             | (222, 322)                                                  | Heavy             |
| (37-53)                                     | Moderate          | (80, 174)                                                   | Moderate          |
| (3-22)                                      | Light             | (5, 48)                                                     | Light             |

**Table 2 Qualitative value of water level**

| Monthly average water level $h$ (m) | Qualitative value | Monthly change of water level $\Delta h$ (m/month) | Qualitative value |
|-----------------------------------|-------------------|-----------------------------------------------------|-------------------|
| (170-175)                         | High              | (8.5, 11.4)                                         | Sharp increase    |
| (160-167)                         | Medium            | (3.5, 4.1)                                          | Moderate increase |
| (145-158)                         | Low               | (0, 1.7)                                             | Slow increase     |
|                                  |                   | (-3.5, -0.5)                                        | Slow drop         |
|                                  |                   | (-5.3, -4.2)                                        | Moderate drop     |
|                                  |                   | (-11.2, -10.5)                                      | Sharp drop        |

**Table 3 Qualitative value of displacement velocity**

| Monthly velocity $v$ (mm/month) | Qualitative value |
|---------------------------------|-------------------|
| (7.5-9.2)                       | High              |
| (5.5-6.8)                       | Medium            |
| (1.1-4.5)                       | Low               |
Six criterions of triggering factors’ threshold of Majiagou landslide are established as depicted in Table 4 by decision tree model. The overall accuracy of the evolution threshold criterion is high, which can be used to predict the landslide deformation.

Criterion 1 indicates that if the reservoir water level is higher than 170 m, the landslide will deform at a low rate when the monthly maximum daily rainfall intensity is less than 21 mm. The accuracy of the rule is 87%.

Criterion 2-3 suggests that the landslide deforms in a medium rate. Criterion 2 can be interpreted as: when the monthly maximum daily rainfall intensity exceeds 23.7mm, the total monthly rainfall intensity is less than 174 mm and the monthly change of water level is smaller than -1.25, the landslide will deform at a medium rate. The accuracy of the criterion can attains as high as 98%. Criterion 3 signifies that when the monthly maximum daily rainfall intensity exceeds 23.7 mm, the total monthly rainfall intensity is more than 174 mm and the monthly change of water level is bigger than -1.25, the landslide will deform at a medium rate.

Criterion 4-5 indicates the landslide deforms at a high rate. Criterion 4 with an accuracy of 84% signifies that when the reservoir water level is less than 153 m and the maximum daily rainfall is less than 23.7 mm, the landslide deforms at a high rate. Criterion 5 can be expressed as follows: when the monthly maximum daily rainfall intensity exceeds 23.7 mm, the total monthly rainfall intensity is more than 174 mm and the monthly change of water level is smaller than -1.25, the landslide deforms at a higher. The accuracy of criterion 5 is as high as 100%.

Compared with criteria 1 and 4, the reservoir water level when landslide deformation rate transmits from low to medium is 153 m. It indicates that when the reservoir water level is high, the deformation rate of Majiagou landslide is low. Whereas, when the reservoir water level drops to 153m, the deformation rate transmits from low to medium, which can be attributed to the hydrostatic pressure. That is, the drop of reservoir water level leads to the decrease of hydrostatic pressure acting on the slope surface, which contributes to the deformation of the landslide. According to criteria 2 and 5, it is founded that the threshold of total monthly rainfall for rapid deformation is 174 mm. Under the action of intense rainfall, the seepage force pointed to the exterior of the landslide will accelerate the deformation of the landslide. In addition, with the infiltration of rainfall, the water content of the soil increases, resulting in the decrease of soil shear strength, which eventually accelerates the deformation of Majiagou landslide. According to criterion 3 and 5, the reservoir water level fluctuation threshold for rapid deformation of Majiagou landslide is - 1.25 m/ month, that is, the reservoir water level drops at a rate of 1.25 m / month. It can be inferred that when the fluctuation rate of the reservoir water level exceeds - 1.25 m / month, the pore water pressure inside the slope shows an inadaptable response. Due to the inadequate dissipation of pore water, the transient seepage occurred inside the slope deposits above the water level. Meanwhile, the transient flow inside slope deposits forms seepage force downslope along the slip direction, thereby accelerating the deformation of the landslide.

| Deformation | NO | Criterion | Accuracy |
|-------------|----|-----------|----------|
| Low         | 1  | if $q \leq 21 \& h > 170$ then $v = low$ | 87 |
| Medium      | 2  | if $q > 23.7 \& \Delta h < -1.25 \& q_{month} \leq 174$ then $v = Medium$ | 98 |
|             | 3  | if $q > 23.7 \& \Delta h \geq -1.25 \& q_{month} > 174$ then $v = Medium$ | 63 |
| High        | 4  | if $q \leq 21 \& h \leq 153$ then $v = high$ | 87 |
|             | 5  | if $q > 23.7 \& \Delta h < -1.25 \& q_{month} > 174$ then $v = high$ | 100 |

5. Conclusion
Taking Majiagou landslide as an example, the real-time monitoring system has been installed and the associated rainfall, reservoir water level and displacement were recorded. Noted that rainfall and reservoir water level are two main triggering factors that govern the deformation of reservoir landslide.
For the sake of estimating the threshold of triggering factors, data mining method including the cluster analysis and decision tree algorithm was proposed. Firstly, the rainfall index, reservoir water level index and landslide deformation index are expressed by two-step clustering method. Then, the landslide deformation evolution state is set as the output term and the external hydrological inducing factor as the input term. The dataset is divided into training dataset and test dataset. Among which, 80% of the data set is used for training model, and the remaining 20% is used for validation. Thus, the decision tree model is established. According to the threshold criterion for deformation evolution of Majiagou landslide based on decision tree model, it is founded that: the reservoir water level when landslide deformation rate transmits from low to medium is 153 m and the threshold of total monthly rainfall for rapid deformation is 174 mm. In addition, the reservoir water level fluctuation threshold for rapid deformation of Majiagou landslide can also be estimated and the critical value is - 1.25 m/month. The study here provides a new and effective method to estimate the triggering threshold and contribute to the prediction and early warning of reservoir landslides.

Acknowledgements

The authors gratefully acknowledge the financial support provided by the National Natural Science Foundation of China (42077238, 41941019).

Reference

[1] Wang F, Li T 2009 Landslide Disaster Mitigation in Three Gorges Reservoir, China. Mt Res Dev 30 184-185.
[2] Xu GL, Li WN, Yu Z, Ma XH, Yu ZZ 2015 The 2 September 2014 Shanshucao landslide, Three Gorges Reservoir, China. Landslides 12 1169-1178.
[3] Paronuzzi P, Bolla A, Rigo E 2016 Brittle and Ductile Behavior in Deep-Seated Landslides: Learning from the Vajont Experience. Rock Mech Rock Eng 49 2389-2411.
[4] Caine Nel 1980 The rainfall intensity: duration control of shallow landslides and debris flows[J]. Geografiska Annaler. Series A, Physical Geography 62 23-27.
[5] Cannon SH., Ellen SD 1985 Rainfall conditions for abundant debris avalanches. San Francisco Bay region, California. California Geology 38 267-272.
[6] Jibson Randall W 1989 Debris flows in southern Puerto Rico[J]. Geological Society of America Special Papers 236 29-56.
[7] Althuwaynee OF, Pradhan B, Park HJ, Lee JH 2014 A novel ensemble decision tree based Chi-squared Automatic Interaction Detection (CHAID) and multivariate logistic regression models in landslide susceptibility mapping. Landslides 11 1063–1078.
[8] Nefeslioglu HA, Sezer E, Gokceoglu C, Bozkir AS, Duman TY 2010 Assessment of landslide susceptibility by decision trees in the metropolitan area of Istanbul, Turkey. Math Probl Eng.
[9] Liang WJ, Zhuang DF, Jiang D, Pan JJ, Ren HY 2012 Assessment of debris flow hazards using a Bayesian network. Geomorphology 171 94–100.
[10] Song YQ, Gong JH, Gao S, Wang DC, Cui TJ, Li Y, Wei BQ 2012 Susceptibility assessment of earthquake-induced landslides using Bayesian network: a case study in Beichuan, China. Comput Geosci 42 189–199.
[11] Ermini L, Catani F, Casagli N 2005 Artificial neural networks applied to landslide susceptibility assessment. Geomorphology 66 327–343.
[12] Zhang L, Shi B, Zhu HH, Yu Nill, Wei GQ 2020a A Machine learning method for inclinometer lateral deflection calculation based on distributed strain sensing technology. Bulletin of Engineering Geology and the Environment, 79 3383–3401.
[13] Zhang L, Shi B, Zhu HH, Yu Bill, Han HM, Fan XD 2021 PSO-SVM-based deep displacement prediction of Majiagou landslide considering the deformation hysteresis effect. Landslides 2021 18 179–193.
[14] Zhang L, Shi B, Zhang D, Sun YJ 2020b Inyang I.H; Kinematics, triggers and mechanism of Majiagou landslide based on FBG real-time monitoring. Environmental Earth Science 79
200.