August 11, 2014

To Executive Editors Erwin Zehe, Alberto Guadagnini & Hubert H.G. Savenije

To editor Thom Boogard

To the four anonymous reviewers of the manuscript and SL Gariano

Hydrology and Earth System Sciences

Re: Revised Manuscript "Derivation and evaluation of landslide triggering thresholds by a Monte Carlo approach" by David J. Peres and Antonino Cancelliere

Dear Editors,
We appreciate the efforts you and the reviewers have invested in our manuscript (MS). We are pleased to submit a revised version in response the reviewers' insightful comments.
We have responded to reviewer comments in the discussion phase, here we just indicate point by point the way that the MS has been modified to respond to the comments by reviewers. Referee comments are in *Arial* italics, our responses are in normal Times New Roman font.

Comments by Referee#1

This is an interesting paper, that combines in an innovative way existing modeling components for the construction of synthetic time series of rainfall, for rainfall infiltration and the build up of a water table in a slope, and for slope stability analysis. Overall, I have found the paper sound and of potential interest to a large audience of hydrologists and geomorphologists. However, there are quite a few issues that need to be clarified in the paper. I have divided my comments into "editorial", general comments and specific comments. The latter are keyed to page and line numbers. The distinction does not imply a ranking of the importance or relevance of the comments. It should not be too difficult for the authors to address my comments, and respond to my queries.

Thank you for your interest in the paper and you comments.

Overall, the paper reads well, but in places the text is difficult to read, and follow. Other parts of the text are not really necessary, and could be deleted, shortened, or moved to an appendix. As an example, the description of the single modelling components (rainfall time series, TRIGRS modelling, slope stability modelling), does not add much to what is already known in the literature. The description of these model components should be shortened, or moved in specific appendixes, where they can be properly described.

We have extended the description of the Monte Carlo methodology and moved the model-components sections into appendixes.

The Abstract and the Conclusions are not fully clear, and should be rewritten. In the Abstract should state briefly the purpose of the research, the principal results and the major conclusions, and must be able to stand by itself. In the Conclusions, the authors should list only the main conclusions and relevant findings of their work.

We have rewritten the abstract and the conclusions.

In the text, the authors use the term "hyetograph". A hyetograph is a graphical representation of the distribution of rainfall over time i.e., a rainfall record or a rainfall time series. Indeed, for their work the authors have used rainfall records, and not hyetographs. This should be clarified throughout the text.

Where the term “hyetograph” appears for the first time we have briefly explained its meaning in our work: “In fact, rainfall events represented by the same pair (*D*, *I*) may correspond to totally different event time-histories (hyetographs) that thus may or may not result in triggering.”

In section 3, the definition of the threshold and the evaluation of the performance of the threshold should be separated in two different sections, or sub-sections at least.

Section 3 is now split into two subsections.

Quality of the figures should be improved. Text and labels in the figures are small, and can be difficult to read. When modifying the figures the authors should consider their final size in the journal, considering that figures can occupy one or two columns. Some of the figures (e.g. Figures 4, 5) are very small, and difficult to read. Parts of the charts (e.g., in Figure 4) do not show data. These parts can be removed from the charts, to make them larger. Authors should consider that use of colours is possible, I believe with no extra cost, in HESS. Some of the figures may improve significantly is colours are used.

Figures have been modified according to referee suggestions to improve their readability.

Abstract, Line 1, “Rainfall thresholds are the basis of early warning systems able to promptly warn about the potential triggering of landslides in an area”. This is a vague, and partly misleading statement. First, rainfall thresholds are not the only basis for landslide early warning systems. Second, “prompt” warning or alarm is not base solely on the exceedence (or not exceedance) of a threshold. A substantial amount of human judgment is involved in giving a warning, or an alarm. [Note that the same comment holds for the first sentence of the Introduction].

The sentence to which referee refers has been modified as follows:
In the abstract: “Assessment of landslide triggering rainfall thresholds plays a key role in early warning systems aimed at warning the potential occurrence of landslides in prone areas.”
In the introduction: “Rainfall thresholds indicating landslide triggering play a key role within early warning systems aimed at warning the potential occurrence of landslides in prone areas.”

Introduction, page 2761, lines 19, 20. “Reliability of thresholds derived by the analysis of observed data is generally limited by the quality and availability of such data”. There are other factors that influence the reliability of rainfall thresholds, including the uncertainty associated with the definition of the thresholds. See e.g. doi:10.1016/j.geomorph.2011.10.005.

We have modified the sentence as follows:
“Many factors of uncertainty affect the reliability of empirical thresholds, such as rainfall temporal and spatial variability, uncertain knowledge of the triggering instants, simplicity of threshold equation that does not include all control variables and statistical issues as well (Peruccacci et al., 2012). Nonetheless, it can be argued that most of the uncertainty stems from the availability and quality of the data used to derive the thresholds (Glade et al., 2000; Berti et al., 2012).”

Introduction, page 2761, line 27. “ . . . critical duration D . . .”. The authors should be aware that “critical” is used with different meanings in the literature related to the definition of rainfall thresholds. See e.g., doi:10.1016/j.enggeo.2004.01.007, Govi and Sorzana (1980), Heyerdahl et al (2003), doi: 10.1007/s00703-007-0262-7.

At the first time “critical” appears we have added a specification of its meaning within our MS:
“This has a direct consequence on threshold derivation, because critical (where ”critical” here means ”corresponding to landslide triggering”)”

Introduction, page 2763, lines 16-17. " . . . thus casting some doubts on the use of parametric power-law as a proper functional form in deriving rainfall thresholds.". A conclusion of the work is that the power law threshold model fits well the modeled data. The authors should comment this finding, in view of the findings of e.g., Rosso et al., 2006; Salciarini et al., 2008.

Manuscript has been modified in several points to more clearly explain when and why the ID power law fits well modeled data (Rosso et al., 2006; Salciarini et al., 2008 do not analyse the effect of rainfall intensity temporal variability within events). In the point indicated by the reviewer the following text has been added:
"In other words, because such thresholds were derived from a physically-based model, this may be interpreted as an evidence that the use of the power law form is not supported by a physically-based standpoint. Nevertheless, in such studies the meteorological aspects were analysed in a simplistic way, because the thresholds do not consider variability of rainfall intensity during events and the initial conditions are not computed as a function of rainfall time-history preceding the current event.".

Introduction, page 2764, line 4. "For this last point we adopt a precise rainfall identification criterion." Precise criterion? What does this mean? It seems to me that the authors have established a criterion, and have used it. There is no evidence that this criterion is accurate (or precise).

The sentence has been removed because we feel that it is indeed a level of detail not needed in the introduction

Introduction, page 2764, line 18. "From their study . . ." Unclear how is "they".

The sentence has been completed as follows:
"From their study they conclude that beta-shaped non-uniform hyetographs have a stronger destabilizing effect than uniform hyetographs of the same volume, since associated return period of $FS \leq 1$ resulted higher in this last case."

Introduction, page 2764, line 21. Provide references for the NSRP model.

References for the NSRP model were added: Neyman and Scott, 1958; Kavvas and Delleur, 1975; Cox and Isham, 1980; Rodriguez-Iturbe et al., 1987a

Monte Carlo modelling, page 2765, lines 8-9, "A stochastic rainfall model, calibrated on observations at a representative site, is used to generate a 1000-years long hourly rainfall time series." This is a tricky point that may ingenerate some confusion. The rainfall time series is "virtual", as it is its length of 1000 years. It is not a climatic series that actually spans a 1000-year period. The difference is significant. As far as I can tell, no climatic information was used to generate the series, and the series was generated ignoring known of possible changes in the climatic signal. This should be clarified by the authors.

This issue has been clarified at the end of the "Rainfall stochastic model" section by the following sentence:
"Though calibration is conducted taking into account seasonality, by calibrating the model separately for the various homogeneous season within the year (section 4.2), it is noteworthy to point out that the generated series is globally stationary, and consequently eventual annual non-stationarity due to climate change, is not taken into account. In other words, the generated series represents possible realizations of the rainfall events under current climate conditions, the final aim being of deriving thresholds suitable under the present climate and not to assess how climate change may affect them."

Monte Carlo modelling, page 2765, line 14. Why have the authors decided to use a day (24 h) to separate rainy from dry (non-rainy) periods? Monte Carlo modelling, page 2765, line 20. What is this "basal boundary"?

At the end of section "Monte Carlo synthetic data generation" the following explanation has been added to better explain why we have choose 24 hours to define individual rainfall events:
"Regarding the choice of the inter-event time $\Delta t_{\min}$ – an issue that is the focus of some works in literature (e.g. Restrepo-Posada and Eagleson, 1982; Bonta and Rao, 1988) – we have followed an approach analogous to that used by Balistrocchi et al. (2009) and Balistrocchi and Bacchi (2011), for which the inter-event time may be assumed as the minimum time needed to avoid overlapping of the response produced by two subsequent rainfall events. To this end we considered that the temporal peak of pressure head due to an individual rainfall event may be reached, as mentioned above, at an instant significantly after rainfall ceases. Hence a criterion for selecting the inter-event time has been that of choosing a value that approximates the dry time interval that contains the peak pressure head response relatively to all the N_{RE} simulated rainfall-events. In our case, from preliminary simulations a $\Delta t_{\min}$ = 24 h appeared suitable for the hydraulic and geotechnical soil properties which are considered in this work (see Sect. 3)."
Please note the related sentence added at section "3.1-Triggering and non-triggering rainfall identification":
"From a general standpoint, this procedure may be disconnected from the way event separation has been performed to compute the triggering instants with the methods described in Sect. 2. Nevertheless, for consistency with the event separation criterion that is considered in Monte Carlo simulation methodology, it is preferable that the procedure for identification of triggering and non-triggering events is based on the same inter-event time $\Delta t_{\min}$ used in Monte Carlo simulations."

"Basal boundary" has been changed into "soil-bedrock interface", which is more understandable.

Monte Carlo modelling, page 2766, lines 21-22. What is this "storm origin"? This is not at all clear? "Storms" and "rainfall" can be very different, and the authors should make this clear.

Though use of the term "storm" is correct since it is reported in several works on stochastic models, to avoid confusion with commonly-used meanings of words like "storm" and "rainfall", we have modified the MS in this part by changing "storms" into "clusters".

Monte Carlo modelling, page 2767, line 3. What is this "cell origin"? Again, this is not clear. Monte Carlo modelling, page 2767, line 3. What is this "cell origin"? Again, this is not clear.

The term "cluster" was used instead of storm and "cell origin" was changed into "cluster origin". Some other few details have been added to section "rainfall stochastic model" to better explain the concepts of stochastic modeling of rainfall by the NSRP model.

Monte Carlo modelling, page 2768, line 4. Say something about the "assumptions of Rosso et al. (2006)."

This part has been modified as follows to explain assumptions of Rosso et al. 2006:
"This latter model is derived from a mass conservation equation of soil water coupled with the Darcy's law used to describe seepage flow, where for simplicity the soil volumetric strain is neglected (the variation of porosity with pressure head is assumed null). A similar conceptualization is the basis of the model proposed by Rosso et al. (2006)."

Monte Carlo modelling, page 2769, line 1. "The ratio A/B is the well-known specific upslope contributing area A/B" ? Text is unclear. What does it mean that A/B is . . .A/B?

This part has been modified as follows:
"The ratio *A/B*, which can be computed based on a Digital Terrain Model (DTM), is the well-known specific upslope contributing area, an important variable on which the topographic control on shallow landslide triggering depends (Montgomery and Dietrich, 1994). For instance, it is $A/B = BN_d$, where N_d is the number of cells draining into the local one, if one determines flow paths via the non-dispersive single direction (D8) method (O'Callaghan and Mark, 1984). Other methods consider multiple flow directions (cf. Holmgren, 1994)."

Threshold derivation, page 2771, lines 4-6. "For a hillslope of given properties, Monte Carlo simulations lead to a series of computed failures, i.e. time instants at which the factor of safety drops below the value of 1." Although the statement is correct, I see a conceptual potential problem. It is well known that so called "deterministic models", including TRIGRS, underestimate the stability conditions of the areas where they are applied. Indeed, it is common that the application of TRIGRS in a study area results in widespread (predicted) instability, which does not match the actual (real) abundance of event landslides (that is typically less, or much less than what is predicted by TRIGRS, or other similar models). Reasons for this behaviour are manifold, and not really important for this study. What is important is the fact that if TRIGRS predicts "instability" (e.g. FS < 1), not necessarily a slope failure occurs. This has an impact on the number of true positive and true negatives, and the related analyses. The author should comment this issue in their work.

The following sentence was added to the paper to take into account the issues mentioned by the reviewer:
"A triggering rainfall may be associated to each down-crossing, though it is noteworthy to point out that some uncertainty is present in the link between the actual failure of the slope and its theoretical instability. Nevertheless, following several works in literature (e.g., Iverson, 2000; Rosso et al., 2006; Baum et al., 2010) this uncertainty has been not taken into account here, though it may affect at a certain degree the way that rainfall events are classified as triggering and non-triggering and the subsequent ROC-based analysis (Sect. 3.2)."

Threshold derivation, pages 2771-2. I can think of additional reasons for the scatter of empirical points in a I/D chart that the two listed by the authors. One is the natural variability of rainfall induced landslides. It is well known that a simple (possibly too simple) threshold model, like the I,D model adopted in this work, cannot capture the (large!) natural variability and complexity of rainfall induced landslides. Stating that (only) two factors control the joint presence of I/D events that have and have not resulted in landslides in an I/D log-log plot is too simplistic, and (partly) misleading.
The following sentence was added about this issue:
"It is noteworthy to highlight that the real uncertainty associated with this threshold generally yields different – likely worse – performances of that assessed here, since uncertainty factors are more than the ones related to the stochastic nature of rainfall listed at the beginning of this section."

ROC analysis, pages 2771-2774. Most of what is written in this section of the text is not really new, or innovative. This part of the text can be shortened considerably. In the very (very!) large literature on the assessment of forecasts using contingency tables, the zillions of related performance indexes, ROC plots, etc., there is indeed confusion and significant overlaps. Indeed, this does not help. However, attempts to limit the confusion exist. The authors should consider doi: 10.1175/WAF1031.1 and specifically doi: 10.1175/2009WAF2222300.1. The performance index "Delta" used by the authors is not really new, but well known in the literature. I have notices that a public comment to this paper has pointed out the issue already. The authors are advised to check the references listed in this public comment http://www.hydrol-earthsyst-sci-discuss.net/11/C624/2014/hessd-11-C624-2014.pdf.

The ROC analysis part has been shortened taking into account the comments received by S.L. Gariano. Paper has been modified in all necessary parts. It is ironic to see that we are not alone in rediscovering "old" indexes (see interesting review paper by Murphy 1996)!

ROC analysis, pages 2773, line 21. " . . . model deterministic thresholds . . .". I have a problem with this definition of a "deterministic threshold". A (pseudo-)deterministic model like the one used in this work, does not result in a "deterministic" threshold, necessarily. The authors should clarify the meaning of "deterministic threshold", and if their definition implies that the threshold is accurate, clear-cut, or fuzzy. This is crucial for the paper.

Some modifications in the terminology were given in the new manuscript. In particular we distinguish between "stochastic-input physically-based thresholds" and "constant-intensity input physically-based thresholds". This should help to avoid confusion to the readers.

Investigated area and data, page 2774, lines 7-10. More information should be given on the location and size of the study area. Figure 3 is not really sufficient. Accurate or approximate location of the considered landslides should be shown in Figure 3. Depending on the location of the landslides, use of a single rain gauge may be reasonable, or not. This should be clarified.

The Section 4 presents detailed information on the case-study area, including an approximate location of the considered landslides:
"The area has been hit by highly-damaging diffused shallow landslides, in the last decade.
Precisely, widespread landslide events occurred in this area on: (I) 15 September 2006 (areas #4, #5, #6 and #7), (II) 25 October 2007 (area #7), (III) 24 September 2009 (areas #1, #2, #3, #4, #5) and (IV) 1 October 2009 (area #7). The areas indicated into brackets have been derived from newspapers archives (cf. e.g. http://gazzettadelsud.virtualnewspaper.it/gdsstorico/), which also present further information on the events."
Moreover detailed location of slides occurred on the last 1 October 2009 event has been added to the MS by Fig. 3. Fiumedinisi raingauge is the only reliable one available in the area for NSRP calibration, and the validation tests of Sect. 5.2 seem to support the reasonability of its use for the area indicated in Fig.2.

Investigated area and data, page 2774, lines 15-17. "Based on a preliminary analysis of monthly statistics, six homogeneous rainfall seasons have been identified: (i) September and October, (ii) November, (iii) December, (iv) January–March, (v) April and (vi) May–August." More should be said on how these "homogeneous rainfall seasons" were identified. Is this statistically significant, given the very short period covered by the rainfall time series? This needs to be clarified.

Figure 4 has been added to better explain the way the homogenous seasons were identified.

Results and discussion, page 2775, line 5. Is the number of 19,826 rainfall events in a (virtual) 1000-year period realistic, or not? This is an "average" of 20 rainfall events per year? Is this reasonable for the study area? How does it compare with the number of real rainfall events in the considered period?

The apparently low number of events derives from cut of under-leakage events (explained now in section MC) and cut of non-rainy seasons. Number of rainfall events per year is practically the same for both the observed and the simulated series when subject to the same preprocessing.
The MS has been modified to clarify this issue:
At section 2 (MC meth.): "Some of the generated rainfall events are removed from the analysis because, according to the hydrological model they will produce no significant variation of pressure head distribution, being their instantaneous (hourly) intensity too low. In particular the events

having maximum intensity less than imin are removed from the analysis. We assume $i_{\min}$ equal to the leakage flux limit, given by $c_d K_S(1-cos^2\ \delta)$, c_d being the vertical leakage ratio, K_S the saturated hydraulic conductivity and δ the slope of the hillslope (see Appendix B)."

At section 4.2 (Rainfall data): "From the assumed inter-event time $\Delta t_{\min}$ = 24 h and soil properties of Tab. 3 the number of rainfall events results N_{RE} = 19826 (in average 19.83 events per year). This number results from initial 28751 events then becoming after cutting the events with hourly intensities always below i_{min} = $c_d K_S(1-cos^2\ \delta)$= 2.975mm/h. These values are statistically comparable to the ones on the observed series (19.18 events/year from 28.91 events/year previously to the cut of under-leakage events)."

At section 5.2 (Validation…) "In particular we have derived from the series the rainfall events with the same criterion adopted in Monte Carlo simulations.Yet the events in the months neglected there (April-August) and the events with intensities below the leakage flow $c_d K_S(1-cos^2\ \delta)$ were not removed here in the observed record, for the test to be unbiased to this preprocessing of data.".

Results and discussion, page 2775, lines 12-13. The count of Positives (N) and Negatives (N) can be misleading. See previous comment: Threshold derivation, page 2771, lines 4-6.

See above response to previous comment

Results and discussion, page 2775, lines 1-2. It is not clear to me how the contributing area, and the specific catchment areas were determined. In spatial modelling, this is usually done exploiting a DEM, of a given resolution. However, this is not the case in this work. This should be clarified.

Figure 3 and section 4 explains how A/B was derived in our work based on a 5 m resolution DTM.

Results and discussion, page 2776, lines 28-29. ". . . but may still be acceptable." Why? What do you mean, exactly?

The expression "still acceptable" has been removed from the MS, and the value of TSS is reported. One can think as a threshold with acceptable performance if TSS>0.80.

Results and discussion, page 2777. Eq. (12). How does this threshold compare to similar thresholds for the same area, of for nearby areas. As an example, thresholds have been recently proposed for Calabria, to the N and NE of the study area. See: doi:10.5194/nhess-14-317-2014. Other thresholds may be available for Sicily, or for similar areas.

Comparison of the derived threshold with the threshold by Gariano et al 2013 has been added to the analysis (section 5.2 and figure 5b). In particular in Sect. 5.2 we had written:

"Comparison with other thresholds may also help in understanding how reliable the performed analysis is. Gariano et al. (2013) proposed for Sicily the threshold $E = 10.4D^{0,22}$, where $E = I\times D$ is cumulative event rainfall, and hence threshold is equivalent to $I = 10.4D^{-0.78}$. This threshold has been derived considering only observed triggering events and it is corresponding to exceedance frequency of 1%. It is firstly interesting to notice that the exponent is practically equal to the one that results from our analyses (a_2 = −0.8). Furthermore, as can be seen from Figure 5b this threshold exceeds one triggering event of the MC simulated data, which equals the 1% of the triggering-rainfall dataset (see Table 5: 0.01×(TP+FN) = 0.01×(104+11) = 1.15). This result is a further support to the validity of the performed Monte Carlo analysis and highlights the importance to take into account non-triggering rainfall in assessing threshold performance."

Comments by Referee#2

General comment
The paper deals with the definition of rainfall Intensity-Duration thresholds to be used as tools for early warning of shallow landslides. Such a topic is surely of interest for the readership of HESS, and is somewhat innovative, in the sense that it makes use of already known models and approaches in an innovative way. Overall, the manuscript is well written and concise. Nonetheless, some parts are maybe even too concise and some information is missing, making in some cases difficult to judge the significance of the major obtained results (as I better explain in the following detailed comments). Therefore, in my opinion, moderate revision is needed before the manuscript could be accepted for publication in HESS.

Thank you, your comments were very useful for improving our work.

Detailed comments
As my comment is being posted after the comments made by two other readers, I prefer not to repeat all the already raised issues (in particular, in section 3, the potentially misleading symbols used in the ROC analysis, which description could be shortened because is not novel).

ROC section was revised to account for literature on the subject.

Section 2 (page 2765, lines 10-16). The Authors should discuss the implication of the choice of the inter-event time. In fact, some of their conclusions about the validity of the identified thresholds have to do with the role of the initial (pre-event) conditions on the triggering of the landslide. It is quite obvious that the choice of the inter-event time affects the "memory" of the previous event at the beginning of the new one. Such a memory depends also on the hydraulic properties of the soil cover under study (see another related comment below), so I expect that, for a given soil cover, the (arbitrary) choice of the inter-event time may hide the effects of the previously fallen rainfall. Indeed, the small number of events identified in the synthetic rainfall series used for the Monte Carlo simulation (less than 20 rainfall events a year) makes me think that the choice of a dry interval of at least 24 hours results in few long-lasting events separated by dry intervals long enough to allow the drainage of most of the previously infiltrated water from the soil cover (especially considering the high value of the hydraulic conductivity assumed for either the regolith or the fractured bedrock).

At the end of section "Monte Carlo synthetic data generation" the following explanation has been added to better explain why we have choose 24 hours to define individual rainfall events:
"Regarding the choice of the inter-event time $\Delta t_{\min}$ – an issue that is the focus of some works in literature (e.g. Restrepo-Posada and Eagleson, 1982; Bonta and Rao, 1988) – we have followed an approach analogous to that used by Balistrocchi et al. (2009) and Balistrocchi and Bacchi (2011), for which the inter-event time may be assumed as the minimum time needed to avoid overlapping of the response produced by two subsequent rainfall events. To this end we considered that the temporal peak of pressure head due to an individual rainfall event may be reached, as mentioned above, at an instant significantly after rainfall ceases. Hence a criterion for selecting the inter-event time has been that of choosing a value that approximates the dry time interval that contains the peak pressure head response relatively to all the N_{RE} simulated rainfall-events. In our case, from preliminary simulations a $\Delta t_{\min}$ = 24 h appeared suitable for the hydraulic and geotechnical soil properties which are considered in this work (see Sect. 3)."
Please note the related sentence added at section "3.1-Triggering and non-triggering rainfall identification":
"From a general standpoint, this procedure may be disconnected from the way event separation has been performed to compute the triggering instants with the methods described in Sect. 2.
Nevertheless, for consistency with the event separation criterion that is considered in Monte Carlo

simulation methodology, it is preferable that the procedure for identification of triggering and non-triggering events is based on the same inter-event time $\Delta t_{\min}$ used in Monte Carlo simulations."

Section 3 (page 2773, equation (9) and following discussion). As already pointed out by comments made by others, the chosen objective function is not novel. A possible improvement of such a commonly adopted approach, could be taking into account that in many real cases it is not obvious that a false alarm and a missing alarm have the same importance (the losses deriving from a missing alarm may be much more serious than the costs deriving from a false alarm), and it would be more effective for the decision-maker to define an objective function which accounts for such weights.

ROC section was revised to account for literature on the subject.
The following sentence was added regarding about possible adoption of different weights for quantities in the confusion matrix:
"Different weights may be given to the TP, TN, FP and FN, as 275 pointed out by Peirce (1884) itself, in order to account for the fact that a FN is more harmful than a FP (see also Peres and Cancelliere, 2012, 2013). Since data on the possible weights to assume are usually scarce, here we prefer to proceed in a more simple and standard manner, where this different weighting is not considered.".

Section 4 (page 2774, lines 13-20). More information should be provided about the calibration of the NRSP model, especially because the resulting synthetic series may affect the significance of the obtained I-D threshold (see the above comment about the effects of the choice of the minimum inter-event dry interval). Such concern about the NSRP calibration is motivated by the sentence at the beginning of page 2778, where the Authors say that in five years 190 events were recorded: nearly 40 events per year, which is around the double of the average yearly number of events of the synthetic generated series.

The following material/parts have been added to the manuscript as response to this comment:
-Figure 4 to explain how homogeneous seasons were identified;
-At section 4.2: "From the assumed inter-event time $\Delta t_{\min}$ = 24 h and soil properties of Tab. 3 the number of rainfall events results N_{RE} = 19826 (in average 19.83 events per year). This number results from initial 28 751 events then becoming after cutting the events with hourly intensities always below $i_{\min} = c_d K_s(1-\cos^2 \delta)$ = 2.975mm/h. These values are statistically comparable to the ones on the observed series (19.18 events/year from 28.91 events/year previously to the cut of under-leakage events)."
-At section 5.2: In particular we have derived from the series the rainfall events with the same criterion adopted in Monte Carlo simulations. Yet the events in the months neglected there (April-August) and the events with intensities below the leakage flow $c_d K_s(1-\cos^2 \delta)$ were not removed here in the observed record, for the test to be unbiased to this preprocessing of data.

Section 4 (page 2774, lines 22-26). Much more information about soil properties should be provided (it is not even written which kind of soil is studied). It seems that the Authors consider the obtained thresholds representative of an area as large as several tens of square kms. The variability of soil properties and slope morphology within such a large area could completely reduce the obtained results to a mere modeling exercise.
The value of the critical wetness ratio corresponding to the assumed geotechnical soil properties and slope geometry should be given here.

Section 4 on the study area has been extended to better explain the validity of the assumptions made in deriving thresholds. Results reported in the validation section 5.2 support the validity of the threshold for the area bounded in Figure 2 by a black line and composed by 7 sub-areas. The conducted sensitivity analysis (Section 5.3) enables to assess consequences of variability of soil properties, that may be observed due to spatial variability.

Regarding this last point, the following sentence has been added to Section 4:

"Spatial variability of each of the parameters could be included in our model simulations. Nonetheless, detailed information on how the properties are distributed spatially is unavailable. Hence we preferred to carry out a sensitivity analysis varying the following soil properties the hydraulic conductivity K_S, the leakage ratio c_d, the soil depth d_{LZ} according to Tab. 4 and the critical wetness ratio in the range $0 \leq \zeta_{CR} \leq 1$. This way to proceed enables to better analyse the way model results are influenced by these variables rather than assuming that they are distributed spatially with interpolating laws of difficult validation. Since slope mainly affects slope stability (Eq. B5) rather the infiltration process, variation of slope is indirectly taken into account by variation of ζ_{CR}. It is noteworthy to write that an alternative approach may be of considering model parameters generated according to a probability distribution, as proposed by the TRIGRS-P modification of the TRIGRS code, developed by Raia et al. (2014)."

The value of critical wetness ratio (0.4645) has been added to table 3.

Section 5. The results of the sensitivity analysis to variations of geotechnical soil properties and soil cover thickness are quite interesting, but they should be completed also with the analysis of the effects of variations of the hydraulic properties (in particular the hydraulic conductivity) and, even more, of the ratio cd between the hydraulic conductivity of the fractured bedrock and that of the soil cover (as far as I understand, arbitrarily set to 0.1 in absence of experimental data). Indeed, as the failure is a-priori assumed to occur at the soil bedrock interface, I expect the results to be extremely sensitive to the variation of such parameters.

Sensitivity analysis (new section 5.3) has been extended based on the suggestions of the referee. Conclusions and other parts of the paper have been updated to account for the extended sensitivity analysis.

Editorial issues

Page 2769, last three lines, and page 2770, first line. The explanation of the meaning of A, B and their ratio is unclear and should be reformulated.

The sentence has been modified as follows:

"The ratio *A/B*, which can be computed based on a Digital Terrain Model (DTM), is the well-known specific upslope contributing area, an important variable on which the topographic control on shallow landslide triggering depends (Montgomery and Dietrich, 1994). For instance, it is $A/B = BN_d$, where N_d is the number of cells draining into the local one, if one determines flow paths via the non-dispersive single direction (D8) method (O'Callaghan and Mark, 1984). Other methods consider multiple flow directions (cf. Holmgren, 1994)."

Page 2774, lines 9-10. Possibly the events of 25 October 2009 and 1 October 2009 are the same event (with wrong dates): somewhere else in the paper it is written that only four landslides occurred during the considered period.

Correct dates are now reported in the MS. The 1 march 2011 event was deleted from the validation test because we did not have simultaneous rainfall. Now, to avoid confusion, it was also deleted in the case-study section, because not relevant for the paper. Now everywhere in the paper only 4 landslide events are mentioned.

Page 2776, line 8. Replace "correspond" with "corresponds".

Done.

Page 2778, line 19. Insert “as” between “soon” and “soil”.

Sentence is not present in the revised MS.

Page 2779, line 2. Delete the word “for” after “Traditionally”.

Sentence is not present in the revised MS.

Comments by Referee#3

The manuscript is well written and describes an interesting and innovative approach to obtaining landslide triggering thresholds. The paper would benefit by clarification of a few minor points:
p. 2765, line 14. Please explain why 24 hours was selected as the time interval between separate storms. Was it an arbitrary selection or is this interval related to observed rainfall patterns for the study area? Does it have any connection with soil drainage rates for the study area?

As responded to the other reviewers, explanation of this choice has been added to the MS at the end of Section 2.

p. 2768, line 17, and elsewhere, change "indefinite slope model" to "infinite slope model"

Done.

p. 2769, line 23 to p. 2770, line 1, what is meant by "a lamination effect?"

The term "lamination effect" is wrong, and does not appear in the new MS. The phenomenon has been instead explained as follows in the revised MS: "The solution to Richards' equation provides the pore pressure profile in the unsaturated zone, and a flux to the saturated zone $q(d_u, t)$. Because of the partial absorption of water within the unsaturated zone, this flux results damped and smoothed respect to the infiltrating flux at the ground surface (cf. Figure 8 of Baum et al., 2008)."

p. 2771, section 3. Note that Staley et al. 2013 have recently published a similar approach, applying ROC analysis to instrumental data for deriving thresholds.
Staley, D.M., Kean, J.W., Cannon, S.H., Schmidt, K.M., Laber, J.L., (2013) Objective definition of rainfall intensity–duration thresholds for the initiation of post-fire debris flows in southern California. Landslides, 10(5):547-562.

Citation of Staley 2013 has been added in the introduction.

p. 2776, lines 3 - 25. The finding described here seems consistent with intense, short-duration rainfall being mainly responsible for inducing shallow landslides. If I am interpreting Figure 6 correctly, periods of higher intensity rainfall, sometimes following hours of low-intensity rainfall was a major factor in landslide triggering during most of your observed events. If so, then perhaps high-intensity rainfall during storms should be the primary focus of efforts to improve early-warning thresholds for shallow landslides.

Thank you for this suggestion, future research will be oriented to understand if rainfall thresholds may be improved by using other explanatory variables than the mean intensity of rainfall events. The sentence on future research at the end of the conclusions includes from a general perspective this aspect:
"Further ongoing research is oriented to introduce additional information in the derivation of the thresholds, such as antecedent precipitation as well as indexes representative of <u>the shape of the hyetograph</u>."

p. 2778, line 18, change "as soon soil" to "as soon as soil".

Sentence is no longer present in the revised MS.

Fig. 4. The flattening of the curve at long duration for the deterministic threshold shown in Fig. 4 results from competition between drainage and decreasing infiltration rates in the TRIGRS model for unsaturated infiltration. As the ratio of infiltration rate to Ks decreases, infiltration rate eventually becomes so small that pressure head cannot rise sufficiently to produce a factor of safety less than 1.

What reviewer states is true. The sensitivity analysis relative to the c_d parameter (leakage ratio), shows that as c_d increases ("competition between drainage and infiltration" moves in favour to drainage) the number of points triggered decreases with c_d (see Fig. R3-1 below which compares two situations that differ only for the c_d ratio).
We feel that the importance of the c_d ratio is often underrated in TRIGRS use (for instance Raia et al 2014 did not included it in the aleatory variables).

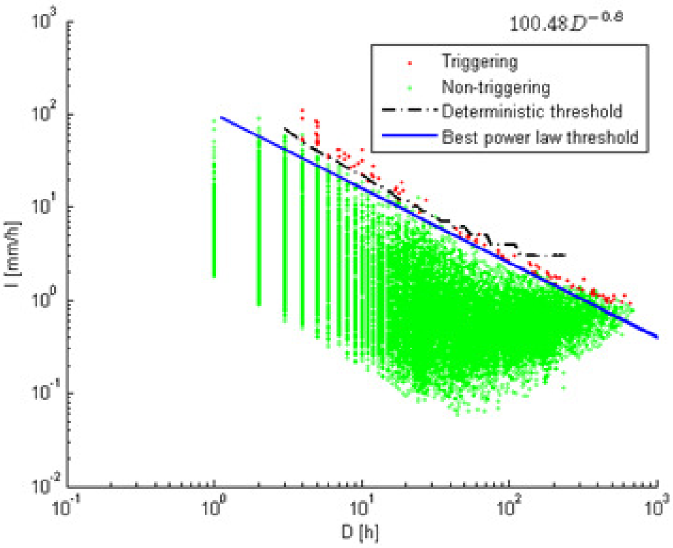

Fig. R3-1a. Triggering and non-triggering points for cd = 0.05 and Ks= 72 mm/h ζ_{cr} = 0.5, d_LZ=2m

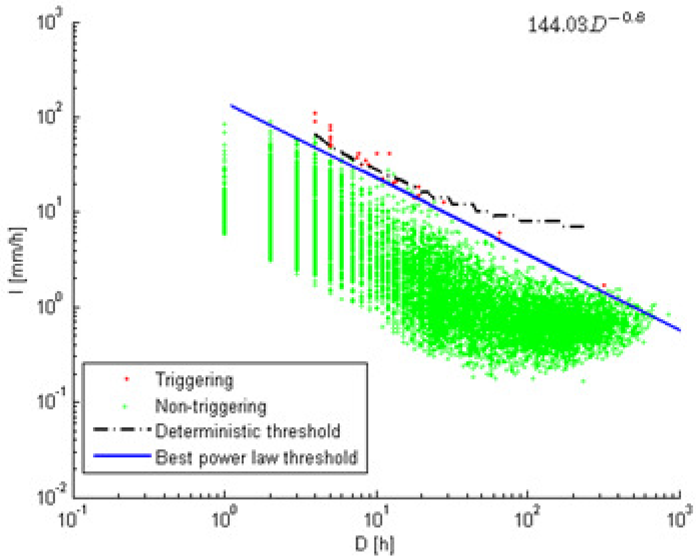

Fig. R3-1b Triggering and non-triggering points for cd = 0.20 and Ks= 72 mm/h ζ_{cr} = 0.5, d_LZ=2m

p. 2781, please add the following reference:
Baum, R. L., and J. W. Godt (2013), Correction to "Estimating the timing and location of shallow rainfall induced landslides using a model for transient, unsaturated infiltration", J. Geophys. Res. Earth Surf., 118, doi:10.1002/jgrf.20100.

Citation of Baum and Godt 2013 has been added where appropriate.

Comments by Referee#4

General comments
This paper deals with the definition of rainfall intensity-duration thresholds based on a combined approach which uses Montecarlo simulations to generate synthetic rainfall series, and a physically-based model to simulate pore-pressure dynamics and estimate the factor of safety of hillslope. Specifically, the Authors introduce a new way to assess the quality of empirically-based power-low ID thresholds. The paper is within the scope of HESS and surely of interest for both hydrologists and geomorphologists. However, for my perspective, this work - as it stands - suffers from two major limitations: 1) The scientific methods and assumptions are not always valid (at least for the "hydrological" part of the work); 2) The results are not sufficient to support the interpretations and conclusions of the work.
The revised MS presents clarifications about the concerns by this referee. In particular (1) The A/B = 0 case has been removed, and is more appropriately indicated as "no-memory" $\psi_0 = 0$ case. The paper introduces this case just to isolate the uncertainty due to rainfall intensity variability. (2) We have extended the section regarding the case study area and explained the validity and the data assumptions on which threshold derivation is based. More details are given below.

About the first point, the assumption to consider two different time-scales for vertical and lateral flow is valid if the ratio between soil depth and the square-root of the upslope contributing area is small: d_lz/Aˆ0.5«1. Based on this assumption, the Authors used the TRIGRS model (1D model) to simulate vertical rainfall infiltration during rainfall
events, and a drainage model to simulate lateral flow after the end of rainfall. The Authors agree with the hypothesis (section 2.2), but then, paradoxically, they perform their experiments by varying the specific upslope area from 0 (d_lz/Aˆ0.5=infinity) to 20/B. For my view this is a strong weakness of the paper, as the major conclusions of the work are related exactly to this point ("power-law ID equations can adequately represents the triggering conditions. . .. . ..for a hillslope with small specific contributing area").

The "A/B = 0 case" has been removed, and is more appropriately indicated as "" $\psi_0 = 0$ case". This case is introduced in the paper to isolate the uncertainty due to rainfall intensity variability from influence of initial conditions. Conclusions and others parts of the MS have been modified to clarify this aspect and conditions under which ID power-law thresholds may be valid based on our results. In particular for the conclusions:
"[…] the following conclusions can be drawn: (1) Variability of intensity during rainfall events influences significantly rainfall Intensity and Duration associated to landslide triggering. In particular constant-intensity input thresholds perform conservatively only for low rainfall durations, while the opposite occurs for events of longer duration. On the other hand, when a time variable rainfall-rate event is considered model, the simulated triggering points may be separated with a very good approximation (True Skill Statistic results close to 1) from the non-triggering ones by a power-law ID equation. This indicates that this widely-used model is adequate to represent the triggering part due to transient infiltration produced by rainfall events. Thus this gives a physically-based justification for such a widely-used threshold form, which results valid when landslide occurrence is mostly due to that part. This depends, for a given rainfall climate, mostly on the timing of recession of the saturated zone occurring during dry inter-event intervals (in our model represented by the constant τ_M), but also on the other soil hydraulic and geotechnical parameters, and in particular on soil depth d_{LZ}, which must not be to shallow, and critical wetness ratio ζ_{CR}, that must be not to low. For instance, for the case study area the I-D power law threshold performs with a TSS > 0.80 when it is $\tau_M \leq 3$ days and $d_{LZ} \geq 1.5$ m and $\zeta_{CR} > 0.50$".

About point 2, the Authors say that ". . .the proposed methodology is applied to the landslide-prone area of Peloritani Mountains, Northeastern Sicily, Italy.". Actually, for my perspective, this work represents a modeling exercise on an artificial-uniqueplanar hillslope, characterized by 40 degrees

constant slope, uniform soil depth, and mechanical- and hydrological- properties. I am not sure that results and conclusions can be generalized and validated against the real events (figure 5), and I am wondering why the Authors did not use the "real" landscape (digital elevation model) of the landslide-prone area of Peloritani Mountains to perform their experiments.
Therefore, in my opinion a major revision is needed before the manuscript can be considered ready for publication in HESS.

More information on the case study area is provided in the revised MS (Sect. 4). In particular the following parts are relevant for this comment:
"From the analysis of slopes δ and within the slide areas, based on a pre-event DTM at a 5 m resolution, it results that the most populated class of A/B is centered on the value of 10 m, while the mean slope within the range of theoretical potentially unstable slopes 29◦ ≤ δ ≤ 47◦ results slightly lower than 40°. Also, the values of A/B = 10 m and δ = 40◦ correspond to a portion of the Peloritani Mountains for which it starts to be worthed to issue landslide early warnings. Hence these values may be adopted for the successive derivation of a threshold for the area (see Sect. 5).

And after, regarding a possible spatially distributed application:
"Spatial variability of each of the parameters could be included in our model simulations. Nonetheless, detailed information on how the properties are distributed spatially is unavailable. Hence we preferred to carry out a sensitivity analysis varying the following soil properties the hydraulic conductivity K_S, the leakage ratio c_d, the soil depth d_{LZ} according to Tab. 4 and the critical wetness ratio in the range $0 \leq \zeta_{CR} \leq 1$. This way to proceed enables to better analyse the way model results are influenced by these variables rather than assuming that they are distributed spatially with interpolating laws of difficult validation. Since slope mainly affects slope stability (Eq. B5) rather the infiltration process, variation of slope is indirectly taken into account by variation of ζCR. It is noteworthy to write that an alternative approach may be of considering model parameters generated 340 according to a probability distribution, as proposed by the TRIGRS-P modification of the TRIGRS code, developed by Raia et al. (2014)."

The obtained threshold on the Peloritani Mountains is validated against available data and other thresholds (sect. 5.2). We don't understand his perplexity on the validation test. The following sentence at section 5.2 clarifies the meaning and validity of the test:
"This is the best result one can obtain from this test, but it is perhaps noteworthy to clarify that it is expected that in the long period the same test will not perform without errors, consistently with the Monte Carlo simulations and the way the threshold was derived."

Specific comments
Abstract (page 2760, lines 21-23): This conclusion "power-law ID equations can adequately represents the triggering conditions. . .. . ..for a hillslope with small specific contributing area") is not consistent with the hypothesis of the "TRIGRS-Drainage" model (see General comments). Paradoxically, This is also what the Authors argue at section 2.2. This point absolutely needs to be addressed.

See response above. Here we add that also the abstract has been modified to better clarify paper conclusions:
"This indicates that the ID power-law equation is adequate to represent the triggering part due to transient infiltration produced by rainfall events of variable intensity and thus gives a physically-based justification for this widely-used threshold form, which results valid when landslide occurrence is mostly due to that part. These conditions are more likely to occur in hillslopes of low specific upslope contributing area, relatively high hydraulic conductivity and high critical wetness ratio. Otherwise, rainfall time-history occurring before single rainfall events influences landslide

triggering, determining whether a threshold based on rainfall Intensity and Duration only may be sufficient or it needs to be improved by the introduction of antecedent rainfall variables."

Introduction (page 2762, lines 14-20): I think that this part should be moved into the results and discussion section.

The sentence has been modified as follows to less appear as a result: "Since many landslides, especially the most devastating shallow rapidly-moving ones, may be triggered by rainfall events of few hours (cf., e.g., Highland and Bobrowsky, 2008), use of daily rainfall for threshold derivation in these cases <u>is quite questionable."</u>

Section 2 (page 2765, lines 8-9): Please, clarify what you mean with "a representative site". As also stated by referee #1, all this part needs to be clarified.

Section on the NSRP model has been modified to clarify this part.

Section 2 (page 2765, lines 18-20): This is an assumption that the Authors make. After a long dry period, pore pressure at the soil-bedrock interface may also assume negative values. Are the Authors assuming a steady-state initial pore-pressure (i.e., suction head) profile?

The part has been modified to take into account this comment:
"For the analysed case-study area and many similar cases, it may be assumed that at the beginning of each hydrological year the water table is at the basal boundary, because an almost totally-dry season comes before <u>(this may be a slightly conservative assumption, since pressure head at the soil-bedrock interface may assume negative values after a long dry season)."</u>

Are the Authors assuming a steady-state initial pore-pressure (i.e., suction head) profile?

Regarding this point the following part was added to the MS :

"The initial pressure head distribution above the water table is computed accordingly with assumptions of the transient vertical infiltration model (see next section B2), letting the steady (initial) surface flux $I_{ZLT} = 0$, which yields the following equation (see Baum et al., 2010; Baum and Godt, 2013):
$\psi(Z, t = 0) = -(d_u - Z)\cos\delta - 1/\alpha$ (B3)
for depths $Z \leq d_u$, d_u being the depth to the top of the capillary fringe and α the parameter of Gardner's (1958) exponential soil-water characteristic curve (cf. Fig. B.1)."

Section 2 (page 2766, lines 5): Please, clarify that you are making the assumption that the failure surface coincides with the soil-bedrock interface. Is this consistent with the landslide-prone area of Peloritani Mountains?

The sentence was modified as follows:
"In this scheme the failure occurs at the basal boundary $Z = d_{LZ}$, because pressure head results maximum at that depth.".
The assumption is consistent to soil stratification in the Peloritani area (see section 4):
"Core samples collected in the area indicate the presence of a surficial debris material dLZ = 2m deep covering a fractured bedrock strata.".

Section 2.1 (page 2767, lines 8): Please define cdf.

Done.

Section 2.2.2 (page 2769, line 17): I do not see alpha_1 in the equation.

Equation has been changed to a more appropriate version in which alpha_1 appears.

Section 2.2.2 (page 2769, line 23): This is a very "particular" explanation. I would remove this sentence from the text.

The sentence was better explained to appear less "particular" (peculiar), as follows:
"Because of the partial absorption of water within the unsaturated zone, this flux results damped and smoothed respect to the infiltrating flux at the ground surface (cf. Figure 8 of Baum et al., 2008)."

Section 2.2.3 (page 2771, lines 1-2): Actually, the equation tell us that for gentle slope (and/or thin soil depth) the critical wetness ratio is also high with poor mechanical soil-properties. I would remove this sentence.

Sentence has not been removed, yet it has been changed as follows:
"The ζ_{CR} varies from 0 to 1, respectively for an unconditionally unstable and a unconditionally stable hillslope (Montgomery and Dietrich, 1994), and hence it indicates the natural degree of stability of the hillslope."

Section 4 (page 2774, lines 23-26): I think this is a too simplistic assumption ("we consider as representative for the case study area a hillslope of slope d=40_ and. . .."). What about the characteristics (slope angle, soil depth, etc.) of the four (are the Authors sure on this value?) landslides that the Authors used to validate the derived threshold? Much more information about the investigated area and data should be provided in the paper.

More information about the investigated area and data is be provided in the paper. In particular the assumption of δ=40° and A/B = 10 m is based on analysis of a 5x5 DTM, as explained in the revised MS and the added figure 3.

Section 5 (Results and discussion): The ID thresholds are derived by considering a "rigid" hillslope configuration and this makes the comparison between model results and reality quite flawed. What about to use the "real" landscape (at least a portion where landslides occurred) instead of the "rigid" hilllslope schematization?

It is not totally clear to us why the referee thinks that "comparison between model results and reality quite flawed". The following sentence was added to the MS in the validation test part to avoid misinterpretations:
"This is the best result one can obtain from this test, but it is perhaps noteworthy to clarify that it is expected that in the long period the same test will not perform without errors, consistently with the Monte Carlo simulations and the way the threshold was derived."

Comments by SL Gariano

This paper focuses on the determination of rainfall thresholds for the forecasting of rainfall induced landslides, and provides interesting food for thoughts. Use of stochastic rainfall data coupled with a slope stability model ensures a good balance between the classical empirical and physically based approaches. The ROC-based criterion
offers several quantitative values â˘A ´ Nâ˘A ´N that contribute to make less subjective the identification of the thresholds.
Concerning the ROC-based analysis, I would like to highlight the following aspects. In the (wide!) literature on ROC analyses, the same indexes are shown with different names and acronyms. This is unfortunate. To avoid possible misunderstandings, Barnes et al. (2009) have recommend that Authors consider using the following nomenclature:
- Probability of Detection, POD = TP / (TP + FN),
- Probability of False Detection, POFD = FP / (FP + TN), and
- Probability of False Alarms, POFA = TP / (TP + FP).
I recommend that the authors adopt this nomenclature, and that they change their
"TPR, FPR, and PRE" with "POD, POFD and POFA".
Further, the index that the authors call "Δ" [capital Delta] is known as the Hanssen- Kuipers discriminant, and was originally introduced by Pierce in 1884. It is also called True Skill Statistic (see e.g., Hanssen & Kuipers, 1965; Wilks, 1995; Stephenson 2000; Accadia et al., 2003). To avoid unnecessary confusion, the authors should consider using the notation "HK" (or "TSS"), instead of "Δ".

Thank you for your comments. Paper was modified in many parts to account for ROC literature, and relevant citations were added.

Sincerely,

David Johnny Peres
DICAR – Deparment of Civil Engineering and Architecture
Via S. Sofia, 64
95123 – Catania (Italy)
Tel: +39 095 738 2711
Fax: +39 095 738 2748
email: djperes@dica.unict.it