

## ***Interactive comment on “Derivation and evaluation of landslide triggering thresholds by a Monte Carlo approach” by D. J. Peres and A. Cancelliere***

**D. J. Peres and A. Cancelliere**

djperes@dica.unict.it

Received and published: 27 June 2014

Dear S.L. Gariano,

We would like to thank you for your interest in the manuscript (MS), and for your useful comment.

Concerning the ROC-based analysis, or perhaps better said "signal detection theory"-based analysis, we fully agree that one can get lost in the related wide literature, and naively think existing performance indexes as new, as occurred in our case for the " $\Delta$ " index, which has to be called more properly True Skill Statistic and denoted by TSS, and was introduced by Pierce in the far 1884. Future versions of the manuscript will account for these facts, as well as citing the other articles you suggested.

C2073

Regarding the other terminology issues, we agree only partially with your comments. First of all, for the PRE vs. the POFA case,  $PRE = \frac{TP}{TP+FP}$  which differs from the "Probability Of False Alarms" which is  $POFA = \frac{FP}{TP+FP}$  (there is a mistake in your comment, see Barnes et al, 2009). Hence the MS presents correct nomenclature of the  $\frac{FP}{TP+FP}$  statistic.

Regarding the other two ratios, we are far from being sure that we have done a thorough terminology research, but several names are present in literature for the two other statistics  $\frac{TP}{TP+FN}$  and  $\frac{FP}{FP+TN}$ . For the first, we have found: *sensitivity*, *true positive rate*, *hit rate*, *recall* and *probability of detection* (the one you suggested to use). The same thing occurs for the second statistic, for which we have: *false positive rate*, *1-specificity*, *false positive rate*, *false alarm rate* (the one to which corrigendum by Barnes et al., 2009 is referred to) and finally the *Probability of False Detection* (the one you suggested to use).

In our opinion, as also stated by Barnes et al. (2009), what it is really important is "that authors make explicit how their verification statistics are calculated in their manuscripts".

Speaking specifically in relation to our MS, there are two points which I would like to consider. Firstly, I have some doubts if using the term "probability" is fully correct. In fact one can speak of "probability" with reference to a "population" and not to a "sample" - as in our MS, even though it is large (1000 years of simulations). In this latter case, the term "frequency", intended as an "estimate of probability" is more appropriate. Hence, one may think at the use of the terms FOD "Frequency Of Detection" and FOFD "Frequency Of False Detection", as well as FOFA "Frequency of False Alarms", but maybe it is just not the case to introduce new names in this already confusing literature context.

Secondly, the use of the term "alarm" in the indexes would be somewhat inappropriate in our work, because we are not analysing a set of data coming from an actual

C2074

early warning system, where "false alarms" occurred (the term "false positives" is more appropriate).

Concluding, for future presentations of our work list all the names of the indexes that we have encountered in literature, excluding the ones that contain the word "alarm". The ones with the term "probability" will not be excluded, because they are widely used in literature, even though it would be perhaps more correct to use "frequency" instead of "probability".

---

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 11, 2759, 2014.

C2075