Engineering Geology 117 (2011) 52–61

Contents lists available at ScienceDirect

Engineering Geology

journal homepage: www.elsevier.com/locate/enggeo

Logistic regression model for predicting the failure probability of a landslide dam

Jia-Jyun Dong [a,b,*], Yu-Hsiang Tung [a], Chien-Chih Chen [b], Jyh-Jong Liao [c], Yii-Wen Pan [c]

[a] Graduate Institute of Applied Geology, National Central University, Jungli, 32001, Taiwan
[b] Graduate Institute of Geophysics, National Central University, Jhongli 32001, Taiwan
[c] Department of Civil Engineering and Hazard Mitigation Research Center, National Chiao Tung University, Hsinchu 30010, Taiwan

ARTICLE INFO

Article history:
Received 12 February 2010
Received in revised form 23 September 2010
Accepted 2 October 2010
Available online 13 October 2011

Keywords:
Landslide dam
Failure probability
Logistic regression
Geomorphic variables

ABSTRACT

Landslides may obstruct river flow and result in landslide dams; they occur in many regions of the world. The formation and disappearance of natural lakes involve a complex earth–surface process. According to the lessons learned from many historical cases, landslide dams usually break down rapidly soon after the formation of the lake. Regarding hazard mitigation, prompt evaluation of the stability of the landslide dam is crucial. Based on a Japanese dataset, this study utilized the logistic regression method and the jack-knife technique to identify the important geomorphic variables, including peak flow (or catchment area), dam height, width and length in sequence, affecting the stability of landslide dams. The resulting high overall prediction power demonstrates the robustness of the proposed logistic regression models. Accordingly, the failure probability of a landslide dam can also be evaluated based on this approach. Ten landslide dams (formed after the 1999 Chi-Chi Earthquake, the 2008 Wenchuan Earthquake and 2009 Typhoon Morakot) with complete dam geometry records were adopted as examples of evaluating the failure probability. The stable Tsao-Ling landslide dam, which was induced by the Chi-Chi earthquake, has a failure probability of 27.68% using a model incorporating the catchment area and dam geometry. On the contrary, the Tangjiashan landslide dam, which was artificially breached soon after its formation during the Wenchuan earthquake, has a failure probability as high as 99.54%. Typhoon Morakot induced the Siaolin landslide dam, which was breached within one hour after its formation and has a failure probability of 71.09%. Notably, the failure probability of the earthquake induced cases is reduced if the catchment area in the prediction model is replaced by the peak flow of the dammed stream for these cases. In contrast, the predicted failure probability of the heavy rainfall-induced case increases if the high flow rate of the dammed stream is incorporated into the prediction model. Consequently, it is suggested that the prediction model using the peak flow as causative factor should be used to evaluate the stability of a landslide dam if the peak flow is available. Together with an estimation of the impact of an outburst flood from a landslide-dammed lake, the failure probability of the landslide dam predicted by the proposed logistic regression model could be useful for evaluating the related risk.

1. Introduction

After the formation of a landslide dam, the natural lake may be quickly breached with an outburst flood and debris flow that results in a catastrophic disaster. Schuster and Costa (1986) reported that half of landslide dams fail within 10 days based on 63 cases from the literature. For example, the 25 August 1933 Deixi Earthquake resulted in the Deixi landslide dam in south-Central China, which was breached on 7 October and killed at least 2423 people (Li et al., 1986). It is always a great challenge for hazard mitigation because proper actions should be performed within a limited time.

Rapid assessment of the landslide-dam stability is one of the crucial steps for decision-making to reduce the related disasters. Geomorphic approaches are widely used to correlate the dam, river, impoundment characteristics and landslide-dam stability (Swanson et al., 1986; Costa and Schuster, 1988; Casagli and Ermini, 1999; Ermini and Casagli, 2003; Korup, 2004). Among them, Ermini and Casagli (2003) suggested the use of the geomorphic index DBI, which combines three important variables (dam height H, dam volume V and catchment area A) to evaluate the stability of a landslide dam, where $DBI = \log[(H \cdot A)/V]$. By incorporating these simplistic geomorphologic analyses, GIS-based modeling can be used to evaluate the potential for river blockages, upstream flooding and related hazards of outburst floods due to the probable sudden failure of the landslide dam (Clerici and Perego, 2000; Korup, 2005).

Dong et al. (2009a) compiled a Japanese dataset (Tabata et al., 2002) with complete records of the characteristics of landslide dams and a worldwide dataset (Ermini and Casagli, 2003) with dam height, dam volume and catchment area for documented landslide dams. Based on the Japanese dataset (Tabata inventory) consisting of 43 well

* Corresponding author. No. 300, Jungda Rd., Jungli, Taoyuan, 32001, Taiwan. Tel./fax: +886 3 4224114.
E-mail address: jjdong@geo.ncu.edu.tw (J.-J. Dong).

0013-7952/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.enggeo.2010.10.004

documented landslide dams, Dong et al. (2009a) utilized discriminant analysis to find out the dominant variables that affect the stability of a landslide dam and to construct a series of multivariate regression models. These variables include peak flow (or catchment area), dam height, width and length, in the order of relative importance. The resulting high prediction power (88.4% of the 43 training cases were correctly classified) and high cross-validation accuracy (86%) in their work demonstrate the robustness of the discriminant models *PHWL* (denoted as Model *PHWL*_Dis herein)

$$D_s = -2.94\log(P) - 4.58\log(H) + 4.17\log(W) + 2.39\log(L) - 2.52 \quad (1)$$

and *AHWL* (denoted as Model *AHWL*_Dis herein)

$$D_s = -2.62\log(A) - 4.67\log(H) + 4.57\log(W) + 2.67\log(L) + 8.26, \quad (2)$$

where D_s is the discriminant score; *P*, *H*, *W*, *L*, *A* are the peak flow, dam height, width, length and catchment area, respectively. To validate the index-based graphic approach, Dong et al. (2009a) further used the 84 worldwide dataset (training set) to build a model *AHV* with three variables (log-transformed catchment area log(A), dam height log(H) and dam volume log(V)). The discriminant model *AHV*_Dis is

$$D_s = -2.13\log(A) - 4.08\log(H) + 2.94\log(V) + 4.09. \quad (3)$$

The overall prediction power of the *AHV*_Dis was 88.1%, while its cross-validation accuracy was 83.3%. Using Eq. (3), the 37 Japanese landslide dams (target set) collected by Tabata et al. (2002) were classified into either stable or unstable. The overall prediction power was 70.1% for Model *AHV*_Dis. It appears that their discriminant model *AHV*_Dis has a better performance than the index-based graphic approach (which had an overall prediction power 64.9%).

The risk assessment and management may involve (1) estimating the risk level, (2) judging whether the risk level is acceptable and (3) exercising appropriate countermeasures to reduce the risk if the risk level is not acceptable (Dai et al., 2002). Korup (2005) considered that the risk level of the breach of a landslide dam is a function of the probability of an outburst flood from the landslide-dammed lake and the probability of spatial impact by the outburst flood. Hence, a prompt evaluation for the occurrence chance (probability) of dam breaching is essential. The logistic regression is a well established technique that can evaluate the occurrence probability of a catastrophic event (Begueria, 2006). The present work utilizes the Tabata inventory and Ermini–Casagli inventory (Supplementary Table S1 and S2 in Dong et al., 2009a) to construct the logistic regression models for predicting the failure probability of landslide dams. This work also compared the performance of the logistic regression models with those of the previous DBI index-based graphic model and discriminant models. The relative importance of relevant variables was also evaluated. Ten landslide dams formed after recent catastrophic earthquakes and heavy rainfall were adopted as examples for evaluating their failure probability. Finally, the implication and limitations of the proposed model on risk assessment of a landslide dam are discussed.

2. Methodology

Logistic regression is a widely used statistical approach (e.g., Carrara et al., 1991; Atkinson and Massari, 1998; Chung and Fabbri, 1999; Lee and Min, 2001; Dai and Lee, 2002, 2003; Ohlmacher and Davis, 2003; Can et al., 2005; Ayalew and Yamagishi, 2005; Chang et al., 2007a,b; Greco et al., 2007). As a first step, the landslide dams in the dataset can be categorized into two groups, (1) a stable group and (2) an unstable group, on the basis of a non-linear "logistic regression function" of a set of selected variables. The occurrence probability of landslide dam failure (i.e., the failure probability of the landslide dam) can then be calculated based on a logistic regression model. Next, the model performance can be evaluated through the confusion matrix and relative operating characteristic (ROC) diagram. Finally, the relative importance of individual variables can be sorted by the jackknife technique. Brief descriptions of these methods are as follows.

2.1. Categorization of dataset by logistic regression analysis

Logistic regression is useful when the dependent variable is categorical (e.g., presence or absence) and the explanatory (independent) variables are categorical, numerical, or both (Menard, 2002). An odds ratio P_s, representing the probability of a landslide dam remaining stable, is defined as

$$P_s = \frac{1}{1 + e^{-L_s}}, \quad (4)$$

where L_s is a certain linear combination of the influencing variables as follows:

$$L_s = b_0 + b_1x_1 + b_2x_2 + \ldots + b_nx_n = \ln\left(\frac{P_s}{1-P_s}\right). \quad (5)$$

In Eq. (5), x_i ($i = 1 \sim n$) is the independent variable, b_i ($i = 0 \sim n$) is the regression coefficient for the sample data and n is the number of independent variables. $L_s = \ln\left(\frac{P_s}{1-P_s}\right)$ is the logarithm of the odds ratio, called "logit." The condition $L_s = 0$ corresponds to the condition that the probability of a landslide-dam failure is 50%. If a dam with variables x_i has $L_s > 0$ (or $P_s > 50\%$), it is classified into the stable group. Otherwise, it is classified into the unstable group. A failure probability of the landslide dam is further defined as follows:

$$P_f = 1 - P_s = \frac{e^{-L_s}}{1 + e^{-L_s}}. \quad (6)$$

In this study, we adopted the identical variables *P*, *H*, *W*, *L*, *A* (which are also included in the two discriminant models *PHWL*_Dis (Eq. (1)) and *AHWL*_Dis (Eq. (2))) to construct the logistic regression models. The training dataset comprised 43 well-documented landslide dams in the Tabata inventory (Supplementary Table S1 in Dong et al., 2009a). We also used the 84 worldwide dataset (training set; Supplementary Table S2) to build a logistic regression model with three variables including log(A), log(H) and log(V), which are identical to the index-based graphic model (DBI index) and the discriminant model *AHV*_Dis (Eq. (3)). With Tabata's inventory (including 37 cases) as the target set (none in the training set), we compared the prediction performance of the landslide-dam stability by the index-based graphic model, discriminant model and logistic regression model. The detailed information of the inventories, statistics and correlation analysis of the geomorphologic variables, reliability and the process for selecting the significant variables to construct the statistical models can be found in Dong et al. (2009a).

2.2. Performance evaluation of the logistic regression model

The concept of a confusion matrix (Table 1) is often used to examine the performance of a prediction model. The proportion of correctly classified observations ($(a+d)/N$) is calculated to illustrate the prediction ability of the proposed statistical model where N is the number of landslide dam cases, a is the true positive and d is the true negative. Cross-validation can also be used to examine the reliability and robustness of the proposed model (Carrara, et al., 2008). We randomly split the dataset (43 cases) into (1) the training set (17 unstable and 5 stable) and (2) the target set (17 unstable and 4 stable).

Table 1
Confusion matrix. a: true positives; b: false positives (error type I); c: false negative (error type II); d: true negatives. N $(=a+b+c+d)$ is the total number of data sets.

Predicted	Observed	
	Stable dam	Unstable dam
Stable	a	b
Unstable	c	d

By using the target set, we then evaluated the predictive success of the model built on the training set. The proportion of correctly classified observations was calculated to illustrate the prediction ability of the proposed statistical models.

Alternatively, the prediction performance of a predictive model can be evaluated via the ROC (relative operating characteristic) diagram. The ROC diagram method has been widely used to measure the prediction potential of landslide susceptibility models (e.g., Chung and Fabbri, 2003; Chen et al., 2007; Carrara et al., 2008; Lee et al., 2008a,b; Dong et al., 2009b). In the ROC diagram, the false alarm rate (FAR) is on the horizontal axis while the hit rate (HR) is on the vertical axis. HR is the fraction of positive occurrences of dam failure that is correctly predicted, while FAR is the fraction of incorrectly predicted cases that did not occur (Swets, 1988). A larger area under the ROC Curve (AUC) indicates better model prediction; the index AUC ranges from 0.5 (for models with no predictive capability) to 1.0 (for models with perfect predictive power).

2.3. Jack-knife technique

The jack-knife technique (Swan and Sandilands, 1995) can be utilized to sort the importance of the relevant variables in the logistic regression model. We eliminate the variables in the logistic regression model one by one. M sub-models are created if the logistic regression model has M variables. The greater the prediction ability of the sub-model is (compared with the original model), the more important the eliminated variable is. Consequently, the relative importance of each variable contributing to the landslide-dam stability in the proposed models can be sorted.

The aforementioned logistic regression analyses were carried out with the commercial statistical package, SPSS (Statistical Package for the Social Sciences).

3. Results

3.1. Logistic regression models PHWL_Log and AHWL_Log

Logistic regression models (*PHWL*_Log and *AHWL*_Log) with four variables P (or A), H, W and L were built on the basis of 34 unstable dams and 9 stable dams as follows:

$$L_s = -2.55\log(P) - 3.64\log(H) + 2.99\log(W) + 2.73\log(L) - 3.87 \quad (7)$$

$$L_s = -2.22\log(A) - 3.76\log(H) + 3.17\log(W) + 2.85\log(L) + 5.93 \quad (8)$$

where P (or A), H, W and L are the peak flow (or catchment area), dam height, width and length, respectively. If a dam with variables x_i results in $L_s>0$ (or $P_s>0.5$), then it belongs to the stable group; otherwise, it belongs to the unstable group. Fig. 1 (a) and (b) shows the classification results for the logistic regression models *PTHWL*_Log and *ATHWL*_Log. For model *PTHWL*_Log, 6 out of the 9 stable landslide dams and 32 out of the 34 unstable landslide dams were correctively classified (Figure 1(a)). For model *ATHWL*_Log, 6 out of the 9 stable landslide dams and 33 out of the 34 unstable landslide dams were correctively classified (Figure 1 (b)). The overall prediction powers (percentage of landslide dams correctly classified) for models *PHWL*_Log (Eq. (7)) and *AHWL*_Log (Eq. (8)) were 88.4% and 90.7%, respectively. The cross-validation accuracy of models *PHWL*_Log and *AHWL*_Log was 85.7% and 77.3%, respectively. The confusion matrix of the logistic regression models are illustrated in Table 2. Fig. 2 shows the ROC curves of the proposed model. The AUC = 0.948 and AUC = 0.925 for models *PHWL*_Log and *AHWL*_Log, respectively, indicate that both of the proposed models were able to categorize the landslide dams into stable and unstable groups with high success rates.

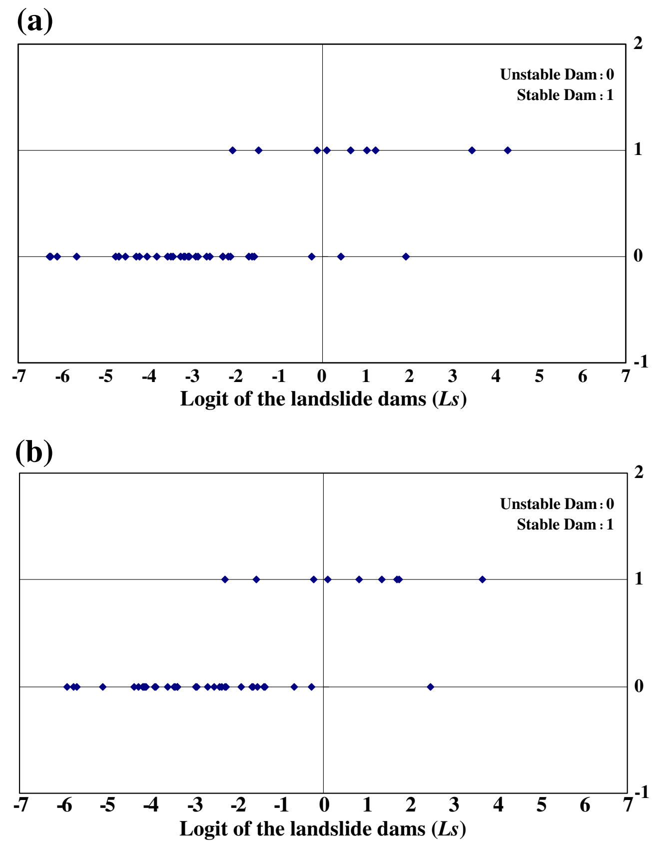

Fig. 1. Logit (L_s) distribution ($L_s>0$ stable, $L_s<0$ unstable) of (a) the Model *PHWL*_Log, and (b) the Model *AHWL*_Log for the stability of 43 landslide dams.

3.2. Logistic regression model AHV_Log

The 84 worldwide dataset (the training set; Supplementary Table S2) was used to build the logistic regression model with three variables

Table 2
Confusion matrix of the logistic regression models.

Model	Actual groups	Number of landslide dams	Predicted group membership	
			Group 1 (stable)	Group 2 (unstable)
*PHWL*_Log	Group 1 (stable)	9	6 (66.7%)	3 (33.3%)
	Group 2 (unstable)	34	2 (5.9%)	32 (94.1%)
	Percentage of landslide dams correctly classified: 88.4 (whole dataset; 43 cases) 85.7 (cross-validation)			
*AHWL*_Log	Group 1 (stable)	9	6 (66.7%)	3 (33.3%)
	Group 2 (unstable)	34	1 (2.9%)	33 (97.1%)
	Percentage of landslide dams correctly classified: 90.7 (whole dataset; 43 cases) 77.3 (cross-validation)			

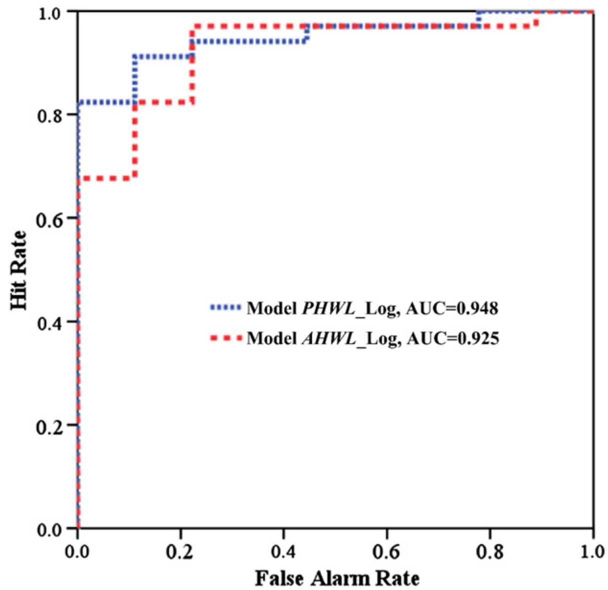

Fig. 2. ROC curves of the proposed logistic regression models.

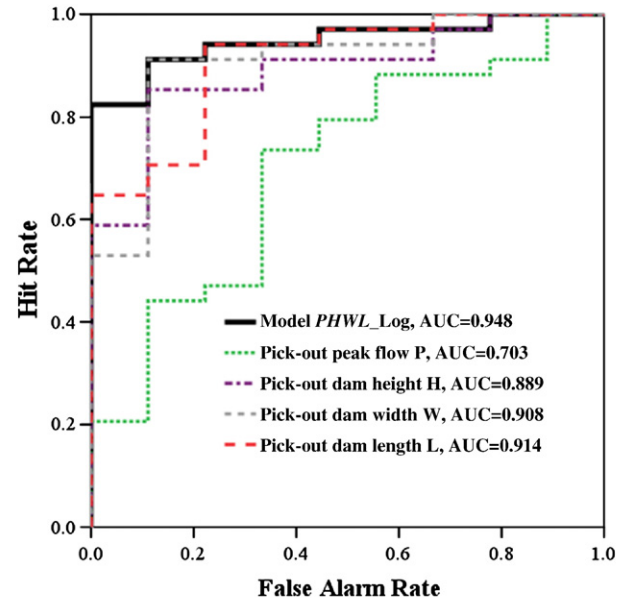

Fig. 3. ROC curves of Model *PHWL*_Log and 4 sub-models.

(*A*, *H* and *V*). Model *AHV*_Log obtained from the logistic regression is as follows.

$$L_s = -4.48\log(A) - 9.31\log(H) + 6.61\log(V) + 6.39. \quad (9)$$

The overall prediction power (success rate) of *AHV*_Log was 89.3%. The cross-validation accuracy was 85.7%. The AUC of the Model *AHV*_Log was 0.951. Using Eq. (9), the 37 Japanese landslide dams in Tabata's inventory (the target set; not in the training set) were classified as either stable or unstable landslide dams. The overall prediction power (prediction rate) was 76.9% for model *AHV*_Log. Comparing with the overall prediction powers of 70.1% and 64.9%, respectively, for the discriminant model *AHV*_Dis and for the index-based graphic model (DBI<2.75 stable; DBI>3.08 unstable), the logistic regression model seems to have a better ability to categorize landslide dams (as stable or unstable).

3.3. Importance of the factors affecting landslide-dam stability

The jack-knife technique was utilized to examine the relative importance of each variable in the predictive models. We eliminated one of the four variables one by one and established four logistic regression sub-models. The ROC curves were derived and the AUCs were calculated. Fig. 3 shows the ROC curves of model *PHWL*_Log and the 4 sub-models. The AUC for the case with the peak flow variable eliminated was 0.703; it was also the lowest one among all of the 4 sub-models. The result clearly indicates that the peak flow is the most significant variable in model *PHWL*_Log. The relative importance of the variables are peak flow (AUC = 0.703), dam height (AUC = 0.889), dam width (AUC = 0.908) and dam length (AUC = 0.914), in that order.

For model *AHWL*_Log, the relative importance of the variables are catchment area (AUC = 0.703), dam height (AUC = 0.859), dam length (AUC = 0.878) and dam width (AUC = 0.899), in that order (as shown in Figure 4).

3.4. Failure probability of landslide dams in Tobata inventory

The failure probability of a landslide dam, P_f, could be derived from the logit L_s using Eq. (6). Fig. 5 (a) and (b) show the failure probability of landslide dams in the Tobata inventory as calculated by the logistic regression models (*PHWL*_Log and *AHWL*_Log), respectively. Fig. 5(a) shows that 94.1% of the 34 stable landslide dams (P_f>50%) and 66.7% of the 9 unstable landslide dams (P_f<50%) were correctively classified by model *PHWL*_Log. For model *AHWL*_Log, 97.1% of the 34 unstable landslide dams and 66.7% of the 9 stable landslide dams were correctively classified (Figure 5 (b)). Notably, 31 out of 34 unstable landslide dams (about 90%) had a failure probability greater than 80% according to model *PHWL*_Log. For model *AHWL*_Log, about 85% of unstable landslide dams had a failure probability greater than 80%.

Fig. 6 shows the contour planes for P_f = 2%, 10%, 50%, 90% and 98% as predicted by model *AHV*_Log (the 84 worldwide dataset was used as the training set). The three-dimensional plot is for a view angle parallel to the strike of the contour planes. We calculated the number of stable and unstable cases of the 37 Japanese landslide dams in the Tabata inventory (the target set; not in the training set) within P_f = 0–2%, P_f = 2–10%, P_f = 10–50, P_f = 50–90%, P_f = 90–98% and

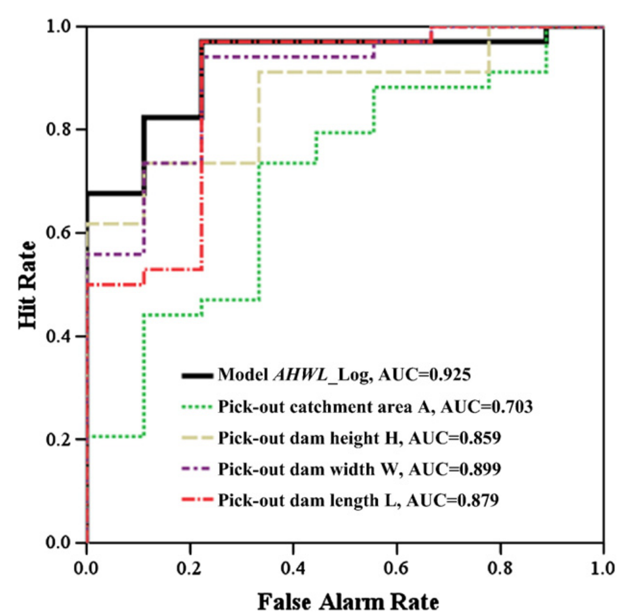

Fig. 4. ROC curves of Model *AHWL*_Log and 4 sub-models.

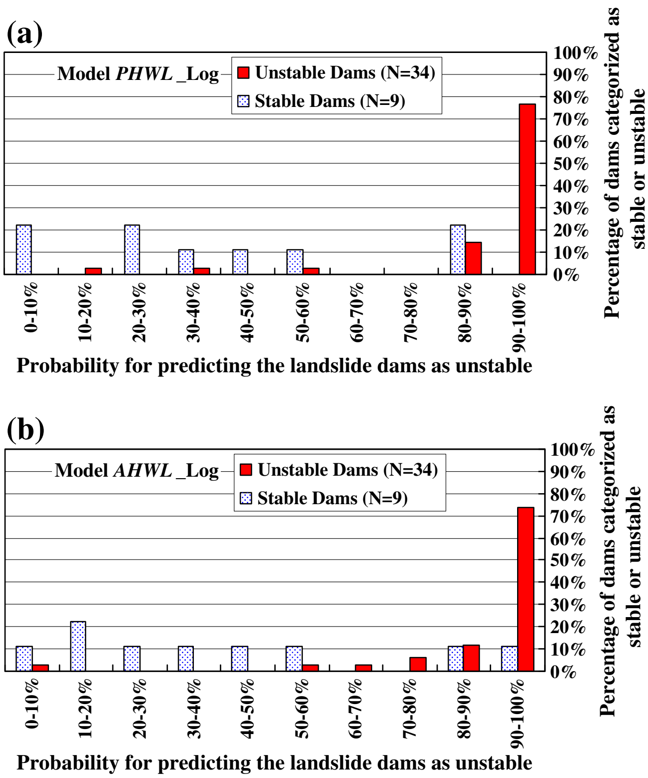

Fig. 5. Histogram showing the distribution of failure probability (P_f) of the 43 landslide dams. (a) Model *PHWL*_Log; (b) Model *AHWL*_Log.

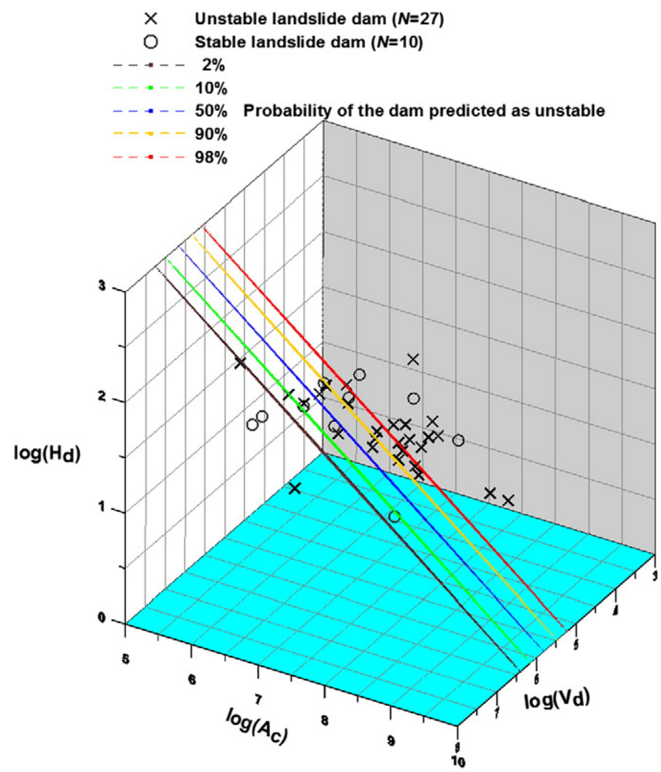

Fig. 6. Contour planes for 2%, 10%, 50%, 90%, and 98% of failure probability of a landslide dam predicted by the Model *AHV*_Log. The stable and unstable landslide dams in Tabata's inventory (37 target dataset) and the failure probability of Tangjiashan and Tsao-Ling landslide dam are also shown.

P_f = 98–100%. The results are shown in Fig. 7. Ten unstable dams and three stable dams fell within the range of failure probability P_f = 98–100%. Eight unstable dams and one stable dam fell within the range of failure probability P_f = 90–98%. It is qualitatively supported that the logistic regression model is capable of predicting the failure probability as well as categorizing a landslide dam into the stable or unstable group.

4. Application to the landslide dams induced by recent catastrophic events

Landslide dam inventories with more than 250 cases triggered by the 2008 Wenchuan earthquake were recently published (Cui et al., 2009; Xu et al., 2009; Yin et al., 2009). However, as is not uncommon, the geomorphic characteristics of the landslide dams were incompletely documented. Only 19 cases with dam height, length, width and dam volume were completely reported (Yin et al., 2009). None of the inventories include the data for the catchment area at the upstream of the landslide dams, which is one of the critical characteristics related to the dam stability (Dong et al., 2009a).

In this work, a digital elevation model derived from the Shuttle Radar Topographic Mission (SRTM) with a precision of 90 m was used to estimate the catchment area of the 19 aforementioned landslide dams. The locations of these landslide dams were provided by Cui et al. (2009). Among the 19 cases, only 8 landslide dams were reasonably located and thus the catchment areas estimated with confidence.

Together with the 8 Wenchuan cases, two Taiwanese landslide dam cases were also adopted to demonstrate the capability of the proposed logistic regression model for calculating the failure probability. The first one is the Tsao-Ling landslide dam formed during the 1999 Chi-Chi earthquake; it was eventually filled up with sediments and the dam remains stable (Li et al., 2002; Lee and Lin, 2006). The second one is the Siaolin landslide dam triggered by the heavy rainfall in 2009 due to Typhoon Morakot; it was rapidly breached after its formation (Dong et al., submitted). Table 3 lists the geomorphologic characteristics required for the prediction models. Data from different sources are sometimes inconsistent. In addition, some of the dam heights were provided as a range. In all the cases, the smallest values of the dam height were adopted to evaluate the dam stability with models *AHWL*_Log and *PHWL*_Log and the DBI index. The input values of the dam geometry are the bold data in Table 3.

The Tsao-Ling landslide dam was categorized as a stable dam. The failure probability of the Tsao-Ling landslide dam evaluated by model *AHWL*_Log was 27.68% (Table 3). On the other hand, the Siaolin and Tangjiashan landslide dams were classified as unstable. Evaluated by model *AHWL*_Log, the failure probability of the naturally breached Siaolin landslide dam was 81.91%, while the failure probability of the artificially breached Tangjiashan landslide dam was greater than 99%. Among the other seven landslide dams triggered by the Wenchuan earthquake, three dams were categorized in the stable group (failure

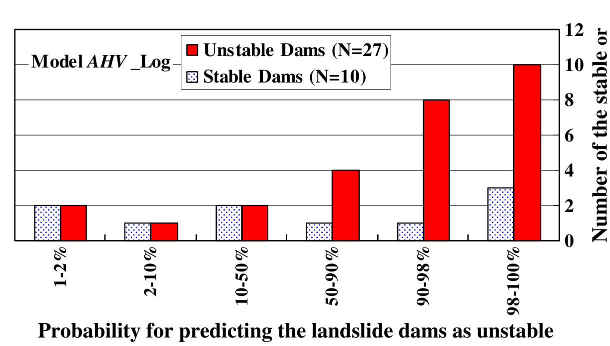

Fig. 7. A comparison of relative frequencies of stable and unstable landslide dams of the target data set (37 Japanese cases) and dam failure probabilities (P_f) calculated from the Model *AHV*_Log developed from the training data set (84 worldwide cases).

Table 3
Geomorphologic parameters and the failure probabilities predicted by logistic regression models of several landslide dams.

Triggering events	Landslide dam	Geomorphologic parameters						Predicted stability and failure probability of landslide dams		
		P (m^3/s)	A (10^{6}m^2)	H (m)	W (m)	L (m)	V (10^6 m^3)	*AHWL*_Log: L_s(P_f%)	*PHWL*_Log: L_s(P_f%)	DBI (stability)
1999 Chi-Chi Earthqrake	Tsao-Ling		**162.0**[1]	**50**–110[1]	**5000**[1]	**600**[1]	**120**[1]	0.96(27.68)		1.83(S)
2008 Wenchuan Earthquake	Tangjiashan	**88.9**[a]	**3937.5**	**80**–120[2] 82–124[3] 82–124[4]	**800**[2] 800[3] 803[4]	**612**[2] 600[3] 611[4]	**20.4**[2] 20.4[3]	−5.38(99.54)	−0.59(64.34)	4.19(U)
	Kuzhuba-downstream		**3952.4**	**60**[2] 60[4]	**200**[2]	**300**[2]	**0.17**[2]	−7.71(99.96)		6.14(U)
	Shibangou		**19.6**	60[2] **30**–75[3]	**450**[2]	**800**[2]	**8.1**[2] 15.0[3]	0.87(29.53)		1.86(S)
	Donhekou	<**100**[2]	**24.9**	**15**–25[2] 20[3]	**700**[2]	**500**[2]	**10.0**[2] 12.0[3]	1.80(14.19)	2.62(6.79)	1.57(S)
	Hongshihe		**27.0**	40[2] **30**–50[3]	**500**[2]	**400**[2]	**1.0**[2] 4.0[3]	−0.15(53.74)		2.91(U/S)
	Guantan		**177.8**	**60**[2] 60[3]	**120**[2]	**200**[2]	**1.44**[2] 1.2[3]	−5.92(99.73)		3.87(U)
	Xiaogonjian-upstream		**369.1**	72[2] **62**–72[3] 62[4]	**172**[2]	**120**[2]	**1.6**[3]	−6.82(99.89)		4.16(U)
	Yingxiuwan-Taipingyi HPS		**1907.1**	**18**[2]	**300**[2]	**200**[2]	**1.0**[2]	−4.98(99.32)		4.54(U)
2009 Typhoon Morakot	Siaolin	**2974**[5]	**354.0**[5]	**44**[5]	**1554**[5]	**760**[5]	**13.18**[5]	−0.90(71.09)	−1.30(78.58)	3.07(U/S)

[1]Li et al. (2002); [2]Yin et al. (2009); [3]Xu et al. (2009); [4]Cui et al. (2009); [5]Dong et al. (submitted).
[a]Estimated based on a nearby Xiaojiahe landslide dam (catchment area = 159.8 km^2, flow rate = 4 m^3/s in May) (Yin et al., 2009).
U: unstable; S: stable; U/S: unclassified.

probability 12%–46%) and four dams were categorized in the unstable group (failure probability > 99%).

5. Discussion

5.1. Comparison of the logistic regression models and discriminant models

Table 4 compares the overall prediction power of the logistic regression models and the discriminant models. It appears the proposed logistic regression models *PHWL*_Log (AUC = 94.8%) and *AHWL*_Log (AUC = 92.5%) were able to categorize the landslide dams into the stable and unstable groups with high success rates. The AUCs of the logistic regression models were slightly higher than the AUCs of the discriminant models (see Table 4). The models containing the peak flow variable (*PHWL*_Dis and *PHWL*_Log) were superior to the models with catchment area instead (*AHWL*_Dis and *AHWL*_Log). It is speculated that the peak flow should have a direct impact on the stability and erosion of a landslide dam. On the other hand, the catchment area can only implicitly account for the potential peak flow that may flow downstream and threaten the landslide-dam stability. The difference in the predicted failure probability of landslide dams using different models will be further discussed in Section 5.2.

Notably, the proportion of correctly classified observations ($(a+d)/N$) of the proposed logistic regression models derived from cross-validation is lower than that of the discriminant models. In addition, the overall prediction power of the logistic models derived from cross-validation is not as good as that of the discriminant models. However, the error of type II "c" (false negative; model failed to predict the landslide dam instability) is lower than that of the discriminant models (Table 4). Begueria (2006) indicated that the senses of false negatives "c" and false positives "b" (error of types I) with respect to risk assessment could be significantly different. The logistic regression analysis has a lower error of type II "c" (5.9% and 2.9%), which indicates a high chance that the model will identify an unstable landslide dam (94.1% and 97.1%, in Table 4). It may also imply that the logistic regression models are conservative in predicting the stability of landslide dams compared with the discriminant models proposed by Dong et al. (2009a). For catastrophic hazards induced by the collapse of landslide dams, the necessity for correctly classifying an unstable landslide dam may be more crucial than correctly classifying a stable landslide dam.

Regarding the relative importance of the variables affecting the stability of landslide dams, the jack-knife technique identified the peak flow and the catchment area to be the most important ones in the *PHWL*_Log and *AHWL*_Log models, respectively. The dam height was the second most important variable contributing to the landslide-dam stability in the proposed logistic regression models. The peak flow, catchment area and dam height are all negative factors contributing to the stability of landslide dams. On the other hand, the dam width and dam length are positive factors. In model *PHWL*_Log, the dam width is more important than the dam length. In model *PHWL*_Log, the importance of the dam length is greater than that of the dam width. The significance of the variables contributing to the stability of landslide dams for various models is listed in Table 4. The relative importance of variables of landslide-dam stability for the logistic regression models are almost identical to the ones derived

Table 4
Comparison between the performance of the logistic regression models and the discriminant models (Dong et al., 2009a).

Model	Predicted	Observed		Proportion of correctly classified observations $(a+d)/N$		AUC
		Stable	Unstable	Whole data set	Cross-validation	
*PHWL*_Log	Stable	66.7%	33.3%	88.4%	85.7%	0.948
	Unstable	5.9%	94.1%			
	Significance of the variables contributed to the stability of landslide dam: $P>H>W>L$					
*AHWL*_Log	Stable	66.7%	33.3%	90.7%	77.3%	0.925
	Unstable	2.9%	97.1%			
	Significance of the variables contributed to the stability of landslide dam: $P>H>L>W$					
*PHWL*_Dis	Stable	77.8%	22.2%	88.4%	86.0%	0.935
	Unstable	8.8%	91.2%			
	Significance of the variables contributed to the stability of landslide dam: $P>H>W>L$					
*AHWL*_Dis	Stable	77.8%	22.2%	88.4%	86.0%	0.905
	Unstable	8.8%	91.2%			
	Significance of the variables contributed to the stability of landslide dam: $P>H>W>L$					

Table 5
Empirical approach of risk assessment with respect to outburst flooding of a landslide dam (Cui et al., 2009; Xu et al., 2009; Yin et al., 2009).

Index	Grade			
	Extreme high risk (EHR)	High risk (HR)	Medium risk (MR)	Low risk (LR)
Dam height (m)	>100	50–100	25–50	<25
Water storage capacity ($\times 10^6$ m^3)	100	10–100	1–10	<1
Composition of dam materials	[a]Group 1	Group 2	Group 3	Group 4

[a] Group 1: soil and fragments of rock; Group 2: soil and fragments of rock with a few boulders and blocks; Group 3: boulders and blocks with little soil and rock fragments; Group 4: boulders and blocks.

using the standardized canonical discriminant coefficient (SCDC) of variables in the discriminant models (Dong et al., 2009a); the only exception is the reverse order of the dam length and dam width for model *AHWL*_Log.

5.2. Predicted failure probability using different hydrology factors

As demonstrated in Section 5.1, the models containing the peak flow variable are superior to the models with catchment area instead. The peak flow of a stream is dominated by many factors, such as morphological, hydro-geological and meteorological characteristics. Obviously, the use of the catchment area in place of peak flow is not able to fully reflect those influences. In particular, the peak flow of the dammed stream during the life span of an earthquake-induced landslide dam will be quite different than a heavy rainfall-induced one.

The estimated peak flows during the life spans of the Tangjiashan, Donhekou and Siaolin landslide dams are shown in Table 3. The estimated peak flow was low (<100 m^3/s) for the Wenchuan earthquake-induced landslide dams, whereas the peak flow of the heavy rainfall-induced Siaolin cases was relatively large (2974 m^3/s). The failure probability predicted by model *PHWL*_Log was lower than that predicted by model *AHWL*_Log for the earthquake-induced cases. For example, the failure probability of the earthquake-induced Tangjiashan landslide dam decreases from 99.54% to 64.34% if model *AHWL*_Log is replaced by model *PHWL*_Log. Contrarily, the predicted failure probability of the heavy rainfall-induced Siaolin landslide dam increases from 71.094% to 78.58% if the high flow rate replaces the catchment area. Hence, it is suggested that model *PHWL*_Log should be used to evaluate the stability of a landslide dam as long as the variable *P* (peak flow) is available. Nevertheless, it is often imperative to evaluate the dam stability as soon as possible. Consequently, model *AHWL*_Log is still useful in practice because the peak flow is often unclear right after the formation of a landslide dam.

5.3. Risk assessment of a landslide dam

One of the key issues for decision-making for hazard mitigation is to estimate the risk related to the landslide dams promptly. Korup (2005) critically reviewed the problems relating to the complexity for risk (including hazard and vulnerability) assessment of a landslide dam. It was suggested that the hazard induced by a dam-break flood is a function of (1) the probability of the landslide event, (2) the probability of dam and lake formation given the landslide event, (3) the probability of an outburst flood from an existing landslide-dammed lake (time-variant) and (4) the probability of spatial (downstream) impact due to an outburst flood from a landslide-dammed lake. Yet, an applicable empirical approach to categorize the degree of risk by using simple geomorphic parameters is preferred, at least in the initial period of risk evaluation. For example, the risk level for the landslide dams induced by the Wenchuan earthquake were classified into four grades, (1) extremely high risk (EHR), (2) high risk (HR), (3) medium risk (MR) and (4) low risk (LR), by taking into account the landslide dam height (*H*), lake capacity (V_l) and dam materials (Cui et al., 2009; Xu et al., 2009; Yin et al., 2009). Table 5 shows the matrix for a quick qualitative risk assessment of the landslide dams induced by the Wenchuan earthquake. The proposed empirical approach is rather simple and time saving and fits the requirement for emergency decisions making. However, the influence of the landslide dam failure and the impact (consequence) of an outburst flood from a landslide-dammed lake are not evaluated separately.

Table 6 shows the risk levels of ten landslide dams evaluated by the empirical approach (Table 5). The evaluated "high risk" of the Tsao-Ling landslide dam is mainly due to the dam height and large water storage capacity. Because the failure probability is very low

Table 6
Classified risk level of the selected landslide dams.

Triggering events	Landslide dam	H (m)	V_l (10^6 m^3)	Dam material	Risk level; based on the empirical approach (Table 5)	Status
1999 Chi-Chi Earthqrake	Tsao-Ling	50	40[1]	Group 4	HR	Filled up with sediments and the dam remains stable
2008 Wenchuan Earthquake	Tangjiashan	80	250[2] 300[3] 302[4]	Group 3	EHR	Possible whole collapse can be caused by intensive rainfall
	Kuzhuba-downstream	60	2[3] 2[4]	Group 4	MR	The whole stability is high, but can be affected by the possible failure of Tangjiashan landslide dam
	Shibangou	30	20[2] 11[3]	Group 3	HR	Low stability for permeability, high susceptibility for dam collapse
	Donhekou	15	10[2] 10[3]	Group 1	MR	Dam collapse and overflowing. Collapse can be caused by upstream landslide dam failure
	Hongshihe	30	3[2] 1.2[3]	Group 3	MR	Low stability for permeability, high susceptibility for dam collapsing partially or in whole
	Guantan	60	5[2] 10[3]	Group 1	HR	Low stability for permeability, high susceptibility for dam collapse
	Xiaogonjian-upstream	62	11[2] 11[3] 11[4]	Group 3	HR	Low stability for permeability, high susceptibility for dam collapse
	Yingxiuwan-Taipingyi HPS	18	2[2]	–	MR	Dammed half of the river. Overflow at low point
2009 Typhoon Morakot	Siaolin	44	9.9[5]	Group 1	MR	Breached and failed soon after its forming

[1]Lee and Lin (2006); [2]Yin et al. (2009); [3]Xu et al. (2009); [4]Cui et al. (2009); [5]Dong et al. (submitted).
EHR: extremely high risk; HR: high risk; MR: medium risk; LR: low risk.

according to the proposed logistic regression models, a risk level of "medium risk" seems more suitable for the Tsao-Ling case. On the contrary, the catastrophic Siaolin case is classified as "medium risk" by this empirical approach. It would be more appropriate to classify the Siaolin case as "high risk" to "extremely high risk" because of its high failure probability. Together with an estimation of the impact of the outburst flood from a landslide-dammed lake, the failure probability of the landslide dam predicted by the proposed logistic regression model can be useful for evaluating the related risk.

5.4. Limitations of the models

The proposed models for evaluating the stability of landslide dams do have their limitations. First of all, with a set of databases (such as landslides), it is possible to develop rigorous susceptibility models and to generate a susceptibility map. However, these actions do not guarantee that the predictions will coincide with reality in the field for general applications (Dai and Lee, 2003). The proposed logistic regression models, as well as the index––based graphic approach or the discriminant models, should always be used with caution for their applications in regions beyond where the inventories cover.

Secondly, using logistic regression analysis, the failure probability assigned to any landslide dam is the probability that the dam pertains to one of two groups, namely (i) stable dams or (ii) unstable dams, given the set of variables used in the model (Guzzetti et al., 2005). However, P_s in Eq. (4) is not a probability in a strict sense because the time-scale over which an event (such as failure of a landslide dam) is expected to occur is not explicitly given (Atkinson and Massari, 1998). The failure probability of an existing landslide-dammed lake can change with increasing lake volume, repeated strong seismic ground motion, or rapid changes in water level (Korup, 2005). In addition, outburst floods from upstream natural dams may cause overtopping and breaching of otherwise stable landslide dams downstream, which is difficult to quantify in terms of conditional probability (Korup, 2005). Further study to expand the evaluation capability of landslide-dam failure probability to a spatial-temporal context is needed.

Thirdly, in the Tabata's inventory, the dataset of stable landslide dam is a smaller dataset (20.1% of the total cases) compared with the unstable ones (79.9% of the total cases). It should be noted that the common statistical multivariate regression procedures, such as the logistic regression, ought to work with groups that are more or less equal in size (Begueria and Lorente, 1999). In the domain of political sciences, King and Zeng (2001) reported that ordinary logistic regression may sharply underestimate probabilities if the number of presences in the population is tens to thousands of times smaller than the number of absences. It is possible that the proposed models could over-predict the failure probability of landslide dams. Using rare event logistic regression, which has proven successful in landslide susceptibility analysis (Van Den Eeckhaut et al., 2006), may be one of the possible solutions. The influence of sampling on the evaluation of the stability of landslide dams using the logistic regression remains to be addressed in the future.

Finally, the accuracy of the geomorphic variables to be input into the model is always problematic. For example, it is difficult to determine the correct dam height H in the proposed models for the cases shown in Table 3. Fig. 8 shows the sensitivity of the failure probability due to the dam height. The horizontal axis is a normalized dam height H_N, which is defined as

$$H_N = \frac{H - H_{\min}}{H_{\max} - H_{\min}} \tag{10}$$

where the $H_{\max}$ and $H_{\min}$ are the reported maximum and minimum dam heights, respectively. For the Tangjiashan landslide dam, the reported height is within 82–124 m. The failure probability (>99%) will not change too much if the dam height increases from 82 m to 124 m. In the case of the Tsao-Ling landslide dam, the dam height is 50 m on the upstream side and 110 m on the downstream side. According to the logistic regression model *AHWL*_Log, the failure probability of Tsao-Ling landslide dam will significantly increase from 27.68% to 58.11% if the dam height increases from 50 m to 110 m. Because the geometry of landslide dams is often complex, the geomorphic approach should be used with caution. For model development and prediction, the definition of dam geometry deserves some deliberation. For the Siaolin landslide dam, the adopted dam height of 44 m was the deposition depth at the breaching point (Dong et al., submitted). It would be proper to define the dam height as the maximum depth of lake, which is better related to the dam height at its breaching point.

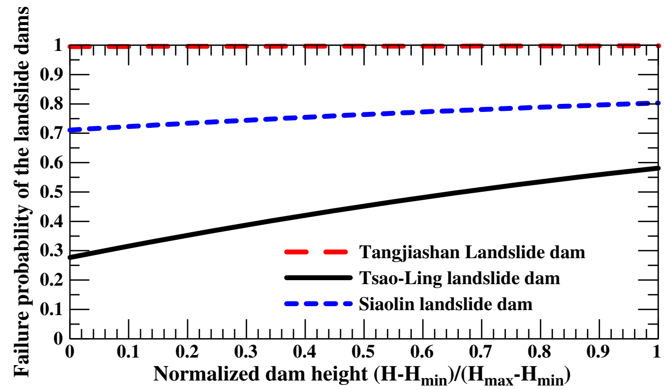

Fig. 8. The sensitivity analysis for the dam height on the failure probability of Tsao-Ling and Tangjiashan landslide dam using Model *AHV*_Log.

The aforementioned limitations all lead to an identical key issue: the incompleteness and uncertainty of the landslide dam dataset leads to uncertainty in the classification of landslide dams (into stable and unstable groups) using geomorphic approaches (Korup, 2004). Despite the limitations, the proposed models are still useful for quantitative evaluation of the failure probability of landslide dams in a logical approach. These models may be valuable for decision-makers to choose a proper countermeasure based on risk level. It is expected that worldwide records of landslide dams and complete geomorphic parameters will be accumulated quickly after new techniques, such as high resolution airborne LiDAR, become widely used. Additionally, more data availability and GIS-based geospatial extrapolation capability should expand the scope for future research on the formulation of regional susceptibility models for landslide-driven stream blockages based on catchment parameters.

6. Conclusions

Based on 43 Japanese cases as the training dataset, logistic regression models for the quantitative prediction of landslide-dam stability were presented. The proposed models *PHWL*_Log and *AHWL*_Log were able to categorize the landslide dams into stable and unstable groups with high success rates. Model *PHWL*_Log (AUC = 94.8%) was slightly superior to model *AHWL*_Log (AUC = 92.5%). Yet, model *AHWL*_Log may be more useful in practice because peak flow information is not always available in the early stage after dam formation.

The log-transformed peak flows (or alternatively, the catchment area) are identified as the most important geomorphic variables influencing the stability of a landslide dam. The log-transformed dam height, with a negative contribution to the stability of a landslide dam, is the second most significant variable. The log-transformed dam width and length have a similar positive effect on a dam's stability;

their relative importance found in this study is in agreement with the results derived from the discriminant analysis.

Compared with the discriminant models, the logistic regression models have a slightly better ability to categorize the landslide dams (into stable and unstable groups). Furthermore, the lower false negative (error type II; model fails to predict the landslide dam instability) predicted by the logistic regression models reveals the conservative nature of the model in categorizing a landslide dam into the unstable group.

In addition to the classification of landslide dams into the stable and unstable groups, the failure probability can also be evaluated based on the proposed logistic regression models. The present work took ten case histories of landslide dams (two in Taiwan and eight in China) to demonstrate the capability for the evaluation of failure probability by the logistic regression models. The stable Tsao-Ling landslide dam, formed after the 1999 Chi-Chi earthquake, had a failure probability 27.68% as predicted by model *AHWL*_Log. On the other hand, the artificially breached Tangjiashan landslide dam, formed after the 2008 Wenchuan earthquake, had a failure probability as high as 99.53%. Typhoon Morakot induced the Siaolin landslide dam, which was breached within one hour after its formation had a failure probability of 71.09%. In the cases presented, the failure probability of an earthquake-induced landslide dam decreased if the prediction model *AHWL*_Log was replaced by model *PHWL*_Log. In contrast, the failure probability of a heavy rainfall-induced case predicted by model *PHWL*_Log is higher than that by *AHWL*_Log because a high flow rate is incorporated into the prediction model (*PHWL*_Log).

Finally, the empirical approach for classifying the risk level of landslide dams based on the Wenchuan experience was tested on some cases in Taiwan. According to this empirical approach, the stable Tsao-Ling landslide dam was categorized as high risk level, whereas the catastrophic Siaolin case was classified as medium risk level. It would be more suitable if both the failure probability and the impact due to the outburst flood from the landslide-dammed lake could be considered separately for classifying the risk level of a landslide dam. Therefore, a simple model describing the failure probability of landslide dams, such as the proposed logistic regression model, is necessary for classifying the risk level related to a landslide dam breach.

In summary, the proposed models can be used for evaluating the risk associated with outburst floods from landslide-dammed lakes. These models can be used as an evaluation tool for decision-making concerning hazard mitigation actions, especially when the allowable time is limited.

Acknowledgements

The presented work was supported by the National Science Council of the Republic of China (Taiwan) under Contracts No. NSC 99-2625-M-009-004-MY3. This support is gratefully acknowledged. The writers also appreciate the constructive comments from the Editor and two anonymous reviewers.

References

Atkinson, P.M., Massari, R., 1998. Generalised linear modelling of susceptibility of landsliding in the Central Apennines, Italy. Computers & Geosciences 24, 373–385.
Ayalew, L., Yamagishi, H., 2005. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. Geomorphology 65, 15–31.
Begueria, S., 2006. Validation and evaluation of predictive models in hazard assessment and risk management. Natural Hazards 37, 315–329.
Begueria, S., Lorente, A., 1999. Landslide hazard mapping by multivariate statistics; comparison of methods and case study in the Spanish Pyrenees. The Damocles Project Work, Contract No EVG1-CT 1999-00007. Technical Report.
Can, T., Nefeslioglu, H.A., Gokceoglu, C., Sonmez, H., Duman, T.Y., 2005. Susceptibility assessments of shallow earthflows triggered by heavy rainfall at three catchments by logistic regression analyses. Geomorphology 72, 250–271.
Carrara, A., Cardinali, M., Detti, R., Guzzetti, F., Pasqui, V., Reichenbach, P., 1991. GIS techniques and statistical models in evaluating landslide hazard. Earth Surface Processes and Landforms 16, 427–445.
Carrara, A., Crosta, G., Frattini, P., 2008. Comparing models of debris-flow susceptibility in the alpine environment. Geomorphology 94, 353–378.
Casagli, N., Ermini, L., 1999. Geomorphic analysis of landslide dams in the Northern Apennine. Transactions of the Japanese Geomorphological Union 20, 219–249.
Chang, K.T., Chiang, S.H., Hsu, M.L., 2007a. Modeling typhoon- and earthquake-induced landslides in a mountainous watershed using logistic regression. Geomorphology 89, 335–347.
Chang, K.T., Chiang, S.H., Lei, F., 2007b. Analysing the relationship between typhoon-triggered landslides and critical rainfall conditions. Earth Surface Processes and Landforms 33, 1261–1271.
Chen, C.C., Tseng, C.Y., Dong, J.J., 2007. New entropy based method for variables selection and its application to the debris-flow hazard assessment. Engineering Geology 94, 19–26.
Chung, C.F., Fabbri, A.G., 1999. Probabilistic prediction models for landslide hazard mapping. Photogrammetric Engineering and Remote Sensing 65, 1389–1399.
Chung, C.F., Fabbri, A.G., 2003. Validation of spatial prediction models for landslide hazard mapping. Natural Hazards 30, 451–472.
Clerici, A., Perego, S., 2000. Simulation of the Parma River blockage by the Corniglio landslide Northern Italy. Geomorphology 33, 1–23.
Costa, J.E., Schuster, R.L., 1988. The formation and failure of natural dam. Geological Society of America Bulletin 100, 1054–1068.
Cui, P., Zhu, Y., Han, Y., Chen, X., Zhuang, J., 2009. The 12 May Wenchuan earthquake-induced landslide lakes: distribution and preliminary risk evaluation. Landslides 6, 209–223.
Dai, F.C., Lee, C.F., 2002. Landslide characteristics and slope instability modelling using GIS, Lantau Island, Hong Kong. Geomorphology 42, 213–228.
Dai, F.C., Lee, C.F., 2003. A spatiotemporal probabilistic modeling of storm-induced shallow landsliding using aerial photographs and logistic regression. Earth Surface Processes and Landforms 28, 527–545.
Dai, F.C., Lee, C.F., Ngai, Y.Y., 2002. Landslide risk assessment and management: an overview. Engineering Geology 64, 65–87.
Dong, J.J., Tung, Y.H., Chen, C.C., Liao, J.J., Pan, Y.W., 2009a. Discriminant analysis of the geomorphic characteristics and stability of landslide dams. Geomorphology 110, 162–171.
Dong, J.J., Lee, C.T., Tung, Y.H., Liu, C.N., Lin, K.P., Lee, J.F., 2009b. The role of the sediment budget in understanding debris flow susceptibility. Earth Surface Processes and Landforms 34, 1612–1624.
Dong, J.J., Li, Y.S., Kuo, C.Y., Sung, R.T., Li, M.H., Lee, C.T., Chen, C.C., 2010. The formation and breach of a short-lived landslide dam at Siaolin village, Taiwan - Part I: Post-event reconstruction of dam geometry. Engineering Geology submitted.
Ermini, L., Casagli, N., 2003. Prediction of the behaviour of landslide dams using a geomorphological dimensionless index. Earth Surface Processes and Landforms 28, 31–47.
Greco, R., Sorriso-Valvo, M., Catalano, E., 2007. Logistic regression analysis in the evaluation of mass movements susceptibility: the Aspromonte case study, Calabria, Italy. Engineering Geology 89, 47–66.
Guzzetti, F., Reichenbach, P., Cardinali, M., Galli, M., Ardizzone, F., 2005. Probabilistic landslide hazard assessment at the basin scale. Geomorphology 72, 272–299.
King, G., Zeng, L., 2001. Logistic regression in rare events data. Political Analysis 9, 137–163.
Korup, O., 2004. Geomorphometric characteristics of New Zealand landslide dams. Engineering Geology 73, 13–35.
Korup, O., 2005. Geomorphic hazard assessment of landslide dams in South Westland, New Zealand: fundamental problems and approaches. Geomorphology 66, 167–188.
Lee, K.T., Lin, Y.T., 2006. Flow analysis of landslide dammed lake water sheds a case study. Journal of the American Water Resources Association 42, 1615–1628.
Lee, S., Min, K., 2001. Statistical analysis of landslide susceptibility at Yongin, Korea. Environmental Geology 40, 1095–1113.
Lee, C.T., Huang, C.C., Lee, J.F., Pan, K.L., Lin, M.L., Dong, J.J., 2008a. Statistical approach to earthquake-induced landslide susceptibility. Engineering Geology 100, 43–58.
Lee, C.T., Huang, C.C., Lee, J.F., Pan, K.L., Lin, M.L., Dong, J.J., 2008b. Statistical approach to storm event-induced landslides susceptibility. Natural Hazards and Earth System Sciences 8, 941–960.
Li, T., Schuster, R.L., Wu, J., 1986. Landslide dams in south-central China. In: Schuster, R.L. (Ed.), Landslide Dam: Processes Risk and Mitigation: American Society of Civil Engineers. Geotechnical Special Publication, 3, pp. 146–162.
Li, M.H., Hsu, M.H., Hsieh, L.S., Teng, W.H., 2002. Inundation potentials analysis for Tsao-Ling landslide lake formed by Chi-Chi earthquake in Taiwan. Natural Hazards 25, 289–303.
Menard, S., 2002. Applied Logistic Regression Analysis, 2nd (Ed.). Sage, Thousand Oaks, CA.
Ohlmacher, G.C., Davis, J.C., 2003. Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA. Engineering Geology 69, 331–343.
Schuster, R.L., Costa, J.E., 1986. A perspective on landslide dams. In: Schuster, R.L. (Ed.), Landslide Dam: Processes Risk and Mitigation: American Society of Civil Engineers. Geotechnical Special Publication, 3, pp. 1–20.
Swan, A.R.H., Sandilands, M., 1995. Introduction to Geological Data Analysis. Blackwell Scientific.
Swanson, F.J., Oyagi, N., Tominaga, M., 1986. Landslide dam in Japan. In: Schuster, R.L. (Ed.), Landslide Dam: Processes Risk and Mitigation: American Society of Civil Engineers. Geotechnical Special Publication, 3, pp. 131–145.

Swets, J.A., 1988. Measuring the accuracy of diagnostic systems. Science 204 (4857), 1285–1293.

Tabata, S., Mizuyama, T., Inoue, K., 2002. Natural Landslide Dams Hazards, Kokonshoin, Tokyo. (in Japanese).

Van Den Eeckhaut, M., Vanwalleghem, T., Poesen, J., Govers, G., Verstraeten, G., Vandekerckhove, L., 2006. Prediction of landslide susceptibility using rare events logistic regression: a case-study in the Flemish Ardennes (Belgium). Geomorphology 76, 392–410.

Xu, Q., Fan, X.M., Huang, R.Q., Van Westen, C., 2009. Landslide dams triggered by the Wenchuan Earthquake, Sichuan Province, south west China. Bulletin of Engineering Geology and the Environment 68, 373–386.

Yin, Y., Wang, F., Sun, P., 2009. Landslide hazards triggered by the 2008 Wenchuan earthquake, Sichuan, China. Landslides 6, 139–152.