Stochastic Environmental Research and Risk Assessment
https://doi.org/10.1007/s00477-024-02753-9

ORIGINAL PAPER

Hybrid method for rainfall-induced regional landslide susceptibility mapping

Shuangyi Wu[1] · **Huaan Wang**[2] · **Jie Zhang**[3] · **Haijun Qin**[4]

Accepted: 29 May 2024

Abstract
Landslide susceptibility maps can provide important information for managing regional landslide risks. Traditionally, data-driven and physically-based models are widely used for rainfall-induced landslide susceptibility mapping, but each method has limitations. In this study, a hybrid method that integrates a data-driven model and a physically-based model is proposed for rainfall-induced landslide susceptibility mapping, where the uncertainty in the soil properties can be explicitly considered. The proposed method is illustrated with landslide susceptibility mapping in Shengzhou County, Zhejiang Province, China. Logistic regression is used as the data-driven model, and the regional assessment of rainfall-induced landslides model (RARIL) is used as the physically-based model. Three hybrid models are developed. Hybrid model I, which considers soil parameters uncertainty, is compared with hybrid models II and III, which do not consider it. Results indicate that all the three hybrid models outperform the conventional logistic regression and RARIL models. Notably, hybrid model I, which considers the soil parameters uncertainty, outperforms hybrid models II and III, which do not consider it.

Keywords Logistic regression · Hybrid model · Soil parameters uncertainty · Landslide susceptibility mapping

1 Introduction

Rainfall-induced landslides often cause enormous damage and property losses worldwide (e.g., Zhang et al. 2011; Wei et al. 2019; Lu et al. 2023; Mondini et al. 2023). As these landslides are often distributed over large regions, developing landslide susceptibility maps (LSMs) has been considered a useful tool for the risk management of these landslides (e.g., Long et al. 2021; Gong et al. 2022; Su et al. 2023). At present, two types of models have been widely used for developing the LSMs, i.e., the data-driven models rely on the statistical analysis of previous landslide events, and physically-based models through regional stability analysis of the slopes (e.g., Merghadi et al. 2020; Ji et al. 2022). The data-driven methods consider the evolution of landslides to predict slope failure (Cascini et al. 2022), while also assuming that conditions resembling those of historical landslide areas may evolve into future slope instability (e.g., Yalcin., 2008; Gong et al. 2022), and various techniques have been used for learning it, such as logistic regression (e.g., Zhao et al. 2019), random forest (e.g., Sun et al. 2021; Fu et al. 2023), convolutional neural network (e.g., Fang et al. 2020; Wang et al. 2021a; Wei et

✉ Jie Zhang
cezhangjie@tongji.edu.cn

Shuangyi Wu
sywu@tongji.edu.cn

Huaan Wang
wanghuaan@gedi.com.cn

Haijun Qin
qinhaijun@tongji.edu.cn

[1] Department of Geotechnical Engineering, Tongji University, 1239 Siping Road, Shanghai 200092, China

[2] China Energy Engineering Group Guangdong Electric Power Design Institute Co., Ltd., Guangzhou 510663, China

[3] Key Laboratory of Geotechnical and Underground Engineering of Ministry of Education, Department of Geotechnical Engineering, Tongji University, 1239 Siping Road, Shanghai 200092, China

[4] China Railway Design Co., Tianjin 300142, China

Published online: 18 September 2024

al. 2024). However, the applicability of data-driven models may be limited by the quality and quantity of available historical landslide data. Unlike the data-driven models, the physically-based model typically consists of a hydrological and a slope stability model (e.g., Vieira et al. 2018), which allows realistic modeling of the physical mechanism of slope failure. Several physically-based models have been developed for rainfall-induced regional LSMs, such as shallow slope stability model (e.g., Montgomery and Dietrich 1994; Abraham et al. 2023), transient rainfall infiltration and grid-based regional slope-stability model (e.g., Baum et al., 2002; Ma et al. 2021; Wei et al. 2021) and the regional assessment of rainfall-induced landslides model (RARIL) (Yang et al. 2022). Nevertheless, physically-based models mostly are created by simplifying the real physical process and giving assumptions (Medina et al. 2021). Additionally, the soil parameters necessary for physically-based models on a regional scale could be difficult to obtain (e.g., Sun et al. 2024) and are often treated as fixed values during the application of such models (e.g., Pradhan and Kim 2016).

To overcome the limitations of the two types of models, the hybrid model, which intends to take advantage of the two types of models, has also been explored recently as a promising tool to achieve more interpretable and accurate LSMs (e.g., Nguyen et al. 2022; Wei et al. 2023). For instance, Nguyen et al. (2022) adopted a matrix-based approach to combine the outcomes of data-driven and physically-based models. Wei et al. (2023) integrated the factor of safety (FOS) value derived from the physically-based model into the data-driven model, serving as a new conditioning factor to replace the initial geological conditioning factor. These hybrid models improved the interpretability of landslide susceptibility analysis and produced more accurate LSMs. Nevertheless, in these hybrid models, the soil parameters uncertainty is not considered. The soil properties in a region, are indeed hard to determine accurately and could be associated with a significant amount of uncertainty (e.g., Jelínek and Wagner 2007; Cao et al. 2016).

The object of this study is to propose a novel hybrid model for landslide susceptibility analysis with explicit consideration of the uncertainty in the soil properties. This paper first introduces the framework for constructing hybrid models. Then, the data-driven model and physically-based model is introduced, followed by the three hybrid models, i.e., one model considering the soil parameter uncertainty, another model using the same data-driven framework but does not consider soil parameter uncertainty, and a matrix-based model also does not consider soil parameter uncertainty. Finally, the conventional data-driven, physically-based model and the three hybrid models are illustrated and compared with a case. We believe that the proposed method offers a new way to combine the advantages of data-driven and physically-based models for more accurate rainfall-induced landslide susceptibility assessment.

2 Methodology

Model hybridizing provides a novel method for improving landslide susceptibility assessment. As the data-driven model is trained to find the relationship between conditioning factors and landslide occurrence, one possible way to develop the hybrid model is to train the data-driven model by considering the prediction from the physically-based model as an additional conditioning factor. As such, the information provided by the physically-based model can naturally be integrated into the hybrid model. To consider the soil parameters uncertainty in the physically-based model, the uncertain soil parameters can be treated as random variables, and the stability of the slopes can be thus measured through the failure probability. In such a case, the failure probability of the landslide calculated through a physically-based model can be considered as the additional conditioning factor to be incorporated in the conventional data-driven model.

To implement the above idea, a data-driven model and a physically-based model should be first selected. As an illustration, the logistic regression (e.g., Yilmaz 2009; Zhan et al. 2023) will be used in this study to develop the data-driven model as it is simple and be widely used, and RARIL (Yang et al. 2022) will be used as the physically-based model for its ability to explicitly consider uncertainty in the soil properties through reliability analysis. Figure 1 shows the process of the proposed hybrid models, which consists of three steps, i.e., construction of a data-driven module, establishment of a physically-based model, and development of the hybrid model through the combination of the two types of models. In the following, each of the above steps will be introduced.

2.1 Data-driven model

To develop the hybrid model, a data-driven technique should be first selected. As an illustration, the logistic regression will be used in this study as the data-driven model. In the logistic regression, the relationship between the probability of landslide occurrence and the conditioning factors (e.g., Zhao et al. 2019) can be expressed as follows:

$$p = \frac{1}{1+e^{-Z}} = \frac{e^{Z}}{1+e^{Z}} \tag{1}$$

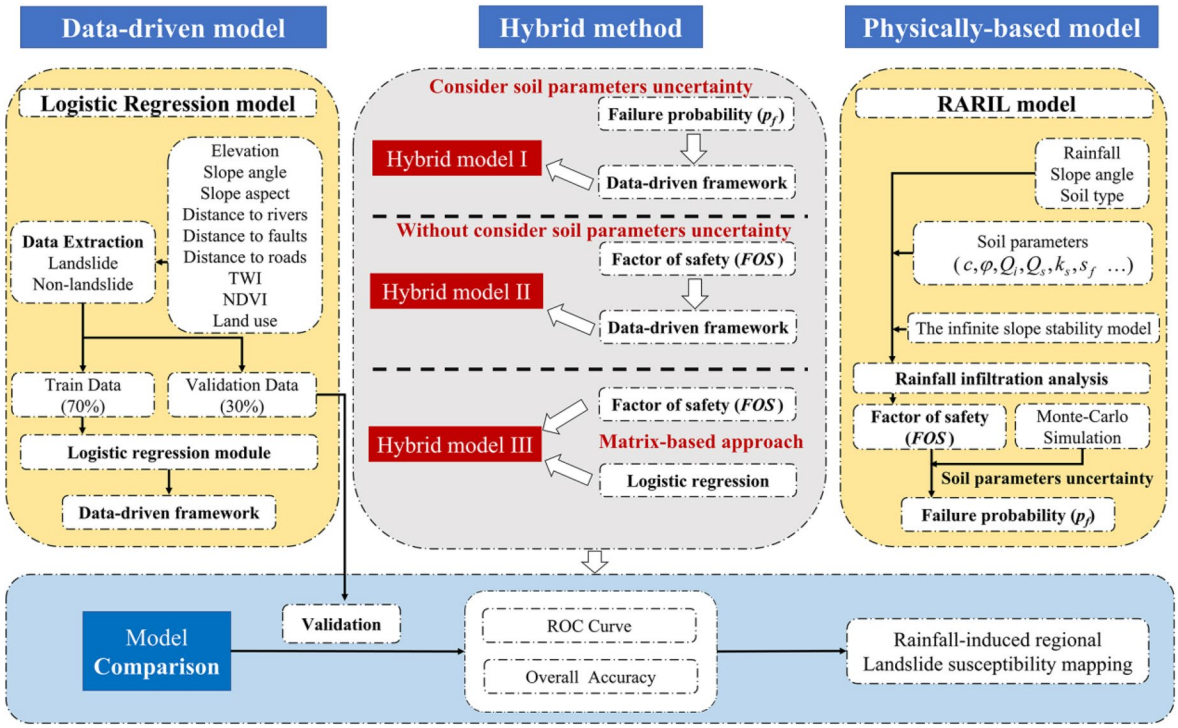

Fig. 1 Illustration of the development process of hybrid models

where p is the estimated landslide occurrence probability, and Z is the linear combination of conditioning factors as follows:

$$Z = w_0 + \sum_{i=1}^{n} w_i x_i \tag{2}$$

where x_i is the ith conditioning factor, w_i is the regression coefficient of the ith conditioning factor, and n is the number of conditioning factors.

To calibrate Eq. (1), a landslide inventory should be first compiled. Then, the parameters in Eq. (2) can be calibrated through the method like maximum likelihood estimation (e.g., Chen et al. 2016). To standardize the raw data in regression model, the Min-Max normalization method can be adopted (e.g., Zhang et al. 2021) as shown below:

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{3}$$

where x is the raw data, x^* is the normalized data, and $x_{\max}$ and $x_{\min}$ are the maximum and minimum values of the conditioning factor x.

2.2 Physically-based model

As an illustration, RARIL, as suggested by Yang et al. (2022), is used as the physically-based model in this study. In RARIL model, the digital terrain is first partitioned into square grids with 30 m resolution, and the stability of each grid is assessed through slope stability analysis during the rainfall infiltration process. In particular, the modified Green-Ampt model (Zhang et al. 2014) is used for rainfall infiltration analysis, and the infinite slope stability model (Fig. 2a) (e.g., Zhang et al. 2014; Lu et al. 2023) is used to calculate the FOS of each cell. Figure 2b shows the water content profile of Green-Ampt model. In the modified Green-Ampt model, the calculation of the infiltration rate after ponding occurs, i, is performed (Chen and Young 2006) as follows:

$$i = k_s\left[\frac{s_f(Q_s - Q_i)}{I} + \cos\alpha\right] \tag{4}$$

where k_s is the soil saturated permeability, s_f is the suction head at the wetting front, Q_s is the water content of the soil in the wetted zone, Q_i is the initial water content of the soil, and α is the slope angle. Based on Eq. (4), the cumulative infiltration, I, can be solved through a two-step procedure (Zhang et al. 2014). As presented in Fig. 2b, the depth of the wetting front, Z_w, can then be calculated as follows:

$$Z_w = \frac{I}{(Q_s - Q_i)\cos\alpha} \tag{5}$$

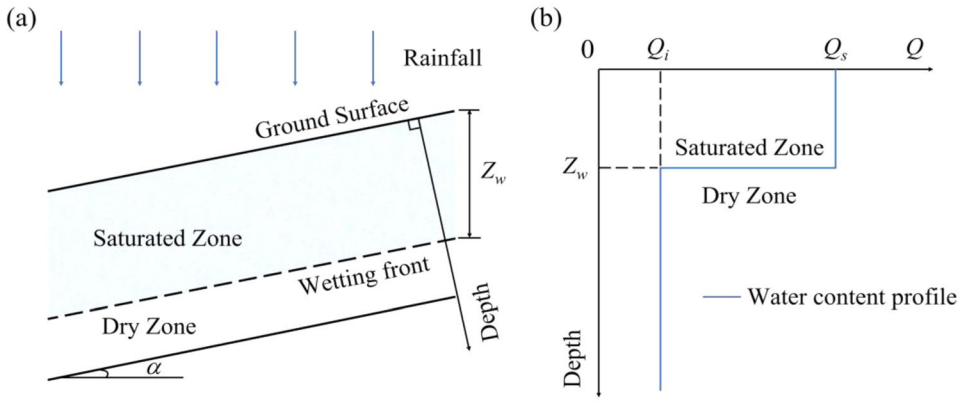

Fig. 2 (**a**) The infinite slope stability model under rainfall infiltration. Modified from Zhang et al. 2014, (**b**) The water content profile of the modified Green-Ampt model. Modified from Kim et al. 2014

Assuming that rainfall-induced landslides occur along the wetting front (e.g., Cho and Lee 2002; Vanacker et al. 2003; Zhang et al. 2014; Zhang et al. 2020; Chen et al. 2021), the FOS of the slope, F_s, can then be generated using the following equation (Fredlund et al. 1978):

$$F_s = \frac{c+\gamma_s Z_w \cos^2\alpha \tan\varphi + s_f \gamma_w \tan\varphi}{\gamma_s Z_w \sin\alpha \cos\alpha} \tag{6}$$

where c and φ are the effective cohesion and effective friction angle of the soil, γ_s is the unit weight of soil, and γ_w is the unit weight of water.

One challenge for applying the physically-based model in a region is that the soil parameters are difficult to determine accurately (e.g., Wang et al. 2015). Let $\boldsymbol{\theta}$ denote uncertain soil parameters in the RARIL model, and $f(\boldsymbol{\theta})$ denote the probability density function (PDF) of $\boldsymbol{\theta}$. Let $F_s(\boldsymbol{\theta}, t)$ denote the FOS of the slope at time t evaluated based on Eq. (6). The failure probability at time t, i.e., $p_f(t)$, can then be expressed below (e.g., Tobutt 1982; Zhang et al. 2014):

$$p_f(t) = \iiint [F_s(\theta, t)]\, f(\theta) \mathrm{d}\theta \tag{7}$$

where $J[F_s(\boldsymbol{\theta}, t)]$ is an indicator function that measures the stability of slope defined as follows:

$$J[F_s(\theta, t)] = \begin{cases} 0, F_s(\theta, t) > 1 \\ 1, F_s(\theta, t) \leqslant 1 \end{cases} \tag{8}$$

Let $\boldsymbol{\theta}^k$ denote the kth sample of $\boldsymbol{\theta}$. The failure probability in Eq. (7) can then be estimated based on Monte Carlo simulation as follows:

$$p_f(t) \approx \frac{1}{N} \sum_{k=1}^{N} J[F_s(\theta^k, t)] \tag{9}$$

where N is the number of samples.

2.3 Hybrid model

The purpose of the hybrid model is to take advantage of both the data-driven and physically-based models (e.g., Oliveira et al. 2017; Strauch et al. 2019). In this study, the prediction from the physically-based model is first considered as an additional conditioning factor for constructing the data-driven model, through which the information obtained through the physically-based model can be taken into account in the hybrid model. When using logistic regression to develop the hybrid model, Eq. (2) can be revised as follows:

$$Z = w_0 + \sum_{i}^{n} w_i x_i + w_{n+1} x_{n+1} \tag{10}$$

where x_{n+1} represents the additional conditioning factor in the hybrid model. Two hybrid methods are first considered to develop the hybrid model, i.e., hybrid model I, where the failure probability calculated based on Eq. (9) is considered as the additional conditioning factor x_{n+1}, and hybrid model II, where the FOS calculated based on Eq. (6) is considered as an additional conditioning factor x_{n+1}. To develop the two hybrid models, one can first evaluate the failure probability or FOS of the slopes using the physically-based model. Then, the hybrid model can be developed based on logistic regression as given by Eq. (10) through taking failure probability or FOS as the additional conditioning factor.

Furthermore, to provide another comparison, this paper also adopts a matrix-based hybrid method (Hybrid method III) by integrating the data-driven and physically-based classification results. This method has been frequently used in previous studies by constructing a matrix where rows represent landslide susceptibility classes derived from the physically-based model, and columns represent landslide susceptibility classes obtained from the data-driven models (e.g., Chowdhury and Flentje 2003; Nguyen et al. 2022; Xue et al. 2024). As shown in Fig. 3, each landslide

susceptibility results are further divided into five classes: Very High, High, Middle, Low, and Very Low. In this study, the matrix-based hybrid method results are considered as the lower-class value between the landslide susceptibility class from data-driven and physically-based models.

2.4 Model evaluation

In this study, the overall accuracy, the receiver operating characteristic curve (ROC), and the relative landslide density index (R-index) will be used as quantitative measures to compare the prediction capacity of the different landslide susceptibility analysis models (e.g., Shahabi et al. 2014; Wei et al. 2021; Miao et al. 2023). When comparing the model predictions with the observations, it is considered a true positive (TP) case if the landslide does occur at the predicted landslide location. If the landslide does not occur at the predicted landslide location, it is considered a false positive (FP) case. If the landslide does not occur at the predicted non-landslide location, it is considered a true negative (TN) case. If the landslide occurs at the predicted non-landslide location, it is considered a false negative (FN) case. Using these definitions, the overall accuracy, true positive rate (TPR), and false positive rate (FPR) can be determined below:

$$\text{Overall Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \tag{11}$$

$$\text{TPR} = \frac{\text{TP}}{\text{TP}+\text{FN}} \tag{12}$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP}+\text{TN}} \tag{13}$$

As can be seen from the above definition, the overall accuracy is a measure of model correctness. It represents the proportion of correct predicted values made by the model (e.g., Liu et al. 2023). The TPR describes the proportion of landslide locations that are classified correctly as landslide locations by the model, while the FPR represents the proportion of non-landslide locations that the model incorrectly classifies as landslide locations (e.g., Cantarino et al., 2019). In this study, the predictions from the landslide susceptibility analysis models are in terms of probabilities of landslides. To assess the prediction capacity of different models, a probability threshold should be first specified, i.e., a landslide is predicted to occur if the probability is greater than the probability threshold and vice versa. As the probability threshold changes, the values of TP, FP, TN, and FN also change, and hence the values of overall accuracy, TPR, and FPR value change. The relationship between TPR and FPR is often known as ROC. From the ROC curve, the area under the curve (AUC) can be calculated (e.g., Cantarino et al., 2019), which is between 0 and 1. The value of AUC is often used to access the prediction capacity of a model (e.g., Bradley 1997) and is classified as average (0.6–0.7), good (0.7–0.8), very good (0.8–0.9), and excellent (0.9-1) (e.g., Woodard et al. 2023).

The R-index, which measures the relative landslide density, is often used to validate the quality of LSMs (Baeza and Corominas 2001), as defined below:

$$\text{R} = (n_i/N_i)/\sum(n_i/N_i) \times 100 \tag{14}$$

where n_i and N_i are the number of landslide samples and total cells in the susceptibility class i. Based on Eq. (14), the R-index can be calculated for each susceptibility class. It is expected that, when the susceptibility class is low, the chance of occurrence of landslides in the low susceptibility region is also low, and hence the R-index should be small. On the other hand, when the susceptibility class is high, the R-index should also be large. Therefore, the value of R-index is expected to increase as the susceptibility class changes from low to high, so it can be used to measure the classification capacity of a LSM (Shahabi et al. 2014). In the following, the conventional data-driven, physically-based model and the three hybrid models for developing the landslide susceptibility assessment work will be illustrated with a case and validated with the above evaluation methods.

		Landslide susceptibility class from physically-based model				
		Very High	High	Middle	Low	Very Low
Landslide susceptibility class from data-driven model	Very High	Very High	High	Middle	Low	Very Low
	High	High	High	Middle	Low	Very Low
	Middle	Middle	Middle	Middle	Low	Very Low
	Low	Low	Low	Low	Low	Very Low
	Very Low	Very Low	Very Low	Very Low	Very Low	Very Low

Fig. 3 Landslide prediction matrix integrating data-driven and physically-based models

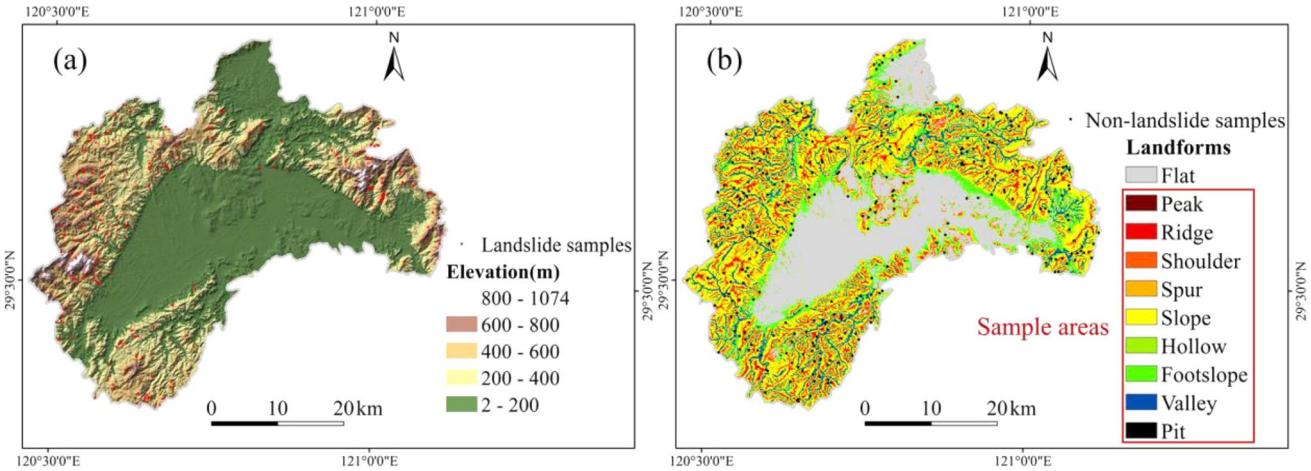

Fig. 4 (**a**) Elevation and landslide samples distribution in the Shengzhou County, (**b**) Non-landslide samples distribution

3 Case study

3.1 Engineering background

Covering a total area of 1789.6 km^2, Shengzhou County is located in the east of Zhejiang province, China. Figure 4a shows the elevation of the terrain in Shengzhou County. As shown in Fig. 4a, its terrain tilts from northwest to southeast with altitude varying from 1074 m to 2 m. Rainfall serves as the major triggering factor for landslides in this region. On August 2nd, 2018, the typhoon "Jongdari", which brought extremely heavy precipitation (e.g., Zhan and Xie 2022), attacked Shengzhou County, leading to many landslides. Landslide inventory is an essential component of LSMs (e.g., Harp et al. 2011; Wang et al. 2021b). Based on Google Earth and GF-1 remote sensing data, 240 landslide polygons were detected by comparing the satellite figures before and after the typhoon-related rainfall events. For further analysis, all landslide polygons were converted into landslide grids (30 m resolution). These 318 resulting grids are then serves as representative landslide samples, as illustrated in Fig. 4a. To build a dataset for model analysis, an equal number of non-landslide samples are also needed (e.g., Liu et al. 2022; Chang et al. 2023). Considering that there are large flat terrains in Shengzhou County with almost no observed landslides, we first employed the r.geomorphon module to exclude these areas (Jasiewicz and Stepinski 2013). Furthermore, we create a 100 m buffer zone around landslide points (Süzen and Doyuran 2004), thus ensuring most of the non-landslide sample area exclusion from regions of landslide occurrences (Fig. 4b). The combined landslide/non-landslide dataset is further randomly split into two parts, with 70% for training and the remaining 30% for validation. (e.g., Razavi-Termeh et al. 2023; Saha et al. 2023).

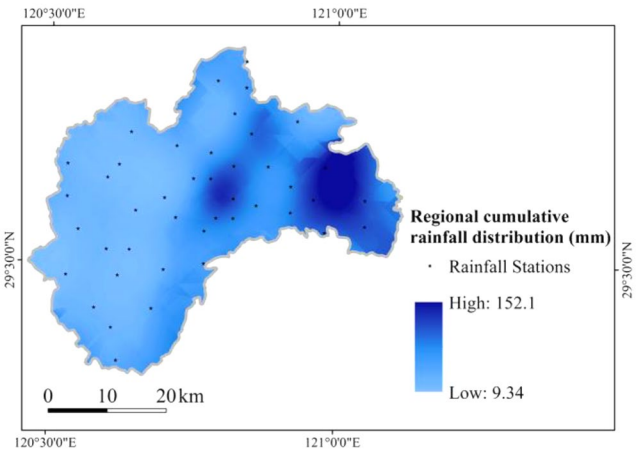

Fig. 5 Distribution of the rainfall stations and 29-hours cumulative rainfall data

3.2 Rainfall data

Shengzhou Meteorological Bureau set up a monitoring network consisting of 43 rainfall stations in the study area. Figure 5 shows the distribution of 43 rainfall stations in this region. Figure 6 shows the regional maximum hourly precipitation in all rainfall stations and maximum cumulative precipitation collected by rainfall station in this region. As presented in Fig. 6, the rainfall in this region lasts nearly 29 h and brings a maximum cumulative precipitation exceeding 150 mm.

Based on rainfall data collected from 43 rainfall monitoring stations, as shown in Fig. 5, the distribution of rainfall over the region can be derived through the Kriging interpolation method (e.g., Cressie 1990; Olea,1999; Chen et al., 2014; Zhao et al. 2019; Wang et al. 2023). Let $R(\mathbf{u}_i)$ denote the cumulative rainfall at the location of the ith rainfall station $\mathbf{u}_i$, and let $R(\mathbf{u}_0)$ represent the rainfall data at a location $\mathbf{u}_0$ where the cumulative rainfall needs to be interpolated.

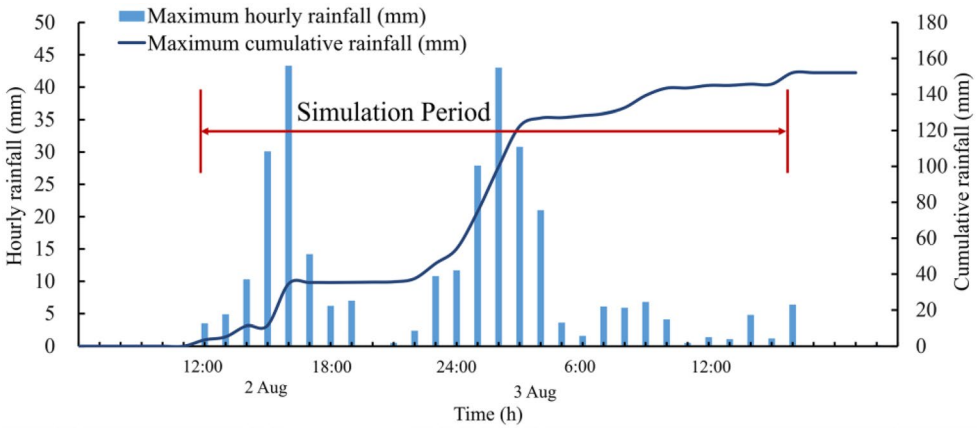

Fig. 6 Records of the maximum hourly precipitation in this region and maximum cumulative precipitation in the rainfall station

Based on the kriging method, $R(\mathbf{u}_0)$ can then be expressed by a linear combination of $R(\mathbf{u}_i)$ as follows (e.g., Chen et al., 2014):

$$R(\mathbf{u}_0) = \sum_{n_r}^{i=1} \lambda_i R(\mathbf{u}_i) \tag{15}$$

where n_r is the number of rainfall stations in the region, and λ_i is the optimal weight with the sum weights for all n_r stations equal to 1. With the rainfall data collected from the 43 rainfall stations, the 29 h cumulative rainfall data at the regional scale can then be interpolated, as shown in Fig. 5. In RARIL, the simulation period spans the entire duration of the rainfall event.

3.3 Conditioning factors

While many studies have been carried out to identify conditioning factors for developing LSMs (e.g., Reichenbach et al. 2018), consensus on the selection of which conditioning factors remains elusive (e.g., Kavzoglu et al. 2015). In general, the conditioning factors used should both consider the characteristics of landslides and the availability of the data (e.g., Kavzoglu et al. 2015; Zhao et al. 2019). In this study, nine conditioning factors are considered to build the basic data-driven model through logistic regression, including elevation, slope angle, slope aspect, distance to rivers, distance to faults, distance to roads, topographic wetness index (TWI), normalized difference vegetation index (NDVI), and land use, which are shown in Fig. 7a-i, respectively. In addition to the above nine conditioning factors, the soil type layer (available at https://www.fao.org/) is also obtained, as shown in Fig. 7j, which will be used to derive soil parameters as input for RARIL. Details of the selected conditioning factors are presented in Table 1. Considering that some environmental conditioning factors have time-variant characteristics, especially when applying during a typhoon event (Lee et al. 2008), such as NDVI and TWI values, the NDVI is derived from a Landsat 8 satellite image captured shortly before the typhoon event. Similarly, the digital terrain model with a 30 m resolution, which is used to extract the TWI, is obtained before the date of the typhoon event. The elevation, slope angle, slope aspect, and distance to rivers are also extracted from the same digital elevation model. Furthermore, a land use map provided by Globeland30 is used to get the land use layer (available at http://www.globallandcover.com/) and a road map provided by Bigemap (available at http://www.bigemap.com/) is used to generate the value of distance to roads and Zhejiang province geological map at a scale of 1:500,000 (available at http://www.ngac.org.cn) is used to extract the value of distance to faults. Consistent with the resolution of the conditioning factors, all the models in this study have a 30 m resolution.

In the following, different models will be used to assess the landslide susceptibility for Shengzhou County.

4 Results

4.1 Logistic regression model for landslide susceptibility mapping

To build a logistic regression model, the created dataset is adopted. Using nine basic conditioning factors as the features, a logistic regression model is developed based on Eqs. (1) and (2) through the training dataset. Then, the prediction capacity of the logistic regression model is checked with the validation dataset. Table 2 shows the regression coefficients of the logistic regression model. Generally, after Min-Max normalization of the raw data, a greater absolute value of the regression coefficient indicates that the conditioning factor has a greater impact on the landslide probability (e.g., Shahabi et al. 2014). Table 2 shows that the slope angle has the greatest regression coefficient (i.e., 7.908),

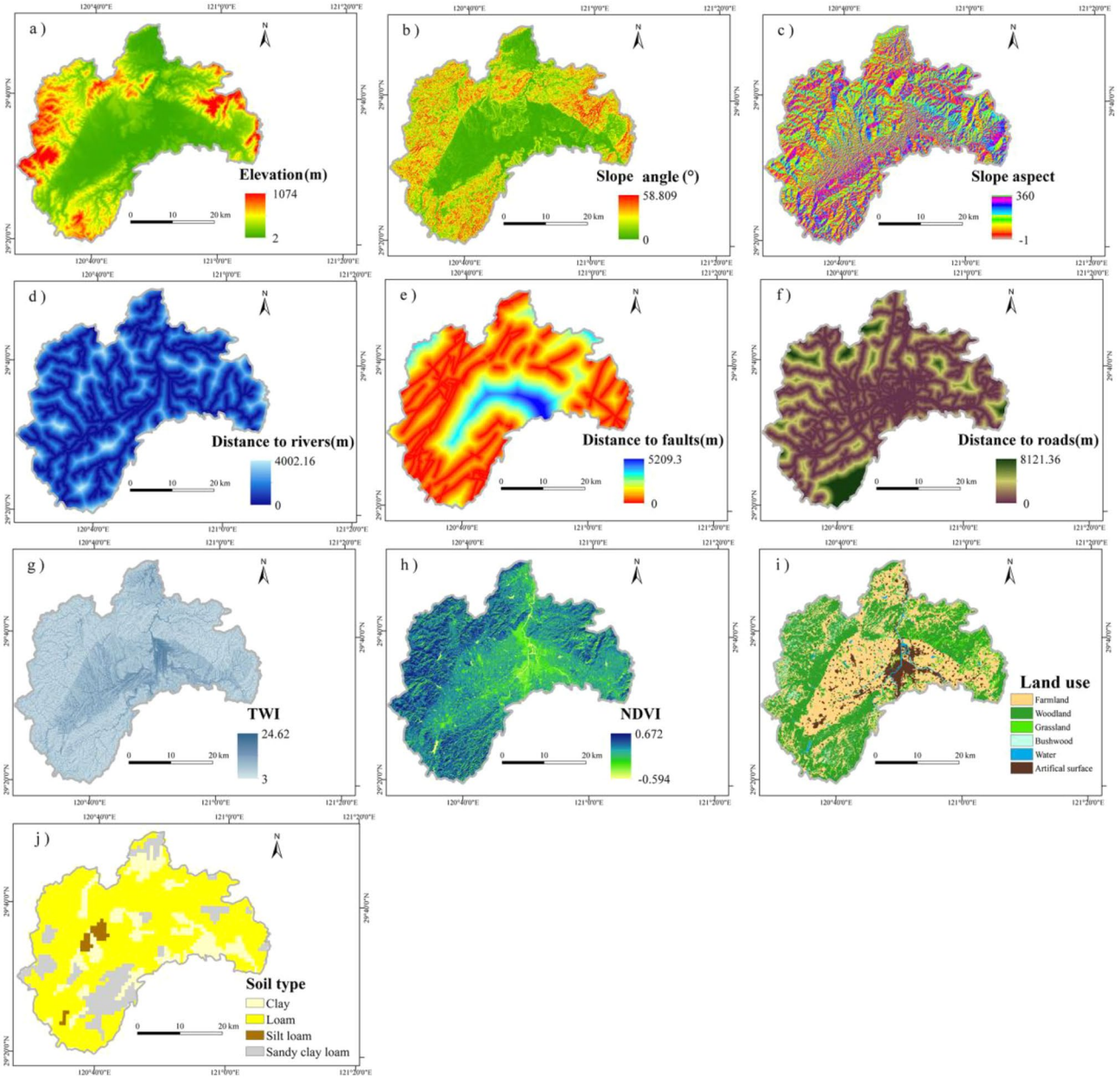

Fig. 7 Landslide conditioning factors: (**a**) elevation, (**b**) slope angle, (**c**) slope aspect, (**d**) distance to rivers, (**e**) distance to faults, (**f**) distance to roads, (**g**) TWI, (**h**) NDVI, (**i**) land use, (**j**) soil type

indicating that it has the greatest impact on the landslide probability.

When the data-driven method is used for developing LSM, the susceptibility can be divided into five classes through the landslide probability using the equal interval method, i.e., very low (0-0.2), low (0.2–0.4), middle (0.4–0.6), high (0.6–0.8), and very high (0.8-1), respectively (e.g., Baeza et al. 2016). Figure 8a shows the LSM derived based on the logistic regression model using the above susceptibility classification method. Table 3 presents the number of landslides (n_i) and number of cells (N_i) in each susceptibility class i, which can then be used to calculate the R-index using Eq. (14). The R-index values corresponding to susceptibility classes of very low, low, middle, high, and very high in the logistic regression model are 0.59, 4.01, 9.29, 25.58, and 60.53, respectively. It seems that as the susceptibility class increases from very low to very high, the R-index value which measures the landslide density also increases rapidly, indicating that the landslide susceptibility class is consistent with the observed landslide phenomenon.

Table 1 Information on landslide conditioning factors

Conditioning factors	Description
Elevation (m), Fig. 7a	Vertical height of a location on the earth's surface
Slope angle (°), Fig. 7b	The inclination of the slope surface to the horizontal plane
Slope aspect, Fig. 7c	The direction of the slope projected onto a horizontal plane
Distance to rivers (m), Fig. 7d	Distance to the nearest river
Distance to faults (m), Fig. 7e	Distance to the nearest fault
Distance to roads (m), Fig. 7f	Distance to the nearest road
TWI, Fig. 7g	An index to qualify the potential of water accumulation
NDVI, Fig. 7h	An index to qualify the vegetation cover
Land use, Fig. 7i	Type of land utilized
Soil type, Fig. 7j	Soil classification and categorization

Table 2 Coefficients of the logistic regression, hybrid models I and II

Conditioning factors	Coefficient	Logistic regression	Hybrid model I	Hybrid model II
Intercept	w_0	-0.681	0.133	2.160
Elevation	w_1	1.032	0.632	0.600
Slope angle	w_2	7.908	5.142	4.827
Slope aspect	w_3	-0.618	-0.735	-0.736
Distance to rivers	w_4	-1.218	-2.110	-1.295
Distance to faults	w_5	-0.601	-1.663	-0.905
Distance to roads	w_6	-3.025	-2.366	-2.705
TWI	w_7	-0.184	-0.634	-0.022
NDVI	w_8	-0.818	-0.461	-0.766
Land use	w_9	-0.992	-0.920	-0.963
Failure probability (p_f)/ Factor of safety (FOS)	w_{10}		44.763	-2.689

4.2 RARIL model for landslide susceptibility mapping

To implement the RARIL model, equal weight is initially assigned to all analysis grids and soil property samples (Horton et al. 2013), ensuring that each component contributes equally to the overall representation of terrain and slope stability analysis. Then the soil parameters in the region is determined. In this study, the soil parameters are determined based on its soil type, and further be divided into four types, i.e., clay, loam, silt loam, and sandy clay loam, as shown in Fig. 7j. The statistics and distributions of c, φ, Q_i, Q_s, k_s, and s_f for different types of soil used in this study are summarized in Table 4. As shown in Table 4, the initial parameter space for soil properties is provided, specifying that c and φ follow a normal distribution, Q_i and Q_s follow a uniform distribution, and k_s and s_f follow a lognormal distribution. Then, the Monte Carlo simulation generates a set of random samples from these probability distributions, and each sample represents a possible combination of soil property values. These samples are further used to calculate the failure probability of each grid under the rainfall infiltration process. In particular, the cohesion and the friction angle of the soils are determined based on the Geotechdata Database (available at https://www.geotechdata.info/). The values of Q_i and Q_s variables are derived from Zhang et al. (2018), and the values of k_s and s_f are derived from Mays (2011).

Based on the Geotechdata Database, the unit weights used for different soils are as follows: 18 kN/m^3 for clay, 20 kN/m^3 for loam, 19.5 kN/m^3 for silt loam and 21 kN/m^3 for sandy clay loam. This study also adopts a unit weight of water (γ_w) of 9.81 kN/m^3. With the above soil parameters, the failure probability of each cell in the study area during the rainfall event can be calculated through Monte Carlo simulation based on Eq. (9). For a physically-based model, Yang et al. (2022) suggested that the susceptibility of a cell can be considered as low, middle, and high when the failure probability is less than 0.05, between 0.05 and 0.1, and above 0.1, respectively. Based on the above landslide susceptibility classification method, a LSM can also be generated for the study area, as shown in Fig. 8b. Table 3 also presents the R-index for each susceptibility class corresponding to the RARIL model. As shown in Table 3, a notable high R-index (71.71) is observed in the high susceptibility class, indicating a reasonable classification capacity of the RARIL model.

4.3 Hybrid model I for landslide susceptibility mapping

By employing the nine basic conditioning factors and incorporating the failure probability value obtained in Sect. 4.2 as an additional conditioning factor, a logistic regression model can be trained and the coefficients are shown in Table 2. It is interesting to see that while in the logistic regression model and hybrid model II, the slope angle has the greatest coefficient, in the hybrid model I, the failure probability estimated based on RARIL has the greatest coefficient. The LSM generated by hybrid model I is subsequently reclassified using the equal interval method, as shown in Fig. 8c. Table 3 presents the R-index for each susceptibility class corresponding to hybrid model I. It is interesting that for the hybrid model I, its R-index value in the very high landslide susceptibility class is 80.46, which is greater than all other models, indicating that it can classify landslides more accurately.

4.4 Hybrid model II for landslide susceptibility mapping

The hybrid model II is created by treating the FOS value calculated through RARIL as the additional conditioning

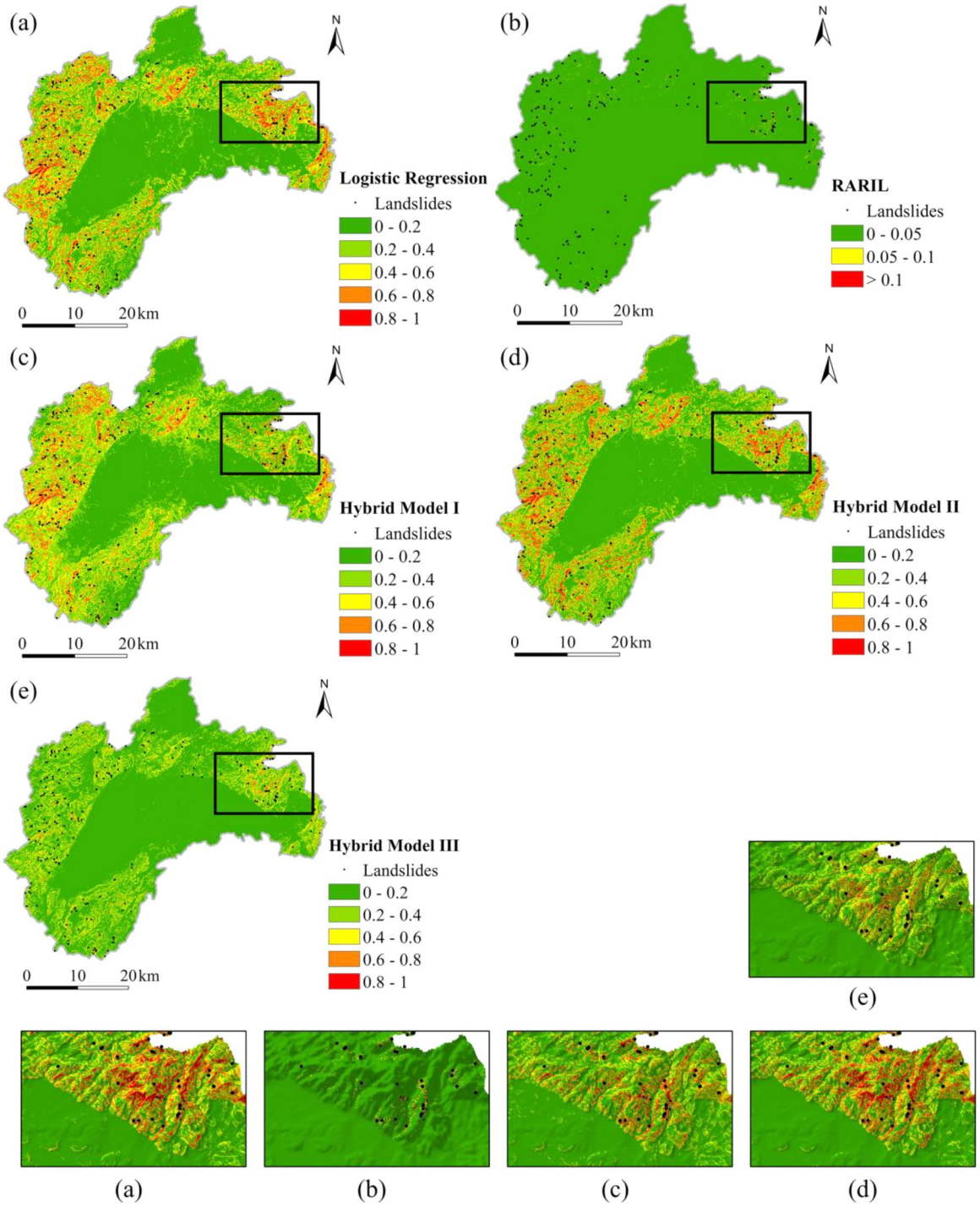

Fig. 8 Landslide susceptibility maps derived using five models: (**a**) Logistic regression, (**b**) RARIL, (**c**) Hybrid model I, (**d**) Hybrid model II, (**e**) Hybrid model III

factor for logistic regression. To implement hybrid model II, the FOS value generated by RARIL without considering soil parameters uncertainty is needed. Hence, the FOS is calculated by RARIL using the mean values of the random variables as listed in Table 4. Subsequently, using Eq. (6), RARIL is adopted to calculate the FOS distribution of the study area under the typhoon-related rainfall event, as illustrated in Fig. 9.

Using nine basic conditioning factors and incorporating the FOS value as an additional conditioning factor, hybrid model II is built through logistic regression. Table 2 also presents the coefficient of the ten conditioning factors of

Table 3 R-index comparison for the logistic regression, RARIL, and three hybrid models

Method	Probability of landslide	Susceptibility class	Number of landslide samples in susceptibility class i (n_i)	Number of cells in susceptibility class i (N_i)	R-index
Logistic regression	0-0.2	Very Low	14	1,009,497	0.59
	0.2–0.4	Low	41	433,533	4.01
	0.4–0.6	Middle	64	291,854	9.29
	0.6–0.8	High	110	182,137	25.58
	0.8-1	Very High	89	62,271	60.53
RARIL	0-0.05	Low	114	1,937,308	0.49
	0.05–0.1	Middle	98	29,579	27.80
	0.1-1	High	106	12,405	71.71
Hybrid model I	0-0.2	Very Low	12	948,930	0.32
	0.2–0.4	Low	49	582,987	2.10
	0.4–0.6	Middle	58	289,709	4.99
	0.6–0.8	High	55	113,033	12.13
	0.8-1	Very High	144	44,633	80.46
Hybrid model II	0-0.2	Very Low	18	1,138,824	0.59
	0.2–0.4	Low	33	334,286	3.71
	0.4–0.6	Middle	42	248,381	6.36
	0.6–0.8	High	92	187,202	18.48
	0.8-1	Very High	133	70,599	70.85
Hybrid model III	0-0.2	Very Low	47	1,419,613	0.25
	0.2–0.4	Low	77	348,319	1.65
	0.4–0.6	Middle	80	177,224	3.36
	0.6–0.8	High	90	31,708	21.13
	0.8-1	Very High	24	2428	73.61

hybrid model II. As can be seen from Table 2, the slope angle exhibits the highest significance in hybrid model II which shows that the FOS still not surpass the influence of slope angle with this hybrid method. Based on the equal interval method, a new LSM is created, as shown in Fig. 8d, and the R-indexes for different landslide susceptibility classes are calculated and shown in Table 3. As can be seen from Table 3, the R-index values corresponding to susceptibility classes of very low, low, middle, high, and very high in the hybrid model II are 0.59, 3.71, 6.36, 18.48, and 70.85, respectively, exhibiting an accurate trend of increasing landslide intensity with the landslide susceptibility class.

4.5 Hybrid model III for landslide susceptibility mapping

As mentioned in Sect. 2.3, hybrid model III which employed a matrix-based approach, is also included for comparison. This method integrates the data-driven landslide susceptibility classes derived by logistic regression (Sect. 4.1) with FOS values obtained from the RARIL model without considering the soil parameters uncertainty (Sect. 4.4). As can be seen from Fig. 9, FOS values are divided into five classes: FOS < 1 (Very High), 1 < FOS < 1.2 (High), 1.2 < FOS < 1.5 (Middle), 1.5 < FOS < 2 (Low), and FOS > 2 (Very Low) (e.g., Wei et al. 2023). Following the matrix-based approach illustrated in Fig. 3, these inputs are combined to generate a new LSM, as shown in Fig. 8e. The corresponding R-index values for each susceptibility class are also presented in Table 3. This hybrid model also demonstrates an accurate trend of increasing relative landslide intensity index with increasing susceptibility class: very low (0.25), low (1.65), middle (3.36), high (21.13), and very high (73.61).

5 Discussion

5.1 Comparison of the conventional and hybrid models

In this section, the five landslide susceptibility models are compared in terms of overall accuracy and the ROC curve with the validation dataset. To calculate the overall accuracy of the five models based on Eq. (11), the landslide probability threshold beyond which a landslide is considered to occur needs to be determined. For the conventional logistic regression model, hybrid models I and II, which are constructed based on logistic regression, and hybrid model III, which is constructed based on a matrix approach, the probability threshold is set to 0.5 (e.g., Nefeslioglu et al. 2008). For RARIL, the probability threshold is selected as 0.05 (Yang et al. 2022). Table 5 shows the overall accuracy of the five models. As shown in Table 5, hybrid model I gets the highest overall accuracy with a value of 0.80, followed

Table 4 Statistics of soil parameters

Soil type	c (kPa)		φ (°)		Q_i		Q_s		k_s (m/s)		s_f (m)	
	μ	cov	μ	cov	lower	upper	lower	upper	μ	cov	μ	cov
Clay	8	0.15	21	0.20	0.166	0.286	0.380	0.538	1.71×10^{-6}	0.10	0.316	0.10
Loam	11	0.30	29	0.10	0.059	0.193	0.302	0.504	1.54×10^{-6}	0.10	0.089	0.10
Silt loam	15	0.30	28	0.12	0.048	0.182	0.346	0.532	2.14×10^{-6}	0.10	0.167	0.10
Sandy clay loam	10	0.20	24	0.20	0.108	0.230	0.324	0.440	1.53×10^{-6}	0.10	0.219	0.10
Distribution	Normal		Normal		Uniform		Uniform		Lognormal		Lognormal	

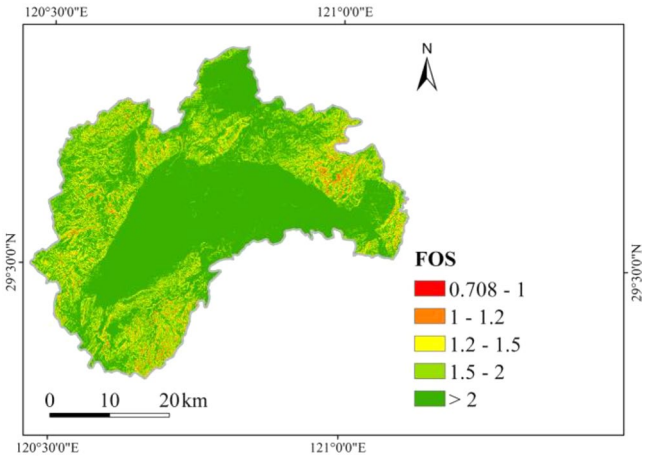

Fig. 9 Factor of safety (FOS) calculated by the RARIL

Table 5 The overall accuracy of the five models

Model	Overall Accuracy
Logistic regression	0.74
RARIL	0.70
Hybrid model I	0.80
Hybrid model II	0.77
Hybrid model III	0.75

by hybrid model II (0.77), hybrid model III (0.75), logistic regression (0.74), and RARIL (0.70). It is obvious that all the three hybrid models outperform the conventional logistic regression and RARIL model. Moreover, after considering the soil parameters uncertainty, hybrid model I achieved the highest overall accuracy.

Figure 10 presents the ROC curves of the five models. As shown in Fig. 10, the AUC values of the conventional logistic regression model, RARIL model, hybrid model I, hybrid model II and hybrid model III are 0.82, 0.79, 0.89, 0.85, and 0.85, respectively. Also can be seen that all the hybrid models outperform the conventional data-driven and physically-based models. Among the three hybrid models, the hybrid model I has the greatest AUC compared to the other models, indicating that considering the soil parameters uncertainty can improve the prediction capacity of landslide susceptibility analysis in this study.

5.2 Improvement of prediction capacity through model hybridization

The comparison of five models reveals the potential of integrating data-driven and physically-based approaches for the improvement of model prediction capacity. In this study, we proposed a hybrid method that can considers soil parameter uncertainty and compared it with two other models that does not consider it. Hybrid models II and III, which integrate the RARIL model with fixed soil parameters to generate FOS

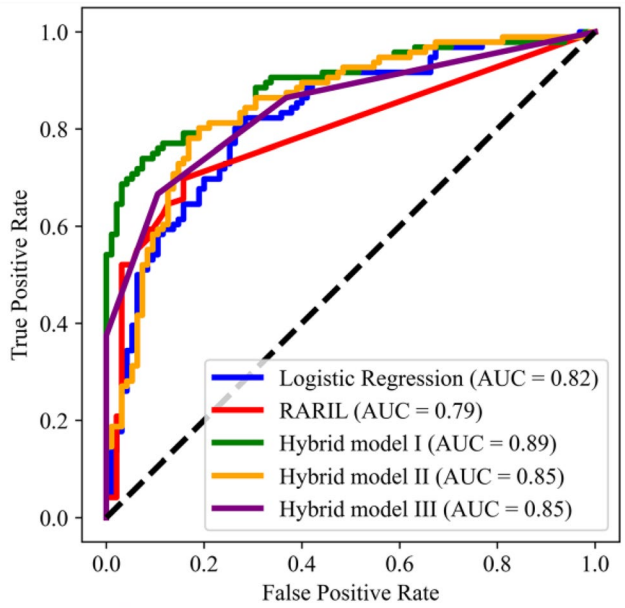

Fig. 10 ROC curves for the logistic regression, RARIL, and three hybrid models

values/classes, exhibited similar AUC values. This suggests that combining the two types of model results using either a logistic regression framework or a matrix-based approach does not significantly impact predictions in this study. Furthermore, hybrid model I, which considers soil parameter uncertainty with the RARIL model to generate failure probability values, shows a notable improvement in prediction capacity compared to the other hybrid models that do not consider it. This can be explained that after consider the inherent uncertainty in soil properties at a regional scale, hybrid model I provides a more realistic reflection of actual soil conditions in its physically-based model inputs. Moreover, through considering the soil parameters uncertainty and simulating more soil samples to calculate failure probability values, the proposed hybrid model can further reduce the randomness of calculation than using FOS as inputs, thus reaching a better prediction capacity in the hybrid results.

5.3 Event-induced landslide susceptibility mapping

In many regions affected by typhoons, typhoon-related heavy rainfall event can lead to landslides on a large regional scale. Sometimes, as one typhoon-related heavy rainfall event just passes, another follows, and rainfall-induced regional landslides in this region may occur again. This phenomenon is quite common. However, fewer typhoon-related rainfall-induced regional landslide susceptibility assessments have been carried out. Previously, in most rainfall-induced landslide susceptibility assessment studies, data-driven models based on historical landslide inventories and conditioning factors have been commonly used, but their triggering factors have seldom been emphasized. Due to the lack of temporal information in multi-source historical landslide inventories, rainfall factors are typically treated similarly to other environmental or geological conditioning factors by using annual average rainfall data. Therefore, we aimed to address this issue by considering rainfall-related results calculated by the physically-based model as an additional conditioning factor to assess the influence of typhoon-related rainfall events on a regional scale. This can also be seen as the event-induced landslide susceptibility mapping. Further, there is published research about that. Dai and Lee (2003) proposed a method using real rainfall data as an independent variable in typhoon-induced shallow landslide susceptibility assessment which provides new insight into rainfall-induced regional landslide susceptibility mapping. Subsequently, Lee et al. (2008) conducted study with four typhoon rainfall-induced landslide inventories, confirming the effectiveness of the method in predicting subsequent typhoon-induced regional landslide events in nearby areas. In this study, we select Shengzhou County as an illustration, as it was significantly affected by typhoon "Jongdari". The proposed hybrid method incorporates a data-driven model and a physically-based model. Initially, the data-driven model adopted a typhoon event-related landslide inventory to establish a basic framework. Subsequently, the physically-based model, which does not consider the actual typhoon-related landslide inventory data, relies solely on the regional terrain data, slope angle, soil type, and the triggering factor - rainfall, to predict the landslide-prone area through reliability analysis.

While the data-driven model uses the landslide inventory caused by a typhoon event, which can also be seen as historical landslide data in the subsequent typhoon events in this region, the physically-based model's results are primarily influenced by the rainfall event, showing no need for actual typhoon-related landslide inventory data. This also gives this study the potential for application. It may further be applied to the study area experiencing subsequent rainfall events by updating the rainfall event inputs in the physically-based model to generate new failure probability value. With the data-driven model establishing a basic framework, coupled with the inputs from the updated failure probability value, thus a new landslide susceptibility maps can be created with the new rainfall event.

6 Summary and conclusions

Traditionally, data-driven model and physically-based models are widely used for landslide susceptibility mapping, but each method has limitations. This paper introduces

a novel hybrid method that combines the data-driven and physically-based models, aiming to improve the prediction capacity of rainfall-induced regional landslide susceptibility mapping by explicitly considering uncertainty in soil properties. The proposed method is implemented with a case study in Shengzhou County, Zhejiang Province, China. In the case study, logistic regression serves as the data-driven model, and RARIL serves as the physically-based model. Three hybrid models are developed: hybrid model I, which explicit consider the soil parameters uncertainty, hybrid model II and hybrid model III, which do not consider it and with different hybrid types. The comparison results indicate that all the hybrid models outperform conventional logistic regression and RARIL model. Notably, hybrid model I, which consider soil parameter uncertainty, exhibits greater prediction capacity compared to hybrid model II and hybrid model III that do not consider it, demonstrating the importance of considering uncertainty in soil properties for more accurate landslide susceptibility mapping.

Appendix

For our case study in Shengzhou County, Zhejiang Province, China, the digital terrain is partitioned into 1979292 grid cells with 30 m resolution, the computation time on a Windows 11 computer is as follows:

Data-driven model (logistic regression): within 1 min.
Physically-based model (RARIL): within 2 h.
Hybrid model I: within 1 min.
Hybrid model II: within 1 min.
Hybrid model III: within 1 min.

The computations were performed on a computer with the following specifications:

CPU: Intel Core i9-13900 K @ 3.00 GHz.
RAM: 64GB.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00477-024-02753-9.

Acknowledgements This research was supported by The National Key Research and Development Program of China (No. 2021YFB2600500), The National Natural Science Foundation of China (42072302, 52025094), and Fundamental Research Funds for the Central Universities.

Author contributions All authors contributed to the study conception and design. Shuangyi Wu designed and conducted the experiments, analyzed the results, and wrote the initial draft of the paper. Huaan Wang participated in data analysis. Jie Zhang conducted the final review and revision of the paper. Haijun Qin contributed to editing and visualization. All authors read and approved the final manuscript.

Data availability Data will be made available on reasonable request.

Declarations

Competing interests The authors declare no competing interests.

References

Abraham MT, Vaddapally M, Satyam N, Pradhan B (2023) Spatio-temporal landslide forecasting using process-based and data-driven approaches: a case study from Western Ghats, India. CATENA 223:106948. https://doi.org/10.1016/j.catena.2023.106948

Baeza C, Corominas J (2001) Assessment of shallow landslide susceptibility by means of multivariate statistical techniques. Earth Surf Proc Land 26(12):1251–1263. https://doi.org/10.1002/esp.263

Baeza C, Lantada N, Amorim S (2016) Statistical and spatial analysis of landslide susceptibility maps with different classification systems. Environ Earth Sci 75(19):1318. https://doi.org/10.1007/s12665-016-6124-1

Baum RL, Savage WZ, Godt JW (2002) TRIGRS—A Fortran program for transient rainfall infiltration and grid-based regional slope-stability analysis. US Geol Surv open-file Rep 424:38. https://pubs.usgs.gov/of/2008/1159/

Bradley AP (1997) The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recogn 30(7):1145–1159. https://doi.org/10.1016/S0031-3203(96)00142-2

Cantarino I, Carrion MA, Goerlich F, Martinez Ibañez A (2019) A ROC analysis-based classification method for landslide susceptibility maps. Landslides 16(2):265–282. https://doi.org/10.1007/s10346-018-1063-4

Cao Z, Wang Y, Li D (2016) Site-specific characterization of soil properties using multiple measurements from different test procedures at different locations—A bayesian sequential updating approach. Eng Geol 211:150–161. https://doi.org/10.1016/j.enggeo.2016.06.021

Cascini L, Scoppettuolo MR, Babilio E (2022) Forecasting the landslide evolution: from theory to practice. Landslides 19(12):2839–2851. https://doi.org/10.1007/s10346-022-01934-3

Chang Z, Huang F, Huang J, Jiang SH, Liu Y, Meena SR, Catani F (2023) An updating of landslide susceptibility prediction from the perspective of space and time. Geosci Front 14(5):101619. https://doi.org/10.1016/j.gsf.2023.101619

Chen L, Young MH (2006) Green-Ampt infiltration model for sloping surfaces. Water Resour Res 42:W07420. https://doi.org/10.1029/2005WR004468

Chen HX, Zhang LM (2014) A physically-based distributed cell model for predicting regional rainfall-induced shallow slope failures. Eng Geol 176:79–92. https://doi.org/10.1016/j.enggeo.2014.04.011

Chen T, Niu R, Jia X (2016) A comparison of information value and logistic regression models in landslide susceptibility mapping by using GIS. Environ Earth Sci 75:867. https://doi.org/10.1007/s12665-016-5317-y

Chen X, Zhang L, Zhang L, Zhou Y, Ye G, Guo N (2021) Modelling rainfall-induced landslides from initiation of instability to post-failure. Comput Geotech 129:103877. https://doi.org/10.1016/j.compgeo.2020.103877

Cho SE, Lee SR (2002) Evaluation of surficial stability for homogeneous slopes considering rainfall characteristics. J Geotech GeoEnviron Eng 128(9):756–763. https://doi.org/10.1061/(ASCE)1090-0241(2002)128:9(756)

Chowdhury R, Flentje P (2003) Role of slope reliability analysis in landslide risk management. Bull Eng Geol Environ 62(1):41–46. https://doi.org/10.1007/s10064-002-0166-1

Cressie N (1990) The origins of kriging. Math Geol 22(3):239–252. https://doi.org/10.1007/BF00889887

Dai FC, Lee CF (2003) A spatiotemporal probabilistic modelling of storm-induced shallow landsliding using aerial photographs and logistic regression. Earth Surf Process Landf 28(5):527–545. https://doi.org/10.1002/esp.456

Fang Z, Wang Y, Peng L, Hong H (2020) Integration of convolutional neural network and conventional machine learning classifiers for landslide susceptibility mapping. Comput Geosci 139:104470. https://doi.org/10.1016/j.cageo.2020.104470

Fredlund DG, Morgenstern NR, Widger RA (1978) The shear strength of unsaturated soils. Can Geotech J 15(3):313–321. https://doi.org/10.1139/t78-029

Fu Z, Wang F, Dou J, Nam K, Ma H (2023) Enhanced absence sampling technique for data-driven landslide susceptibility mapping: a case study in Songyang County, China. Remote Sens 15(13):3345. https://doi.org/10.3390/rs15133345

Gong W, Hu M, Zhang Y, Tang H, Liu D, Song Q (2022) GIS-based landslide susceptibility mapping using ensemble methods for Fengjie County in the Three Gorges Reservoir Region, China. Int J Environ Sci Technol 19(8):7803–7820. https://doi.org/10.1007/s13762-021-03572-z

Harp EL, Keefer DK, Sato HP, Yagi H (2011) Landslide inventories: the essential part of seismic landslide hazard analyses. Eng Geol 122(1):9–21. https://doi.org/10.1016/j.enggeo.2010.06.013

Horton P, Jaboyedoff M, Rudaz BEA, Zimmermann M (2013) Flow-R, a model for susceptibility mapping of debris flows and other gravitational hazards at a regional scale. Nat Hazards Earth Syst Sci 13(4):869–885. https://doi.org/10.5194/nhess-13-869-2013

Jasiewicz J, Stepinski TF (2013) Geomorphons—a pattern recognition approach to classification and mapping of landforms. Geomorphology 182:147–156. https://doi.org/10.1016/j.geomorph.2012.11.005

Jelínek R, Wagner P (2007) Landslide hazard zonation by deterministic analysis (Veľká Čausa landslide area, Slovakia). Landslides 4(4):339–350. https://doi.org/10.1007/s10346-007-0089-9

Ji J, Cui H, Zhang T, Song J, Gao Y (2022) A GIS-based tool for probabilistic physical modelling and prediction of landslides: GIS-FORM landslide susceptibility analysis in seismic areas. Landslides 19(9):2213–2231. https://doi.org/10.1007/s10346-022-01885-9

Kavzoglu T, Sahin EK, Colkesen I (2015) Selecting optimal conditioning factors in shallow translational landslide susceptibility mapping using genetic algorithm. Eng Geol 192:101–112. https://doi.org/10.1016/j.enggeo.2015.04.004

Kim J, Lee K, Jeong S, Kim G (2014) GIS-based prediction method of landslide susceptibility using a rainfall infiltration-groundwater flow model. Eng Geol 182:63–78. https://doi.org/10.1016/j.enggeo.2014.09.001

Lee CT, Huang CC, Lee JF, Pan KL, Lin ML, Dong JJ (2008) Statistical approach to storm event-induced landslides susceptibility. Nat Hazards Earth Syst Sci 8(4):941–960. https://doi.org/10.5194/nhess-8-941-2008

Liu Q, Tang A, Huang D, Huang Z, Zhang B, Xu X (2022) Total probabilistic measure for the potential risk of regional roads exposed to landslides. Reliab Eng Syst Saf 228:108822. https://doi.org/10.1016/j.ress.2022.108822

Liu Q, Tang A, Huang D (2023) Exploring the uncertainty of landslide susceptibility assessment caused by the number of non–landslides. CATENA 227:107109. https://doi.org/10.1016/j.catena.2023.107109

Long J, Liu Y, Li C, Fu Z, Zhang H (2021) A novel model for regional susceptibility mapping of rainfall-reservoir induced landslides in jurassic slide-prone strata of western Hubei Province, Three Gorges Reservoir area. Stoch Env Res Risk Assess 35:1403–1426. https://doi.org/10.1007/s00477-020-01892-z

Lu M, Zheng J, Zhang J, Huang H (2023) On assessing the probability of rainfall-induced slope failure during a given exposure time. Acta Geotech 18(3):1255–1267. https://doi.org/10.1007/s11440-022-01655-w

Ma S, Shao X, Xu C, He X, Zhang P (2021) MAT.TRIGRS (v1.0): a new open-source tool for predicting spatiotemporal distribution of rainfall-induced landslides. Nat Hazards Res 1(4):161–170. https://doi.org/10.1016/j.nhres.2021.11.001

Mays LW (2011) Water resources Engineering. Wiley, New York

Medina V, Hürlimann M, Guo Z, Lloret A, Vaunat J (2021) Fast physically-based model for rainfall-induced landslide susceptibility assessment at regional scale. CATENA 201:105213. https://doi.org/10.1016/j.catena.2021.105213

Merghadi A, Yunus AP, Dou J, Whiteley J, ThaiPham B, Bui DT, Avtar R, Abderrahmane B (2020) Machine learning methods for landslide susceptibility studies: a comparative overview of algorithm performance. Earth Sci Rev 207:103225. https://doi.org/10.1016/j.earscirev.2020.103225

Miao F, Zhao F, Wu Y, Li L, Török Á (2023) Landslide susceptibility mapping in Three Gorges Reservoir area based on GIS and boosting decision tree model. Stoch Env Res Risk Assess 37:2283–2303. https://doi.org/10.1007/s00477-023-02394-4

Mondini AC, Guzzetti F, Melillo M (2023) Deep learning forecast of rainfall-induced shallow landslides. Nat Commun 14(1):2466. https://doi.org/10.1038/s41467-023-38135-y

Montgomery DR, Dietrich WE (1994) A physically based model for the topographic control on shallow landsliding. Water Resour Res 30(4):1153–1171. https://doi.org/10.1029/93WR02979

Nefeslioglu HA, Gokceoglu C, Sonmez H (2008) An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility maps. Eng Geol 97(3–4):171–191. https://doi.org/10.1016/j.enggeo.2008.01.004

Nguyen BQV, Song CH, Kim YT (2022) A hybrid physical and machine learning model for assessing landslide spatial probability caused by raising of ground water table and earthquake in Atsuma, Japan—case study. KSCE J Civ Eng 26(8):3416–3429. https://doi.org/10.1007/s12205-022-1656-2

Olea RA (1999) Geostatistics for engineers and earth scientists. Kluwer Academic, Norwell

Oliveira SC, Zêzere JL, Lajas S, Melo R (2017) Combination of statistical and physically based methods to assess shallow slide susceptibility at the basin scale. Nat Hazards Earth Syst Sci 17(7):1091–1109. https://doi.org/10.5194/nhess-17-1091-2017

Pradhan AMS, Kim YT (2016) Evaluation of a combined spatial multi-criteria evaluation model and deterministic model for landslide susceptibility mapping. CATENA 140:125–139. https://doi.org/10.1016/j.catena.2016.01.022

Razavi-Termeh SV, Hatamiafkoueieh J, Sadeghi-Niaraki A, Choi SM, Al-Kindi KM (2023) A GIS-based multi-objective evolutionary algorithm for landslide susceptibility mapping. Stoch Env Res Risk Assess 1–26. https://doi.org/10.1007/s00477-023-02562-6

Reichenbach P, Rossi M, Malamud B, Mihir M, Guzzetti F (2018) A review of statistically-based landslide susceptibility models. Earth Sci Rev 180:60–91. https://doi.org/10.1016/j.earscirev.2018.03.001

Saha A, Villuri VGK, Bhardwaj A (2023) Development and assessment of a novel hybrid machine learning-based landslide susceptibility mapping model in the Darjeeling Himalayas. Stoch Env Res Risk Assess 1–24. https://doi.org/10.1007/s00477-023-02528-8

Shahabi H, Khezri S, Ahmad BB, Hashim M (2014) Landslide susceptibility mapping at central zab basin, Iran: a comparison between analytical hierarchy process, frequency ratio and logistic regression models. CATENA 115:55–70. https://doi.org/10.1016/j.catena.2013.11.014

Strauch R, Istanbulluoglu E, Riedel J (2019) A new approach to mapping landslide hazards: a probabilistic integration of empirical and physically based models in the North Cascades of Washington,

USA. Nat Hazards Earth Syst Sci 19(11):2477–2495. https://doi.org/10.5194/nhess-19-2477-2019

Su C, Wang B, Lv Y, Zhang M, Peng D, Bate B, Zhang S (2023) Improved landslide susceptibility mapping using unsupervised and supervised collaborative machine learning models. Georisk: Assess Manage Risk Eng Syst Geohazards 17(2):387–405. https://doi.org/10.1080/17499518.2022.2088802

Sun D, Xu J, Wen H, Wang D (2021) Assessment of landslide susceptibility mapping based on bayesian hyperparameter optimization: a comparison between logistic regression and random forest. Eng Geol 281:105972. https://doi.org/10.1016/j.enggeo.2020.105972

Sun Y, Zhang J, Wang H, Lu D (2024) Probabilistic thresholds for regional rainfall induced landslides. Comput Geotech 166:106040. https://doi.org/10.1016/j.compgeo.2023.106040

Süzen ML, Doyuran V (2004) Data driven bivariate landslide susceptibility assessment using geographical information systems: a method and application to Asarsuyu catchment, Turkey. Eng Geol 71(3-4):303–321. https://doi.org/10.1016/S0013-7952(03)00143-1

Tobutt DC (1982) Monte Carlo simulation methods for slope stability. Comput Geosci 8(2):199–208. https://doi.org/10.1016/0098-3004(82)90021-8

Vanacker V, Vanderschaeghe M, Govers G, Willems E, Poesen J, Deckers J, De Bievre B (2003) Linking hydrological, infinite slope stability and land-use change models through GIS for assessing the impact of deforestation on slope stability in high Andean watersheds. Geomorphology 52(3–4):299–315. https://doi.org/10.1016/S0169-555X(02)00263-5

Vieira BC, Fernandes NF, Augusto Filho O, Martins TD, Montgomery DR (2018) Assessing shallow landslide hazards using the TRIGRS and SHALSTAB models, Serra do Mar, Brazil. Environ Earth Sci 77(6):1–15. https://doi.org/10.1007/s12665-018-7436-0

Wang H, Yang T, Zhang P, Liu F, Liu H, Niu P (2023) Landslide susceptibility prediction considering rock integrity and stress state: a case study. Bull Eng Geol Environ 82(7):259. https://doi.org/10.1007/s10064-023-03250-z

Wang H, Zhang L, Luo H, He J, Cheung RWM (2021a) AI-powered landslide susceptibility assessment in Hong Kong. Eng Geol 288:106103. https://doi.org/10.1016/j.enggeo.2021.106103

Wang H, Zhang L, Yin K, Luo H, Li J (2021b) Landslide identification using machine learning. Geosci Front 12(1):351–364. https://doi.org/10.1016/j.gsf.2020.02.012

Wang Y, Zhao T, Cao Z (2015) Site-specific probability distribution of geotechnical properties. Comput Geotech 70:159–168. https://doi.org/10.1016/j.compgeo.2015.08.002

Wei Z, Lü Q, Sun H, Shang Y (2019) Estimating the rainfall threshold of a deep-seated landslide by integrating models for predicting the groundwater level and stability analysis of the slope. Eng Geol 253:14–26. https://doi.org/10.1016/j.enggeo.2019.02.026

Wei X, Zhang LL, Luo JY, Liu DS (2021) A hybrid framework integrating physical model and convolutional neural network for regional landslide susceptibility mapping. Nat Hazards 109(1):471–497. https://doi.org/10.1007/s11069-021-04844-0

Wei X, Zhang L, Gardoni P, Chen Y, Tan L, Liu D, Du C, Li H (2023) Comparison of hybrid data-driven and physical models for landslide susceptibility mapping at regional scales. Acta Geotech 1–24. https://doi.org/10.1007/s11440-023-01841-4

Wei X, Gardoni P, Zhang L, Tan L, Liu D, Du C, Li H (2024) Improving pixel-based regional landslide susceptibility mapping. Geosci Front 101782. https://doi.org/10.1016/j.gsf.2024.101782

Woodard JB, Mirus BB, Crawford MM, Or D, Leshchinsky BA, Allstadt KE, Wood NJ (2023) Mapping landslide susceptibility over large regions with limited data. J Geophys Research: Earth Surf 128(5). https://doi.org/10.1029/2022JF006810

Xue Z, Feng W, Yi X, Dun J, Wu M (2024) Integrating data-driven and physically based landslide susceptibility methods using matrix models to predict reservoir landslides. Adv Space Res 73(3):1702–1720. https://doi.org/10.1016/j.asr.2023.11.014

Yalcin A (2008) GIS-based landslide susceptibility mapping using analytical hierarchy process and bivariate statistics in Ardesen (Turkey): comparisons of results and confirmations. CATENA 72(1):1–12. https://doi.org/10.1016/j.catena.2007.01.003

Yang S, Tan Z, Chen H, Zhang J (2022) Analysis of instability disaster of rainfall induced shallow landslides at the regional scale based on the modified green-ampt model. Bull Geol Sci Technol 41(2):221–229. https://doi.org/10.19509/j.cnki.dzkq.2022.0048

Yilmaz I (2009) Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: a case study from kat landslides (Tokat—Turkey). Comput Geosci 35(6):1125–1138. https://doi.org/10.1016/j.cageo.2008.08.007

Zhan C, Xie M (2022) Exploring the link between ozone pollution and stratospheric intrusion under the influence of tropical cyclone Ampil. Sci Total Environ 828:154261. https://doi.org/10.1016/j.scitotenv.2022.154261

Zhan W, Baise LG, Moaveni B (2023) An uncertainty quantification framework for logistic regression based geospatial natural hazard modeling. Eng Geol 324:107271. https://doi.org/10.1016/j.enggeo.2023.107271

Zhang LL, Zhang J, Zhang LM, Tang WH (2011) Stability analysis of rainfall-induced slope failure: a review. Proc Institution Civil Engineers-Geotechnical Eng 164(5):299–316. https://doi.org/10.1680/geng.2011.164.5.299

Zhang J, Huang HW, Zhang LM, Zhu HH, Shi B (2014) Probabilistic prediction of rainfall-induced slope failure using a mechanics-based model. Eng Geol 168:129–140. https://doi.org/10.1016/j.enggeo.2013.11.005

Zhang Y, Schaap MG, Zha Y (2018) A high resolution global map of soil hydraulic properties produced by a hierarchical parameterization of a physically based water retention model. Water Resour Res 54(12):9774–9790. https://doi.org/10.1029/2018WR023539

Zhang J, Zhu D, Zhang S (2020) Shallow slope stability evolution during rainwater infiltration considering soil cracking state. Comput Geotech 117:103285. https://doi.org/10.1016/j.compgeo.2019.103285

Zhang K, Zhang K, Cai C, Liu W, Xie J (2021) Displacement prediction of step-like landslides based on feature optimization and VMD-Bi-LSTM: a case study of the Bazimen and Baishuihe landslides in the three gorges, China. Bull Eng Geol Environ 80(11):8481–8502. https://doi.org/10.1007/s10064-021-02454-5

Zhao Y, Wang R, Jiang Y, Liu H, Wei Z (2019) GIS-based logistic regression for rainfall-induced landslide susceptibility mapping under different grid sizes in yueqing, southeastern China. Eng Geol 259:105147. https://doi.org/10.1016/j.enggeo.2019.105147