Hydrol. Earth Syst. Sci. Discuss., 3, S918–S925, 2006
www.hydrol-earth-syst-sci-discuss.net/3/S918/2006/

Hydrology and Earth System Sciences Discussions

HESSD

3, S918–S925, 2006

Interactive Comment

Interactive comment on “Optimal estimator for assessing landslide model efficiency” *by* J. C. Huang and S. J. Kao

J. C. Huang and S. J. Kao

Received and published: 16 September 2006

This paper presents a new index, named as the Modified Success Rate (MSR), to assess efficiency of landslide prediction. The authors claim that the proposed MSR avoids over-prediction of unstable cells over total actual landslide sites because it includes the prediction performance of stable cells too. The modification in the MSR is sought by adding an equally weighted efficiency of prediction of stable cells. Since any new index can have new scope and potential to be used in various ways, the proposed MSR has potential to attract interest of possible readers. The mathematics and results are quite simple but the paper presents it in a complex way. On the other hand, the paper does not contain essential details about the model. So, this paper is likely to lose satisfaction of its potential readers. Following are some of my concerns, which should be addressed before the paper is ready to publish.

General concerns - Motivation of the study is particularly focused into the weakness of SR, which avoids including stable cells in efficiency measurement. The paper does not describe why the stable cells have to be included in the landslide model prediction. A straight-forward understanding of landslide prediction efficiency is to take account of unstable cells, as measured by SR, whereas stable cells do not resemble to landslides. The paper is not clear on what extra advantage would result by forcing the new index to account the stable cells in estimation of landslide prediction.

Appreciate reviewer's comment. We have added much more statements emphasizing advantage of our new index. In the second paragraph of the Introduction of our revised version, we also give an example to illustrate an optimal index should be able to detect over-prediction (page 2, line 16, "... a simulation that perfectly predicts all shallow landslides but with 80% watershed area been predicted as unstable gives no discrimination power"). We also add a new figure (Figure 3) to highlight the inter-exclusive feature in success rates between stable and unstable cell predictions while we apply SHALSTAB model. In the first two paragraphs in Sec. 4.2, we addressed the situation of over- and under-prediction in a real case study and point out the necessity that both stable and unstable cell predictions should be considered in performance measure.

- Kappa index has been criticized repeatedly as a "stern". The outcome of Kappa is not surprising particularly in the test similar to this study. Kappa is not a good estimator where the portion of sample exhibiting actual agreement is mutually exclusive to the portion of sample exhibiting expected agreement.

We have eliminated "stern" throughout this paper. Now we just indicate its improperness while severing as an index for landslide model performance measure. The improperness of Kappa is discussed in Stochastic Analyses. We indicate that Kappa may provide inconsistent results as landslide cover changes. Such a systematic shift in performance value precludes its across-watershed and/or inter-event applications (page 8, line 16˜20).

HESSD

3, S918–S925, 2006

Interactive Comment

- Performance of model or efficiency of prediction is usually judged in relative terms. If a, b, and c are three cases of predictions such that they yield SR(a)$<$SR(b)$<$SR(c). If the same cases yield Kappa(a)$<$Kappa(b)$<$Kappa(c) or MSR(a)$<$MSR(b)$<$MSR(c), then all of these indices are apparently presenting the same result in different form. In fact, the paper contains this information but it is not presented well.

We follow reviewer's suggestion discussing them in sequence at the bottom of Sec. 4.2 (page 14, line 17).

Specific comments - The “unstable” and “stable” cells are not defined.

We define both at the beginning of Sec. 2 in this new version (page 3, line 8).

- Random distribution of landslides on the maps would be a crude way for evaluation of prediction efficiency because it does not represent the case that agrees with nature (1129-17).

That artificial landslide map is mainly for testing response of each method to simulations. We do not try to represent the nature; yet, a wide range of success rates for both stable and unstable cell predictions is suitable for testing response of estimator to model simulations. Meanwhile, no matter what technique is applied main results will be the same. To make the entire story complete, we applied a case study (i.e., real model framework with a natural landslide map from mountainous watershed) to illustrate the advantages of MSR in performance measure.

- It is not clear under what basis the MSR is assigned equal weights for prediction of stable and unstable cells. If one proposes unequal weights based on the proportion of stable cells to total cells or unstable cells, what would be response of the authors?

According to this comment, we assign different pairs of weighting factors in MSR and make a new figure (Figure 3b) in this version to illustrate how MSR is affected by weighting factors and to convince readers that equal weight is the optimal for MSR (page 14, line 1˜12).

HESSD

3, S918–S925, 2006

Interactive Comment

EGU

- 1129-line 23, what is “landslide aggregation”?

The term is incorrect. In this new version, we use “cell” aggregation instead of “landslide” aggregation.

- 1130-line 4-16 requires explanation and re-phrasing as it is hard to comprehend message from this portion. For example, model results, parameter a and b, spacing of parameters, etc. is suddenly introduced in the text with no backgrounds. The details about the model and the role of parameters are missing, which might frustrate readers.

We rewrite the paragraph and introduce more background information for model application (page 6, line 9).

- What is the “method-derived model efficiency? 1130-line 19.

We eliminate this confusing term “method-derived” in this version.

- Authors emphasize the use “C++ language”, which does not seem giving any specific advantage (1127-13).

It is just a statement for our method.

- The procedure used to generate landslide susceptibility map is very confusing. Adopting a randomized stochastic field might serve the purpose of testing SR versus MSR instead of using the generated maps. The paper fails to clarify about advantage of using the generated landslide susceptibility map. (1129-17)

As replied earlier to Dr. Foufoula, the success rate in unstable cell prediction (b) is not exactly the same as “success rate in landslide site prediction” since a landslide may contain more than one cells. The advantage of generated landslide susceptibility maps can illustrate the responses of the three methods in all possible predictions. From the responses of the generated susceptibility maps, the characteristics of the three methods can be uncovered.

- The scale of a cell used in the study is not mentioned. Use of different scale of the

HESSD
3, S918–S925, 2006
Interactive Comment

cell might pose challenge on the analysis and/or interpretation. (1129-17)

In stochastic analyses (line 1129-17 in original version), we create a 20*20 matrix and generate different sizes of unstable area for artificial landslide maps. The unit is cell, which is scale independent.

- It is not clear why authors have preferred using kriging method to interpolate contour patterns, which has resulted curled lines for no good reason. (1131-3)

In fact, the curled feature is only partially due to the interpolation method. The main cause for curled lines is that former SR is calculated based on unit of landslide number rather than the unit of cell (page 7, line 18).

- It is not clear why authors believe the MSR avoids over-prediction of unstable cells looking onto its less sensitivity on one axis. Can same logic be presented as the MSR avoids under-prediction of unstable cells too? (1131-20)

This comment is well taken. We put pretty much emphases and added a new figure (Figure 3b) on discussing the under- and over-prediction. Field experiences about landslide over-prediction is recognized and adopted as criteria to examine the optimal weighting factor for MSR. Results indicate that MSR can avoid both over- and under-prediction.

- 1132-6to18 and 19-28 is difficult to understand. I would suggest clarifying the parameters, calibration process and direction of calibration in detail.

The "model efficiency as guidance for model calibration" is designed originally to discuss trial and error process in the model application. However, this part is conceptually correct yet does not really exist in real model applications. According to this comment, we eliminated the part using stochastic analyses for discussion and added a new figure instead (Figure. 2) to discuss the calibration in real application.

- The authors' logic stating that linear response of the MSR to both stable and unstable cell errors lead to optimality of the estimator is questionable. This raises a question

HESSD

3, S918–S925, 2006

Interactive Comment

whether authors are looking for an optimal estimator or linearly responding estimator. A linearly responding estimator is not necessarily preserve optimality. (1133-10)

Reviewer is correct. This part of discussion regarding “correct direction” and linearly responding estimator has been removed and replaced by balance between under- and over- prediction (in Sec. 4.2).

- No description of SHALSTAB model is presented. Present brief description of the model, the specific methodology adopted in this study, similarities and differences with respect to original SHALSTAB etc. may be added for better clarity (1133-15).

More descriptions regarding SHALSTAB model are added (in Sec. 4.1), and we add one more reference (Huang et al. NHESS) that readers can refer to details.

- Technical details has been omitted in many parts of the paper, e.g. DEM resolution is missing; how spatial patterns of C is derived from GIS and satellite images is omitted; how and why R & T are chosen randomly are not mentioned; how GIS gave internal friction angle (or is it simply derived from geological data, if so, what are they?) is not described; and so on. Without having these details, the landslide modeling part of the analysis is hard to conceive.

As suggested, we address those details of landslide modeling and the data preparation in sec 4-1.

- It is unclear how the model and its parameters are being effective to test a real case of landslide versus generated landslide susceptibility maps using random parameters while comparing SR and MSR.

In stochastic analyses (Figure 1), we test the responses of indicators over full range performance. In the case study, all parameters are set as wide as possible. In this version, we added a new figure (Figure 2), in which we plotted the 4000 simulations for comparison. This new figure reveals the inter-exclusive behavior in stable and unstable cell predictions (page 12, line 10). The plot also exhibits wide ranges of performance

HESSD

3, S918–S925, 2006

Interactive Comment

for both stable and unstable cell predictions.

- A smaller range of MSR (0.5-0.75) as opposed to the wide range of SR (0.1-0.9) seems coming from the halved weight of SR and halved weight of its complement. If SR is small, it is likely to have smaller percentage coverage of unstable cells. The complement of SR, other half in the MSR, would then be larger yielding a large MSR value. This would bring the lower range of the MSR near 0.5. However, at the higher side of SR, the differences between the index of actual agreement (Ka) and the index of expected agreement (Ke) would play role to limit the higher side of MSR at around 0.75. This dynamic is easily understood, which authors have elaborated (1135-1to16). However, the paper fails to clarify on specific gain by knowing this dynamics of index in landslide prediction.

We added a new figure (Figure 2) to discuss the inter-exclusive feature in real landslide modeling case. A new section “balance between over- and under-prediction” was added to explain dynamics of MSR.

- It is not clear why authors claim in the conclusion that use of only SR may retrieve improper parameter combination. On what basis is it possible to claim that the use of MSR would retrieve proper parameter combination?

We thank the reviewer’s reminder. The term “parameter combination” has been removed in this version. As K. Beven suggested even we can retrieve the “optimal” model simulation by using some kind of index we can hardly prove the parameter combination in it is optimal due to the complexity caused by model structure error, spatial heterogeneity in parameters and parameter co-linearity. We added Huang et al. (2006) for reference, in which the inter-correlation of parameters and equifinality problem in model application.

- The notion of “reliable and sensitive measure in model efficiency” can be argued as a subjective judgment of authors in favor of their claim because the MSR is not investigated enough by testing its reliability and sensitivity.

HESSD

3, S918–S925, 2006

Interactive Comment

In our real case application, MSR is definitely more reliable and sensitive when compared to the former SR. Thus we keep the sentence “The case study in Chi-Jia Wan watershed demonstrates that MSR is able to overcome over-prediction problem and to provide much reliable and sensitive measure in model performance in comparison to the old SR method”.

- Explanations of figure 2c and 2f are not clear.

The part of discussion is removed.

- What do the “correct direction” and “wrong direction” refers in figure 3?

The part of discussion is removed.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 3, 1125, 2006.