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 #1

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

The manuscript of Peres and co-authors entitled "Influence of uncertain identification of triggering rainfall on the assessment of landslide early warning thresholds" is an interesting well-structured and well-written manuscript that addresses a very important scientific question that is within the scope of NHESS. However, it needs some minor revisions prior to be published.

Thanks again for the comments. Please see the following point by point replies.

Specific Comments

1 - The exercise presented along the manuscript is based on synthetic data, which are easier to control and monitor. However, the exercise has the drawback of reporting a single ideal slope. So, there is also a matter of scale when we compare the obtained results with most rainfall thresholds reported in literature that were built to be applied and interpreted at the regional scale. May be this is not enough discussed along the manuscript.

As correctly stated by the reviewer, to refer to synthetic data allows to isolate factors of uncertainty to test their influence on a issue of interest – on ID thresholds in the case of our manuscript.
It is certainly true that mostly thresholds are determined by analyzing rainfall-landslide data from multiple locations within a region. This means that the properties of unstable slopes change from landslide to landslide. Clearly this heterogeneity impacts on the performances of regional thresholds. This is a problem of empirical thresholds, and an additional source of uncertainty. To analyze this source of uncertainty in combination with that related to uncertain knowledge of triggering rainfall events, is out of the scope of our MS, and may be the scope of further research. A comment on this will be added to the text. An outlook in the conclusions mentioning this issue will be added as well.

2 - Within the simulation of uncertainty in triggering instant and the reporting of the landslide, authors establish the 'Observer's day' as lasting from the 6pm of Day D-1 to 6pm of the Day D. The explanation of this option is not clear. Although the reporting of a landslide in newspapers is usually delayed in relation to the actual triggering instant, the information about the timing of triggering may be quite precise namely in those cases where landslide generated severe human and/or economic damages. Apparently, this was not considered in the definition of the 'Observer's day'.

The 'Observer's day' is assumed as lasting from the 6pm of Day *D*-1 to 6pm of the Day *D*. This is justified by normal working hours at day *D*, plus the fact that what happens before in the night is reported in newspapers (and similar sources) from the next morning. The choice of 6pm rather than another hour is quite arbitrary, but a different choice would not affect significantly our results. A small discussion on this will be added to the revised MS.

We agree with the referee that in real datasets there may be a portion of triggering instants known precisely. We preferred to do not consider "mixed scenarios" where small and big errors coexist in certain proportions. It may not be difficult to add those scenarios, but we believe that this would not add substantial changes to the conclusions of the manuscript, or even result in less clear findings. Mixed scenarios would produce impacts that are intermediate between two/three of the considered RS, depending on their percentage.

On this point we also refer to the reply to comment 5 of referee #2.

3 - Quite interesting, figures 6 a), 6 b) and 7 c) are very similar. Comparing figure 6a) and 6b) one can conclude that working at the daily scale the knowledge of exact timing of the landslide triggering is not essential, providing the reporting Day (D) is correct. In addition, when the daily rainfall depth is measured form 09:00 AM to 09:00AM it is clear that most of the rain that falls in the day D will be registered in the day D+1. Therefore, it is normal that threshold (c) corresponding to Scenario RS2 (Day D+1) in figure 7 is virtual similar to the Scenario RS1 (Day D) and RS0 (actual triggering instant) in figure 6. In the opinion of the reviewer, this topic should de discussed more in detail in the paper.

We agree with the reviewer about the comparison of Fig. 6a with Fig. 6b. Stronger comments will be added to the MS following the suggestion of the referee, though the point of the reviewer is already stated in the MS at two points: P6 L17-18; P8L17-18.
Relatively to comparison between Fig. 7c and Fig. 6b, we agree with the reviewer that there is a compensation of errors in this case, as already commented in the MS P7 L3-4. In the conclusions there it is also mentioned that this implicates that the analyzer should check if the original data are affected by this systematic error, and eventually compensate for it (P8 L29-30): "the data analyst has to be aware of possible shifts/delays in the rainfall accumulation interval"
However a more explicit suggestion for the "analyzer" to check and correct for this error will be added.

4 - Although this information is contained on Figures 8 and 9, the equations of thresholds could be provided in a summary table, allowing for a more easy comparison.

In the revised MS, Figures 8 and 9 will be replaced by Tables with the same information.

5 - When performing the exercise for the daily scale that is summarized in figure 6 and 7, a contradiction exists, between figures and text (page 6 line 35) on the assumed Smin. In figure caption it is referred Smin = 0 mm whereas in text is referred Smin = 5 mm.

The actual adopted value is S_{min} = 0 mm. This will be corrected in the revised MS

6 - In figure 10 authors present the "correct thresholds". However, it is not given the information on the considered Umin and Smin parameters.

The actual adopted value is S_{min} = 0.2 mm and U_{min} = 24 h for the correct thresholds determined from hourly data, and S_{min} = 0 mm and U_{min} = 1 day for those determined from daily data. This will be specified in the caption of the figure.

Technical corrections

In figure 2, the time scale should be respected. The position of 6pm in Day D and Day D-1 is not correctly scaled. Add the notation RS0 in figure 2.

This will be fixed for the revised MS

Figure 3 The aggregation of data within figure 3 should be clearer. Rain gauge D+1appear two times; why? The total amount of rain measured on calendar days and raingauge days is not the same. Authors should acknowledge this difference and explain why.

This will be fixed. Improved figure and a more detailed caption will appear in the revised MS

Table 3 Some rainfall event identification instead of Some event identification.

This will be fixed for the revised MS

Reference of the paper of Nikolopoulos et al needs to be corrected in reference list.

This will be fixed for the revised MS

Page 2. Line 26 Rodriguez-Iturbe et al., 1987a, 1987b instead of Rodriguez-Iturbe et al., 1987; Rodríguez-Iturbe et al., 1987. Introduced a and b in the reference list.

This will be fixed for the revised MS

Page 2, line 31 Baum and Godt, 2010, instead of Baum et al., 2010 ?

This will be fixed for the revised MS. Correct is Baum et al., 2010. Mistake is in reference (third author missing)

Page 3. Line 7 Schilirò et al., 2015a, 2015b, 2016; instead of Schilirò et al., 2015,2016; Schilirò et al., 2015;

This will be fixed for the revised MS

Page 3, line 38 Guzzetti et al 1997, 1998 are missing in reference list.

This will be fixed for the revised MS

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).

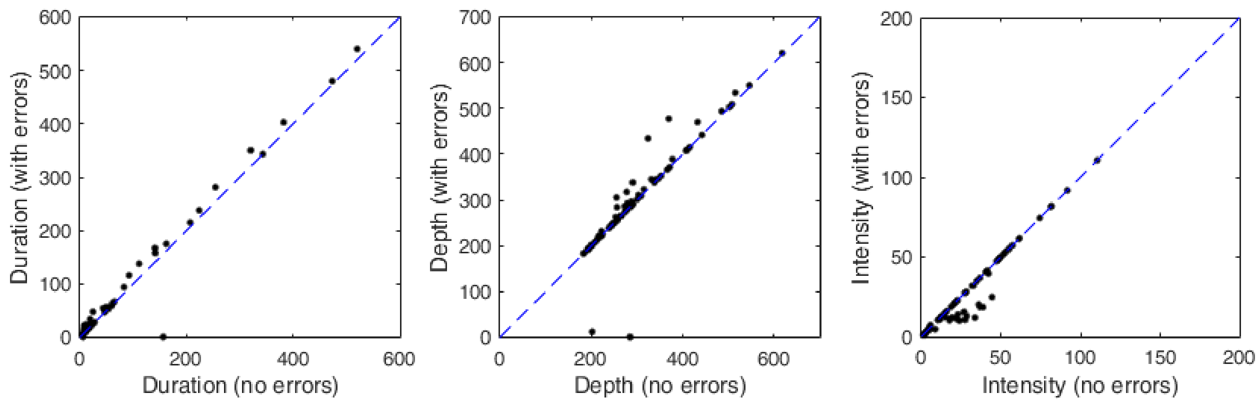


Fig R1- Comparison of triggering event characteristics for scenarios RS0 and RS1 in the case of hourly data and S_{min} = 0 and U_{min} = 24h (cf. Fig 5a and 5b of the MS)

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.

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.

List of modifications related to comments by Referee #1

Please notice that page and line numbers are those of the revised version of the MS

Referee comment	**Modifications**
General Comments The manuscript of Peres and co-authors entitled "Influence of uncertain identification of triggering rainfall on the assessment of landslide early warning thresholds" is an interesting well-structured and well-written manuscript that addresses a very important scientific question that is within the scope of NHESS. However, it needs some minor revisions prior to be published.	-
Specific Comments 1 - The exercise presented along the manuscript is based on synthetic data, which are easier to control and monitor. However, the exercise has the drawback of reporting a single ideal slope. So, there is also a matter of scale when we compare the obtained results with most rainfall thresholds reported in literature that were built to be applied and interpreted at the regional scale. May be this is not enough discussed along the manuscript.	P3 L11-14 The application to a hillslope of definite characteristics enables us to isolate the impact of the uncertainty in triggering rainfall identification; regional determination of thresholds do contain also factors of uncertainty related to the heterogeneity of landslide characteristics; the assessment of this combined uncertainty is out of the scope of our present analysis.
2 - Within the simulation of uncertainty in triggering instant and the reporting of the landslide, authors establish the 'Observer's day' as lasting from the 6pm of Day D-1 to 6pm of the Day D. The explanation of this option is not clear. Although the reporting of a landslide in newspapers is usually delayed in relation to the actual triggering instant, the information about the timing of triggering may be quite precise namely in those cases where landslide generated severe human and/or economic damages. Apparently, this was not considered in the definition of the 'Observer's day'.	P4 L 1-2 […] this choice is an attempt to resemble usual working hours, and the fact landslides occurring by night may be reported the morning after P4 L14-15 The two parameters, T_O and T_A, can be set to simulate a range of scenarios, for which real situations may represent intermediate cases P8 L 34- 38To this aim, we have investigated the effect of a set of hypothesized scenarios of landslide information retrieval and interpretation which can induce errors in the identification of instants of landslide occurrence. Moreover, we have analysed how the impact of reasonable scenarios may vary in dependence of rainfall aggregation

	(hourly or daily), and of rainfall event identification criteria. Real situations may be a mixture of the considered scenarios, and thus the impacts are presumably intermediate between the ones hypothesized.
3 - Quite interesting, figures 6 a), 6 b) and 7 c) are very similar. Comparing figure 6a) and 6b) one can conclude that working at the daily scale the knowledge of exact timing of the landslide triggering is not essential, providing the reporting Day (D) is correct. In addition, when the daily rainfall depth is measured form 09:00 AM to 09:00AM it is clear that most of the rain that falls in the day D will be registered in the day D+1. Therefore, it is normal that threshold (c) corresponding to Scenario RS2 (Day D+1) in figure 7 is virtual similar to the Scenario RS1 (Day D) and RS0 (actual triggering instant) in figure 6. In the opinion of the reviewer, this topic should de discussed more in detail in the paper.	P1 L 19-21 The analysis shows that the impacts of the above uncertainty sources can be significant, especially when errors exceed one day or the actual instants are after the erroneous ones. P 7 L 24-27 There is, however, the possibility that the errors due to rainfall aggregation and reporting landslide time interval compensate for each other, as in the case of scenario RS2 (delayed reporting of landslides), Fig. 7c (notice that this plot is similar to Fig. 6b). If the analyser is aware of the rainfall-aggregation shift, then he should correct as much as possible for this error – in this specific case, by shifting the entire daily rainfall dataset one day forward. P9 L13-14 From our analysis no significant impacts seem to be induced by the use of daily data; however, it is of fundamental importance to check, and correct where possible, for the presence of delays in the rainfall accumulation interval
4 - Although this information is contained on Figures 8 and 9, the equations of thresholds could be provided in a summary table, allowing for a more easy comparison.	Figures 8 and 9 of previous MS have been removed and replaced by Tables 6 and 7 showing the same information
5 - When performing the exercise for the daily scale that is summarized in figure 6 and 7, a contradiction exists, between figures and text (page 6 line 35) on the assumed Smin. In figure caption it is referred Smin = 0 mm whereas in text is referred Smin = 5 mm.	P7 L16 Figure 6 shows the results of calibration obtained with correctly-aggregated daily rainfall data and smin = 0 and umin = 1 day
6 - In figure 10 authors present the "correct thresholds". However, it is not given the information on the considered Umin and Smin parameters.	The missing information was added to Figure's caption (Fig.8 in the revised MS)
Technical corrections In figure 2, the time scale should be respected. The position of 6pm in Day D and Day D-1 is not correctly scaled. Add the notation RS0 in figure 2.	Figure 2 has been corrected as suggested

Figure 3 The aggregation of data within figure 3 should be clearer. Rain gauge D+1appear two times; why? The total amount of rain measured on calendar days and raingauge days is not the same. Authors should acknowledge this difference and explain why.	The figure has been improved and corrected. Caption has been integrated with more information: Figure 1: Aggregation of rainfall data from hourly to daily time scale: daily rainfall depths on the top row result from correct aggregation; those on the bottom row from shifted aggregation, as occurs for the Italian Hydrological Bulletins (Annali Idrologici). The shift is due to manual collection of data in early decades of operation of the monitoring network; the presence of the shift is still continued, in spite of installation of automatic rain gauges, to preserve homogeneity of the entire historical time series.
Table 3 Some rainfall event identification instead of Some event identification.	Fixed
Reference of the paper of Nikolopoulos et al needs to be corrected in reference list.	Fixed
Page 2. Line 26 Rodriguez-Iturbe et al., 1987a, 1987b instead of Rodriguez-Iturbe et al., 1987; Rodríguez-Iturbe et al., 1987. Introduced a and b in the reference list.	Fixed
Page 2, line 31 Baum and Godt, 2010, instead of Baum et al., 2010 ?	Fixed
Page 3. Line 7 Schilirò et al., 2015a, 2015b, 2016; instead of Schilirò et al., 2015,2016; Schilirò et al., 2015;	Fixed
Page 3, line 38 Guzzetti et al 1997, 1998 are missing in reference list.	Correct citation is Guzzetti et al 2007, 2008

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.

List of modifications related to comments by Referee #2

Please notice that page and line numbers are those of the revised version of the MS

Referee comment	Modifications
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.	
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.	P3 L11-14 The application to a hillslope of definite characteristics enables us to isolate the impact of the uncertainty in triggering rainfall identification; regional determination of thresholds do contain also factors of uncertainty related to the heterogeneity of landslide characteristics; the assessment of this combined uncertainty is out of the scope of our present analysis.
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.	P3 L27-28 Table 2 shows some characteristics of the 1000-year long synthetic databases, which do not change among the different scenarios illustrated in the following section. P6 L19-21 One advantage of the TSS is that it includes all the entries of the confusion matrix, and thus its maximization yields thresholds that result in a good trade-off between correct and wrong warnings/non-warnings.

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. 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.	P6 L34 –L40 The presence of small delay reporting errors (RS1), has little impacts on the position of triggering rainfall points (Fig. 5b), which in general are shifted slightly down along the intensity axis; this is related to the higher durations produced by positive errors in triggering instants, combined with an induced decrease of mean rainfall event intensities – a general behavior exhibited by extreme events (cf. the negative slope of well-known rainfall intensity-duration-frequency curves, see Bogaard and Greco, 2018). Only two rainfall events (the 2.5 % of triggering events) are highly-impacted, being moved to a duration of 1 hour. The latter and mainly the former effect, contribute to slightly flatten the threshold for TSS maximization (decrease of β το 0.7)
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	P4 L 8-9 These errors are implicitly random, since though t_i' are deterministically chosen, the actual instant t_i varies in an aleatory fashion according to rainfall time history. P5 L12-14 Automatic procedures have the advantage of being objective and reproducible, and thus more scientifically sound than subjective judgment (Melillo et al., 2015; Vessia et al., 2014); nevertheless, algorithms are suitable to reproduce the latter with a certain level of fidelity (Berti et al., 2012). P4 L19, L23, L28 "random in the range" has been added P2 L2-3. In their analysis, only information with an accuracy at least of one day was retained from the larger available dataset. Still for this trimmed dataset, triggering instants were available with high precision (minute or hour) only for the 37.3% of the data, being the day or part of it available for the majority (27.6% and 35.1%, respectively).

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)	P2 L19-25 We then fictitiously introduce errors in the triggering instants and in the rainfall series based on hypothetical scenarios of landslide data retrieval and analysis, and analyse the implications on the accuracy of ID thresholds. Quality of information available in real datasets is generally intermediate of that corresponding to the hypothesized scenarios. These scenarios are combined with different criteria for event rainfall identification, and different aggregations of rainfall data (hourly and daily, and daily in the presence of a shift due to manual collection of data), so the effects of these other two sources of uncertainty are analysed as well (items *i)* and *ii)* of the above list). P4 L 14-15 The two parameters, T_O and T_A, can be set to simulate a range of scenarios, respect to which real situations may represent intermediate cases P8 L37-38 Real situations may be a mixture of the considered scenarios, and thus the impacts are presumably intermediate between the ones hypothesized.
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.	P7 L 26-27 If the analyser is aware of the rainfall-aggregation shift, then he should correct as much as possible for this error – in this specific case, by shift the entire daily rainfall dataset one day forward. P9 L 13-14 From our analysis no significant impacts seem to be induced by the use of daily data; however, it is of fundamental importance to check, and correct where possible, for the presence of delays in the rainfall accumulation interval, that is if precipitation reported for a given day is the total amount occurred in a shifted period

Influence of uncertain identification of triggering rainfall on the assessment of landslide early warning thresholds

David J. Peres[1], Antonino Cancelliere[1], Roberto Greco[2], and Thom A. Bogaard[3]

[1]Department of Civil Engineering and Architecture, University of Catania, Catania, Italy
[2]Dipartimento di Ingegneria Civile Design Edilizia e Ambiente, Università degli Studi della Campania Luigi Vanvitelli, Aversa (CE), Italy
[3]Water Resources Section, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, the Netherlands

Correspondence to: D J. Peres (djperes@dica.unict.it)

Abstract

Uncertainty in rainfall datasets and landslide inventories is known to have negative impacts on the assessment of landslide-triggering thresholds. In this paper, we perform a quantitative analysis of the impacts ~~that the~~of uncertain knowledge of landslide initiation instants ~~have~~ on the assessment of ~~landslide~~rainfall intensity-duration landslide early warning thresholds. The analysis is based on ~~an ideal~~a synthetic database of rainfall and landslide ~~data~~information, generated by coupling a stochastic rainfall generator and a physically based hydrological and slope stability model~~.~~, and therefore error-free in terms of knowledge of triggering rainfall instants. This dataset is then perturbed according to hypothetical “reporting scenarios”, that allow to simulate possible errors in landslide triggering instants~~,~~ as ~~d~~retrieved from historical archives. The impact of these errors is analysed ~~by combining~~jointly using different criteria to single-out rainfall events from a continuous series and ~~different~~two typical temporal aggregations of rainfall (hourly and daily). The analysis shows that the impacts of the above uncertainty sources can be significant~~.~~, especially when errors exceed one day or the actual instants follow the erroneous ones. Errors ~~influence thresholds in a way that they are~~ generally lead to underestimated thresholds, i.e. lower than those that would be obtained from an error-free dataset. Potentially, the amount of the underestimation can be enough to induce an excessive number of false positives, hence limiting possible landslide mitigation benefits. Moreover, the uncertain knowledge of triggering rainfall, limits the possibility to set up links between thresholds and physio-geographical factors.

1. Introduction

Thresholds estimating rainfall conditions correlated to landslide occurrence are useful for landslide early warning systems (Guzzetti et al., 2007; Highland and Bobrowsky, 2008; Sidle and Ochiai, 2013). Commonly, thresholds are derived by empirical approaches based on the direct statistical analysis of historical rainfall series and landslide inventories, from which a line roughly separating triggering from non-triggering conditions is drawn. Among the various thresholds types, precipitation intensity~~ and ~~-duration power -law thresholds (hereafter referred to as ID thresholds), introduced by Caine (1980), have been derived for many regions of the Earth, and are still considered as a valid empirical model (Caracciolo et al., 2017; Gariano et al., 2015; Peruccacci et al., 2017; Vennari et al., 2014), though they ~~have~~are affected by several theoretical and practical limitations (Bogaard and Greco, 2018)~~.~~.

Thresholds derived for different geographical areas vary significantly, and some attempts have been made to find a rationale underlying this variability, by linking threshold parameters to physio-geographical and climatic features (Guzzetti et al., 2007, 2008). Nevertheless, rainfall and landslide data quality issues, reported in almost all of the papers on threshold determination, are known to potentially hamper the assessment of this link. As reported in many studies, the triggering instants available from real landslide inventories are imprecise. For instance, Guzzetti et al. (2007, 2008) reported that in a global database of 2626 landslides, the vast majority (68.2 %) had no explicit information on the date or the time of occurrence of slope failure;

for most of the remaining events only the date of failure was known, and more precise information was available just for 5.1% of landslides. These issues are confirmed with reference to an updated dataset of landslides occurred in Italy (Peruccacci et al., 2017). ~~In this case~~In their analysis, only information with an accuracy at least of one day was retained from the larger available dataset. Still for this trimmed dataset, triggering instants were available with high precision (minute or hour) only for the 37.3% of the data, being the day or part of it available for the majority (27.6% and 35.1%, respectively).

Other data art~~e~~ifacts include: *i*) rainfall measurement delays related to manual collection of data; *ii*) different criteria to identify rainfall events; *iii*) lack of completeness of landslide catalogues; *iv*) imprecise location of landslides, or precipitation measurements available at a significant distance apart from the location of failure. Though there is a general agreement that these factors affect the accuracy of rainfall thresholds, a quantification of the influence of these data quality issues on landslide triggering thresholds has been carried out in the literature only partially. In particular, to the authors knowledge, only the effect of rain gauge location and of the density of rainfall networks (point *iv*) has been analysed (Nikolopoulos et al., 2014)~~(Nikolopoulos et al 2014),~~, showing that the use of rainfall measured at some distance from debris flow location can lead to an underestimation of the triggering thresholds.

Quantitative assessments of the influence of the sources of errors listed above are difficult to be based on observational datasets, since it cannot be ensured that these are immune of errors. In this paper we capitalize on the synthetic rainfall-landslide data set of a preceding study (Peres and Cancelliere, 2014), to quantify the effects of the imprecise identification of triggering rainfall on the assessment and performances of landslide triggering thresholds. The dataset is in principle "error-free" in the sense that the instants of landslide triggering are exactly known, as well as the triggering rainfall time history. We then fictitiously introduce errors in the triggering instants and in the rainfall series based on ~~realistic~~hypothetical scenarios of landslide data retrieval and analysis, and analyse the ~~effect~~implications on the accuracy of ID thresholds. Quality of information available in real datasets is generally intermediate of that corresponding to the hypothesized scenarios. These scenarios are combined with different criteria for event rainfall identification, and different aggregations of rainfall data (hourly and daily, and daily in the presence of a shift due to manual collection of data), so the effects of these other two sources of uncertainty are analysed as well~~.~~ (items *i*) and *ii*) of the above list). The synthetic data used for our analyses are based on characteristic for hillslopes in the landslide-prone region of Peloritani Mountains, in ~~North-eastern~~Northeastern Sicily, Southern Italy.

2 Dataset: generation of synthetic rainfall and landslide data

We refer to the dataset ~~built~~developed in Peres and Cancelliere (2014~~) has been used here as reference.~~). Here we provide a basic description of the methodology used for its ~~development~~generation, which includes the following steps:

- *Synthetic generation of hourly rainfall time series:* A seasonal Neyman-Scott Rectangular Pulses (NSRP) stochastic rainfall model (Cowpertwait et al., 1996; Rodríguez-Iturbe et al., 1987a, 1987b) is used for the generation of 1000-years of hourly rainfall data. The model is calibrated on approximately 9 years of hourly observations from the Fiumedinisi rain gauge located in the area (Fig. 1).
- *Computation of hillslope pressure-head response*: A two-state hydrological model is used for the computation of pressure head. State 1 and 2 are activated separately during rainfall events and during no-rain intervals, respectively. Rainfall events are defined as a section of the rainfall series preceded and followed by no rainfall for a minimum time interval of 24 hours. Within state 1 the TRIGRS-v2 model (Baum et al., 2010) is applied, which is based on the Richards' equation for mono-dimensional vertical infiltration with a Gardner negative exponential soil water characteristic curve. This is the least simplified form of the Richards' equation for which an analytical solution has been derived so far (Srivastava and Yeh, 1991). A leakage flux at the soil-bedrock interface is

considered, assuming the vertical hydraulic conductivity of the bedrock strata $c_D = 0.1$ times the saturated conductivity K_S of the pervious soil layer. Within state 2 a linear reservoir water table recession model is activated to simulate sub-horizontal drainage, and is used to compute water table height at the beginning of the next passage to state 1. A linear reservoir scheme computes a drainage flow that depends on the water table level, determining a negative-exponential decay of pressure head at the bottom of the regolith layer, with recession constant τ_M.

- *Derivation of virtual landslide occurrence times*: An infinite slope model to compute the factor of safety F_S for slope stability is applied. For this schematization, failure surface coincides with the regolith-bedrock interface. The time instants at which a downward crossing of $F_S = 1$ occurs are assumed to be the instants at which landslides are triggered.

The data set is generated considering soil hydraulic and geotechnical properties ~~reported~~shown in Tab. 1 that can be considered representative of hillslopes in the Peloritani Mountains landslide-prone area (see Fig. 1). ~~This~~The application to a hillslope of definite characteristics enables us to isolate the impact of the uncertainty in triggering rainfall identification; regional determination of thresholds do contain also factors of uncertainty related to the heterogeneity of landslide characteristics; the assessment of this combined uncertainty is out of the scope of our present analysis. The Peloritani area has been affected several times by catastrophic shallow landslide phenomena in the past; including the 1 October 2009 disaster, which has been analysed and described in several studies (Cama et al., 2017; Schilirò et al., 2015a, 2015b, 2016; Stancanelli et al., 2017)~~.~~. A morphological analysis of the catastrophic landslides occurred on 1 October 2009, has shown that a reasonable value of the recession constant for the specific case study area is $\tau_M = 2.$~~7~~75 days (Peres and Cancelliere, 2014). Nevertheless, for the purposes of this study, we focus our analysis mainly on the hypothetical case of no pressure head memory ($\tau_M = 0$), so ~~to isolate~~that the main source of ~~impact of~~ uncertainty ~~in~~considered in threshold determination is that related to identification of triggering rainfall events. In other words, in the “ideal” simulations described above, the only uncertainty present is that of rainfall intra-event intensity variability, which is relatively small, so that a landslide-triggering threshold expressed in terms of rainfall duration and intensity performs almost perfectly (Peres and Cancelliere, 2014). For completeness, we however present a secondary analysis~~, in~~ including antecedent rainfall memory, for which $\tau_M = 2.$~~7~~75 days – a value determined from the analysis observed landslides (see Peres and Cancelliere, 2014)~~.~~. Table 2 shows some characteristics of the 1000-year long synthetic databases, which do not change among the different scenarios illustrated in the following section.

3 Methodology

3.1 Simulation of uncertainty in triggering rainfall identification ~~uncertainty~~

As already mentioned, the available triggering instants from real landslide inventories are seldom precise. On the other hand, the instants at which landslides are triggered are known exactly (on hourly resolution) for the ~~ideal~~ synthetic series~~,~~ illustrated in previous ~~section~~Sect. 2. We then introduce errors ~~in the triggering instants~~to this synthetic dataset by hypothesizing the way such an information may be retrieved from newspapers, and similar resources (blogs and fire brigades reports), which in fact are the main primary sources available to build landslide historical inventories (e.g., Guzzetti and Tonelli, 2004).

We suppose that only the date of the landslide is reported, ~~and so is done~~ with some delay~~.~~ (See Fig. 2). For a landslide to be reported on day *D*, it has to be ~~observed~~spotted within a time interval ~~that goes from~~we denote as the ~~night preceding that~~“observers’ day ~~to the end of its working hours (the “observer day”).~~” *D’*. Then the user of the landslide archive (the analyser), makes an interpretation of the available information, i.e. chooses an instant of the reported day of landslide occurrence to seek backwards for the triggering rainfall.

~~Based on the above reasoning, we simulate the errors induced by the use of these sources by distinguishing an observation day *D'*, that ends at hour T_O of day *D*, and an analyser time, T_A (Fig. 2).~~In particular, the *i*-th landslide observed at t_i within the ~~observation~~observers' day *D'*, i.e. hours [T_O – 24 h, T_O] of day *D*, is assumed by the analyser to be triggered T_A hours after the start of day *D* (civil day *D* starts at 00:00). The observer day is made of the hours in which observers can report a landslide on day *D*. We assume that the observer day is given by hours going from 6 pm of day *D*–1 to 6 pm of day *D* (T_O = 18 h~~).~~); this choice is an attempt to resemble usual working hours, and the fact landslides occurring by night may be reported the morning after. The analyser time is the instant of landslide triggering as considered by who analyses the data (the "analyser") to derive landslide-triggering thresholds, counted from the beginning of day *D*. This way to process the data introduces a sampling error and a shift between the actual instant at which the generic landslide *i* is triggered, t_i , and that assumed by who analyses the data, t_i'. Hence the error for the *i*-th landslide is given by:

$$e_i = t_i' - \cancel{t_i}\, t_i \qquad (1)$$

These errors are implicitly random, since though t_i' are deterministically chosen, the actual instant t_i varies in an aleatory fashion according to rainfall time history.

A positive error can be in general considered as more ~~probable~~likely than a negative, since landslides are typically reported after some time they have occurred (Guzzetti et al., 2007, 2008; Peres and Cancelliere, 2013)~~(Guzzetti et al 1997, 1998;~~. This, however, does not exclude the possibility of a significant number of negative errors, because of temporal shifts in rainfall data, as will be discussed later.

The two parameters, T_O and T_A, can be set to simulate ~~different realistic~~a range of scenarios, for which real situations may represent intermediate cases. We perform our analysis based on four scenarios (which include the "ideal" one), hereafter referred to as landslide information "reporting scenarios" (RS), and illustrated in Fig. 2:

1. *Ideal scenario* RS0 (T_O = 0, T_A = 0; e_i = 0 for all landslides). This is the ~~ideal~~error-free scenario (described in Sect. 2), ~~without errors,~~ that is considered ~~as a reference~~ for ~~measuring errors in~~definition of the actual instants of landslide triggering~~ instants in the database simulated by the three following scenarios.~~, t_i.
2. *Small delay reporting* RS1 (T_O = 18 h, T_A = 24 h; random in the range $0 \le e_i \le 30$ hours). A landslide occurring within the interval from night hours of *D* – 1 until the evening of day *D* (i.e. within the observers' day *D'*) will be reported at day *D*. Here we suppose that the analyst attributes the landslide *at the end* of day *D* (T_A = 24 hours), i.e. searches the triggering event backwards from that instant.
3. *Large delay reporting* RS2 (T_O = 18 h, T_A =48 h; random in the range $0 \le e_i \le 54$ hours). This scenario is similar to the previous, but here larger errors are hypothesized. We suppose that the landslide occurring during the observers' day *D'* is reported on day *D* + 1, which is also erroneously assumed by the analyser as the day at which the landslide was triggered. He then attributes the landslide at the end day *D* + 1 (T_A = 48 hours). These timing errors may also be likely when landslides occur on weekends.
4. *Anticipated reporting* RS3 (T_O = 18 h, T_A = 0 h; random in the range $-18 \le e_i \le 6$ hours): This case is the same of RS1, but here analyst searches the triggering event backwards *from the beginning* of day *D*, i.e. at 00:00 (instead that at 24:00).

Within the context of sampling errors, another point is related to the way *rainfall* data is collected, specifically for daily data manually measured until some decades ago. A significant amount of papers derive landslide triggering thresholds using daily rainfall data (Berti et al., 2012; Leonarduzzi et al., 2017; Li et al., 2011; Terlien, 1998). In an ideal situation rainfall intensity should be aggregated from 00:00 to 23:59, i.e. over a ~~"~~civil "calendar day", as illustrated in Fig. 3. With reference to manual collection of rainfall data, this requires that ~~raingauge~~rain gauge should be read at midnight of each day, which is an uncomfortable hour. Manual collection of daily data is usually carried out at easier hours. For instance, in Italy, where the widest source of information are the Hydrological Bulletins (locally known as *Annali Idrologici*), the operator would measure the rainfall collected in the rainfall bucket every day at 9~~:00~~ am. Thus, daily rainfall in a given day is the amount of rainfall occurred in the 24 hours preceding 9 am of the same day. As illustrated in Fig. 3, in this case the reported daily rainfall amounts can be dramatically different than actual (see also Caracciolo et al., 2017).

Identification of triggering rainfall is uncertain also because of the different criteria that one can apply to isolate rainfall events from a continuous time series – Tab. 3 lists a range of criteria adopted in the literature. Here we analyse how the different criteria can impact the identification of triggering rainfall, both in the case that uncertainty in the triggering instants is present (datasets RS1-RS3) or not (dataset RS0).

The automatic procedure we adopt for isolating events is as follows (see sketch on Fig. 4). First, a minimum rainfall threshold $s_{\min}$ is applied to all rainfall pulses at the fixed temporal aggregation. This means that from the original series a new one is obtained, where precipitation pulses less than $s_{\min}$ are replaced by zeros. In the sketch, these pulses are colored in light gray. Afterwards, rainfall events are singled-out when separated by zero-rain intervals longer than $u_{\min}$. This parameter is the most important parameter for ~~definition~~the identification of rainfall events. With the aim of quantifying how the impact of the errors implied by the different reporting scenarios changes with rainfall identification criteria, various pairs of $s_{\min}$ and $u_{\min}$ have been set (see Sect. 3.3). The described algorithm defines the rainfall event regardless it is associated or not to a landslide. For attributing a rainfall event to a landslide, the cases where the triggering instant is within a dry or a wet period~~,~~ should be analysed separately. In the first case, the landslide is associated to the whole closest event occurring before the landslide, in the other case it is to the part of rainfall event occurring before the triggering instant. Automatic procedures have the advantage of being objective and reproducible, and thus more scientifically sound than subjective judgment (Melillo et al., 2015; Vessia et al., 2014); nevertheless, algorithms are suitable to reproduce the latter with a certain level of fidelity (Berti et al., 2012).

Finally, triggering rainfall identification uncertainty is simulated by combining the reporting scenarios, different parameters of the rainfall event identification algorithm, and three rainfall aggregation schemes (hourly, daily correct and daily shifted). This results in twenty-eight combinations for each recession constant value τ_M (see Tab. 4).

3.2 Threshold definition, calibration and testing performance

Seventeen different landslide-triggering threshold types based on rainfall characteristics have been proposed in the literature in the period 1970-2006 (according to the list reported at rainfallthresholds.irpi.cnr.it, last date accessed ~~11 Sept. 2017~~15 Jan. 2018). In spite of this variety, the most widely used is the rainfall intensity-duration (ID) threshold, as 96 out of 125 (about 77 %) are of this type, if one includes equivalent rainfall depth-duration (ED) thresholds. Therefore, our analysis adopts this threshold type, which may be defined as follows:

$$I = \alpha D^{-\beta} \qquad (2)$$

where I [L/T] is the mean rainfall event intensity, D [L] is the rainfall event duration (both defined according to scheme of Fig. 4~~).~~); $\alpha, \beta > 0$ are respectively the intercept and slope parameters of the threshold. ED thresholds are equivalent to IDs, since rainfall intensity I is the ratio between event rainfall E (the total depth of a rainfall event) and its duration D; so ~~an~~ they can be converted in the ID type ~~by~~ just by subtracting 1 to the exponent of duration.

The procedures for the identification of best threshold parameters have historically increased their complexity through time. Early works have considered lower boundary curves of the triggering events traced with subjective criteria (Caine, 1980). Then more objective procedures have been then proposed, still based on the triggering events only, such as the so-called “frequentist” method (e.g., Brunetti et al., 2010). ~~Finally,~~ More advanced approaches are currently used, and are derived from the analysis of both triggering and non-triggering events. These procedures are more transparent than methods based on triggering events only, as the uncertainty of the thresholds can be assessed through indices based on the confusion matrix~~,~~ or the Receiver-operating characteristics (ROC), that is, in terms of the count of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) (Tab. 5). More importantly, these methods are also more robust, since the presence of non-triggering data points makes the choice of the threshold less sensitive to possible errors in the attribution of triggering rainfall event duration and intensity. Here we use these methods of recent ~~methods~~application, implicitly assuming that the impact of the uncertainty under analysis is likely to be higher on thresholds derived from procedures based on triggering rainfall only.

Best-thresholds can be calibrated by maximizing their performances expressed in terms of suitable metrics. One widely used metric is the True Skill Statistics (Ciavolella et al., 2016; Peres and Cancelliere, 2014; Staley et al., 2013) originally proposed by Peirce (1884) :

$$\mathrm{TSS} = \frac{\mathrm{TP}}{\mathrm{TP+FN}} - \frac{\mathrm{FP}}{\mathrm{TN+FP}} \qquad (3)$$

An apparently alternative approach is given by Bayesian analysis (Berti et al., 2012). Indeed, this approach can be interpreted as a special case of the ROC analysis, since Bayesian a-posteriori probability ~~equals~~is equivalent to the ROC-based Precision (PRE):

$$P(L|R) = \frac{P(R|L)P(L)}{P(R)} = \frac{\frac{\mathrm{TP}}{\mathrm{TP+FN}}\frac{\mathrm{TP+FN}}{\mathrm{N_T}}}{\frac{\mathrm{TP+FP}}{\mathrm{N_T}}} = \frac{\mathrm{TP}}{\mathrm{TP+FP}} = \mathrm{PRE} \qquad (4)$$

where:

$P(L|R)$ = probability of landslide occurrence given rainfall exceeding the threshold (*a posteriori* probability),

N_T = total number of rainfall events (triggering and non-triggering),

$P(R) = (\mathrm{TP} + \mathrm{FP})/N_T$ = probability of rainfall events exceeding the threshold,

$P(L) = (\mathrm{TP} + \mathrm{FN})/N_T$ = (*a priori*) probability of landslide occurrence,

$P(R|L) = \mathrm{TP}/(\mathrm{TP+FN})$ = probability of rainfall event exceeding the threshold, given that a landslide has occurred (known as the *likelihood*).

Different papers discuss advantages and disadvantages of various indices proposed in natural-hazard forecasting, as one single index is not sufficient to fully describe the confusion matrix (Frattini et al., 2010; Murphy, 1996; Stephenson, 2000). Nevertheless, the choice of a single index is essential to keep the calibration procedure simple, i.e. a single-objective optimization problem. Hence, ~~we~~ here we calibrate thresholds by maximizing the TSS. One advantage of the TSS is that it includes all the entries of the confusion matrix, and thus its maximization yields thresholds that result in a good trade-off between correct and wrong warnings/non-warnings.

Once thresholds for each RS scenario are derived, the TSS and the confusion matrix ~~in general~~ provide a measure of the uncertainty inherent the data, as assessable by who derives the threshold~~,~~ and is not aware of the errors~~ that could be present in the data~~. On the other hand, it is also of interest to test how a threshold derived from erroneous data may perform when, after its determination, it is applied to precise monitored data, and thus ~~mostly~~potentially free of the errors present in the threshold calibration data set. To do so, the calibrated thresholds are applied to the ~~ideal~~error-free synthetic data set~~.~~ (Sect. 5). The performances in this test are indicative of the impacts of errors when thresholds are actually used.

4 Impact of ~~errors~~uncertain identification of triggering rainfall on threshold calibration

4.1 Hourly data

Results relative to the use of hourly data are shown in Fig. 5**,** for a given separation algorithm (s_{min} = 0.2 mm, u_{min} = 24 h). For the reference dataset RS0, there is a negligible overlapping between triggering and non-triggering events (Fig. ~~5a).~~5a), due to intra-event rainfall intensity variability. In fact in this case the best ID threshold ($I = 101\ D^{-0.80}$) performs almost perfectly, with a TSS of 0.99 (for u_{min} = 24 h). The presence of small delay reporting errors (RS1), has little impacts on the position of ~~the~~ triggering rainfall points (Fig. ~~5b). Two~~5b), which in general are shifted slightly down along the intensity axis; this is related to the higher durations produced by positive errors in triggering instants, combined with an induced decrease of mean rainfall event intensities – a general behavior exhibited by extreme events (cf. the negative slope of well-known rainfall intensity-duration-frequency curves, see Bogaard and Greco, 2018~~are shifted~~). Only two rainfall events (the 2.5 % of triggering events) are highly-impacted, being moved to a duration of 1 hour~~, which contributes~~. The latter and mainly the former effect, contribute to slightly flatten the threshold for TSS maximization (decrease of ~~|β|~~β to 0.7). When high delay sampling errors are present (RS2), the effects may not be negligible as in the previous case, as more ~~erroneous~~highly-impacted rainfall events are present, now also for significant durations (up to 24 h in the plot, Fig. 5c). These ~~errors~~erroneous data points are difficult to be identified by an analyser, and thus their impact on threshold determination can be significant, and lead to a lower slope and intercept, i.e. an underestimation of the threshold, which changes to I~~=19D~~ = 19 $D^{-0.50}$ (reference is $I = 101\ D^{-0.80}$). The impact of these errors may be more dramatic when thresholds are assessed making use of triggering rainfall events only, following "traditional", less robust, approaches.

Negative errors, introduced by an anticipation of the real landslide instant (RS3), can have very high impacts, as can be seen from the relative plot in Fig. 5d, and the loss of the correct position of many of the triggering points. The best threshold corresponds to TSS = 0.49, which reflects the high degree of uncertainty implied by this kind of data errors.

4.2 Daily data

Shallow landslides can be triggered by rainfall events that are only ~~some~~few hours long (Bogaard and Greco, 2016; Highland and Bobrowsky, 2008; Sidle and Ochiai, 2013), and various studies have shown that the impact of small scale intra-event rainfall intensity variability can have a significant effect on landslide triggering (D'Odorico et al., 2005; Peres and Cancelliere, 2014, 2016). Hence, apart from the errors in the dataset, it is of interest to see how the ~~passage~~change from hourly to daily data may affect threshold determination. This can be done by comparing thresholds determined from the hourly and daily datasets.

Figure 6 shows the results of calibration obtained with correctly-aggregated daily rainfall data and s_{min} = ~~5 mm/day,~~0 and u_{min} = 1 day. As can be seen from the plots, the impact of delayed reporting of landslides (errors RS1 and RS2) is less significant than with hourly data. In fact, though α and β are lower than those determined from hourly data, the threshold determined from daily data passes more or less in the same zone for durations in their range of validity, D > 1 day. This is because the smaller slope β in the log-log plane compensates the smaller intercept α. The effect of anticipating landslide time location (RS3) has also here high impacts on the thresholds, Fig. 6d.

Figure 7 plots the results relative to daily rainfall data affected by a delay in the aggregation interval, as present for instance in Italian datasets, and related to ~~use~~availability of data from non-automatic rain gauges. The impacts of this systematic rainfall error can be high (Fig. 7a, b, and d). There is, however, the possibility that the errors due to rainfall aggregation and reporting landslide time interval compensate for each other, as in the case of scenario RS2 (delayed reporting of landslides), Fig. ~~7c.~~ 7c (notice that this plot is similar to Fig. 6b). If the analyser is aware of the rainfall-aggregation shift, then he should correct as much as possible for this error – in this specific case, by shifting the entire daily rainfall dataset one day forward.

4.3 Possible effects of rainfall separation criteria and antecedent rainfall

~~Figure 8~~Table 6 shows the results obtained by setting the parameters of rainfall event separation algorithms, in the ~~(a)~~ hourly, ~~(b)~~ daily correct, and ~~(c)~~ daily shifted ~~data~~aggregation cases. From the TSS values ~~shown in Fig. 8a~~obtained for hourly data, it can be seen that the impact of RS1 and RS2 ~~errors~~ increases with decreasing minimum interarrival value u_{min}. ~~In the case of~~For RS3~~ error~~, differences obtained with different u_{min} are not relevant, since the performances are poor in general (TSS a~~r~~bou~~nd~~t 0.5). In the case of daily data~~ (Figs. 8 b and c)~~, the importance of different criteria for separating events (values of the minimum daily rainfall threshold s_{min}), are relatively lower than in the hourly data case. Though differences in the TSS are not significant, this may not be true for the thresholds parameters, which can vary significantly. In fact, ~~a higher s_{min} results in~~ higher thresholds are obtained from an increase of s_{min}, because of the ~~removal~~decrease of the number of days ~~of with below a given rainfall amount~~counted as rainy.

The behaviour related to hourly data, ~~may be due~~is related to the fact that, by choosing lower u_{min}, events get generally shorter ~~and more numerous~~, and thus it is more likely that a landslide event is attributed to only a part of the actual triggering event. In this case the effect of preceding rainfall events cannot be neglected in general. In other words, our analysis suggests that the choice of the u_{min} is crucial, and must be based on the timescales of the hydrological processes governing landslide triggering, in terms of long and short term responses (Iverson, 2000). This means that the effect of different criteria for rainfall separation is somehow related to that of antecedent precipitation. The effects of antecedent precipitation is specifically taken into account performing Monte Carlo simulations with $\tau_M = 2.75$ days (results shown in ~~Fig~~Tab. 9). For this simulation, ~~no matter what is~~regardless of the rainfall separation time interval, the initial water table height measured from the bottom of the soil column is in general greater than zero, becoming negligible after a dry interval of $3\tau_M$ ~~=~~ = 3×2.75 = 8.5 days~~.~~ (exponential decay). As can be seen, the results are qualitatively similar to the no-memory case; the main difference is that lower TSS are obtained for the added uncertainty due to antecedent conditions, and the thresholds are lower, since less event rainfall is needed in average ~~needed~~ to trigger a landslide ~~due to~~because of non-zero initial wetness conditions.

5 Impact of ~~errors~~uncertain identification of triggering rainfall on threshold use

Thresholds determined based on historical datasets are then meant to be used ~~within~~for early warning systems, when, consequently, more detailed meteorological and landslide monitoring is set up. This means that ~~the~~it is reasonable to hypothesize that after thresholds are determined ~~with real datasets, affected by errors, are then~~, they are subsequently applied to high quality datasets, which suffer less ~~suffering~~ of the limitations and errors present in datasets used for threshold ~~determination,~~calibration, generally not ~~initially~~ conceived for that specific purpose. This might induce to modify the thresholds in view of the new data, but this is a process ~~can~~whose implementation may take several years. Hence, with the aim of determining which would be the consequences of building an early warning system with thresholds derived from historical data with errors, Fig. ~~10~~8 shows a visual comparison between the thresholds determined in the various numerical experiments and the ideal hourly dataset, for results related to the hourly (Fig. ~~10a~~8a) and daily data sets (Fig. ~~10b~~8b). For sake of clarity, it may be worthwhile to remember that the dataset of triggering and non-triggering points has been used in calibrating the thresholds only ~~in~~for the RS0 scenario ~~(no errors),~~, with hourly data~~,~~ and u_{min} = 24 h~~,~~ and s_{min} = 0.2 mm (the related threshold is shown ~~in~~on Fig. ~~10a~~8 as a thick black line). Thus the other thresholds are tested against this ideal dataset, which ~~is not~~differs from the one used for their calibration.

The plots show that the presence of errors can induce a significant variability of thresholds which is totally unrelated to the different characteristics of a site (i.e. the geomorphological, hydraulic, geotechnical and land use characteristics). This allows to ~~draw the hypothesis~~speculate that a significant part of the variability of landslide triggering-thresholds reported in literature (cf., Guzzetti et al., 2007) may be ~~related~~due to the sources of uncertainty here discussed. As a consequence, it is challenging to seek for links between the variability of physio-geographical characteristics and those of thresholds, as determined from different sites.

The presence of errors in the landslide dataset yields thresholds that are in general underestimated, i.e. lower than the correct ones. Many thresholds on Fig. 8 are significantly lower than the correct ones, and the number of false positives can be relatively high, and not balanced by true positives. A good trade-off between correct and wrong predictions is essential for the success of an early warning system, since with an high number of false alarms the so-called cry-wolf effect may take place, inducing the populations not to take precautionary actions when warnings are issued (Barnes et al., 2007).

6 Conclusions

~~According to several studies, landslide inventories do not provide precise triggering instants information. In this paper,~~We have analysed and discussed the possible effects of ~~this problem~~uncertain triggering rainfall identification on the assessment of empirical landslide early warning ID thresholds~~ for landslide initiation have been analysed and discussed~~, capitalizing on ~~an ideal~~a synthetic rainfall-landslide dataset generated by Monte Carlo simulation. To this aim, we have investigated the effect of a set of hypothesized ~~reasonable~~ scenarios of landslide information retrieval and interpretation~~, that~~ which can induce errors in the identification of instants of landslide occurrence. Moreover, we have analysed how the impact of reasonable scenarios may vary in dependence of rainfall aggregation (hourly or daily), and of rainfall event identification criteria. Real situations may be a mixture of the considered scenarios, and thus the impacts are presumably intermediate between the ones hypothesized.

The errors in the time instants can be, in an algebraic sense, positive or negative, according to whether ~~the~~a landslide is reported after its actual occurrence or before, respectively. Following literature, positive errors are more likely than negative, since ~~it is typical that~~typically a landslide is reported some time after its actual occurrence. Our analyses have shown that if ~~these~~such errors are limited to less than 30 h (about one day~~,~~), their impacts on the threshold may be relatively low; yet if the delay is higher, impacts can be significant. Negative errors, though less probable, can also exist, based on how an analyst interprets the information retrieved from landslide historical archives. The impact of these errors can be dramatic, as the location of triggering ~~-~~events in the log$D$ – log$I$ plane can be completely ~~modified~~altered. Errors in landslide triggering instants lead to triggering events that are shorter than the actual ones, so that their effect is to induce an incorrect identification of triggering rainfall for short durations. For higher durations (>1 day), the location of triggering events seems to be more robust, except when negative errors are present. This behavio~~u~~r induces a flattening of the ~~P~~ID thresholds (i.e. a lower slope ~~|β|)~~β) and an underestimation of the position parameter of the threshold (i.e. a lower intercept α).

The impact of reporting errors can change significantly in dependence of the algorithm adopted for rainfall event identification. Specifically, a shorter "maximum dryness" interval for event separation induces an increase of the impacts of all ~~kind of landslide time~~ reporting ~~errors.~~scenarios.

~~When thresholds are determined from daily data, the data analyst has to be aware of possible shifts/~~From our analysis no significant impacts seem to be induced by the use of daily data; however, it is of fundamental importance to check, and correct where possible, for the presence of delays in the rainfall accumulation interval, that is if precipitation reported for a given day is the total amount occurred in a shifted period (e.g., within the 24 hours preceding 9 am of that day rather than before midnight). Such a kind of shift affects, for instance, the Italian Hydrological Annual Reports, which constitute the largest rainfall data collection in Italy. The impacts of these shifts are potentially dramatic.

Overall, the presence of reporting errors in landslide triggering instants brings to lower thresholds, making them less suitable to set up of landslide early warning systems, as they can lead to a high number of false alarms, generating a misbelief by populations that are expected to benefit from their implementation. Similar effects have been found as a consequence of rainfall measurement uncertainty on thresholds (Nikolopoulos et al., 2014)~~. Just~~. These two sources of errors– always present in observed datasets – are alone enough to generate an uncertainty in thresholds assessment that is of significant magnitude. These results bring to the conclusion that the uncertainty inherent the available data can jeopardize the possibility to find a physically based rationale underlying the variability of empirical landslide-triggering thresholds across different sites. In other words, with the quality of current available data, attempts of relating thresholds to climate and other regional characteristics can be very difficult. An improvement of landslide and rainfall monitoring – e.g. rainfall, soil moisture and landslide satellite data, as well as landslide data crowd-sourcing (Guzzetti et al., 2012; Strozzi et al., 2013; Wan et al., 2014) – may be a step forward for overcoming these problems. Once accurate rainfall-landslide data are available, standardized methodologies have to be implemented to derive the thresholds, in order to allow their comparisons and to link their variability to site-specific landslide susceptibility factors.

Aknowledgements ~~David J. Peres~~. *DJP was supported by post-doctoral contract on "Studio dei processi idrologici relative a frane superficiali in un contesto di cambiamenti climatici" (Analysis of landslide hydrological processes in a changing climate), at University of Catania. Part of the work was developed during his three-month stay as a visiting researcher at the Water Resources Section of TUDelft. The authors thank the comments by two anonymous reviewers, which helped to considerably improve the paper.*

References

Barnes, L. R., Gruntfest, E. C., Hayden, M. H., Schultz, D. M. and Benight, C.: False Alarms and Close Calls: A Conceptual Model of Warning Accuracy, Weather Forecast., 22(5), 1140–1147, doi:10.1175/WAF1031.1, 2007.

Baum, R. L., Godt, J. W. and Savage, W. Z.: Estimating the timing and location of shallow rainfall-induced landslides using a model for transient, unsaturated infiltration, J. Geophys. Res., 115(F3), F03013, doi:10.1029/2009JF001321, 2010.

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

Bogaard, T. and Greco, R.: Invited perspectives: Hydrological perspectives on precipitation intensity-duration thresholds for landslide initiation: proposing hydro-meteorological thresholds, Nat. Hazards Earth Syst. Sci., 18(1), 31–39, doi:10.5194/nhess-18-31-2018, 2018.

Bogaard, T. A. and Greco, R.: Landslide hydrology: from hydrology to pore pressure, Wiley Interdiscip. Rev. Water, 3(3), 439–459, doi:10.1002/wat2.1126, 2016.

Brunetti, M. T., Peruccacci, S., Rossi, M., Luciani, S., Valigi, D. and Guzzetti, F.: Rainfall thresholds for the possible occurrence of landslides in Italy, Nat. Hazards Earth Syst. Sci., 10(3), 447–458, doi:10.5194/nhess-10-447-2010, 2010.

Caine, N.: The Rainfall Intensity-Duration Control of Shallow Landslides and Debris Flows, Soc. Swedish Ann. Geogr. Geogr. Phys., 62(1), 23–27, 1980.

Cama, M., Lombardo, L., Conoscenti, C. and Rotigliano, E.: Improving transferability strategies for debris flow susceptibility assessment: Application to the Saponara and Itala catchments (Messina, Italy), Geomorphology, 288, 52–65, doi:10.1016/J.GEOMORPH.2017.03.025, 2017.

Caracciolo, D., Arnone, E., Conti, F. Lo and Noto, L. V.: Exploiting historical rainfall and landslide data in a spatial database for the derivation of critical rainfall thresholds, Environ. Earth Sci., 76(5), 222, doi:10.1007/s12665-017-6545-5, 2017.

Ciavolella, M., Bogaard, T., Gargano, R. and Greco, R.: Is there Predictive Power in Hydrological Catchment Information for Regional Landslide Hazard Assessment?, Procedia Earth Planet. Sci., 16, 195–203, doi:10.1016/j.proeps.2016.10.021, 2016.

Cowpertwait, P. S. P., O'Connell, P. E., Metcalfe, A. V. and Mawdsley, J. A.: Stochastic point process modelling of rainfall. I. Single-site fitting and validation, J. Hydrol., 175(1–4), 17–46, doi:10.1016/S0022-1694(96)80004-7, 1996.

D'Odorico, P., Fagherazzi, S. and Rigon, R.: Potential for landsliding: Dependence on hyetograph characteristics, J. Geophys. Res. Earth Surf., 110(1), 1–10, doi:10.1029/2004JF000127, 2005.

Frattini, P., Crosta, G. and Carrara, A.: Techniques for evaluating the performance of landslide susceptibility models, Eng. Geol., 111(1–4), 62–72, doi:10.1016/j.enggeo.2009.12.004, 2010.

Gariano, S. L., Brunetti, M. T., Iovine, G., Melillo, M., Peruccacci, S., Terranova, O., Vennari, C. and Guzzetti, F.: Calibration and validation of rainfall thresholds for shallow landslide forecasting in Sicily, southern Italy, Geomorphology, 228, 653–665, doi:10.1016/j.geomorph.2014.10.019, 2015.

Guzzetti, F. and Tonelli, G.: Information system on hydrological and geomorphological catastrophes in Italy (SICI): a tool for managing landslide and flood hazards, Nat. Hazards Earth Syst. Sci., 4(2), 213–232, doi:10.5194/nhess-4-213-2004, 2004.

Guzzetti, F., Peruccacci, S., Rossi, M. and Stark, C. P.: Rainfall thresholds for the initiation of landslides in central and southern Europe, Meteorol. Atmos. Phys., 98(3–4), 239–267, doi:10.1007/s00703-007-0262-7, 2007.

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

Guzzetti, F., Mondini, A. C., Cardinali, M., Fiorucci, F., Santangelo, M. and Chang, K. T.: Landslide inventory maps: New tools for an old problem, Earth-Science Rev., 112(1–2), 42–66, doi:10.1016/j.earscirev.2012.02.001, 2012.

Highland, L. M. and Bobrowsky, P.: The Landslide Handbook — A Guide to Understanding Landslides, Reston, Virginia., 2008.

Leonarduzzi, E., Molnar, P. and McArdell, B. W.: Predictive performance of rainfall thresholds for shallow landslides in Switzerland from gridded daily data, Water Resour. Res., 53(8), 6612–6625, doi:10.1002/2017WR021044, 2017.

Li, C., Ma, T., Zhu, X. and Li, W.: The power-law relationship between landslide occurrence and rainfall level, Geomorphology, 130(3–4), 221–229, doi:10.1016/j.geomorph.2011.03.018, 2011.

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

Murphy, A. H.: The Finley Affair: A Signal Event in the History of Forecast Verification, Weather Forecast., 11(1), 3–20, doi:10.1175/1520-0434(1996)011<0003:TFAASE>2.0.CO;2, 1996.

Nikolopoulos, E. I., Crema, S., Marchi, L., Marra, F., Guzzetti, F. and Borga, M.: Impact of uncertainty in rainfall estimation on the identification of rainfall thresholds for debris flow occurrence, Geomorphology, 221, 286–297, doi:10.1016/J.GEOMORPH.2014.06.015, 2014.

Peirce, C. S.: The numerical measure of the success predictions, Science (80-.)., 4(93), 453–454, doi:10.1126/science.ns-4.93.453-a, 1884.

Peres, D. J. and Cancelliere, A.: Defining rainfall thresholds for early warning of rainfall-triggered landslides: The case of North-East Sicily, in Landslide Science and Practice: Global Environmental Change, vol. 4, edited by C. Margottini, P. Canuti, and K. Sassa, pp. 257–263., 2013.

Peres, D. J. and Cancelliere, A.: Derivation and evaluation of landslide-triggering thresholds by a Monte Carlo approach, Hydrol. Earth Syst. Sci., 18(12), 4913–4931, doi:10.5194/hess-18-4913-2014, 2014.

Peres, D. J. and Cancelliere, A.: Estimating return period of landslide triggering by Monte Carlo simulation, J. Hydrol., doi:10.1016/j.jhydrol.2016.03.036, 2016.

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

Pizziolo, M., Del Maschio, L., Gozza, G. and Pignone, S.: Determinazione di soglie pluviometriche per l’innesco di frane in Emilia - Romagna, Geol. dell’Emilia-Romagna, 29, 21–27, 2008.

Rappelli, F.: Definizione delle soglie pluviometriche d’innesco di frane superficiali e colate torrentizie: accorpamento per aree omogenee, Rapp. per IReR, Milano, 2008.

Rodríguez-Iturbe, I., de Power, B. F. and Valdés, J. B.: Rectangular pulses point process models for rainfall: Analysis of empirical data, J. Geophys. Res., 92(D8), 9645, doi:10.1029/JD092iD08p09645, 1987a.

Rodríguez-Iturbe, I., Cox, D. R. and Isham, V.: Some Models for Rainfall Based on Stochastic Point Processes, Proc. R. Soc. A Math. Phys. Eng. Sci., 410(1839), 269–288, doi:10.1098/rspa.1987.0039, 1987b.

Saito, H., Nakayama, D. and Matsuyama, H.: Relationship between the initiation of a shallow landslide and rainfall intensity-duration thresholds in Japan, Geomorphology, 118(1–2), 167–175, doi:10.1016/j.geomorph.2009.12.016, 2010.

Schilirò, L., Esposito, C. and Scarascia Mugnozza, G.: Evaluation of shallow landslide-triggering scenarios through a physically based approach: An example of application in the southern Messina area (northeastern Sicily, Italy), Nat. Hazards Earth Syst. Sci., 15(9), 2091–2109, doi:10.5194/nhess-15-2091-2015, 2015a.

Schilirò, L., De Blasio, F. V., Esposito, C. and Scarascia Mugnozza, G.: Reconstruction of a destructive debris-flow event via numerical modeling: The role of valley geometry on flow dynamics, Earth Surf. Process. Landforms, 40(14), 1847–1861, doi:10.1002/esp.3762, 2015b.

Schilirò, L., Montrasio, L. and Scarascia Mugnozza, G.: Prediction of shallow landslide occurrence: Validation of a physically-based approach through a real case study, Sci. Total Environ., 569–570, 134–144, doi:10.1016/j.scitotenv.2016.06.124, 2016.

Segoni, S., Rosi, A., Rossi, G., Catani, F. and Casagli, N.: Analysing the relationship between rainfalls and landslides to define a mosaic of triggering thresholds for regional-scale warning systems, , 2637–2648, doi:10.5194/nhess-14-2637-2014, 2014a.

Segoni, S., Rossi, G., Rosi, A. and Catani, F.: Landslides triggered by rainfall: A semi-automated procedure to define consistent intensity-duration thresholds, Comput. Geosci., 63, doi:10.1016/j.cageo.2013.10.009, 2014b.

Sidle, R. C. and Ochiai, H.: Landslides: Processes, Prediction, and Land Use, Water Resources Monograph., 2013.

Srivastava, R. and Yeh, T.-C. J.: Analytical solutions for one-dimensional, transient infiltration toward the water table in homogeneous and layered soils, Water Resour. Res., 27(5), 753–762, doi:10.1029/90WR02772, 1991.

Staley, D. M., Kean, J. W., Cannon, S. H., Schmidt, K. M. and Laber, J. L.: Objective definition of rainfall intensity-duration thresholds for the initiation of post-fire debris flows in southern California, Landslides, 10(5), 547–562, doi:10.1007/s10346-012-0341-9, 2013.

Stancanelli, L. M., Peres, D. J., Cancelliere, A. and Foti, E.: A combined triggering-propagation modeling approach for the assessment of rainfall induced debris flow susceptibility, J. Hydrol., 550, 130–143, doi:10.1016/j.jhydrol.2017.04.038, 2017.

Stephenson, D. B.: Use of the "Odds Ratio" for Diagnosing Forecast Skill, Weather Forecast., 15(2), 221–232, doi:10.1175/1520-0434(2000)015<0221:UOTORF>2.0.CO;2, 2000.

Strozzi, T., Ambrosi, C. and Raetzo, H.: Interpretation of aerial photographs and satellite SAR interferometry for the inventory of landslides, Remote Sens., 5(5), 2554–2570, doi:10.3390/rs5052554, 2013.

Terlien, M. T. J.: The determination of statistical and deterministic hydrological landslide-triggering thresholds, Environ. Geol., 35(2–3), 124–130, doi:10.1007/s002540050299, 1998.

Vennari, C., Gariano, S. L., Antronico, L., Brunetti, M. T., Iovine, G., Peruccacci, S., Terranova, O. and Guzzetti, F.: Rainfall thresholds for shallow landslide occurrence in Calabria, southern Italy, Nat. Hazards Earth Syst. Sci., 14(2), 317–330, doi:10.5194/nhess-14-317-2014, 2014.

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.

Wan, Z., Hong, Y., Khan, S., Gourley, J., Flamig, Z., Kirschbaum, D. and Tang, G.: A cloud-based global flood disaster community cyber-infrastructure: Development and demonstration, Environ. Model. Softw., 58, 86–94, doi:10.1016/j.envsoft.2014.04.007, 2014.

List of figures

Figure 1: Location of the Peloritani Mountains area in Sicily, Italy, and of the Fiumedinisi rain gauge.

Figure 2: Sketch illustrating simulation of uncertainty in triggering instants likely present in landslide inventories built from newspapers or similar sources. The black numbered circles indicate one of the reporting scenarios (RS), which may induce a random error $e = t' - t$ in landslide triggering instants. In particular, a landslide that occurs within the observers' day, is reported at day *D* and attributed to the end of the same day (small delay reporting scenario, RS1) or to its beginning (anticipated reporting scenario, RS3). It can be reported also at day *D*+1 and then attributed to the end of it (large delay reporting scenario RS2). These scenarios can be described in terms of two parameters: T_O = the ending hour of observers' day, and T_A = the triggering instant, referred to hours 00:00 of day *D*, assumed by an analyser who interprets the newspaper-like information.

Figure 3: Aggregation of rainfall data from hourly to daily time scale: daily rainfall depths on the top row result from correct aggregation; those on the bottom row from shifted aggregation, as occurs for the Italian Hydrological Bulletins (Annali Idrologici). The shift is due to manual collection of data in early decades of operation of the monitoring network; the presence of the shift is still continued, in spite of installation of automatic rain gauges, to preserve homogeneity of the entire historical time series.

Figure 4: Sketch illustrating the algorithm for the identification of triggering and non-triggering rainfall events, and relative parameters s_{min} and u_{min}. When a landslide is triggered in a dry period, it is attributed to the whole event preceding it; otherwise, only the part of the event preceding the landslide triggering instant is considered. For non-triggering rainfall (the first one in the sketch), duration and intensity are computed considering the entire rainfall event.

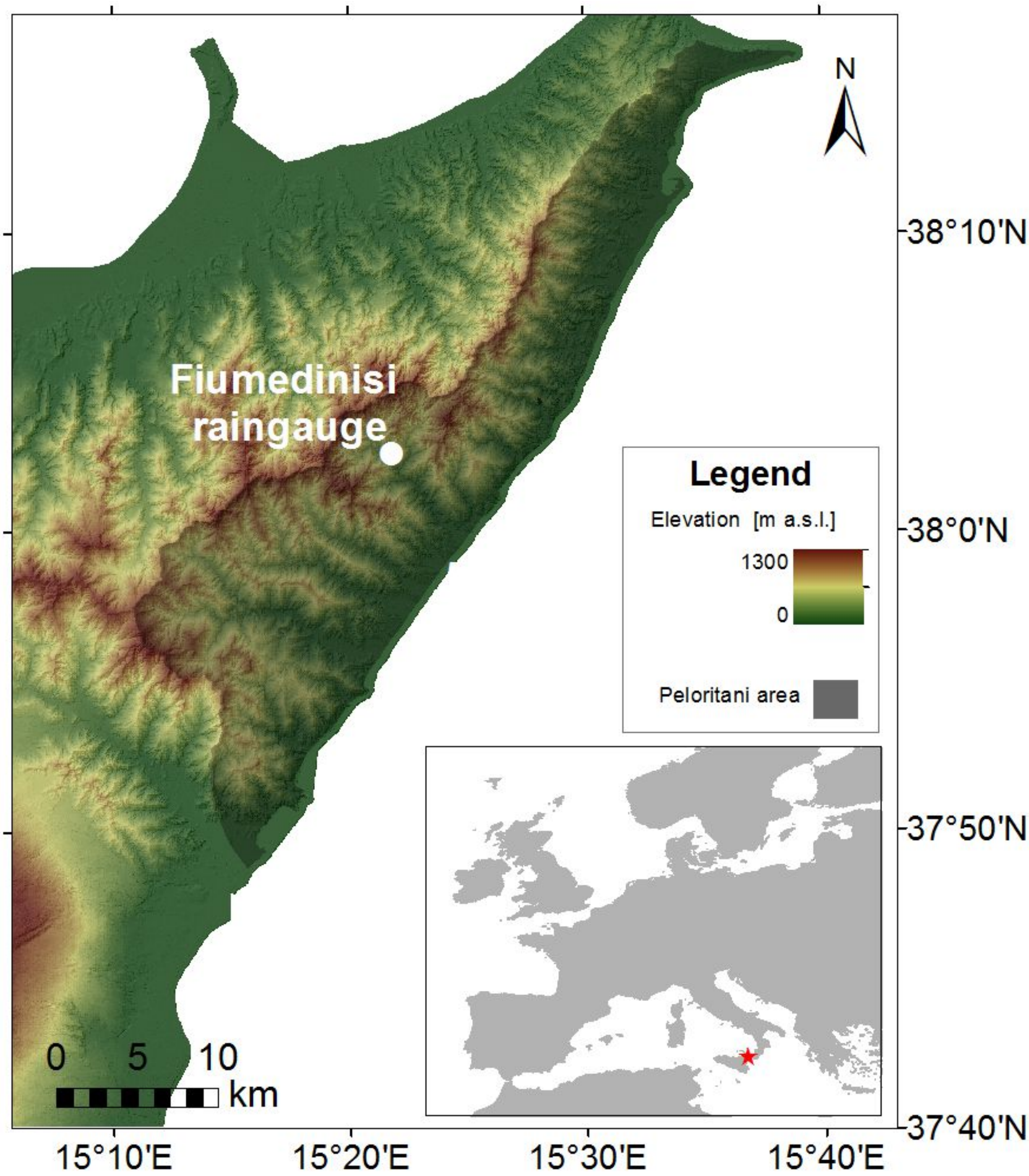
Figure 5: Scatter plot, in the double-logarithmic rainfall duration-intensity plane, of triggering and non-triggering events for *hourly* data and separation algorithm parameters u_{min} = 24 h, s_{min} = 0.2 mm. Thresholds correspond to the maximum performance in terms of True Skill Statistic. The plots show outcomes relative to a) reference RS0, and b-d) various erroneous reporting scenarios (RS1, RS2, RS3).

Figure 6: Scatter plot, in the double-logarithmic rainfall duration-intensity plane, of triggering and non-triggering events for *daily* data and separation algorithm parameters u_{min} = 1 day, s_{min} = 0. Thresholds correspond to the maximum performance in terms of True Skill Statistic. The plots show outcomes relative to *a*) reference RS0, and *b-d*) various erroneous reporting scenarios (RS1, RS2, RS3).

Figure 7: Scatter plot, on the double-logarithmic rainfall duration-intensity plane, of triggering and non-triggering events for daily data with aggregation shift as in the Italian rainfall databases. Separation algorithm parameters are: u_{min} = 1 day, s_{min} = 0 mm. Thresholds correspond to the maximum performance in terms of True Skill Statistic. The plots show outcomes relative to *a*) reference RS0, and *b-d*) various erroneous reporting scenarios (RS1, RS2, RS3).

Figure 8: Comparison of thresholds, calibrated in the various scenarios and event identification parameters, with the correct hourly dataset. Thresholds determined with *a*) hourly and *b*) daily data (both correct and with aggregation shift), are distinguished. Correct thresholds are relative to the following event identification parameters: u_{min} = 24 h, s_{min} = 0.2 mm, and u_{min} = 1 da, s_{min} = 0 mm, for hourly and daily data respectively. These plots are representative of how thresholds calibrated with uncertain information of triggering rainfall data may perform in early warning systems that use high quality rainfall and landslide monitoring.

List of tables


Fiumedinisi
raingauge
N
Legend
Elevation [m a.s.l.]
1300
0
Peloritani area
0 5 10
km
38°10'N
38°0'N
37°50'N
37°40'N
15°10'E
15°20'E
15°30'E
15°40'E

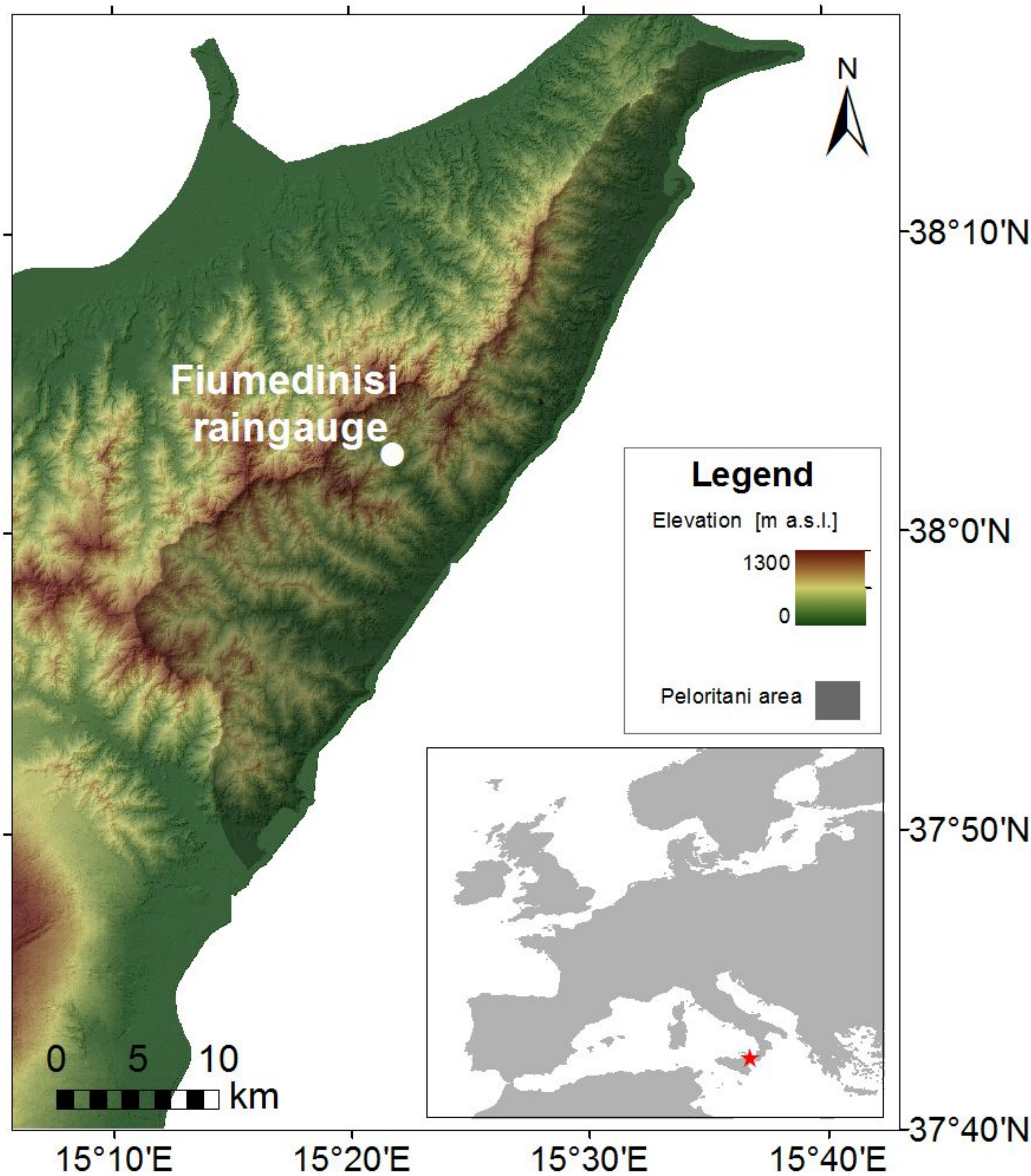


Figure 1: Location of the Peloritani Mountains area in Sicily, Italy, and of the Fiumedinisi rain gauge.

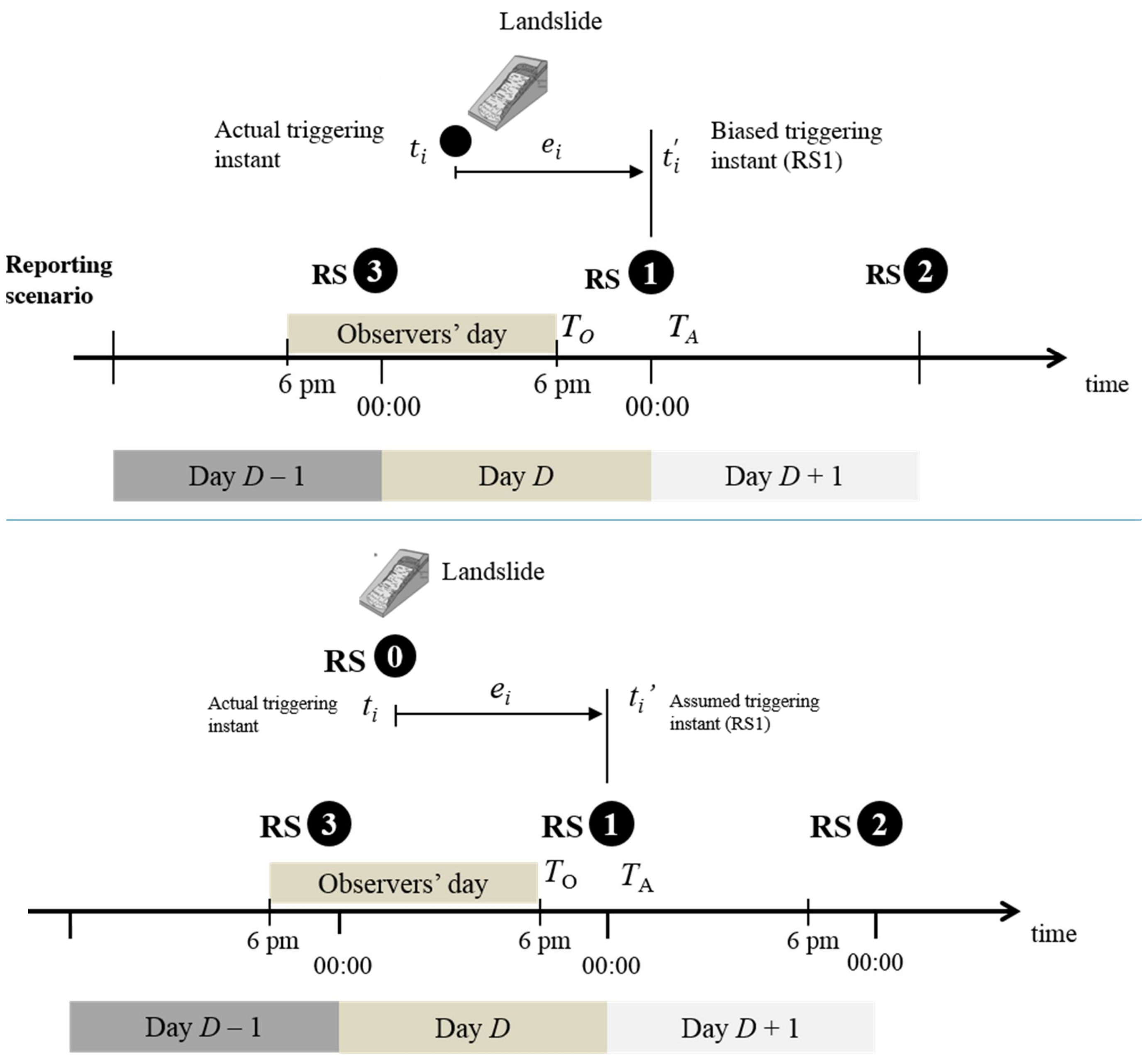


Figure 2: Sketch illustrating simulation of uncertainty in triggering instants likely present in landslide inventories built from newspapers or similar sources. The black numbered circles indicate one of the reporting scenarios (RS), ~~each inducing~~which may induce a random errors $e = t' - t$ in ~~the~~ landslide triggering instants. In particular, a landslide that occurs within the observers' day, is reported at day D and attributed to the end of the same day (small delay reporting scenario, RS1) or to its beginning (anticipated reporting scenario, RS3). It can be reported also at day D+1 and th~~us~~en attributed to the end of it (large delay reporting scenario RS2). These scenarios can be described in terms of two parameters: T_O = the ending hour of observers' day, and T_A = the triggering instant, referred to hours 00:00 of day D, assumed by an analyser who interprets the newspaper-like information.

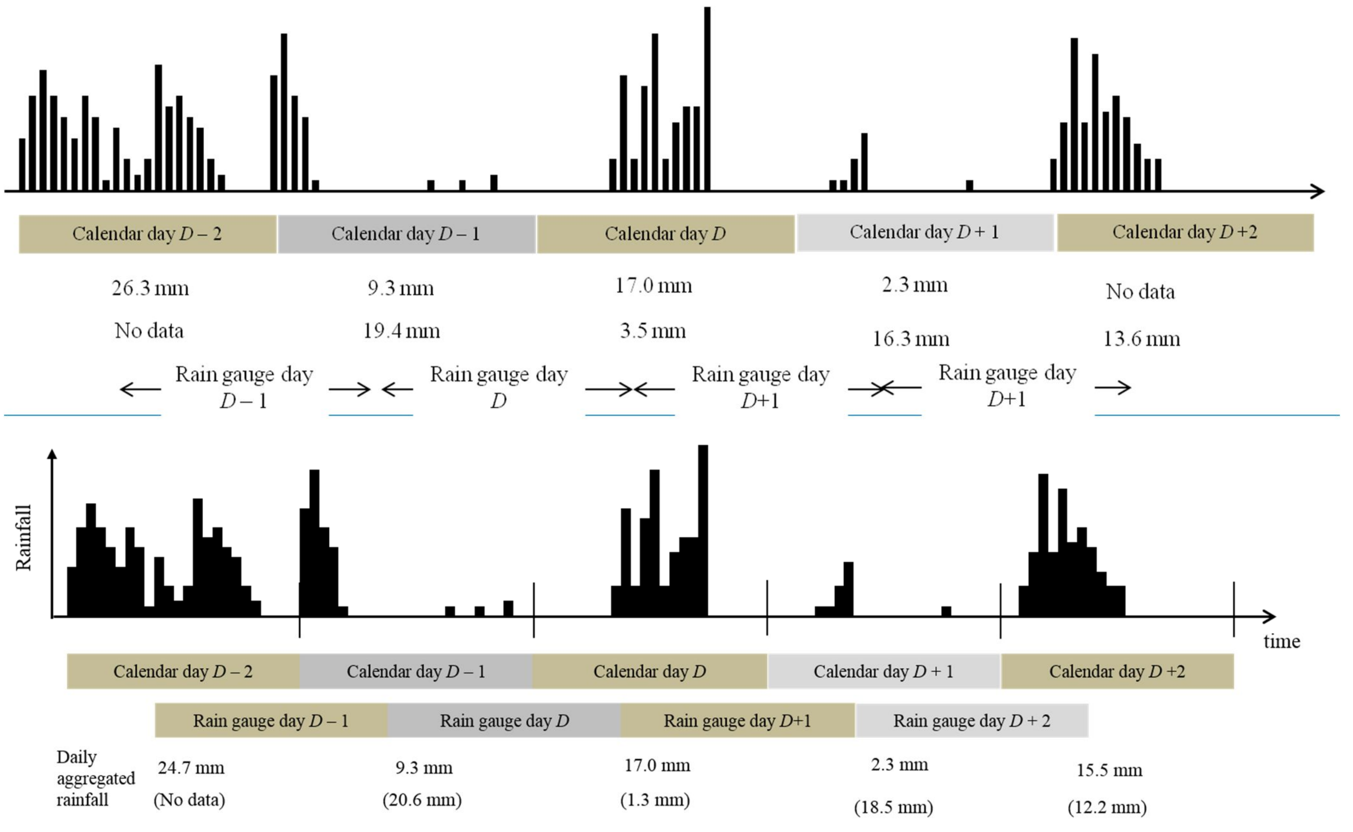


Figure 3: Aggregation of rainfall data from ~~the~~ hourly to ~~the~~ daily time scale: daily rainfall depths on the top row result from correct aggregation; those on the bottom row from shifted aggregation, as ~~present in the Italian Hydrological Bulletins data~~occurs for the Italian Hydrological Bulletins (Annali Idrologici). The shift is due to manual collection of data in early decades of operation of the monitoring network; the presence of the shift is still continued, in spite of installation of automatic rain gauges, to preserve homogeneity of the entire historical time series.

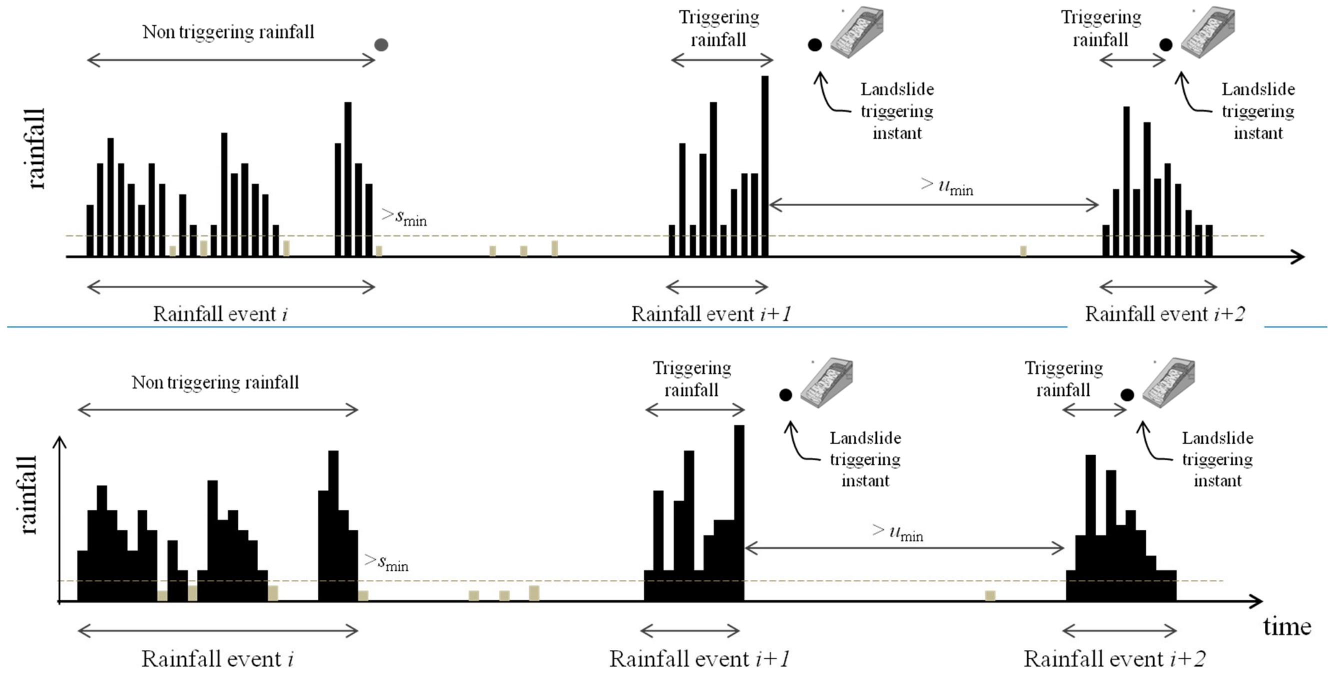
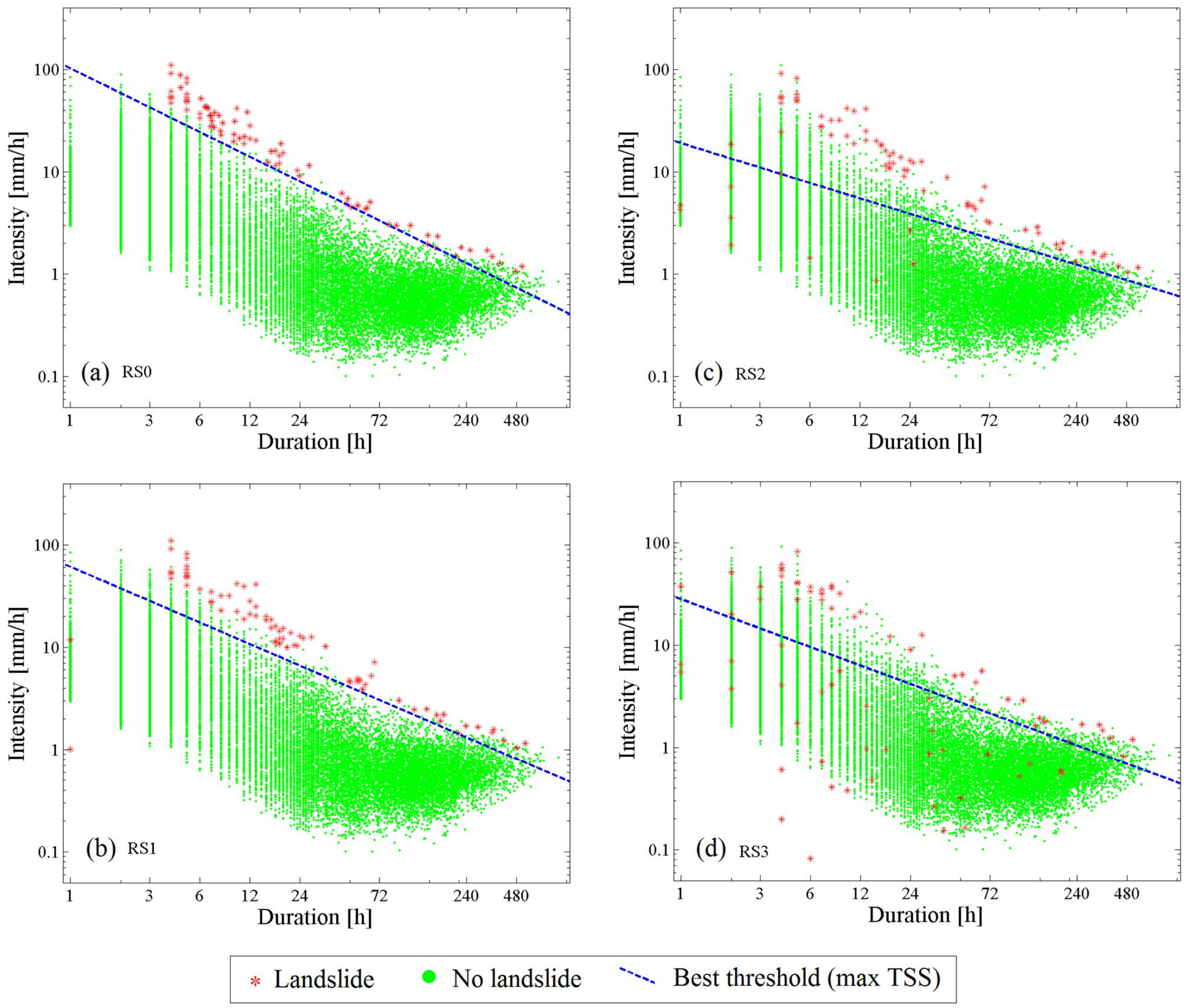


Figure 4: Sketch illustrating the algorithm for the identification of triggering and non-triggering rainfall events, and relative parameters s_{min} and u_{min}. When a landslide is triggered in a dry period, it is attributed to the whole event preceding it; otherwise, only the part of the event preceding the landslide triggering instant is considered. For non-triggering rainfall (the first one in the sketch), duration and intensity are computed ~~referring to~~considering the ~~whole~~entire rainfall event.



(a) RS0
(b) RS1
(c) RS2
(d) RS3
Intensity [mm/h]
Duration [h]
0.1
1
10
100
1
3
6
12
24
72
240
480
Landslide
No landslide
Best threshold (max TSS)

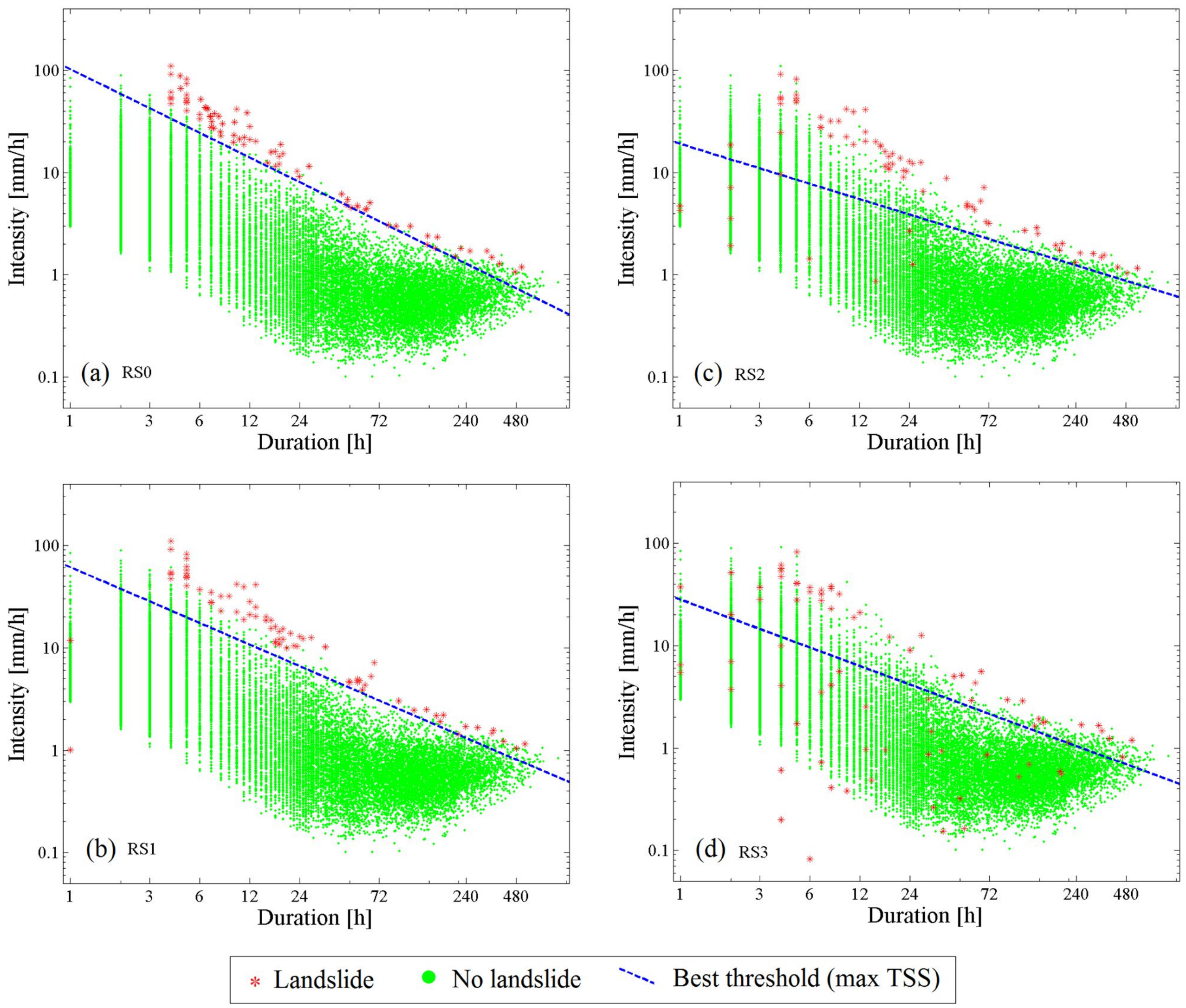


Figure 5: Scatter plot, in the double-logarithmic rainfall duration-intensity plane, of triggering and non-triggering events for *hourly* data and separation algorithm parameters $u_{\min}$ = 24 h, ~~and~~ $s_{\min}$ = 0.2 mm~~/hour~~. Thresholds correspond to the maximum performance in terms of True Skill Statistic. The plots show outcomes relative to a) reference RS0, and b-d) various erroneous reporting scenarios (RS1, RS2, RS3).

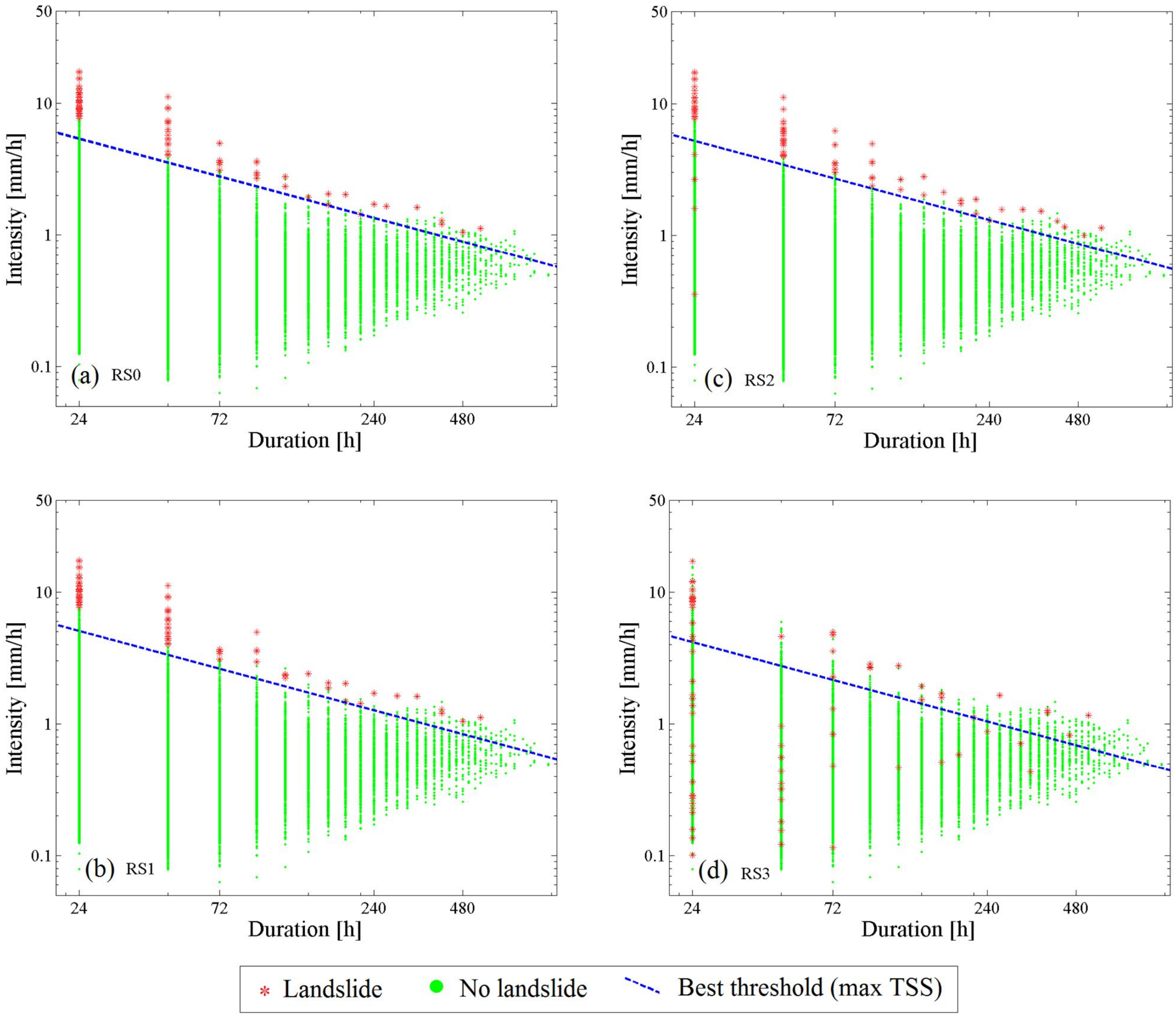


Figure 6: Scatter plot, in the double-logarithmic rainfall duration-intensity plane, of triggering and non-triggering events for *daily* data and separation algorithm parameters $u_{\min}$ = 1 day ~~and~~, $s_{\min}$ = 0 ~~mm~~. Thresholds correspond to the maximum performance in terms of True Skill Statistic. The plots show outcomes relative to *a*) reference RS0, and *b-d*) various erroneous reporting scenarios (RS1, RS2, RS3).

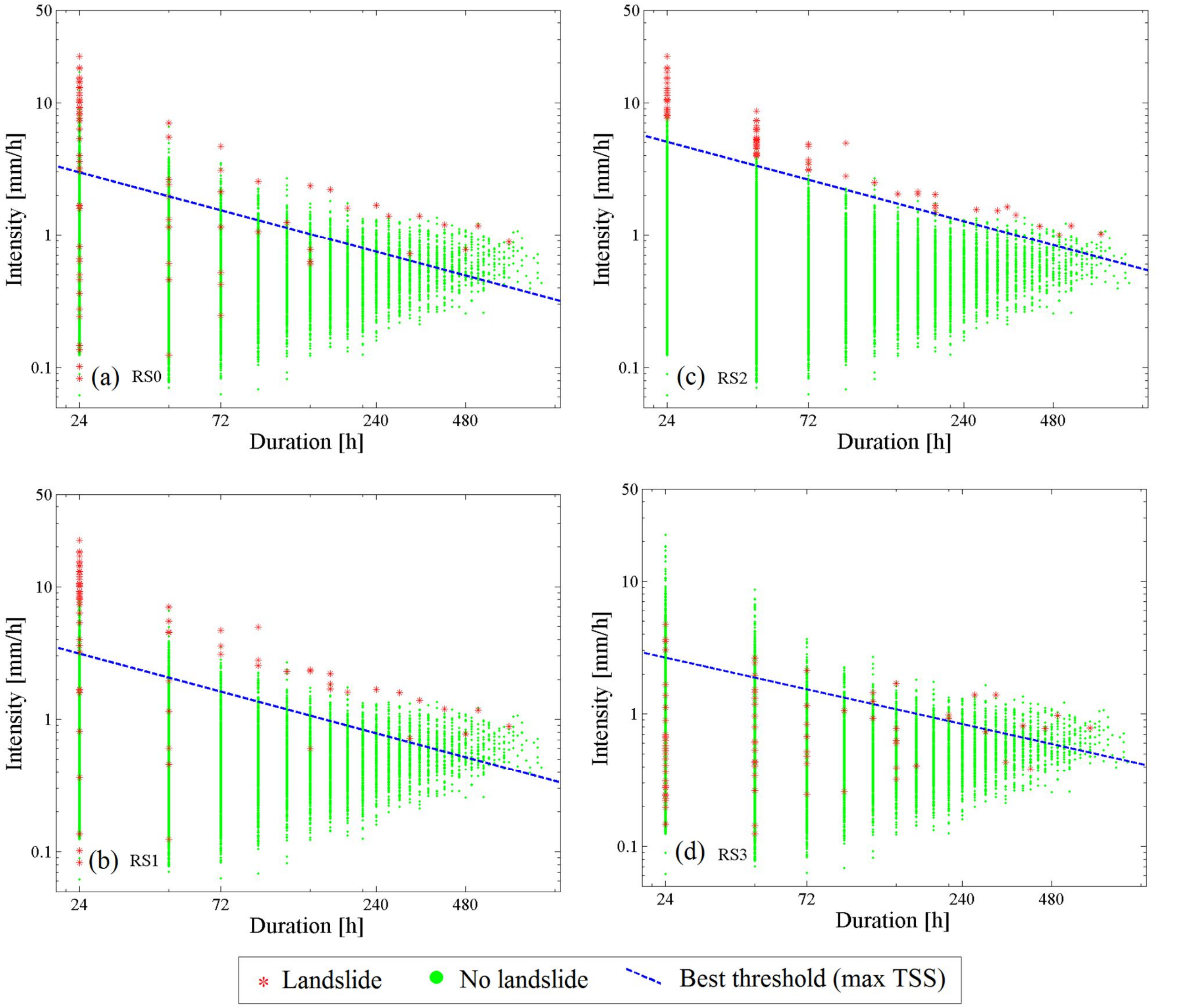
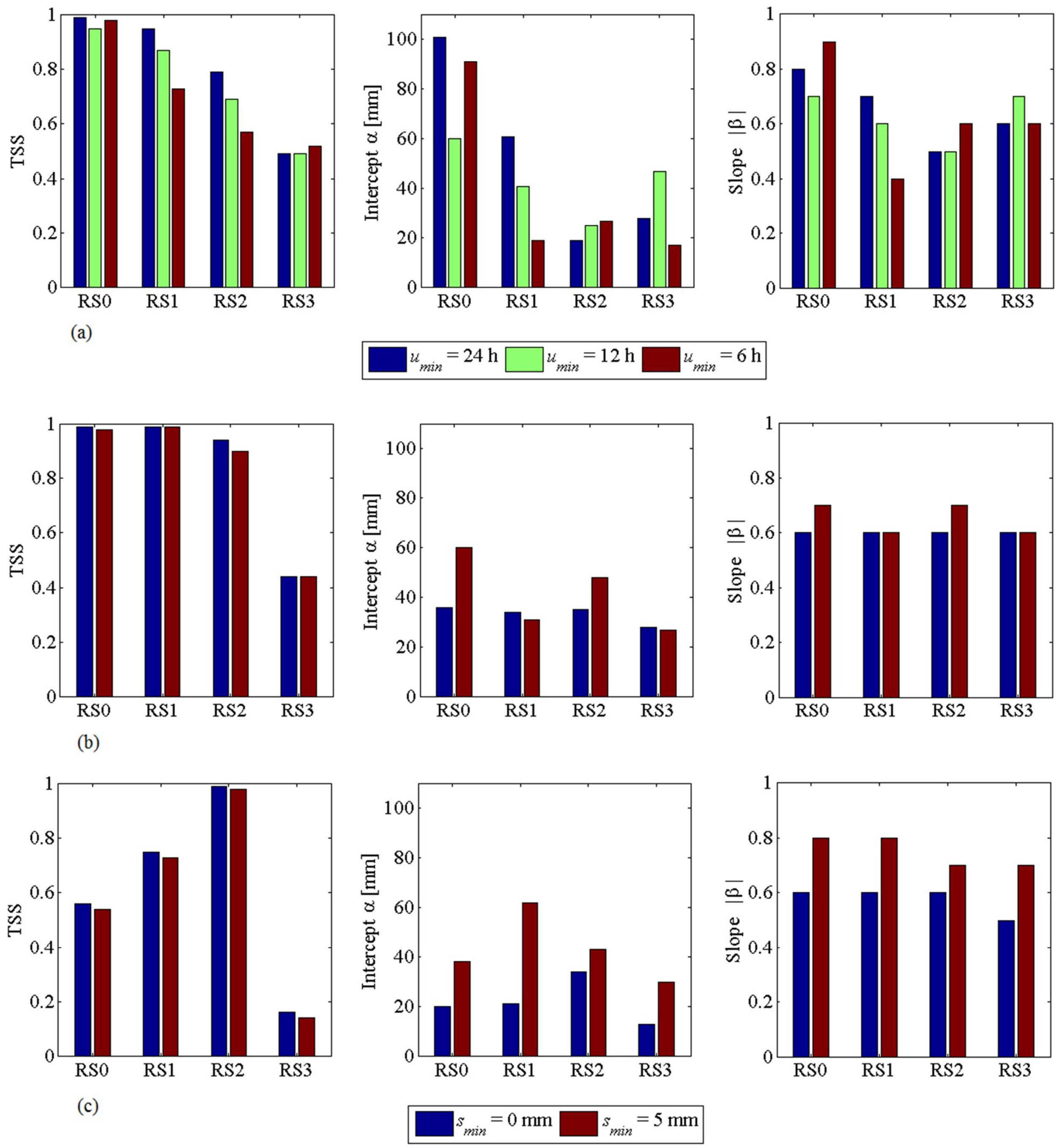
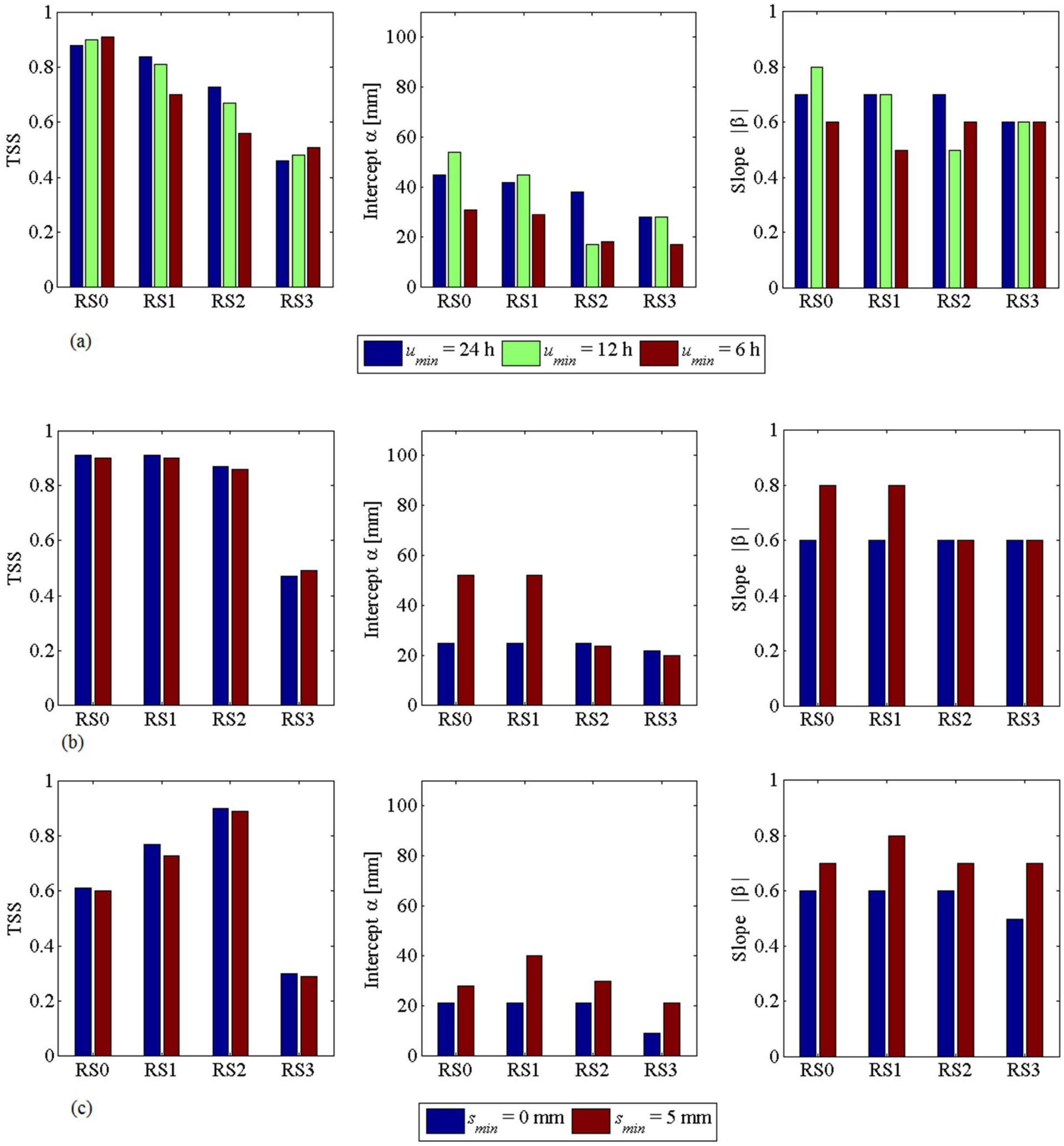


Figure 7: Scatter plot, ~~in~~on the double-logarithmic rainfall duration-intensity plane, of triggering and non-triggering events for daily data with aggregation shift as in the Italian rainfall databases. Separation algorithm parameters are: u_{min} = 1 day~~ and~~, s_{min} = 0 mm. Thresholds correspond to the maximum performance in terms of True Skill Statistic. The plots show outcomes relative to *a*) reference RS0, and *b*-*d*) various erroneous reporting scenarios (RS1, RS2, RS3).



~~Figure : Threshold calibration results for all simulations. Plots show the value of the maximum TSS, and of best-threshold intercept α and slope β parameters, for different rainfall event identification algorithms and datasets: a) hourly resolution data, b) daily resolution and c) daily resolution rainfall data with aggregation shift errors. Case of nulled effects of antecedent precipitation (τ_M = 0).~~



~~Figure : Same as Figure 8 but taking into account the presence of pressure head memory (recession constant τ_M = 2.7 days).~~

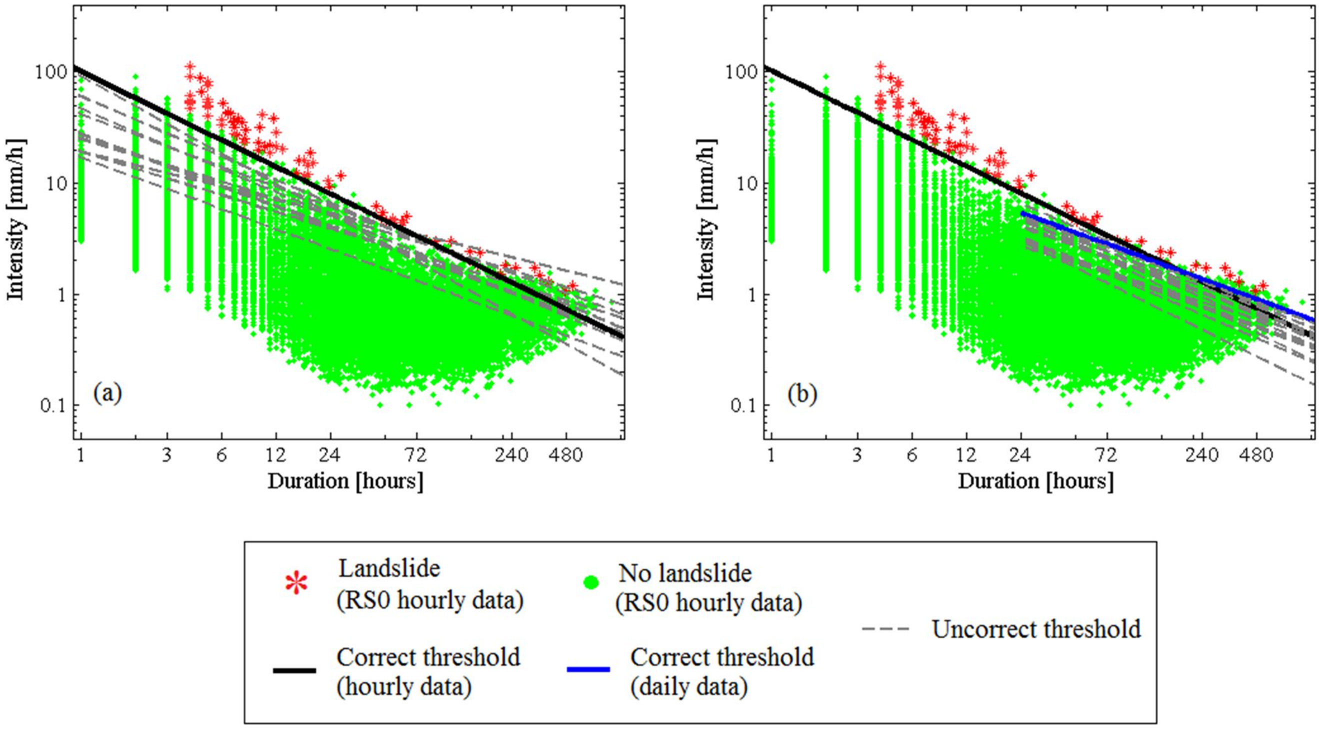


Figure 8: Comparison of thresholds, calibrated in the various scenarios and event identification parameters, with the correct hourly dataset. Thresholds determined with *a*) hourly and *b*) daily data (both correct and with aggregation shift), are distinguished. Correct thresholds are relative to the following event identification parameters: u_{min} = 24 h, s_{min} = 0.2 mm, and u_{min} = 1 da, s_{min} = 0 mm, for hourly and daily data respectively. These plots are representative of how thresholds calibrated with uncertain information of triggering rainfall data may perform in early warning systems that use high quality rainfall and landslide monitoring.

Table 1: Soil and morphological properties of a representative hillslope in the Peloritani Mountains area, Sicily, Italy (after Peres and Cancelliere, 2014).

Variable	Units	Value
Soil friction angle ϕ'	[°]	37
Soil cohesion c'	[kPa]	5.7
Unit weight of soil γ_s	[N/m^3]	19000
Saturated soil water content θ_s	[–]	0.35
Residual soil water content θ_r	[–]	0.045
Saturated soil hydraulic conductivity K_s	[m/s]	0.00002
Saturated soil hydraulic diffusivity D_0	[m^2/s]	0.00005
Gardner soil characteristic curve parameter α_0 (*)	[1/m]	3.5
Soil depth d_{LZ}	[m]	2
Terrain slope δ	[°]	40
Basal drainage leakage ratio c_D (*)	[–]	0.1

(*) See Baum et al. (2010) for details

Table 2: Some characteristics of the ideal Monte Carlo simulation dataset.

Variable	**Value**
Number of simulated years	1000
Number of rainfall events	19 826
Number of landslide events for $\tau_M = 0$	81
(return period)	12.3
Number of landslide events for $\tau_M = 2.$~~7~~75	115
(return period)	8.7

Table 3: Some rainfall event identification algorithms found in the literature.

Reference	~~**Discretization**~~ **Aggregation**	**Algorithm parameters**	
		s_{min}	u_{min}
Pizziolo et al., 2008	daily	5 mm	1 day
Berti et al., 2012 (*)	daily	2 mm, *or* 1mm, *or* 2/3 mm	1 day *or,* 2 days *or,* 3days
Rappelli, 2008	hourly	1 mm	12 h
Melillo et al., 2015; Vessia et al., 2014 (**)	hourly	0.2 mm	3 h
			6h
Saito et al., 2010	hourly	1 mm	24 h
Segoni et al., 2014a, 2014b	hourly	0	$u_{\mathrm{min}} = 10 \div 36$ h selected so that threshold performances were optimized
Brunetti et al., 2010; Peruccacci et al., 2017	sub-hourly	0 mm	2 days (May-Sept) 4 days (Oct-Apr)
Peres and Cancelliere, 2014	hourly	0.2 mm	24 h
Nikolopoulos et al., 2014	hourly	0.2 mm	24 h

(*) More precisely "the algorithm scans a rainfall time series and detect the rainfall events using a moving-window technique: a new event starts when the precipitation cumulated over D_T days exceeds a certain threshold E_T, and ends when it goes below this threshold. For instance, if $D_T = 3$ days and $E_T = 2$ mm, the rainfall event starts when the cumulative rainfall exceeds 2 mm in 1, 2, or 3 days (that is if 2 mm are exceeded on the first day, the rainfall starts at day 1). Then, the rainfall event stops when it rains less than 2 mm in 3 days; the end of the event is defined as the last of the three days in which the rainfall is greater than zero". $D_T = 3$ days and $E_T = 5$ mm were chosen.
(**) The algorithm can be only approximately expressed in terms of s_{min} and u_{min}. In particular, the algorithm additionally excludes "sub-events" having a total event rainfall below a seasonally variable threshold

Table 4: Set-up of the numerical experiments. Each set of algorithm parameters is considered for the four hypothesized landslide reporting-scenarios.

Aggregation	Event identification algorithm parameters
Hourly	u_{min} = 24 h, s_{min} = 0.2 mm
	u_{min} = 12 h, s_{min} = 0.2 mm
	u_{min} = 6 h, s_{min} = 0.2 mm
Daily correct and daily shifted (Italian database)	u_{min} = 1 day, s_{min} = 0 mm/day
	u_{min} = 1 day, s_{min} = 5 mm/day

Table 5: Confusion matrix for evaluation of landslide-triggering thresholds (assumed here to be of the ID type: $I = f(D)$).

		Actual	
		Landslide (POS = TP + FN)	No landslide (NEG) (NEG = FP + TN)
Predicted	Landslide (POS'): $I \geq f(D)$ (POS' = TP + FP)	true positives, TP	false positives, FP
	No landslide (NEG'): $I<f(D)$ (NEG' = FN + TN)	false negatives, FN	true negatives, TN

Table 6: Threshold calibration results for all simulations, in the case of nulled effects of antecedent precipitation ($\tau_M = 0$).

			RS0			RS1			RS2			RS3		
Aggregation	u_{min} [h]	s_{min} [mm]	TSS	α [mm/h]	β	TSS	α [mm/h]	β	TSS	α [mm/h]	β	TSS	α [mm/h]	β
Hourly	24	0.2	0.99	101	0.80	0.95	61	0.70	0.79	19	0.50	0.49	28	0.60
	12	0.2	0.95	60	0.7	0.87	41	0.6	0.69	25	0.5	0.49	47	0.7
	6	0.2	0.98	91	0.9	0.73	19	0.4	0.57	27	0.6	0.52	17	0.6
Daily	24	0	0.99	36	0.6	0.99	34	0.6	0.94	35	0.6	0.44	28	0.6
	24	5	0.98	60	0.7	0.99	31	0.6	0.9	48	0.7	0.44	27	0.6
Daily (Shifted)	24	0	0.56	20	0.6	0.75	21	0.6	0.99	34	0.6	0.16	13	0.5
	24	5	0.54	38	0.8	0.73	62	0.8	0.98	43	0.7	0.14	30	0.7

Table 7: Threshold calibration results for all simulations, when antecedent precipitation memory is present (τ_M = 2.75 days).

			RS0			RS1			RS2			RS3		
Aggregation	u_{min} [h]	s_{min} [mm]	TSS	α [mm/h]	β	TSS	α [mm/h]	β	TSS	α [mm/h]	β	TSS	α [mm/h]	β
Hourly	24	0.2	0.88	45	0.7	0.84	42	0.7	0.73	38	0.7	0.46	28	0.6
	12	0.2	0.9	54	0.8	0.81	45	0.7	0.67	17	0.5	0.48	28	0.6
	6	0.2	0.91	31	0.6	0.7	29	0.5	0.56	18	0.6	0.51	17	0.6
Daily	24	0	0.91	25	0.6	0.91	25	0.6	0.87	25	0.6	0.47	22	0.6
	24	5	0.9	52	0.8	0.9	52	0.8	0.86	24	0.6	0.49	20	0.6
Daily (Shifted)	24	0	0.61	21	0.6	0.77	21	0.6	0.9	21	0.6	0.3	9	0.5
	24	5	0.6	28	0.7	0.73	40	0.8	0.89	30	0.7	0.29	21	0.7