General response

We would like to thank the Editor and both reviewers for their careful evaluation of our manuscript and for their constructive comments. We have carefully considered all suggestions and believe that they have helped us improve the quality and clarity of the manuscript.

In addition to the point-by-point responses provided below, we have thoroughly revised the manuscript to improve its overall readability, language, and flow. We also corrected typographical errors, clarified several passages, and improved the consistency of the text where appropriate.

Furthermore, during the revision process, we identified a slight temporal offset in the meteorological dataset used for the analyses. After correcting this issue, we recomputed the rainfall metrics and recalculated the rainfall thresholds presented in the manuscript. This correction resulted in minor changes to some of the values and a slight shift in the position of some data points in Figure 4. However, these changes do not affect the overall results, interpretations, or conclusions of the study.

The original comments of the reviewers are reproduced below in bold black, our responses are provided in blue, and the corresponding modifications applied to the manuscript are indicated in orange.

Response to Reviewer #1

The manuscript presents a solid and well-structured analysis of the impact of extreme rainfall events on landslide rainfall thresholds and susceptibility modelling. The dataset is comprehensive, and the methodological framework (CTRL-T and Random Forest) is appropriate and carefully implemented. The results are clear and provide meaningful insights into how extreme events can alter statistical models used for landslide prediction. Overall, the manuscript is suitable for publication after moderate revisions. However, several aspects could be further strengthened to improve clarity, robustness, and broader applicability.
First, the discussion section should be slightly expanded to better generalize the findings beyond the study area, particularly addressing whether similar effects of extreme rainfall events on thresholds and susceptibility can be expected in other climatic and geomorphological contexts.

Response:

We thank the reviewer for this suggestion. In response, we have expanded the discussion by adding a paragraph at the end of section *6.4 Changes in susceptibility maps due to Storm Alex landslides* to explicitly address the potential generalization of our findings to other climatic and geomorphological contexts, supported by literature.

Modifications:

➔ The following paragraph has been added at the end of the section 6.4 :

"Beyond the specific case of the Alpes-Maritimes region, our results also suggest broader implications regarding the role of extreme rainfall in landslide triggering. The observed activation of areas with lower pre-event susceptibility supports the hypothesis that extreme hydrometeorological forcing can partially override the usual controls exerted by soil and terrain properties on slope stability. Similar effects may occur in other regions exposed to short-duration, high-intensity rainfall events, such as in mediterranean and tropical environments.

Achu et al. (2024) investigated the impact of landslides triggered during extreme rainfall events on the performance of machine learning models and susceptibility maps using a Deep Neural Network approach in the southern Western Ghats (Kerala, India), a humid tropical region affected by recurrent high-intensity monsoon rainfall. Their results showed that a model trained without landslides associated with extreme events remained capable of identifying landslides triggered during such events. Nevertheless, incorporating these landslides into the susceptibility modelling increased the spatial extent of low and extreme susceptibility classes (+120% and +32% relative increase, respectively), while the moderate and high susceptibility classes decreased (-15% and -42% relative increase, respectively). The area associated with the very low susceptibility class does not present a significant change.

Our results partially agree with those of Achu et al. (2024). Similar to their observations, the inclusion of landslides triggered during Storm Alex leads to an increase in the area associated with low susceptibility values (<0.2), increasing from 27% in WoSAL to 31% in WSAL (+15% of relative increase), and a decrease in the area associated with intermediate susceptibility values (0.2–0.6), decreasing from 62% to 57% (–8% of relative decrease) (Fig. 10). However, in our study, the proportion of the study area associated with high susceptibility values (>0.6) slightly increases from 11% to 12% (+9% of relative increase). Overall, the magnitude of these variations remains less pronounced than that reported by Achu et al. (2024). These differences may partly result from the use of different susceptibility thresholds to compare spatial susceptibility distributions. In Achu et al. (2024), susceptibility maps are divided into five classes, although the specific classification thresholds are not detailed. In contrast, our reported percentages are based on a simplified three-class partitioning of the continuous susceptibility values using fixed thresholds (0.2 and 0.6). These thresholds were selected to highlight the ranges where differences between the two maps are most pronounced in Fig.

10. In addition, contextual factors are likely to play a role. Indeed, the influence of extreme rainfall events on susceptibility patterns likely depends on both the regional geomorphological context and the rarity of the extreme rainfall event. While Achu et al. (2024) indicate that their extreme rainfall events were associated with return periods exceeding 100 years, Storm Alex was characterized by a millennial return period. Nevertheless, both studies show that incorporating landslides triggered by extreme rainfall events modifies susceptibility patterns."

Second, the limitations of the study should be more explicitly acknowledged, especially regarding the spatial representativeness of the Storm Alex landslides, the temporal inconsistency of the inventory, and the assumptions made in rainfall–landslide matching (e.g., fixed occurrence time).

Response:

We have revised the manuscript to better highlight the limitations of the study. Regarding the spatial representativeness of the Storm Alex landslides, we have expanded the discussion in Section *6.1 Quality of landslide inventory* to better highlight their strong spatial clustering and the fact that they are concentrated within a limited area. The issue of temporal inconsistency in the inventory, i.e. differences in temporal precision, is also addressed in this section, with a reminder that only landslides with sufficient temporal accuracy were used for the computation of rainfall thresholds.

Finally, the question of the use of a fixed occurrence time, is already discussed in the section *5.1 ED threshold*. This paragraph has been slightly expanded to further address this point and to explicitly account for the reviewer's comment.

Modifications:

➔ The following paragraph has been added to the section 6.1 :

"It should also be noted that Alex landslides exhibit a strong spatial clustering, mainly concentrated within the Vésubie, Tinée and Roya valleys. This clustering primarily reflects the actual spatial extent of the storm impacts, as these valleys correspond to the areas that experienced the most intense rainfall during Storm Alex (Fig. 2). Most Alex-induced landslides were identified through orthophotography comparison using the post-Alex orthophoto, which specifically covered these heavily affected sectors (Fig. 3). Although it is possible that a limited number of Alex-induced landslides located outside the orthophoto extent have not been identified, such cases are expected to remain minor given the localized nature of the storm and the concentration of the most extreme rainfall within the surveyed valleys. Consequently, the strong spatial clustering of Alex landslides is considered to reflect the actual spatial distribution of landslide occurrence during the event, rather than being an artefact of the mapping procedure. To further limit the influence of this clustering, the

analysis of the distributions of predisposing factors (Fig. 8) was restricted to the spatial extent covered by the post-Alex orthophoto for both Alex and non-Alex landslides. Despite this spatial restriction, clear differences between Alex and non-Alex landslides were still observed, suggesting that the identified contrasts are not only related to the spatial concentration of the Alex inventory, but reflect actual differences in the geomorphological conditions under which landslides were triggered during the storm."

- ➔ The following sentence has been added to Section 6.1 to address the inconsistency of the inventory :

"Only landslides with reliable dating (daily or hourly resolution) were retained for this analysis, representing about 40% of the total inventory (Fig. 3d) "

- ➔ The section *5.1 ED threshold* has been completed as follow:

"Recall that the default hour of occurrence of landslides is set at 12:00 in the CTRL-T algorithm. This simplification may affect the estimation of the rainfall conditions associated with each event, particularly for short-duration or high-intensity rainfall, where the timing of the triggering rainfall relative to the landslide occurrence can significantly affect the calculated event duration and cumulative rainfall. To assess the influence of this arbitrary timing, sensitivity tests were performed by changing the assumed landslide occurrence time (00:00, 06:00, 12:00, 18:00, and 23:00) (Fig. A4). This resulted in noticeable changes in the obtained thresholds. However, in all cases, the threshold including the Alex landslides remained higher than the threshold computed without them, by a factor of 1.1 to 1.5 "

Third, a brief perspective on future research directions would be valuable, for example the need for non-stationary modelling frameworks, event-based calibration strategies, or hybrid approaches distinguishing between ordinary and extreme triggering conditions.

Response:

The subject of non-stationary modelling frameworks, i.e. dynamic susceptibility varying with actual precipitation conditions, is already addressed in the perspectives section. The conclusion has nevertheless been expanded to include additional future research directions, in line with the other suggestions raised in this comment.

Modifications:

- ➔ The paragraph of the conclusion talking about perspectives has been revised as follow :

"From a methodological perspective, several prospects emerge from this study. Approaches explicitly considering rainfall event return periods, for instance by adapting model calibration or by using separate modelling frameworks according to rainfall return periods, could improve model robustness

under extreme rainfall conditions. Additionally, our results point out the importance of jointly considering predisposing (e.g., slope, lithology, land use) and triggering factors (e.g., rainfall intensity and duration) in landslide susceptibility assessment. Statistical models integrating both predisposing factors and triggering variables would presumably better capture the full range of conditions under which landslides occur, thereby improving hazard assessment and risk management strategies. Incorporating additional parameters, such as antecedent soil moisture, snowmelt indicators, or effective rainfall rather than event-based rainfall metrics, could further enhance the predictive capability of these models. Finally, performing analyses by landslide type (e.g., flow-like landslides, mid-slope landslides and cut-slope landslides) could also be relevant, as different processes are controlled by distinct conditioning and triggering mechanisms; however, such stratification inevitably reduces the number of available landslides for model calibration and validation, potentially limiting statistical robustness and requiring larger or longer-term inventories."

Finally, moderate clarifications could be added regarding the influence of non-landslide sampling strategy and the potential uncertainty introduced by converting landslide polygons into points. With these moderate improvements, the manuscript will be further strengthened and make a valuable contribution to the field.

Response:

We have addressed these points in the revised manuscript. Clarifications regarding the influence of the non-landslide sampling strategy are provided in section *4.2.2 Non-landslides samplings*, as well as the potential uncertainty introduced by converting landslide polygons into points, have been added respectively to the section *6.1 Quality of the landslide inventory*.

Modifications:

➔ The section 4.2.2 has been revised as follow:

"In the context of supervised landslide susceptibility mapping, the selection of non-landslide samples is a crucial step that can strongly influence model performance and final results (Fu et al., 2025). In particular, it controls the statistical contrast between landslide and non-landslide conditions used to train the RF model. Sampling strategies that preferentially select non-landslide locations in clearly stable areas may increase class separability and produce more contrasted susceptibility patterns. In contrast, sampling designs that rely on a broader or less constrained selection of non-landslide locations may include areas close to past landslide locations.

A commonly adopted approach in landslide susceptibility studies is to generate non-landslide points outside a buffer zone around known landslides (Gu et al., 2023, 2024). Building on this principle, and aiming to capture the geomorphological variability of the study area without imposing deterministic assumptions on terrain stability, we adopted a random sampling approach within a spatially constrained domain. This domain was designed to ensure comparable reporting conditions and reduce potential bias. Specifically, it was defined as a 1 km buffer around major roads, while excluding areas located within a 150 m radius of recorded landslides (Fig. A3). This design reduces potential reporting bias, as landslides occurring near infrastructures are more likely to be documented.

Within this constrained domain, non-landslide points were generated using random sampling. To account for variability in non-landslide selection, 50 distinct non-landslide datasets were generated for each scenario (i.e., with and without Alex landslides), keeping landslide points fixed while randomly generating non-landslide points within the valid domain. A balanced 1:1 ratio between landslide and non-landslide points was maintained to avoid class imbalance.

Non-landslide points were first generated for the WSAL dataset (i.e. including Alex landslides). Each dataset therefore includes 4,200 landslide points and 4,200 non-landslide points. To generate the corresponding WoSAL datasets (i.e. excluding Alex landslides), we started from the WSAL datasets. In each dataset, the 1,319 Alex landslides were removed, and an equal number of non-landslide points was randomly excluded. This approach avoids regenerating non-landslide points, which could otherwise introduce spatial inconsistencies, and ensures maximum comparability between WSAL and WoSAL datasets. The full procedure is summarized in Fig. 5."

➔ The following paragraph has been added at the end of the section 6.1:

"Moreover, for random forest (RF) modelling, landslides originally mapped as polygons were simplified as single points for the RF modelling. This approach ensures a homogeneous treatment of the inventory, although it may introduce some uncertainties in the values of predisposing factors (e.g., slope) attributed to each landslide."

Response to Reviewer #2

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 2006 Corine Land Cover dataset was chosen as it best represents the mid-point of the study period covered by the landslide inventory (see Sect. 3.3), 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 by 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 reported between the inventory and the subset used for thresholds computation.

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 methodology implemented in the CTRL-T algorithm (Melillo et al., 2015, 2018). First, for each grid point of the COMEPHORE rainfall dataset, precipitation time series were segmented into rainfall events and sub-events, defined as periods of continuous rainfall separated by minimum dry intervals. During the warm season, the minimum dry interval was set to 48 h for rainfall events and 6 h for sub-events. During the cold season, these durations were multiplied by a factor *R*, defined as the ratio between warm-season and cold-season actual evapotranspiration. Following Barthélemy et al. (2024), the start (*sws*) and end (*ews*) of the warm season were estimated from the SIM climatic dataset and used to define the seasonal parameters applied to the COMEPHORE precipitation series. Across the study area, *sws* varies from February to June, whereas *ews* ranges from August to September depending on local climatic conditions. Accordingly, the local *R* factor ranges from 1 to 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.

Second, 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 aggregating duplicated rainfall–landslide in a single conditions. This aggregation resulted in a final set of 518 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 1,332 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 518 aggregated MPRCs derived from CTRL-T, as quantile regression does not assume a normal distribution of the residuals.

Note that the aggregation 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.

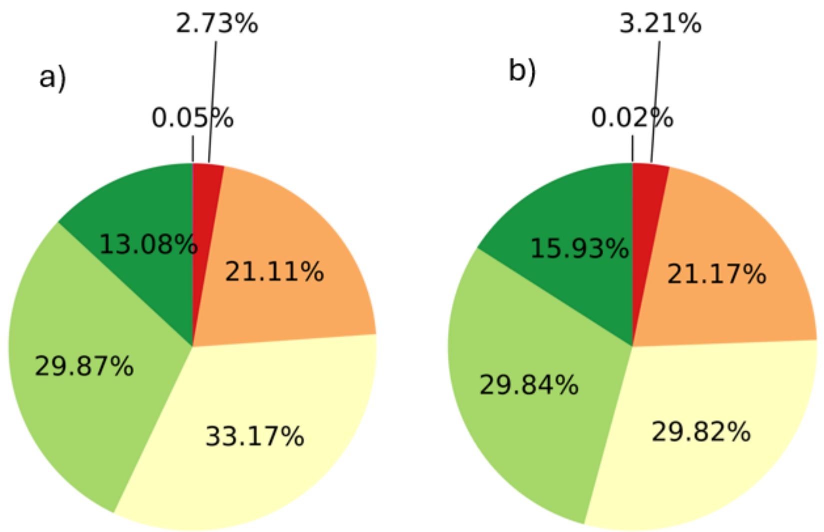
- ➔ The threshold estimation methodology has been updated to rely on quantile regression.
- ➔ The revised results show that rainfall thresholds estimated with and without Storm Alex are similar.
- ➔ The effect of the clustering is discussed in the section *6.2 Impact of Storm Alex landslides on rainfall 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