Dear Editor and Dr. Francesco Marra,

Thank you very much for your useful comments and suggestions.

In this document, you will find a detailed explanation of the changes made to the original manuscript to meet your suggestions.

For the sake of clarity, we used the following text styles:

black, italics:	reviewer comment
blue, plain text:	our reply
blue, italics:	revised text

Best regards

Elena Ioriatti

Mauro Reguzzoni

Edoardo Reguzzoni

Andreas Schimmel

Luca Beretta

Massimo Ceriani

Matteo Berti

Line 35: I think the reference to Nikolopoulos & al here is misplaced as this paper does not aim at developing or testing empirical thresholds.

Thank you for your observation. The reference to Nikolopoulos et al. 2014 is not fully aligned with the purpose of this sentence. In the revised manuscript, we have removed this citation.

Lines 43-49: Temporal resolution of the rainfall data also constitutes an important factor for empirical thresholds (Marra & al 2019; Gariano & al, 2020).

Thank you for your valuable suggestion. We have included temporal resolution as an additional factor influencing the reliability of rainfall thresholds in the revised text.

The reliability of rainfall thresholds defined with an empirical approach can be influenced by several sources of uncertainty, including the spatial distribution of rain gauges, the criteria used to define individual rainfall events, and the temporal resolution of rainfall data. Marra et al. (2016) and Nikolopoulos et al. (2014) highlighted that rain gauge networks with limited spatial coverage can underestimate rainfall during convective storms. This may lead to thresholds that do not accurately reflect triggering conditions. Another key source of uncertainty lies in the method used to identify discrete rainfall events from continuous data (Melillo et al., 2015). A common approach is to use a minimum inter-event time, but there are still no clear criteria for determining its optimal duration (Dunkerley, 2008). The temporal resolution of rainfall data is another important factor, as coarse resolution has been shown to systematically underestimate depth–duration (ED) thresholds (Marra, 2019; Gariano et al., 2020).

Furthermore, we have specified the temporal resolution of our rainfall data at lines 150–151 (Sect. 3.3):

Rainfall data with a temporal resolution of 5 minutes were collected between late spring and early autumn in 2021, 2022, and 2023.

Lines 149-150: This is an unnecessary level of detail for a basic analysis. It sounds more like a technical report than a scientific paper.

Thank you for your comment. The sentence has been removed in the revised manuscript.

Lines 173-176: it seems that the classification is still done by an operator. I suggest removing this and simply state that the classification was done based on an operator.

Thank you for the comment. We have clarified that event classification was performed by an operator. We also consider it important to specify that a script with a user interface was employed to automatically record the start and end times of each observed process in a structured dataset, as this detail is relevant for replicating the method. Indeed, performing the classification manually and transcribing the start and end times of each process is time-consuming and prone to errors.

Event classification was operator-based and supported by a script that displays the images and automatically generates a structured dataset with class labels and corresponding time intervals. In an initial attempt, classification was carried out manually by reviewing each image, recording start and end times, and assigning class labels. This method proved inefficient and error-prone due to manual transcription. The script improved the process by providing an interface with simple controls that allowed the operator to classify each image, automatically creating a continuous time-series dataset that recorded the class and the corresponding start and end times of each observed process. Although classification still required expert input, the script greatly reduced transcription errors and accelerated the overall workflow.

Lines 182-183: "the operator became more adept…" does this mean the quality of the classification changes over time? What are the implications for the analyses? Would two different operators do the same classification, would we get the same results? How would these potential differences affect the AUC? Would a sensitivity analysis to these subjective choices help quantifying the uncertainties related to the proposed method?

Thank you for raising this point and for noting the potential ambiguity of that sentence. We recognise that image classification inevitably involves a degree of subjectivity. To minimise this, we defined classes to be as objective as possible, while acknowledging that the analysis necessarily remains operator based. To reduce this subjectivity, two of the authors jointly examined a large number of cases and established shared classification criteria. Following this initial training, these criteria were applied consistently across all images, thereby limiting operator bias. We did not perform a sensitivity analysis, as the classification was considered sufficiently robust.

Some classification uncertainties arose during rainfall, when visibility was compromised by dense fog or water droplets on the camera lens. In such instances, it was sometimes helpful to examine images of the channel before and after the event to determine whether sediment transport, along with associated erosion or deposition, had occurred. Weather conditions, such as sunny or cloudy days and the time of day, could also lead to misinterpretations. For example, in shaded areas water flow was often less visible, while in bright sunlight reflections on the water surface could create the illusion of higher discharge. To address these challenges, two of the authors jointly reviewed a large number of cases and established shared criteria to distinguish actual changes in discharge from lighting or visibility artefacts. These criteria were then applied consistently across all images to ensure comparability and limit operator bias.

Lines 191-192: again, this is an unnecessary level of detail for a basic analysis

Thank you for the suggestion. The sentence has been modified removing that we used a custom script.

Each precipitation event was characterised by the highest class observed during its duration.

Section 3.4: it is not mentioned what features are examined for this dimensionality reduction. This is crucial information. It turns out from section 4.2 that basically this is a 2-dimensional clustering with duration and average intensity. Should be stated here.

Thank you for this valuable comment. We have revised the text to explicitly state that rainfall duration and average intensity were defined a priori as the predictor variables.

In this study, we did not apply dimensionality reduction from a larger feature set; instead, rainfall duration and average intensity were defined a priori as the predictor variables for the analysis.

Lines 219-223: if I understand this correctly, this means that the threshold was optimised to maximise the AUC. I don't understand why the 0.1 steps in log scale are needed for this. Seems like a lot of details for an optimisation.

We apologise for the oversight and thank you for your comment. We have revised the text to clarify that TH2 is a new threshold, specifically defined to identify debris flows, and not an optimisation of TH1. TH2 was obtained by shifting the threshold TH1 upward along the intensity axis with increments of α of 0.1, while maintaining the same slope β. The threshold TH2 was selected as the one, among all candidates generated at each 0.1 step, that maximised AUC. The choice of 0.1 increments was a practical way to explore the threshold space with sufficient resolution.

Two thresholds were defined to separate the observed hydrological responses: TH1 is a lower threshold distinguishing low flow (C1) from high flow, high flow with sediment transport and debris flow (C2, C3, C4), while TH2 is an upper threshold distinguishing debris flow (C4) from all other classes (C1, C2, C3).

[…]

TH2 was derived by keeping the scaling exponent β (slope) of TH1 and iteratively increasing the coefficient α in increments of 0.1, which in the log–log form of $I = \alpha \cdot D^{\beta}$ corresponds to shifting the intercept $\log_{10}\alpha$ upward. The model performance was evaluated at each step by using the Area Under the Receiver Operating Characteristic Curve (AUC). The final threshold was selected as the value that maximized AUC, ensuring the best separation between debris flows and non-debris flows.

Figure 7: this is trivial, the text in 3.5 is enough to understand that you computed the min, max and mean values across the different monitoring stations.

We agree that the information in Section 3.5 already describes the procedure; however, we would prefer to keep Figure 7 as it makes it immediately clear that for each rainfall event recorded by the UNIBO rain gauge and at least one Hortus rain gauge, we calculated the minimum, maximum, and mean values. These values were then used to plot the error bars shown in Figure 13.

Section 4.2 lots of methods here. The first sentence (lines 277-279) is a repetition of what stated in section 3.4 and should be removed. Lines 280-285 instead provide important information on the methods that should be moved to section 3.4.

Thank you for the suggestion. We have removed these sentences from the Results section and revised Section 3.4 to provide this information more effectively.

3.4 Rainfall threshold definition using the Linear Discriminant Analysis (LDA)

Two thresholds were defined to separate the observed hydrological responses: TH1 is a lower threshold distinguishing low flow (C1) from high flow, high flow with sediment transport and debris flow (C2, C3, C4), while TH2 is an upper threshold distinguishing debris flow (C4) from all other classes (C1, C2, C3). To determine TH1, the method of Linear Discriminant Analysis (LDA) was applied to the dataset, treating low-flow events (C1) as non-triggering ("False") and all other responses as triggering (C2, C3, C4, "True"). LDA is a statistical method for dimensionality reduction and feature selection that identifies a linear combination of input variables to optimally separate triggering and non-triggering classes (Fisher, 1936; Ramos-Cañón et al., 2016). In this study, we did not apply dimensionality reduction from a larger feature set; instead, rainfall duration and average intensity were defined a priori as the predictor variables for the analysis. LDA was then applied to the rainfall events, with the aim of identifying a discriminant axis that maximizes between-class variance and minimizes within-class variance, as described by the objective function J(w) in Eq. (1) […]

It would be very useful here to know that is the advantage of using this LDA clustering over other methods for defining thresholds that are more common in literature. In particular, the way TH2 is calculated resembles a lot the frequentist approach by Brunetti et al 2010 in which a slope in log(D)-log(I) coordinates is kept constant and the intercept is changed to match some condition (here to optimize the AUC and there to leave a pre-defined proportion of observed events below). In both cases, the question on whether the same slope should be used is a fundamental one. Perhaps it should be discussed in the frame of this method: what is the hydrological reasoning behind using the same slope?

The main reason we employed LDA rather than the frequentist approach of Brunetti et al. (2010) is the limited number of events available, which did not allow for a robust probabilistic analysis of the type used in their study. LDA offers an alternative by defining the threshold through statistical separation of predefined classes, which can be applied even with relatively small datasets.

As for the question of using the same slope for the two thresholds, previous studies on runoff-generated debris flows (e.g. Berti and Simoni, 2005; Simoni et al., 2020; Berti et al., 2020) have shown that thresholds tend to display similar slopes, at least for the short-duration events typical of debris-flow initiation. This reflects the fact that both runoff generation and debris mobilization are controlled by the same hydraulic process, namely the concentration of overland flow and its transformation into channelized flow. This point has been clarified in Section 3.4 of the Methods. Moreover, new analyses have been performed on the hydrological interpretation of the two thresholds using the SCS Curve Number (CN) rainfall–excess model combined with the SCS dimensionless Unit Hydrograph (CN–UH method; Soil Conservation Service, 1972) and the results added as Section 5.3 of the Discussion.

3.4 Rainfall threshold definition using the Linear Discriminant Analysis (LDA)

[…]

For the debris-flow threshold (TH2), events classified as debris flows (C4) were treated as "triggering" ("True"), while all other classes (C1, C2, C3) were treated as "non-triggering" ("False"). The LDA method was not applied in this case because the limited number of debris flows made it difficult to reliably estimate within-class variance and class means for a stable discriminant axis. Moreover, the strong imbalance between classes biases the separation boundary, as the dominance of the majority class shifts the boundary toward the minority class, reducing classification accuracy. TH2 was derived by keeping the scaling exponent β (slope) of TH1 and iteratively increasing the coefficient α in increments of 0.1, which in the log–log form of $I = \alpha \cdot D^{\beta}$ *corresponds to shifting the intercept* $\log_{10}\alpha$ *upward. The model performance was evaluated at each step by using the Area Under the Receiver Operating Characteristic Curve (AUC). The final threshold was selected as the value that maximized AUC, ensuring the best separation between debris flows and non-debris flows. Although assuming parallelism between TH1 and TH2 is methodologically convenient, it can be questioned from a hydrological perspective, as the rainfall duration–intensity relationship may differ between flow-depth increases and debris-flow mobilization. However, for runoff-generated debris flows, studies have shown that the two thresholds display similar slopes, at least for the short-duration events that typically trigger debris flows (Berti and Simoni, 2005; Simoni et al., 2020; Berti et al., 2020). This similarity arises because runoff generation and the mobilization of channel debris are both expressions of the same hydraulic process: the concentration of overland flow within the catchment and its transformation into channelized flow.*

5.3 Hydrological interpretation of rainfall thresholds

A major strength of our method, which relies on monitoring data from many rainfall events, is the ability to identify thresholds not only for debris-flow initiation but also for earlier stages of hydrological response. The lower threshold, TH1, which separates events that do not change channel flow depth from those that cause a measurable increase, is particularly relevant from a hydrological standpoint. It marks the point at which rainfall surpasses the catchment's initial losses, producing overland flow on exposed rock surfaces and shallow runoff along talus-slope drainage lines, and ultimately supplying water to the main debris-flow channel. This empirical threshold can be further supported by a simple hydrological analysis that improves understanding of catchment response.

Figure 15 compares the UNIBO TH1 threshold with the theoretical runoff discharge computed at the UNIBO monitoring station using the SCS Curve Number (CN) rainfall–excess model combined with the SCS dimensionless Unit Hydrograph (CN–UH method; Soil Conservation Service, 1972). A similar approach was applied by Gregoretti et al. (2016) and Berti et al. (2020) to evaluate rainfall excess in debris-flow initiation zones of alpine catchments. Input data for the analysis are listed in Table 5. The key parameter of the method is the Curve Number, which defines the watershed's potential maximum retention and directly controls runoff *generation. We derived a composite CN as the area-weighted average of three values assigned to exposed bedrock, old landslides, and debris deposits that characterize the basin upstream of the UNIBO station (Fig. 3). A sensitivity analysis was carried out using minimum and maximum CN values for each unit, derived from USDA-SCS lookup tables and from values back-calculated by Bernard et al. (2025) for three monitored basins in the Eastern Italian Alps. All analyses assumed normal antecedent moisture conditions (AMC II), and the time of concentration was estimated with Kirpich's formula (Kirpich, 1940).*

The results show a fairly good agreement between empirical and theoretical thresholds. In particular, this agreement is clear for high CN values, which reflect low infiltration capacity. In these cases, the zero-discharge line marking the onset of channel runoff coincides with the lower boundary of the blue triangles, which indicate visible increases in flow depth recorded on video. Nevertheless, the theoretical TH1 threshold is steeper than the empirical one. This discrepancy arises from the simplified assumptions of the SCS-CN abstraction model. As highlighted by Berti et al. (2020), under the assumption of constant initial loss the model behaves like a simple "bucket," where the catchment begins to spill once its storage capacity is filled. In such conditions, the theoretical slope of the runoff-initiation threshold is –1, compared with –0.56 for the empirical threshold. The gentler empirical slope suggests that initial losses increase with rainfall duration, likely due to long-term infiltration into weathered rock or debris, an effect not represented in this simplified analysis.

With regard to the empirical debris-flow threshold (TH2), the model indicates that debris mobilization corresponds to a peak runoff discharge of about 2–3 m³/s. These values appear much higher than the critical surface discharge values reported by Gregoretti and Dalla Fontana (2008) and Berti et al. (2020) in similar geological settings, which are typically below 0.2 m³/s. However, it should be emphasized that the runoff discharges in Fig. 15 are computed at the UNIBO station, not in the initiation area, where the contributing headwater catchment is considerably smaller. More relevant to our analysis is the fact that the empirical threshold TH2 is roughly parallel to a theoretical line of equal-runoff discharge, again supporting the physical basis of the threshold identified from monitoring data. Although the discharge contours do not exactly match the slope of the runoff-initiation line, the discrepancy is minor and difficult to detect in empirical datasets. Consequently, the simplified assumption of slope similarity between TH1 and TH2 remains theoretically founded.

Table 5. Parameters adopted for the SCS-CN and SCS Unit Hydrograph (SCS–UH) analysis at the UNIBO monitoring station. The table reports basin descriptors, land-cover/soil units with corresponding Curve Numbers (CN), and hydrological parameters used for runoff and hydrograph computation.

Parameter		Value
Basin characteristics	Basin area (m^2)	2052904
	Basin length (m)	2856
	Basin height (m)	1800
Land cover	Rock area (m^2)	1245455
	Landslide area (m^2)	328718
	Debris area (m^2)	478731
	Rock Curve Number [min–max]	85-95
	Landslide Curve Number [min–max]	60-70
	Debris Curve Number [min–max]	70-80
	Composite Curve Number [min-max]]	77-87
Hydrological parameters	Potential Maximum Retention, S (mm) [min-max] *S=(25400/CN)-254*	38-76
	Initial Abstractions, Ia (mm) [min-max] *Ia=0.2S*	8-15
	Time of Concentration, Tc (h) *from Kirpich's formula*	0.18

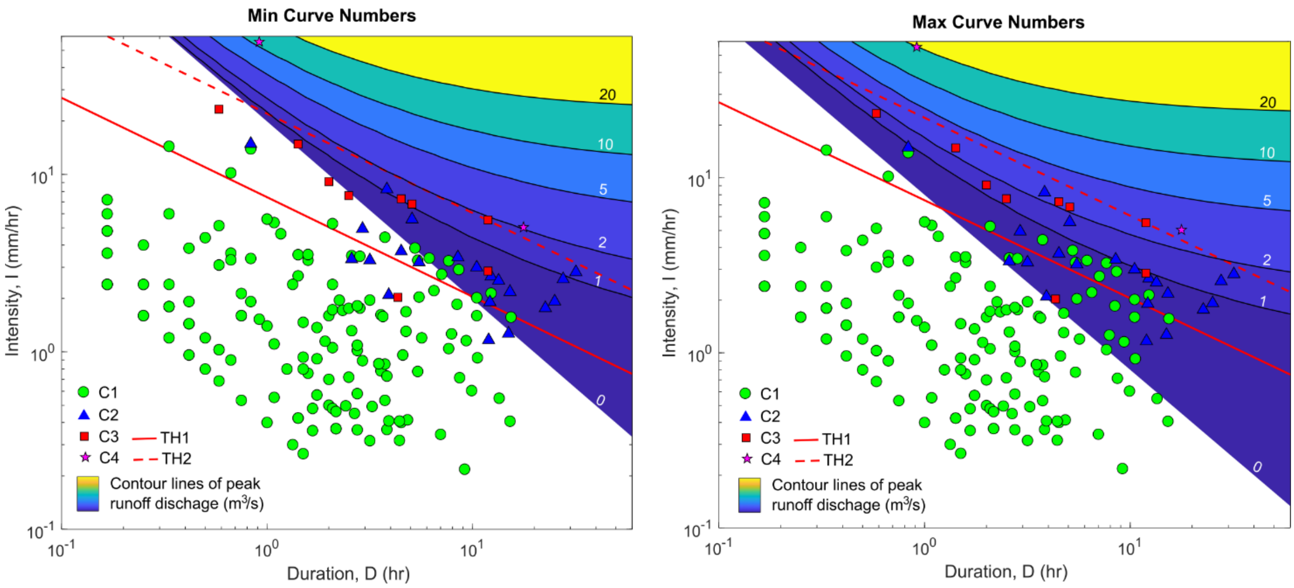
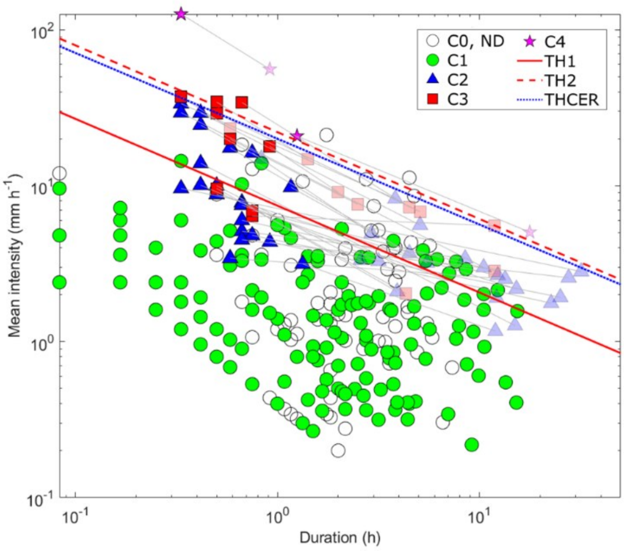


Figure 15. Contour maps of peak runoff discharge obtained with the SCS–UH method for (left) minimum CN values and (right) maximum CN values. Empirical observations of rainfall events are superimposed, with symbols indicating event classification. The comparison illustrates the sensitivity of theoretical runoff estimates to Curve Number selection.

Figures 9, 10 and 11: it is notable that the separation between C1 events and other events is better at short durations and then the different events merge for longer durations. One could claim this is because longer-duration events likely include short-duration bursts with higher intensities that determine the hydrological response. What is the time of concentration of the catchment? Given the fact that 2 hours separation are considered enough for separating events (and, therefore, antecedent conditions are relatively negligible), does it make sense to average intensities over durations longer than the time of concentration? It would be useful commenting on this aspect.

Many thanks for the insightful comment. The time of concentration (Tc) range between 11 min (Kirpich) and 17 min (Giandotti). For all rainfall events that showed a basin response (C2, C3, C4), we extracted the burst and recalculated the intensity. It should be noted that the extraction of the burst introduces an additional degree of subjectivity. The figure shows the new data in magenta, red, and blue, while the corresponding previously identified events are shown in transparency. We observed that discrimination among classes does not improve; in particular, classes C2 and C3 remain mixed, as was the case with the previously identified events in Fig. 9 of the manuscript. Although in theory, for durations longer than Tc, keeping the intensity constant, discharge should be constant (the triggering intensity should remain constant, i.e., the threshold would be horizontal for D>Tc), in practice this is unlikely because rainfall is not uniformly distributed across the basin. Therefore, as duration increases, it is more likely that the rainfall cell moves and covers a larger portion of the basin, producing a greater discharge. Because the threshold represents the set of (I, D) pairs that generate the same hydrological response, even for D > Tc a decrease in the critical intensity required to maintain that discharge is observed, resulting in a negatively sloped threshold. Accordingly, considering durations longer than Tc remains physically meaningful in this setting, because the evolving areal coverage of precipitation cells can increase discharge over time. Therefore, for a fixed target discharge (i.e., the same hydrological response), the critical intensity Ic required to achieve it decreases as duration increases.



4.1 Classified rainfall events

[…]

The separation of C2, C3, and C4 events from C1 events is clearer at short durations than at longer durations. One could argue that this occurs because long-duration events include short high-intensity phases (bursts) that control the hydrological response, and that it is therefore not appropriate to consider durations exceeding the time of concentration (Tc). For durations close to Tc the entire catchment contributes to runoff; consequently, at the intensity Ic associated with Tc, durations greater than Tc should not further increase discharge. One would therefore expect the threshold to be horizontal for D > Tc, indicating an approximately constant basin response.

In practice, however, rainfall is not uniformly distributed across the catchment. As duration increases, the precipitation cell is more likely to shift and cover a larger fraction of the basin, producing greater discharge. Because the threshold represents the set of (I, D) pairs that generate the same hydrological response, even for D > Tc a decrease in Ic required to maintain that discharge is observed, resulting in a negatively sloped threshold. Accordingly, considering durations longer than Tc remains physically meaningful in this setting, because the evolving areal coverage of precipitation cells can increase discharge over time. Therefore, for a fixed target discharge (i.e., the same hydrological response), the critical intensity Ic required to achieve it decreases as duration increases.

Lines 334-336: in addition to the percent change of beta e alfa, it would be useful to know the largest percent changes in the intensities for the range of durations that are considered useful for the triggering in the area.

Following the proposed approach, we computed the largest percentage changes in rainfall intensity among the five gauges for durations considered relevant for landslide triggering in the study area. Specifically, for D = 0.5 h the maximum percentage change is 137.8%, while for D = 1 h it is 107.0%. These values have now been reported in the revised manuscript.

The largest percent changes in intensity among the five rain gauges for the duration of 30 minutes is 137.8%, while for the duration of 1 hour is 107.0%.

Figure 12: how are the regressions and the related uncertainties computed? Usual linear regression models assume no error on the variable used in the x axis and homoschedastic variables on the y axis, which is not necessarily the case here, since there is complete symmetry between UNIBO and the other stations.

We decided to adopt the UNIBO rain gauge as the reference, considering it as the ground truth and assessing all comparisons against it. We now state this choice explicitly in the revised manuscript.

Rainfall data recorded at the Hortus stations were compared against the UNIBO reference (treated as ground truth) using ordinary least-squares linear regression for each rainfall-event characteristic. Figure 12 illustrates the differences in precipitation amount (a), duration (b), and mean intensity (c) between the reference UNIBO station and the four Hortus stations, which are positioned upslope (H1, H2, H3) and downslope (H4). Each plot includes the 95% confidence bands of the regression lines, shown in different colours for each rain gauge.

Line 350: how is this statistical significance calculated?

Thank you for pointing this out. The earlier statement incorrectly inferred statistical significance from whether the plotted confidence band appeared to exclude the 1:1 line. We have corrected this mistake in the text.

For the mean intensity, when UNIBO intensity exceeds 15 mm h^{-1}, the 95% confidence band for H1–H3 lies entirely below the 1:1 line, suggesting a systematic negative bias. By contrast, the band for H4 still overlaps the 1:1 line, indicating approximate agreement also at higher intensities (Fig. 12c).

We also corrected this part of the manuscript:

In general, total precipitation measurements show good agreement between UNIBO and Hortus rain gauges. For higher cumulative totals, the 95% confidence bands for H1 and H2 lie predominantly above the 1:1 line, whereas H3 overlaps to it and H4 lies below it. These patterns are consistent with an elevation effect: at higher altitudes (H1, 1,330 m; H2, 1,248 m) the fitted mean lies above the 1:1 line, whereas at lower altitudes (H3, 770 m; H4, 695 m) it is close to or below the 1:1 line (Fig. 12a).

The 95% confidence bands for H1, H2, and H3 overlap the 1:1 line across the observed range, suggesting approximate agreement in event duration relative to the UNIBO gauge, whereas the band for H4 does not overlap the 1:1 line for event durations longer than 25 hours. The band for H1 is wider than the others, indicating greater uncertainty in how event durations at H1 relate to UNIBO, whereas H4 shows a narrower band, suggesting a tighter relationship (Fig. 12b).

Lines 368-370: how are these random samples taken? Uniform distributions over x and y? Normal distributions? Is the correlation between I and D considered? Since I is calculated from D, there is a correlation between the variables that must be accounted for in such an analysis (lower D in one station implies that higher I is more likely than lower I, etc.). ID and ED thresholds are equivalent from several points of view, but not from this one. I believe the blue area UBR in Figure 13 cannot be interpreted, and the conclusion that "the impact of spatial variability on the threshold definition is moderate" cannot be stated unless the points above are clarified and, if necessary, amended.

Random samples are taken following a uniform distribution over *x* and *y* because we consider all precipitation values within the maximum–minimum duration and maximum–minimum intensity intervals across the five rain gauges to be equally probable. Although there is indeed a correlation between duration and intensity, we believe that the blue area (UBR) is still valid because the random samples are drawn from within a distribution of real measured data. This important information on the distribution and the correlation between the variables has been added to the revised manuscript.

To further explore this aspect, 10,000 random simulations were performed, in which for each rainfall event a random point was sampled within the uncertainty rectangle defined by the minimum and maximum observed values of duration and intensity. The samples were generated following a uniform distribution, so that each value within these ranges had the same probability of being selected. For each simulation, the corresponding TH1 threshold was calculated using the LDA method, resulting in an uncertainty band (Uncertainty Band for

Rainfall variability, UBR; Fig. 13b). Although duration and intensity are correlated, this approach ensures that the sampling still reflects the variability captured in the measured data, since all random points are confined to the ranges derived from the observations.

New references:

Bernard, M., Barbini, M., Berti, M., Boreggio, M., Simoni, A., and Gregoretti, C.: Rainfall-Runoff Modeling in Rocky Headwater Catchments for the Prediction of Debris Flow Occurrence, Water Resources Research, 61, e2023WR036887, https://doi.org/10.1029/2023WR036887, 2025.

Berti, M., and Simoni, A.: Experimental evidences and numerical modelling of debris flow initiated by channel runoff, Landslides, 2, 171-182, https://doi.org/10.1007/s10346-005-0062-4, 2005.

Berti, M., Bernard, M., Gregoretti, C., and Simoni, A.: Physical Interpretation of Rainfall Thresholds for Runoff-Generated Debris Flows, J. Geophys. Res. Earth Surf., 125, https://doi.org/10.1029/2019JF005513, 2020.

Gariano, S. L., Melillo, M., Peruccacci, S., and Brunetti, M. T.: How much does the rainfall temporal resolution affect rainfall thresholds for landslide triggering?, Nat Hazards, 100, 655–670, https://doi.org/10.1007/s11069-019-03830-x, 2020.

Gregoretti, C. and Fontana, G. D.: The triggering of debris flow due to channel-bed failure in some alpine headwater basins of the Dolomites: analyses of critical runoff, Hydrol. Process., 22, 2248–2263, https://doi.org/10.1002/hyp.6821, 2008.

Gregoretti, C., Degetto, M., Bernard, M., Crucil, G., Pimazzoni, A., De Vido, G., Berti, M., Simoni, A., and Lanzoni, S.: Runoff of small rocky headwater catchments: Field observations and hydrological modeling, Water Resour. Res., 52, 8138–8158, https://doi.org/10.1002/2016WR018675, 2016.

Kirpich, Z. P.: Time of concentration of small agricultural watersheds, Civ. Eng., 10, 362, 1940.

Marra, F.: Rainfall thresholds for landslide occurrence: systematic underestimation using coarse temporal resolution data, Nat Hazards, 95, 883–890, https://doi.org/10.1007/s11069-018-3508-4, 2019.

Simoni, A., Bernard, M., Berti, M., Boreggio, M., Lanzoni, S., Stancanelli, L. M., and Gregoretti, C.: Runoff-generated debris flows: Observation of initiation conditions and erosion–deposition dynamics along the channel at Cancia (eastern Italian Alps), Earth Surf. Process. Landforms, 45, 3556–3571, https://doi.org/10.1002/esp.4981, 2020.

Soil Conservation Service (SCS): Section 4: Hydrology, in: National Engineering Handbook, U.S Department of Agriculture, Washington DC, 1972.