A hybrid machine-learning model to estimate potential debris-flow volumes

Jian Huang[1,2], Tristram C. Hales[2], Runqiu Huang[1], Nengpan Ju[1], Qiao Li[1], Yin Huang[1]
1. State Key Laboratory of Geohazard Prevention and Geoenvironment Protection
Chengdu University of Technology, Chengdu, Sichuan 610059, China
2. School of Earth and Ocean Sciences, Cardiff University, UK

Abstract: Empirical-statistical models of debris-flow are challenging to implement in environments where sedimentary and hydrologic triggering processes change through time, such as after a large earthquake. The flexible and adaptive statistical methods provided by machine learning algorithms may improve the quality of debris flow predictions where triggering conditions and the nature of sediment that can bulk flows varies with time. We developed a hybrid machine-learning model of future debris-flow volumes using a dataset of measured debris-flow volumes from 60 catchments that generated post-Wenchuan Earthquake (M_w 7.9) debris flows. We input topographic variables (catchment area, topographic relief, channel length, distance from seismic fault, and average channel gradient) and the total volume of co-seismic landslide debris into the PSO-ELM_AdaBoost machine-learning model, created by combining Extreme learning machine (ELM), particle swarm optimization (PSO) and adaptive boosting machine learning algorithm (AdaBoost). The model was trained and tested using post-2008 M_w 7.9 Wenchuan Earthquake debris flows, then applied to understand potential volumes of post-earthquake debris flows associated with other regional earthquakes (2013 M_w 6.6 Lushan Earthquake, 2010 M_w 6.9 Yushu Earthquake). We compared the PSO-ELM_Adaboost method with different machine learning methods, including back-propagation neural network (BPNN), support vector machine (SVM), ELM, PSO-ELM. The Comparative analysis demonstrated that the PSO-ELM_Adaboost method has a higher statistical validity and prediction accuracy with a mean absolute percentage error (MAPE) less than 0.10. The prediction accuracy of debris-flow volumes triggered by other earthquakes decreases to 0.11 - 0.16 (absolute percentage error), suggesting that once calibrated for a region this method can be applied to other regional earthquakes. This model may be useful for engineering design to mitigate the risk of large post-earthquake debris flows.

Keywords: debris flow; machine-learning model; estimated volume; prediction

1. Introduction

Co-seismic landslides triggered by strong earthquakes can act as sources for post-earthquake debris flows (Chen 2011; Fan et al. 2019). In the regions hardest-hit by the M_w 7.9 Wenchuan Earthquake, where up to 3 km^3 of landslide material was deposited, post-earthquake debris flows have been prevalent and appear to be triggered by the remobilization of landslide sediment (X. Fan et al., 2018b). The deposits from co-seismic landslides can act as sources for post-seismic debris flows that occur with greater frequency and magnitude than pre-earthquake debris flows (Tang et al. 2012; Yu et al. 2013). Catastrophic debris flows continue to occur during periods of extreme rainfall, with notable events occurring in 2008, 2010, 2013, and 2019. The Zhouqu debris flow on August 7, 2010 (C. Tang et al., 2011), Hongchun and Wenjia debris flows on August 13, 2010 (Q. Xu et al., 2012), and Wenchuan debris flows on August 20, 2019 caused notable socio-economic losses. Understanding potential post-earthquake debris-flow volumes is crucial for mitigating losses during the post-earthquake reconstruction.

Post-earthquake debris flows are typically fast-moving, sediment-water mixtures that initiate in one of three ways, either as new landslides, from remobilization of co-seismic landslide debris, or remobilization of in-channel sediment. Here, we define debris flows broadly as mass movements of mixtures of poorly sorted sediment and variable amounts of water during a rainstorm in a catchment (Iverson, 1997). The dynamics of debris flows that control their volume is complex and strongly depend on the rate of bulking that occurs as the flow mobilize and transport landslide and in-channel sediment (Iverson et al. 2011; Horton et al. 2019). While there is no simple relationship between topographic metrics and the mechanisms of bulking; catchment morphometry, geology and hydroclimatic conditions have been used to estimate the potential distribution of debris-flow volumes on debris flow fans (de Haas and Densmore 2019). Empirical relationships between debris-flow volumes and topography have been established in different hydro-climatic contexts including to estimate debris-flow volumes associated with wildfires (Santi et al., 2008) and extreme rainfall (Chang et al., 2011). Simoni et al. (2011) validated that debris-flow volume, inundated area and cross-sectional area have mutual relations based on the vast majority of cases. Marchi et al. (2019) also suggested that there was a weak, but significant correlation between debris-flow volume and catchment area. The evidence from previous work suggests that topographic metrics provide a first-order control on debris-flow volumes such that topographical models may be useful for hazard mapping and analysis.

Machine learning offers a potentially new method to improve debris-flow volume prediction. They are purely statistical in their implementation and make no assumptions regarding triggering conditions. Machine-learning approaches make predictions or decisions based on sample data, known as "training data" without being explicitly programmed (Bishop 2006). Machine learning methods can include neural networks, non-linear regression, and other methods to optimise data for predictive purposes (Table 1). These methods have been applied in landslide assessment and displacement forecasting, are seen as being efficient and reliable measures of these parameters (e.g. Mennis and Guo 2009; Tien Bui et al. 2016; Zhou et al. 2018). For example, Fanos et al. (2018) proposed and evaluated a hybrid model using machine-learning methods and GIS for potential rockfall source identification with an accuracy of 0.92 based on training data and 0.96 on validation data. Kern et al., (2017) proposed an advanced model using machine learning to improve the ability to accurately predict debris flow events in wildfire-prone intermountain western United States. Debris-flow volume is one of the most important parameters to evaluate a potential hazard. Particularly, when designing any protection measures, an acceptable volume-estimation of debris flow has to be defined. However, little existing research attempts to identify potential volumes of debris flows relate to post-earthquake topographic metrics and co-seismic landslide debris. Many studies contain data on the estimation of debris-flow volumes using empirically-based models to correlate debris-flow volume with morphometric catchment characteristics (de Haas and Densmore, 2019; Gartner et al., 2008; Ma et al., 2013; Marchi et al., 2019; Chang et al., 2011 and references therein). Debris-flow volumes calculated using these methods may overestimate the actual volumes by up to two orders of magnitude (Rickenmann, 1999).

After earthquakes, remobilization of co-seismic and in-channel debris increases the potential for debris flows that are of greater volume than has previously been experienced (Fan et al., 2019b). Under these conditions, debris flow hazard depends both on a changing frequency of triggering precipitation (Marra et al., 2017) and a changing magnitude and frequency distribution of debris-flow volumes (R. L. Fan et al., 2018). There has been a significant focus, particularly after the 2008 Wenchuan earthquake, on the first part of this problem (X. Fan et al., 2018b), yet despite this work, both hard and soft engineered structures are often inundated by debris flows that are many times their design capacity. By focusing on the volume part of the problem, we can develop tools that can be used to better understand the scale of debris flows that are possible in a catchment. Machine learning methods allow us to examine their predictive capacity for debris-flow volume, in order to support the engineering design to reduce losses and costs following an earthquake. Prediction of debris flow volume is important for post-earthquake hazard assessment and mitigation because their size and frequency are strongly affected by the total deposited materials in catchments (Bovis and Jakob, 1999).

Table 1. Introduction of several major machine-learning algorithms

Name	Description	Application	Literature
Back propagation-based neural network (BPNN)	A neural network composed of three layers (input, hidden and output), is simply a gradient – descent algorithm that uses to minimize the total error or mean error of target.	Mapping and prediction tool in the geotechnical engineering field, etc.	Neaupane and Achet (2004); Dou et al. (2015); Yang et al. (2019); etc.
Support vector machine (SVM)	A non-linear regression forecasting method, in which the input variables are mapped into a high-dimensional linear feature space through a non-	Landslide susceptibility, displacement forecast model and volume of	Marjanović et al. (2011); Zhou et al. (2016); Xu et al. (2012a); Zhu et al.

	linear transformation.	debris flow prediction, etc.	(2018); etc.
Extreme learning machine (ELM)	A novel training algorithm for single-hidden-layer feedforward neural networks with randomly assigned input weights and biases. The only unknown parameter is the output weights.	Classification and regression problems, etc.	Ding et al. (2015); Cao et al. (2016); Guan et al. (2018); Yoan et al. (2010); etc.
Adaptive boosting machine learning algorithm (AdaBoost)	An adaptive boosting machine learning algorithm, which was designed to facilitate cooperation among weak predictors and to cope with classification problems among the weak predictors.	Landslide susceptibility model, debris flow prediction and facial recognition programs, etc.	Tien Bui et al. (2016a); Pai et al. (2014); Liu et al. (2015); Kadavi et al. (2018); etc.

Here we use morphological features and co-seismic deposits collected from 60 debris-flow catchments in the hardest-hit region by the Wenchuan earthquake, to (1) determine the significant components for volume-estimation of future debris flows, based on correlation analysis and dimensionality reduction to figure out the indeterminate relations between morphometric parameters, volume of deposited materials in each catchment and potential debris-flow volume; (2) propose a hybrid machine-learning model to improve the performance of model computations and reduce the sensitivity of the model to the variations in different conditioning factors, which is composed of extreme learning machine (ELM), particle swarm optimization (PSO)and adaptive boosting machine learning algorithm (AdaBoost); and (3) compare with other machine-learning models (BNPP, SVM, ELM and PSO-ELM), and validate using debris flows triggered in Ludian and Yushu earthquake. The proposed model, therefore, can be suitable and helpful for post-earthquake debris flow assessment and mitigation design-volume estimation.

2. Study area

Longmen Shan, a steep mountain range at the edge of the Tibetan Plateau in Southwestern China has been selected as the study area (Fig. 1). This region comprises 60 typical catchments over 1.2×10^4 km^2 from Yinxiu Town to Beichuan County, through a rugged mountain range with elevations varying between 407m and 6100m above sea level (a.s.l.). Ridges and valleys generally trend NE direction, parallel to the geologic structure, along which the slope gradients are up to 69°, with more than half of the slopes being steeper than 36° (X. Fan et al., 2018a). The mountain range is bisected by major axial drainage basins, such as the Min and Mianyuan Rivers that also act as the main transport routes through the mountains. Debris flows tend to occur in smaller first to fifth order catchments that intersect with these main stem rivers. Beneath the vegetation and thin soil cover, the rocks of the mountainous areas consist mainly of basalt, granite, phyllite, dolomite, limestone, sandstone and shale, and other types of rocks, which range from Precambrian to Cretaceous in age and have a highly fractured and weathered feature. Large thrust earthquakes that generate co-seismic landslides are common in this landscape (Fan et al., 2019a). Post-earthquake debris flows associated with the 2008 M_w 7.9 Wenchuan earthquake have occurred in every monsoonal season since the earthquake. The debris flows initiate in catchments where intense seismic shaking (intensities of XI and X) has greatly increased the volume of deposited materials available to be mobilized (Fig. 1). Catchments often produce debris-flow events more than once, e.g. Wenjia catchment in Fig. 2.

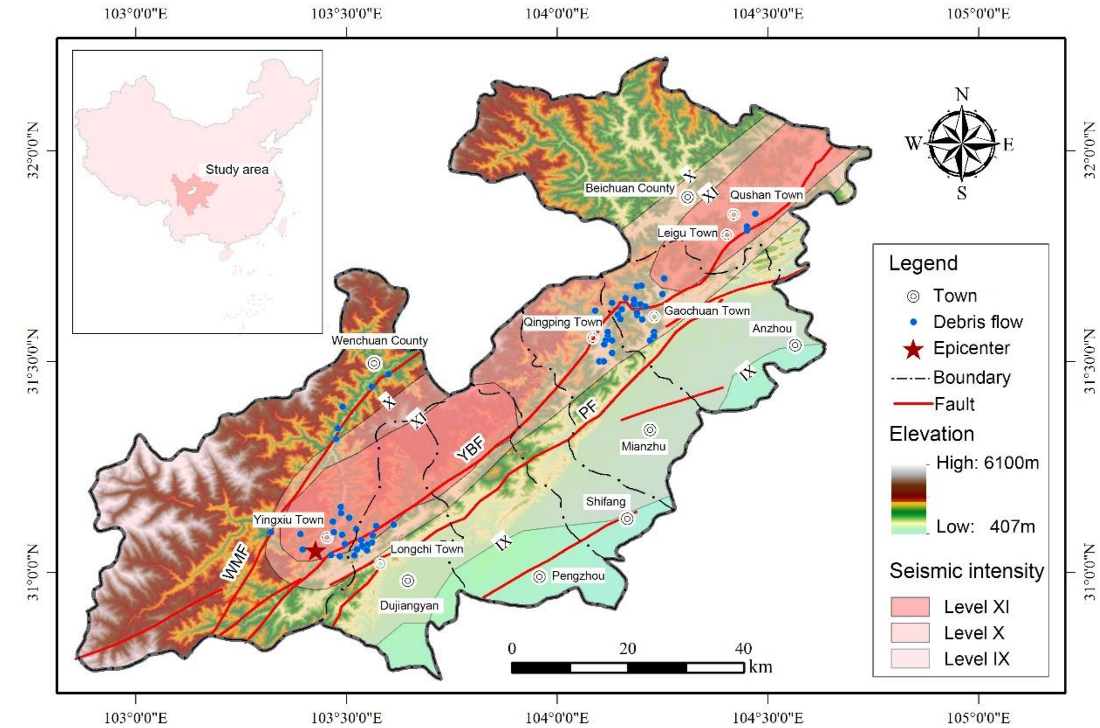


Fig. 1. Map of the study area with topography background based on 30m DEM, county boundaries, the epicenter of the Wenchuan Earthquake and faults (WMF Wenchuan-Maowen fault, YBF Yingxiu-Beichuan fault, PF Pengguan fault.). We collected data from 60 catchments that have experienced debris flows in ten years after the earthquake (blue dots). These debris flow catchments are concentrated close to the faults (red line) and in areas of high seismic intensity (pink polygons).

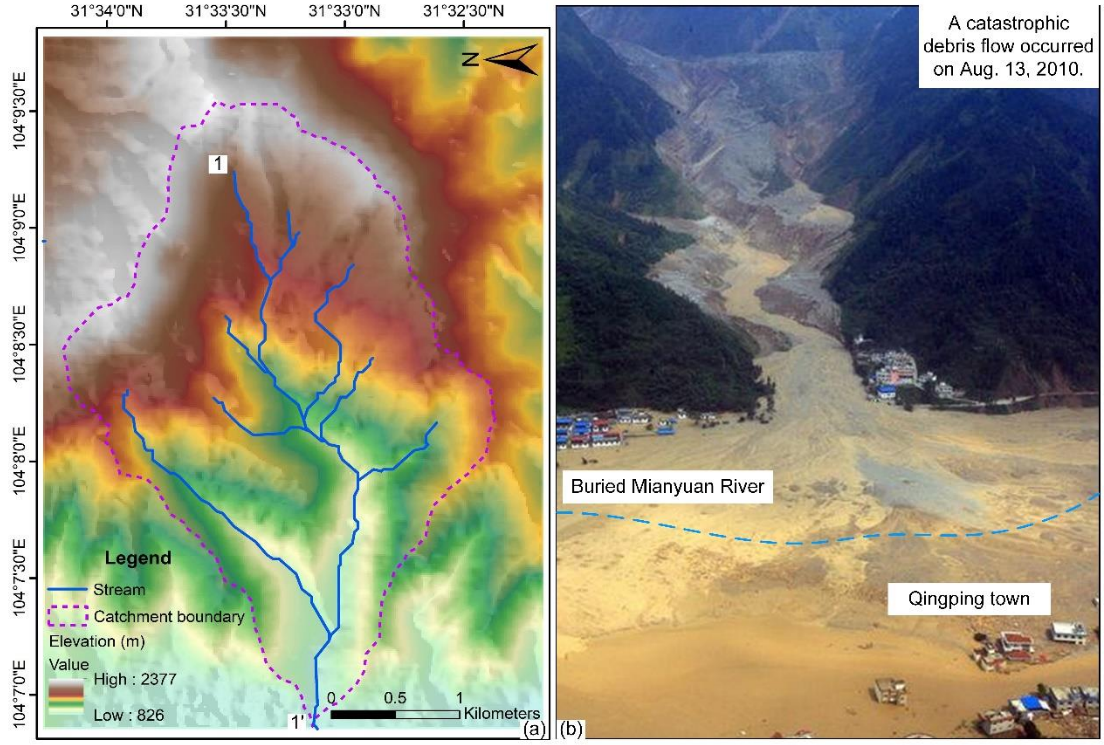


Fig. 2. Examples of the extremely large volume debris-flow events that occurred in the Wenjia catchment region, Qingping Town. (a) The topography of Wenjia catchment based on 10m DEM with a basin boundary and main channel (1-1'). (b) The biggest catastrophic debris flow event occurred on 13-08-2010.

3. Material and Methodology

3.1 Datasets

We focus on 60 typical debris-flow catchments with abundant previous work during the period 2008-2018 in the selected study area. A debris-flow catchment is defined as any first to fifth order catchment that has experienced a debris flow at the catchment mouth. The historical debris-flow events were water-laden masses of soil and fragmented rocks that travelled long distances in the areas with significant gully topography, called gully-type debris flow (Yu et al., 2014). The data are randomly distributed in six different geographical regions, and then a stratified random sampling method is applied to divide the collected data into two datasets with a 9:1 ratio. The first one (54 catchments) is used for building the model whereas the second one (6 catchments randomly split from each area, as shown in Fig. 1 and Table 2) is used for the model testing, respectively. Ten percent of the data using for testification is restricted presently by a limitation of data (Area-6 has only three catchments).

For each catchment, we paid attention to the four morphological factors that have been highlighted in previous studies (Gartner et al., 2008; de Haas and Densmore 2019; Marchi et al. 2019) as having the potential to control debris-flow volume, including catchment area (A), topographic relief (H), channel length (L), and average channel gradient (J). These were all measured from the junction of the tributary catchment and the main axial river drainage using standard algorithms in ArcGIS using 10- and 30-meter digital elevation models (sources of DEMs are from UAV photogrammetry and SRTM, respectively), as most vulnerable linear infrastructure, towns and villages are located along these main rivers. We calculated the main channel length (L) along the main channel from the basin out let to the start of stream within the drainage network derived from the DEM by a flow accumulation threshold (Fig. 2a), H is the change in elevation between the basin mouth and the highest point in the catchment, J is the ratio of H/L. Distance from seismic fault (D) serves as a proxy for the intensity of seismic shaking and frequency of co-seismic landslide deposits (Huang and Li, 2009), as both attenuate rapidly away from thrust faults like the Yingxiu-Beichuan fault. We also used two metrics that were to account for the observations that many of the debris-flows in the Wenchuan area were initiated in co-seismic landslide debris. V is the total volume of co-seismic landslide debris generated in each catchment. Debris-flow volume (V_0) is the debris-flow magnitude defined by Marchi and D'Agostino (2004), means the total volume of debris discharged during a single event, irrespective of the number of surges. Detailed information on debris-flow volumes are difficult to determine due to the wide territorial extent and to the long-time span of the dataset. Most of the data are derived from the available technical journals, e.g. Tiantao (2014), Wentao (2015) and Tang et al. (2010), reports and unpublished documents organized and produced by the local relevant authority management agencies. Fig.3 shows the frequency distributions of some morphometric parameters, and co-seismic landslide volumes in catchments, which are integrated to be the dataset foundation for subsequent determination of significant components. These debris flow events mostly occurred in catchments smaller than 10 km^2, with topographic relief between 500 and 2,000 m, channel lengths less than 7 km, at less than 6 km from the seismic fault. There is a wide range in co-seismic landslide volume, although volume (< 500×10^4 m^3) correspond to 75.0% of the total samples.

Table 2. Dataset of debris-flows in the study area summarized, Area-1: Gaochuan (No.1 - No. 21), Area-2: Qingping (No. 22 - No. 29), Area-3: Yingxiu (No. 30 - No. 42), Area-4: Road 213 (No. 43 - No. 47), Area-5: Longchi (No. 48 - No. 57), Area-6: Beichuan (No. 58 - No. 60). The acronyms mean separately: A - Catchment area, H - Topographic relief, L - Channel length, D - Distance from seismic fault, J - Average channel gradient, V - Total volume of co-seismic landslide debris and V_0 - Debris-flow volume.

No.	A (km^2)	H (m)	L (km)	D (km)	J (‰)	V (10^{4}m^3)	V_0 (10^{4}m^3)
1	14.74	882	4.56	1.46	193.42	485.76	115.11
2	0.69	565	1.93	2.87	292.75	49.53	10.68
3	1.54	374	2.22	3.19	168.47	25.58	6.20

4	0.65	180	1.45	0.57	124.14	7.01	2.80
5	8.17	1150	3.94	0.06	291.88	1603.25	285.39
6	8.91	970	3.10	4.42	312.90	920.70	213.13
7	3.83	640	2.41	0.88	265.56	195.90	58.81
8	0.50	470	1.20	2.08	391.67	42.84	8.24
9	3.25	920	2.61	2.54	352.49	996.46	182.44
10	0.10	188	0.58	1.44	324.14	11.60	2.82
11	2.22	754	1.43	0.85	527.27	375.70	65.23
12	0.72	531	0.74	0.46	717.57	35.10	7.86
13	30.84	910	4.01	3.42	226.93	2000	800
14	2.17	1000	1.15	2.52	869.57	27.66	12.32
15	3.05	698	2.17	1.60	321.66	299.21	69.92
16	0.86	972	1.46	8.86	665.75	64.56	18.12
17	2.97	1027	1.12	10.16	916.96	121.23	66.35
18	3.66	865	1.26	5.30	686.51	592.74	166.35
19	0.74	529	0.78	2.50	678.21	112.51	8.56
20	4.06	772	2.80	5.50	275.71	405.10	104.38
21	0.72	643	0.68	0.26	945.59	45.36	6.35
22	0.86	585	0.74	1.55	790.54	34.51	5.34
23	0.53	463	0.67	1.29	691.04	29.84	4.31
24	0.62	532	0.84	0.17	633.33	35.21	5.38
25	0.69	546	0.74	0.57	737.84	30.64	4.71
26*	7.81	1490	4.90	3.60	304.08	1580.20	657.30
27	1.36	1177	2.59	0.87	454.44	334.30	156.80
28	5.72	986	3.66	0.16	269.40	432.66	108.61
29	8.43	963	4.02	3.80	239.55	964.86	292.10
30	5.24	876	2.08	7.60	421.15	234.64	51.42
31	6.51	886	4.98	10.50	177.91	287.52	67.93
32	6.27	1870	5.60	16.89	333.93	358.26	74.12
33	5.35	1288	3.60	0.13	357.78	358.14	98.40
34	10.70	1842	5.82	0.42	316.49	1151.41	218.72
35	2.18	1820	2.72	0.72	669.12	322.30	80.40
36	54.2	2900	14.2	0.47	204.23	2180.57	505.34
37	0.06	984	1.58	0.65	622.78	222.00	77.30
38	7.50	935	5.20	4.39	179.81	647.48	150.50
39	2.18	1220	2.68	4.93	455.22	122.30	92.51
40	5.21	1678	3.40	2.51	493.53	321.32	89.16
41	1.21	1596	1.14	2.07	140.00	13.80	3.62
42	10.39	1453	5.51	2.61	263.70	727.04	108.91
43	16.49	2382	6.20	0.68	384.19	742.68	194.52
44	1.12	1000	3.25	1.36	307.69	35.93	8.21
45	0.97	1600	1.87	2.10	853.33	48.84	15.30
46	0.29	1580	2.10	1.97	752.38	17.22	4.38
47	21.70	2952	8.90	0.57	331.69	366.67	136.07
48	0.46	650	1.30	1.68	500.00	31.63	8.97
49	8.63	1605	4.45	1.49	360.67	858.94	250.06
50	1.98	965	2.44	0.92	395.29	81.40	28.89
51	1.54	1002	2.35	1.53	426.38	77.60	16.30
52	0.68	952	1.95	1.71	488.21	191.29	46.67
53	8.32	1668	4.76	3.86	350.42	136.02	29.25
54	0.20	434	1.21	0.91	358.68	21.46	6.29
55	0.21	440	1.26	1.82	349.21	19.17	2.54

56	0.29	460	1.35	0.85	340.74	26.13	7.94
57	0.64	660	1.98	4.80	333.33	59.90	15.60
58	1.55	1120	4.01	0.45	279.30	270.16	94.50
59	9.80	1162	4.51	9.58	257.65	1754.64	162.64
60	36.77	1203	11.92	7.04	100.92	1200.26	311.93

* Wenjia catchment in Qingping town (Fig. 2).

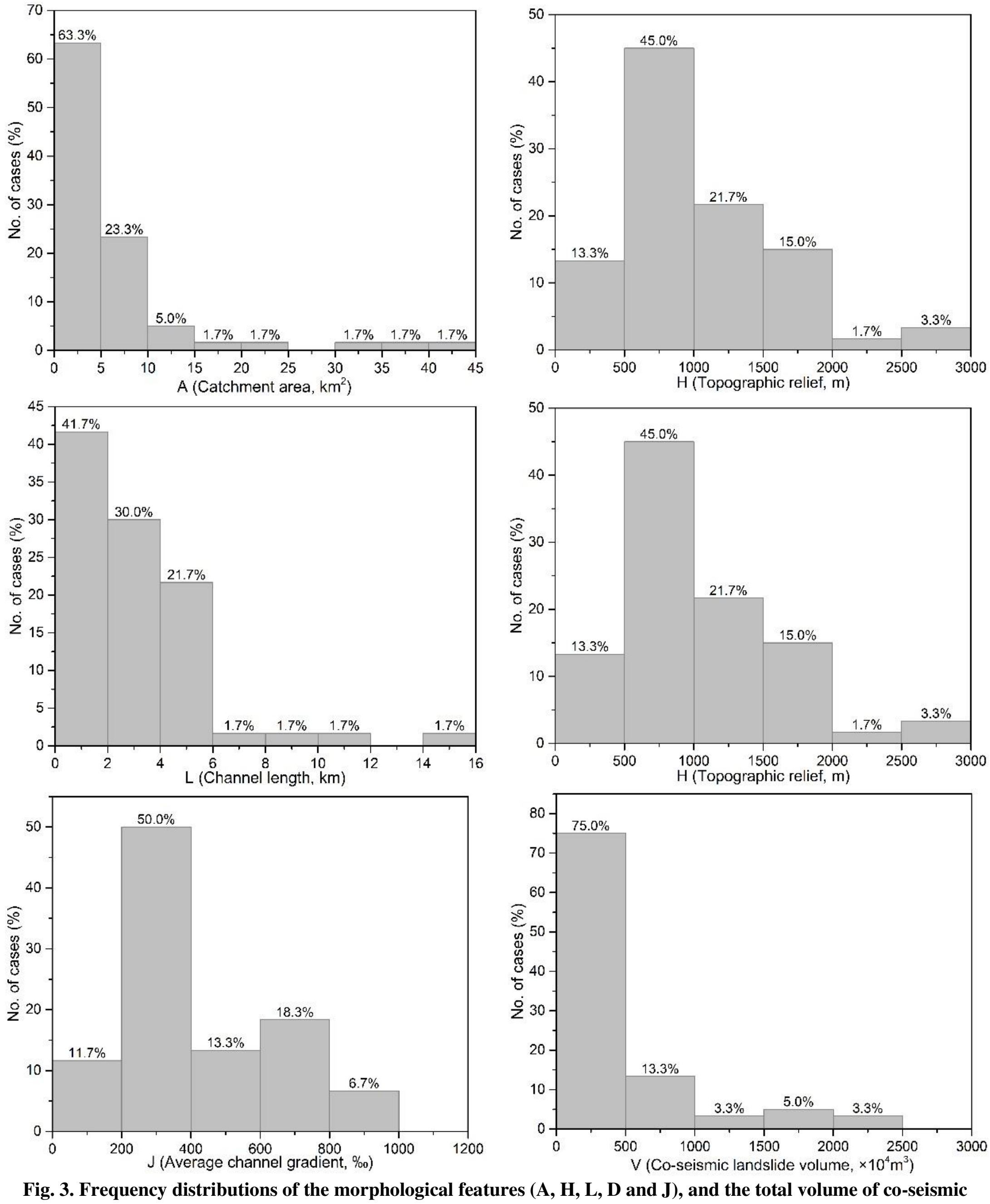


Fig. 3. Frequency distributions of the morphological features (A, H, L, D and J), and the total volume of co-seismic landslide debris (V) in the study region

3.2 Workflow of the proposed hybrid model

The hybrid model was based on AdaBoost mechanism (Freund and Schapire, 1995), a machine learning meta-algorithm that adjusts adaptively to the errors of the weak hypotheses returned by a weak learning algorithm, supported by ELM model (Ding et al., 2015) and PSO algorithm (El-Shorbagy and Hassanien, 2018) for parameter optimization. The processing steps are summarized in Fig. 4. First of all, correlation analysis and dimensionality reduction were applied to determine significant components for the model building. Then, the hybrid model for volume estimation of potential debris flows was trained and tested. Finally, the model was applied to estimate the volumes of debris flows measured after the M_w 6.9 2010 Yushu earthquake and the M_w 6.6 2013 Lushan earthquake.

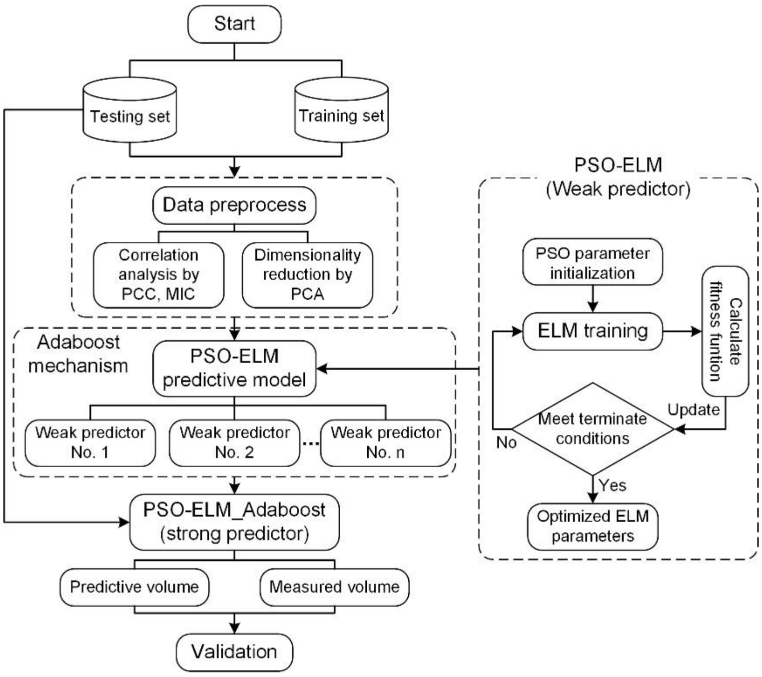


Fig. 4. The overall framework and methodological workflow of the hybrid model

3.2.1 Data preprocess

The first step of data processing is to determine the significant and independent components for model building by correlation analysis (using both linear and non-linear relationships) and dimension reduction methods. Pearson correlation coefficient (PCC) and maximal information coefficient (MIC) are selected and applied to explore the correlations between single factors (A, H, L, D, J and V) and debris-flow volume (V_0). PCC is a well-established measurement of correlation, with a range of +1 (perfect correlation) to -1 (perfect but negative correlation) and 0 denoting the absence of a relation (Adler and Parmryd, 2010). MIC proposed by Reshef et al. (2011) as a measurement of the correlation between two variables. The value of MIC is normalized from 0 to 1, and the larger value indicating a much stronger association between two variables. Principal component analysis (PCA) is one of most common methods to reduce the number of possibly correlated variables into a small number of newly uncorrelated variables, and orthogonal to each other (Nandi et al., 2016). Thus, PCC, MIC and PCA are used to finish data preparation for the model building in subsequent research.

3.2.2 Hybrid model of PSO-ELM_AdaBoost

AdaBoost, a group intelligent predictor, is designed to facilitate mutual cooperation among weak predictors and to cope with forecasting problems among these weak predictors (Pai et al., 2014). So, we can present a hybrid model by AdaBoost mechanism composed of suitable weak predictors. Compared to BPNN and SVM, ELM has a fast learning speed and strong generalization performance (Ding et al. 2014). Therefore, the ELM model is used to identify the weak predictors of the hybrid model, moreover, an evolutionary computational algorithm (PSO) minimizes the loss function by optimizing the weights and thresholds.

(1) Extreme learning machine (ELM)

ELM is a single-hidden layer feedforward neural network (SLFN), basically composed of three layers: the input layer, the hidden layer, and the output layer (Ding et al. 2015). The hidden layer output matrix can be computed by a random assignment to input layer weight matrix and hidden layer biases, such as least-square linear regression. ELM model can be expressed as Eq. (1) and Eq. (2).

$$f_M(x_j) = y_j, \forall j \tag{1}$$

$$\sum_{i=1}^{M} \beta_i G(w_i, b_i, x_j) = t_j, j = 1,2,\cdots,N \tag{2}$$

Set the training set (x_i, t_i), the hidden node output function $G(w, b, x)$, and the number of hidden nodes M. Where x_j represents the input parameters, w_i is the weight vector connecting the ith hidden node, and β_i is the weight vector connecting the ith hidden node and the output nodes. The ELM training contains three steps: (a) randomly generate the weight vector w that connects the input layer and the hidden layer, and generate the hidden layer bias; (b) calculate the hidden layer output matrix H by Eq. (3); (c) calculate the output weight, $\hat{\beta} = H^{+}T$, where H^{+} is the Moore-Penrose generalized inverse of the hidden layer output matrix H, T is the expected output matrix.

$$H(w_1, \cdots, w_L; b_1, \cdots, b_L; x_1, \cdots, x_N) = \begin{bmatrix} G(w_1, b_1, x_1) & \cdots & G(w_L, b_L, x_1) \\ \vdots & \ddots & \vdots \\ G(w_1, b_1, x_N) & \cdots & G(w_L, b_L, x_N) \end{bmatrix}_{N\times L} \tag{3}$$

(2) Particle swarm optimization (PSO)

PSO is a population-based stochastic optimization method with a concise performance and intelligent background (El-Shorbagy and Hassanien, 2018). Inspired by the feeding behavior characteristics of bird flocks, PSO is frequently used to solve the optimization problem. In the PSO model, each particle is described by three basic parameters (position, speed and fitness value), which represents a solution for the target problem. The pursuit process of PSO is implemented through a loop iteration, in which the global best solution can be achieved by adjusting the trajectory of each particle toward its own best location and the entire swarm (Ab Talib and Mat Darus, 2016). Considering that prediction accuracy of ELM may be strongly influenced in modelling, PSO, therefore adopted to determine the appropriate parameters and the model named as PSO-ELM.

(3) The model and performance evaluation

To implement the model, we followed these steps: (a) We initially assigned an equal weight $\{D_t(i)\}$ to each dataset $\{X_i\}$. (b) Then the PSO-ELM based predictor P_t forecast the debris-flow volume series $\{X_i\}$, and the corresponding overall error $\{e_t\}$ by Eq. (4),

$$\begin{cases} e_i = \frac{|X_i - \hat{X}_i|}{X_i}, i = 1,2,\cdots,n \\ e_t = \frac{1}{n}\sum_{i=1}^{n} e_i \end{cases} \tag{4}$$

(c) We computed the series weights for the built predictor P_t: $W_t = \frac{1}{2} ln\,(\frac{1-e_t}{e_t})$ and updated the sampling weights $\{D_t(i)\}$ of the series $\{X_i\}$ by Eq. (5),

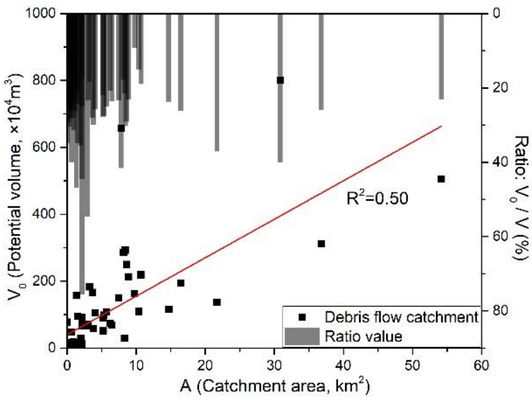
$$\begin{cases} D_t(i) = \frac{D_{t-1}(i)\beta_t^{-e_i}}{Z_t} \\ \beta_t = \frac{e_t}{1-e_t} \end{cases} \tag{5}$$

where Z_t is the normalizing impact which realizes $\sum_{i=1}^{n} D_t(X_i) = 1$. The procedure of steps (a to c) was repeated until all the PSO-ELM based predictors are executed (T). Finally, we summarized all the PSO-ELM based predictors (P_t) in the Adaboost framework to form the final strong predictor (P): $P = \sum_{t=1}^{T} W_t P_t$. Then, root mean square error (RMSE) and mean absolute percentage error (MAPE) provide an assessment of the proposed hybrid model performance.

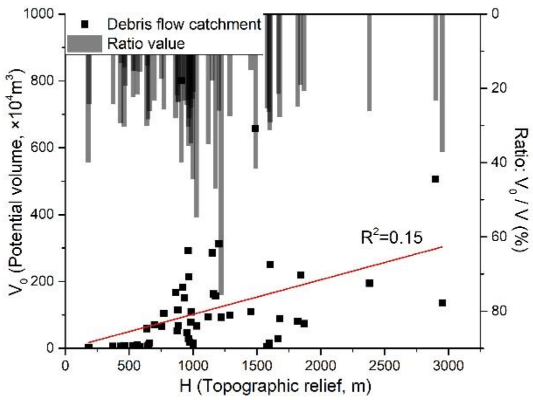
4. Results

4.1 Correlation analysis

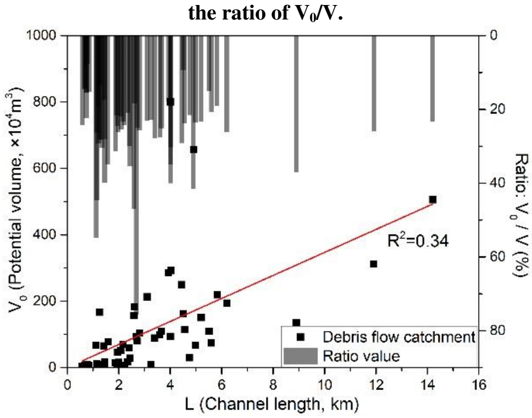
Correlation analysis demonstrates positive linear correlations between debris-flow volume and catchment area, catchment length, and total volume of co-seismic deposits (Fig. 5). V and V_0 were well correlated ($V_0 = 0.26V + 1.04, Pvalue = 0.000, R^2 = 0.80$), while morphologic factors all demonstrate positive correlations, but with lower R^2 values ($A: V_0 = 11.51A + 39.80, Pvalue = 2.09E-10, R^2 = 0.50$; $H: V_0 = 0.10H - 1.72, Pvalue = 0.002, R^2 = 0.15$; $L: V_0 = 34.66L + 0.380, Pvalue = 1.26E-6, R^2 = 0.34$). D has a $Pvalue = 0.86$ higher than 0.05 which indicates strong evidence for the null hypothesis. J has an opposite relation ($J: V_0 = -0.25J + 214.09, Pvalue = 0.006,\ R^2 = 0.13$). Meanwhile, issues of autocorrelation still existed among these factors, e.g. L, H and J. In order to reduce attribute characteristics and meet an assumption of mutual independence among factors in Modelling, dimensionality reduction was applied to ensure the significant components in this work.



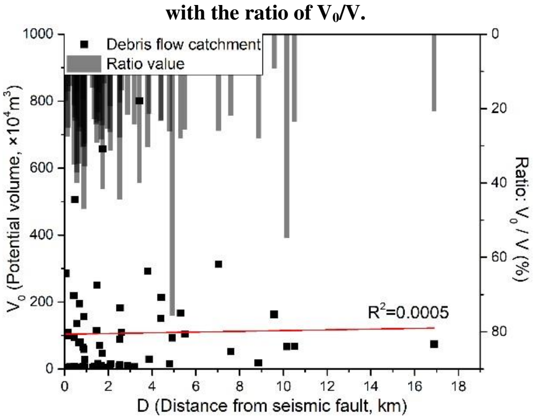
(a) Correlation between A, V_0 and V_0/V. There is a linear correlation between A and V_0, but little correlation with the ratio of V_0/V.



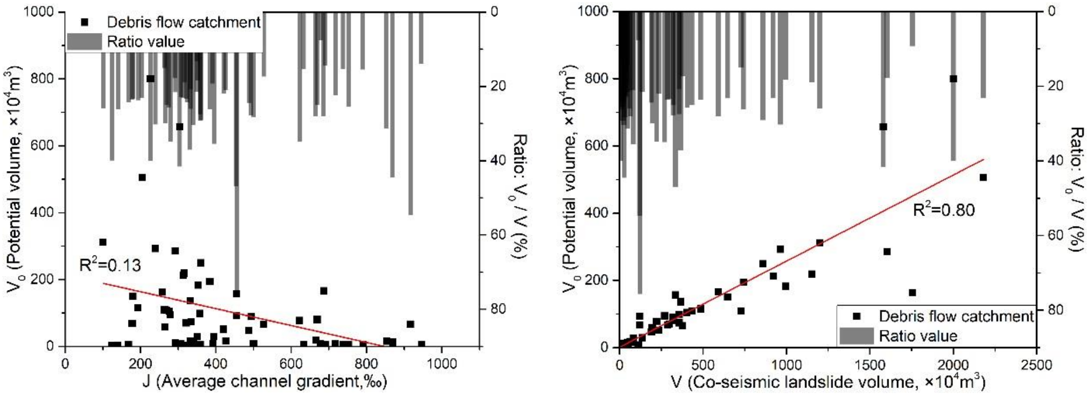
(b) Correlation between H, V_0 and V_0/V. There is a little linear correlation between H and V_0, but little correlation with the ratio of V_0/V.



(c) Correlation between L, V_0 and V_0/V. There is a little linear correlation between L and V_0, but little correlation with the ratio of V_0/V.



(d) Correlation between D, V_0 and V_0/V. There is little linear correlation between D and V_0, but little correlation with the ratio of V_0/V.



(e) Correlation between J, V_0 and V_0 / V. There is a little linear correlation between J and V_0, but little correlation with the ratio of V_0 / V.

(f) Correlation between V, V_0 and V_0 / V. There is a linear correlation between V and V_0, but little correlation with the ratio of V_0 / V.

Fig. 5. Graphs plotted with each catchment's morphological features (A, H, L, D and J), the total volume of co-seismic landslide debris (V) and potential volume of debris flow (V_0), and the ratio (V_0/V).

4.2 Determination of significant components

To measure the correlation both linear and nonlinear between two variables, Pearson correlation coefficient (PCC) and maximal information coefficient (MIC) were applied to calculate the correlations between morphological features (A, H, L, D and J), the total volume of co-seismic landslide debris (V) and debris-flow volume (V_0) (Table 3). It can be seen that correlations obtained from MIC are more significant than those of PCC. PCC even calculates a negative value from the average channel gradient (J). Correlations of MIC show that the sensitivity order is V > A > L> H >J > D, which is consistent with the result from preliminary analysis on the raw dataset.

Table 3. Result of correlation analysis by PCC and MIC

Factor	PCC	MIC	Sensitivity
Catchment area (A)	0.710	0.828	2
Topographic relief (H)	0.389	0.634	4
Channel length (L)	0.579	0.690	3
Distance from seismic fault (D)	0.023	0.251	6
Average channel gradient (J)	-0.354	0.342	5
Co-seismic landslide volume (V)	0.892	0.967	1

Principle component analysis was used to reduce dimensionality (Table 2). Four significant components (*P-1* to *P-4*) have already had a cumulative ratio of up to 94.2 % (Table 4). Subsequently, these four mutually independent significant components (*P-1* to *P-4*) became input parameters of machine learning models. New datasets were produced by a matrix multiplication from the eigenvector (right part of Table 4) and the raw datasets (Table 2).

Table 4. Eigenvalues and eigenvectors of the significant components

Component	Eigenvalue	Ratio* (%)	Cumulative ratio (%)	Eigenvector / Factor	λ_{p-1}	λ_{p-2}	λ_{p-3}	λ_{p-4}
P-1	2.930	48.829	48.829	A	0.308	0.070	0.012	-0.217
P-2	1.025	17.089	65.918	H	0.249	0.437	0.400	-0.051
P-3	0.974	16.230	82.148	L	0.330	0.008	0.053	-0.252
P-4	0.723	12.049	94.197	D	0.050	-0.678	0.700	0.240

P-5	0.302	5.040	99.238	J	-0.182	0.564	0.538	0.325
P-6	0.046	0.762	100.000	V	0.199	0.018	-0.292	1.052

Note: * Ratio is obtained by the equation of eigenvalue/variance.

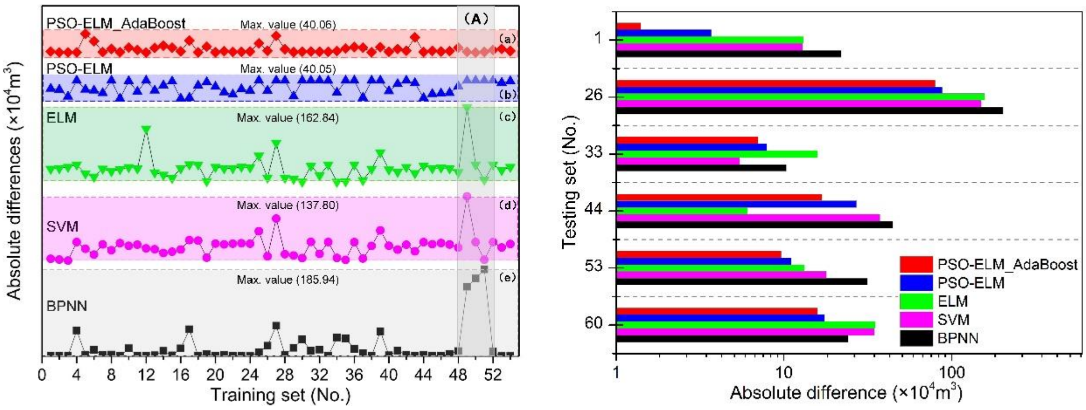
4.3 Results of machine learning modelling

After selecting the four significant components for debris-flow volume estimation, different machine learning models (BPNN, SVM, ELM, PSO-ELM and PSO-ELM_AdaBoost) were trained and tested. Detailed criteria and parameter-setting in the models are shown in Table 5. BPNN composed of a three-layer model is available in MATLAB's built-in toolbox for the debris-flow volume estimation, based on the same training and testing set with parameters setting (Table 5). The SVM model is developed and executed by MATLAB program, and the tuning parameters of the SVM (Table 5) are determined by fivefold cross-validation for its advantage on the average exacted prediction error and circumvents the overfitting problem (Ge et al., 2018). In the ELM model, as shown in Table 5, the only parameter-setting determined by Sigmoid function is the number of neurons in the hidden layer. PSO-ELM model is using PSO algorithm to optimize the connection weight between the input layer and the hidden layer, and the threshold value of the hidden layer neuron in the ELM model. The parameters-setting of PSO-ELM can be found in Table 5. The parameters-setting of PSO-ELM_Adaboost are the same to PSO-ELM, but take ten times the number of iterations (PSO-ELM model).

Table 5. Parameters-setting in different machine learning models

Name	Parameters in modelling
BPNN	A three-layer BPNN composed of an input layer (4 neurons), hidden layer (4 neurons), and an output layer. Initial learning rate (LR) = 0.05, Number of epochs = 5000, Root mean square error (RMSE) = 0.01.
SVM	Radial basis kernel function (RBF) was adopted as the network kernel function. Penalty factor (c) and the parameter of kernel function (g) were 5.6569 and 0.0625, respectively. These tuning parameters of SVM are determined by the cross-validation method, as shown in Appendix Fig. A-1.
ELM	Sigmoid was adopted as an activation function to find the optimal number of hidden layer nodes for the cyclic analysis from 1 to 100. MSE is below 0.03, Number of hidden nodes is up to 57 (Appendix Fig. A-2).
PSO-ELM	The number of hidden nodes is up to 40. In PSO, acceleration coefficients $c_1=1.5$, $c_2=1.7$, inertia weight $w=1$ and *sizepop=20, maxgen=100*. Other parameters are the same as the ELM model (Appendix Fig. A-3).
PSO-ELM AdaBoost	Number of iterations: T = 10. The other parameters are the same as PSO-ELM.

Results can be seen in Fig. 6. The predicted values obtained using PSO-ELM and PSO-ELM_AdaBoost exhibit better agreement with observations than the other models (max. value=40.05, 40.06 < 162.84, 137.80 and 185.94, respectively). The wave range of differences (zone a-e) show a similar conclusion, and PSO-ELM_AdaBoost has much better performance. There is also a good agreement between the estimated and measured debris-flow volumes between the training and testing set, with the lowest error value of 20.14×10^4 m^3 (RMSE) and 8.15 % (MAPE) in the proposed PSO-ELM_AdaBoost model. The lowest of average run time is 0.042 s by ELM model, which is significantly faster than other models. The coupled models (PSO-ELM and PSO-ELM_AdaBoost) are better than single model (BPNN, SVM and ELM) on the prediction performance, but to require longer run times. PSO-ELM_AdaBoost is nearly 3 seconds slower due to its more complex network architecture.



(a) comparison diagram by the training set, zone a-e display the wave range in each model, zone A displays a comparison among the models.

(b) comparison diagram by testing set

Fig. 6. Comparative analysis between predicted and measured value (Absolute difference in Y-axis (a) and X-axis (b) are defined as an absolute value between each predicted value and measured value in different models. X-axis (b) used a logarithmic coordinate.)

We conducted a sensitivity analysis on our training data, by removing the largest one data in the testing set (No. 26 in Table 2 and Fig. 6b). Doing this decreases the uncertainty in the model result. The effect of this large value varied based on the structure of the machine-learning model imposed, with BPNN and PSO-ELM_AdaBoost having a much large effect than SVM, ELM, and PSO-ELM. In order to test the generalization ability of this proposed hybrid model and evaluate whether it is helpful to the volume estimation of future debris flow trigged by other earthquakes. We ran three further tests on debris-flows associated with different earthquakes; Lengmu catchment and Zhonggang catchment in Lushan earthquake of M_w 6.6 (20 April 2013) and Buqinglong catchment in Yushu earthquake of M_w 6.9 (14 April 2010).

Based on the three further tests from similar basic morphological features and measured loose material source, the predicted values of the potential volume of future debris flow are larger than the measured ones (Table 6). The absolute percentage error (APE) exceeds 11%, which is larger than the result of the testing set (8.15%). It seems that the calibrated model performs well within the Longmen Shan, a region of diverse geology and topography. The increased error from the three cases of the application indicates that there still have limitations to apply this proposed model to other debris-flow catchments in different seismic regions. The reason is debris-flow volume likewise influenced by the expected rainfall conditions and lithologic characteristics of the catchment. These factors should be considered in the subsequent study. In spite of these limitations, the hybrid model is a helpful role in volume estimation of future debris flows. This comes with two advantages for future planning for debris flows; that the model accuracy improves with additional data and produces acceptable accuracy for the engineering design to protect the safety and property in the seismic areas. The first advantage reflects the flexibility of machine learning methods. The relationships between debris-flow volume and the factors framed within a neural network allows the prediction accuracy to improve as new data are introduced to the model. The frequency-size distribution of debris flows can be incorporated with the previous work on the other landslide types (Malamud et al., 2004) to quantify the severity of post-earthquake debris-flow events. As such, this allows the model to become smarter with time. The second is related to model predictive capacity. After the Wenchuan earthquake, post-earthquake debris flows occurred suddenly and at a magnitude never experienced within the region. As part of the 3-year post-earthquake reconstruction plan, many hundreds of engineered debris flow structures were created to mitigate the effects of these hazardous events. Many of these structures were designed without a clear understanding of the potential volumes of debris flows that could be produced after this event. Our model provides a simple application for the estimation of the potential for the largest debris-flow volumes from a catchment. Given the correlations with the volume of co-seismic landslide debris and catchment size and slope, there is significant potential for this method to be used for the development of engineering structures that can mitigate the largest possible debris-flows. Furthermore, this model can be transferred between earthquake events within the same region. This suggests that a well-calibrated regional machine learning model could potentially act as a useful debris-flow volume prediction tool for the immediate aftermath of an earthquake.

Table 6. Model validation in other seismic regions

Name	A (km^2)	H (m)	L (km)	D (km)	J (‰)	V ($10^4 m^3$)	V_0 (Measured value, $10^4 m^3$)	V_0 (Predicted value by PSO-ELM_Adaboost, $10^4 m^3$)	APE (%)
Buqinglong catchment	19.80	1016	3.03	8.60	335.31	90.05	42.32	47.11	11.32
Lengmu catchment	9.44	2048	3.98	15.10	514.57	381.87	68.60	76.46	11.46
Zhonggang catchment	17.76	2235	9.72	17.48	229.94	600.42	96.05	111.53	16.12

5. Discussion and Conclusion

Our machine learning model predicts the magnitude of that largest single debris-flow event in a given catchment by analyzing the characteristics of 60 debris-flow catchments in the hardest-hit regions of the Wenchuan earthquake in the decade after the earthquake. The presented model (PSO-

ELM_AdaBoost) composed of four-significant components demonstrates an uncertainty in the prediction of the largest possible debris-flows in a catchment of between 11 % and 16 %. A comparison of the developed model with the existing semi-empirical function originated from worldwide debris-flow events (de Haas and Densmore 2019; Simoni et al. 2011; Ma et al. 2013) and the fitting function between the total volume of co-seismic landslide debris (V) and debris-flow volume (V_0) using the three cases trigged by other earthquakes is made. The results are shown in Fig. 7. The volume from the fitting function is much lower than the measured value in Buqinglong catchment but much higher in Lengmu catchment, and even beyond the figure's boundary in Zhonggang catchment (157×10^4 m^3). The proposed hybrid model is evidently supported by over-prediction to larger volumes than the measured ones, otherwise, there is also under-prediction to smaller volumes by the empirical-statistical functions. The estimated accuracy of the machine-learning model is better than the empirical models with specific consideration of the total volume of co-seismic landslide debris, and training dataset in the given regions. The standard errors of estimated volumes by Ma et al. (2013) are much larger than ones by de Haas and Densmore (2019), for the possible reason that the empirical formulas have only considered the statistical relationship between loose material volume and debris-flow volume. Similarly, the equation of linear regression between V and V_0 obviously has a poor performance on volume estimation of debris flows.

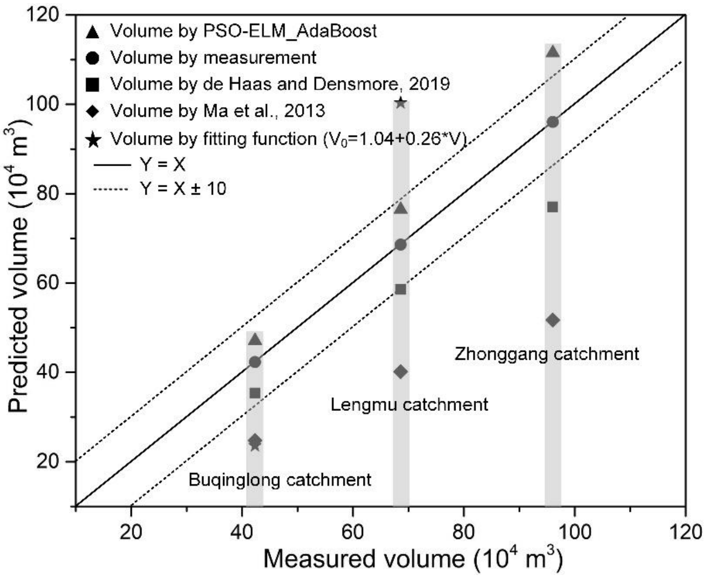


Fig. 7. Comparison of the volume-estimations of debris flows by PSO-ELM_AdaBoost, measurement and semi-empirical function. The fitting function is between the total volume of co-seismic landslide debris (V) and debris-flow volume (V_0) in Fig. 5 (f). The solid black line indicates a perfect fit and dotted lines represent the plus and minus one specific tolerance value (e.g., 10×10^4 m^3).

For a given catchment, the debris-flow volume depends on the amount of sediment available and the potential of the flow to mobilize and transport along the debris-flow path, therefore it might be regarded as a function of catchment morphometry, geology and hydroclimatic conditions (de Haas and Densmore, 2019). The debris-flow volumes have a relation of proportional growth with basin area which supplies loose materials from widespread sediment sources, especially along the main channel. Co-seismic landslides increase the amount of loose debris on slopes and in gully floors. As a consequence, debris flow frequency increases, and large debris flows occur at lower rainfall rates in the years immediately following strong earthquakes (Ma et al., 2013; Guo et al., 2016). Therefore, even though there is close correlation between the volume of loose material available within the basin after the earthquake and the debris-flow volume (X. Guo et al., 2016), other parameters (e.g., rainfall intensity or slope gradient) still can influence the debris-flow magnitude, especially during a single event. After the M_w 7.6 Chi-Chi earthquake in Taiwan, Chen et al. (2011) presented a recovery equation to describe the variation in the rainfall threshold for triggering debris flow after the earthquake and evaluated the recovery period. However, similar attempts in the Wenchuan context have been challenging to implement. In the instance of such large epistemic uncertainty, machine learning models may provide a useful alternative to process-based or semi-empirical models, particularly in the specific case of debris flow volumes. The machine learning method makes no assumptions about parameter correlations, instead compiles a range of data and produces the most optimised result. In these specific cases, where triggering and bulking conditions are changing rapidly, our work demonstrates that machine learning methods may be a powerful tool to aid hazard mitigation. Hence, we strongly recommended using the presented model to estimate the volumes of debris flows with careful attention to the specific circumstance. In practical usage, therefore, a tolerance value (e.g. 10×10^4 m^3 in Fig. 7) can be included in the volume estimation of future debris flows, which has been proved to have a much better estimation performance combined with the proposed machine-learning model.

While we do not explicitly calculate the frequency for the largest potential landslide volume, our modelling implicitly calculates the largest potential debris flow that will occur within the 10 years after a strong earthquake. Evidence from a number of large earthquakes has demonstrated that high debris flow rates are common immediately after an earthquake, decaying to background rates within 4 – 10 years (Marc et al., 2019). Hence our modelling has an implicit timescale of within 10 years after a large earthquake. This timescale is particularly important for earthquake recovery, as post-earthquake debris flows can affect vulnerable and displaced communities. In the post-Wenchuan earthquake case, the building of debris flow check dams primarily occurred within 3 years of the earthquake, based on standard, linear equations that dramatically underestimated the potential volume of debris flows and were often inundated (Chuan Tang et al., 2011). There is a strong desire for better predictive capacity of hazard volumes in these key few years after an earthquake. We demonstrate the potential power of machine learning as a tool that can be translated, albeit with an increase in uncertainty, to earthquake events in similar topographic, geologic, and hydrologic settings. Thus, our application of machine learning presents an alternative to more traditional methods for estimating debris flow susceptibility. However, as the model does not include a specific temporal component, it does not attempt to model debris flow hazard for a particular catchment.

In conclusion, the type of machine learning model chosen affects the robustness of the model result, with the hybrid model (PSO-ELM_AdaBoost) showing the strongest correlations with the measured volumes in the test data. Importantly, the uncertainty does not decrease when applied to debris-flows associated with different earthquakes of different magnitudes in the same tectonic setting (the collision region of the India Plate with the Eurasia Plate). This result suggests the machine-learning methods could prove useful as initial estimates of debris-flow potential after earthquakes.

Acknowledgements

This research work was financially supported by the State Key Laboratory of Geo-hazard Prevention and Geo-environment Protection (Chengdu University of Technology) (Grant No. SKLGP2017Z006, SKLGP2015Z006) and Creative Research Groups of China (Grant No. 41521002), International Cooperation (NSFC-RCUK_NERC), Resilience to Earthquake-induced landslide risk in China (grant no. 41661134010; NE/N012240/1). The authors also give great thanks to the editor and reviewers for their work to improve the quality of the paper.

Appendix

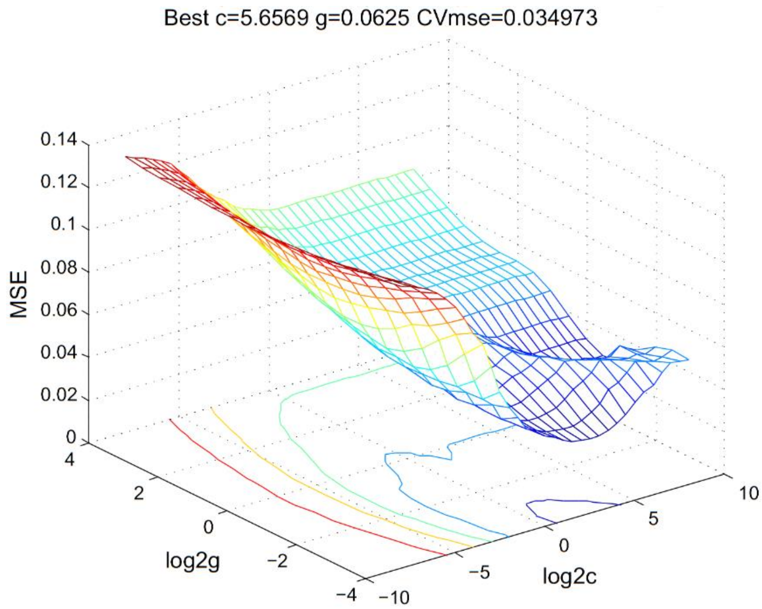


Fig. A-1. Parameters determination in SVM model

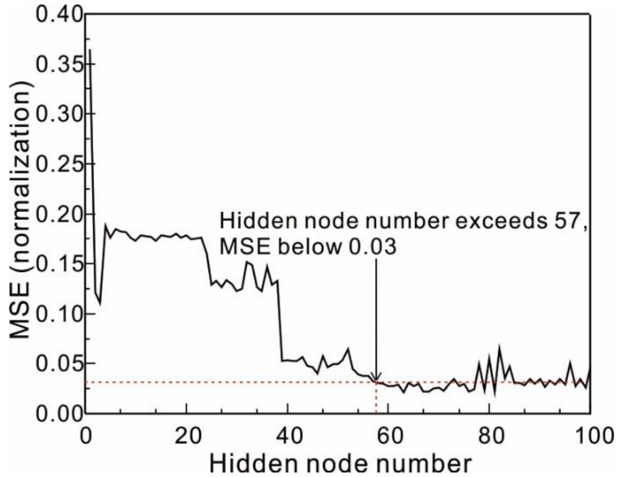


Fig. A-2. Parameters determination in ELM model

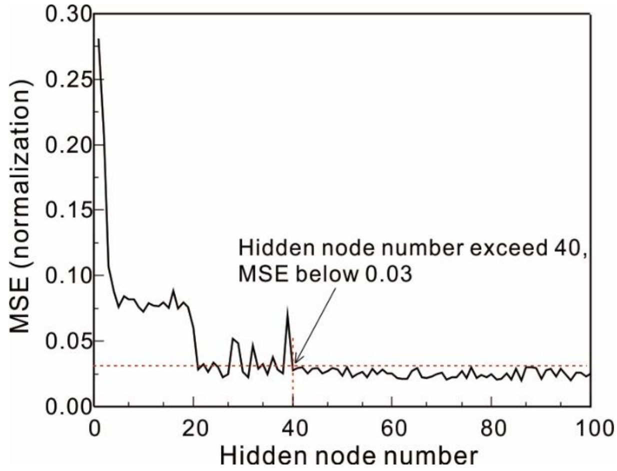


Fig. A-3. Parameters determination in PSO-ELM model

References

Ab Talib, M.H., Mat Darus, I.Z., 2016. Intelligent fuzzy logic with firefly algorithm and particle swarm optimization for semi-active suspension system using magneto-rheological damper. J. Vib. Control 23, 501–514. https://doi.org/10.1177/1077546315580693

Adler, J., Parmryd, I., 2010. Quantifying colocalization by correlation: The pearson correlation coefficient is superior to the Mander's overlap coefficient. Cytom. Part A 77, 733–742. https://doi.org/10.1002/cyto.a.20896

Bovis, M.J., Jakob, M., 1999. The role of debris supply conditions in predicting debris flow activity. Earth Surf. Process. Landforms 24, 1039–1054. https://doi.org/10.1002/(SICI)1096-9837(199910)24:11<1039::AID-ESP29>3.0.CO;2-U

Cao, Y., Yin, K., Alexander, D.E., Zhou, C., 2016. Using an extreme learning machine to predict the displacement of step-like landslides in relation to controlling factors. Landslides 13, 725–736. https://doi.org/10.1007/s10346-015-0596-z

Chang, C.-W., Lin, P.-S., Tsai, C.-L., 2011. Estimation of sediment volume of debris flow caused by extreme rainfall in Taiwan. Eng. Geol. 123, 83–90. https://doi.org/10.1016/j.enggeo.2011.07.004

Chen, H., Lin, G., Lu, M., Shih, T., Horng, M., Wu, S., Chuang, B., 2011. Effects of topography , lithology , rainfall and earthquake on landslide and sediment discharge in mountain catchments of southeastern Taiwan. Geomorphology 133, 132–142. https://doi.org/10.1016/j.geomorph.2010.12.031

Chen, J.C., 2011. Variability of impact of earthquake on debris-flow triggering conditions: Case study of Chen-Yu-Lan watershed, Taiwan. Environ. Earth Sci. 64, 1787–1794. https://doi.org/10.1007/s12665-011-0981-4

Chuan T, Jun D, Jingtao L, 2010. Remote sensing images based observational analysis on charcters of debris flow source areas in Beichuan county of Wenchuan Earthquake epicenter region. Journal of Engineering Geology, 18(1):1-7 (in Chinese)

de Haas, T., Densmore, A.L., 2019. Debris-flow volume quantile prediction from catchment morphometry. Geology 47, 791–794. https://doi.org/10.1130/g45950.1

Ding, S., Xu, X., Nie, R., 2014. Extreme learning machine and its applications. Neural Comput. Appl. 25, 549–556. https://doi.org/10.1007/s00521-013-1522-8

Ding, S., Zhao, H., Zhang, Y., Xu, X., Nie, R., 2015. Extreme learning machine: algorithm, theory and applications. Artif. Intell. Rev. 44, 103–115. https://doi.org/10.1007/s10462-013-9405-z

Dou, J., Yamagishi, H., Pourghasemi, H.R., Yunus, A.P., Song, X., Xu, Y., Zhu, Z., 2015. An integrated artificial neural network model for the landslide susceptibility assessment of Osado Island, Japan. Nat. Hazards 78, 1749–1776. https://doi.org/10.1007/s11069-015-1799-2

El-Shorbagy, M.A., Hassanien, A.E., 2018. Particle Swarm Optimization from Theory to Applications. Int. J. Rough Sets Data Anal. 5, 1–24. https://doi.org/10.4018/ijrsda.2018040101

Fan, R.L., Zhang, L.M., Wang, H.J., Fan, X.M., 2018. Evolution of debris flow activities in Gaojiagou Ravine during 2008–2016 after the Wenchuan earthquake. Eng. Geol. 235, 1–10. https://doi.org/10.1016/j.enggeo.2018.01.017

Fan, X., Domènech, G., Scaringi, G., Huang, R., Xu, Q., Hales, T.C., Dai, L., Yang, Q., Francis, O., 2018a. Spatio-temporal evolution of mass wasting after the 2008 Mw 7.9 Wenchuan earthquake revealed by a detailed multi-temporal inventory. Landslides 15, 2325–2341. https://doi.org/10.1007/s10346-018-1054-5

Fan, X., Juang, C.H., Wasowski, J., Huang, R., Xu, Q., Scaringi, G., van Westen, C.J., Havenith, H.B., 2018b. What we have learned from the 2008 Wenchuan Earthquake and its aftermath: A decade of research and challenges. Eng. Geol. 241, 25–32. https://doi.org/10.1016/j.enggeo.2018.05.004

Fan, X., Scaringi, G., Domènech, G., Yang, F., Guo, X., Dai, L., He, C., Xu, Q., Huang, R., 2019a. Two multi-temporal datasets that track the enhanced landsliding after the 2008 Wenchuan earthquake. Earth Syst. Sci. Data 11, 35–55. https://doi.org/10.5194/essd-11-35-2019

Fan, X., Scaringi, G., Korup, O., West, A.J., van Westen, C.J., Tanyas, H., Hovius, N., Hales, T.C., Jibson, R.W., Allstadt, K.E., Zhang, L., Evans, S.G., Xu, C., Li, G., Pei, X., Xu, Q., Huang, R., 2019b. Earthquake-Induced Chains of Geologic Hazards: Patterns, Mechanisms, and Impacts. Rev. Geophys. 57. https://doi.org/10.1029/2018RG000626

Fanos, A.M., Pradhan, B., Mansor, S., Yusoff, Z.M., Abdullah, A.F. bin, 2018. A hybrid model using machine learning methods and GIS for potential rockfall source identification from airborne laser scanning data. Landslides 15, 1833–1850. https://doi.org/10.1007/s10346-018-0990-4

Freund, Y., Schapire, R.E., 1995. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting, in: European Conference on Computational Learning Theory. Springer Berlin Heidelberg, pp. 23–37.

Gartner, J.E., Cannon, S.H., Santi, P.M., Dewolfe, V.G., 2008. Empirical models to predict the volumes of debris flows generated by recently burned basins in the western U.S. Geomorphology 96, 339–354. https://doi.org/10.1016/j.geomorph.2007.02.033

Ge, Y., Chen, H., Zhao, B., Tang, H., Lin, Z., Xie, Z., Lv, L., Zhong, P., 2018. A comparison of five methods in landslide susceptibility assessment: a case study from the 330-kV transmission line in Gansu Region, China. Environ. Earth Sci. 77. https://doi.org/10.1007/s12665-018-7814-7

Guan, S., Zhu, Y., Zhou, L., Deng, H., Nan, X., 2018. Prediction of Landslide Displacement Using EMD-PSO-ELM with Multiple Factors. Proc. - 7th Int. Conf. Digit. Home, ICDH 2018 230–235. https://doi.org/10.1109/icdh.2018.00048

Guo, Xiaojun, Cui, P., Li, Y., Fan, J., Yan, Y., 2016. Temporal differentiation of rainfall thresholds for debris flows in Wenchuan earthquake-affected areas. Environ. Earth Sci. 75, 1–12. https://doi.org/10.1007/s12665-015-5031-1

Guo, X., Cui, P., Li, Y., Zou, Q., Kong, Y., 2016. The formation and development of debris flows in large watersheds after the 2008 Wenchuan Earthquake. Landslides 13, 25–37. https://doi.org/10.1007/s10346-014-0541-6

Horton, A.J., Hales, T.C., Ouyang, C., Fan, X., 2019. Identifying post-earthquake debris flow hazard using Massflow. Eng. Geol. 258. https://doi.org/10.1016/j.enggeo.2019.05.011

Huang, R., Li, W., 2009. Development and distribution of geohazards triggered by the 5.12 Wenchuan Earthquake in China. Sci. China, Ser. E Technol. Sci. 52, 810–819. https://doi.org/10.1007/s11431-009-0117-1

Iverson, R.M., 1997. The physics of debris flows. Rev. Geophys. 35, 245–296.

Iverson, R.M., Reid, M.E., Logan, M., LaHusen, R.G., Godt, J.W., Griswold, J.P., 2011. Positive feedback and momentum growth during debris-flow entrainment of wet bed sediment. Nat. Geosci. 4, 116–121. https://doi.org/10.1038/ngeo1040

Kadavi, P.R., Lee, C.-W.W., Lee, S., 2018. Application of Ensemble-Based Machine Learning Models to Landslide Susceptibility Mapping. Remote Sens. 10, 1252. https://doi.org/10.3390/rs10081252

Kern, A.N., Addison, P., Oommen, T., Salazar, S.E., Coffman, R.A., 2017. Machine Learning Based Predictive Modeling of

Debris Flow Probability Following Wildfire in the Intermountain Western United States. Math. Geosci. 49, 717–735. https://doi.org/10.1007/s11004-017-9681-2

Liu, H., Tian, H.Q., Li, Y.F., Zhang, L., 2015. Comparison of four Adaboost algorithm based artificial neural networks in wind speed predictions. Energy Convers. Manag. 92, 67–81. https://doi.org/10.1016/j.enconman.2014.12.053

Ma, C., Hu, K., Tian, M., 2013. Comparison of debris-flow volume and activity under different formation conditions. Nat. Hazards 67, 261–273. https://doi.org/10.1007/s11069-013-0557-6

Malamud, B.D., Turcotte, D.L., Guzzetti, F., Reichenbach, P., 2004. Landslide inventories and their statistical properties 711, 687–711. https://doi.org/10.1002/esp.1064

Marc, O., Behling, R., Andermann, C., Turowski, J.M., Illien, L., Roessner, S., Hovius, N., 2019. Long-term erosion of the Nepal Himalayas by bedrock landsliding : the role of monsoons , earthquakes and giant landslides. Earth Surf. Dyn. 7, 107–128.

Marchi, L., Brunetti, M.T., Cavalli, M., Crema, S., 2019. Debris-flow volumes in northeastern Italy: Relationship with drainage area and size probability. Earth Surf. Process. Landforms 44, 933–943. https://doi.org/10.1002/esp.4546

Marchi, L., D'Agostino, V., 2004. Estimation of debris-flow magnitude in the Eastern Italian Alps. Earth Surf. Process. Landforms 29, 207–220. https://doi.org/10.1002/esp.1027

Marjanović, M., Kovačević, M., Bajat, B., Voženílek, V., 2011. Landslide susceptibility assessment using SVM machine learning algorithm. Eng. Geol. 123, 225–234. https://doi.org/10.1016/j.enggeo.2011.09.006

Marra, F., Destro, E., Nikolopoulos, E.I., Zoccatelli, D., Dominique Creutin, J., Guzzetti, F., Borga, M., 2017. Impact of rainfall spatial aggregation on the identification of debris flow occurrence thresholds. Hydrol. Earth Syst. Sci. 21, 4525–4532. https://doi.org/10.5194/hess-21-4525-2017

Mennis, J., Guo, D., 2009. Spatial data mining and geographic knowledge discovery-An introduction. Comput. Environ. Urban Syst. 33, 403–408. https://doi.org/10.1016/j.compenvurbsys.2009.11.001

Nandi, A., Mandal, A., Wilson, M., Smith, D., 2016. Flood hazard mapping in Jamaica using principal component analysis and logistic regression. Environ. Earth Sci. 75, 1–16. https://doi.org/10.1007/s12665-016-5323-0

Neaupane, K.M., Achet, S.H., 2004. Use of backpropagation neural network for landslide monitoring: a case study in the higher Himalaya. Eng. Geol. 74, 213–226. https://doi.org/10.1016/j.enggeo.2004.03.010

Pai, P.F., Li, L.L., Hung, W.Z., Lin, K.P., 2014. Using ADABOOST and Rough Set Theory for Predicting Debris Flow Disaster. Water Resour. Manag. 28, 1143–1155. https://doi.org/10.1007/s11269-014-0548-8

Reshef, D.N., Reshef, Y.A., Finucane, H.K., Grossman, S.R., McVean, G., Turnbaugh, P.J., Lander, E.S., Mitzenmacher, M., Sabeti, P.C., 2011. Detecting novel associations in large data sets. Science (80-.). 334, 1518–1524. https://doi.org/10.1126/science.1205438

Rickenmann, D., 1999. Empirical Relationships for Debris Flows. Nat. Hazards 19, 47–77.

Santi, P.M., deWolfe, V.G., Higgins, J.D., Cannon, S.H., Gartner, J.E., 2008. Sources of debris flow material in burned areas. Geomorphology 96, 310–321. https://doi.org/10.1016/j.geomorph.2007.02.022

Simoni, A., Mammoliti, M., Berti, M., 2011. Uncertainty of debris flow mobility relationships and its influence on the prediction of inundated areas. Geomorphology 132, 249–259. https://doi.org/10.1016/j.geomorph.2011.05.013

Tang, C., Rengers, N., Van Asch, T.W.J., Yang, Y.H., Wang, G.F., 2011. Triggering conditions and depositional characteristics of a disastrous debris flow event in Zhouqu city, Gansu Province, northwestern China. Nat. Hazards Earth Syst. Sci. 11, 2903–2912. https://doi.org/10.5194/nhess-11-2903-2011

Tang, C., Van Asch, T.W.J., Chang, M., Chen, G.Q., Zhao, X.H., Huang, X.C., 2012. Catastrophic debris flows on 13 August 2010 in the Qingping area, southwestern China: The combined effects of a strong earthquake and subsequent rainstorms. Geomorphology 139–140, 559–576. https://doi.org/10.1016/j.geomorph.2011.12.021

Tang, Chuan, Zhu, J., Ding, J., Cui, X., Chen, L., Zhang, J., 2011. Catastrophic debris flows triggered by a 14 August 2010 rainfall at the epicenter of the Wenchuan earthquake. Landslides 8, 485–497. https://doi.org/10.1007/s10346-011-0269-5

Tiantao L, 2014. Study on the Debris flow characteristics and initiation mechanisms in the Wenchuan Earthquake areas.Master thesis in Chengdu University of Technology(in Chinese)

Tien Bui, D., Ho, T.-C.C., Pradhan, B., Pham, B.-T.T., Nhu, V.-H.H., Revhaug, I., 2016a. GIS-based modeling of rainfall-induced landslides using data mining-based functional trees classifier with AdaBoost, Bagging, and MultiBoost ensemble frameworks. Environ. Earth Sci. 75, 1–22. https://doi.org/10.1007/s12665-016-5919-4

Tien Bui, D., Tuan, T.A., Klempe, H., Pradhan, B., Revhaug, I., 2016b. Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. Landslides 13, 361–378. https://doi.org/10.1007/s10346-015-0557-6

Wentao G, 2015. Study on provenance characteristics and initiation mechanism of debris flows in Meizoseismal areas of Wenchuan Earthquake-taking study areas, Gaochuan village, Qingping village and yinxiu town as example. Master thesis in Chengdu University of Technology(in Chinese)

Xu, C., Dai, F., Xu, X., Lee, Y.H., 2012. GIS-based support vector machine modeling of earthquake-triggered landslide susceptibility in the Jianjiang River watershed, China. Geomorphology 145–146, 70–80. https://doi.org/10.1016/j.geomorph.2011.12.040

Xu, Q., Zhang, S., Li, W.L., Van Asch, T.W.J., 2012. The 13 August 2010 catastrophic debris flows after the 2008 Wenchuan earthquake, China. Nat. Hazards Earth Syst. Sci. 12, 201–216. https://doi.org/10.5194/nhess-12-201-2012

Yang, B., Yin, K., Lacasse, S., Liu, Z., 2019. Time series analysis and long short-term memory neural network to predict landslide displacement. Landslides 16, 677–694. https://doi.org/10.1007/s10346-018-01127-x

Yoan, M., Sorjamaa, A., Bas, P., Simula, O., Jutten, C., Lendasse, A., Miche, Y., Sorjamaa, A., Bas, P., Simula, O., Jutten, C., Lendasse, A., Yoan, M., Sorjamaa, A., Bas, P., Simula, O., Jutten, C., Lendasse, A., 2010. OP-ELM: Optimally Pruned Extreme Learning Machine. IEEE Trans. Neural Networks 21, 158–162. https://doi.org/10.1109/tnn.2009.2036259

Yu, B., Ma, Y., Wu, Y.F., 2013. Case study of a giant debris flow in the Wenjia Gully, Sichuan Province, China. Nat. Hazards 65, 835–849. https://doi.org/10.1007/s11069-012-0395-y

Yu, B., Zhu, Yuan, Wang, T., Chen, Y., Zhu, Yunbo, Tie, Y., Lu, K., 2014. A prediction model for debris flows triggered by a runoff-induced mechanism. Nat. Hazards 74, 1141–1161. https://doi.org/10.1007/s11069-014-1234-0

Zhou, C., Yin, K., Cao, Y., Ahmed, B., 2016. Application of time series analysis and PSO–SVM model in predicting the Bazimen landslide in the Three Gorges Reservoir, China. Eng. Geol. 204, 108–120. https://doi.org/10.1016/j.enggeo.2016.02.009

Zhou, C., Yin, K., Cao, Y., Intrieri, E., Ahmed, B., Catani, F., 2018. Displacement prediction of step-like landslide by applying a novel kernel extreme learning machine method. Landslides 15, 2211–2225. https://doi.org/10.1007/s10346-018-1022-0

Zhu, X., Ma, S. qi, Xu, Q., Liu, W. de, 2018. A WD-GA-LSSVM model for rainfall-triggered landslide displacement prediction. J. Mt. Sci. 15, 156–166. https://doi.org/10.1007/s11629-016-4245-3