Response to Reviewer #2

We sincerely thank the reviewer for the thorough evaluation of the manuscript and for the constructive and encouraging comments. Our responses to each remark and comment are provided below in blue.

General comment
This study addresses a relevant question: how do landslides triggered by extreme rainfall events affect the calibration of statistically-based prediction tools. Using the case of Storm Alex, the authors demonstrate quantitatively that including such events in landslide inventories substantially alters both spatial and temporal prediction tools. The work is clear, well-written and structured. Figures are clear and useful. The results are supported by the data. The discussion is useful to highlight the strengths and limitations of the work. The sensitivity analysis on the assumed hour of landslide occurrence is valuable and strengthens the robustness of the rainfall threshold results. However, the implications for operational early warning system design could be discussed more explicitly in the conclusions.
Overall, the manuscript is interesting for NHESS readers and it can be published after major revisions.
I've found three main methodological issues in (1) the use of land cover map; (2) the definition of non-landslide sample and (3) the calculation of the rainfall thresholds. These should be addressed before the manuscript can be reconsidered for publication. Moreover, I've found some other minor technical comments.

Predisposing factors:
I wonder why the authors used the 2006 release of the Corine Land Cover, while more recent releases are available. Using a more recent Corine release would allow a better representation of the land cover conditions in the occurrence of the Alex storm.

Response:

The available Corine Land Cover datasets correspond to the years 1990, 2000, 2006, 2012, and 2018. A single dataset had to be selected for the analysis, and the 2006 dataset was chosen as it represents the median period of our landslide inventory. Changes in land cover between 2006 and later releases (2012 and 2018) are relatively minor in the study area, and when they occur, they affect only a limited number of landslides in the inventory. Moreover, land cover does not appear to be a dominant controlling factor in our case, as the majority of landslides are concentrated within only two classes (forests and scrub and/or herbaceous vegetation).

We specified why the 2006 Corine Landcover was chosen.

Modifications:

- ➔ The following sentence has been added in the section *3.2 Predisposing factor:*

"The 2016 Corine Land Cover dataset was chosen as it best represents the mid-point of the temporal period covered by the landslide inventory, ensuring temporal consistency between land cover information and observed landslide occurrences."

Landslide inventory:
The used inventory is comprehensive and accurate. However, a known problem of landslide inventories is their completeness, particularly if the dataset dates back to previous centuries, as in this case. From Figure 3b, it is clear that the inventory completeness is different before and after 1850. My suggestion is to remove all data previous to 1850 and consider only post-1850 data. The completeness and reliability of the inventory would benefit from this.

Response:

Indeed, 12 landslides in the inventory are dated prior to 1850 in Figure 3. These events have limited spatial precision (kilometer to hectometer scale). Consequently, due to their age and insufficient spatial accuracy, these pre-1850 landslides were not included in the susceptibility and threshold analyses. We acknowledge that this may have led to some confusion, as Figure 3 originally represented the complete landslide inventory rather than the combinations of subsets used in the analyses (see Section 3.3.2). To improve clarity, Figure 3 has been revised to display only the landslides considered in the analyses, therefore excluding both kilometric-scale events and riverbank landslides that were not used in any part of the modelling workflow. Additionally, the temporal distribution of landslides for each subset has been included in Figure 4.

Modifications:

- ➔ We have revised figures 3 and 4. The text associated with these figures has been revised accordingly.

Something not clear to me: at line 195 the authors state that 1655 landslides were directly associated with Storm Alex, while at line 213 they state that 1335 landslides are associated with Storm Alex. Please clarify.

Response:

River bank landslides were systematically excluded from all our analyses. This explains the lower number of landslides associated with Storm Alex reported in line 213: the difference corresponds entirely to river bank landslides that were initially counted but subsequently removed from the dataset. After the revisions to Figures 3 and 4 in response to the preceding comment, this issue is now clearer, as the same number of Storm Alex landslides is consistently reported.

ED rainfall thresholds:

- The warm and cold seasons should be defined.

Response:

The warm and cold seasons are now clearly defined in Section 4.1 of the revised manuscript.

Modifications :

- ➔ The paragraph talking about warm and cold seasons has been revised as follow:

"The triggering conditions, namely cumulative rainfall (E) and rainfall duration (D), were determined for each landslide following the same methodology as the CTRL-T algorithm. First, at each grid point of the COMEPHORE rainfall dataset, raw precipitation data were segmented into rainfall events and sub-events, defined as periods of continuous precipitation separated by dry intervals of minimum duration. During the warm season, these minimum dry intervals were set to 48 h for rainfall events and 6 h for sub-events. These durations are multiplied by an R factor during the cold season to account for seasonal differences in evapotranspiration. This R factor is computed as the ratio between potential evapotranspiration during the warm and cold seasons. Following Barthélemy et al. (2024), the start and end of the warm season (SWS and EWS, respectively) were estimated for each grid point using the SIM climatic dataset. In the study area, SWS varies from February to June, whereas EWS ranges from August to September depending on local climatic conditions. Consequently, the R factor varies spatially between 1 and 3. "

Non-landslide sampling:

Why the author selected the same number of landslide and non-landslides points? Usually, in susceptibility and hazard analyses, an imbalance among landslide and non-landslide points is adopted, with a larger number of non-landslide points. See e.g. Steger et al. (2023, 2024); Nocentini et al. (2024). Indeed, this is much more consistent with physical reality, given that landslides are rare phenomena in both space and time. I suggest to use an imbalanced dataset of non-landslides points.

Response:

While it is true that landslides are rare phenomena in both space and time, our choice of a balanced dataset is motivated by methodological considerations related to the modeling approach.

First, the Random Forest algorithm is known to be sensitive to class imbalance. In such cases, the model tends to favor the majority class, leading to biased predictions and a reduced ability to correctly identify areas prone to landslides.

SSecond, although imbalanced datasets are often appropriate in temporal hazard modelling, where the objective is to represent the rarity of triggering events, the aim of this study is landslide susceptibility mapping, which does not explicitly account for the temporal dimension. In this context, the objective is to discriminate between stable and unstable areas based on their geomorphological characteristics, rather than to replicate the real-world imbalance between landslide and non-landslide occurrences. For this reason, a balanced sampling strategy is commonly adopted in susceptibility modelling studies (e.g. Nocentini et al., 2024; Taalab et al., 2018; Akinci et al., 2020).

Result – ED thresholds
Looking at Figure 6, I see some issues.
First, I see some points in the graph with very low ED conditions (i.e. all points with cumulative rainfall lower than 10 mm, even with long durations). I think that these conditions are not realistic – probably due to errors in the rainfall products used or in the landslide timing of location, or rainfall underestimations. In any case, I suggest to remove these points, given that they represent rainfall conditions that can't realistically trigger landslides.
Second, the uncertainty region of the WSAL threshold is larger than that of WoSAL one, despite the larger number of points (1704 for WSAL and 371 for WoSAL, if I have well understood). This can be done to two causes:

1) the distribution of the residuals for WSAL dataset is not Gaussian, as partially mentioned by the authors in the discussion;

2) the 1333 ED conditions for the Alex storm are mostly repeated several times, i.e. data points corresponding to the same rainfall conditions (the same landslide and the same weather cell) are repeated multiple times. Also this is partially mentioned in the discussion.

If these two issues are confirmed, I have to say that both are not acceptable in the application of the frequentist method included in the CTRL-T tool. Indeed, if the point distribution is not gaussian, the method can't be applied. Moreover, data points repetition should be avoided, because this alters the statistics behind the method used to define the thresholds. As an example, if the same ED condition is associated to more than one landslide, this is usually plotted and used in the calculations only once.
I suggest to check these issues and correct them if necessary.

Response:

We thank the reviewers for these remarks that led us to improve the ED threshold analyses in the manuscript.

A minimum cumulative rainfall threshold of 5 mm is implemented in the CTRL-T algorithm to avoid unrealistic ED conditions. However, we acknowledge that this value may still be too low. In response to this suggestion, we re-analysed the dataset and increased the minimum rainfall threshold to 10 mm in order to retain only more physically meaningful rainfall–landslide conditions.

We also recognise the issue of repeated ED conditions resulting from multiple landslides triggered by identical rainfall inputs. We agree that retaining such duplicates may introduce a strong imbalance, particularly for events that produced a large number of landslides, such as Storm Alex. To address this, we re-ran the analysis using the MPRC dataset after removing duplicated rainfall–landslide conditions. This clustering resulted in a final set of 515 MPRCs (E, D) couples used for threshold computation, including 201 associated with Storm Alex, compared to an initial dataset of 1,704 MPRCs, of which 1,333 were related to Storm Alex.

Following this step, we recalculated the thresholds using the CTRL-T frequentist approach and observed that the residuals exhibited a bimodal distribution when Storm Alex was included. As a result, we finally chose not to rely on the frequentist method implemented in CTRL-T for threshold estimation. Instead, we adopted a 5% quantile regression approach, while still using the 515 clustered MPRCs derived from CTRL-T, as quantile regression does not assume a normal distribution of the residuals.

Note that the clustering of MPRCs resulted in much closer threshold estimates between the cases with and without Storm Alex. Consequently, the *Results, Discussion* and *Conclusions* sections have been revised accordingly.

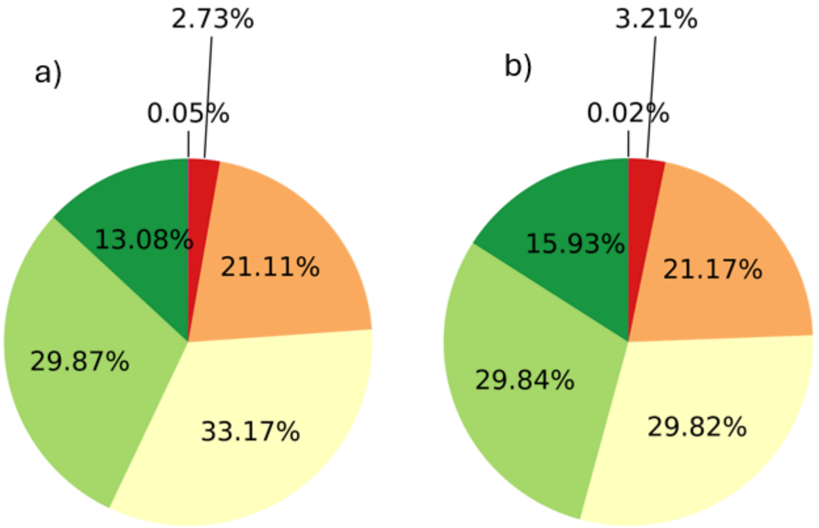
Modifications :

- ➔ Figure 6 has been replaced to present rainfall thresholds computed using the 5% quantile regression approach instead of the CTRL-T frequentist method.
- ➔ The threshold estimation methodology has been updated to rely on quantile regression.
- ➔ The revised results show that rainfall thresholds estimated with and without Storm Alex are similar.
- ➔ The effect of the clustering is discussed in the section *6.2 Impact of Storm Alex landslides on rainfall triggering thresholds* and the thresholds obtained without clusterisation are added in supplementary materials.

Results – Landslide susceptibility maps:
- Figure 9: I suggest adding two pie charts to show the percentage of cells in each susceptibility class. Moreover, I would use only one decimal digit. Please use point as decimal separator.

Response:
The inclusion of pie charts of susceptibility classes does not provide additional insight, as it does not reveal meaningful differences in the spatial extent of susceptibility classes between WoSAL and WSAL maps, and depends strongly on how continuous susceptibility values are discretised. For this reason, we instead compare the maps using the continuous distribution of susceptibility values through kernel density estimation (KDE), which more effectively

captures variations in susceptibility. The distribution of susceptibility classes between WoSAL and WSAL maps are presented below, but are not included in the revised manuscript.



Distribution of susceptibility class areas within the Alpes-Maritimes region (a) for the WoSAL map (without Storm Alex landslides) and (b) for the WSAL map (including Storm Alex landslides).

Technical corrections

- Please correct "rainfalls" with "rainfall" everywhere in the text.
- L95-97: Please correct the percentage values. Their sum is now 105%.
- L102: Please correct "mm.yr-1"
- L118: I would say "The storm was responsible…"
- L196: Please correct "Ales"
- L355: Perhaps "44" should be corrected with "4"
- L400: Not clear why those references are mentioned in this line.
- Figures 9 and 11: Please use points as decimal separators.

Response:

Thank you for these technical corrections.

Note that the percentage values are not intended to sum to 100%, as they refer to partially overlapping categories. In particular, sedimentary rocks represent 70% of the territory, within which marly and clayey formations account for 14% of the territory for example. To avoid any confusion, the text has been slightly revised to clarify this point.

Modifications :

➔ The paragraph in the section *2.1 Study area* referring to percentage values of bedrocks has been revised as follow:

"The harmonized geological map of the Alpes-Maritimes (Gonzales, 2008) indicates that the territory exhibits a highly diverse geology, with a majority (77%) of sedimentary rocks (Fig. 1b). The latter include marly and clayey formations (14% of the territory), which weather into clay-rich soils that are highly sensitive to water saturation and alteration. In addition, recent unconsolidated deposits such as moraines (3% of the territory), alluvial deposits (5% of the territory) and scree or debris cone deposits (20% of the territory), represent unstable materials that can further increase the susceptibility to landslides.

- ➔ Figures 9 and 11 has been revised to consider your remarks
- ➔ All the other technical corrections were considered in the revised version of the manuscript

References:

Nocentini, N. et al (2024) Regional-scale spatiotemporal landslide probability assessment through machine learning and potential applications for operational warning systems: a case study in Kvam (Norway). Landslides 21, 2369–2387 (2024). https://doi.org/10.1007/s10346-024-02287-9

Taalab, K. et al. (2018). Mapping landslide susceptibility and types using Random Forest. Big Earth Data, 2(2), 159–178. https://doi.org/10.1080/20964471.2018.1472392

Akinci, H.; Kilicoglu, C.; Dogan, S. Random Forest-Based Landslide Susceptibility Mapping in Coastal Regions of Artvin, Turkey. ISPRS Int. J. Geo-Inf. ***2020****, 9, 553. https://doi.org/10.3390/ijgi9090553*