Nat. Hazard Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2017-328
Influence of uncertain identification of triggering rainfall on the assessment of landslide early warning thresholds, by David J. Peres, A. Cancelliere, R. Greco and T.A. Bogaard.

Reply to Referee #2

We thank the referee for reviewing our manuscript (MS). In the following we answer point by point to his constructive comments. Referee comments are in Times new roman (black) typesetting, our responses in Arial (blue) typesetting.

- *The authors*

GENERAL COMMENTS

In this manuscript the authors investigate the effects of uncertain knowledge of the timing of landslide occurrence on the definition of intensity duration rainfall thresholds. The study is based on synthetic rainfall data and virtual landslide events. Thresholds are defined using the True Skill Statistic as optimization criterion. The work is carried out for one ideal slope in the Peloritani Mountains in Sicily (IT). Overall the paper is well written, with a clear structure and objective. I believe it could benefit from some more elaborations on some of the aspects presented, mentioned here below. I recommend minor revisions before publication on the journal.

Thanks again to the referee for his comments, to which we reply in the "Specific Comments" section.

SPECIFIC COMMENTS

1 – On the line of what already mentioned by Anonymous Referee #1, the study is purely focused on one ideal slope and synthetic data. The authors could discuss how this might make the results transferable to a real situation, when regions are considered and heterogeneities come in to play. This with respect especially to the difference in the scale and the use of virtual landslides.

As we stated in the reply to referee #1, the use of synthetic data allows to isolate and test the effect of landslide triggering thresholds of single and controlled factors of uncertainty. When regions are considered, heterogeneities come in to play, which means additional sources of uncertainty in landslide threshold determination, which would make less clear the effects on the threshold of the source of uncertainty considered here. It is out of the scope of our MS to combine these two different sources of uncertainty. This will be more clearly stated in the revised paper, and discussed briefly.

2 – The authors should report the total number of landslides as well as of non-triggering events considered. While this probably changes with the different parameters for the definition of the events, it would be useful to give an idea of the "robustness" of the results, that is whether the change of just few events among different scenarios would affect or not the threshold. Although the TSS considers both triggering and non-triggering events, the less the triggering events the more their relative importance on the definition of the threshold.

Perhaps the information required by the referee is already shown in Table 2 of the MS: the number of landslides is 81 (115) and the number of non-triggering *rainfall* events is 19826 – 81 = 19745 (19711) for τ_M=0 (τ_M = 2.7 days). These numbers do not change when different scenarios and different parameters for the definition of the rainfall events (U_{min} and S_{min}) are applied. Hence the

effect on the TSS mentioned by the referee in not present, and does not affect the comparison of scenarios in terms of threshold determination and relative performances.

3 – The authors could elaborate more on how the threshold was defined, as the results are difficult to explain without this information. An example is the change going from the case shown in Figure 5a to 5b. The "two rainfall events shifted to a duration of 1 h" (line 18-19 page 6 in the text) cannot be responsible for the lowering of the threshold intercept or slope as they are not correctly captured by the threshold but are "missed". So either some other triggering events changed causing the decrease of the threshold or the threshold shouldn't have changed. All this is true unless the authors gave somehow weight also to the distance from the threshold. If being just below the threshold or well below the threshold makes a difference in the TSS, then yes those points could be responsible for the change and you should ignore this comment, but it would be helpful if the method would be explained.

We thank the referee for his suggestion to include more details on threshold determination. These will be added to the MS to better clarify how the TSS determines threshold position. However, in contrast to the referee's reasoning, Figure 5a and 5b differ for more than just the "two rainfall events shifted to a duration of 1 h" (line 18-19 page 6 in the text): the rainfall intensity and duration of generally *all* triggering events changes. Though these changes are relatively small, they still affect the position of the TSS-optimized thresholds. In other words, it is true that the TSS does not "weight the distance from the threshold", and so it is also true that only two points cannot be responsible for a significant change in threshold parameters and performances. It is rather the fact that *all the* triggering points in general change, though slightly. The figure below (Fig. R1) compares duration, depth and intensity of triggering events relative to the data in Fig. 5a ("no errors", RS0 hourly) and Fig. 5b ("with errors", RS1).

These details will be clarified in the revised MS (possibly with the addition of Fig R1).

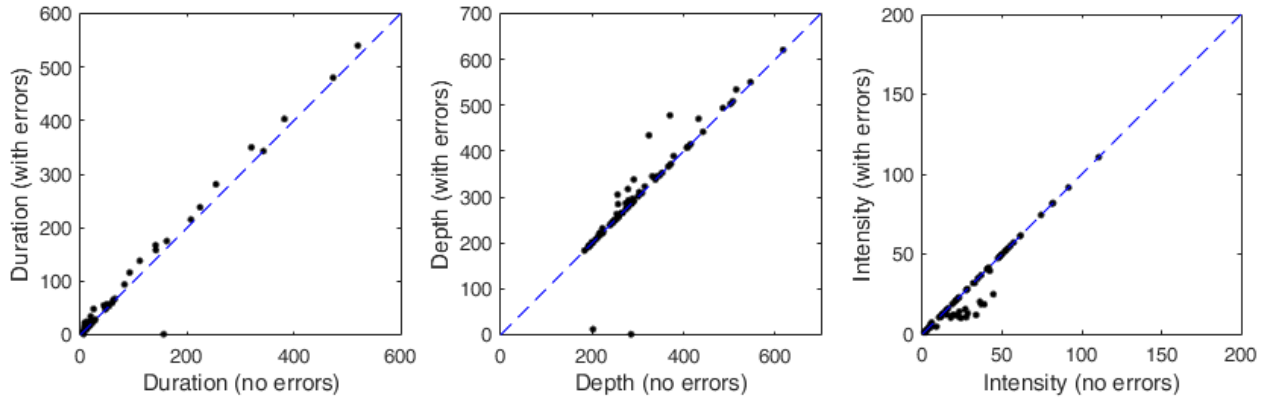

Fig R1- Comparison of triggering event characteristics for scenarios RS0 and RS1 in the case of hourly data and S_{min} = 0 and U_{min} = 24h (cf. Fig 5a and 5b of the MS)

4 – It seems that in general the points in the ID plane always move down (or left) in all the different scenarios. One would expect that sometime the landslides occur during intense rainfall storms and therefore including some extra hours actually could increase the intensity and duration.

We thank the referee for this comment, which will help to clarify some aspects of the obtained results. In fact, while, as a consequence of errors in the triggering instants, the rainfall event duration *T* may increase and the total rainfall *depth H* too, their ratio (rainfall intensity *I*) seldom increases. This is well known from rainfall extreme event analysis – the so-called intensity-duration-frequency (IDF) curves have always negative slope (see, for instance Bogaard and Greco, 2017): this is related to the fact that the higher the duration, the lower the mean rainfall intensity tends to be. Again, Fig R1 can be looked at as a confirmation of this behavior. Moreover, the few events that may have an

increased T and $I = H/T$, have a lower influence on threshold determination than the majority, which present decreased duration and intensity. This is not only because the events with increasing intensity are few, but also because the optimal threshold position is more sensitive to changes in the lower part of the cloud of triggering points (related to lower intensities), which partly mix up with the upper part of the non-triggering cloud. On the other side, the triggering points with increased intensity are usually not originally mixed up with the non-triggering cloud, and thus their change seldom determines a variation of maximum TSS.

These aspects will be shortly detailed in the revised manuscript.

Refs.

Bogaard, T., Greco, R., 2017. Invited perspectives. A hydrological look to precipitation intensity duration thresholds for landslide initiation: proposing hydro-meteorological thresholds. Nat. Hazards Earth Syst. Sci. Discuss. 1–17. https://doi.org/10.5194/nhess-2017-241

5 – The authors could explain better how the different scenarios are then used and corresponding triggering events selected. In fact, the scenarios are explained very well, but it is unclear how the events are then constructed. Is e_i randomly selected for each virtual landslide within the range defined for each scenario? Are then the results shown only one possible realization? Or is the wrong timing always fixed to Ta (that is always midnight, either 0, 24 or 48)? In other words, is the triggering event always the one happening at midnight or the last one that happened just before then? That wouldn't be a very realistic case because one would either try to find out at least whether it was morning or afternoon, or choose the most intense event within the day (which would then result in an overestimation of the threshold, but probably would still better than taking midnight rain) or choose the typical timing of landslides. Also for an available database, not for all entries timing or at least part of the day would be unknown (for the example you report in line40 page1 to line2 page2, only 27.7% of the cases would fall in this case, of only day know)

The following may serve as clarification in respect to the above referee comments.

Within the RS1-RS3 scenarios, we assume that the analyzer attributes the landslide to a day. The most conservative option is to do so by searching the rainfall event backwards from the end of the day (24h in RS1 or 48h in RS2), the least conservative is to do it from the beginning (0h in RS3). With our scenarios we consider a range of possibilities respect to which real scenarios (datasets) may represent intermediate cases. Our objective is not to analyze the complex subjective process that the analyzer may adopt in searching for triggering rainfall. Indeed, subjective criteria have been criticized by several researchers (e.g. Berti et al, 2013; Vessia et al, 2014; Melillo et al., 2015 – papers already in MS references) in favor of automatic procedures, which are more objective and thus more scientifically sound. Interestingly, in the paper by Berti et al. (2013), an automatic algorithm is calibrated based on decisions taken by a group of "expert analyzers". Thus automatic procedures can proxy "expert analyzer" behavior, with the added advantage of reproducibility.

In order to clarify the origin of errors e_i, perhaps it is useful to more explicitly specify the difference between the real triggering date t_i and the one at which the analyzer considers the landslide triggered t_i' (that generally differs from t_i, because of the limited information available). It is the latter that is discretized at midnights; the former is determined by rainfall time history and thus is random. Thus errors $e_i = t_i' - t_i$ are implicitly random.The ranges indicated within brackets are the maximum and minimum values of the errors in the given scenario.

Regarding the last part of the referee comment, line 40 page1 - line 2 page 2 reports the study of Peruccacci et al. (2017), which indicates errors that are always less than 1 day. As already

commented in the MS (P6 L17-18; P8L17-18) and discussed also in the reply to reviewer #1, our analyses show that errors of such amount do not affect significantly threshold determination and performances. Hence, other elaborations are not needed to simulate consequences of situations similar to those reported by Peruccacci et al. (2017). The study of Peruccacci et al. (2017) reports a relatively high precision of data, because the events are selected from a larger dataset covering a whole nation (Italy), *explicitly requiring* high accuracy. This will be specified in the revised MS. Especially when dealing with regions of smaller extension (as it is more usual), the data quality requirements can be less restrictive, to retain a significantly numerous dataset. Moreover, the referee should note that we cited also Guzzetti et al. (2008), which reports (for a global dataset) a way lower precision. They reported that the vast majority of events (68.2%) had no explicit information on the date or the time of occurrence of slope failure, while for most of the remaining events only the date of failure was known; more precise information was available just for 5.1% of the events. It is out of the scope of the paper to reproduce errors occurred in specific datasets used in landslide triggering threshold assessments performed by others. Our scenarios represent a range of possibilities, respect to which real datasets may likely represent intermediate cases.

The revised MS will include some sentences aimed at making more clear what discussed above.

Refs.

Peruccacci, S., Brunetti, M.T., Gariano, S.L., Melillo, M., Rossi, M., Guzzetti, F., 2017. Rainfall thresholds for possible landslide occurrence in Italy. Geomorphology 290, 39–57. https://doi.org/10.1016/j.geomorph.2017.03.031

Berti, M., Martina, M. L. V, Franceschini, S., Pignone, S., Simoni, A. and Pizziolo, M.: Probabilistic rainfall thresholds forlandslide occurrence using a Bayesian approach, J. Geophys. Res. Earth Surf., 117(4), 1–20, doi:10.1029/2012JF002367, 2012.

Vessia, G., Parise, M., Brunetti, M. T., Peruccacci, S., Rossi, M., Vennari, C. and Guzzetti, F.: Automated reconstruction of rainfall events responsible for shallow landslides, Nat. Hazards Earth Syst. Sci., 14(9), 2399–2408, doi:10.5194/nhess-14-2399-2014, 2014.

Melillo, M., Brunetti, M. T., Peruccacci, S., Gariano, S. L. and Guzzetti, F.: An algorithm for the objective reconstruction ofrainfall events responsible for landslides, Landslides, 12(2), 311–320, doi:10.1007/s10346-014-0471-3, 2015.

Guzzetti, F., Peruccacci, S., Rossi, M., Stark, C.P., 2008. The rainfall intensity-duration control of shallow landslides and debris flows: An update. Landslides 5, 3–17. https://doi.org/10.1007/s10346-007-0112-1

6 – The case of the Italian rainfall dataset is presented in which precipitation for the day D is collected for the 24h preceding 9am of day D. Wouldn't one use this dataset by shifting it by one day? So that precipitation of day D is between 9am of day D and 9am of day D+1? Surely there will still be some error as it still wouldn't match with the day definition, but this would probably be more meaningful.

We agree with the referee on this point. By the case of the "Italian rainfall datasets" we show what are the consequences of being unaware of the aggregation shift. Of course, if the analyzer is aware of this artifact, he would try to exploit the dataset at best, i.e. by shifting the original data as mentioned by the referee. And indeed in the conclusion this is what we want to stress in (p8 lines 29-33: "when threshold are determined from daily data, the data analyst has to be aware of possible shifts/delays in the rainfall accumulation interval, that is, if precipitation reported for a given day is the total amount occurred in a shifted period"). When corrected as the referee suggests, one would obtain low impacts. Nevertheless, we believe that the issue of shifted rainfall amounts deserves to be explicitly discussed, as is done in our MS. This because, apart from few papers (only Caracciolo et al., 2017, to our knowledge), most of the papers focused on the determination of landslide triggering thresholds in Italy (for which this shift can be present), do not report any relative correction. From this we may infer that in a significant number of studies the analyzer was not aware of the shift, since it would have been otherwise mentioned. There is no need for doing additional elaborations, as the results would be quite similar to those obtained in Fig. 7c (cf. also answer to referee #1).
More detailed discussion on these issues will be added to the revised MS.