Exploring the potential relationship between the occurrence of debris flow and landslide

Zhu Liang[1], Changming Wang[1], Donghe Ma[2] and Kaleem Ullah Jan Khan[1]

[1]College of Construction Engineering, Jilin University, 130000 Changchun, People's Republic of China;

[2]China Water Northeastern Investigation, Design and Research Co.Ltd.

E-mail:wangcm@jlu.edu.cn

Abstract: The aim of the present study is to explore the potential relationship between debris flow and soil slide by establishing susceptibility zoning maps (SZM) separately with the use of random forest. Longzi County, located in Southeastern Tibet, where historical landslides occurred commonly, was selected as the study area. The work has been carried out with the following steps: **(1)** An inventory map consisting of 448 landslides (399 soil slides and 49 debris flows) was determined; **(2)** Slope units and 11 conditioning factors were prepared for the susceptibility modelling of landslide while watershed units and 12 factors for debris flow; **(3)** SZM were constructed for landslide and debris flow, respectively, with the use of random forest; **(4)** The performance of two models were evaluated by 5-fold cross-validation using relative operating characteristic curve (ROC), area under the curve (AUC) and statistical measures; **(5)** The potential relationship between soil slide and debris flow was explored by the superimposition of two zoning maps; **(6)** Gini index was applied to determined the major factors and analyze the difference between debris flow and soil slide; **(7)** A combined susceptibility map with two kinds of disaster

was obtained. Two models had demonstrated great predictive capabilities, of which accuracy and AUC was 87.33%, 0.902 and 85.17%, 0.892, respectively. The loose sources need by the debris flow were not necessarily brought by the landslides although most landslides can be converted into debris flow. The area prone to debris flow did not promote the occurrence of landslide. A susceptibility zoning map composed of two or more natural disasters is comprehensive and significant in this regard, which provides valuable reference for researches of disaster-chain and engineering applications.

Key words: Landslide; Debris flow; Susceptibility; Random forest; Potential relationship

1. Introduction

Soil slide and debris flow are two kinds of natural phenomenon mainly occurring in mountainous areas, which pose considerable threats to people, industries, and the environment directly or indirectly. Generally, damages can be decreased to a certain extent by predicting the likely location of future disasters (Pradhan, 2010). Thus, extensive research has been conducted for the prediction and susceptibility assessment of soil slide and debris flow.

In geomorphology, a “landslide” is the movement of a mass of rock, debris or earth down a slope, under the influence of gravity (Cruden and Varnes, 1996). According to different variables, landslides can be divided into different types (Varnes, 1978). Debris flow is a specific type of landslide, which can be defined as (Hungr et al. 2013): ‘‘Very rapid to extremely rapid surging flow of saturated debris in a steep channel”. Most of debris flows are runoff generated (Ma et al., 2018). Generally, slides that occur on a steep slope and become disaggregated as they tumble

down can transform into debris flows if they contain sufficient water for saturation (Huang et al., 2020). Debris flow usually occurs on a channel bed for the entrainment into abundant runoff of debris supplied by deep or shallow slides of slopes incised by the channel (Imaizumi et al.2019; Zhou et al., 2019). Therefore, slides may provide sufficient material source for the occurrence of debris flow and most of the slides are accompanied by debris flow. In the past, it is not clear the way the potential relationship between debris flow and landslide is approached through the separated susceptibility analysis (Alessandro et al., 2015; Guzzetti et al., 2005). In addition, some scholars made separate evaluations of slides and debris flow (Park et al., 2011; Haydar et al., 2016). Some scholars have proposed a coupled model of landslide-debris flow (Chiang et al., 2012; Gomes et al., 2013). However, not every slide has evolved into a debris flow and the material source of the debris flow is not necessary coming from slides. The formation and manifestations of different types of landslides are different, especially debris flow, which is a kind of "wet flow" (Varnes, 1978). In other words, there is no determined connection between debris flow and other types of landslide. Therefore, the potential relationship between debris flow and other types of landslide need further exploration.

Besides, the conditioning factors and mapping units involved in the susceptibility assessment different kinds of landslides are not identical. Especially slope and water content are the most critical factors controlling movements of debris flow (Takahashi 2007). Therefore, it is more reasonable to evaluate the susceptibility of different kinds of landslides separately. As an example, one landslide inventory map includes only one type of landslide, as does debris flow.

The methods of susceptibility assessment can be broadly classified as qualitative or quantitative (Aleotti et al., 1999). Several methods and approaches have been proposed and tested

to ascertain susceptibility, such as physical-based approaches (Carrara et al., 2008), heuristic methods (Blais et al., 2016) and statistically-based approaches (Reichenbach et al., 2018). In addition, new machine learning models, such as neural networks (Park et al.,2013), support vector machines (Colkesen et al.,2016) and random forest (RF) (Zhu et al., 2020a), have also been applied.

The Longzi County in Southeastern Tibet is always exposed to slides and debris flow hazard because of climatic and topographic conditions, which is chosen as the study area. The purpose of the present study is to explore the potential relationship between the occurrence of debris flow and soil slide by establishing susceptibility zoning maps separately with the use of random forest. It also provides a reference for the study of landslide-debris flow, a common disaster chain.

2. Materials

2.1 Study area

The study area located in Longzi Township, Longzi County, Southeastern Tibet is bounded by longitudes of 92°15'E and 92°45'E, latitudes of 28°10'N and 28°30'N (Fig.1). It covers a surface of about 535 km^2 with a population of more than 6000. The study area belongs to a semi-arid temperate monsoon climate with the annual rainfall of 279 mm, mainly concentrated in May to September. The seismic intensity within the area has a degree of VIII on the modified Mercalli index.

The study area belongs to the zone of stratigraphic division of the Northern Himalayan block. The strata is mainly composed of Mesozoic Cretaceous, Jurassic, Triassic, and Cenozoic units. Three types of lithology were mainly observed during our field investigation: Siltstone from the

Laka Formation (K_1l); Conglomerates from the Weimei Formation (J_3w) and Quaternary slope wash (Q_4^{el+dl}) from the Cenozoic strata.

Main common disasters in the study area mainly consist of rain-fed high frequency debris flows and landslides, which destroyed and flooded roads, bridges, farmlands, villages, etc., causing great economic losses.

2.2 Landslide and debris flow inventory

The statistically-based susceptibility models are based on an important assumption: future landslides have more chances to occur again under the conditions which led to the landslides past and present (Varnes, 1984; Furlani and Ninfo, 2015). Therefore, a complete and accurate inventory map is the key for model training and validation. In this study, data comes from historical records (1970~2010), field surveys from 2000~2003 (**Fig.2 and Fig.3**) and interpretation of Google Earth images carried out in Google Earth pro 7.1 (**Fig.4**). Finally, a total of 399 soil slides and 49 debris flow locations with positive label were recorded and mapped (**Fig.1**) and the same number of non-landslide points with negative were selected randomly from the landslide-free area. The landslide locations were recorded as a point which were showed in **Fig.1**.

2.3 Mapping units

The selection of the mapping unit is an important pre-requisite for susceptibility modelling (Guzzetti, 2006). The main mapping units commonly used for landslide and debris flow susceptibility assessment are grid cells (Reichenbach et al., 2018). Despite its popularity and operational advantages, grid-cells have clear drawbacks for susceptibility modelling (Guzzetti et al., 1999). There is no physical relationship between a grid-cell or a group of grid-cells and slope,

while slope units can make up for this deficiency (Reichenbach et al., 2018). Depending on the landslide type, a slope unit may correspond to an individual slope, an ensemble of adjacent slopes or a small catchment (Reichenbach et al., 2018). The geometry of the debris flow is tortuous and complex, which is not suitable to represent it with a regular grid unit. In the present study, adjacent slope units were applied to the susceptibility assessment of soil slide. First-order sub-catchments, which is also called watershed unit, was applied to the susceptibility of debris flow (Francesco et al., 2015; Zhu et al., 2020b). Accordingly, the study area was divided into 1003 slope units for the modeling of soil slide or 174 watershed units for debris flow.

2.4 Controlling factors and mapping

The selection of evaluation parameters is another key prerequisite to ensure that the model is accurate and reasonable. With reference to previous studies (Ahmed et al., 2016; Xu et al., 2013; Braun et al., 2018), there area lot of difference between the factors used by different landslide susceptibility assessments. In this paper, 11 controlling factors are selected for the susceptibility assessment of landslide, including distance to fault, distance to road, distance to river, annual rainfall, slope angle, aspect, plan curvature, profile curvature, topographic wetness index, elevation and maximum elevation difference. The occurrence of debris flow emphasizes the indispensability of provenance, topography and triggering factors. Moreover, availability, reliability, and practicality of the factor data were also considered (van Westen et al., 2008). Accordingly, a total of 12 controlling factors, including basin area, main channel length, normalized difference vegetation index (NDVI), drainage density, roundness, melton, average gradient of main channel, slope angle, maximum elevation difference, annual rainfall, distance to

fault and elevation were selected to fully reflect the characteristics of the watershed for the susceptibility assessment of debris flow. Detailed information on conditioning factors is shown in **Fig.5a~5m**. A brief description of each controlling factor is given below.

2.4.1 Factors used in landslide susceptibility assessment

Aspect reflects sunshine duration and rainfall, which is frequently used as landslide controlling factor (Dai and Lee, 2002) and was reclassified into 8 classes (**Fig. 5g**). Plan curvature and profile curvature which reflect the relief of the terrain were both considered and reclassified into six classes (**Fig. 5b and 5e**). Generally, faults, rivers and roads play a key role in the occurrence of landslides as landslides are more likely disturbed around faults, rivers and roads (Taskin et al., 2015) which were reclassified into seven classes using an interval of 1500m (**Fig. 5i~k**). Topographic wetness index (TWI) belong to a hydrological variable that reflects both slope and soil moisture content (Wilson and Gallant 2000) and was reclassified into five classes (**Fig. 5h**).

2.4.2 Factors used in debris flow susceptibility assessment

NDVI reflects the vegetation conditions in the area and was reclassified into 5 classes (**Fig. 6b**). Drainage density is the ratio of the total drainage length to the watershed area and was reclassified into six classes (**Fig.6 g**). Roundness refers to the ratio of the area of a basin to the area of a circle with the same circumference and was reclassified into six classes (**Fig.6 d**). Melton ratio refers to the ratio of the degree of undulation in the watershed to the square root of the arithmetic area of the watershed (Melton, 1965), which is reclassified into seven classes (**Fig. 6a**). Basin area was reclassified into four classes and main channel length are represented by the same

graph considering the correlation between the two controlling factors (**Fig.6h**). Average gradient of main channel, which is the ratio of the maximum elevation difference of main channel to its linear length, was reclassified into six classes (**Fig. 6j**).

2.4.3 Factors used in both landslide and debris flow

Rainfall is the only triggering factor to be considered for both landslide and debris flow in this paper, which was reclassified into six classes (**Fig. 5a and Fig. 6c**). Slope angle which reflects kinetic energy conditions is frequently employed in both landslide and debris flow susceptibility mapping and was reclassified into six classes (**Fig. 5f and Fig. 6i**). Maximum elevation difference also reflects the kinetic energy condition and is reclassified into 6 classes using an interval of 200m (**Fig. 5c and Fig. 6e**). Elevation affects the rainfall and vegetation (**Fig. 5d and Fig. 6f**), which has also been used by many authors and was reclassified into five classes in the study (Ayalew and Yamagishi, 2005; Pourghasemi et al. 2013a, b).

The values of 18 controlling factors were classified by processing the raw data in the ArcGIS 10.2 platform. Morpholigical and topographic related factors were derived from the DEM with a resolution of 30 × 30 m (http://www.gscloud.cn). Geological related factors were extracted from 1:50000 geological maps (http://www.ngac.org.cn/). Rainfall is one of the most important external factors inducing landslides and debris flow, which was determined by ordinary kriging interpolation in ArcGIS based on 11 precipitation stations data near the study area collected from the China Meteorological Administration. Roads networks are provided by Landsat 8 LOI images (2018.8.12).

3. Methods

3.1 Sampling strategy and performance assessment

Statistical models for landslide susceptibility zoning reconstruct the relationships between dependent and independent variables using training data set, and verify these relationships using validation sets (Guzzetti et al., 2006a,b), which usually implies the partitioning of the inventory in subsets. The sampling strategy affects the results of the susceptibility map (Yilmaz, 2010). The partition of landslide inventory is approached based on temporal, spatial or random criteria, (Chung and Fabbri, 2003) and among them that of one time random selection is the most used (Reichenbach et al., 2018). However, there is a need for a more reliable estimation of the model performance. The ability of the models to classify independent test data was elaborated using a k-fold cross validation procedure (k=5 in this paper) (James et al., 2013).

The computation of the area under the curve (AUC) is the most popular metrics to estimate the quality of model, which has been applied for ROC curves (Green and Swets, 1966). It is one of the most commonly used indicators. Three statistical metrics as accuracy, sensitivity, and specificity are generally applied to assess the performance of the landslide susceptibility models (Tien Bui et al. 2016).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{FP + TN} \quad (1)$$

where True Positives (*TP*), i.e., cells predicted unstable and observed unstable, True Negatives

(*TN*), i.e., cells predicted stable and observed stable, False Positives (*FP*), i.e., cells predicted unstable but observed stable and False Negatives (*FN*), i.e., cells predicted stable but observed unstable.

3.2 Random Forests

Random forest (RF) is a powerful ensemble-learning method and was first introduced by Breiman (2001). Bagging technique is applied to select random samples of variables and observations as the training data set at each node of the tree for the modeling of RF. Unselected cases (out of bag) are used to calculate the error of the model (OOB Error). The increase in OOB error is proportional to the importance of the predictive variable (Breiman and Cutler 2004). There are no restrictions on the types of variables, either numerical or categorical. RF has the ability to reduce errors caused by unbalanced data, which is suitable for susceptibility assessment.

The number of trees and the number of predictive variables used to split the nodes are two user-defined parameters required to grow a random forest (Ahmed et al.,2016). Cross-Validation were applied to optimized the hyper-parameters before application (Schratz et al., 2019). Scikit-learn package (Pedregosa et al.,2011) in the programming software python version 3.7 was used for the modeling. Gini index (the larger the value of the obtained result, the greater the contribution to the occurrence of landslide) (Breiman,2001) was applied to analyze the relative importance of conditioning factors for both soil slide and debris flow.

4. Results and verification

4.1 Landslide susceptibility mapping results

The predictive accuracy, ROC curves and AUC values of the RF models using training data set were showed in **Table 1** and **Fig. 7**. The RF model ensured a satisfactory performance for classifying landslides with a sensitivity value of 91.62%. In terms of the classification of non-landslides zones, specificity value also reached 89.06. An AUC value ranges from 0.5 to 1 and equals to 1 indicates perfect prediction accuracy (Vorpahl et al., 2012). The RF model had great performance in terms of AUC, with a value of 0.976. Standard error (St.), confidence interval (CI) at 95% and significance (Sig.) were applied as three evaluation statistics. All these results indicated that the models achieved a reasonable goodness-of-fit, for which the values were reasonably small.

Verifying the generalization ability of the model is a key step in prediction models as shown in **Table 2** and **Fig. 7**. Accordingly, the values of sensitivity and specificity were 88.69% and 86.05%, respectively. The model also achieved a great performance in terms of AUC with a value of 0.902. In comparison with the training dataset, the accuracy and AUC values have slightly decreased, but still perform well.

The landslide susceptibility map was reclassified into five classes: very low (0~0.2), low (0.2~0.4), moderate (0.4~0.6), high (0.6~0.8), very high (0.8~1) by using the equal spacing method (**Fig.8**). The map should satisfy two spatial effective rules: (1) The existing disaster points should belong to the high-susceptibility class and **(2)** The high-susceptibility class should cover only small areas (Bui et al. 2012). The number of units belonging to very high class accounted for

17% of the total number of units (**Fig.9**). Disaster points were mostly in the dark (red or orange) areas. The units belonging to moderate class accounted for the smallest proportion, at 13% of the total number of units (**Fig.9**).

The controlling factors with significant effects were selected and normalized as shown in **Table 3**. The weight values of slope angle, distance to fault, plan curvature and topographic wetness index were 0.21, 0.19, 0.17, 0.13 respectively, which was closely related to the occurrence of landslide. The weight values of distance to road, maximum elevation difference, profile curvature and elevation are less than 0.1 as 0.08, 0.08, 0.06, and 0.05, respectively (**Fig.10**).

4.2 Debris flow susceptibility mapping result

The debris flow susceptibility model perform well with a very high sensitivity and specificity values as 87.80% and 88.89%, respectively. In terms of accuracy and AUC, the model had also a great prediction performance with the value of 88.57% and 0.967 (**Fig.7**). Three evaluation statistics also indicate a reasonable goodness-of-fit for the model.

Table 2 shows that the values of sensitivity and specificity were 85.71% and 84.62%, which were slightly decreased compared to the training model. However, the model had achieved a great performance in terms of AUC, with value of 0.892.

The number of units belonging to very high-class accounted for 15% of the total number of units while the units belonging to high-class accounted for the smallest proportion at 13%. More than half of the units (58%) belong to on a low or very low-class (**Fig.9**). Disaster points were mostly in the dark (Bright or deep red) areas (**Fig.8**).

The weight values of main channel length, roundness and slope angle were 0.25, 0.16, 0.14

respectively and these factors have significant influence on the occurrence of debris flow (**Table 4**). The weight values of elevation, maximum elevation difference, melton and basin area are close to 0.1, which are 0.13, 0.12, 0.1, and 0.1 respectively (**Fig.10**).

4.3 Analysis and comparison of landslide and debris flow susceptibility

It is worth comparing the two susceptibility zoning maps. In terms of prediction accuracy, the values of sensitivity, specificity and AUC of landslide model were slightly higher than that of debris flow. However, both models achieved high predictive performance. Therefore, the landslide and debris flow susceptibility assessment models based on RF are reliable. The purpose of the present study is to explore the potential relationship between landslides and debris flows by establishing the respective susceptibility zoning maps. Figure 11 shows the overlapping areas between debris flow and landslide in high or very high-class of susceptibility zoning map. It can be seen that most of the areas with high or very high-class in the map of debris flow are covered with landslides. However, there are also non-overlapping areas between the two zoning maps. There are 23 watershed units belonging to high-class in the debris flow susceptibility zoning map (**Fig.8**), of which 17 units correspond to with high or very high-class slope units in the landslide zoning map (**Table 5**). In addition, there are 4 watershed units covered with low or very low class slope units. In the same way, 19 watershed units belonging to very high-class are covered with high or very high-class slop units and 4 watershed units with low or very low-class slop units. In other words, more than 70% of the high or very high-class watershed units are covered with high or very high-class slope units. However, there are still 30% of watershed units with high or very

high-class without the distribution of slope units in corresponding grades. It validated the previous view that most of landslides can be transformed into debris flows.

Factor analysis which is a tool for dimensionality reduction and exploring the major factors was applied to further analyze the reasons for the difference (Zhu et al., 2020a). 36 watershed units with distribution of high or very high-grade slope units were taken as model 1 and the left 8 watershed units as model 2 (**Table 5**). The KMO (Kaiser-Meyer-Olkin) and Sig. testing are two statistical parameters which ensured the feasibility before application, which are provided by SPSS. The closer the KMO statistic is to 1, the stronger the correlation between variables and the better the effect of factor analysis. The KMO values were 0.766 and 0.643 respectively, which indicated that the correlation between variables was obvious and suitable for factor analysis (**Table 6)**. In model 1, the cumulative contribution rate of the first three factors (F1, F2 ,F3) reached to 83.6% (**Table 7**), while the first four factors (C1, C2, C3 and C4) reached to 80.5% for model 2 (**Table 8**). The first common factor mainly highlighted the information of basin area, main channel length and maximum elevation difference. Similarly, the second and the third common factor highlighted the information of slope angle and elevation and roundness, respectively. The difference between the two models (model 1 and model 2) is that the second model has the fourth common factor (**Table 8**), which emphasized the effects of rainfall and distance to the fault. The transformation from a landslide to a debris flow often occurs during heavy rainfall (Takahashi, 1978), and the landslides are the source area. But landslides are not the only source of debris flows. The loose material distributed in the basin is not necessarily caused by landslide.

In turn, we analyze the distribution of high or very high-class slope units in watershed units.

292 The landslide zoning map was put at the bottom floor and the debris flow zoning map on the top
293 floor (**Fig. 11**). There are 167 slope units belonging to high-class, of which 68 units (accounting
294 for about 40%) are distributed in the area of high or very high-class watershed units in the debris
295 flow zoning map (**Table 9**). Besides, 69 slope units (accounting for about 41%) are distributed in
296 the area of low or very low-class watershed units. Similarly, 53 slope units (accounting for about
297 30%) belonging to very high-class are distributed in the area of high or very high-class watershed
298 units and 88 slope units (accounting for about 50%) in low or very low-class slop units (**Table 9**).
299 Comparing with the extent of the landslide affecting the debris flow, the impact of the debris flow
300 on the landslide is not obvious. It indicated that the area prone to debris flow does not promote the
301 occurrence of landslides.

302 Finally, we took the center of gravity of 1,003 slope units as the potential hazard points and
303 spread them over 174 watershed units. Thus, a combining susceptibility zonation map for
304 landslide and debris flow was obtained (**Fig.11**). The darker the color, the higher the class of
305 susceptibility will be. It can be seen that the susceptibility in the south is generally higher than that
306 in the north, and the area in the southwest is disaster-prone. The northeast and central locations in
307 the area are less likely to be affected by landslides and belong to low-susceptibility areas. Green or
308 yellow dots, which refer to slope units with very low or low- class in the landslide zoning map,
309 mainly distributed in light-colored areas but there are also quite a few green or yellow dots
310 distributed in dark areas, which means that the occurrence of debris flow not necessarily depend
311 on landslides. Blue or black spots are mainly distributed in dark areas but there are also quite a
312 few blue or black spots distributed in dark light areas, which means that landslide is not the only
313 condition for debris flow to occur. Most of the watershed units are distributed with two or more

colored dots, which means that there would be multiple slope units with different susceptibility class in the same watershed. According to the combining susceptibility zoning map of landslide and debris flow, the study area can be divided into 4 categories: **(1)** Low or very low-class watershed units coupled with low or very low-class slope units; **(2)** Low or very low-class watershed units coupled with high or very high-class slope units; **(3)** High or very high-class watershed units coupled with low or very low-class slope units; **(4)** High or very high-class watershed units coupled with high or very high-class slope units. We assume that the occurrence of landslides can bring rich sources of debris flow, thereby promoting or aggravating the outbreak of debris flow, that is, forming a landslide-debris flow disaster chain. Therefore, the susceptibility assessment of the landslide-debris flow chain in the study area can be roughly divided into three classes, which are low, moderate and high (**Table 10**).

5. Discussion

5.1 Method used for modeling

Many researchers have used different statistically-based methods to evaluate the susceptibility of landslides or debris flows. Logistic regression and discriminant analysis are the most popular methods to use in traditional multivariate statistical analysis (Teigila, 2015; Abdelaziz et al., 2020). The performance of new learning machines, such as support vector machines and neural networks, has also been verified. RF, as a newly integrated learning machine, has been little used until now for susceptibility analysis of landslide and debris flows (Chen et al., 2017; Zhang et al., 2017). Actually, RF have powerful data processing capabilities and can simultaneously solve problems such as high-dimensional, unbalanced and data loss, which are common in geological disaster assessment. Most importantly, RF can compare the important differences between features and have ability to reduce errors caused by unbalanced data and, which achieved strong generalization properties (Zhu et al., 2020a, c).

5.2 Potential relationship between landslide and debris flow

There is a certain similarity in the evaluation of the susceptibility of landslide and debris flow as the selection of controlling factors and the application of modeling strategies. Therefore, some researchers have neglected the difference between landslide and debris flow i.e to express two different disasters with the same susceptibility zoning map (Ciurleo et al., 2016; Ciurleo et al., 2017; Persichillo et al., 2017). However, similarity does not always mean consistency. Many researchers have previously conducted studies into the debris flow mobilization from shallow landslide using a coupled methodology (Wang et al., 2013; Fan et al., 2017). However, not every

landslide evolves into a debris flow, which means that the analysis process is highly selective or uncertain. In the same way, the source of the debris flow is not limited to landslide. There, the potential relationship between landslide and debris flow needs to be discussed more reasonably and effectively. In this study, the corresponding influencing factors and mapping units are selected to establish landslide and debris flow susceptibility zoning maps, respectively. The potential relationship between landslide and debris flow is explored in two ways: **1)** Superimposing the high or very high-class susceptibility areas in the two maps; **2)** Transforming the slope units into points and distributed them on the watershed units. The relationship between landslide and debris flow is illustrated by the distribution of slope units of different grades on the watershed units with different prone grades. Different kinds of landslides should be evaluated respectively conditioning that conditioning factors and scale varies.

5.3 Necessity and feasibility of combining multiple natural disaster susceptibility zoning maps

Previous studies on susceptibility zoning mapping of disaster have agreed that one disaster corresponds to one map. However, it will cause some confusion in practical. For example, multiple disasters may be bred simultaneously in a watershed unit. For another case, the probability of one kind of disaster like debris flow in a watershed is negligible, while another disaster like rockfall occurs frequently. Therefore, we need to combine multiple zoning maps at the same time to give a comprehensive evaluation, which is arduous to achieve. On the one hand, the prediction accuracy and error of different zoning maps should be similar or even consistent. On the other hand, the dimensions of the mapping unit should be consistent or complementary.

The fact that the appropriate prediction method (like RF applied in this study) and mapping units applied to the two disasters makes it possible to merge the two zoning maps. Disaster risk is higher in landslide-debris flow chain, causing significant loss of life and property. Therefore, two natural disasters with potential relationship are simultaneously reflected in the same susceptibility zoning map, which can better guide the implementation of engineering, such as landslide-debris flow disaster chain.

6. Conclusion

In this study, susceptibility assessment models for landslide and debris flow are established through RF, respectively and the performance of the models are excellent in terms of accuracy and goodness of fit. The potential relationship between landslide and debris flow is discussed by the superimposition of two zoning maps and the following conclusions can be drawn:

(1) The landslide and debris flow susceptibility mapping based on RF models have great performance of accuracy and goodness-of-fit and have the ability to analyze the relative importance of different impact factors, which is suitable for the evaluation of natural disasters;

(2) There is no potential relationship between the occurrence of the two considered phenomena. Although most landslides will be converted into debris flow, the landslides are not necessarily the source of debris flow, and the loose sources carried by the debris flow are not necessarily brought by the landslides. On the other hand, the impact of the debris flow on the landslide is also not obvious.

(3) A susceptibility zoning map composed of two or more natural disasters is more comprehensive and significant, which provides valuable reference for researchers and engineering

applications.

Data availability. The data used to support the findings of this study are included within the article.

Author contributions. ZL was responsible for the writing and graphic production of the paper. CMW was responsible for the revision of the paper. DHM was responsible for calculation. KUJK was responsible for the translation.

Competing interests. The authors declare that they have no conflict of interest.

Special issue statement. This article is part of the special issue“Resilience to risks in built environments”. It is not associated with a conference.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 41972267 and 41572257).

References:

Ayalew L, Yamagishi H (2005) The Application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda–Yahiko Mountains, central Japan. Geomorphology 65:15–31. doi:10.1016/j. geomorph. 2004.06.010

Alessandro Trigila, Carla Iadanza, Carlo Esposito, et al. Comparison of Logistic Regression and Random Forests techniques for shallow landslide susceptibility assessment in Giampilieri (NE Sicily, Italy). 2015, 249:119-136.

Ahmed Mohamed Youssef, Hamid Reza Pourghasemi, Zohre Sadat Pourtaghi, et al. Erratum to:

Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. 2016, 13(5):1315-1318.

Anika Braun, Elias Leonardo Garcia Urquia, Rigoberto Moncada Lopez, et al. Landslide Susceptibility Mapping in Tegucigalpa, Honduras, Using Data Mining Methods. 2018.

Abdelaziz Merghadi, Ali P. Yunus, Jie Dou, Jim Whiteley, Binh ThaiPham,Dieu Tien Bui, Ram Avtar and Boumezbeur Abderrahmane. Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. 2020, 207

Breiman, L.,2001. Random Forests. Machine learning, 45(1): 5-32. doi:https://doi.org/10.1023/A: 10 1093340

Breiman L, Cutler A (2004) http://www.stat.berkeley.edu/users/Breiman/RandomForests/ccpapers. H-Tml

Bui DT, Pradhan B, Lofman O, Revhaug I, Dick OB (2012) Landslide susceptibility assessment in the Hoa Binh Province of Vietnam: a comparison of the Levenberg-Marquardt and Bayesian regularized neural networks. Geomorphology. doi:10.1016/j.geomorph.2012.04.023

Blais-Stevens A, Behnia P (2016) Debris flow susceptibility mapping using a qualitative heuristic method and flow-R along the Yukon Alaska Highway Corridor, Canada. Nat Hazard Earth Syst Sci 16(2):449–462.

Chung, C.F., Fabbri, A.G., 2003. Validation of spatial prediction models for landslide hazard mapping. Nat. Hazards 30,451–472.

Carrara A, Crosta G, Frattini P (2008) Comparing models of debris-flow susceptibility in the alpine environment. Geomorphology 94:353-378

Chunxiang Wang, Hideaki Marui, Gen Furuya, et al. Two Integrated Models Simulating Dynamic Process of Landslide Using GIS. 2013.

Chung, C.F., Fabbri, A.G., 2003. Validation of spatial prediction models for landslide hazard mapping. Nat. Hazards 30, 451–472.

Colkesen, I., Sahin, E.K., Kavzoglu, T., 2016. Susceptibility mapping of shallow landslides using kernel-based Gaussian process, support vector machines and logistic regression. African Earth Sciences]–>J. Afr. Earth Sci. 118, 53–64.

Chen, W., Xie, X., Wang, J., Pradhan, B., Hong, H., Bui, D.T., Duan, Z. and Ma, J., 2017. A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. Catena, 151, pp.147-160.

Cruden, D.M., Varnes, D.J., 1996. Landslide types and processes. In: Turner, A.K., Schuster, R.L. (Eds.), Landslides, Investigation and Mitigation, Special Report 247. Transportation Research Board, Washington D.C., pp. 36–75 ISSN: 0360-859X, ISBN: 030906208X.

Dai, F.C., Lee, C.F., 2002. Landslide characteristics and slope instability modelling using GIS, Lantau Island, Hong Kong. Geomorphology 42, 213–228.

D. W. Park, N. V. Nikhil, S. R. Lee. Landslide and debris flow susceptibility zonation using TRIGRS for the 2011 Seoul landslide event. 2013, 13(11):2833-2849.

Furlani, S., Ninfo, A., 2015. Is the present the key to the future? Earth-Sci. Rev. 142 (C),38–46.

Francesco Bregoli, Vicente Medina, Guillaume Chevalier, et al. Debris-flow susceptibility assessment at regional scale: Validation on an alpine environment. 2015, 12(3):437-454.

Green, D.M., Swets, J.M., 1966. Signal Detection Theory and Psychophysics. John Wiley and Sons, New York ISBN: 0-471-32420-5.

453 Guzzetti, F., Carrara, A., Cardinali, M., Reichenbach, P., 1999. Landslide hazard evaluation: a
454 review of current techniques and their application in a multi-scale study, Central Italy.
455 Geomorphology 31, 181–216.
456 Guzzetti. F, M. Galli, P. Reichenbach, et al. Landslide hazard assessment in the Collazzone area,
457 Umbria, Central Italy. 2006, 6(1):115.
458 Guzzetti, F., Galli, M., Reichenbach, P., Ardizzone, F., Cardinali, M., 2006a. Landslide hazard
459 assessment in the Collazzone area, Umbria, central Italy. Nat. Hazard. Earth Syst. Sci. 6,
460 115–131.
461 Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M., Galli, M., 2006b. Estimating the
462 quality of landslide susceptibility models. Geomorphology 81, 166–184.
463 Gomes, R. A. T., Guimaraes, R. F., Carvalho Júnior, O. A., Fernandes, N. F., and Amaral Jr., E. V.:
464 Combining spatial models for shallow landslides and debris flows prediction, Remote Sens.,
465 5,2219–2237, 2013
466 Hungr, O., Leroueil, S., Picarelli, L., 2013. The Varnes classification of landslide types, an update.
467 Landslides 11 (2), 167–194.
468 Huang Xiaohu, Guo Fei, Deng miaolin, Yi Wu, Huang Haifeng.Understanding the deformation
469 mechanism and threshold reservoir level of the floating weight-reducing landslide in the
470 Three Gorges Reservoir Area, China. Landslides DOI 10.1007/s10346-020-01435-1.
471 Haydar Y. Hussin, Veronica Zumpano, Paola Reichenbach, et al. Different landslide sampling
472 strategies in a grid-based bi-variate statistical susceptibility model. 2016, 253:508-523.
473 Imaizumi F, Masui T, Yokota Y, Tsunetaka H, Hayakawa YS, Hotta N. 2019.Initiation and runout
474 characteristics of debris flows urges in Ohya landslide scar, Japan. Geomorphology, 2019

339:58–69.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning. Springer, New York, p. 441.

Lee, S., Dan, N.T., 2005. Probabilistic landslide susceptibility mapping on the Lai Chau province of Vietnam: focus on the relationship between tectonic fractures and landslides. Environmental Geology 48, 778–787.

Linfeng Fan, Peter Lehmann, Brian McArdell, et al. Linking rainfall-induced landslides with debris flows runout patterns towards catchment scale hazard assessment. 2017, 280:1-15.

Liqiang Tong, Wensheng Qi, Guoying An,Chunling Liu. Remote sensing survey of major geological disasters in the Himalayas[J].Journal of engineering geology,2019,27(03):496.

Melton M A． The Geomorphic and Paleoclimatic Significance of Alluvial Deposits in Southern Arizona: A Reply［J］. The Journal of Geology，1965，73(1) : 102 －106.

Mariantonietta Ciurleo, Michele Calvello, Leonardo Cascini. Susceptibility zoning of shallow landslides in fine grained soils by statistical methods. 2016, 139:250-264.

Mariantonietta Ciurleo, Leonardo Cascini, Michele Calvello. A comparison of statistical and deterministic methods for shallow landslide susceptibility zoning in clayey soils. 2017, 223:71-81.

Maria Giuseppina Persichillo, Massimiliano Bordoni, Claudia Meisina, et al. Shallow landslides susceptibility assessment in different environments. 2017, 8(2):748-771.

Ma C., Deng J., Wang R. Analysis of the triggering conditions and erosion of a run off triggered debris flow in Mi yun County, Beijing, China. Landslide, 2018, DOI

10.1007/s10346-018-1080-3

P. Aleotti, R. Chowdhury. Landslide hazard assessment: summary review and new perspectives. 1999, 58(1):21-44.

Pradhan, B., 2010. Landslide susceptibility mapping of a catchment area using frequency ratio, fuzzy logic and multivariate logistic regression approaches. J. Indian Soc. Remote Sens. 38 (2), 301–320.

Pedregosa,F., Varoquaux,G., Gramfort, A., et al., 2011. Scikit-Learn: Machine Learning in Python. Journal of machine learning research, 12(10): 2825-2830.

P. Vorpahl, H. Elsenbeer, M. Märker, and B. Schröder, "How can statistical models help to determine driving factors of landslides?" Ecol. Model., vol. 239, pp. 27–39, 2012

Pourghasemi HR, Moradi HR, Fatemi Aghda SM (2013a) Landslide susceptibility mapping by binary logistic regression, analytical hierarchy process, and statistical index models and assessment of their performances. Nat Hazards 69:749–779. doi:10.1007/s11069-013-0728-5

Pourghasemi HR, Pradhan B, Gokceoglu C, Mohammadi M, Moradi HR (2013b) Application of weights-of-evidence and certainty factor models and their comparison in landslide susceptibility mapping at Haraz watershed, Iran. Arab J Geosci 6(7):2351–2365. doi:10.1007/s12517-012-0532-7

Reichenbach,P.,Rossi,M.,Malamud,B.D.,et al.,2018.A Review of Statistically-Based Landslide Susceptibility Models. Earth-Science Reviews, 180(5): 60-91. doi:https://doi.org/10.1016/j.earscirev. 2018. 03. 001

Shou-Hao Chiang, Kang-Tsung Chang, Alessandro C. Mondini, et al. Simulation of event-based landslides and debris flows at watershed level. 2011, 138(1):306-318.

519 Takahashi, T., 1978. Mechanical characteristics of debris flow. Journal of the Hydraulics Division
520 104, 1153–1169.
521 Takahashi T(2007) Debris Flow:Mechanics, Prediction and Countermeasures. Taylor & Francis,
522 448 p. ISBN 978-0-415-43552-9
523 Teigila A, et al. Comparison of Logistic Regression and Random Forests techniques for shallow
524 landslide susceptibility assessment in Giampilieri (NE Sicily, Italy)][J]. Geomorphology
525 Science Letter, 2015
526 Taskin Kavzoglu, Emrehan Kutlug Sahin, Ismail Colkesen, Selecting optimal conditioning factors
527 in shallow translational landslide susceptibility mapping using genetic algorithm,Engineering
528 Geology,Volume192, 2015, Pages 101-112.
529 Varnes DJ (1978) Slope movement types and processes, in Schuster, R.L., and Krizek, R.J., eds.,
530 Landslides: Analysis and control, National Research Council, Washington, D.C.,
531 Transportation Re- search Board, National Academy Press, Special Report 176, p. 11–33
532 Varnes, D.J., IAEG Commission on Landslides and other Mass-Movements, 1984. Landslide
533 Hazard Zonation: A Review of Principles and Practice. The UNESCO Press, Paris (63 pp).
534 van Westen CJ, Castellanos E, Kuriakose SL (2008) Spatial data for landslide susceptibility,
535 hazard, and vulnerability assessment: an overview. Eng Geol 102(3–4):112–131
536 Wilson JP, Gallant JC (2000) Digital terrain analysis. In: Wilson JP, Gallant JC (eds) Terrain
537 analysis. John Wiley & Sons, New York, pp 1–27
538 Xu WB, Yu WJ, et al. (2013) Debris flow susceptibility assessment by GIS and information value
539 model in a large-scale region, Sichuan Province (China).Natural Hazards 65(3):1379-1392.
540 Yilmaz, I., 2010. The effect of the sampling strategies on the landslide susceptibility mapping by

conditional probability and artificial neural networks. Environ. Earth Sci. 60,505–519.

Zhang, K., Wu, X., Niu, R., Yang, K. and Zhao, L., 2017. The assessment of landslide susceptibility mapping using random forest and decision tree methods in the Three Gorges Reservoir area, China. Environmental Earth Sciences, 76(11), pp.1-20.

Zhou, W., Fan.J., Tang, C., Yang, G. (2019)Empirical relationships for the estimation of debris flow runout distances on depositional fans in the Wenchuan earthquake zone. Journal of Hydrology, 577, https://doi.org/10.1016/j. jhydrol. 2019. 123932

Zhu Liang, Changming Wang, , Zhi-Min Zhang and Kaleem-Ullah-Jan Khan. A comparison of statistical and machine learning methods for debris flow susceptibility mapping. Stoch Environ Res Risk Assess (2020a) https://doi.org/10.1007/s00477-020-01851-8

Zhu Liang, Changming Wang, Songling Han, Kaleem Ullah Jan Khan, and Yiao Liu. Classification and susceptibility assessment of debris flow based on a semi-quantitative method combination of the fuzzy C-means algorithm, factor analysis and efficacy coefficient.Nat. Hazards Earth Syst. Sci., 20, 1287–1304, 2020b https://doi.org/10.5194/nhess-20-1287-2020

Zhu Liang, Wang Changming and Kaleem-Ullah-Jan Khan. Application and comparison of different ensemble learning machines combining with a novel sampling strategy for shallow landslide susceptibility mapping. Stoch Environ Res Risk Assess (2020c). https://doi.org/10.1007/s00477-020-01893-y

Table 1 Models' performance using training dataset

Metrics	Landslide	Debris flow
TP（%）	88.71	87.80

TN（%）	91.89	88.89
FP（%）	11.29	12.20
FN（%）	8.11	11.11
Sensitivity（%）	91.62	88.77
Specificity（%）	89.06	87.93
Accuracy（%）	90.65	88.57
AUC	0.976	0.967

561 **Table 2** Models' performance using verification dataset

Metrics	Landslide	Debris flow
TP（%）	85.56	85.71
TN（%）	89.09	84.62
FP（%）	14.44	14.29
FN（%）	10.91	15.38
Sensitivity（%）	88.69	84.79
Specificity（%）	86.05	85.55
Accuracy（%）	87.33	85.17
AUC	0.902	0.892

562 **Table 3** Variables importance assigned for landslide

Test group	Slope angle	Distance to fault	Plan curvature	Topographic wetness index	Distance to road	Maximum elevation difference	Profile curvature	Elevation

Landslide	0.21	0.19	0.17	0.13	0.08	0.07	0.06	0.05

563 **Table 4** Variables importance assigned for debris flow

Test group	Main channel length	Roundness	Slope angle	Elevation	Maximum elevation difference	Melton	Basin area
Debris flow	0.25	0.16	0.14	0.13	0.12	0.1	0.1

564 **Table 5** The overlap number of debris flow and landslide height and very high-class mapping units

Debris flow \ Landslide	Very low	Low	High	Very high
High	3/23	1/23	5/23	12/23
Very high	2/26	2/26	8/26	11/26

565 **Table 6** Statistical parameters of the two models

Statistical parameters \ Model	Model 1	Mode 2
KMO	0.766	0.643
Sig.	0.001	0.003

566 **Table 7** The correlation coefficients between common factors and primitive variables

Factor	F1	F2	F3
NDVI	0.386	-0.336	-0.621
Basin area	0.897	-0.007	0.041
Main channel length	0.984	0.046	-0.023
Slop angle	-0.223	0.829	0.455
Maximum elevation difference	0.744	0.66	0.011

Rainfall	-0.768	0.33	0.201
Average gradient of main channel	-0.753	0.544	0.106
Drainage density	-0.844	0.06	0.015
Roundness	0.331	0.14	0.818
Elevation	0.133	0.846	0.382
Distance to fault	-0.16	0.211	0.421
Melton	-0.625	0.737	0.149
Contribution rate (%)	41.2	24.7	16.7
Accumulative contribution (%)	41.2	65.9	83.6

567 **Table 8** The correlation coefficients between common factors and primitive variables

Factor	C1	C2	C3	C4
NDVI	0.042	-0.079	-0.279	-0.813
Basin area	0.802	-0.344	0.057	0.009
Main channel length	0.885	0.126	-0.196	0.227
Slop angle	0.009	0.748	0.58	-0.057
Maximum elevation difference	0.801	0.434	-0.128	0.144
Rainfall	0.197	-0.076	-0.487	0.637
Average gradient of main channel	-0.744	0.205	0.15	-0.23
Drainage density	-0.776	-0.176	-0.267	0.117
Roundness	-0.014	0.022	0.896	-0.002
Elevation	0.34	0.746	0.25	0.326

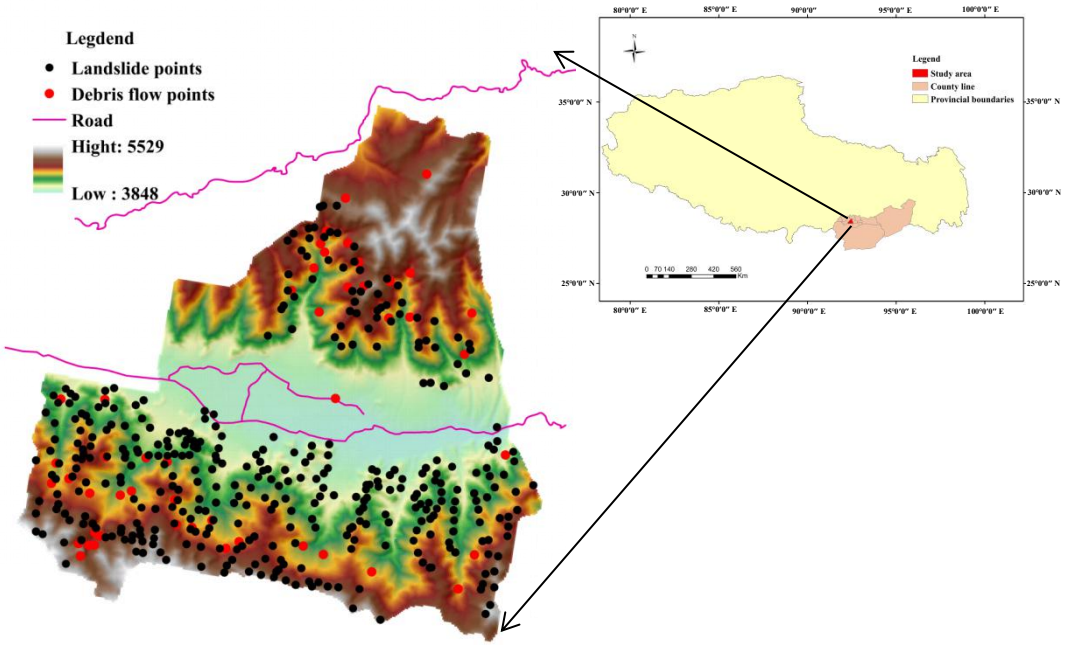
Distance to fault	0.31	0.289	-0.344	0.757
Melton	-0.182	0.932	-0.192	0.061
Contribution rate (%)	29.2	20.3	15.2	15.8
Accumulative contribution (%)	29.2	49.5	64.7	80.5

568 **Table 9** The overlap number of landslide and debris flow height and very-high class mapping units

Landslide \ Debris flow	Very low	Low	High	Very high
High	36/167	33/167	25/167	43/167
Very high	48/179	40/179	25/179	28/179

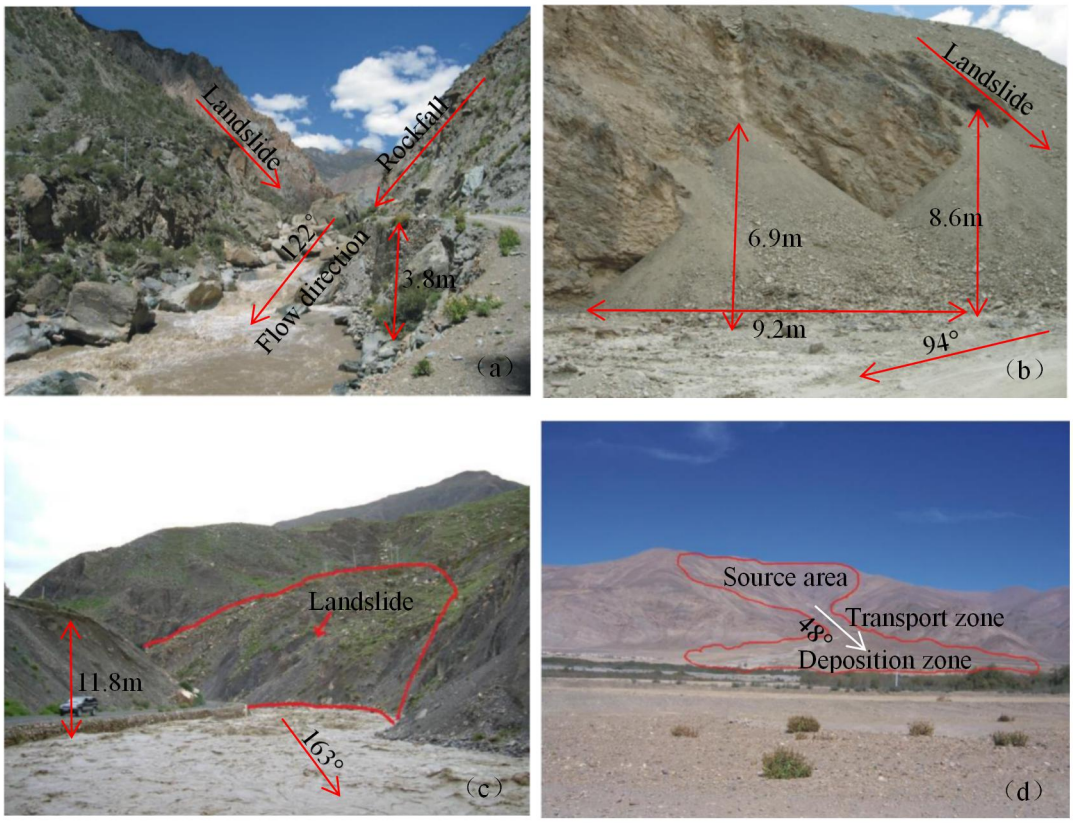
569 **Table 10** Comprehensive evaluation of landslide-debris flow susceptibility

Landslide \ Debris flow	Low or Very low	High or Very high
Low or Very low	Low	Moderate
High or Very high	Moderate	High



570

571 **Fig.1.** Location map of the study area showing landslide and debris flow inventory.

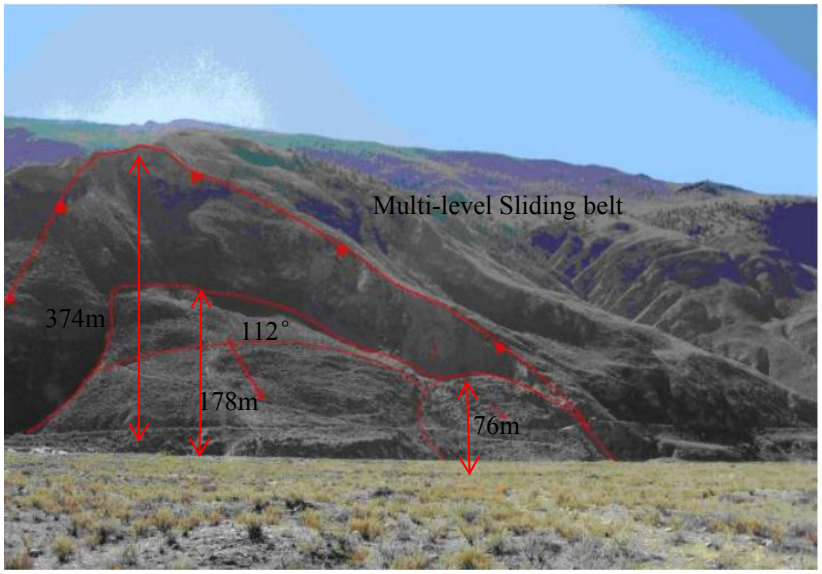


572

573

574 **Fig.2.** Photos of landslide or debris flow: **(a)** Lunba landslide in a tributary; **(b)** Zhenqiong landslide in

575 Jiayu village; **(c)** Debris flow in Misha Township; **(d)** Debris flow in Lelong Village.



576

577 **Fig.3.** Multistage landslide in Xiongqu village

578

579 **Fig.4.** Stereo remote sensing map of landslides in Longzi Township (Tong et al., 2019): **（a）** Landslides

580 in Longzi town; **(b)** Landslides in Malu town.

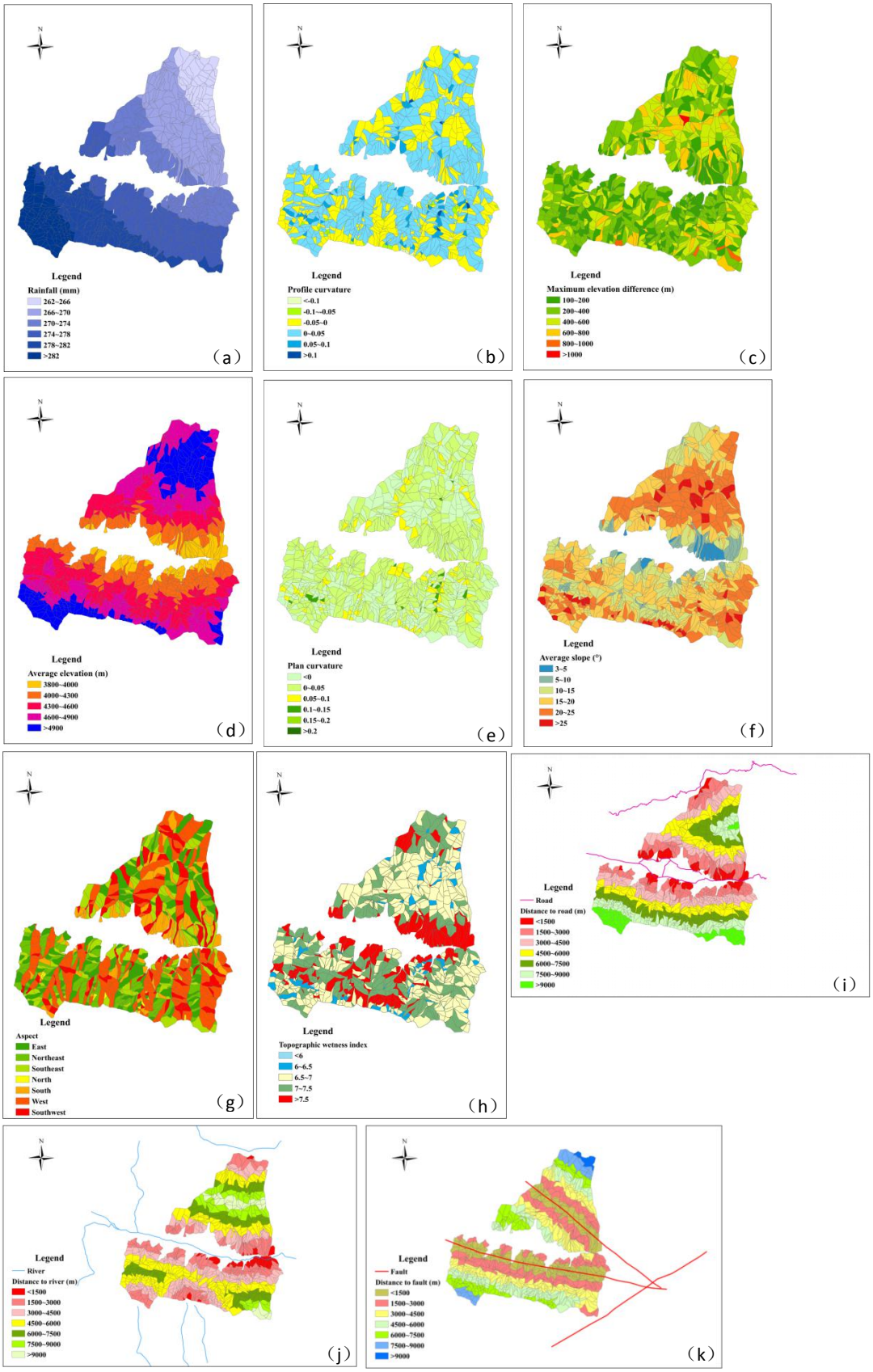
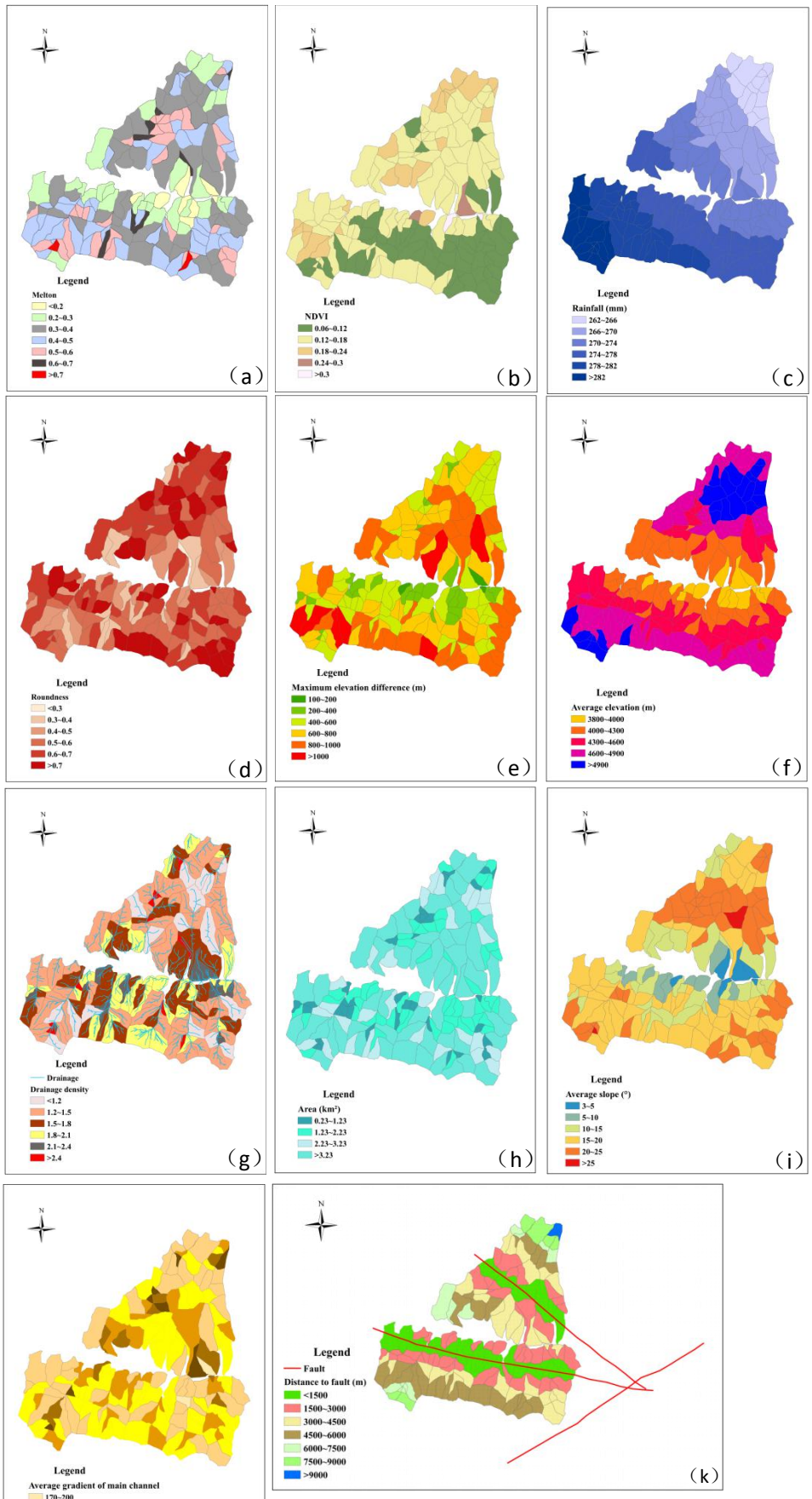


Fig.5. Study area thematic maps for landslide: （**a**）Rainfall; （**b**）Profile curvature; （**c**）Maximum elevation difference; （**d**）Average elevation; （**e**）Plan curvature; （**f**）Average slope; （**g**）Aspect;

584 （**h**）Wetness;（**i**）Distance to road;（**j**）Distance to river;（**k**）Distance to fault.

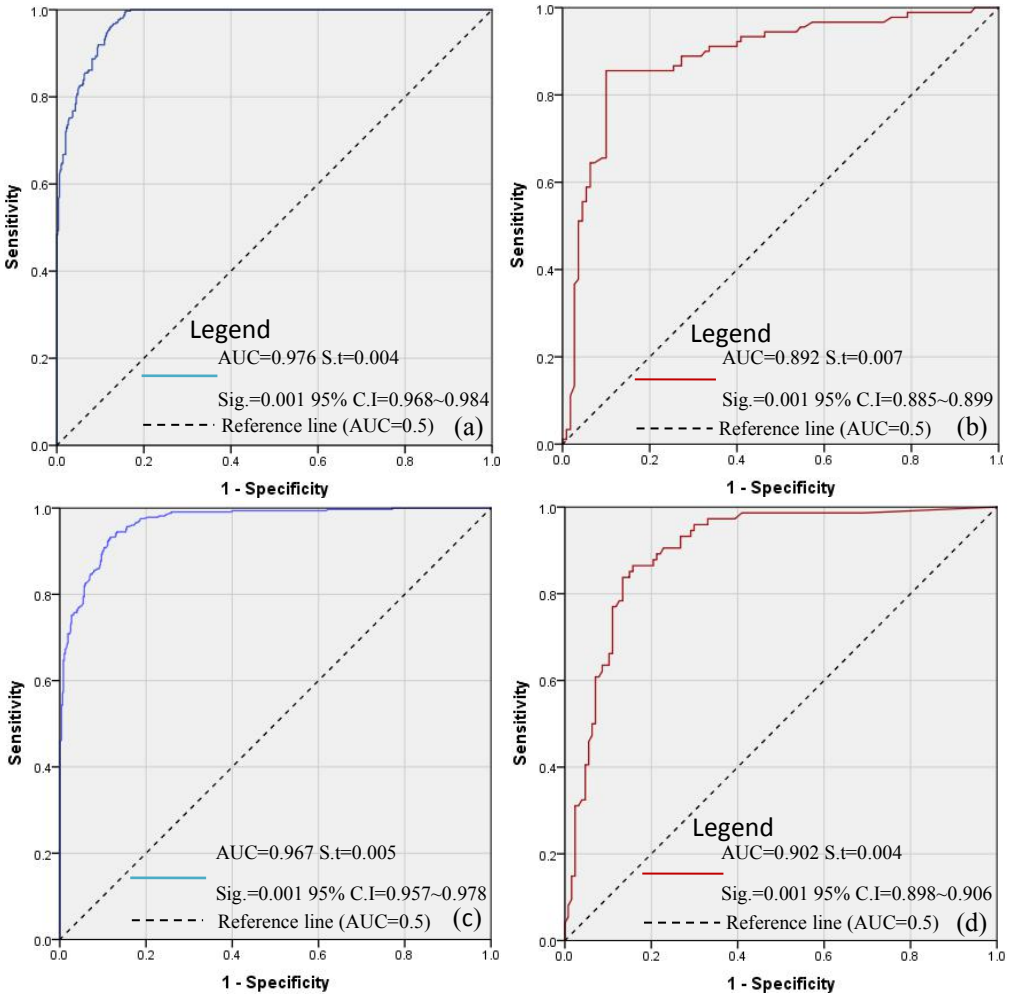

Legend
Melton
<0.2
0.2~0.3
0.3~0.4
0.4~0.5
0.5~0.6
0.6~0.7
>0.7
(a)
Legend
NDVI
0.06~0.12
0.12~0.18
0.18~0.24
0.24~0.3
>0.3
(b)
Legend
Rainfall (mm)
262~266
266~270
270~274
274~278
278~282
>282
(c)
Legend
Roundness
<0.3
0.3~0.4
0.4~0.5
0.5~0.6
0.6~0.7
>0.7
(d)
Legend
Maximum elevation difference (m)
100~200
200~400
400~600
600~800
800~1000
>1000
(e)
Legend
Average elevation (m)
3800~4000
4000~4300
4300~4600
4600~4900
>4900
(f)
Legend
Drainage
Drainage density
<1.2
1.2~1.5
1.5~1.8
1.8~2.1
2.1~2.4
>2.4
(g)
Legend
Area (km²)
0.23~1.23
1.23~2.23
2.23~3.23
>3.23
(h)
Legend
Average slope (°)
3~5
5~10
10~15
15~20
20~25
>25
(i)
Legend
Average gradient of main channel
170~200
200~400
400~600
600~800
800~1000
>1000
(j)
Legend
Fault
Distance to fault (m)
<1500
1500~3000
3000~4500
4500~6000
6000~7500
7500~9000
>9000
(k)

585

586 **Fig.6.** Study area thematic maps for debris flow: （**a**）Melton; （**b**）NDVI; （**c**）Rainfall; （**d**）Roundness;

587 （**e**）Maximum elevation difference; （**f**）Average elevation; （**g**）Drainage density; （**h**）Area; （**i**）

588 Average slope; （**j**）Average gradient of main channel; （**k**）Distance to fault.



589

590 **Fig. 7.** Analysis of ROC curve for the two susceptibility maps: **(a)** Success rate curve of landslide using

591 the training dataset; **(b)** Prediction rate curve of landslide using the validation dataset; **(c)** Success rate

592 curve of debris flow using the training dataset; **(d)** Prediction rate curve of debris flow using the

593 validation dataset.

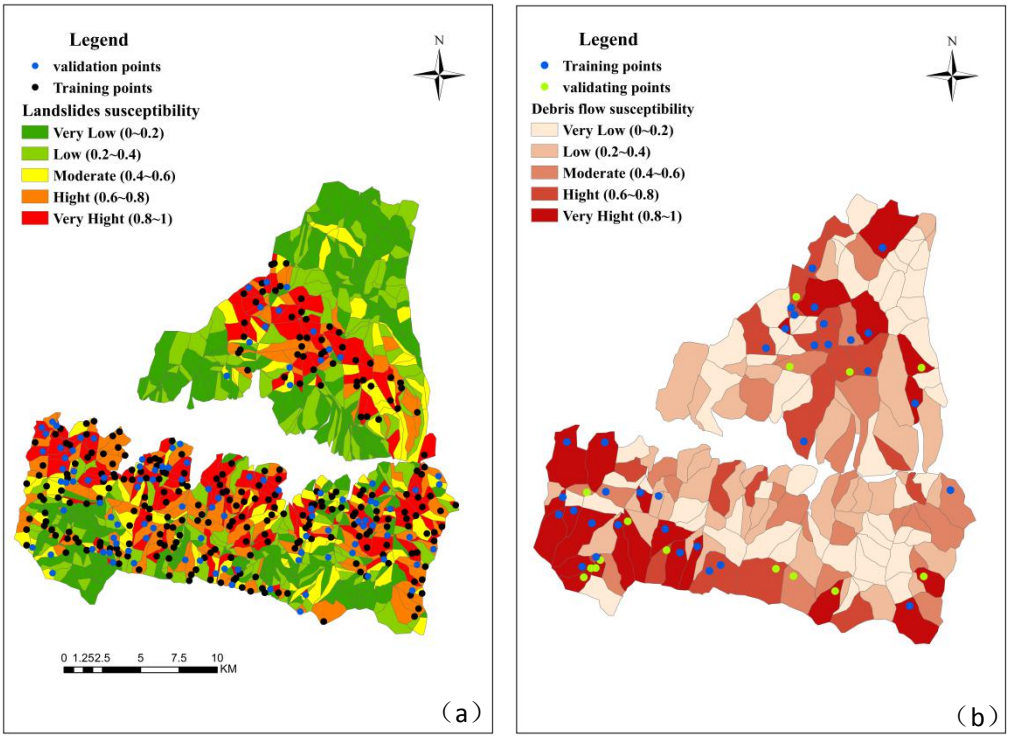


Fig.8. Susceptibility maps: （**a**） Landslide susceptibility zoning map; （**b**） Debris flow susceptibility zoning map.

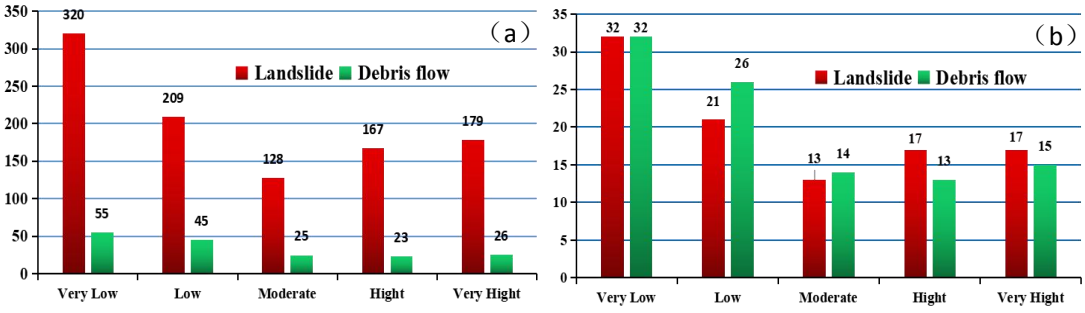
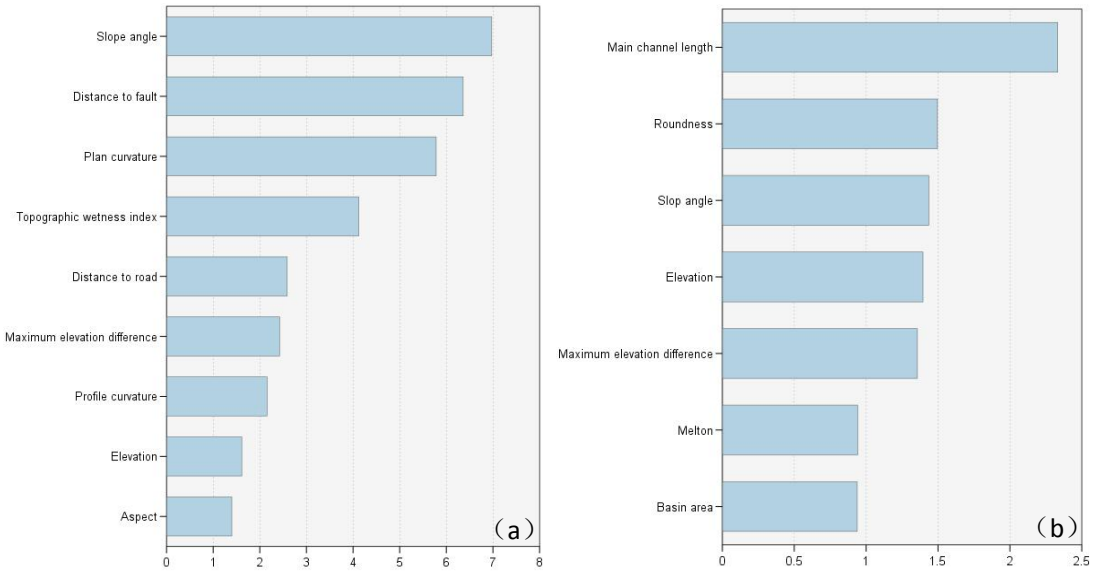
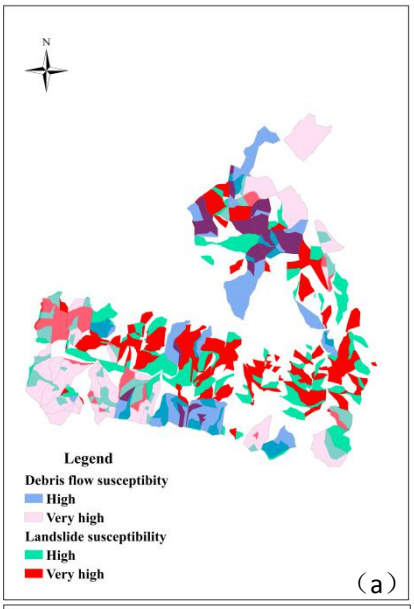


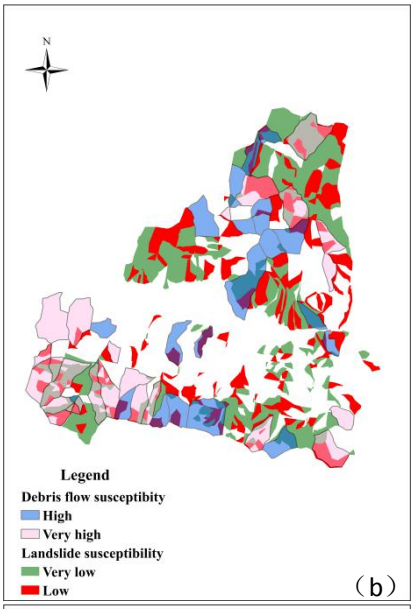
Fig. 9. Numbers and percentage of units in different susceptibility classes for landslide and debris flow: **(a)** Numbers of units in different susceptibility classes for landslide and debris flow; **(b)** Percentages of different susceptibility classes for landslide and debris flow.

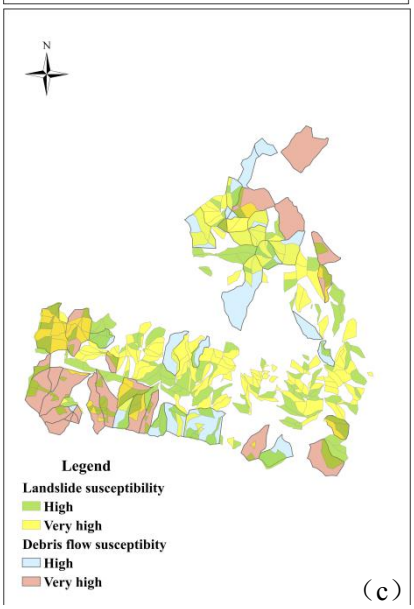
601 

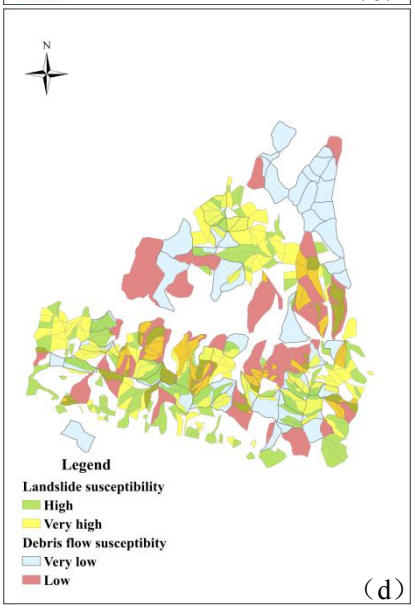
602 **Fig.10.** Parametric importance graphics obtained from RF model: **(a)** Parametric importance graphics

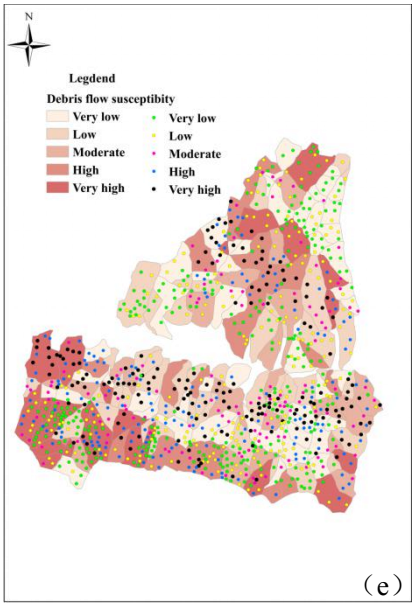
603 of landslide; **(b)** Parametric importance graphics of debris flow.


N
Legend
Debris flow susceptibity
High
Very high
Landslide susceptibility
High
Very high
(a)


N
Legend
Debris flow susceptibity
High
Very high
Landslide susceptibility
Very low
Low
(b)


N
Legend
Landslide susceptibility
High
Very high
Debris flow susceptibity
High
Very high
(c)


N
Legend
Landslide susceptibility
High
Very high
Debris flow susceptibity
Very low
Low
(d)


N
Legdend
Debris flow susceptibity
Very low
Low
Moderate
High
Very high
Very low
Low
Moderate
High
Very high
(e)

604

Fig.11. Landslide-debris flow susceptibility maps: **(a)** Height and very high-class watershed units with high or very high slope units; **(b)** High or very high-class watershed units with low or very low slope units; **(c)** High or very high-class slope units with high or very high-class watershed units; **(d)** Mapping units.