Hydrol. Earth Syst. Sci., 10, 957–965, 2006
www.hydrol-earth-syst-sci.net/10/957/2006/

Hydrology and Earth System Sciences

Optimal estimator for assessing landslide model performance

J. C. Huang and S. J. Kao

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

Received: 20 April 2006 – Published in Hydrol. Earth Syst. Sci. Discuss.: 23 June 2006
Revised: 24 October 2006 – Accepted: 23 November 2006 – Published: 14 December 2006

Abstract. The commonly used success rate (SR) in evaluating cell-based landslide model performance is based on the ratio of successfully predicted landslide sites over total actual landslide sites without considering the performance in predicting stable cells. We proposed a modified SR (MSR), in which the performance of stable cell prediction is included. The advantage of MSR is to avoid over- and under-prediction while upholding the stable sensitivity throughout all simulated cases. Stochastic analyses are conducted by using artificial landslide maps and simulations with a full range of performances (from worst to perfect) in both stable and unstable cell predictions. Stochastic analyses reveal mathematical responses of estimators to various model results in calculating performance. The Kappa method, which is commonly used for satellite image analysis, is improper for landslide modeling giving inconsistent performance when landslide coverage changes. To examine differences among SR and MSR in real model application, we applied the SHALSTAB model onto a mountainous watershed in Taiwan. Case study shows that stable and unstable cell predictions are inter-exclusive in SHALSTAB model. The optimal estimator should compromise landslide over- and under-prediction. According to our 4000 simulations, the best simulation generated by MSR projects 83 hits over 131 actual landslide sites while the unstable cells cover only 16% of the studied watershed. By contrast, despite the fact that the best simulation deduced from SR projects 120 hits over 131 actual landslide sites, this high performance is only obtained when unstable cells cover an incredibly high landslide cover (∼75%) of the entire watershed exhibiting a significant landslide over-prediction.

Correspondence to: S. J. Kao
(sjkao@gate.sinica.edu.tw)

1 Introduction

Landslides triggered by intensive rainfall cause serious damages and result in thousands of deaths and billion dollars of property losses every year. To mitigate these damages, deterministic and non-deterministic (stochastic) models have been developed to generate landslide susceptibility maps to assess the degree of risk (e.g. Ayalew and Yamagishi, 2005; Bora et al., 1998; Pack et al., 2001). To ensure the model effectiveness, all modeled landslide maps must be compared with actual landslide map, which is the end-product resulted from the real landform processes, for validation. Thus, a proper index or an estimator for measuring performance (or efficiency) is essential. The estimator may also serve as a likelihood measure in model calibration.

Previous studies (Montgomery and Dietrich, 1994; Dietrich et al., 1995; Borga et al., 1998; Duan and Grant, 2000; Lee, 2005) had been using "success rate" (hereafter, SR) to evaluate the model performance. The SR is defined as a ratio of how many actual landslide sites are successfully predicted. However, this former SR does not include the success (or failure) in stable cell prediction inherited in the model, thus, precludes the detection of over-prediction of slope failure. For land management an applicable simulation should be defined as one that enables to identify the maximum number of landslides with minimum percentage of land coverage been predicted fail. For example, a simulation that perfectly predicts all shallow landslides but with 80% watershed area been predicted as unstable gives no discrimination power. In other words, detecting over-prediction is crucial in evaluating the performance of modeling.

In this study, we modify the SR by incorporating the success of stable cell prediction into performance estimation. We generate artificial landslide maps and landslide susceptibility maps to examine responses of SR and MSR to model results that have wide degree of performance under different landslide patterns. The Kappa statistic, which is used for

Published by Copernicus GmbH on behalf of the European Geosciences Union.

measuring inter-observers, and logic concepts in analyzing image similarity are discussed. For further annotation, we present a case study to demonstrate differences between SR and MSR methods in real model applications and the advantages of proposed MSR.

2 Methods and logics in image similarity analysis

As a predicted landslide map through cell-based models is generated, only two categories are defined: unstable cell and stable cell. Unstable cell represents a cell been predicted as landslide cell. When comparing the actual landslide maps with the predicted maps, four types of outcomes are possible: Type 1) actual landslide cells are predicted as unstable cells; Type 2) actual landslide cells are predicted as stable cells; Type 3) actual stable cells are predicted as unstable cells; and Type 4) actual stable cells are predicted as stable cells. Type 1 and Type 4 are considered as success in model prediction, while both Type 2 and Type 3 are regarded as failure in prediction. Generally speaking, the performance assessment in landslide model is comparable to the similarity measurement in image analysis. Thus, we invoked the Kappa statistic for comparison (Longley et al., 2001). Below, we review two existing indices in assessing image similarity and proposed a new index, called the modified SR (hereafter MSR).

2.1 The former success rate

The former success rate calculation only takes Type 1 outcomes into account and ignores the other three types. The equation of SR:

$$\text{SR} = \frac{\text{number of successfully predicted landslides}}{\text{total number of actual landslides}}. \quad (1)$$

Note that in this equation number of landslide instead of number of cells is used in performance calculation (a landslide usually contains more than one cell). A successfully predicted landslide is defined as $\geq$1 cell is predicted as unstable within the respective landslide zone. 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. Nevertheless, only Type 1 outcome is considered in this equation.

2.2 The Kappa statistic

The Kappa value is a precise index. It quantitatively measures the magnitude of agreement among inter-observers (e.g. Viera and Garrett, 2005). The equation of Kappa is:

$$\text{Kappa} = \frac{K_a - K_e}{1 - K_e} \quad , \quad (2)$$

where K_a is the actual agreement and K_e is the expected agreement. The concept of this calculation is based on the difference between how much agreement is actually presented (actual agreement) compared to how much agreement would be expected by chance alone (expected agreement). When Kappa is applied to landslide prediction actual agreement is the sum of the probability of Type 1 and Type 4 outcomes. Theoretically, the maximum value of actual agreement is one; thus, the lower the actual agreement the less outcomes of Type 1 and/or Type 4. Expected agreement is the sum of the expected values of predicted actual unstable and stable cells. The expected agreement explains the expected value of the two inter-observers in each class in a random space. Two images are in moderate agreement when the Kappa value is larger than 0.4. Perfect congruence is present when Kappa value equals 1.0. Note that, all calculations in Kappa are based on cell, no landslide triggering or transporting mechanisms are considered.

2.3 Modified success rate (MSR)

Here we introduce Type 4 outcome, the success of stable cell prediction, into the former SR calculation. The SR and performance of stable cell prediction are equally weighted. The MSR is defined by the following equation:

$$\text{MSR}=0.5 \cdot \text{SR} + 0.5 \cdot \frac{\text{successfully predicted stable cells}}{\text{total number of actual stable cells}}. \quad (3)$$

The performance value derived by MSR ranges from 0.0$\sim$1.0. The incorporation of Type 4 outcome promotes the role of stable area prediction in measuring model performance, and thus substantially reduces the potential of landslide over-prediction. The weighting factor of 0.5 is assigned according to results from stochastic test and real case application (see below). In MSR, the first component (SR) is still calculated based on landslide number rather than cell number since the conventional concept of SR contains landslide triggering mechanism, which is worthwhile keeping.

3 Stochastic analyses

3.1 Generate artificial landslide maps and simulations

Stochastic analysis is applied to examine the differences among these three methods on evaluating model performance. We generate 3 artifical landslide maps in a 20$\times$20 matrix. The three maps have landslide coverage approximately 5.0%, 10.0%, and 15.0%, respectively, of the whole matrix. The area percentage is limited to 15% since landslide area rarely occupied >15% of the total area in most natural watersheds (Carrara et al., 1995). Based on the three maps, we further assign three degrees of cell aggregation to each of them (see Table 1; the degree of cell aggregation is presented quantitatively by the ratio of total number of landslide cells to total number of landslide sites). Thus, a total of 9 artificial landslide maps is created. The three maps with different degrees of cell aggregation at a fixed landslide cover are used

to test cell aggregation effects. This effect needs to be examined since cell aggregation might lead to changes in landslide numbers, which is the basis of the former SR and our MSR methods. Features of the nine artificial landslide maps are summarized in Table 1. Those maps serve as actual landslide maps for model performance calculation.

To examine the response of estimators in all possible simulations (i.e. landslide susceptibility map), we generate artificial simulations based on given actual landslide maps. Based on each landslide map, we utilize a generator with dual-parameter, "a" and "b", to create susceptibility maps with a full range of possible simulations in both stable and unstable cell predictions. The two parameters "a" and "b" represent the success rate (from 0.0∼1.0) of stable and unstable cell prediction, respectively. The interval for both "a" and "b" are set to be 0.05; thus, 441 susceptibility maps upholding different degree of model performance are generated for each of the 9 original landslide maps. For instance, a totally failed landslide susceptibility map (i.e., all stable cells are predicted to be unstable and all unstable cells are predicted to be stable) would be generated by a parameter set of (0.0, 0.0). On the contrary, a perfect match case will be derived by a parameter set of (1.0, 1.0), whereas (0.0, 1.0) represents a map full of unstable cells; therefore this map totally fails in stable cell prediction but perfectly predicts unstable cells. Those artificial simulations are compared with their respective artificial landslide map and performances are measured separately by the three methods, thus, responses of each method to simulations with different performance can be revealed.

3.2 The response of estimators to simulation results

Model performances derived by the three methods are presented in contour patterns (i.e. response surface; Fig. 1). In each case, the performance value (Z) against success rate of stable (X) and unstable (Y) cell prediction are plotted. The x-axis is defined as the ratio of the total number of successfully predicted stable cells to total number of actual stable cells, and the y-axis as the ratio of the number of successfully predicted unstable cells to the total number of unstable cells (in fact, the values X and Y are, respectively, the parameter "a" and "b" in our generator). (Note that definition of y-axis is not the same as that of SR, which is calculated in units of landslide number.) The interpolation and contour pattern are obtained by using Kriging. The contour pattern serves as a response surface and enables us to evaluate the performance distribution with respect to prediction errors along both axes.

Cell aggregation shows insignificant effects on contour patterns. By contrast, distinctive contour patterns among methods are found (Figs. 1a, d and g). Thus, we only present contour patterns at middle level (Pattern 2 in Table 1) of cell aggregation respective to different landslide coverage (Fig. 1).

Among those contours, isopleths show curled feature for SR (Figs. 1a, b, c) and MSR (Figs. 1g, h, i) methods. The curves mainly result from different methods of Z calculation, which is in the unit of landslide number instead of cell number. By contrast, contours for the Kappa index show a non-curled feature, since Kappa's calculation is based on cell unit (slightly curled due to interpolation). Among the three methods, only isopleths by Kappa (Figs. 1d, e, f) can be derived from analytical solution. (Note: no analytical solutions can be obtained for SR and MSR expect under the condition of no cell aggregation, which means that each landslide is composed of only one single cell.)

Table 1. Basic information of artificial landslide maps in a 20×20 matrix.

Landslide coverage	∼5.0%	∼10.0%	∼15.0%
Pattern 1	23/16	41/21	62/25
Pattern 2	21/15	42/32	58/28
Pattern 3	17/15	38/29	59/34

* Numerator and denominator represent total number of landslide cells and total number of landslide sites, respectively. The degree of cell aggregation: Pattern 1>Pattern 2>Pattern 3.

For the SR method (Figs. 1a, b and c), the contour lines distribute horizontally with the Z values increasing as Y increases regardless of changes along the x-axis. The former success rate is obviously insensitive to errors in predicting stable cells that are equally important in landslide models. Thus, modelers could obtain a very high performance value even with the worst performance in stable cell prediction (i.e., X=0).

On the other hand, the Kappa index contour shows diagonal pattern implying its sensitivity to both axes. Zero isopleth is the result of equivalence between actual agreement and expected agreement connecting the coordinate (0.0, 1.0) and (1.0, 0.0) diagonally. Apparently, the Kappa index will give performance value of zero when modelers have a complete success on only one single side prediction. Higher performance values (Z) appear at the very upper right corner approaching (1.0, 1.0) where only near-perfect success at both axes would be located. Around 90% Z values in all Kappa contours are lower than 0.4 (grayish band, 0.4 to 0.7, marks the moderate performance values in Fig. 1). Meanwhile, as the size of landslide coverage changes the tilted isopleths shift implying that Kappa-derived performance values are affected by the landslide coverage in the map. The degree of tilting is determined by the relative proportion of unstable/stable cell number in the map. Such a systematic shift in performance value precludes its across-watershed and/or inter-event applications since inconsistent model performance will be derived at fixed success rates (see the reference point in Figs. 1d, e and f).

Contour patterns derived from MSR (Figs. 1g, h and i) are similar to the results from Kappa, in which the higher

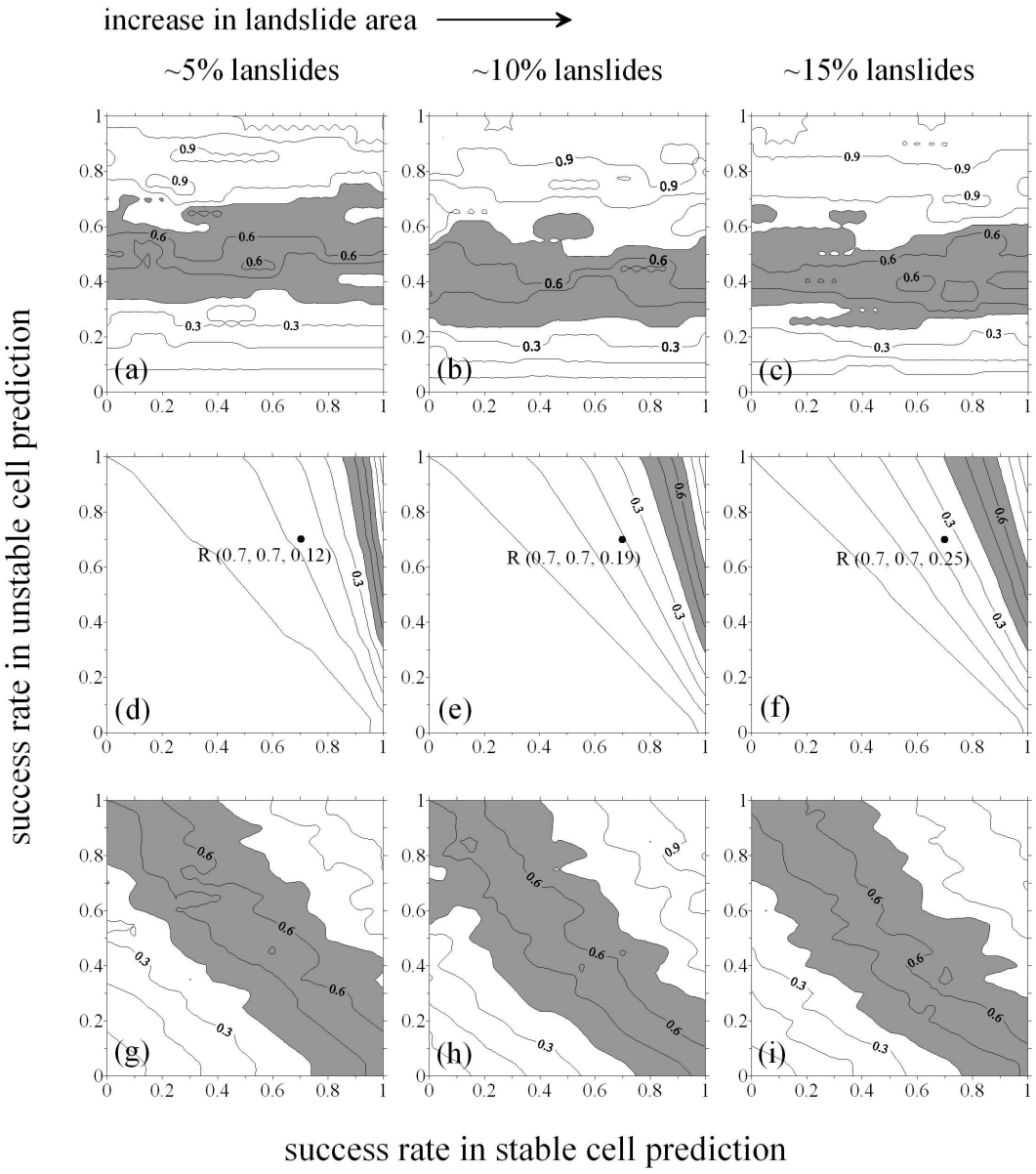

Fig. 1. The contour plots of model performance values derived from SR **(a, b, c)**, Kappa **(d, e, f)** and MSR **(g, h, i)** methods. Landslide coverage increases rightward from 5% to 15%. Bands of moderate performance, 0.4–0.7, are shadowed for reference. A reference point with fixed coordinate is given in (d), (e) and (f) (see text).

performance values occur at the upper right corner yet with smoother gradient toward (1.0, 1.0). Compared with Kappa, the contour pattern of MSR is consistent throughout different sizes of landslide coverage, therefore, provides a consistent performance value at fixed success rates. The moderate performance distributes diagonally from (0.0, 1.0) to (1.0, 0.0), which differs from that of SR and Kappa in absolute value and distribution pattern. Performance value of 0.5 is given in MSR even if a case that totally fails in one side prediction (i.e. in X = success rate of stable cells) yet has a complete success on the other side (i.e. in Y = success rate of unstable cells). Compared to the former SR, the MSR is sensitive to both axes; potentially, it is useful to avoid over-prediction of unstable cells (see the case study presented below).

Through such stochastic analyses, the response of each estimator to full-range model simulations is revealed. Previous studies (Carrara et al., 1995; Borga et al., 1998) have suggested that two extra errors should be included in assessing the accuracy of slope stability models: 1) landslide does not occur at a predicted unstable site (Type 3); and 2) slope failure takes place at predicted stable sites according to complex triggering processes (Type 2). According to our

stochastic analyses, Kappa, which considers all outcomes in calculation as suggested by previous studies, is supposedly the best performance indicator. Yet Kappa's calculation is based on cell instead of landslide site. Therefore, the absolute value by Kappa may shift systematically as landslide coverage changes. Kappa, apparently, is not suitable for landslide modeling. However, the applicability and advantage of MSR in real landslide model evaluation need further tests. Below is a case study to examine the correlation between model predictions and performance values obtained by SR and MSR.

4 A case study application

In this section, we apply SHALSTAB model onto Chi-Jia-Wan, a mountainous watershed in Taiwan (Fig. 4). A good landslide database and sufficient geology and vegetation information is available for this watershed. Since the focus is on the differences in two model-performance methods, the model details and parameter optimization processes are only briefly discussed and can be found in Huang et al. (2006).

4.1 SHALSTAB model and operation

The model used is that proposed by Mongomery 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. (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). The governing equation in the SHALSTAB is:

$$\mathrm{FS} = \frac{C + (1 - \frac{R}{T}\frac{a}{\sin\theta} \cdot \frac{\rho_w}{\rho_s}) g \cos^2\theta \cdot \tan\phi}{\rho_s g Z \sin\theta \cdot \cos\theta} \quad , \tag{4}$$

where FS is the factor of safety. Landslide occurs when FS<1. Thus in all simulated landslide susceptibility maps; unstable cell is defined as FS<1. The term, ρ_w / ρ_s, is the density ratio of water to soil, Z is the soil depth (L), θ is the slope gradient, a represents the specific contributing area (L), and $\tan\phi$ is the internal friction angle of the slope material. C is the effective cohesion (kpa), a combination of soil and root cohesion. R is the rainfall intensity (L/T), and T is the soil transmissivity (L^2/T). The term, $\frac{R}{T} \cdot \frac{a}{\sin\theta}$, is the soil wetness related to pore water pressure.

We set the density ratio (water to soil ratio) to be 0.4 and the soil depth (Z) to be 1.5 m according to field observations reported by Cheng (2003). Variables a and θ are determined based on DEMs. Thereby, only three process-related parameters (C, R/T, and ϕ) remain unknown. Practical steps for calibrating parameters often start with random combinations of parameters (e.g. Duan and Grant, 2000; Zhou et al., 2003). Model performance estimators are used to evaluate model outcomes, thus, acting as sorters to retrieve optimal simulations.

In this case study, we fix a reasonable range for the three parameter values (C, R/T, and ϕ) in the SHALSTAB (Table 2) and use random number generator to create 4000 parameter combinations under the assumption of uniform probability distribution for all three parameters; thus, 4000 landslide susceptibility maps with full range parameter combinations are generated. The spatial pattern of C is derived from satellite image (NDVI values from SPOT imagery). The transfer function is shown in Table 2 and detailed in Huang et al. (2006). The internal friction angles (ϕ) range from 30 to 45 degree according to the GIS and geological datasets. Since R/T fluctuates over orders-of-magnitude through time, the R/T ratio in the entire watershed is the randomly selected R (range 1–20 mm/h) divided by randomly selected T (range 0.001–10 m^2/h) to create a wide spectrum of the hydrological term (R/T ratio). The inter-correlations among the three parameters are discussed elsewhere (Huang et al., 2006).

The 4000 landslide susceptibility maps are validated by comparing with the actual landslide map taken from the database in the Industrial Technology Research Institute in Taiwan (Industrial Technology Research Institute, 1998). The SR and MSR are applied separately to measure the performance of individual simulations.

4.2 Balance between over- and under-prediction

Using the response surface mentioned earlier in Sect. 3, we plot the success rate of unstable cell prediction against success rate of stable cell prediction for all 4000 simulation cases in Fig. 2. The data points distribute over a wide range in both axes indicating a variety of results produced by the SHALSTAB model. The perfect simulation which means the exact match between predicted map and actual landslide map is referred to the point (1.0, 1.0) in Fig. 1. However, the scatter plot reveals an inverse relationship that the success rate of stable cell prediction decreases as the success rate in unstable cell prediction (y-axis) increases. Such an inverse correlation is attributed to the inter-exclusive feature between stable and unstable cell predictions (i.e. an improvement in one prediction results in deterioration in the other). Apparently, if we only track the success of landslide prediction, we may obtain biased model outcomes; which means the stable cell prediction should be properly weighted.

Theoretically, the most successful simulation should be characterized as that with the most number of actual landslides predicted with the least amount of area been predicted to be unstable (Casadei et al., 2003). Over-prediction occurs when extra stable cells are modeled as unstable cells to promote landslide prediction. However, under-prediction may also occur in the opposite situation. In other words, the optimal simulation should meet the balance between over- and under- prediction. In fact, too many over-predicted unstable cells in stable area imply that the model does not adequately

Table 2. Model parameters: the ranges and assumptions of probability distribution.

Parameter	Definitions	Range	Distribution
C (x, y)	The effective cohesion (in kpa). $C(x,y)=C_{\min}+C_{\text{interval}} \cdot \frac{\text{NDVI}(x,y)+1}{2}$	$C_{\min}$: 0.0∼20.0 C_{interval}: 0.0∼30.0	Uniform
R/T	A compound parameter of rainfall intensity and transmissivity. R (mm/h), and T (m^2/h)	R: 1.0∼20.0 T: 0.001∼10.0 R/T: 10^{-6}∼10^{1}	Uniform
ϕ	The internal friction angle (in degree).	30∼45	Uniform

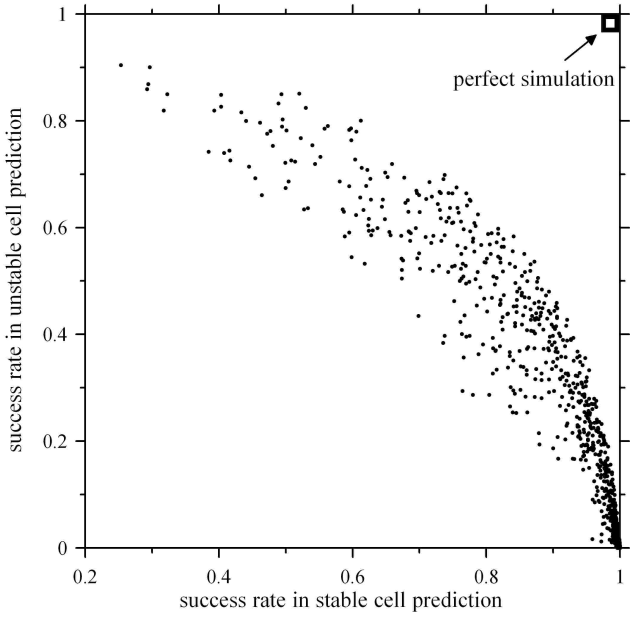

Fig. 2. Simulation results for 4000 landslide simulations in the Chi-Jia-Wan watershed in Taiwan. X-axis is success rate of stable cell prediction. Y-axis is success rate of unstable cell prediction.

grasp landslide mechanism in the specific environmental setting. Casadei et al. (2003) has emphasized that one should avoid over-predicting unstable cells in using the slope stability model. Stand on this point; the stable cell prediction definitely needs to be included while measuring the landslide model performance.

When measuring the model performance, reaching the balance (i.e. determining the weighting) between the number of stable cells and landslide site predictions is the next question. Experiences in field observation shed light on the solution. Typically, the percentage of landslide coverage rarely occupied >15% in most natural watersheds (Carrara et al., 1995). In Taiwan, the landslide coverage is rarely >10%. In most watersheds landslide covers are <5%. This indicates a proper simulation should give >80% of stable cell coverage in total watershed area. If it is below this threshold, over-prediction is likely to occur.

Here we plot SR- and MSR- derived performance values against the coverage percentage of predicted stable area for those 4000 runs (Fig. 3a) to evaluate if MSR may properly detect over-prediction. In Fig. 3a, the SR shows a continuous increasing trend as the coverage percentage of stable area decreases. Yet, MSR-derived performance values exhibit a dome-shape, of which both low and high stable cell coverage make low performance values. The positive trend at left where stable cell coverage <80% (landslide over-prediction likely occurs) results from a progressive prediction in stable cell. Such progression inhibits the unstable cell prediction. On the other hand, the MSR-derived performance values start to decrease as the stable cell coverage keeps going higher than 90%, where landslide under-prediction likely occurs. The best simulation derived by MSR would appear around the crest of the dome, where the proper coverage (80–90%) for stable cell prediction is located (Fig. 3a). This dome-shape pattern indicates that MSR has discrimination power to detect over- and under-prediction.

On the other hand, we assign three pairs of values, A(0.3, 0.7), B(0.5, 0.5) and C(0.7, 0.3), as weighting factors to the two components in MSR (i.e., former SR and success rate of stable cell prediction, respectively) to examine whether the pair of (0.5, 0.5) is the optimal weighting factors in MSR formulation (Eq. 3). Three distinctive patterns are revealed in Fig. 3b due to the various pairs of weighing factors. For those simulations with stable cell coverage >90% (right dashed line in Fig. 3b), the A-pair-MSR gives the highest performance values and the C-pair-MSR gives the lowest values among the three. For those simulations with stable cell coverage <80%, A-pair-MSR gives much lower performance values while C-pair-MSR provides consistently high performance values. Apparently, A-pair-MSR is much sensitive to over-prediction but not sensitive to under-prediction. On the contrary, C-pair-MSR is sensitive to under-prediction, yet shows no discrimination power for simulations with over-prediction. The performance value derived by B-pair-MSR (black triangle in Fig. 3b) is likely the optimal estimator among the three weighting pairs, providing sensitivity for both stable and unstable cell prediction.

For further demonstration, three susceptibility maps are presented in Fig. 4 (marked by (I), (II), (III) in Fig. 3b).

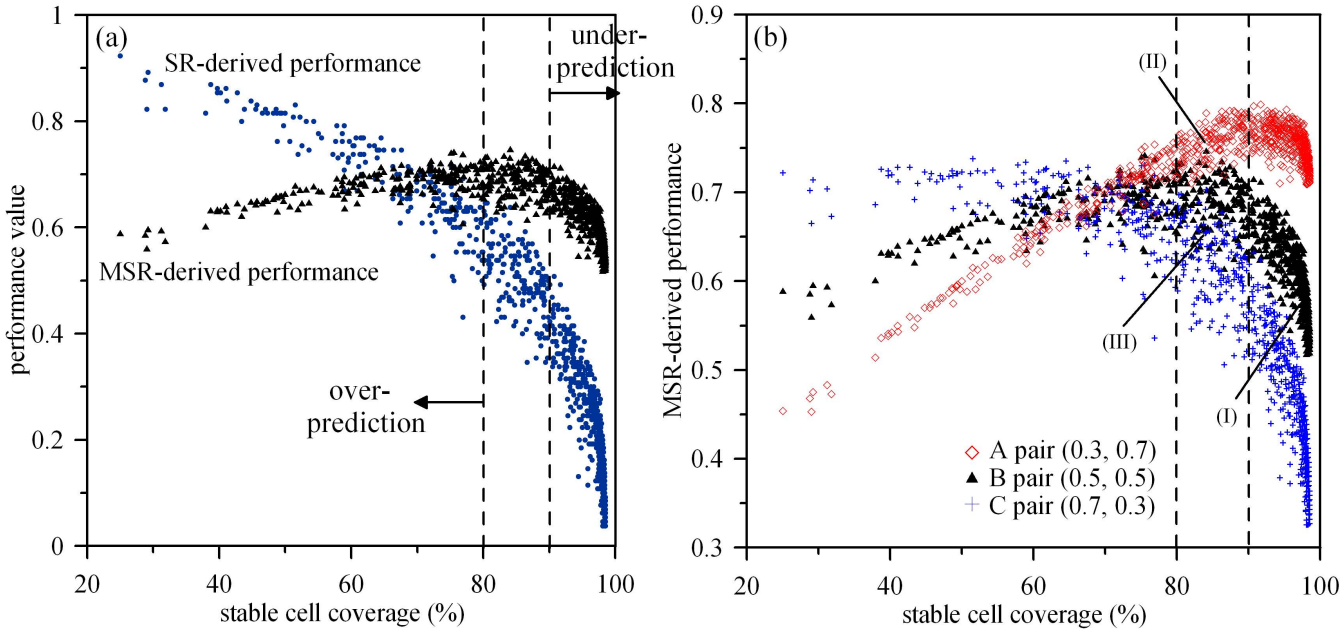

Fig. 3. **(a)** Scatter plot of 4000 model simulations. SR- (○) and MSR-derived (▲) performance values are plotted against the percentage of stable cell coverage. **(b)** Scatter plot for A-pair-(◇), B-pair-(▲) and C-pair-MSR (+) (see text) derived performances. Three examples marked by (I), (II) and (III) are illustrated in Fig. 4. Vertical dashed lines mark 80% and 90% of stable cell coverage.

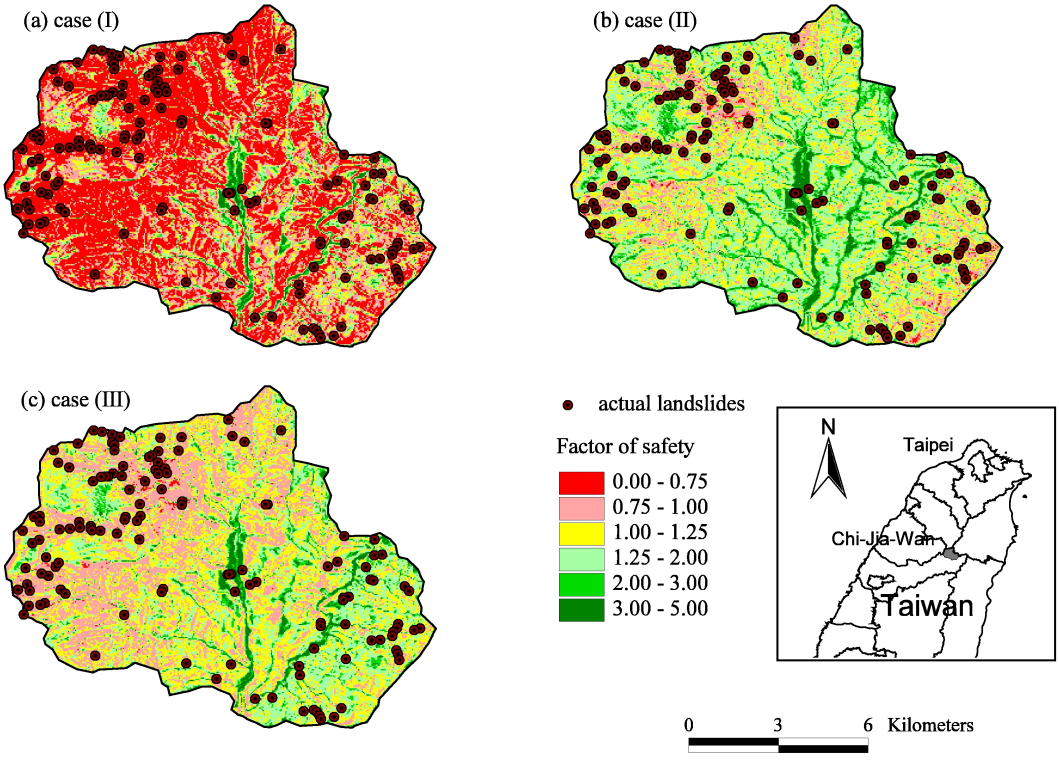

Fig. 4. Landslide susceptibility maps of the three simulation cases labeled in Fig. 3b. Location map of the watershed is shown at the lower right. Dots stand for actual landslide sites. Color scale for FS values is shown. FS<1 is classified as landslide in SHALSTAB. The landslide coverage, SR-derived performance, MSR-derived performance and landslide hits are listed in Table 3 for comparison.

Table 3. Performance values and related information of the three cases in Fig. 3b.

	SR	MSR	unstable cell coverage (%)	Landslide hit[#]
(I)	0.92	0.59	75	120/131
(II)	0.64	0.75	16	83/131
(III)	0.64	0.67	31	83/131

#: total number of successfully predicted landslides/total number of actual landslide site

Related information for the three susceptibility maps is listed in Table 3. The performance value given by SR for Case I, II and III are 0.92, 0.64 and 0.64, respectively, while MSR gives performance values of 0.59, 0.75 and 0.67, respectively.

In case I, the simulation hits 120 over 131 actual landslide sites (see Table 3), thus SR-derived performance is as high as 0.92. In spite of its high performance, the coverage of predicted unstable area is as high as 75% of the total watershed area indicating a significant over-prediction (Fig. 4a). By contrast, MSR gives this simulation a much lower performance of 0.59. In case II, 83 over 131 actual landslide sites are successfully predicted with predicted unstable area covering only 16% of the total watershed area (Fig. 4b). In this case, MSR-derived performance (0.75) is higher than the SR-derived performance (0.64). While in case III (Fig. 4c), the same amount of landslide sites (83 over 131) is predicted when compared to case II, in this simulation the unstable area covers 31%, which is much higher than that of the case II. The MSR gives case III a performance value of 0.67, which is lower than that given in case II, whereas the old SR gives the same performance (0.64) for cases II and III. The case II and case III hold the same success rate in landslide site prediction; yet, the two cases differ from each other in terms of stable cell coverage. MSR can easily distinguish the two.

Both the stochastic analysis and the case study demonstrate that stable cell prediction must be considered in model evaluation and equally weighting MSR can efficiently avoid over/under prediction and enhance model calibration. Practical steps for calibrating parameters often start with random combinations of parameters (e.g. Duan and Grant, 2000; Zhou et al., 2003). With MSR serving as a likelihood measure in landslide modeling, we are able to retrieve the best simulation out of abundant model outcomes (Huang et al., 2006).

5 Conclusions

Performance measure is crucial in landslide modeling. The Kappa index is popular in image analysis but is not a proper estimator for landslide model performance evaluation. The former SR method does not consider the success of stable cell prediction; therefore, precludes the detection of over-prediction. Stochastic analyses and the case study in Chi-Jia Wan watershed demonstrates that only the proposed MSR method can compromise over- and under-prediction problems and to provide much reliable measure for model performance. Among the three estimators, the MSR is optimal; it responds to both stable and unstable cell errors, and therefore, may serve as likelihood measure in landslide modeling to, subsequently, retrieve the best simulation out of abundant model outcomes.

Acknowledgements. We acknowledge comments from M. L. Hsu at Natl. Taiwan Univ., Taiwan. The journal reviewers are gratefully acknowledged. Many thanks must be given to E. Foufoula-Georgiou for her critical comments and careful reviews to make this paper more complete.

Edited by: E. Foufoula

References

Ayalew, L. and Yamagishi, H.: The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan, Geomorphology, 65, 15–31, 2005.

Borga, M., Fontana, G. D., and Marchi, L.: Shallow landslide hazard assessment using a physically based model and digital elevation data, Environ. Geol., 32, 81–88, 1998.

Borga, M., Dalla Fontana, G., Gregoretti, C., and Marchi, L.: Assessment of shallow landsliding by using a physically based model of hillslope stability, Hydrol. Processes, 16, 2833–2851, 2002.

Carrara, A., Cardinali, M., Guzzetti, F., and Reichenbach, P.: GIS technology in mapping landslide hazard, in: Geographical Information Systems in Assessing Natural Hazards, edited by: Carrara, A. and Guzetti, F., Kluwer, Dordrecht, Netherlands, pp. 135–175, 1995.

Cheng, Y. L.: Study on Risk Analysis of Slopes Consideration Spatial Variability – A Case Study at the Lisan, Master Thesis, National Chunghsing University, 2003.

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 Surf. Processes Landf., 28, 925–950, 2003.

Dietrich, W. E., Reiss, R., Hsu, M. L., and Montgomery, D. R.: A process-based model for colluvial soil depth and shallow landsliding using digital elevation data, Hydrol. Processes, 9, 383–400, 1995.

Dietrich, W. E. and Montgomery, D. R.: SHALSTAB: A digital terrain model for mapping shallow landslide potential, http://socrates.berkeley.edu/~geomorph/shalstab, 1998.

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, pp. 311–329, 2000.

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., Kao S. J., Hsu, M. L., and Lin, J. C.: Stochastic procedure to extract and to integrate landslide susceptibility maps: An example of mountainous watershed in Taiwan, Nat. Hazards Earth Syst. Sci., 6, 803–815, 2006, http://www.nat-hazards-earth-syst-sci.net/6/803/2006/.

Industrial Technology Research Institute: The management of Dai-Jia Reservoir Watershed-the 4th technique report, The management committee of Dai-Jia Reservoir (in Chinese), 1998.

Lee, S.: Application and cross-validation of spatial logistic multiple regression for landslide susceptibility analysis, Geosciences, 9(1), 63–71, 2005.

Longley, P. A., Goodchild, M. F., and Rhind, D. W.: Geographic Information Systems and Sciences, New York, Wiley, 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.

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, Vancouver, British Columbia, 19–22 February 2001.

Schuster, R. L. and Krizek, R. J.: Landslide: analysis and control, Transportation Research Board Special Report, 176, 1–10, 1978.

Sidle, R. C., Pearce, A. J., and O'Loughlin, C. L.: Hillslope stability and landuse, Washington, DC, American Geophysical Union, Water Resour. Monogr., 11, 140 pp, 1985.

Viera, A. J. and Garrett, J. M.: Understanding interobserver agreement: the Kappa statistics, Fam. Med., 37(5), 360–363, 2005.

Zhou, Q. and Liu, X.: Error assessment of grid-based flow routing algorithms used in hydrological models, Int. J. Geogr. Inf. Sci., 16(8), 819–842, 2002.

Zhou, G., Esaki, T., Mitani, Y., Xie, M., and Mori, J.: Spatial probabilistic modeling of slope failure using an integrated GIS Monte Carlo simulation approach, Eng. Geol., 68, 373–386, 2003.