Hydrol. Earth Syst. Sci. Discuss., 3, S934–S938, 2006
www.hydrol-earth-syst-sci-discuss.net/3/S934/2006/

Hydrology and Earth System Sciences Discussions

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 proposes to revisit a common index for evaluating landslide model efficiency with a new proposed index and presents a simulation study in which these indices are used as metrics in a model parameter optimization framework, for purpose of comparison. Specifically, the commonly used index called the success ratio SR, is the conditional probability of predicting a cell as unstable given that it is truly unstable SR = p (Ĺu|u). The new index, called modified SR or MSR, includes also the conditional probability of predicting a cell stable given that it is indeed stable. The proposed index is MSR = 1/2 [p (Ĺu|u) + p (Ĺs|s)] where the Ĺ indicates prediction (via a model). The paper spends several pages to convey the above message and in my opinion this is not only confusing the issue and effectiveness of the message, but also casts uncertainty in the subsequent presentation.

We apologize for the insufficient description of SR calculation that misleads Dr. Foufoula. The calculation of SR is based on landslide number rather than on cell. A successfully predicted landslide, which often contains more than one cell, is defined as >1 cell that is predicted as unstable within the respective landslide zone. Such inconsistent base in calculation makes curve contour lines in SR and MSR maps. Therefore we rewrote Sec. 2.1 to clarify SR's calculation (page 4, line 2).

A few comments follow. 1. In the simulation study, the authors create landslide maps by choosing the landslide to cover randomly 5, 10, and 15% of the total area - however, we know that there is some coherence in the landslide coverage and ignoring this coherence or spatial dependence is bound to affect the results. Then, a large number of susceptibility maps are generated for each artificial landslide map. Two parameters control the generation of these maps: success in stable and unstable cell prediction, that is p (Ĺu|u) and p (Ĺs|s). My understanding is that what is explicitly prescribed in the artificial susceptibility maps is nothing but the two indices, SR and MSR, we seek to evaluate (see the formulae above). Thus, the performance of each map is already prespecified and it is not surprising that the two performance indices give different results. In fact, no simulation is even needed to determine these performances. For example, note in Fig. 1 that the (g), (h) and (i) plots are nothing than α = p (Ĺu|u) vs. β = p (Ĺs|s) and it is expected that MSR = 1/2 (α+ β) will be around the 1:1 line with a band around it that depends on the unconditional number of landslides. At the same time, the top figures that give only SR are expected to be horizontal bands since SR is not dependent on α = (Ĺs|s) which is the parameter on the horizontal axis.

As replied above, this question is raised due to our insufficient description. However, reviewer is correct under certain condition. Analytical solutions can be obtained for SR and MSR under the condition of no cell aggregation, which means that each landslide is composed of only one single cell. This condition does not exist in most natural watersheds and not in our study.

Firstly, We calculated the success rate separately by using the conventional SR and

HESSD

3, S934–S938, 2006

EGU

the unit of cell in the 40000 simulations in our case study. The two calculations are not the same and most simulations have higher values if calculated basing on SR. Cell aggregation accounts for this scatter pattern.

Secondly, we also calculated the success rate separately by using the conventional SR and the unit of cell in stochastic analyses. Under such assignment (landslide cover from 5 to 15% and low degree of cell aggregation) the patterns are much like that in our case study mentioned above. The positive correlations between site-based and cell-based calculation appear. Such positive correlation indicates we may obtain similar results if we just calculate landslide performance all based on cell. However, the conventional SR, also the first component of MSR, contains landslide triggering mechanism, which is worthwhile keeping. Moreover, under the extreme case when cell aggregation is very strong the two calculations will give distinctive results.

2. I do not see the point in presenting the Kappa index and showing how bad it does.

We thank the reviewer's reminder. We added two sentences, "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 Fig. 1d, 1e and 1f)." to interpret the improperness of Kappa in assessing landslide performance (page 8, line 19). For example, the landslide coverage generally increases with increasing rainfall (driving force). Apparently, kappa can't hold the consistent model performances for two different events in one watershed.

3. The other contribution of this paper is proposing this modified index for adoptive model calibration (i.e., calibrate model parameters based on past observed cases and model performance optimization as judged by these metrics). I have some reservations with the process (pg. 10) of randomly selecting parameter combinations as these parameters are definitely dependent on each other and this can significantly influence the results.

HESSD

3, S934–S938, 2006

The goal of this study is to seek a proper indicator for assessing landslide model performance. Thus we create wide parameter ranges for model testing. The inter-correlation among parameters definitely affects simulations as reviewer said. Thus the sampling size should be large enough. Nevertheless, to retrieve a better simulation is the first step for model calibration. Issues regarding retrieving parameter combination, inter-correlation among parameters, and equifinality problem in model study are presented and discussed in Huang et al. (2006).

4. Some other comments: Ţ The conclusion presented at the last sentence of the abstract seems obvious to me given the formulae of SR and MSR as given in the review above. That is, if the number of true stable cells is small then the extra term that the modified index adds has minimal effect and SR will compare well with MSR, etc. Yet the authors present elaborate simulations to arrive at such conclusions?

We are not clear why Dr. Foufoula asked this question. However, we modified these sentences in the abstract. The last two sentences are now "The best simulation generated by MSR projects 83 hits over 131 actual landslide sites while the unstable cells cover only 16

Ţ The precise number of 3969 landslide maps, etc. escapes me.

We rewrite the paragraph about artificial landslide maps and this number is removed.

Ţ Pages 3 and 4 - confusing discussion of very simple concepts.

We modified Sec. 2-1 to present the concepts clear in the revised version.

Ţ The test on effect of cell aggregation can be interesting but not in the limited way the authors perform it. Unless there is some significant coherence in the landslides (which is true in reality) aggregation effects will be minimal. Given that randomness has been used in the simulations, I believe that the results on aggregation can be misleading (lower bounds, if anything, of reality).

Reviewer is correct. The cell aggregation effect is insignificant in our stochastic test as indicated. However, the cell aggregation test is needed since the performance calculation is based on unit of landslide number rather than on cell.

Overall, I think the paper has some merit (in introducing the prediction of stable cells in the evaluation process), but due to the limitations of the presentation and the simulation results, I would not recommend publication of this paper as-is.

The manuscript is thoroughly revised. Much more emphases are put on real case study and the advantage of introducing stable cell prediction in performance measure. The condition of under- and over-prediction in landslide modeling and proper weighting factors in the two components of MSR are also discussed. .

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