

Stochastic procedure to extract and to integrate landslide susceptibility maps: an example of mountainous watershed in Taiwan

Jr-Chuan Huang¹, Shuh-Ji Kao^{1,*}, Mei-Ling Hsu², and Jiun-Chuan Lin²

¹Research Center for Environmental Changes, Academia Sinica, Taipei, Taiwan

²Department of Geography, National Taiwan University, Taipei, Taiwan

Received: 18 August 2006 – Revised: 15 September 2006 – Accepted: 15 September 2006 – Published: 25 September 2006

Abstract. The Generalized Likelihood Uncertainty Estimation (GLUE) is here incorporated into a deterministic landslide model (SHALSTAB) to generate 4000 landslide susceptibility maps which enclose various combinations of full range parameters. Furthermore, an improved index is adopted into GLUE as a criterion to measure model performance, and through that, 200 maps holding top 5% performance are retrieved. Proper ranges for parameters are obtained through GLUE yet they only perform well if combined appropriately. The 200 better maps are overlapped to construct an integrated landslide susceptibility map. Instead of giving a single parameter set or a single susceptibility map, the merit of extracting and integrating procedure is to envelope uncertainties inherited in model structure and input parameters. Bias due to subjective parameter input is potentially reduced. The entire procedure is applied to the Chi-Jia-Wan, a mountainous watershed in Taiwan. The integrated map shows high-risk area (>50% predicted landslide probability) only occupies 16.4% of the entire watershed while able to correctly identify 60% of the actual landslides. For areas above 2100 m height the map is even more successful (projects 77 of the 98 actual landslides). Interactions among parameters are discussed to highlight the unsolvable equifinality problem and impropriety of presenting a single model result.

1 Introduction

Landslides triggered by torrential rain, which caused tremendous loss of human lives and property, have been a major concern in forest management in many parts of the world (Sidle et al., 1985). To mitigate the loss and to improve watershed management, deterministic landslide models are widely

applied to construct landslide susceptibility maps (Montgomery and Dietrich, 1994; Wu and Sidle, 1995; Burton and Bathurst, 1997; Borga et al., 1998; Pack et al., 1998). The landslide susceptibility in a model is determined by the factor of safety (FS), which represents the relative strength of downslope momentum and resistance of the slope material. Theoretically, all parameters should be measured for model applications; yet, direct field measurement is time and cost-consuming. Though determining parameters are intensively measured, field data is never sufficient for regional applications due to their natural spatial-temporal variability (Wu and Sidle, 1995; Burton and Bathurst, 1997). As a result, parameters are often applied in assumption of homogeneous in space. To improve over-simplification in using single parameter, many studies have applied GIS (Geographic Information System) and complementary information to symbolize the spatial heterogeneity. For example, Pack et al. (2001) gives different friction angles to individual geologic formation basing on GIS database and Rompaey et al. (2005) applies Landsat Satellite imagery to characterize the spatial pattern of vegetation.

Whether model parameters are spatially characterized or a single parameter set being applied, modeled susceptibility maps are evaluated by comparing with actual landslide map for parameter calibration; thus, a single simulation with “optimal” parameter set is presented. Researchers recognized the problem of presenting a single modeling result for the nature of one specific parameter combination simply cannot stand for the complex system (Beven and Freer, 2001; Beven, 2006) and the selected “optimal” parameter combination does not mean each parameter therein is optimal due to inter-compensation among parameters in the model structure. Moreover, to accurately measure the model performance is never an easy task. So far, none well-accepted index has been published to retrieve “optimal” simulation.

Correspondence to: Shuh-Ji Kao
(sjkao@gate.sinica.edu.tw)

In this study, we conduct spatial pattern characterization for landslide modeling. More importantly, we incorporate the generalized likelihood uncertainty estimation (GLUE; Beven and Binley, 1992), in which we adopted a newly proposed index for performance measure, into landslide model to extract a group of better simulations. Proper ranges for parameters which would benefit future model application in analogous environments lacking field observations are thus identified. In an attempt to rule out suspicions whether a single susceptibility map is sufficient a representative, we assimilate better simulations together as one. An integrated map envelops parameter's intrinsic variance (i.e. uncertainty caused when a single value is applied) in spatial and temporal scale, such as seasonal changes in vegetation growth and fluctuation of rainfall intensity. This stochastic procedure is applicable for various models. We apply it to the Chi-Jia-Wan watershed located in central Taiwan by using the SHALSTAB model. The GLUE has been widely used in hydrological modeling to acquire parameter uncertainties and confidence intervals (Campling et al., 2002). However, this is the first attempt to include the GLUE into landslide modeling. Landslide sites are successfully predicted by using the integrated landslide susceptibility map.

2 To generate, extract and integrate landslide susceptibility maps

In the following, we detail the SHALSTAB model and criteria used in the GLUE to show the procedure to generate, extract and integrate landslide susceptibility maps. The flow chart is shown in Fig. 1.

2.1 The SHALSTAB Model

The model is proposed by Montgomery and Dietrich (1994) and named as SHALSTAB later on by Dietrich et al. (1998). SHALSTAB shares similar governing equation with SINMAP (stability index mapping) proposed by Pack et al. in 1998. Both models have been widely applied in mountainous watersheds, such as in Italy (Borga et al., 2002), North America (Dietrich et al., 2001), United Kingdom (Pack et al., 1998), and Taiwan (Hsu 1998). Here we re-coded SHALSTAB by using C++ language. Montgomery and Dietrich (1994) incorporate a steady state hydrological model into the infinite slope model proposed by Hammond et al. (1992). The slope stability equation is as below:

$$FS = \frac{C + (\rho_s Z - \rho_w h)g \cos^2 \theta \cdot \tan \phi}{\rho_s g Z \sin \theta \cdot \cos \theta} \quad (1)$$

where FS is the factor of safety. Theoretically, landslide occurs when $FS < 1$. In the equation C [N/m^2] is the effective cohesion. g [9.81 m/s^2] is the gravitational acceleration. θ is the slope gradient and ϕ is the internal friction angle. ρ_s and

ρ_w [kg/m^3] are the bulk densities for soil and water, respectively. Z [m] stands for the soil depth and h [m] is the water table height above the slip surface.

To model the hydrologic controls on h/Z , Montgomery and Dietrich (1994) adopted a steady state shallow subsurface flow based on TOPMODEL (Beven and Kirkby, 1979) and O'Loughlin's study (1986). According to Darcy's Law, the steady state shallow subsurface flow at any point can be described by rainfall intensity (R) times specific contributing area (a), which should equal to saturated hydraulic conductivity multiplies water table height and $\sin \theta$:

$$R \cdot a = K \cdot h \cdot \cos \theta \cdot \sin \theta. \quad (2)$$

The maximum lateral flux when soil reaches saturation is:

$$T \cdot \sin \theta = K \cdot Z \cdot \cos \theta \cdot \sin \theta, \quad (3)$$

where T is the soil transmissivity (in L^2/T). Combining the above two equations, the soil wetness (w) can be obtained:

$$w = \frac{h}{Z} = \min \left(\frac{R}{T} \cdot \frac{a}{\sin \theta}, 1.0 \right). \quad (4)$$

The soil wetness ranges from 0 to 1.0. Saturated overland flow occurs when soil wetness is greater than 1, yet, this part is simply removed instantaneously from the catchment, and ignored in further computations. Equation (4) can then be substituted into Eq. (1) to calculate the factor of safety:

$$FS = \frac{C + (1 - w \cdot \frac{\rho_w}{\rho_s}) \rho_s g Z \cos^2 \theta \cdot \tan \phi}{\rho_s g Z \sin \theta \cdot \cos \theta}. \quad (5)$$

The water to soil density ratio ranges from 0.35 to 0.7 and was simplified as a constant in most model applications. The range of Z would be set based on field observation. However, to measure the soil depth at watershed scale is practically impossible (Casadei et al., 2003). Therefore, a fixed soil depth is applied in most landslide models (e.g., Wu and Sidle, 1995; Pack et al., 1998; Borga et al., 1998; 2002) though the soil depth is critical in both slope stability and hydrological models (Wu and Sidle, 1995). On the other hand, some modelers develop soil generation model to simulate soil depths at various terrains (e.g. Heimsath et al., 1999; Casadei et al., 2003). The site-specific soil wetness, $\frac{R}{T} \cdot \frac{a}{\sin \theta}$, is determined by the topographic term ($a/\sin \theta$) and the hydrological term (R/T). The topographic term captures the preliminary effects of topography on runoff. The hydrological term depicts the relative magnitude of the precipitation event (represented by R) and the ability of the soil to convey the water downslope (i.e. the transmissivity) (Dietrich et al, 2001). Since no good measurement strategy is available so far for T observations in points to represent spatial pattern (Diodato, 2004) and measurement for rainfall intensity is generally insufficient to cover its spatial heterogeneity, modeler often use R/T ratio to describe this hydrological term.

Spatial patterns of C and ϕ can be configured based on satellite imagery and geological datasets as mentioned earlier. For instance, Van der Knijff et al. (2002) assesses

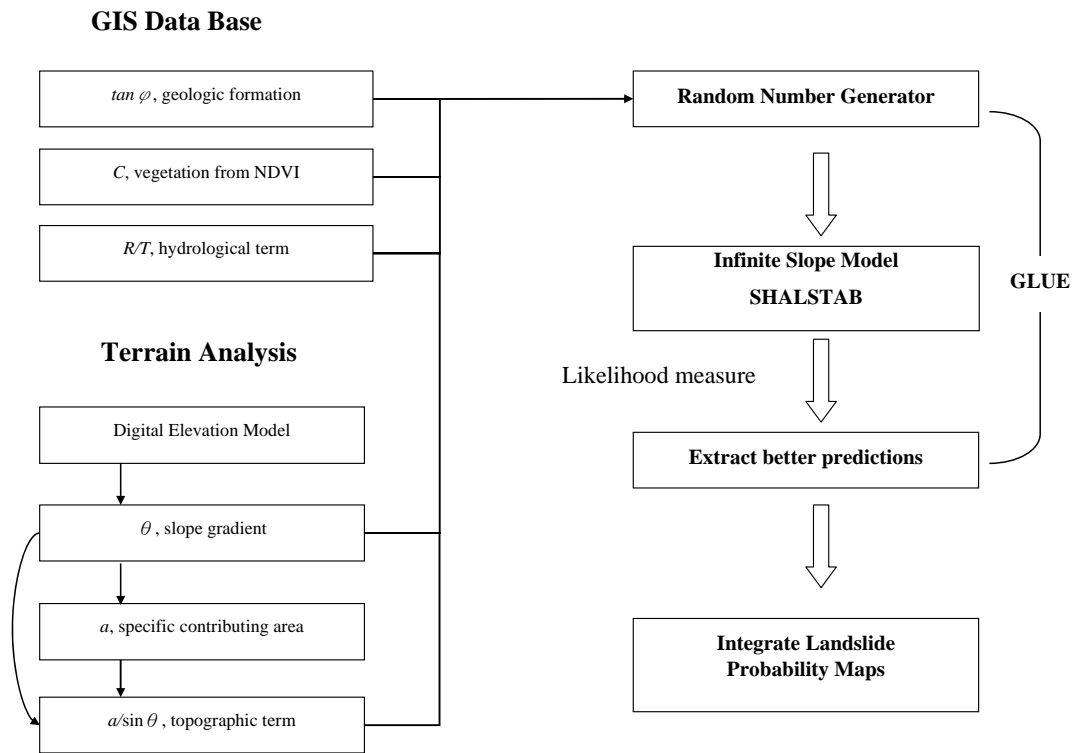


Fig. 1. Research framework of this study. The specific contributing area is generated by infinite flow direction method proposed by Tarboton (1997) and the hydrological term and topographic term are introduced in text.

monthly vegetation cover factor by calculating NDVI-values (Normalized Difference Vegetation Index) from Advanced Very High Resolution Radiometer imagery (AVHRR) (with ground resolution of 1 km²). Rompaey et al. (2005) uses NDVI-value to weigh root strength and assess slope stability in their regional sediment yield model. In this study, we also referred *C* from the NDVI values (see below) to characterize the spatial vegetation. The internal friction angle can be assigned according to geological formations in the GIS database. Based on the SHALSTAB model, we subsequently obtain the landslide susceptibility map.

2.2 Generate and extract susceptibility maps

The GLUE procedure includes a number of requirements (Beven and Binley, 1992):

1. a sampling interval for each parameter,
2. a methodology for sampling the parameter space,
3. a likelihood measure(s) of model performance, and
4. a criterion for acceptance or rejection of model results.

The sampling interval for individual parameters can be set as wide as what is considered feasible based on physical argument or experience (Freer et al., 1996). As for the spacing, the uniform probability distribution is the most popular

method in generating parameter values when insufficient information is available concerning the parameter population for model simulation. In slope stability simulation, adequate sample size is generally recommended from 50 to 2000 sets (Heuvelink, 1993; Duan and Grant, 2000). We generated 4000 parameter sets in our study in order to achieve better representation,. Sampling interval and spacing are illustrated below.

In the GLUE, the most critical step is to accurately measure the model performance. Success rate (*SR*), the ratio of successfully predicted landslide over the number of actual landslide, has often been used (Duan and Grant, 2000). However, traditional *SR* estimation ignores the component of stable cell prediction. Although high *SR* can be achieved the success in stable cell prediction is sacrificed. An improved index for measuring model performance (Modified Success Rate, *MSR*) has been proposed:

$$MSR = 0.5 \cdot SR + 0.5 \cdot \frac{\text{number of successfully predicted stable cell}}{\text{total number of actual stable cell}}, \quad (6)$$

in which *SR* is defined conventionally as mentioned above. The success rate in stable cell prediction is equally weighted. Note that in this equation landslide number instead of cell is used in performance calculation (a landslide usually contains more than one cell). A successfully predicted landslide is

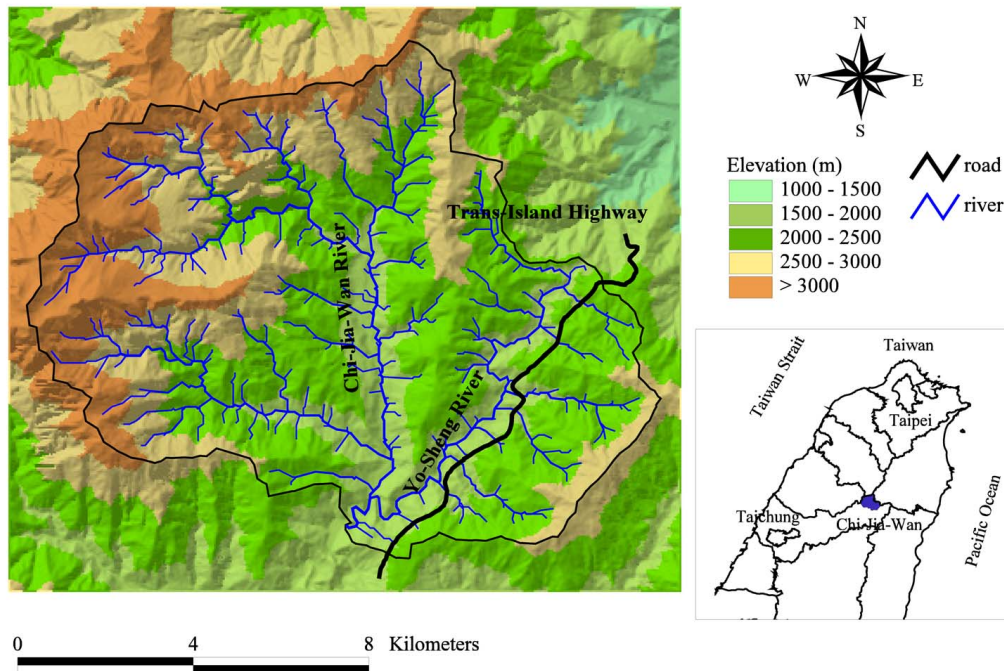


Fig. 2. Topography and stream network of the Chi-Jia-Wan watershed in the central Taiwan. The Trans-Island highway along with the Yo-Sheng River is marked by black curve.

defined as >1 cell is predicted as unstable within the respective landslide site. The reason is that one can only expect the prediction partially overlaps the actual landslide areas that resulted from all complex processes such as triggering and transporting. Details about differences between SR and MSR have been discussed in Huang and Kao (2006).

Based on thousands of simulated cases, Huang and Kao had demonstrated that MSR is more adequate compared to the SR. As a matter of fact, Zhou et al. (2003) has indicated that it is hard to retrieve “optimal” predictions in landslide modeling unless we have a good index for performance measure. The emergence of MSR enables us to develop an objective way to sort better results and to implement parameter range optimization.

As for the criterion for acceptance or rejection, the best 200 out of the 4000 (here the top 5%; depending on researchers) simulations are selected according to MSR-derived model performance. Through GLUE procedure, a group of landslide susceptibility maps is pulled out. Since all landslide susceptibility maps therein are beyond performance threshold, we treat individual map equally without weighting. Meanwhile, proper range is identified for individual parameter based on extracted simulations.

2.3 Integrate susceptibility maps

There are two ways to integrate the top 200 susceptibility maps. The first way is quite straightforward. Overlay the 200 maps together and calculate mean FS for each cell. The sec-

ond is to estimate predicted landslide probability. Boolean variable is applied to every cell distributed in landslide susceptibility map. When the FS value for a cell is less than 1, which is defined as landslide would occur, the boolean variable is assigned as 1, otherwise, 0 is assigned to the cell. Accordingly, we can estimate the cumulative probability of predicted landslide; thus a relative risk measure for landslide potential can be made. Since the 200 FS values for each cell are not always in normal distribution, the above two integration methods have different meanings (see below). Below we applied the proposed stochastic procedure to a mountainous watershed in Taiwan to illustrate the ability of this extraction procedure and the advantage of the integrated landslide susceptibility map.

3 A case study

3.1 Study area

The Chi-Jia-Wan watershed, where we have good landslide database and sufficient geology and vegetation information, locates in central Taiwan with an area of 106 km² and an average slope of 33.8 degree (Fig. 2). It comprises rugged mountainous terrain with elevations ranging from 1131 to 3882 m a.s.l. The entire watershed has experienced alpine climate with mean annual temperature of 14°C. The mean annual rainfall is around 2300 mm, of which more than 50% occurs in the summer brought about by typhoons and

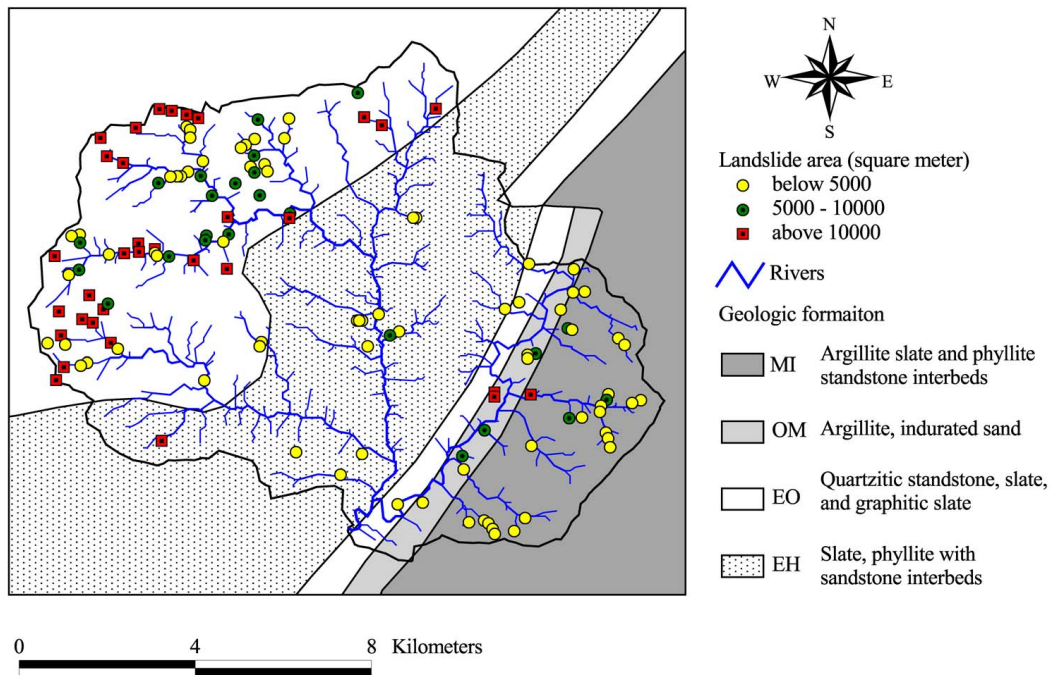


Fig. 3. The spatial patterns of geological formations and landslide sites. The yellow circular, green circular, and red square mark the size of landslide (in m²).

thundershowers. Natural forest covers 87% of the watershed while orchards and vegetable plantations occupy a significant portion of the remaining. Inhabitants are scattered in the watershed but mostly concentrated along the Trans-Island highway passing through the watershed (Fig. 2).

The studied watershed consists of four low-grade metamorphic rocks (MI, OM, EO and EH) of lower Cenozoic formations (lithology shown in Fig. 3). Steep slopes are mantled with shallow clayey or loamy soils underlain by partly weathered regolith. Stream erosion and landslides triggered by heavy rainstorms have formed a deeply dissected landscape. Spatial distribution of those landslides is shown in Fig. 3. According to aerial photos and satellite image, a total of 131 landslides are identified (Industrial Technology Research Institute, 2001). Landslides occur more frequently (62% of recorded landslide sites) in the EO formation and those landslides are significantly larger in size compared to others (30 of 34 large landslides (>10 000 m²) are in the EO Formation). In EO and MI formation landslides locate mostly at channel heads while they occur mostly on riverbanks in OM and EH formation. Field observation suggests that small landslides at the lower elevation (<2100 m) are caused by anthropogenic activities such as road construction and agriculture.

3.2 Input parameters for studied watershed

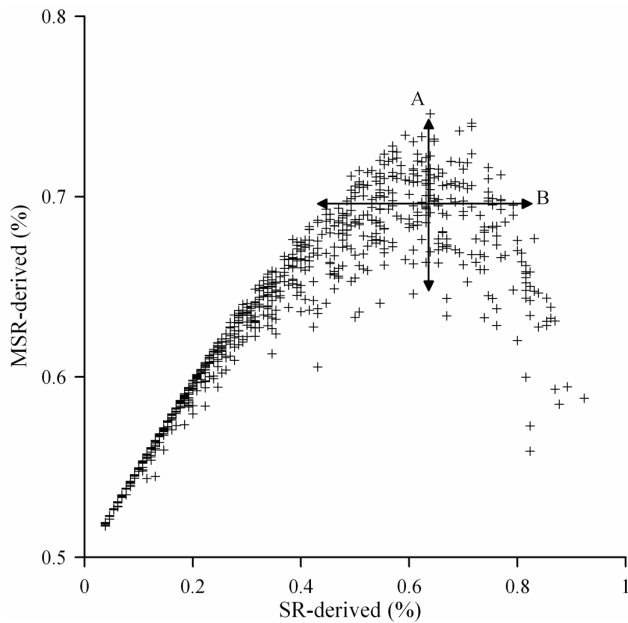
There are 6 essential parameters in the SHALSTAB: specific contributing area (*a*), slope (*θ*), water to soil density ratio

(ρ_w/ρ_s), soil depth (*Z*), internal friction angle (*φ*), and effective cohesion (*C*). Parameter *a* and *θ* are calculated using grided DEMs (40-m resolution), respectively, by infinite direction algorithm (Tarboton, 1997) and Zevenbergen and Thorne’s method (1987). Since we do not have sufficient field observations and no good soil model can properly predict the soil depth in a complex geologic zone in Taiwan, we set 2.5 g/cm³ and 1.5 m, respectively, for the bulk soil density and soil depth in our model based on the field work from Cheng (2003). Thus, only three process-related parameters (*C*, *R/T*, and *φ*) requiring calibration in SHALSTAB model remain unknown.

The intervals of the left three variable parameters are set in reasonable range as wide as possible (see Sect. 2.2). Uniform probability distribution is adopted for the parameter sampling space (Table 1). To accommodate the spatial heterogeneity, the NDVI values retrieved from SPOT-5 are used to configure the spatial pattern of *C*. We define minimum and maximum *C* values (*C*_{min} and *C*_{min} + *C*_{interval}, respectively) to transfer the full spectrum of NDVI value (−1.0~1.0) into effective cohesion for each cell. The *C*_{min} and *C*_{interval} are set to be 0.0~20.0 and 0.0~30.0, respectively, to generate a maximum range of 50.0 kpa cohesion to cover the potential spatial and seasonal changes of vegetation term. Once the *C*_{min} and *C*_{interval} are generated, we can produce a spatial pattern of *C* transferred from NDVI spatial pattern by linear interpolation. For instance, the *C* value in a cell with NDVI=0.2 will equal to 17.0 kpa when *C*_{min}=5.0 and *C*_{interval}=20.0 are selected.

Table 1. Model parameters, their ranges and assumption of probability distribution.

Parameter	Meaning of parameter	Range	Distribution
$C_{(x,y)}$	The effective cohesion [kpa]. Defined as: $C_{(x,y)} = C_{\min} + C_{\text{interval}} \cdot \frac{\text{NDVI}_{(x,y)} + 1}{2}$	C_{\min} : 0~20 C_{interval} : 0~30	Uniform
ϕ	The internal friction angle of corresponding formations [degree].	30~45	Uniform
R/T	A compound parameter of rainfall intensity and transmissivity. R [mm/hr], and T [m ² /hr]	R : 1~20 T : 0.001~10 R/T : 10 ⁻⁶ ~10 ¹	Uniform

**Fig. 4.** The scatter plot of *MSR*- against *SR*-derived model performance for the 4000 runs. Line A and Line B are for reference (see text).

To symbolize the spatial pattern of internal friction angle (ϕ), we assume the pattern of overlying soil closely links to parent rock types. Thus, 30–45 degree is assigned separately for the four lithologic formations based on the GIS database. For R/T ratio in the entire watershed randomly selected R (range 1–20 mm/hr) is divided by randomly selected T (range 0.001–10 m²/hr) to create a wide spectrum of the hydrological term (R/T ratio) since R/T fluctuates over orders-of-magnitude through time. Table 1 denotes three process-related parameters, their ranges and assumption for parameter spacing in this study.

4 Results and discussions

4.1 Model performance

Following *GLUE*, we set a reference value for model performance to define the applicable simulations. Here we demonstrate the advantage of using *MSR* to measure model performance and problems in retrieving optimal model outcome. The 4000 values of model performance derived from newly proposed *MSR* are plotted against the *SR*-derived efficiencies (Fig. 4). *SR*-derived efficiencies range from 0.03 to as high as 0.9 while *MSR*-derived efficiencies fall in a narrower range from 0.53 to 0.78. A dome-shape distribution is shown between the two method-derived efficiencies (Fig. 4). As mentioned earlier, the old *SR* only considers the success of landslide prediction precluding the detection of over-prediction. An applicable simulation should be defined as one that enables to identify the maximum number of landslides with minimum percentage of land predicted fail. For example, a simulation that predicts 100% of mapped failure but with 80% of the watershed area as unstable has no discrimination power. In other words, to avoid over-prediction is important in calibration procedure.

In Fig. 4, *MSR*-derived performance increases together with *SR*-derived performance at the beginning; yet, when *MSR*-derived value (Y-axis) reaches the summit and starts to decrease but *SR*-derived efficiencies keep increasing. The increase in X-axis indicates progress in landslide site prediction while the increase of Y-value indicates better performance in both stable and unstable cell predictions. Obviously, to bring *MSR* in the *GLUE* enables us to avoid over-prediction (Huang and Kao, 2006) and to rank the goodness of simulations. However, even though *MSR* is applied we are still not sure every parameter is optimal in the simulation when the highest performance is obtained since similar model outcome can be achieved through compensation via co-variance in different parameter combinations. This is the concept of equifinality emphasized by Beven (1993, 2006). In terms of *MSR*-derived performance, we show the top 5 parameter sets (Table 2, parameters differ from each other in the 5 cases) as an example to illustrate the dilemma in

Table 2. Parameter values from the top 5 ranked parameter sets.

Rank	C_{-mean} (kpa)	$\text{Log}(R/T)$ (1/m)	ϕ_{-mean} (degree)	Predicted unstable cells	Predicted landslides	Performance (MSR)
Rank 1	6.7	-2.54	40.3	10483	83	0.75
Rank 2	5.5	-2.49	40.0	16128	93	0.74
Rank 3	8.0	-2.42	34.2	16355	93	0.74
Rank 4	7.9	-2.13	38.1	15194	90	0.74
Rank 5	9.7	-2.54	34.1	9024	77	0.73

Total landslides: 131; Total actual stable cells: 64753

selecting the optimal simulation. For instance, the parameter sets for the rank 1 and rank 5 cases have the same hydrological term (R/T), which means similar hydrological magnitude, yet in rank 1 the lower C_{-mean} coupled with higher internal friction angle generates a similar MSR value as that in the rank 5 case. In another words, more than one combination (selected by chance in model), if proper, can produce similar high MSR values; nevertheless, those parameters can not possibly validated site-specifically (since our C and ϕ are distributed parameters for the purpose of convenient discussion we simplify the distributed C and ϕ values as C_{-mean} and ϕ_{-mean} by averaging values in the entire watershed).

On the other hand, since MSR is determined by success in both stable and unstable cell predictions, we may have similar performances for different outcomes (see below). While the top 5 simulations successfully predict 77 to 93 of the 131 actual landslides, a wide range of 9024 to 16355 cells are also predicted unstable in those 5 cases. The number of predicted landslide increases simultaneously as the increasing amount of unstable cells (Table 2). This is a problem cannot be solved mathematically since the model performance is a consequence of stable and unstable cell predictions, which are inter-exclusive within a given area (an improvement in one prediction results in a deterioration in another). In a sense, even we use MSR to measure model performance we cannot tell which one is closest to reality. A notable feature shown in Fig. 4 is that with any given SR-derived performance (similar performance in predicting landslide site) we have a range rather than a fixed value for MSR -derived values (e.g., Line A in Fig. 4) and vice versa (Line B). The interval of Line A denotes that we have many simulations holding a fixed number of predicted landslides yet having different amount of stable cells been predicted (according to Eq. 6). The same circumstance occurs for Line B. Simply speaking, we cannot single out the optimal simulation since there is no reason to exclude those “near-optimal” simulations judging by model performance.

A conclusion can be drawn accordingly that it is improper to present single model result. Therefore, instead of presenting one optimal simulation, we select a group of applicable simulations. Here we define applicable simulations as

when model results have MSR -derived values greater than 0.68 (Line B in Fig. 4; modelers can define the threshold or criterion basing on their judgments), which is the equivalent to the best 5% in the 4000 simulations. By grouping applicable simulations, uncertainties derived from the entire model processing, such as model structural error or input parameter error can be enveloped to approximate to the mean. Below we further integrate these better simulations as one for presentation. Model equifinality problem still exists but is smoothed out through our integration.

4.2 Parameter ranges

As mentioned in 3.2, to accommodate the spatial heterogeneity we used satellite image and GIS lithology map to symbolize, respectively, the spatial patterns of effective cohesion and internal friction angle. After running dynamic parameter ranges, we went back to run the model by using homogenous C_{-mean} and ϕ_{-mean} for the entire watershed to identify the difference in model performance. Results show that model performances are 8~12 % higher when we apply NDVI and geologic theme to configure the spatial patterns of effective cohesion and internal friction angle.

To uncover proper parameter ranges in studied watershed we plot model performances of all 4000 combinations against three corresponding parameters separately. According to our parameter design, values of C_{-mean} range from 0.0 to 40.0. In the condition of >20.0 kpa, MSR values of simulations are consistently as low as 0.52 (thus not show in the Fig. 5). In the condition of $C_{-mean} < 20.0$ kpa, a negative correlation is found against model performance forming a dome shape (Fig. 5a) skewing toward the lower C_{-mean} . Only when C_{-mean} is smaller than 13.0, efficiencies can go beyond 0.68. Better simulations, obviously, only occur when C_{-mean} values fall in the range of 2.5–11.0 kpa (above the solid line; mean=6.4, standard deviation=2.1 for those 200 predictions). After extracting by the GLUE process, the wide-range C is substantially narrowed down to a “proper range” for the study area. Field measurement in the studied watershed reports that soil cohesion in the colluviums ranges from 1.9–3.9 kpa (Cheng, 2003). Some previous studies indicate that root cohesion varies from 0.0–12.5 kpa (Si-

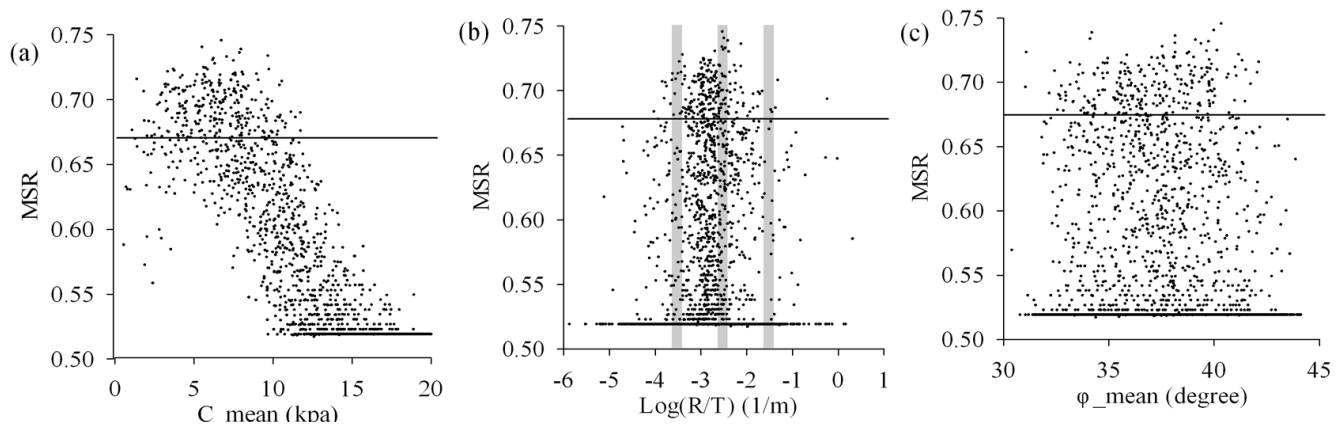


Fig. 5. Scatter plots of *MSR*-derived performances against the individual parameter values. Solid line of *MSR*=0.68 stands for the criterion of the best 200 runs. The gray bars in (b) are used to define three different *R/T* conditions in Fig. 6.

dle, 1991). Combined the root and soil cohesions, the range of effective cohesion will be consistent with our proper range of C_{-mean} (2.5–11.0 kpa). This consistency gives confidence in application of the C_{-mean} to analogue watersheds. Note that, however, many simulations still have low performances even C_{-mean} values fall within the proper range. Apparently, other two parameters in the model work improperly to reduce the model performances.

Figure 5b shows the applicable simulations embrace a wide range $\log(R/T)$ from -4.0 to -1.0 . Outside this range model, simulations never reach high performance. However, this range of $\log(R/T)$ does not guarantee high model performance since the other two parameters may work improperly. The mean value of the 200 pairs of *R/T* is 0.006, which equals -2.2 after \log transformation. This $\log(R/T)$ value is larger when compared with the modeled *R/T* in the case of the west of Nimpkish Lake and in the south of Port McNeill, British Columbia ($-3.3\sim-3.0$; Pack et al., 2001); and the values applied in the California Coast Range, U.S. ($-3.1\sim-2.8$; Dietrich et al., 2001). Since *T* values are similar among those studies and ours, such larger *R/T* ratio may be due to higher rainfall intensity in Taiwan. In the studied watershed, a 14-year historical data shows mean daily rainfall ranges from 3.1 to 8.3 mm; yet, the maximum daily rainfall was up to 358 mm/day; and cumulatively can be up to 626 mm within two days.

Unlike C_{-mean} and *R/T*, the parameter ϕ_{-mean} (Fig. 5c) spreads randomly against model performances showing a flatted-top shape without any notable distribution pattern (mean=37.1, standard deviation=2.4). The ϕ_{-mean} (37.1 degree) agrees well with the field observation (35~40 degree) in the studied area (Cheng, 2003). However, the diffusive and flat-topped pattern may suggest the internal friction angle is not sensitive in this model.

Nevertheless, through the application of the *MSR* in the GLUE procedure we exclude improper parameter ranges for model application in this watershed. We also realize that

governing parameters will interplay among each other and at the same time create complex simulations. Even when we have a proper range for individual parameter, we do not necessarily have high performances if the combination is inadequate. Below, we try to make a 3-D plot to probe the inter-correlation among the three parameters.

4.3 Inter-correlation among parameters

Three variables are used as three axes in a 3-D plot to unravel the response of model performance as parameters are changed at watershed scale. The area average of parameter values within the whole watershed is used for discussion. Discussion by using the area average may smooth out precise inter-correlations among parameters; yet, discussing parameter relationship in distributed approach (taking the mean of parameter value, for example) is well-applied in many applications (e.g. Candela et al., 2005).

We put $\log(R/T)$, which represents the intensity of the external driving force, in Z-axis. The contour of model efficiency (so called response surface) is plotted against C_{-mean} and ϕ_{-mean} at different $\log(R/T)$ levels. Three levels of $\log(R/T)$ are classified, $-1.4\sim-1.6$, $-2.4\sim-2.6$, and $-3.4\sim-3.6$ (Figs. 6a, b, c), to feature heavy, medium and light rainfall intensity conditions. At different rainfall intensities we find different contour patterns for the *MSR*-derived values. As suffering heavy rains (Fig. 6a), higher *MSR* values only occur when C_{-mean} falls in 2.5 to 13.0 kpa as mentioned earlier. Meanwhile, C_{mean} values vary in a narrow range as the changing of ϕ_{-mean} (from 32.0 to 42.0 degree) within the patch of high *MSR*. Correlation between the two parameters is insignificant in the heavy rainfall condition.

At the medium rainfall level, a negative correlation between C_{-mean} and ϕ_{-rmean} is found for the high *MSR* patch. Lower ϕ_{mean} coupled with higher C_{-mean} can better fit the actual landslide map. Two aggregated areas generate higher

MSR; one located around $C_{-mean}=9.7$ and $\phi_{-mean}=34.1$, and the other located around $C_{-mean}=5.5$ and $\phi_{-mean}=40.0$.

At the low rainfall level, we note a significant negative correlation between C_{-mean} and ϕ_{-mean} . The highest *MSR* appears at the Point B (Fig. 6c), where $C_{-mean}=3.7$ and $\phi_{-mean}=39.0$. By contrast, in heavy rainfall condition the highest *MSR* is found at Point A (Fig. 6a), where $C_{-mean}=8.9$ and $\phi_{-mean}=34.3$. Similar performance is achieved basing on quite different two parameter sets.

The three contour maps demonstrate effects of inter-correlation among parameters on model performances. The shear strength in SHALSTAB is a function of C_{-mean} , ϕ_{-mean} and $\log(R/T)$. In light rainfall (insignificant contribution of R/T) condition, co-varied C_{-mean} and ϕ_{-mean} pairs can generate similar shear strength, consequently, their combination can perform similar *MSR*. By contrast, the shear strength tends to be diminished according to enhanced effect of R/T due to heavy rainfall. In this situation, no matter how the ϕ_{-mean} changes, the shear strength would be predominantly influenced by R/T . Thus correlation between C_{-mean} and ϕ_{-mean} becomes indistinct. Parameter interaction inherited in model structure, apparently, creates the “equifinality” problem. The *MSR* value obtained at the Point A ($MSR=0.71$) in heavy rainfall condition and at the Point B ($MSR=0.72$) in light rainfall condition are good examples. Unless we have significant progress in field observations to calibrate all parameter over space and time scale we cannot prove which parameter combination approaches reality the best. The inter-correlation among parameters revealed by model outputs tells us how the model efficiency responds to parameter changes; yet, the correlation is not necessarily a reality in natural environments. Below we suggest assimilating applicable susceptibility maps to envelope inevitable uncertainties both inherited in model structure and derived from parameter variability across space and/or time scale; thus, to approximate the mean condition.

4.4 Integrate landslide susceptibility maps

As mentioned earlier, we overlap maps and take mean for *FS* values in each cell for the 200 better predictions; hereafter, the map is named as mean *FS* map. On the other hand, we use boolean variable to calculate the occurrence probability of $FS < 1$ for each cell in the 200 runs to quantify predicted landslide potential; hereafter, we call it integrated landslide probability map. For each cell we have a mean *FS* value and an occurrence probability for landslide potential. Such integration bundling applicable simulations gives us more objective results in landslide risk assessment.

Both maps are useful and meaningful yet they are not the same. We pull out 200 *FS* values, respectively, from two cells in the watershed as example to show their probability distribution of *FS* values and to examine differences in using the two integration methods. In Fig. 7a, the selected pixel holds *FS* value from 0.6 to 1.3 with a mean of 1.02 in the

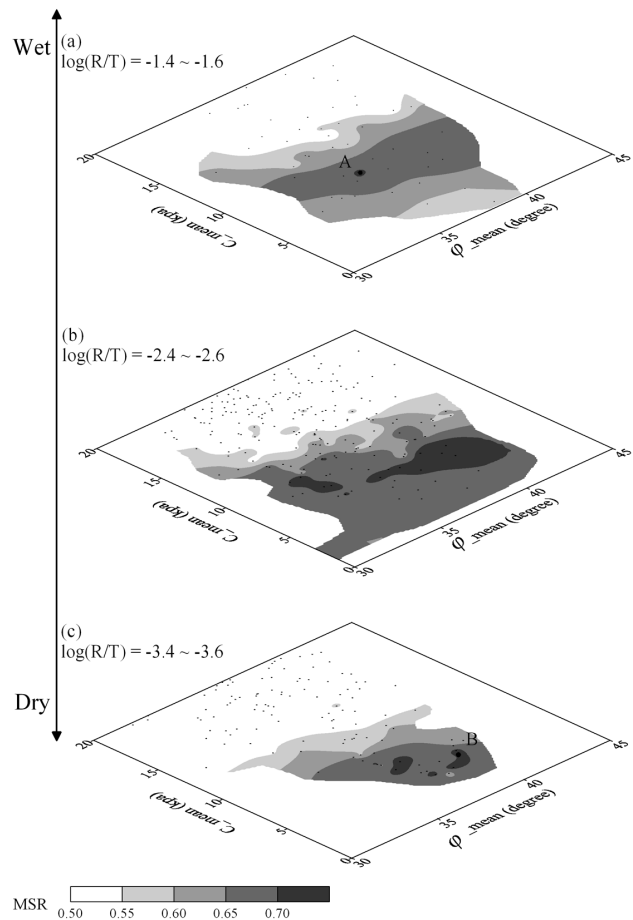


Fig. 6. Contour planes of *MSR*-derived performances against C_{-mean} and ϕ_{-mean} under heavy rainfall ($\log(R/T)=-1.4\sim-1.6$), medium rainfall ($\log(R/T)=-2.4\sim-2.6$), and light rainfall ($\log(R/T)=-3.4\sim-3.6$) conditions. Color scale is shown below. Dots on planes represent coordination of parameter combinations. Extrapolations are blanked for all three conditions. Data points in between the three R/T categories are not shown.

200 runs. The frequency distribution shows 34% of the modeled *FS* values are lower than 1.0. The distribution slightly shifts toward high *FS* direction. The second cell holds the same range of *FS* value (0.6–1.3) yet with different mean *FS* value (0.98) and the frequency distribution pattern shifts toward low *FS* direction, that is, 61% cases contain *FS* values < 1.0 (Fig. 7b). Judging from the mean *FS*, we can only see a slight difference (1.02 and 0.98) between the two cells. By contrast, the occurrence probability gives the first pixel 34% for landslide potential and gives the second cell a significant higher potential of 61% (factor of ~ 2.0). Both integration methods provide consistent result that the first cell is stable and the second is unstable if we take 50% as threshold for landslide prediction. Compared to the mean *FS*, the calculation of occurrence probability magnifies the tiny *FS* signal difference around *FS* of 1, the point we really care about in landslide risk assessment.

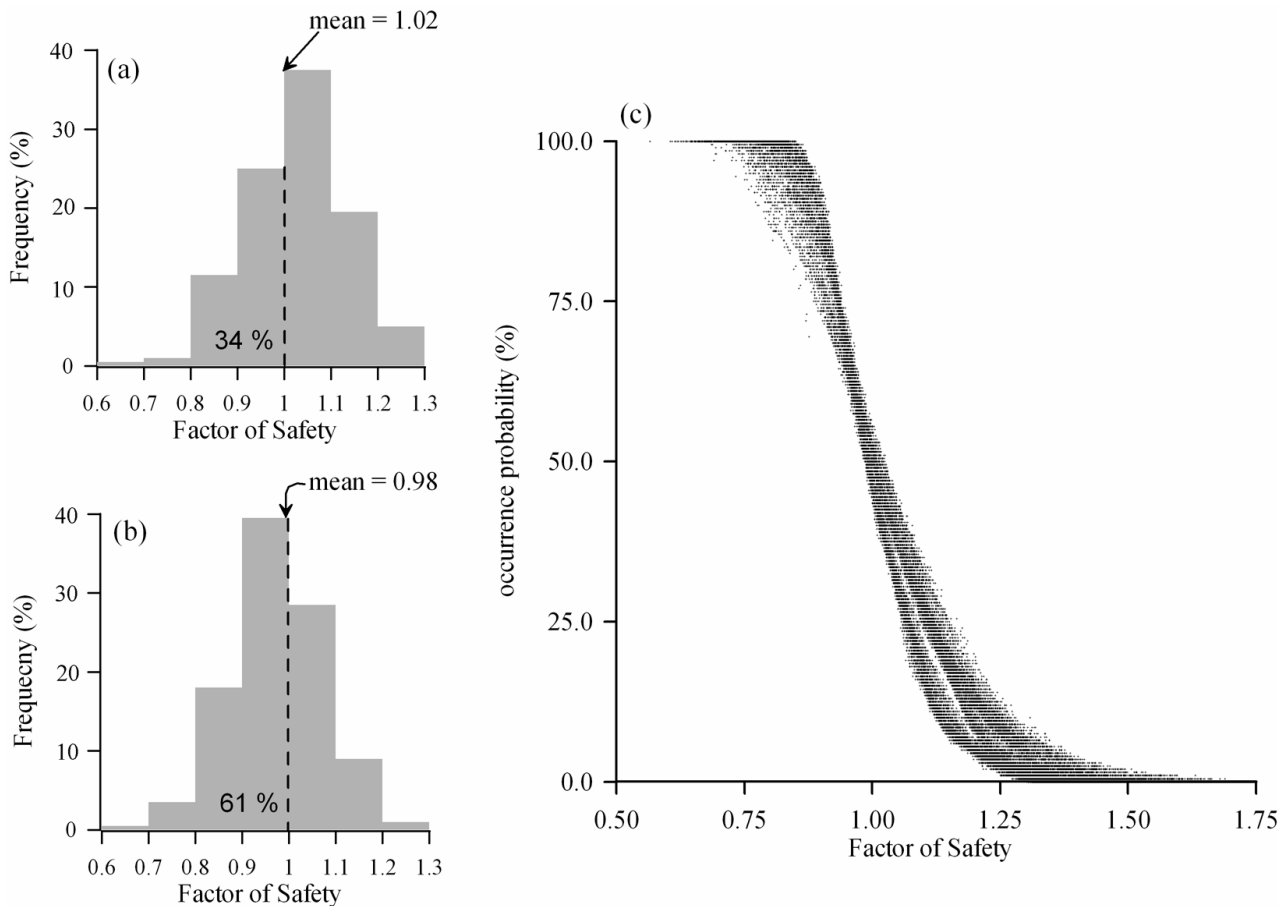


Fig. 7. (a) and (b) are histograms of the 200 *FS* values in the two example cells with similar mean *FS*. Dash lines indicate the mean *FS* value for the cell. (c) The scatter plot of mean *FS* (from mean *FS* map) against the landslide potential (from the integrated landslide probability map) in respective cell over the entire watershed.

From basin-wide perspective (Fig. 7c), an inverse correlation is found between mean *FS* value and occurrence probability for predicted landslide in each respective cell. Confined correlation appears around *FS* value of 1 (~50% for occurrence probability) forming a near-linear correlation. Confined linear correlation in the middle interval indicates both methods can perform fairly consistent result for cells having *FS* values around 1. However, at both ends data points are diffusively curving toward high and low *FS* directions. The diffusive behavior at both ends is caused by the wider range *FS* value and possibly in non-normal distribution. The very stable (referred to $FS > 1.25$) and very unstable ($FS < 0.75$) cells at both ends can be classified by occurrence probability of <10% and >90%, respectively. The using of integrated landslide probability map substantially narrows the divergent *FS* spectrum (from 0.0~5.0 or larger) down to 0.0~100% and magnifies *FS* signals around the middle interval that we are focusing at. Moreover, the using of occurrence probability for landslide potential gives more understandable presentation to the public. Therefore, the Boolean variable method is preferable for integration.

4.5 Comparison with actual landslide map

Figure 8a shows the integrated landslide probability map derived from the best 200 simulations. High values (>90%), which reflect high landslide potential, are mostly disseminated in the EO formation, summarize those areas subjected to landslide potential of: 0–30%, 30–50%, 50–70%, 70–90%, and 90–100%, occupy 75.41%, 8.22%, 5.39%, 4.41%, and 6.57%, respectively, of the entire watershed area. The high-risk area (>50%) embraces only 16.37% of the entire watershed indicating the *MSR* effectively limits the over-prediction (Huang and Kao, 2006). Generally, the proportion of landslide prone area rarely exceeds 11.0% of the entire watershed (depending on regional characteristics) in natural situation (Casadei et al., 2003).

The landslide potential classified by formations is examined by actual landslides in Table 3. For the high-risk area (>50.0%) we find 77 actual landslide sites, which equals 59% (77/131) success. Unexpectedly, 41% actual landslides appeared in areas with low-risk (<30%), of which most landslides are small and most of them located at low elevation

Table 3. The distribution of predicted landslide potential (in %) in the four formations in integrated landslide probability map and number of actual landslides observed in the respective domain.

relative risk measure	<30%	30–50 %	50–70 %	70–90 %	>90%
EO (81)	12	3	5	8	53
EH (14)	10	2	0	2	0
OM (11)	9	0	1	1	0
MI (25)	14	4	1	6	0

Number in parentheses indicates total landslide sites in the respective formation.

in the EH, OM, and MI formations (Fig. 8b) where anthropogenic disturbances and unconsidered mechanisms would often occur. Field observations indicate those landslides are subjected to agriculture, such as vegetable gardens or orchards, particularly in the MI formation. The small landslides in the OM formation are obviously related to road construction. Along river channels in the EH formation, some slope failures are not successful identified due to river lateral erosion, which is not considered in the slope stability model. Meanwhile, we realize that a landslide rasterized by less than 4 cells is hard to be predicted. Only when we have sufficient DEM resolution, those small landslides can be possibly identified. Nevertheless, if we exclude the region below 2100 m, 75% of total actual landslides (74/98) are successfully predicted basing on the criterion of high-risk. Some stable areas fallen in predicted high-risk zones warrant immediate attention.

Through this case study, the SHALSTAB has fully been demonstrated applicable in mountainous watersheds in Taiwan; meanwhile map-extracting procedure is proven beneficial in land management effectively because over-prediction has been reduced significantly. Besides, the integrated landslide probability map provides a decision-making reference. Incorporating socio-economic considerations, land managers therefore may adopt a less strict landuse regulation, i.e., prohibit further development only in areas subjected to the highest risk so that appropriate land use plan or mitigation can be implemented.

5 Conclusions

This is the first study introducing GLUE into a landslide model to generate, extract and integrate landslide susceptibility maps. The new model performance indicator play critical role in the GLUE to avoid over-prediction and to extract better simulations. Although model performance indicator can depict the goodness of model simulations, the attempt to single out one ‘optimal’ simulation is still without success. Applicable (near-optimal) simulations are grouped to envelope uncertainties inherited in model structure and to input param-

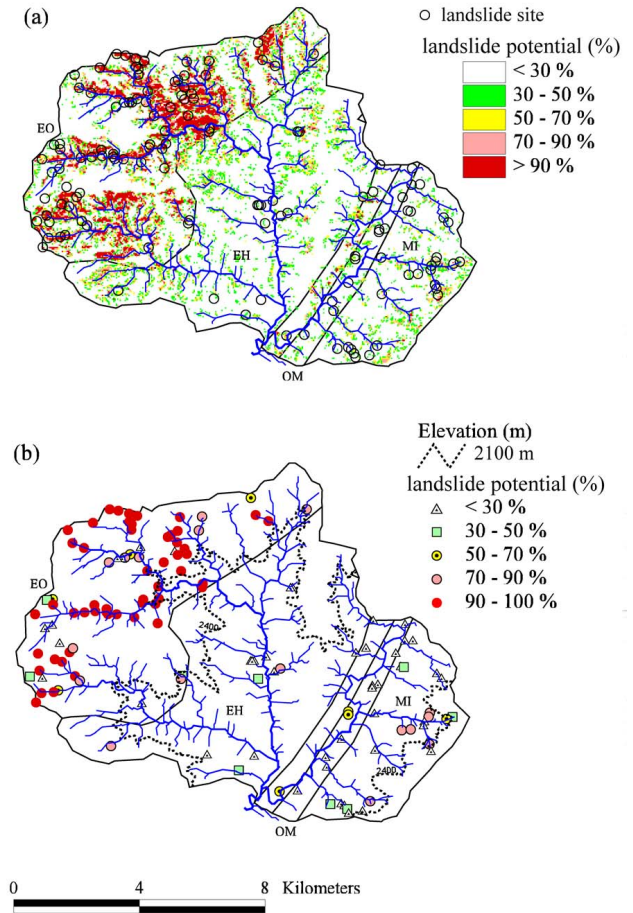


Fig. 8. (a) Integrated landslide probability map and spatial distribution of observed landslide locations. (b) The predicted landslide potential (in %) calculated for each actual landslide site.

eters that are naturally variable in space and time scale. This integration process promotes the representation of landslide susceptibility map. Equifinality problem is not solved but smoothed through this stochastic procedure. The case study in the Chi-Jia-Wan, a mountainous watershed in Taiwan, reveals the applicability of the SHALSTAB model and the integrated landslide probability map. Over-prediction is reduced with MSR by which high-risk areas (with >50% landslide potential) only occupy 16.4% of the entire watershed while 59% of the actual landslides are correctly identified. Excluding the low elevation area with unexpected disturbances, the overall performance is as high as 76%. The stochastic procedure proposed here is clearly defined so that personal bias in risk assessment is minimized and can be applied onto various kinds of landslide models. We also suggest to use integrated landslide probability map since it magnifies FS signals around 1 which acts as a sensitive index for landslide prediction and in the mean time normalizes the wide range FS signals to 0–100%, which is more understandable to the public.

Acknowledgements. This research is funded by the National Science Council of Taiwan, under the project NSC 90-2621-Z-002-018 and 94-2611-M-001-004. The authors are also grateful to Industrial Technology Research Institute and Chuan-Yao Lin for providing, respectively, the landslide data and rainfall records. Meanwhile we are grateful to K. Beven, K. T. Chang and anonymous reviewers for providing a very helpful review of the manuscript.

Edited by: T. Glade

Reviewed by: K. Beven and two other referees

References

- Beven, K.: A manifesto for the equifinality thesis, *J. Hydrol.*, 320(1–2), 18–36, 2006.
- Beven, K. and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modeling of complex environmental systems using the GLUE methodology, *J. Hydrol.*, 249(1–4), 11–29, 2001.
- Beven, K.: Prophecy, Reality and Uncertainty in Distributed Hydrological Modeling, *Adv. Water Resour.*, 16(1), 41–51, 1993.
- Beven, K. J. and Binley, A. M.: The future of distributed models: model calibration and uncertainty prediction, *Hydrol. Processes*, 6, 279–298, 1992.
- Beven, K. and Kirkby, M. J.: A physically based, variable contributing area model of basin hydrology, *Hydrol. Sci. Bull.*, 24, 43–69, 1979.
- Borga, M., Dalla Fontana, G., Gregoretto, C., and Marchi, L.: Assessment of shallow landsliding by using a physically based model of hillslope stability, *Hydrol. Processes*, 16, 2833–2851, 2002.
- Borga, M., Dalla Fontana, G., Da Ros, D., and Marchi, L.: Shallow landslide hazard assessment using a physically based model and digital elevation data, *Environ. Geol.*, 32(2–3), 81–88, 1998.
- Burton, A. and Bathurst, J. C.: Physically based modeling of shallow landslide sediment yield at a catchment scale, *Environ. Geol.*, 35(2–3), 89–99, 1997.
- Candela, A., Noto, L. V., and Aronica, G.: Influence of surface roughness in hydrological response of semiarid catchments, *J. Hydrol.*, 313, 119–131, 2005.
- Campling, P., Gobin, A., Beven, K. J., and Feyen, J.: Rainfall-runoff modeling of a humid tropical catchment: the TOPMODEL approach, *Hydrol. Processes*, 16, 231–253, 2002.
- Casadei, M., Dietrich, W. E., and Miller, N. L.: Testing a model for predicting the timing and location of shallow landslide initiation in soil-mantled landscapes, *Earth Surface Processes and Landforms*, 28, 925–950, 2003.
- Cheng, Y. L.: Study on Risk Analysis of Slopes Consideration Spatial Variability – A Case Study of the Lisan, Master Thesis, National Chung Hsing University, 2003.
- Dietrich, W. E., Bellugi, D., and Real de Asua, R.: Validation of the shallow landslide model, SHALSTAB, for forest management, in: *Land Use and Watersheds: Human influence on hydrology and geomorphology in urban and forest areas*, edited by: Wigmosta, M. S. and Burges, S. J., Amer. Geophys. Union, Water Science and Application, 2, 195–227, 2001.
- Dietrich, W. E. and Montgomery, D. R.: SHALSTAB: A digital terrain model for mapping shallow landslide potential, <http://socrates.berkeley.edu/~geomorph/shalstab>, 1998.
- Diodato, N.: Local models for rainstorm-induced hazard analysis on Mediterranean river – torrential geomorphological systems, *Nat. Hazards Earth Syst. Sci.*, 4, 389–397, 2004, <http://www.nat-hazards-earth-syst-sci.net/4/389/2004/>.
- Duan, J. and Grant, G. E.: Shallow Landslide Delineation for Steep Forest Watersheds Based on Topographic Attributes and Probability Analysis., in: *Terrain Analysis – Principles and Applications*, edited by: Wilson, J. P. and Gallant, J. C., John Wiley & Sons, New York, 311–329, 2000.
- Freer, J., Beven, K., and Ambroise, B.: Bayesian estimation of uncertainty in runoff prediction and the value of data: an application of the GLUE approach, *Water Resour. Res.*, 32(7), 2161–2173, 1996.
- Hammond, C. D., Hall, D., Miller, S., and Swetik, P.: Level I Stability Analysis (LISA): Documentation for version 2.0. Ogden, UT: United States Department of Agriculture, Forest Service, Intermountain Research Station General Technical Report No. 285, 1992.
- Heimsath, A. M., Dietrich, W. E., Nishiizumi, K., and Finkel, R. C.: Cosmogenic nuclides, topography and the spatial variation of soil depth, *Geomorphology*, 27, 151–172, 1999.
- Heuvelink, D. G.: Error propagation in quantitative spatial modeling: Application in GIS, PhD thesis, Utrecht University, Netherlands, 1993.
- Hsu, M. L.: A Grid-Based Model for Predicting Shallow Landslides: A Case Study in Linkou, Taipei, *Eos, Transaction, American Geophysical Union*, 79(24), W25, 1998.
- Huang, J. C. and Kao, S. J.: Optimal Estimator for Assessing Landslide Model Performance, *Hydrol. Earth Syst. Sci. Discuss.*, 3, 1125–1144, 2006, <http://www.hydrol-earth-syst-sci-discuss.net/3/1125/2006/>.
- Industrial Technology Research Institute: Water quality monitoring and management in watershed of Tehchi reservoir the fourth year, Water Resources Agency, Taipei, 2001.
- Montgomery, D. R. and Dietrich, W. E.: A physically based model for the topographic control on the shallow landsliding, *Water Resour. Res.*, 30, 1153–1171, 1994.
- O’Loughlin, E. M.: Prediction of surface saturation zones in natural catchments by topographic analysis, *Water Resour. Res.*, 22, 794–804, 1986.
- Pack, R. T., Tarboton, D. G., and Goodwin, C. N.: The SINMAP Approach to Terrain Stability Mapping, Congress of the International Association of Engineering Geology, Vancouver, British Columbia, Canada, 21–25 September 1998.
- Pack, R. T., Tarboton, D. G., and Goodwin, C. N.: Assessing Terrain Stability in a GIS using SINMAP, in 15th annual GIS conference, GIS 2001, Vancouver, British Columbia, 19–22 February, 2001.
- Rompaey, A. V., Bazzoffi, P., Jones, R. A. A., and Montanarella, L.: Modeling sediment yields in Italian catchments, *Geomorphology*, 65, 157–169, 2005.
- Sidle, R. C., Pearce, A. J., and O’Loughlin, C. L.: *Hillslope Stability and Landuse*. Washington, DC: American Geophysical Union, Water Resour. Monogr., No. 11, 1985.
- Sidle, R. C.: A conceptual model of changes in root cohesion in response to vegetation management, *J. Environ. Qual.*, 20, 43–52, 1991.
- Tarboton, D. G.: A New Method for the Determination of Flow Directions and Contributing Areas in Grid Digital Elevation Models, *Water Resour. Res.*, 33(2), 309–319, 1997.

- Van der Knijff, J., Jones, R. J. A., and Montanarella, L.: Soil erosion risk assessment in Italy, in: Proceedings of the Third International Congress Man and Soil at the Third Millennium, edited by: Rubio, J. L., Morgan, R. P. C., Asins, S., and Andreu, V., Geofoma Ediciones, Logrono, Spain, 1903–1913, 2002.
- Wu, W. and Sidle, R. C.: A distributed slope stability model for steep forested basins, *Water Resour. Res.*, 31, 2097–2110, 1995.
- Zevenbergen, L. W. and Thorne, C. R.: Quantitative analysis of land surface topography, *Earth Surface Processes and Landforms*, 12, 47–56, 1987.
- Zhou, G., Esaki, T., Xie, M., and Mori, J.: Spatial probabilistic modeling of slope failure using an integrated GIS Monte Carlo simulation approach, *Engineering Geology*, 68, 373–386, 2003.