Journal of Rock Mechanics and Geotechnical Engineering xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Rock Mechanics and Geotechnical Engineering

journal homepage: www.jrmge.cn

Full Length Article

Spatiotemporal deformation characteristics of Outang landslide and identification of triggering factors using data mining

Beibei Yang [a], Zhongqiang Liu [b,*], Suzanne Lacasse [b], Xin Liang [c]

[a] School of Civil Engineering, Yantai University, Yantai, 264005, China
[b] Norwegian Geotechnical Institute, Oslo, 0855, Norway
[c] Faculty of Engineering, China University of Geosciences, Wuhan, 430074, China

ARTICLE INFO

Article history:
Received 20 February 2023
Received in revised form
11 July 2023
Accepted 4 September 2023
Available online xxx

Keywords:
Landslide
Deformation characteristics
Triggering factor
Data mining
Three gorges reservoir

ABSTRACT

Since the impoundment of Three Gorges Reservoir (TGR) in 2003, numerous slopes have experienced noticeable movement or destabilization owing to reservoir level changes and seasonal rainfall. One case is the Outang landslide, a large-scale and active landslide, on the south bank of the Yangtze River. The latest monitoring data and site investigations available are analyzed to establish spatial and temporal landslide deformation characteristics. Data mining technology, including the two-step clustering and Apriori algorithm, is then used to identify the dominant triggers of landslide movement. In the data mining process, the two-step clustering method clusters the candidate triggers and displacement rate into several groups, and the Apriori algorithm generates correlation criteria for the cause-and-effect. The analysis considers multiple locations of the landslide and incorporates two types of time scales: long-term deformation on a monthly basis and short-term deformation on a daily basis. This analysis shows that the deformations of the Outang landslide are driven by both rainfall and reservoir water while its deformation varies spatiotemporally mainly due to the difference in local responses to hydrological factors. The data mining results reveal different dominant triggering factors depending on the monitoring frequency: the monthly and bi-monthly cumulative rainfall control the monthly deformation, and the 10-d cumulative rainfall and the 5-d cumulative drop of water level in the reservoir dominate the daily deformation of the landslide. It is concluded that the spatiotemporal deformation pattern and data mining rules associated with precipitation and reservoir water level have the potential to be broadly implemented for improving landslide prevention and control in the dam reservoirs and other landslide-prone areas.

1. Introduction

Landslides constitute a geological hazard across numerous regions worldwide, resulting in catastrophic social and economic consequences (Guzzetti et al., 2012; Froude and Petley, 2018; Mirus et al., 2020; Mandal et al., 2021). The Three Gorges Reservoir Area (TGRA), characterized by large, narrow steep terrain, and unfavorable geology, represents one of the regions of China most prone to landslides (Wang et al., 2021; Guo et al., 2022a; Zhang et al., 2022a). A report of 2009 from the Ministry of Land and Resources of China announced that around 5000 landslides were deforming actively over a length of 2000 km (Yang et al., 2017). Furthermore, more than 500 landslides occurred since the Three Gorges Reservoir (TGR) was first impounded in 2003 (Zhou et al., 2022). Consequently, concentrated endeavors have been undertaken to prevent and control reservoir landslides in this region (Fourniadis et al., 2007; Tang et al., 2019; Yang et al., 2022; Zhang et al., 2022b). One essential initiative was to install monitoring equipment and develop early warning systems for landslides. About 3200 such monitoring systems with multiple equipment have been put into operation (Yin et al., 2010; Park et al., 2019; Guzzetti et al., 2020).

Landslides result from the interactions between triggering events and local natural conditions (Kumar and Anbalagan, 2015; Pasierb et al., 2019). Local natural conditions, referring to inherent factors, include topography, geology, tectonics, and geotechnical materials. Triggering events, are mainly hydrological elements, including precipitation and dam reservoir operation (Gariano and

* Corresponding author.
E-mail address: Zhongqiang.Liu@ngi.no (Z. Liu).

Peer review under responsibility of Institute of Rock and Soil Mechanics, Chinese Academy of Sciences.

https://doi.org/10.1016/j.jrmge.2023.09.030

Guzzetti, 2016; Strauhal et al., 2016; Gariano et al., 2020; Wang et al., 2023). The TGR initiated a periodic fluctuation between elevations of 145 m and 175 m in 2009. The 30-m filling and drawdown of the reservoir significantly altered the hydrogeological conditions of the TGRA, resulting in a reduction in slope stability and inducing a significant quantity of landslides (Weidner et al., 2019; Miao et al., 2022; Zhang et al., 2022c). Moreover, many areas in the TGRA have abundant rainfall with an average annual rainfall exceeding 1000 mm. Fengjie County, which hosts the Outang landslide studied in this paper, is such an example where rainfall impacts slope stability by altering pore water pressure distribution and decreasing slope mass (Chinkulkijniwat et al., 2019; Sun et al., 2021).

Monitored data, including surface displacement, precipitation, and reservoir water level, have been shown to form the basis for dam reservoir-related landslide incidences (Tomás et al., 2016; Pecoraro et al., 2019; Dai et al., 2021). Advances in the technology for monitoring and remote sensing, as well as the interplay of multiple equipment, create large quantities of data. The capacity and efficiency of traditional data analysis processes are challenged, especially as the quantities and complexities of data have multiplied (Wu et al., 2014). Data mining can transform a vast quantity of multi-sourced data into valuable messages in an intelligent way (Lee and Siau, 2002; McClean, 2003). Because of this unique capability, data mining has garnered significant prominence in the exploration of the inherent correlation criteria within datasets and has been introduced in identifying triggers inducing landslide deformation.

For example, Wang et al. (2015) employed data mining techniques to reveal the causal mechanisms, triggering factors, deformation characteristics, and forecast criteria of landslides in the TGRA. Ma et al. (2017), Miao et al. (2021), and Guo et al. (2022b) used the two-step clustering and the Apriori algorithm technology to determine triggers inducing riverbank landslides. Yao et al. (2019) identified triggers resulting in the deformation of the Baijiabao landslide by neighborhood rough set theory. Franceschini et al. (2022) concluded that data mining helped determine the regions and times most exposed to landslides. It should be noted that the cause-and-effect relationship between deformation and triggering factors can vary both temporally and spatially for landslides (Wu et al., 2016). Furthermore, the latest monitoring datasets with shorter intervals have the potential to be used in landslide risk management across various time scales.

The objective of this paper is threefold. Firstly, it aims to develop a comprehensive understanding of the spatiotemporal deformation characteristics of riverbank landslides based on in situ information obtained from field investigation and monitoring systems. Secondly, it seeks to identify the triggers that induce deformation in various parts of a landslide. Thirdly, it analyzes and compares the spatiotemporal deformation characteristics and hydrological triggers of riverbank landslides using monthly and daily time scales for displacement measurements in situ. Through these three objectives, the study will contribute to enhancing our understanding of the deformation mechanism of riverbank landslides and providing valuable insights into movement prediction, stability evaluation, and risk control strategies.

The Outang landslide was a large riverbank landslide in the TGRA, which has undergone deformation since 2003. A monitoring system was established on the landslide with continuous replenishment of new monitoring instruments. The advanced monitoring system provides abundant monitoring data over several years, which is not the case for all the other landslides at the sites. The latest monitoring indicates that the displacement of the Outang landslide is continuing and increasing (so far). A landslide would present a threat to a total of 8950 residents (Chongqing Municipal Planning and Natural Resources Bureau, 2021). Further research on the Outang landslide based on the latest and detailed monitored data is both meaningful and essential (Dai et al., 2016; Guo et al., 2020; Luo and Huang, 2020; Liao et al., 2022; Yan et al., 2022). The novelty of this new study on the Outang landslide resides in: (1) understanding better the on-going displacement patterns (which urgency is important because of the threat to the population); (2) the use of monitoring data with different record frequencies and being able to associate the measurements with different periods of hydrological influence factors; (3) interpretation of the latest data with a combination of machine learning techniques; and (4) discussing the opportunity of the analyses presented in the study to assist in developing strategies for landslide risk control.

The current study focuses on analyzing the spatiotemporal deformation characteristics of the Outang landslide. Data mining techniques, including two-step clustering and the Apriori algorithm, are employed for exploring causal relationships between hydrological factors and landslide deformation. According to the association rules obtained by data mining, the dominant triggering factors for different parts of the Outang landslide are obtained, considering monitoring data over two time scales, monthly and daily. The motivation for the data mining work is that it has a unique potential to provide beneficial insights into landslide zoning prediction and assist in controlling the risk.

2. Materials and methods

2.1. Geological setting of the outang landslide

The Outang landslide lies on the lower valley of the Yangtze River, Fengjie County, China, approximately 177 km upstream from the Three Gorges Dam (Fig. 1). The geographic coordinates are 30°17′12″N and 109°18′26″E. The elevation of the upper edge is 705 m above sea level, whereas its toe has been submerged in the Yangtze River with elevations ranging from 90 m to 102 m. It is bounded laterally by a ridge and a gully. The landslide extends 1800 m from south to north and 900 m from east to west. The thickness of the landslide mass ranges from 2.8 m to 128 m, with an average thickness of 50.8 m. The landslide has an area of 1.8×10^6 m^2 and an estimated volume of 9×10^7 m^3.

After detailed investigations of the Outang landslide by several professional organizations, the landslide is found to have been experienced sliding three times (Fig. 2a). The first sub-unit has an area of 9.2×10^5 m^2 and an elevation of 90–370 m. Two local strong deformation zones appear on either side of the first sub-unit, recorded as eastern and western strong-deformation zones, which are more unstable than the rest of the landslide and still deform with some local movement. The second sub-unit covers the trailing part of the first one. The total area of this sub-unit is approximately 3.2×10^5 m^2, and it covers 250–530 m. The third sub-unit is at the top of the landslide, whereas it stretches over the crown surface of the second one. This sub-unit has an estimated area of 5.4×10^5 m^2, with an elevation from 400 m to 705 m (Guo et al., 2020).

The Outang landslide consists mainly of Quaternary deposits and fractured sandstone from upper to lower, with the latter component dominating the content. The sliding body ranges in thickness from 10 m to 95 m. The bedrock consists of sandstone in the Jurassic System, exhibiting an orientation of 335°–350°/18°–24°. Two incompetent layers (R1 and R3), each contributing as sliding planes, are the interface between the upper sliding body and underlying bedrock (Fig. 2b). The first and second sub-units move along the incompetent layer of R3, while the third one moves along R1. The thicknesses of R1 and R3 are 0.5 m and 0.1 m, respectively.

Please cite this article as: Yang B et al., Spatiotemporal deformation characteristics of Outang landslide and identification of triggering factors using data mining, Journal of Rock Mechanics and Geotechnical Engineering, https://doi.org/10.1016/j.jrmge.2023.09.030

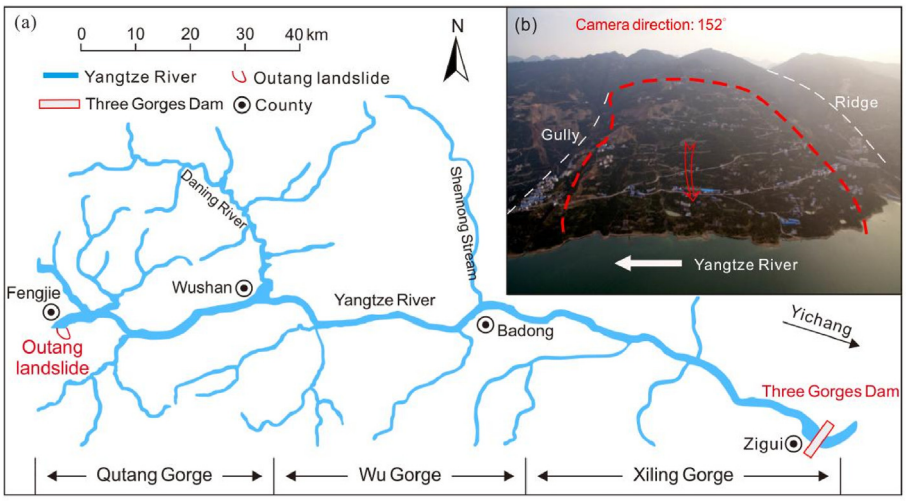

Fig. 1. (a) Map of Three Gorges Dam, the Yangtze River and the Outang landslide; and (b) The Outang landslide.

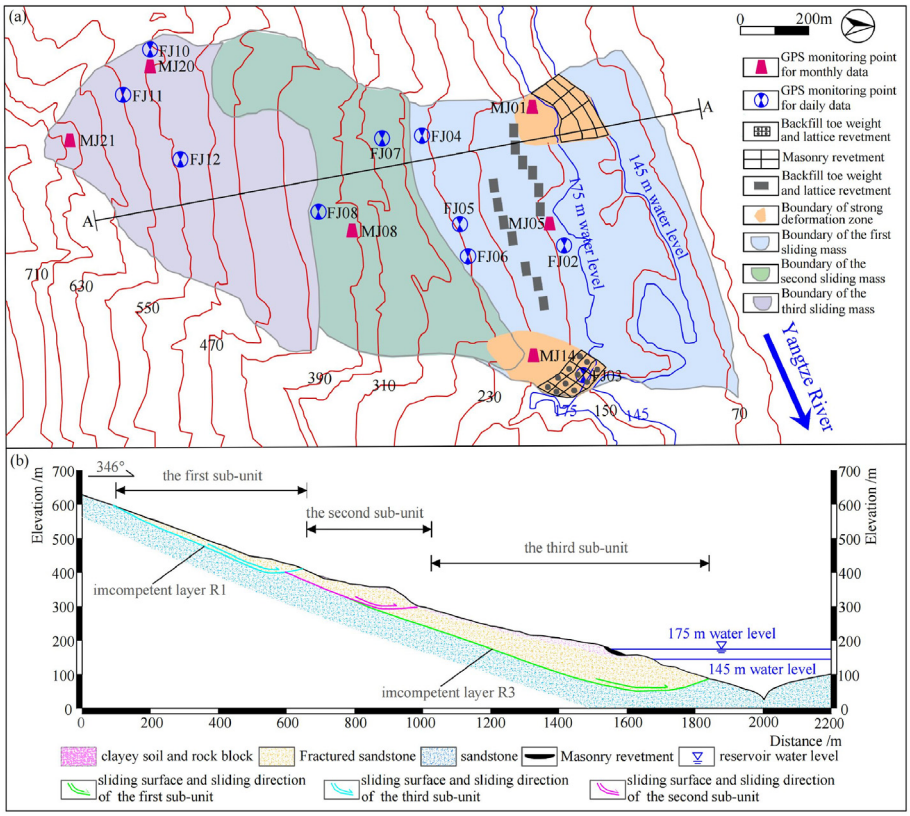

Fig. 2. (a) Topography of the Outang Landslide, and (b) Schematic cross-section *A-A* (from (a)).

These two incompetent layers become deeper from the top to the bottom of the landslide.

2.2. Field observations and monitoring system

2.2.1. Field observations

Three sub-units of the Outang landslide, indicated in Fig. 2 as the first, second, and third sub-units, display different deformation patterns. The collapse occurred at the leading edge of the first sub-unit as the water level at TGR increased to 156 m in 2003. The collapse moved backward more than 5 m compared to the deformation before the impoundment of TGR. The water level went up to 172 m in September 2008, and the deformations of the first sub-unit were mainly surface subsidence and cracks. Several small, shallow earth slides appeared, resulting in damage to houses being built and some rural roads. The second sub-unit was deformed by surface subsidence and soil sliding dispersedly on a small scale. The deformation in the third sub-unit was local soil heaves, road subsidence, and tensile cracks. The trailing edge of the third sub-unit crept forward about 6 m over the past 50 years.

Please cite this article as: Yang B et al., Spatiotemporal deformation characteristics of Outang landslide and identification of triggering factors using data mining, Journal of Rock Mechanics and Geotechnical Engineering, https://doi.org/10.1016/j.jrmge.2023.09.030

The spatiotemporal deformation characteristics of the Outang landslide have been analyzed on the basis of field observations. Time-wise, deformation appeared on the Outang landslide when the TGR initiated impoundment in 2003. Field observations between 2003 and 2012 demonstrated that over 160 cracks were caused by reservoir level change, whereof a sharp increase in crack formation (45 cracks) was observed in 2009 when the dam reservoir was first raised to 175 m (Fig. 3). The number of new cracks reduced significantly after 2012 with approximately only 20 cracks except in some zones of the landslide.

Space-wise, the majority of the cracks occurred mainly at the toe of the first sub-unit. Because of the severe deformation in 2013, a remediation project was undertaken, with backfill toe weight and lattice revetment in the eastern strong-deformation zone and masonry revetment in the western strong-deformation zone (Luo and Huang, 2020). Over the subsequent years, the landslide exhibited macroscopic deformations as displayed in Fig. 4. Besides the appearance of tension cracks on the ground (Fig. 4a and b), a local collapse occurred, which destroyed residential houses and roads (Fig. 4c). Furthermore, deformation was also observed on other sub-units of the Outang landslide (Fig. 4d). This study investigated the spatiotemporal deformation characteristics and identified the dominant factors inducing deformation. The influence of the structures placed for mitigation on the landslide movement was not taken into consideration.

2.2.2. Monitoring system

A monitoring system has been implemented on the Outang landslide and a global positions system (GPS) was included to track its displacement behavior (Fig. 2a). More than 30 automatic continuous GPS monitoring stations were installed, some measuring displacement already from December 2010. Others were installed thereafter. Several monitoring points were relocated or stopped because of construction activities. The automatic continuous GPS monitoring stations, known for their high measurement accuracies, record surface displacement with the precision of 3 mm + 1 ppm in planimetry and 5 mm + 1 ppm in altimetry. The GPS network adopts a set of single-frequency stations to transmit raw data to the GPS receiver every 20 s. The raw data recording relative positions of all measurement stations to datum stations are processed together.

Monthly and daily data at various monitoring locations of the landslide are available from the monitoring system. Herein, the monthly dataset recorded one datum per month from December 2010 to December 2016, while the daily dataset was gathered every 5–10 d from 2013/1/7 to 2016/8/7 and has since then been collected daily up to today. The two datasets could extend the monitoring time periods and cover different time scales, which benefit the analysis of landslide evolution patterns and could be a dominant basis for further risk evaluations. Details of the two datasets are presented in Tables 1 and 2.

2.3. Data mining technology

2.3.1. Two-step clustering

Clustering is a technique commonly applied in artificial intelligence and machine learning (Jain et al., 2000). It refers to the task of identifying patterns and trends of large volumes of data by grouping data points that are similar while separating those that are dissimilar (Lee and Antonsson, 2000). Multiple clustering algorithms have been proposed over the years. Some popular clustering algorithms include K-means, fuzzy clustering, and two-step clustering.

The K-means is commonly easy to implement and suitable for very large data sets, while users need to set the number of clusters in advance (Davies, 2005). Furthermore, its dependence on the initial conditions may cause the algorithm to converge to suboptimal solutions. Fuzzy clustering has an advantage over K-means that allows for overlapping clusters (Izakian and Abraham, 2011). This algorithm assigns each pattern to each cluster with some degree of membership according to the closeness of the data object to the cluster centers (Nayak et al., 2015). Unfortunately, fuzzy clustering is prone to becoming trapped in local optima due to its sensitivity to initialization.

The two-step clustering algorithm was proposed by Chiu et al. (2001). The approach is superior to more traditional techniques mainly due to: (1) it is capable to cluster mixed valued data (numerical and categorical variables); (2) it can handle large datasets efficiently; (3) it determines the optimum number of clusters automatically, following statistical standards, including Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC), rather than based on an arbitrary choice; and (4) it allows to diagnose outliers and noise data in samples. Considering these advantages and previous successful applications of the two-step clustering in landslide deformation analysis (Ma et al., 2017; Miao et al., 2021; Guo et al., 2022b), a two-step clustering algorithm is adopted in this paper.

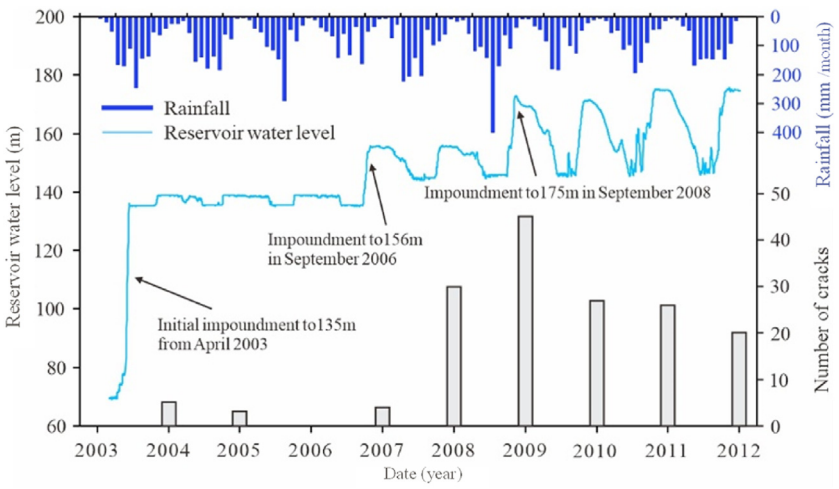

Fig. 3. Relationship between the number of cracks, reservoir water level, and rainfall versus time.

Please cite this article as: Yang B et al., Spatiotemporal deformation characteristics of Outang landslide and identification of triggering factors using data mining, Journal of Rock Mechanics and Geotechnical Engineering, https://doi.org/10.1016/j.jrmge.2023.09.030

Fig. 4. Deformation on the Outang landslide: (a) Tension crack observed on the first sub-unit in 2017, (b) Crack observed on the east strong deformation zone in 2019, (c) Resident house and road destroyed by a local collapse in 2018, and (d) Destroyed drain on the second sub-unit in 2014 (Guo et al., 2020).

The two-step clustering first uses a distance measure to separate groups and then adds a probabilistic analysis to select the optimal subgroup model. In the case where a dataset consists of uniquely numerical variables, Euclidean distance is typically set as a distance measure criterion for clustering. In the case of a dataset containing numerical and categorical variables, log-likelihood is to be selected in the clustering procedure.

The technique is realized in two steps, pre-clustering and clustering (Fig. 5). In the first step, the cases are pre-clustered by a sequential method according to the definition of dense regions in the analyzed attribute space. The number of clusters increases in the pre-clustering process. In the second clustering step, pre-clustering performs statistical merging in a step-by-step manner until all clusters are in one cluster (Benassi et al., 2020).

2.3.2. Apriori algorithm

The Apriori algorithm provides the capacity for searching for frequent itemsets and generating association rules (Agrawal et al., 1993). The minimum support (*minsupp*) and minimum confidence (*minconf*) are fixed to select useful association rules from the set of all possible rules. Fig. 6 describes the implementation process of the Apriori algorithm.

2.3.2.1. Searching for frequent itemsets. A frequent itemset (T) represents a dataset containing an item (a) that has greater than or equal support compared to the minimum support stated by users. In the case where an itemset covers k frequent items, it can be named a frequent k itemset and recorded as L_k.

$$\frac{|T(a)|}{T} \geq minsupp \tag{1}$$

The Apriori algorithm is based on the a priori data property that all subsets of a frequent itemset are always frequent (Abdullah et al., 2008). The principle means that when a k-itemset is not frequent, it is not employed to generate candidates of $k+1$ itemset. This is a critical principle in the process of relation rules generation using the Apriori algorithm. Such an algorithm identifies frequent itemsets according to a layer-by-layer search based on an iteration method.

If we suppose a database marked as D, we can set the support thresholds to n in this example, meaning that one itemset cannot be frequent if its support thresholds are less than n. In addition, C_k and L_k represent the candidate k itemset and frequent k itemset, respectively. The Apriori algorithm is conducted by the following steps: (1) It scans the dataset D and generates candidate itemset C_k, where k starts from unity. (2) The support of each itemset in C_k is calculated and frequent itemset is generated by deleting all the itemsets that the supports are below the threshold. (3) The elements in the frequent itemset L_k are recombined to generate the candidate $k+1$ itemset C_{k+1}. (4) Iteration (Steps 2 and 3) continues until it is unable to generate further frequent itemsets.

In the example shown in Fig. 6, the candidate itemset C_1 is generated by scanning dataset D. However, the support of the element $\{d\}$ in C_1 is below the threshold n and is thus deleted. A frequent itemset L_1 is generated, and it is recombined to give a new candidate itemset C_2. Likewise, the supports of $\{a, b\}$ and $\{a, e\}$ are also below the threshold, so they are removed. A frequent itemset L_2 is obtained after deleting these two itemsets. We can then form a candidate itemset C_3, which contains only one element $\{b, c, e\}$. Since there are no more frequent itemsets that can be generated, the procedure stops at this point.

2.3.2.2. Generating association rules. Association rules are finally obtained from these frequent itemsets by finding all the subsets. Confidence is commonly used to measure the strength of association rules, which is defined as the conditional probability of the consequence of the rule given its antecedent (Borgelt and Kruse, 2002). Among all these rules, valid rules need to be selected according to a criterion expressed as whether the confidence level is above the confidence threshold (*minconf*). In the example in Fig. 6, the frequent itemset L_3 generates 6 rules. Rules 1 and 5 are not useful as their confidence is below the threshold m. If that is the

Please cite this article as: Yang B et al., Spatiotemporal deformation characteristics of Outang landslide and identification of triggering factors using data mining, Journal of Rock Mechanics and Geotechnical Engineering, https://doi.org/10.1016/j.jrmge.2023.09.030

Table 1
Cumulative displacements of monthly monitoring sites.

Location	Monitoring point	Monitoring period	Cumulative displacement (mm)	Average annual displacement rate (mm/year)
Western strong-deformation zone	MJ01	2010/12 −2016/12	686	113
Eastern strong-deformation zone	MJ14	2011/8 −2016/12	426	79
First sub-unit	MJ05	2010/12 −2016/12	234	39
	MJ06	2010/12 −2016/12	214	35
Second sub-unit	MJ07	2010/12 −2016/12	409	67
	MJ08	2010/12 −2016/12	398	65
Third sub-unit	MJ20	2012/5 −2016/12	630	135
	MJ21	2012/5 −2016/12	555	119

Table 2
Cumulative displacements of daily monitoring sites.

Location	Monitoring point	Monitoring period	Cumulative displacement (mm)	Average annual displacement rate (mm/year)
First sub-unit	FJ02	2013/1/7 −2022/3/8	311	34
	FJ03	2013/1/7 −2022/3/8	305	33
	FJ04	2015/7/14 −2022/3/8	311	47
	FJ05	2016/4/24 −2022/3/8	428	73
	FJ06	2013/4/11 −2022/3/8	480	54
Second sub-unit	FJ07	2013/4/11 −2022/3/8	506	57
	FJ08	2015/7/14 −2022/3/8	646	97
Third sub-unit	FJ10	2013/5/22 −2022/3/8	1464	166
	FJ11	2013/5/22 −2022/3/8	1143	130
	FJ12	2013/5/22 −2022/3/8	800	91

case for a certain frequent itemset L, the association criterion can be generated: $L' \rightarrow (L\text{-}L')$.

$$C_{L' \rightarrow (L-L')} = \frac{|T(L)|}{|T(L')|} \geq minconf \tag{2}$$

2.3.3. Data mining process

The data mining process for the Outang landslide consists of the three steps illustrated in Fig. 7.

2.3.3.1. Preparation of data for mining. To begin the data mining process, the first step in the data mining process involves collecting data on a variety of parameters, such as rainfall, reservoir water level, surface displacement, and so on. The original monitoring data undergoes pre-processing by techniques of reviewing, filtering, and sorting to ensure accuracy and completeness. Additionally, some related factors including changes in reservoir levels and cumulative rainfall over a certain period are required to be provided, which can help to identify potential triggers for landslides.

2.3.3.2. Grouping monitoring data into clusters based on numerical values. The Apriori algorithm handles only categorical variables and not numerical ones. Therefore, it is required to cluster the numerical monitoring data into groups before applying the Apriori algorithm to mine the relationship between triggers and movement in the Outang landslide case. To achieve this, the two-step clustering algorithm is adopted to group reservoir level, rainfall, and landslide movement rate into several categories.

2.3.3.3. Exploration of association rules. By using the Apriori algorithm with fixed values for the support threshold (*minsupp*) and confidence threshold (*minconf*), useful association rules are obtained. The association rules between deformation clusters and triggers clusters are crucial for identifying dominant factors triggering landslide deformation. Further explanations for association rules can be found in Section 4.2.

3. Deformation characteristics of outang landslide

3.1. Deformation characteristics of outang landslide based on monthly monitoring data

3.1.1. Spatial deformation characteristics

More than 30 monitoring points with monthly data are distributed on the three sub-units of the landslide including two strong deformation zones. In this study, we select several monitoring stations across the landslide for detailed analysis. The monitoring point recorded as MJ01 is located in the western strong-deformation zone, while MJ14 is located in the eastern strong-deformation zone. MJ05 and MJ06, MJ07 and MJ08, and MJ20 and MJ21 are representatives of the displacements of the first, second, and third sub-units, respectively. Cumulative displacements of these monitoring sites with reservoir water level and rainfall are shown in Fig. 8.

Outang landslide has experienced noticeable deformation. In the first sub-unit, the western strong-deformation zone (MJ01)

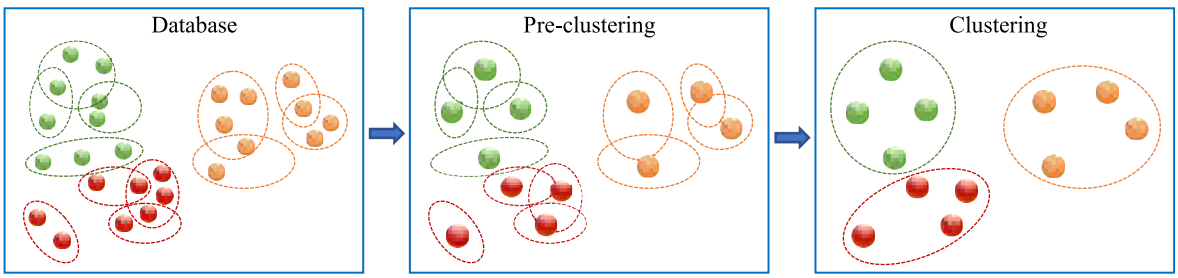

Fig. 5. Implementation process of two-step clustering (He et al., 2022).

Please cite this article as: Yang B et al., Spatiotemporal deformation characteristics of Outang landslide and identification of triggering factors using data mining, Journal of Rock Mechanics and Geotechnical Engineering, https://doi.org/10.1016/j.jrmge.2023.09.030

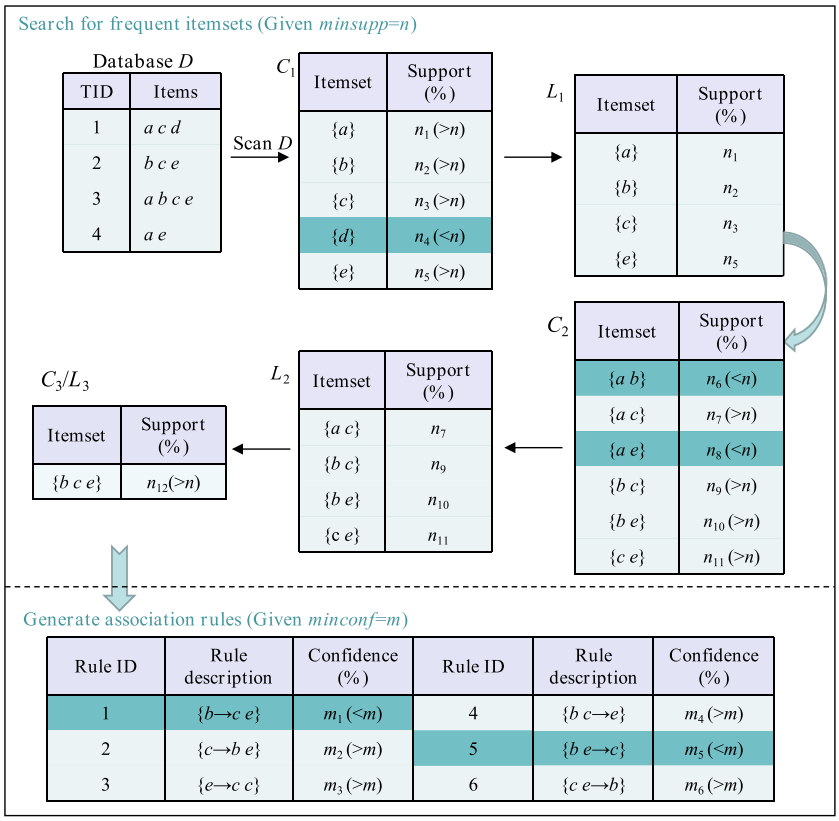

Fig. 6. The algorithm implementation process of Apriori.

exhibits a maximum displacement value of 686.3 mm. The cumulative displacement of the first sub-unit (MJ05 and MJ06) is slightly above 200 mm, representing the lowest displacement among the three sub-units. The second sub-unit (MJ07 and MJ08) has a displacement of about 400 mm, while the third sub-unit (MJ20 and MJ21) has a deformation exceeding 550 mm. From the spatial characteristics, the soil mass of the Outang landslide deforms more intensely from the toe to the trailing edge, except for two strong-deformation zones.

3.1.2. Temporal deformation characteristics

During the monitoring period, displacement is available each month for MJ01, MJ05, MJ08, and MJ21. These displacements are used to analyze the characteristics of landslide deformation with time. Monitoring point MJ01, located in the western strong-deformation zone, had an annual displacement of 121 mm in 2011 and reached a peak in 2012 with a value of 177 mm per year (Fig. 9). A significant reduction in annual displacement occurred in 2013 (76 mm) and 2014 (70 mm), However, larger increment of displacements occurred in the following years. Annual displacements were up to 102 mm and 140 mm in 2015 and 2016, respectively.

MJ05 is located at the toe of the first sub-unit. The maximum (57 mm) and minimum (18 mm) deformation appeared in 2011 and 2013, respectively. The annual displacement in other years was about 40 mm with only small fluctuations. MJ08, representing the second sub-unit, displayed a larger fluctuation of annual displacement than MJ05. The displacement increased by about 70 mm per year in 2011 and 2012. The lowest annual displacement occurred in 2013. Unexpectedly, a larger increment of displacements occurred in 2014 with a value of 88.7 mm. The annual displacements of MJ08 decreased in 2015 and then increased in 2016. The annual displacements were 58 mm and 99 mm in 2015 and 2016, respectively. MJ21 is chosen to represent the third sub-unit. Its deformation trend between 2013 and 2016 was highly consistent with that of MJ08. The difference was that the displacement of MJ21 was significantly higher than that of MJ08. The maximum annual displacement was up to 173 mm in 2016.

3.1.3. Relationships between displacement, rainfall, and reservoir level

Cumulative displacements in Fig. 8 display a step-wise increment, meaning that the landslide deforms sharply between May and September while keeping stable in other months of the year. Fig. 10 summarizes monthly rainfall, reservoir level change rate, and monthly displacements of MJ01, MJ05, MJ08, and MJ21.

When the reservoir level decreased by 6.6 m and the slope experienced a maximum monthly rainfall of 232 mm in June 2011, the displacement of MJ01 and MJ08 had a maximum increment of 30 mm and 25 mm, respectively. The occurrence of a sharp displacement increment was synchronized with rainfall and reservoir level decline, indicating that these two factors induced landslide movement.

The May rainfall reached its highest value of the year, while the landslide deformed only slightly during that month. The monitoring points of MJ01 and MJ08 reached maximum displacement increment in June, indicating one month lag after the largest rainfall. No time lag was observed for the effect of reservoir level drawdown in this duration. The pattern of deformation response to rainfall and reservoir drawdown was interchanged in 2014: there was no time lag in the response to the rainfall, but time lag in the response to the reservoir drawdown. Another noteworthy observation was that the lag in time of the deformation response could differ across the landslide. For example, the displacement of MJ01

Please cite this article as: Yang B et al., Spatiotemporal deformation characteristics of Outang landslide and identification of triggering factors using data mining, Journal of Rock Mechanics and Geotechnical Engineering, https://doi.org/10.1016/j.jrmge.2023.09.030

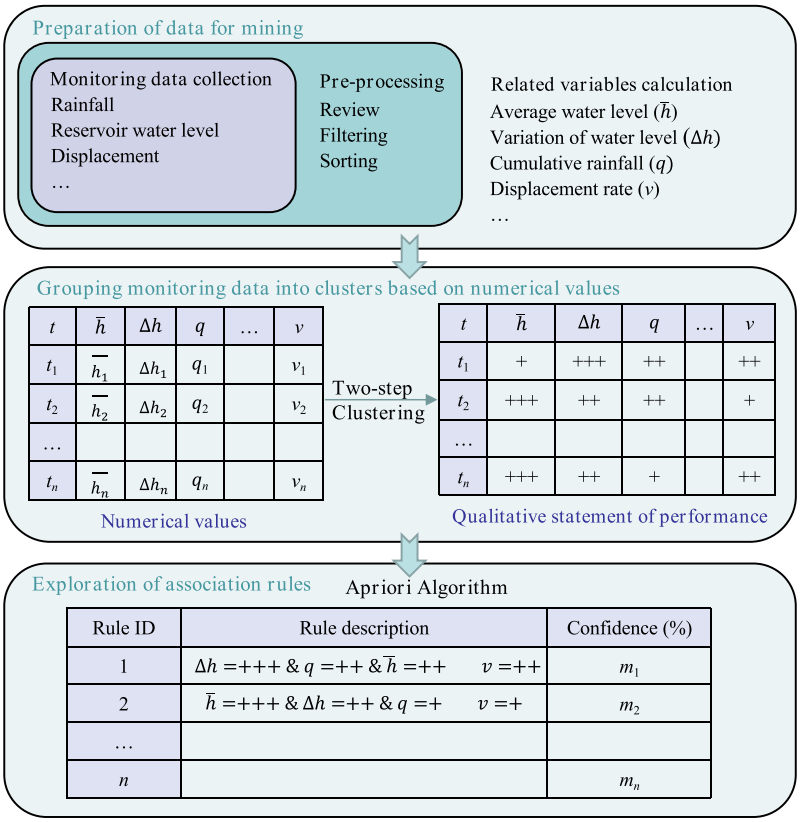

Fig. 7. Flow chart of the data mining process (+, ++, +++ represent the clustering resulting from monitoring data).

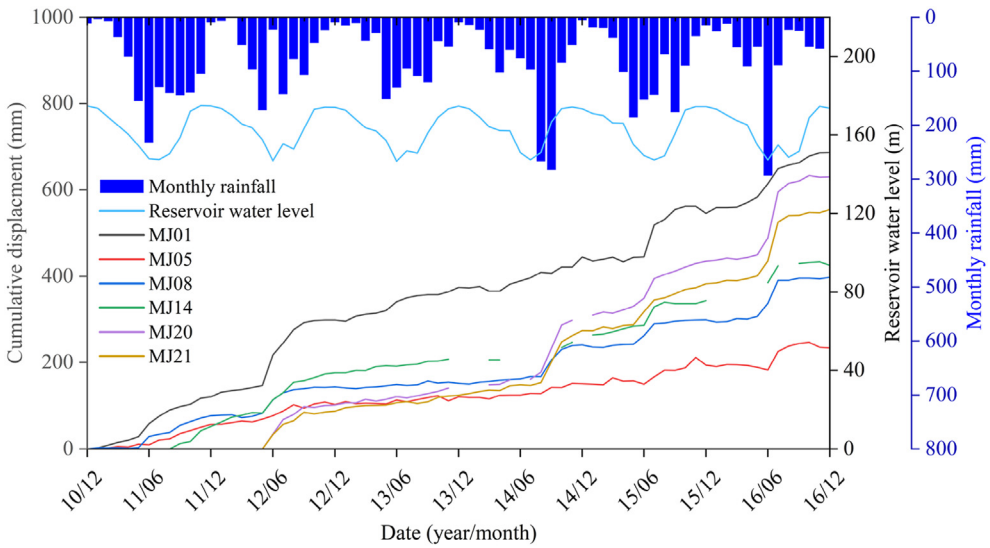

Fig. 8. Cumulative displacements of Outang landslide between 2010 and 2016.

in 2015 lagged the tune of the rainfall for two months, while the lag in time at other monitoring points was only one month.

The reservoir level has been subjected to periodic changes between 145 m and 175 m since 2009. Rapid deformation of the Outang landslide primarily occurs with reservoir drawdown. The timing of the reservoir drawdown is also often close to the time of intense rainfall. The overlapping of these two factors and their hysteresis further compounds the relationships between displacement and triggers.

3.2. Deformation characteristics of outang landslide based on daily monitoring data

3.2.1. Spatial deformation characteristics

To further explore the landslide deformation characteristics, its monitoring system was gradually improved, and 12 stations were installed to track the daily deformation of the Outang landslide. These stations are numbered FJ01 to FJ12 (Fig. 11). The data at points FJ01 and FJ09 were interrupted and are not analyzed in this

Please cite this article as: Yang B et al., Spatiotemporal deformation characteristics of Outang landslide and identification of triggering factors using data mining, Journal of Rock Mechanics and Geotechnical Engineering, https://doi.org/10.1016/j.jrmge.2023.09.030

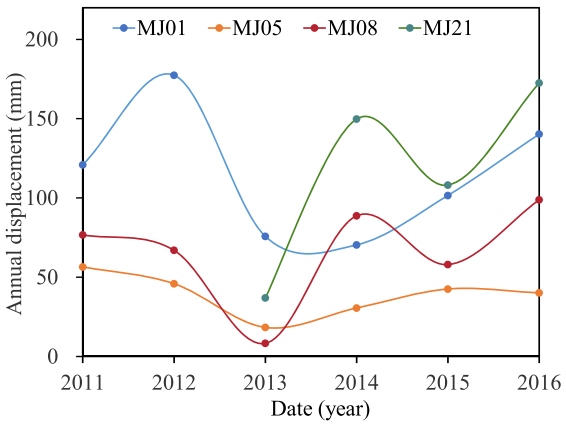

Fig. 9. Annual displacement of MJ01, MJ05, MJ08, and MJ21.

paper. The GPS monitoring stations provide deformation data across the slope. FJ02 to FJ06 are in the first sub-unit of the landslide; FJ07 and FJ08 are in the second sub-unit; and FJ10, FJ11, and FJ12 are in the third sub-unit (Fig. 2a). As of 2022/3/8, the cumulative displacements at the 10 monitoring stations decreased as follows: FJ10>FJ11>FJ12>FJ08>FJ07>FJ06>FJ05>FJ02>FJ04>FJ03. This confirms that the third sub-unit is the least stable of the three and that the first one is the most stable.

(1) In the first sub-unit, there are five monitoring points across this area. The values of cumulative displacements of FJ02, FJ03, and FJ04 are each about 310 mm. FJ05 is next to FJ06, and their cumulative displacements are greater than 420 mm.
(2) In the second sub-unit, FJ08 is located to the east of FJ07, and their cumulative displacements are 646 mm and 506 mm, respectively. The landslide deforms more intensely from west to east in this area.
(3) In the third sub-unit, FJ10 is located on the left boundary; FJ11 is to the east of FJ10; and FJ12 is in the center of the area. The three monitoring points are each at an elevation of about 560 m. However, their cumulative displacements vary from 1464 mm, 1143 mm, to 800 mm, respectively. This area of the landslide deforms more from east to west.

The monitoring results indicate that the deformation behavior differs even in the same sub-unit of the landslide, in addition to the deformation characteristics varying from one sub-unit to the other.

3.2.2. Temporal deformation characteristics

The annual displacements between 2013 and 2021 are summarized in Fig. 12, indicating that the displacements in the three sub-units of the Outang landslide follow similar trends of ups and downs. Yet, the annual displacement shows remarkable variation during the monitoring period. The annual displacements peaked in 2017 and 2021 (and the displacements may increase further in 2022/2023). The lowest annual displacement was observed in 2019. For example, the annual displacement of FJ10 in 2014 and 2015 were 155 mm and 144 mm with a difference of 11 mm. A larger deformation of 202 mm occurred in 2016. Subsequently, peaks appeared in 2017 and 2021 with the amount of 415 mm and 305 mm, which were more than 10 times the displacement in 2019.

3.2.3. Relationships between displacement, rainfall, and reservoir level

The monitoring sites marked FJ06 (in the first sub-unit), FJ08 (in the second sub-unit), and FJ10 (in the third sub-unit), with the largest deformation in each mass of the Outang landslide, were chosen to investigate relationships among landslide displacement, rainfall, and reservoir level. Rainfall and reservoir level were continuously measured on a daily basis, and partial cumulative displacement was obtained at an irregular frequency of 5–10 d. To make the data frequency consistent, daily data between 2017/1/1 and 2021/12/31 were selected. Reservoir level change rate, daily rainfall, and displacement increasing rate of FJ06, FJ08, and FJ10 are shown in Fig. 13. The landslide moved more in 2017 and 2021, and the monitoring data of these two years are analyzed herein.

The annual displacements of FJ06, FJ08, and FJ10 in 2017 were 147 mm, 252 mm, and 415 mm. Two periods of daily displacement increase were observed in 2017. Between 2017/7/10 and 2017/8/5, the maximum displacement rates of FJ06, FJ08, and FJ10 were 2.5 mm, 5.5 mm, and 8.3 mm. During this period, the maximum daily rainfall was 116 mm, and the longest consecutive rain lasted 4 d with a rainfall of 164 mm, accompanied by a drop in reservoir level. Between 2017/9/30 and 2017/10/30, the maximum daily deformation of FJ06, FJ08, and FJ10 was 5.5 mm, 14.3 mm, and 17.4 mm, respectively. During this period, the reservoir level rose to 175 m and was maintained at this level, while total rainfall was 209.5 mm. Notably, between 2017/1/1 and 2017/7/2, daily displacements of these three monitoring points were less than 3 mm/d, in spite of a maximum daily rainfall of 237.5 mm accompanied by reservoir water level dropping at a rate of 0.4 m/d. This phenomenon indicated that neither short-term abundant rainfall nor slow water level drawdown could induce intense deformation for the three sub-unites.

Compared with 2017, the deformation characteristics of 2021 were more uniform. The annual displacements of FJ06, FJ08, and FJ10 were 40 mm, 157 mm, and 305 mm, respectively. A sharp displacement increment of 113 mm occurred between 2021/7/5 and 2021/7/12, and the maximum deformation rate was 34 mm/d. The reservoir level was maintained at about 148 m, with a cumulative rainfall of 734 mm. Continued abundant rainfall may be responsible for the intense deformation of the Outang landslide in 2021.

The analysis indicates that both rainfall and reservoir level contribute to the Outang landslide movement. Because of the hysteresis and combined effects of the two hydrologic factors, it is not easy to identify the triggers controlling landslide deformation. Additionally, the spatiotemporal deformation characteristics of the Outang landslide make the problem even more complex. More efforts are required to separate the influence of rainfall and reservoir level fluctuation on landslide displacements.

4. Hydrological triggers based on data mining technology

4.1. Clustering results

4.1.1. Clustering results of monthly data

Based on the analysis in Section 3.1 and previous research (Ma et al., 2017; Miao et al., 2021; Guo et al., 2022b), six hydrologic causes are chosen to perform data mining. The deformation triggers include monthly average water level ($\overline{h}$), monthly variation of water level (Δh_1), bi-monthly variation of water level (Δh_2), monthly maximum daily drop of water level (Δh_d), monthly cumulative rainfall (q_1), and bi-monthly cumulative rainfall (q_2). The monthly displacement rate at each monitoring location is used to represent landslide movement from the toe to the trailing edge, where MJ01, MJ05, MJ08, and MJ21 are located in the western strong-deformation zone, the first sub-unit, the second sub-unit, and the third sub-unit of the Outang landslide, respectively.

The two-step clustering method is used to group both monthly velocity at each monitoring location and the six hydrological

Please cite this article as: Yang B et al., Spatiotemporal deformation characteristics of Outang landslide and identification of triggering factors using data mining, Journal of Rock Mechanics and Geotechnical Engineering, https://doi.org/10.1016/j.jrmge.2023.09.030

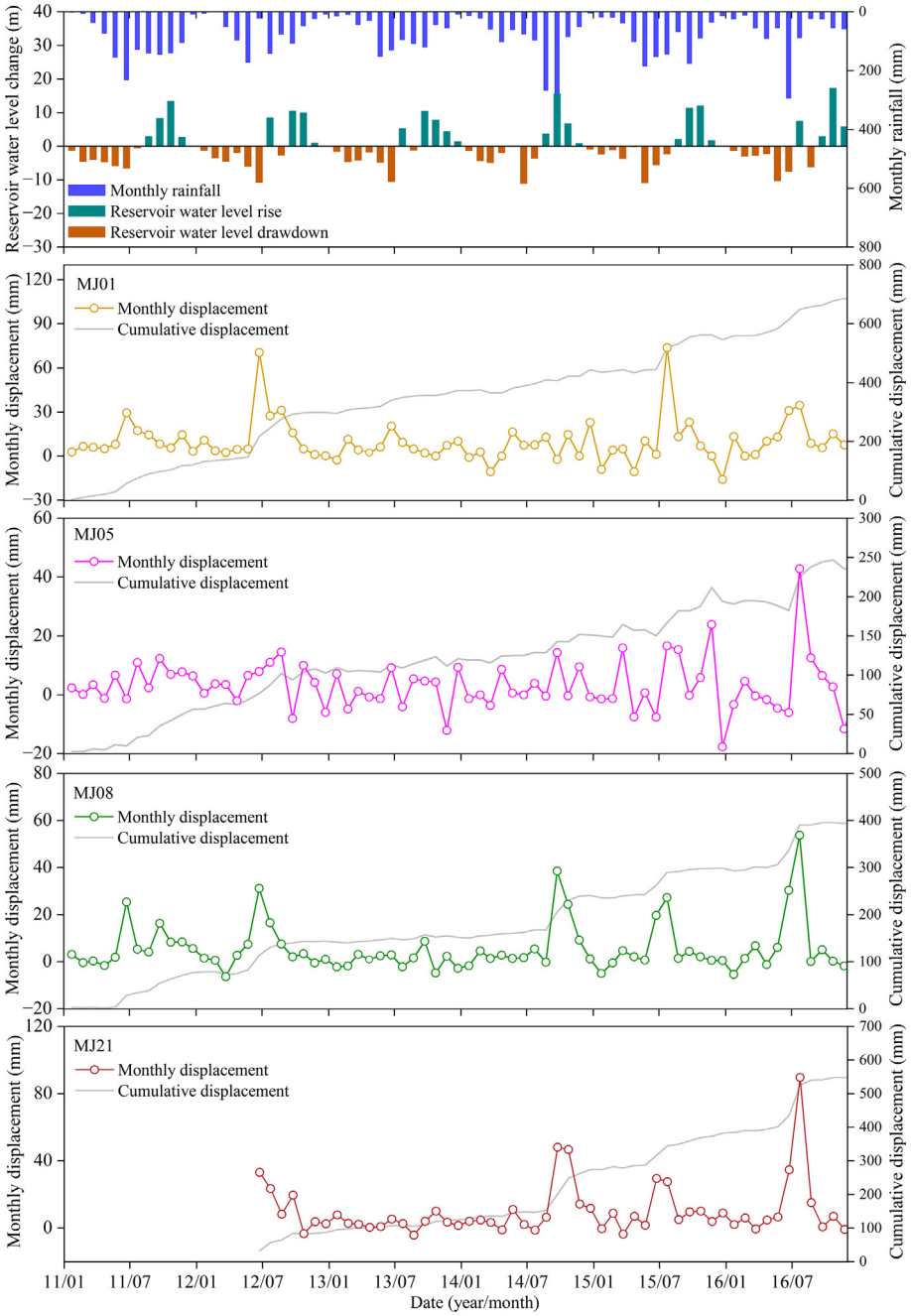

Fig. 10. Monthly rainfall, reservoir level fluctuation, and monthly displacement of MJ01, MJ05, MJ08, and MJ21.

indicators into clusters. During clustering, the number of categories is set between 2 and 10, the Euclidean distance is employed for distance measurement, and the BIC (Bayesian information criterion) is adopted as the clustering criterion. The quality of clustering algorithms can be evaluated by the average silhouette coefficient of all instances. Specifically, the silhouette coefficient quantifies the degree to which each data point within a cluster is well-matched to its own cluster, relative to other clusters. It ranges between −1 and 1, while a higher value indicates better clustering (Aranganayagi and Thangavel, 2007).

These six hydrological indicators are clustered into two to three groups, and each group is equipped with an interval of values (Table 3). The monthly velocity is clustered into two groups (V1 and V2), reflecting steady movement and the distinct movement of the Outang landslide, respectively (Table 4 and Fig. 14). Thus, the clustering group of V2 with a larger monthly velocity (15–89.6 mm/month) is mainly considered to study the inherent correlation between landslide displacement and triggers.

4.1.2. Clustering results of daily data

According to the analysis in Section 3.2 and some references (Yao et al., 2019; Guo et al., 2022b), twelve candidate triggering factors are selected, covering two catalogs. The reservoir level catalog includes eight parameters: daily reservoir level ($\bar{h}^{\mathrm{day}}$), daily variation of water level ($\Delta h_1^{\mathrm{day}}$), 5 d variation of water level ($\Delta h_5^{\mathrm{day}}$), 10 d variation of water level ($\Delta h_{10}^{\mathrm{day}}$), 30 d variation of water level ($\Delta h_{30}^{\mathrm{day}}$), 5 d cumulative drop of water level ($\Delta h_{\mathrm{d-5}}^{\mathrm{day}}$), 10 d cumulative drop of water level ($\Delta h_{\mathrm{d-10}}^{\mathrm{day}}$), and 30 d cumulative drop of water

Please cite this article as: Yang B et al., Spatiotemporal deformation characteristics of Outang landslide and identification of triggering factors using data mining, Journal of Rock Mechanics and Geotechnical Engineering, https://doi.org/10.1016/j.jrmge.2023.09.030

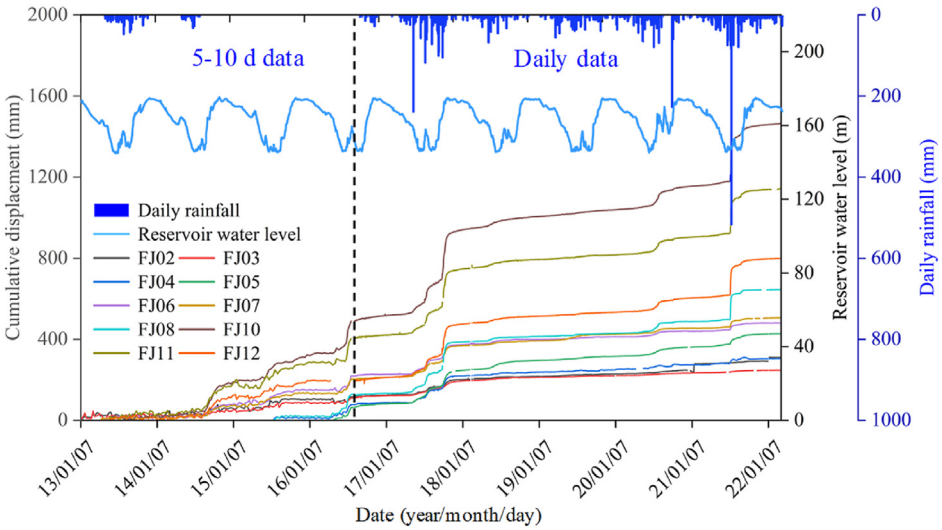

Fig. 11. Cumulative displacements of monitoring sites.

level ($\Delta h_{d\_30}^{day}$). The rainfall catalog includes four parameters: daily rainfall (q_1^{day}), 5 d cumulative rainfall (q_5^{day}), 10 d cumulative rainfall (q_{10}^{day}), and 30 d cumulative rainfall (q_{30}^{day}). The monitoring points FJ06, FJ08, and FJ10 are in the first, second, and third sub-unites and are implemented by data mining. The clustering results of the twelve hydrological indicators are displayed in Table 5. Daily displacement is clustered into two groups, represented as P1 with a low rate of displacement (−6.4–0.4 mm/d) and P2 with a high rate (0.39–44.98 mm/d), as shown in Table 6 and Fig. 15.

4.2. Association criteria mining

4.2.1. Association criteria mining results based on monthly data

The Apriori algorithm is applied to explore the inherent correlation criteria within the monitoring data by setting the confidence threshold to 85% and the support threshold to 1.5% (Ma et al., 2017; Guo et al., 2022b). Six hydrologic factors ($\bar{h}$, Δh_1, Δh_2, Δh_d, q_1, and q_2) and landslide movement rate (v) are programmed as the former item and the consequent item of association rules. Hundreds of association rules are generated. The hydrological causes of landslide movement at stage V2 with large deformation (15–89.6 mm/month) are explored. A few of the typical association rules at these four monitoring stations with varied locations are screened and listed in Fig. 16.

4.2.1.1. The association rules of MJ01. Rules A1 to A4 can be explained by the fact that when the reservoir level fluctuates sharply at the rate of 1.8–5.6 m/month, the landslide tends to pass to the V2 state. Moreover, severe deformation is more likely to appear during periods of low water level (145.55–157.15 m) or drawdown (0.8–1.64 m). These rules include a condition of "slowly-variation" of bi-monthly variation of water level (Δh_2), meaning that this factor does not contribute to severe deformation. It is reasonable to conclude that there is no time lag between reservoir level fluctuation and landslide monthly movement in this sub-unit. Except for Rule A2 containing the "light-cumulative-rainfall" condition, Rule A1, Rule A3, and Rule A4 do not involve rainfall, which indicates that a sharp change in the reservoir water level alone, without rainfall, can trigger significant deformation in this sun-unit.

Rules A5-A10 mean that heavy rainfall (monthly/bi-monthly cumulative rainfall) plays a leading role in inducing landslide movement. The appearance of heavy bi-monthly cumulative rainfall in association rules suggests that there may be a one-month time lag between rainfall and landslide movement. Notably, Rules A5 to A10 each includes a condition of reservoir water level ($\bar{h}$ or h_d) in addition to heavy cumulative rainfall. One can conclude that under intense rainfall, a "medium-drop" of reservoir level by 0.8–1.64 m/month is required for deformation to occur.

4.2.1.2. The association rules of MJ05. The monitoring site MJ05 is in the first sub-unit with an elevation of 185 m, and its association rules reveal a situation similar to MJ01. Rules B1 to B6 display that if the maximum daily drop of water level (Δh_d) is between 2.78 m and 4.12 m, the landslide tends to deform significantly (V2). This phenomenon reveals that the dominant factor inducing landslide large deformation is a sharp reservoir level drawdown.

Rules B7 to B10 are interpreted such that when cumulative rainfall in the previous two months (q_2) achieves an intense (heavy) state, the synergy of other reservoir water can exacerbate landslide deformation. Alternatively, a sharp drop in reservoir water level alone, or intense rainfall accompanied by reservoir water level change, has the potential to cause dramatic deformation in the first sub-unit.

4.2.1.3. The association rules of MJ08. The monitoring point MJ08 is at an elevation of 400 m and is selected as representative of the second sub-unit. The first factor of the antecedents in Rules C1–C10 covers heavy cumulative rainfall (q_1 or q_2), emphasizing the primary role of rainfall in provoking significant deformation in the second mass. Notably, the reservoir water level factors appear in these association rules even though this sub-unit is far away from the river. This may result from the effect of deformation of the leading edge, which is induced by reservoir level change.

4.2.1.4. The association rules of MJ21. Monitoring point MJ21 is situated in the third sub-unit and its elevation is 700 m. The correlation criteria of MJ21 reflect a comparable situation with that of MJ08 and demonstrate the crucial role of rainfall in the deformation of the third sub-unit. Additionally, some reservoir water level factors, such as monthly/bi-monthly water level and monthly average water level, are involved in the deformation of the third sub-unit.

Generally, the longitudinal observation of the correlation criteria of MJ01, MJ05, MJ08, and MJ21 indicates that the

Please cite this article as: Yang B et al., Spatiotemporal deformation characteristics of Outang landslide and identification of triggering factors using data mining, Journal of Rock Mechanics and Geotechnical Engineering, https://doi.org/10.1016/j.jrmge.2023.09.030

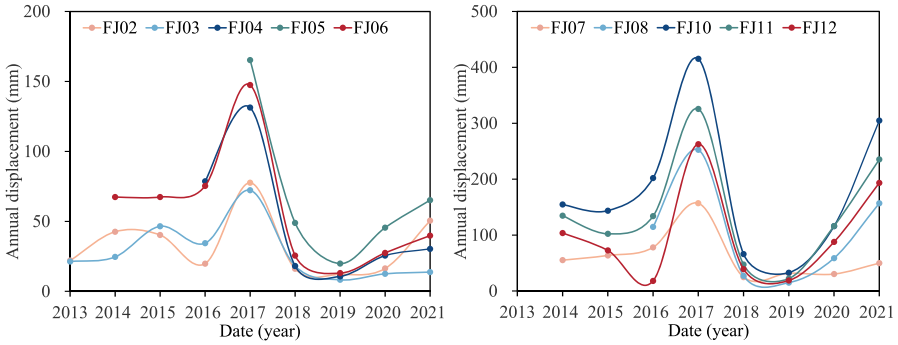

Fig. 12. Annual displacements at monitoring sites.

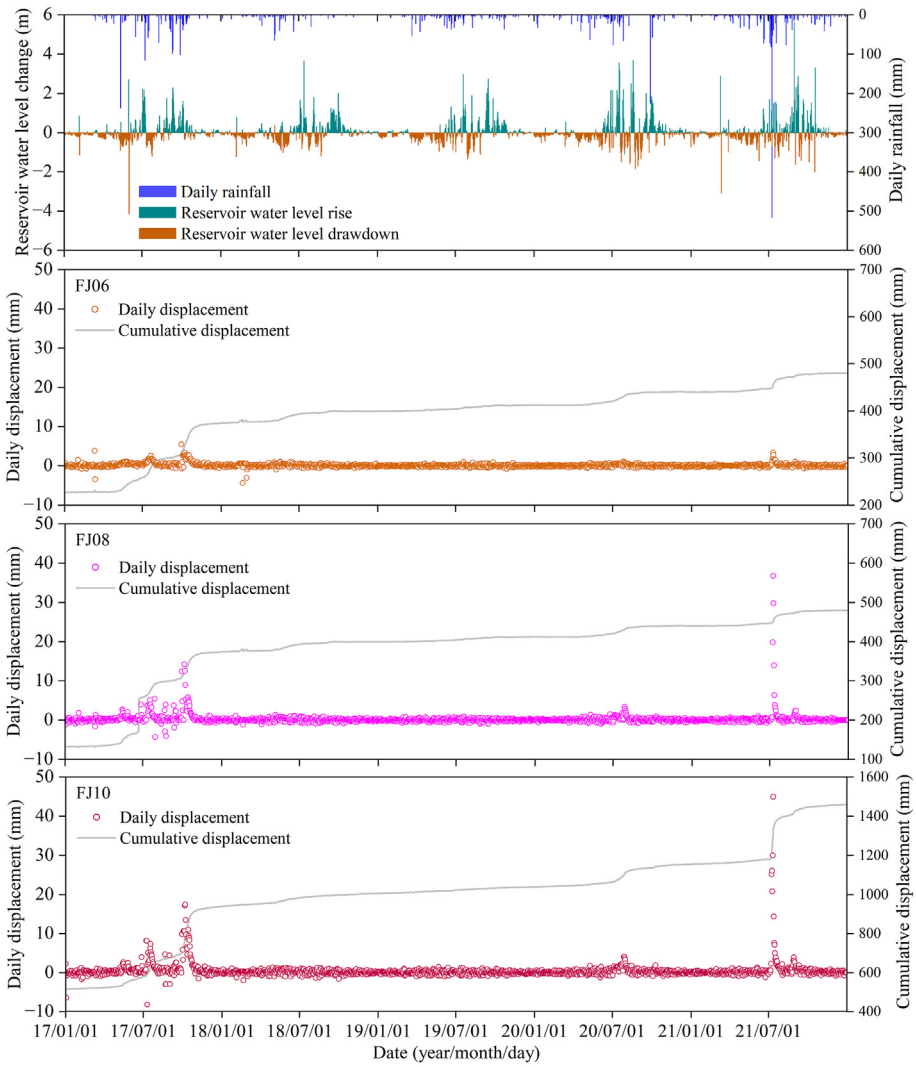

Fig. 13. Displacement rate, daily rainfall, and reservoir water level change of FJ06, FJ08, and FJ10 between 2017 and 2021.

dominant triggering factors of the first, second, and third sub-units of the Outang landslide vary. For the first sub-unit, both rainfall and reservoir level control landslide deformation. However, a sharp drop in reservoir water level alone is capable of aggravating deformation, whereas heavy rainfall requires a concomitant variation in the water level to cause significant deformation. The second and third sub-units have deformation driven primarily by the cumulative monthly rainfall in the past one or two months. Additionally, the reservoir water level also contributes to the deformation of the middle and rear parts of the Outang landslide, through its influence is mainly on the movement of the leading part. This condition may originate from the interaction of deformation among these three sub-units.

Please cite this article as: Yang B et al., Spatiotemporal deformation characteristics of Outang landslide and identification of triggering factors using data mining, Journal of Rock Mechanics and Geotechnical Engineering, https://doi.org/10.1016/j.jrmge.2023.09.030

Table 3
Clustering results of triggering factors for monthly data.

Category	Factor	Clustering result	Qualitative value	Count
Reservoir level	$\overline{h}$ (m)	167.97–174.92	High-water-level	29
		157.61–167.39	Medium-water-level	19
		146.63–157.18	Low-water-level	23
	Δh_1 (m)	0–6.78	Slowly-variation	53
		7.48–17.24	Sharply-variation	18
	Δh_2 (m)	0.02–9.23	Slowly-variation	47
		10.54–23.37	Sharply-variation	24
	Δh_d (m)	0–0.72	Slowly-drop	47
		0.8–1.64	Medium-drop	22
		2.78–4.12	Sharply-drop	2
Rainfall	q_1 (mm)	0–155.1	Light-cumulative-rainfall	64
		171.6–293	Heavy-cumulative-rainfall	7
	q_2 (mm)	6–347.5	Light-cumulative-rainfall	65
		360.4–549.1	Heavy-cumulative-rainfall	6

4.2.2. Association criteria mining results based on daily data

Twelve hydrologic factors ($\overline{h}^{\text{day}}$, Δh_1^{day}, Δh_5^{day}, $\Delta h_{10}^{\text{day}}$, $\Delta h_{30}^{\text{day}}$, $\Delta h_{\text{d}-5}^{\text{day}}$, $\Delta h_{\text{d}-10}^{\text{day}}$, $\Delta h_{\text{d}-30}^{\text{day}}$, q_1^{day}, q_5^{day}, q_{10}^{day}, q_{30}^{day}) are treated as antecedents, while the daily displacement rate is set as the consequent in the rule generator. The association rules for three daily monitoring stations located in distinct areas of the landslide are displayed in Fig. 17.

4.2.2.1. The association rules of FJ06. The monitoring point FJ06 lies in the first sub-unit. Due to the small number of the daily displacement rate of FJ06 belonging to the P2 category, merely four association rules are mined. Rules E1 to E4 suggest that P2 with a high daily displacement rate is prone to appear when the 5-d cumulative drop of reservoir water level ($\Delta h_{\text{d}-5}^{\text{day}}$) ranges from 4.69 m to 9.94 m (sharply-variation), and the 30 d cumulative rainfall (q_{30}^{day}) is in the interval of 135.5–344.4 mm (medium-cumulative-rainfall). It can be concluded that a continuous reservoir level sharp drawdown and prolonged heavy rainfall contribute to aggravating landslide deformation in this sub-unit.

4.2.2.2. The association rules of FJ08. As many as 338 association rules are acquired for FJ08 and ten of them are presented in Fig. 17. Rules F1 to F7 indicate that when cumulative rainfall over 10 d (q_{10}^{day}) reaches a medium state (88–314 mm), the landslide tends to deform more severely. Some factors of reservoir water level (Δh_5^{day} and $\Delta h_{\text{d}-30}^{\text{day}}$) appear in these rules but the water level variations are in a slow state. Rules F8 to F10 disclose that the cumulative drop of reservoir level over 5 d ($\Delta h_{\text{d}-5}^{\text{day}}$) under a sharp reservoir level variation condition can be responsible for the landslide entering the P2 stage. Moreover, the medium cumulative rainfall over 30 d (134–314 mm) plays a role in the movement of the Outang landslide. Alternately, medium rainfall alone ($\Delta h_{10}^{\text{day}}$) or a sharp drop in the reservoir level ($\Delta h_{\text{d}-5}^{\text{day}}$) accompanied by medium cumulative rainfall (q_{30}^{day}) has the potential to initiate severe deformation in the second sub-unit.

4.2.2.3. The association rules of FJ10. The monitoring point FJ10 is situated on the third sub-unit. Ten association rules were selected from over 300 association rules. Among these ten rules, the first 9 rules are related to reservoir water level fluctuation, while the 10th rule is related to rainfall. Rules G1 to G4 indicate that when the 5- or 10-d variation of reservoir water level (Δh_5^{day} and $\Delta h_{10}^{\text{day}}$) is in a sharp variation state or the 30-d variation ($\Delta h_{30}^{\text{day}}$) is in a medium variation state, the landslide could move rapidly. Rules G5 to G9 indicate that a sharp variation of 5- or 10-d cumulative drop in the water reservoir level can induce rapid landslide deformation. In addition, a 30-d cumulative rainfall with the state of "medium-cumulative-rainfall" (q_{30}^{day}) appears in Rules G1 to G9, exhibiting its effect on landslide deformation. Rule G10 has the single factor of rainfall, meaning that continued heavy rainfall over 30 d with 714–1020 mm of precipitation can lead to the landslide into the P2 stage.

The data mining results from the monitoring data on a daily time scale suggest that the leading edge of the landslide deforms significantly under the combined influences of reservoir water level fluctuation and rainfall. The medium cumulative rainfall in 10 d or heavy rainfall in 30 d alone can induce severe deformation of the middle and rear parts of the landslide. Additionally, a medium cumulative rainfall over 30 d is required if the reservoir water level tends to cause deformations in these two parts. Notably, landslide displacement cannot increase immediately with changes in the reservoir water level change and rainfall. It may take 5 d for the drop in water level to trigger landslide deformation and the time interval becomes longer as the distance to the dam reservoir increases from the toe to the trailing edge. A similar situation is observed for rainfall, with the required duration of the rainfall inducing a severe increase in the landslide displacement is 10 d.

5. Discussion

5.1. Spatiotemporal deformation characteristics

Based on the landslide field observations and surficial movement monitoring, the deformation of the Outang landslide varies both temporally and spatially. Temporally, a deformation peak can occur every few years and this phenomenon may relate to the change in external conditions and landslide structure (Zhang et al., 2020). Spatially, the deformation characteristics vary in three subzones of the Outang landslide, as well as within each subzone. Deformation was observed initially in the leading edge while the monitored displacement increased from the toe to the trailing edge thereafter. Furthermore, landslide deformation varies in the same subzone. For example, the west side deforms more severely than the east side in the second mass and the opposite is observed for the third sub-unit.

Spatial variability of deformation can cause localized sliding and result in damage to residents and facilities (Wang et al., 2015; Yao et al., 2019). It is therefore important to distinguish the relationships between the movement of each sub-unit and the triggering factors. The conclusions from this can be beneficial for zoning warnings and risk management of landslides.

5.2. Impact of monitoring data frequency on trigger identification

The monitoring system at the Outang landslide provides both monthly and daily data. These data have been utilized to investigate factors inducing landslide deformation through data mining techniques. The study indicates that different conclusions are drawn if data with different monitored time scales are used. For illustration, MJ08 and FJ08 are two representative monitoring points at the Outang landslide, which are adjacent to each other in the second sub-unit (Fig. 2a). Mining results with association rules for MJ08 indicate that monthly (q_1) and bi-monthly (q_2) cumulative rainfall are significant factors among the six candidate hydrological factors (Fig. 16). On the other hand, for FJ08 with daily data, the prominent factors identified are the 10-d cumulative rainfall (q_{10}^{day}) and 5-d cumulative drop of reservoir water level ($\Delta h_{\text{d}-5}^{\text{day}}$) (Fig. 17). A monitoring system with higher monitoring frequency should therefore be used for a more effective early warning system.

In addition, the relationships between landslide deformation and influencing factors may change because of other factors. It is difficult, maybe unrealistic at this time, to use all the data to research cause-and-effect relationships (Yang et al., 2019).

Please cite this article as: Yang B et al., Spatiotemporal deformation characteristics of Outang landslide and identification of triggering factors using data mining, Journal of Rock Mechanics and Geotechnical Engineering, https://doi.org/10.1016/j.jrmge.2023.09.030

Table 4
Clustering results of the monthly velocity.

Monthly velocity, v (mm/month)	Clustering result
−17.7–14.6	V1
15–89.6	V2

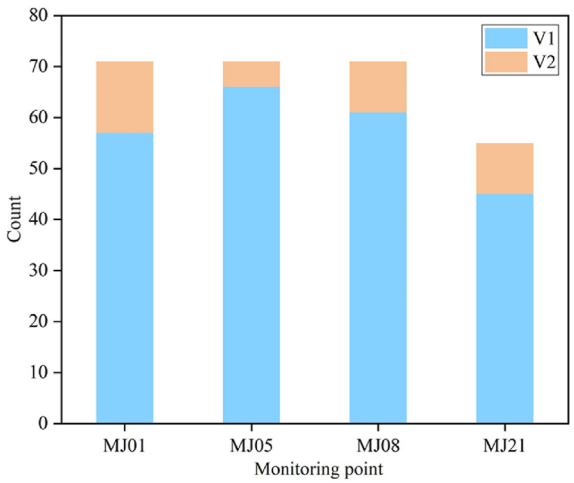

Fig. 14. Clustering results of the monitoring sites with monthly data.

Table 5
Clustering results of triggering factors for daily data.

Category	Factor	Clustering result	Qualitative value	Count
Reservoir water level	$\bar{h}^{day}$ (m/d)	168.27 −175.43	High-water-level	588
		157.21 −168.24	Medium-water-level	482
		145.55 −157.15	Low-water-level	761
	Δh_1^{day} (m)	0–0.58	Slowly-variation	1516
		0.59–1.75	Medium-variation	272
		1.8–5.63	Sharply-variation	43
	Δh_5^{day} (m)	0–4.6	Slowly-variation	1746
		4.69–9.94	Sharply-variation	85
	Δh_{10}^{day} (m)	0–5.71	Slowly-variation	1685
		5.72–16.05	Sharply-variation	176
	Δh_{30}^{day} (m)	0–6.96	Slowly-variation	1296
		6.97–14.41	Medium-variation	413
		14.44–22.17	Sharply-variation	122
	Δh_{d-5}^{day} (m)	0–2.78	Slowly-variation	1727
		2.79–5.96	Sharply-variation	104
	Δh_{d-10}^{day} (m)	0–4.7	Slowly-variation	1706
		4.77–9.41	Sharply-variation	125
	Δh_{d-30}^{day} (m)	0–5.89	Slowly-variation	1187
		5.91–11.81	Medium-variation	544
		11.85–18.05	Sharply-variation	100
Rainfall	q_1^{day} (mm)	0–116	Light-daily-rainfall	263
		226–516.9	Heavy-daily-rainfall	1568
	q_5^{day} (mm)	0–47	Light-cumulative-rainfall	1700
		48–240	Medium-cumulative-rainfall	126
		577.4–658	Heavy-cumulative-rainfall	5
	q_{10}^{day} (mm)	0–85.3	Light-cumulative-rainfall	1718
		88.1–314.1	Medium-cumulative-rainfall	102
		616.8–771.8	Heavy-cumulative-rainfall	10
	q_{30}^{day} (mm)	0–134	Light-cumulative-rainfall	1507
		135.5–344.4	Medium-cumulative-rainfall	294
		714.1–1019.7	Heavy-cumulative-rainfall	30

Table 6
Clustering results of the daily velocity.

Daily velocity, v (mm/d)	Clustering result
−6.4–0.4	P1
0.4–45	P2

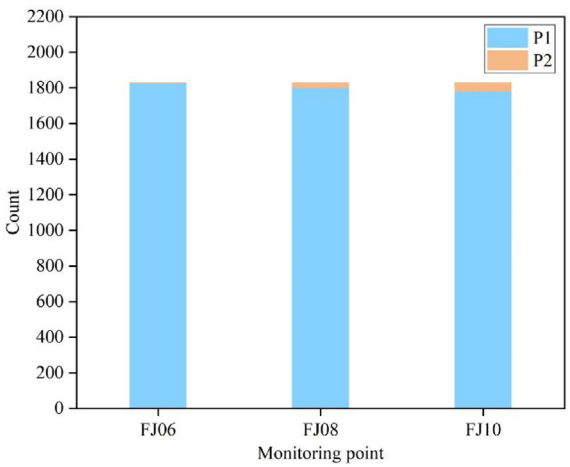

Fig. 15. Clustering results of the monitoring sites with daily data.

5.3. Future work

Previous research finds that there is a difference in landslide deformation resulting from the mechanism and evolution process of a landslide. The sliding material on the west side of the second sub-unit is more fractured than on the east side (Luo and Huang, 2020). Besides the difference in local inherent conditions in the different parts of the landslide and the interaction of deformation between the sub-units, another factor that can contribute to the deformation variability is the response of each sub-unit to external factors. To better understand this variability in landslide deformation, further investigation is needed to determine why different parts of landslides exhibit different responses to external factors.

6. Conclusions

The Outang landslide has been monitored through monthly and daily measurements for several years. The deformation characteristics of the Outang landslide are studied based on the latest monitoring data and site investigations. It is found that landslide deformation varies spatiotemporally due to mainly the different local responses to hydrological parameters (reservoir level change and rainfall). The spatiotemporal characteristics of landslide deformation led to the challenge of finding the causal relationship between hydrological indicators and landslide movement.

A data mining approach, combining the two-step cluster method and the Apriori algorithm, is adopted to identify the causal factor of the Outang landslide movement in the TGRA. The analysis looks into different locations within the Outang landslide, using monitoring data taken at different intervals: “long-term” deformation (monthly) and “short-term” deformation (daily). The longitudinal observation of the correlation criteria suggests that the deformation is driven by both rainfall and water level in the dam reservoir. However, the key triggering factor shifts from reservoir water level to rainfall from the toe to the trailing edge of the Outang landslide. Based on the data mining results of monthly data, the monthly and bi-monthly cumulative rainfall are identified to be the

Please cite this article as: Yang B et al., Spatiotemporal deformation characteristics of Outang landslide and identification of triggering factors using data mining, Journal of Rock Mechanics and Geotechnical Engineering, https://doi.org/10.1016/j.jrmge.2023.09.030

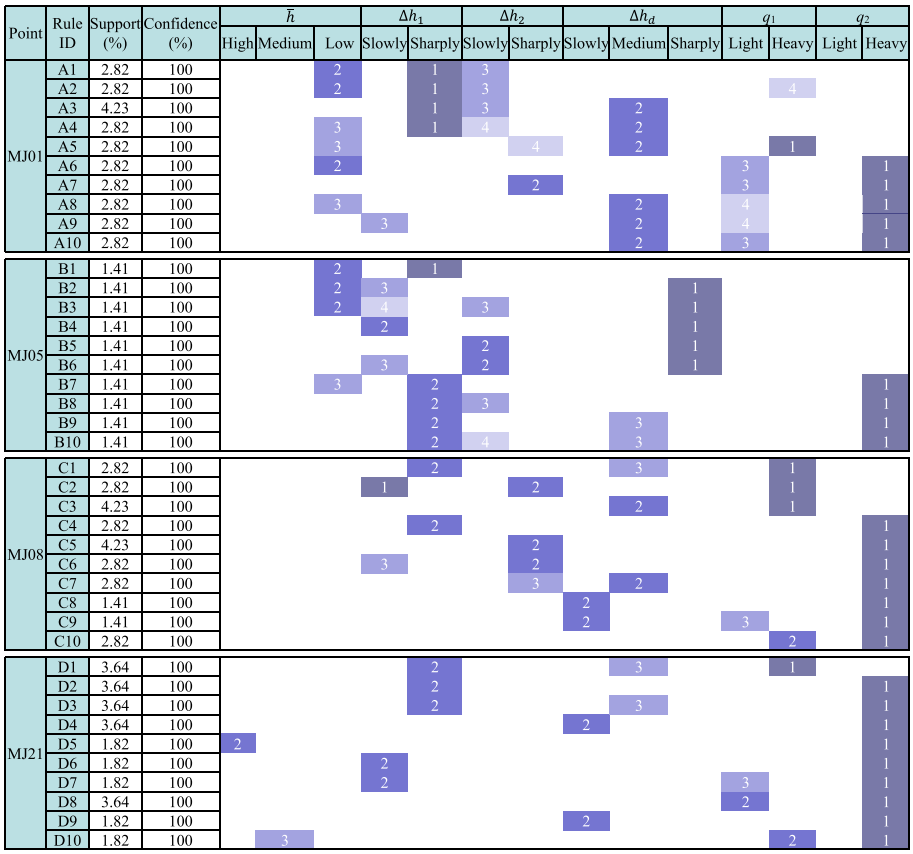

Fig. 16. Association criteria of MJ01, MJ05, MJ08, MJ21 monitoring point. The numbers within the boxes indicate their sequence in a correlation rule. A1 represents an association rule that the antecedent items are "Δh_1 = sharply-variation, $\overline{h}$ = low-water-level and Δh_2 = slowly-variation" and the consequence is monthly displacement rate.

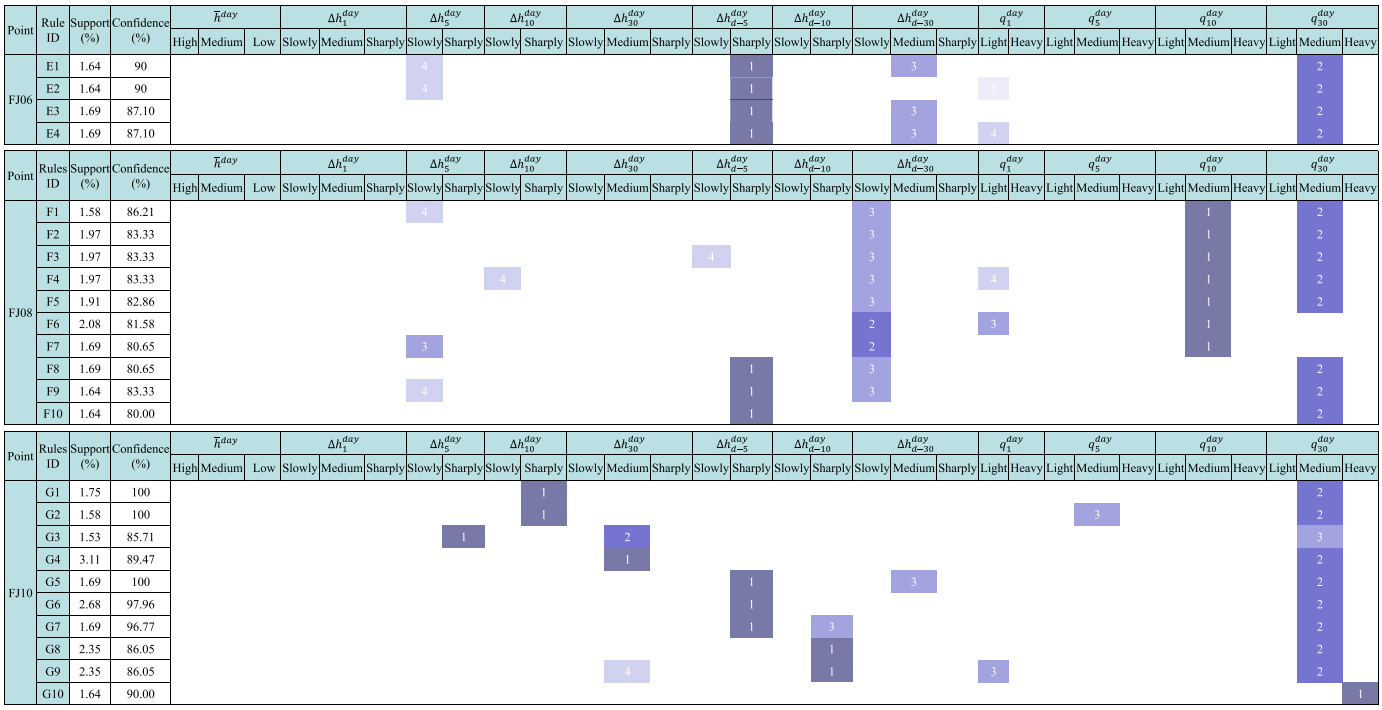

Fig. 17. Association criteria of FJ06, FJ08, FJ10 monitoring points. The numbers within the boxes indicate their sequence in a correlation rule. E1 represents an association rule that the antecedent items are "Δh_{d-5}^{day} = sharply-variation, q_{30}^{day} = medium-cumulative-rainfall, Δh_{d-30}^{day} = medium-variation and Δh_5^{day} = slowly-variation" and the consequence is daily displacement rate.

Please cite this article as: Yang B et al., Spatiotemporal deformation characteristics of Outang landslide and identification of triggering factors using data mining, Journal of Rock Mechanics and Geotechnical Engineering, https://doi.org/10.1016/j.jrmge.2023.09.030

dominant triggering factors. The data mining of the daily rainfall reveals that the 10-d cumulative rainfall and 5-d cumulative drop in reservoir water level dominate as influential hydrological factors. It was found effective to integrate the two datasets covering varying monitoring periods and frequency, thus providing insight into the influence of the different hydrological factors. The information can then be exploited further to form a basis for effective early warning routines.

Understanding spatiotemporal deformation characteristics and identifying triggers on different parts of landslides can contribute to more efficient landslide prevention and control. Research on spatiotemporal deformation and establishing data mining rules relating to rainfall and dam reservoir level, as done in this paper, can be applied to other landslides in the TGRA and in other locations.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by the Natural Science Foundation of Shandong Province, China (Grant No. ZR2021QD032). The first author wishes to thank the China Scholarship Council (CSC) and the Norwegian Geotechnical Institute (NGI) for funding her research period at NGI. The authors wish to thank Dr. Zizheng Guo, Dr. Ting Xiao, and Dr. Fasheng Miao for their assistance in collecting the data.

References

Abdullah, U., Ahmad, J., Ahmed, A., 2008. Analysis of effectiveness of apriori algorithm in medical billing data mining. In: Proceedings of the 4th International Conference on Emerging Technologies. Rawalpindi, Pakistan, pp. 327–331.

Agrawal, R., Imieliński, T., Swami, A., 1993. Mining association rules between sets of items in large databases. ACM SIGMOD Record. ACM 22 (2), 207–216.

Aranganayagi, S., Thangavel, K., 2007. Clustering categorical data using silhouette coefficient as a relocating measure. In: Proceeding of the International Conference on Computational Intelligence and Multimedia Applications (ICCIMA 2007), Sivakasi, India, pp. 13–17.

Benassi, M., Garofalo, S., Ambrosini, et al., 2020. Using two-step cluster analysis and latent class cluster analysis to classify the cognitive heterogeneity of cross-diagnostic psychiatric inpatients. Front. Psychol. 11, 1085.

Borgelt, C., Kruse, R., 2002. Induction of association rules: apriori implementation. In: Proceeding of the 15th Conference on Computational Statistics, Heidelberg, Germany, pp. 395–400.

Chinkulkijniwat, A., Tirametatiparat, T., Supotayan, C., Yubonchit, S., 2019. Stability characteristics of shallow landslide triggered by rainfall. J. Mt. Sci. 16, 2171–2183.

Chiu, T., Fang, D.P., Chen, J., Yao, W., Jeris, C., 2001. A robust and scalable clustering algorithm for mixed type attributes in large database environment. In: Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. CA, USA, pp. 263–268.

Chongqing Municipal Planning and Natural Resources Bureau, 2021. Annual Geological Disaster Prevention and Control Program of Chongqing in 2021. Chongqing Municipal Planning and Natural Resources Bureau, Chongqing.

Dai, Z.W., Yin, Y.P., Wei, Y.J., Lü, T., Luo, J.H., Yao, W., 2016. Deformation and failure mechanism of Outang landslide in three Gorges reservoir area. J. Eng. Geol. 24 (1), 44–55 (in Chinese).

Dai, Z.W., Zhang, C.Y., Wang, L., Fu, Y.P., Zhang, Y., 2021. Interpreting the influence of rainfall and reservoir water level on a large-scale expansive soil landslide in the Danjiangkou Reservoir region. China. Eng. Geol. 288, 106110.

Davies, E., 2005. Machine Vision: Theory, Algorithms, Practicalities, third ed. Morgan Kaufmann Publishers, San Francisco, CA, USA.

Fourniadis, I.G., Liu, J.G., Mason, P.J., 2007. Regional assessment of landslide impact in the Three Gorges area, China, using ASTER data: Wushan-Zigui. Landslides 4, 267–278.

Franceschini, R., Rosi, A., Catani, F., Casagli, N., 2022. Exploring a landslide inventory created by automated web data mining: the case of Italy. Landslides 19 (4), 841–853.

Froude, M.J., Petley, D.N., 2018. Global fatal landslide occurrence from 2004 to 2016. Nat. Hazards Earth Syst. Sci. 18, 2161–2181.

Gariano, S.L., Guzzetti, F., 2016. Landslides in a changing climate. Earth Sci. Rev. 162, 22–252.

Gariano, S.L., Melillo, M., Peruccacci Brunetti, M.T., 2020. How much does the rainfall temporal resolution affect rainfall thresholds for landslide triggering? Nat. Hazards 100, 655–670.

Guo, Z.Z., Chen, L.X., Yin, K.L., Shrestha, D.P., Zhang, L., 2020. Quantitative risk assessment of slow-moving landslides from the viewpoint of decision-making: a case study of the Three Gorges Reservoir in China. Eng. Geol. 273, 105667.

Guo, X.F., Dias, D., Carvajal, C., Peyras, L., Breul, P., 2022a. Three-dimensional probabilistic stability analysis of an earth dam using an active learning meta-modeling approach. Bull. Eng. Geol. Environ. 81, 1–20.

Guo, L.J., Miao, F.S., Zhao, F.C., Wu, Y.P., 2022b. Data mining technology for the identification and threshold of governing factors of landslide in the Three Gorges Reservoir Area. Stoch. Environ. Res. Risk Assess. 36, 3997–4012.

Guzzetti, F., Mondinia, A.C., Cardinali, M., Fiorucci, F., Santangeloab, M., Chang, K.T., 2012. Landslide inventory maps: new tools for an old problem. Earth Sci. Rev. 112 (1–2), 42–66.

Guzzetti, F., Gariano, S.L., Peruccacci, S., et al., 2020. Geographical landslide early warning systems. Earth Sci. Rev. 200, 102973.

He, T., Zhu, S.N., Wang, H., Wang, J.W., Qing, T., 2022. The diagnosis of satellite flywheel bearing cage fault based on two-step clustering of multiple acoustic parameters. Measurement 201, 111683.

Izakian, H., Abraham, A., 2011. Fuzzy C-means and fuzzy swarm for fuzzy clustering problem. Expert Syst. Appl. 38 (3), 1835–1838.

Jain, A.K., Duin, R.R.P.W., Mao, J.C., 2000. Statistical pattern recognition: a Review. IEEE Trans. Pattern Anal. Mach. Intell. 22 (1), 4–37.

Kumar, R., Anbalagan, R., 2015. Landslide susceptibility zonation in part of Tehri reservoir region using frequency ratio, fuzzy logic and GIS. J. Earth Syst. Sci. 124, 431–448.

Lee, C.Y., Antonsson, E.K., 2000. Dynamic partitional clustering using evolution strategies. In: Proceeding of the 26th Annual Conference of the IEEE Industrial Electronics Society. Nagoya, Japan, pp. 2716–2721.

Lee, S.J., Siau, K., 2002. A review of data mining techniques. Ind. Manag. Data Syst. 101 (1), 41–46.

Liao, K., Zhang, W., Zhu, H.H., et al., 2022. Forecasting reservoir-induced landslide deformation using genetic algorithm enhanced multivariate Taylor series Kalman filter. Bull. Eng. Geol. Environ. 81, 104.

Luo, S.L., Huang, D., 2020. Deformation characteristics and reactivation mechanisms of the Outang ancient landslide in the Three Gorges Reservoir, China. Bull. Eng. Geol. Environ. 79, 3943–3958.

Ma, J.W., Tang, H.M., Hu, X.L., et al., 2017. Identification of causal factors for the Majiagou landslide using modern data mining methods. Landslides 14, 311–322.

Mandal, K., Saha, S., Mandal, S., 2021. Applying deep learning and benchmark machine learning algorithms for landslide susceptibility modelling in Rorachu river basin of Sikkim Himalaya, India. Geosci. Front. 12 (5), 101203.

McClean, S.I., 2003. Data mining and knowledge discovery. In: Meyers, R.A. (Ed.), Encyclopedia of Physical Science and Technology, 3rd ed. Academic Press, New York, NY, USA, pp. 229–246.

Miao, F.S., Wu, P.Y., Li, L.W., Liao, K., Xue, Y., 2021. Triggering factors and threshold analysis of baishuihe landslide based on the data mining methods. Nat. Hazards 105, 2677–2696.

Miao, F.S., Wu, Y.P., Török, Á., Li, L.W., Xue, Y., 2022. Centrifugal model test on a riverine landslide in the Three Gorges Reservoir induced by rainfall and water level fluctuation. Geosci. Front. 13 (3), 101378.

Mirus, B.B., Jones, E.S., Baum, R.L., et al., 2020. Landslides across the USA: occurrence, susceptibility, and data limitations. Landslides 17 (10), 2271–2285.

Nayak, J., Naik, B., Behera, H.S., 2015. Fuzzy C-Means (FCM) clustering algorithm: a decade review from 2000 to 2014. Comput. Intell. Data Mining 2, 133–149.

Park, J.Y., Lee, S.R., Lee, D.H., Kim, Y.T., Lee, J.S., 2019. A regional-scale landslide early warning methodology applying statistical and physically based approaches in sequence. Eng. Geol. 260, 105193.

Pasierb, B., Grodecki, M., Gwóźdź, R., 2019. Geophysical and geotechnical approach to a landslide stability assessment: a case study. Acta Geophys. 67, 1823–1834.

Pecoraro, G., Calvello, M., Piciullo, L., 2019. Monitoring strategies for local landslide early warning systems. Landslides 16 (2), 213–231.

Strauhal, T., Loew, S., Holzmann, M., Zangerl, C., 2016. Detailed hydrogeological analysis of a deep-seated rockslide at the Gepatsch reservoir (Klasgarten, Austria). Hydrogeol. J. 24 (2), 349–371.

Sun, P., Wang, H.J., Wang, G., Li, R.J., Zhang, Z., Huo, X.T., 2021. Field model experiments and numerical analysis of rainfall-induced shallow loess landslides. Eng. Geol. 295, 106411.

Tang, H.M., Wasowski, J., Juang, C.H., 2019. Geohazards in the three Gorges reservoir area, China - lessons learned from decades of research. Eng. Geol. 261, 105267.

Tomás, R., Li, Z., Lopez-Sanchez, J.M., Liu, P., Singleton, A., 2016. Using wavelet tools to analyse seasonal variations from InSAR time-series data: a case study of the Huangtupo landslide. Landslides 13, 437–450.

Wang, X.M., Pedrycz, W., Niu, R.Q., 2015. Spatio-temporal analysis of quaternary deposit landslides in the three Gorges. Nat. Hazards 75, 2793–2813.

Wang, J.G., Daniel, S., Liu, Q.B., Su, A.J., Hu, X.L., Blum, P., 2021. Three-dimensional landslide evolution model at the Yangtze River. Eng. Geol. 292, 106275.

Please cite this article as: Yang B et al., Spatiotemporal deformation characteristics of Outang landslide and identification of triggering factors using data mining, Journal of Rock Mechanics and Geotechnical Engineering, https://doi.org/10.1016/j.jrmge.2023.09.030

Wang, L., Wu, C.Z., Yang, Z.Y., Wang, L.Q., 2023. Deep learning methods for time-dependent reliability analysis of reservoir slopes in spatially variable soils. Comput. Geotech. 159, 105413.
Weidner, L., DePrekel, K., Oommen, T., Vitton, S., 2019. Investigating large landslides along a river valley using combined physical, statistical, and hydrologic modeling. Eng. Geol. 259 (4), 105169.
Wu, X.D., Zhu, X.Q., Wu, G.Q., Ding, W., 2014. Data mining with big data. IEEE Trans. Knowl. Data Eng. 26 (1), 97–107.
Wu, X.L., Zhan, F.B., Zhang, K.X., Deng, Q.L., 2016. Application of a two-step cluster analysis and the Apriori algorithm to classify the deformation states of two typical colluvial landslides in the Three Gorges, China. Environ. Earth Sci. 75, 146.
Yan, J.B., Zou, Z.X., Mu, R., et al., 2022. Evaluating the stability of Outang landslide in the Three Gorges Reservoir area considering the mechanical behavior with large deformation of the slip zone. Nat. Hazards 112, 2523–2547.
Yao, W., Li, C.D., Zuo, Q.J., Zhan, H.B., Criss, R.E., 2019. Spatiotemporal deformation characteristics and triggering factors of Baijiabao landslide in Three Gorges Reservoir region, China. Geomorphology 343, 34–47.
Yang, B.B., Yin, K.L., Xiao, T., Chen, L.X., Du, J., 2017. Annual variation of landslide stability under the effect of water level fluctuation and rainfall in the Three Gorges Reservoir, China. Environ. Earth Sci. 76, 564.
Yang, B.B., Yin, K.L., Lacasse, S., Liu, Z.Q., 2019. Time series analysis and long short-term memory neural network to predict landslide displacement. Landslides 16 (4), 677–694.
Yang, B.B., Xiao, T., Wang, L.Q., Huang, W., 2022. Using complementary ensemble empirical mode decomposition and gated recurrent mass to predict landslide displacements in dam reservoir. Sensors 22, 1320.
Yin, Y.P., Wang, H.D., Gao, Y.L., Li, X.C., 2010. Real-time monitoring and early warning of landslides at relocated wushan town, the three Gorges reservoir, China. Landslides 7 (3), 339–349.
Zhang, Y.G., Zhu, S.Y., Tan, J.K., Li, L.D., Yin, X.J., 2020. The influence of water level fluctuation on the stability of landslide in the Three Gorges Reservoir. Arabian J. Geosci. 13, 845.
Zhang, W.G., Li, H.R., Han, L., Chen, L.L., Wang, L., 2022a. Slope stability prediction using ensemble learning techniques: a case study in Yunyang County, Chongqing, China. J. Rock Mech. Geotech. Eng. 2022 14 (4), 1089–1099.
Zhang, W.G., Li, H.R., Tang, L.B., Gu, X., Wang, L.Q., Wang, L., 2022b. Displacement prediction of Jiuxianping landslide using gated recurrent unit (GRU) networks. Acta. Geotech. 17, 1367–1382.
Zhang, W.G., Wu, C.Z., Tang, L.B., Gu, X., Wang, L., 2022c. Efficient time-variant reliability analysis of bazimen landslide in the three Gorges reservoir area using XGBoost and LightGBM algorithms. Gondwana Res. 123, 41–53.
Zhou, C., Cao, Y., Yin, K.L., Intrieri, E., Catani, F., Wu, L.X., 2022. Characteristic comparison of seepage-driven and buoyancy-driven landslides in Three Gorges Reservoir area, China. Eng. Geol. 301, 106590.

Beibei Yang obtained her BSc and PhD degrees in Geological Engineering from China University of Geosciences (Wuhan) in 2013 and 2019, respectively. She is now an associate professor of School of Civil Engineering in Yantai University. She is also a postdoctoral researcher at Norwegian Geotechnical Institute (NGI), a world-renowned research institute in geotechnical engineering. Her research interests include: (i) landslide monitoring and stability evaluation, and (ii) landslide prediction and risk assessment.

Please cite this article as: Yang B et al., Spatiotemporal deformation characteristics of Outang landslide and identification of triggering factors using data mining, Journal of Rock Mechanics and Geotechnical Engineering, https://doi.org/10.1016/j.jrmge.2023.09.030