

Dear Konstantis Alexopoulos, thank you for your reading of our manuscript and for your constructive comments. Below is a point-by-point response to the issues raised. Comments concerning the form are not addressed here but will be incorporated into the revised manuscript.

- First, I am curious as to why authors used massif features instead of watershed features. This might seem counterintuitive to readers, for a study on hydrological impacts. I would think that watersheds are more appropriate features for studying floods, but also landslides, as they may increase the debris input in the hydrographic network, leading to infrastructure damages downstream. I think this choice needs to be addressed/explained better, especially since the authors do refer to watersheds in a couple of occasions in the manuscript.

This point indeed deserves a clearer justification in the article. The watersheds responsible for impacts such as torrential floods or landslides are small in spatial extent. For example, the flooding of the Étançons torrent on 21 June 2024, mentioned in L411–412, involves a watershed of 34.4 km². A subdivision at the scale of such small watersheds would split the study area into a very large number of small units, making it difficult to present synthetic results over the whole French Alps in maps and figures. Moreover, each unit would contain very few, if any, recorded natural hazard events. An approach based on the large watersheds of mountain rivers would not be more relevant either, owing to the mismatch in spatial scale with RTM events. An approach based on the large watersheds of mountain rivers was not chosen in our work, because (1) most events reported in the RTM database concern much smaller spatial scales and (2) it would have introduced a discrepancy with the native geometry of the S2M meteorological dataset, used in this study and considered as a reference for French Alps by Meteo-France.

The details of this justification will be introduced in the revised version of the manuscript.

- Second, while the authors account for total rainfall amount (as well as duration) of ROS events, they don't seem to consider at all the underlying snowmelt contribution. As a result, I'd be more cautious in attributing a flood to ROS, instead of just rainfall, since the contribution of the snowmelt fraction of runoff (which is even accelerated during the ROS event, e.g., Yang et al. (2023)) is unknown. To me at least this appears to be a substantial limitation, which I am sure the readers would like to see addressed, or at the very least discussed.

We agree that this aspect deserves to be addressed. Our study focuses on the meteorological occurrence of rain-on-snow, without accounting for the effect of snowmelt. Our input data are the hydrometeorological forcing, but hydrological analysis

is beyond the scope of our study. This is indeed a limitation, and the effect of snowmelt may vary among impactful ROS events. However, from a risk-prevention perspective, we chose to keep the approach simple by relying on readily available and predictable variables (snow presence and rainfall accumulation).

To address this point, we carried out the suggested processing, dividing the daily ΔSWE by 24 h. Comparing SWE at the beginning and end of each event ($\Delta SWE = SWE_{end} - SWE_{start}$), the snowmelt contribution is highly variable, with many outliers and some events showing a positive ΔSWE (Fig I below). This variability could be explained by the state of the snowpack at the beginning of the event: an unsaturated snowpack can absorb a certain amount of liquid water, resulting in a positive ΔSWE . In addition, according to their definition (Section 3.1), ROS events may include snowfall between rainfall periods. For all ROS events, the median ΔSWE is -2 mm ($Q_1 = -10$ mm, $Q_3 = 3$ mm), and for impactful ROS events it is -13 mm ($Q_1 = -25$ mm, $Q_3 = -2$ mm). The snowmelt contribution for impactful ROS events remains much lower than the total rainfall (median of 56 mm). Although a few events show outliers reaching about 60 mm of snowmelt, rainfall thus remains the main contributor to discharge. Moreover, snowmelt is not the only hydrological role played by snow cover. Its antecedent presence may have wetted the watershed (hydrological yield being higher on a saturated watershed), and, depending on its structure, the snowpack may shorten the time of concentration (Brandt et al. 2022). These effects of snow cover are complex processes that are worth investigating at the watershed scale, but they are beyond the objective of this large-scale study. These elements will be included as a new section of the Discussion in the revised manuscript.

Brandt WT, Haleakala K, Hatchett BJ and Pan M (2022) A Review of the Hydrologic Response Mechanisms During Mountain Rain-on-Snow. *Front. Earth Sci.* 10:791760. doi: [10.3389/feart.2022.791760](https://doi.org/10.3389/feart.2022.791760)

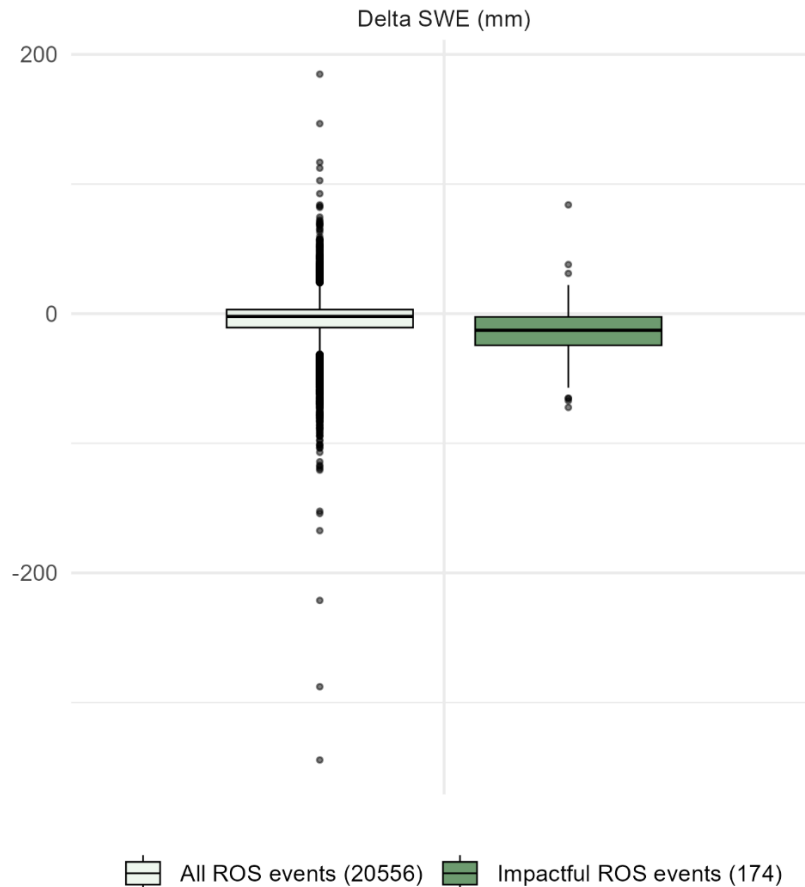


Figure 1 : Boxplots of ΔSWE per ROS event considering all events or restricting to impactful ones, over the 2001–2024

Introduction

Minor

- L18-20: Are the authors describing all ROS events, or maybe just the impactful ones? Would there be a different definition for ROS events where the existing snowpack is not yet close to saturation? I think that that definition given by the authors in L69 is more appropriate and therefore I would encourage a slight rephrasing here.

A rephrasing based on the definition given in L69 is indeed more appropriate. Thank you for this suggestion.

- L72: Could the authors please provide some more clarity as to what is the difference between “Impactful” and “Heavy-rainfall” ROS events? Are the latter also impactful? In the following lines (L73-76) only “all” and “impactful” ROS events are mentioned – “Heavy-rainfall” ones aren’t – causing confusion as to what is the definition, and purpose, of that last category.

Heavy-rainfall ROS events are a proxy for impactful ROS events, needed to study temporal evolution without the reporting bias affecting recorded natural hazard events (L223–227). A clarification will be added as early as the Introduction.

Data

Major

- L85: My main concern regarding the study area and subsequent analysis, is that uses massif features instead of catchment features. This seems like a counterintuitive choice for an analysis of a hydrologically relevant topic, such as flooding (as linked with the RTM database later). While I would strongly encourage this, I understand that changing this now would take a certain amount of effort & time. If the authors chose to keep their analysis at massif scales, then please provide an explanation regarding this choice, as well as its possible implications for the results.

See the response in the preliminary remarks.

- L104-106: It is unclear if the snowpack is calculated with or without taking into account the aspect, which would have a substantial effect on the snowpack properties due to different shading & exposure to incoming shortwave radiation, affecting its melting, which would be highly relevant to the prediction of a ROS event.

S2M provides two datasets: one accounting for aspect and one not. For our study, we used the dataset providing snow simulations on flat terrain, assumed to be a reasonable average estimate of snow cover extent at massif scale. Aspect obviously has a substantial effect on snow presence, but accounting for this source of variability within an exhaustive study of the entire French Alps would have had a new level of complexity (e.g. presenting results for each aspect), while RTM events could not be easily associated with a specific aspect. Indeed, these events may be spatially extended (L209–211) and cover various slope aspect classes., The assumption to ignore aspect-dependence will be more explicit and justified in the revised manuscript.

- Section 2.3: At this point it is still unclear which S2M properties are validated against the observations, as well as whether the S2M reanalysis was assimilated with any of the same weather stations, rendering them unfit for use in the validation stage. Shouldn't the validation happen against an independent dataset? This point is also addressed by the authors at the beginning of the results

section (L247-250) but it would be better if it were moved here or in the methods section.

Thank you for this comment, which highlights an ambiguity in the way we presented this comparison. The purpose of the comparison with observation stations is not to “validate” S2M, but to visualize the differences resulting from the change of spatial scale. We show how ROS event properties differ between the point scale of stations, which reflects what actually occurs, and the properties aggregated at the massif scale. The results are not essential to the study, but we find them interesting. To avoid confusion, and since these results are not used outside Section 4.1, everything related to the comparison with observation stations will be moved to an appendix. Note also that the confidence in massif-scale precipitation of S2M rely on the existing literature about this well-established reanalysis (Durand et al., 1993, 2009 ; Vernay et al., 2022).

Methods

Major

- L181-182: Once again, I am puzzled about the authors’ choice to not analyse the data per hydrological catchment, and I think this requires further explanation in the Section 2.1. Even in L212-213, the authors point out that impact may had been reported on watershed scales, but without explanation, they analyse data in massif scales.

See the response in the preliminary remarks.

- L235: Which statistical test was used to calculate the trends, and which test was used to compare the two 30-year periods? This should also be made clear when reporting results later in the manuscript (especially in figure captions) to improve clarity. In the results section later, I have commented on the authors’ somewhat misleading use of the word ‘trend’ when comparing absolute numbers between two periods, instead of running a formal trend analysis across the time series (e.g., Mann-Kendall, Theil-Sen).

What we refer to as trends in the article is the difference between the means of two 30-year periods (1959–1988 and 1995–2024). This choice was motivated by the strong interannual variability (which rules out any parametric modelling) and by the numerous zeros when results are broken down by elevation band and period of the year. The difference provides a non-parametric, quantitative result. The Mann–Kendall test does not provide this quantitative aspect, and the Theil–Sen estimator yields null results for heavy-rainfall ROS events because of the numerous zeros.

The statistical test used is a permutation test (also called re-randomization test or shuffle test), described in L237–240. It is applied to the annual values that make up the means of the two compared periods. This is a standard approach for comparing the means of two samples.

Minor

- L220-221: Could you please expand further on how this approach accounts for the potential extent of the contributing watershed, as it is currently unclear to me?

This is indeed a bold assumption, implying that what is observed at the massif scale is representative of the small watersheds responsible for impacts. From a meteorological point of view, however, we consider it reasonably acceptable, as ROS events are generally associated with widespread perturbations rather than with localized summer thunderstorms. We retain the elevation band with the highest total rainfall because it represents the most severe conditions experienced within the massif, and therefore the highest potential for triggering natural hazard events. In addition, the runoff generated in this band propagates to lower elevations. We acknowledge that our approach does not account for the fraction of the watershed area affected by rainfall, which depends on the elevation band considered. Assessing this effect would require a hydrological analysis at the watershed scale, which is beyond the scope of this large-scale study.

- L229: Could you please provide some further detail and explanation on how you came about selecting the 50% threshold and what this exactly means for the subsequent analysis?

Regarding the choice of the threshold, we are constrained by the results shown in Fig. 9b. It is not possible to exceed a proportion of impactful ROS events of about 70% based on total rainfall. The choice of 50% allows a very simple interpretation: at this threshold, one ROS event out of two is impactful. It can be inferred that a ROS event exceeding this threshold has a one-in-two chance of being impactful, i.e., of triggering natural hazard events. Furthermore, Fig. 9b shows a fairly linear relationship between the applied threshold and the proportion of impactful ROS events up to a proportion of about 50%. Beyond that, the increase in the proportion of impactful ROS events with the total rainfall threshold is smaller. The 50% value thus appears to be a good compromise between the ability to discriminate impactful ROS events (one ROS event out of two is impactful) and the sample size of heavy-rainfall ROS events, which must remain large enough to analyze their evolution in the remainder of the article. It is also an easily interpretable criterion. The revised manuscript will include a justification based on these elements.

Results

Major

- L330 & Figure 8: The results do appear very clear on the boxplots, but it would be a good idea to also perform a statistical test here, to test for significant differences in the means/variability between all and impactful ROS, for rainfall amount and event duration respectively.

As the data cannot be modelled by a normal distribution, we applied a permutation test, which makes no assumption on the underlying distribution. Let K be the total number of ROS events, among which 174 are impactful ROS events. The test is performed as follows: (i) 174 events are randomly drawn without replacement from the K ROS events; (ii) the mean of the variable of interest (total rainfall or duration) is computed for this random sample. Steps (i) and (ii) are repeated $N = 10\,000$ times, yielding N means that form the distribution of the mean under the null hypothesis that impactful ROS events do not differ from the other ROS events. The p-value is then the proportion of these N means that are greater than or equal to the mean observed for the 174 actual impactful ROS events. We obtain:

- Total rainfall per ROS event:
 - Mean for all ROS events: 18.7 mm
 - Mean for impactful ROS events: 63.0 mm
 - p-value $< 1e-4$
- Duration of ROS events:
 - Mean for all ROS events: 17.0 h
 - Mean for impactful ROS events: 38.9 h
 - p-value $< 1e-4$

These statistical tests will be summarized in the corresponding paragraph (L323-331).

- L368: The authors need to rephrase the “downward trend” part. As far as I understand it, Fig. 12 doesn’t report results from a trend analysis (e.g., Mann-Kendall or Theil-Sen), which should be reported as a function of time (x decline/increase per year/decade) but a comparison between two periods (reported in absolute numbers). Please correct this everywhere where it’s needed, across both text and figure captions (12,13,14,15).

Thank you for pointing out this ambiguity in terminology. This is indeed a difference between two periods. “Trend in ROS events” will be rephrased as “change in ROS events”.

- I think there is a small (& hopefully easy to fix) discrepancy in the hatching (statistical significance) of Figures 13 and 15. Take for example snow cover at the lowest elevation band, in August or July. Is there (or was there) snow cover at that time of year at those altitudes? If not, it might be worth making sure that your code correctly filters data for your prescribed SD & SWE thresholds and excludes them from statistical analysis. These can then appear as ‘greyed out’ areas on the graphs to improve clarity. Same goes for the ROS panels. This might once again be an issue caused by the choice of statistical test, a mention of which I am yet to come across in the manuscript...

As explained in the response in the Methods section, the statistical test is a permutation test. In the absence of ROS events or snow cover, the test remains feasible, since the dataset consists of zeros for every year. This yields a null, non-significant difference. We propose to grey out the areas of the graph where no occurrence is observed over the entire study period. Fig. II below shows the result obtained for Fig. 13a. The same will be applied to Figs. 5, 10b and 15b if this is clearer.

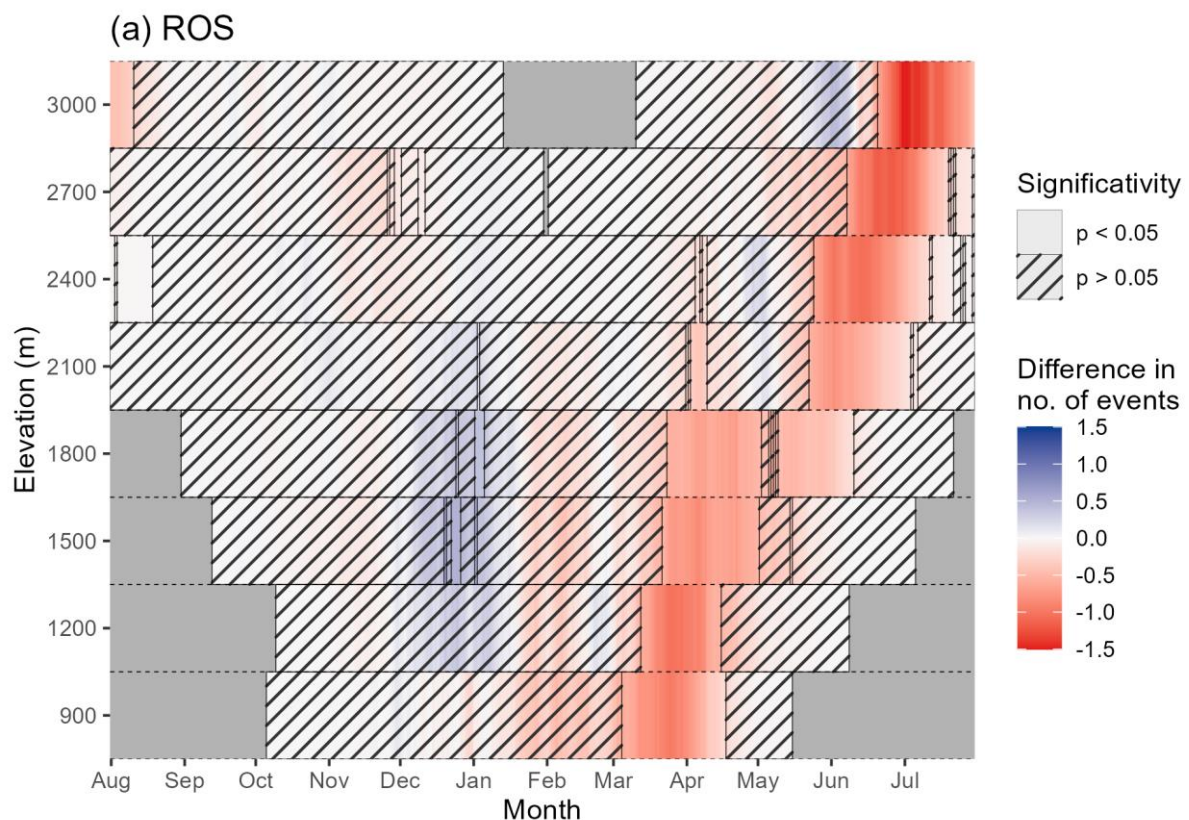


Figure II : Trends in numbers of (a) ROS events, per 30-day window and elevation band. Differences between the 1959–1988 and 1995–2024 periods.

Minor

- L242-250: This section describes the comparison between S2M and weather stations, and should therefore be moved to the Methods section as no results are yet reported here. However, given that the dataset is not independent, it is unclear to me, what the purpose of this comparison is, and if the sensitivity analysis is indeed necessary. After all, wouldn't the temperature and precipitation fields have been modelled across the elevation gradient originally for S2M, using these very observations? How is it expected that the point-scale observation (which would be at a specific elevation and aspect) would be representative for all elevations and aspects of that massif?

See the response in the Data section. Elevation is indeed accounted for in the comparison, as stations are associated with the corresponding elevation bands (L245). Note that in S2M, the precipitation observations are only used to adjust the precipitation analysis at the median elevation of the massif but the vertical gradient is prescribed at a constant ratio of +15% / 1000 m.

- L290: Here it would be very interesting to know the % coverage of those elevation bands.

The two highest elevation bands (2700 m and 3000 m) represent 10.5% of the total area of the study area. This information will be added in the revised manuscript.

- L309-314 & Figure 6: Here the authors present very interesting results on the overall links between ROS events and impact on the ground. I am wondering if the authors had the opportunity to also investigate results separately per category (torrential flood, flood, landslide), to see whether they all follow the same temporal distribution. I really think it would be a very valuable addition to include in the manuscript (including updating figure 6 accordingly).

This is an interesting point. We investigated this, and there are differences in temporal distribution among categories. The results for all RTM T/I/G events (torrential floods, floods, landslides) (Fig. 6) are close to those obtained for torrential floods (T) only (Fig. III below), as these are the most numerous events (699 RTM T events caused by ROS, i.e., 68% of the 1031 RTM T/I/G events caused by ROS). The number of recorded floods (I) is fairly low (53 RTM I events caused by ROS, i.e., 5% of RTM T/I/G events caused by ROS), so the temporal distribution and the proportions obtained (Fig. IV) should be interpreted with caution. For landslides (G) caused by ROS, there is a very strong concentration in December and January (Fig. V-b). They differ notably from torrential floods by a lower number of events in May and, more generally, by a lower proportion of events caused by

ROS (Fig. V-a, note the difference in scales with Fig. III-a). These results will be included as an appendix in the revised manuscript.

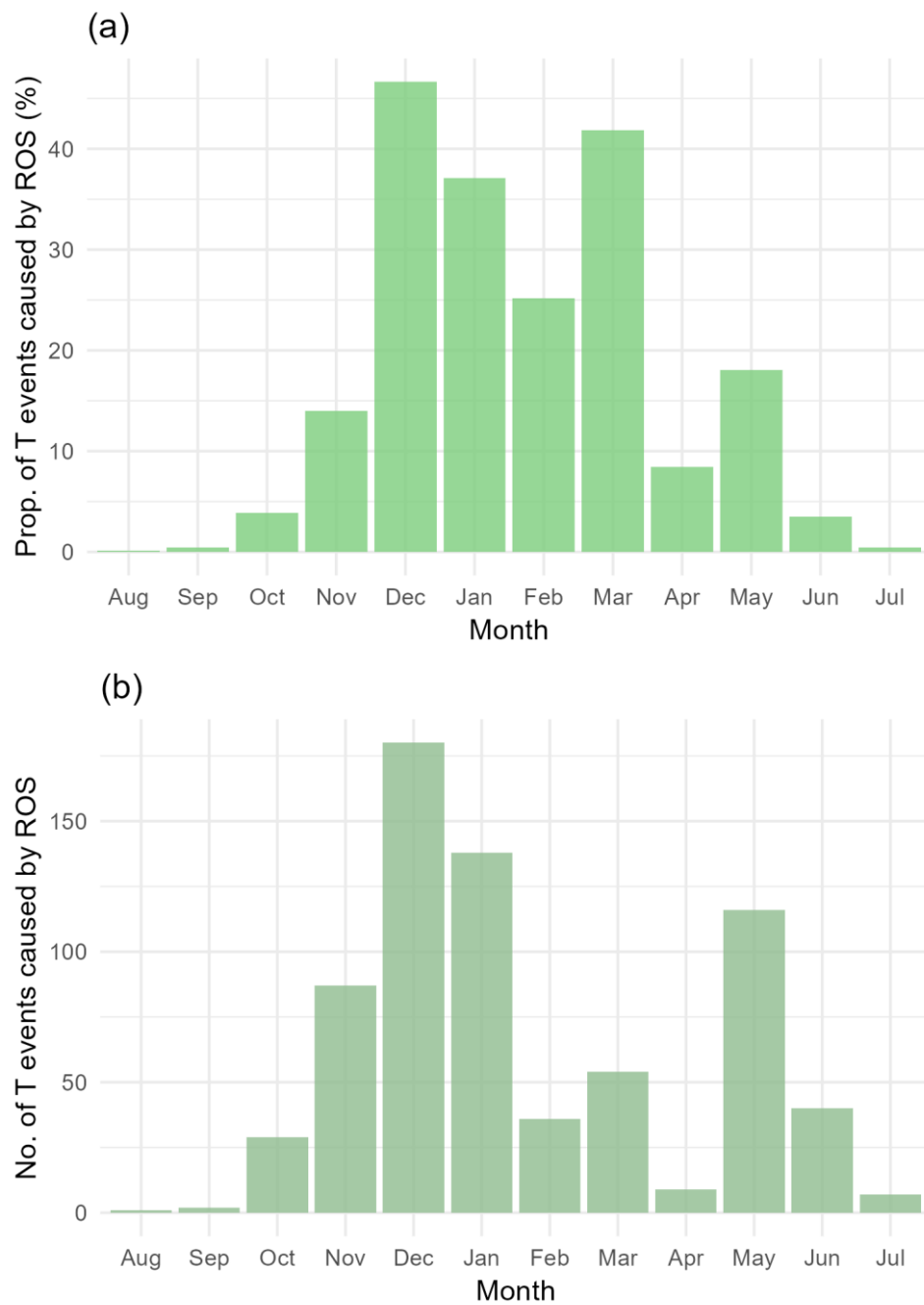


Figure III : (a) Proportion and (b) number of RTM T events (torrential floods) caused by ROS over the 1958-2024 period.

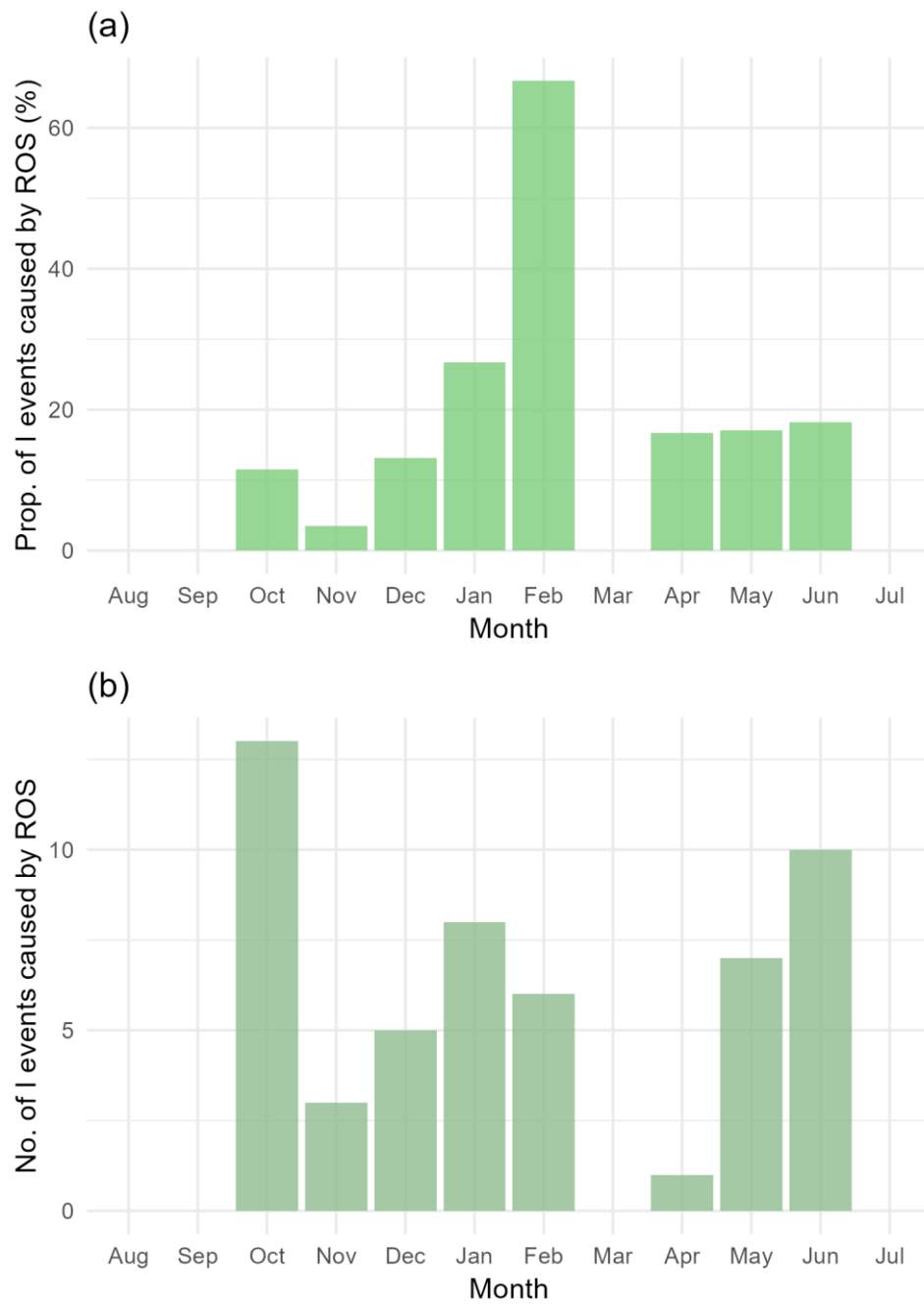


Figure IV : (a) Proportion and (b) number of RTM I events (floods) caused by ROS over the 1958-2024 period.

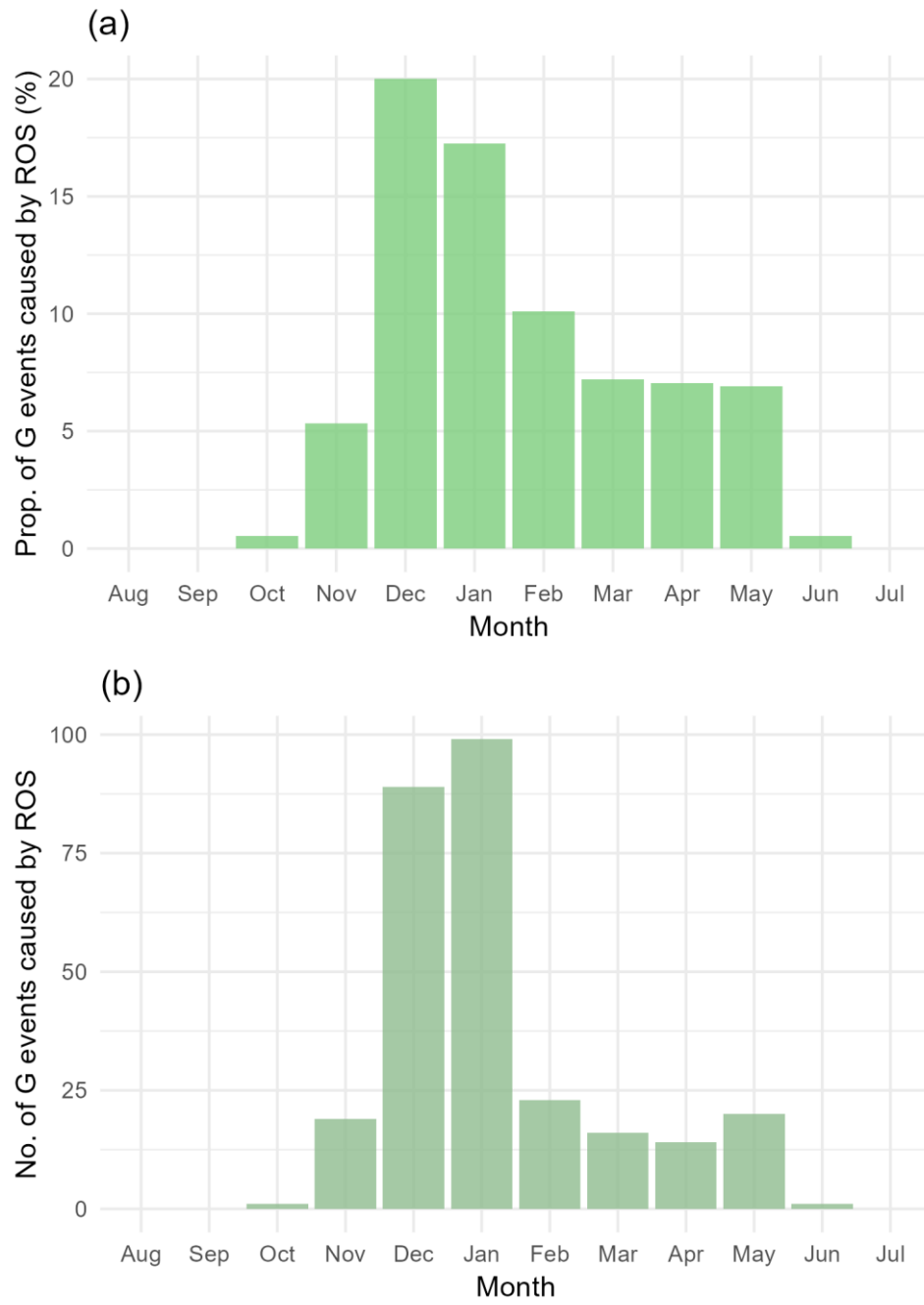


Figure V: (a) Proportion and (b) number of RTM G events (landslides) caused by ROS over the 1958-2024 period.

- L320: I am wondering why Maurienne massif has the highest rate of impact from ROS, considering it doesn't have the highest rate of ROS events. I would be interested to see the authors' thoughts on this disconnect.

This point is addressed in the Discussion, L483–485. The Maurienne massif is characterized by steep slopes and highly degraded terrains that favor debris flows (Martin-Cocher, 1984). As a result, lower total rainfall can be sufficient to trigger hazardous processes in this massif, which explains why its proportion of impactful ROS events is higher than expected from its ROS event frequency alone. We also note that this spatial variability may partly reflect biases in the RTM database, as the Maurienne and Oisans massifs are characterized by high population density and significant economic assets, particularly winter sports resorts (L486–491).

Appendices

Minor

- Figure C1: It would be useful to have maps of means for other variables too as a reference point (e.g., snow cover duration etc.)

These maps are already shown in Fig. 4. Only the map of mean duration of ROS events is not provided, but this variable shows no spatial pattern (L301–302). This map (Fig. VI below) will be added in the appendix. It should be noted that the color scale covers a narrow range: the minimum mean duration per ROS event is 14.4 h in Belledonne and the maximum is 16.6 h in Beaufortain.

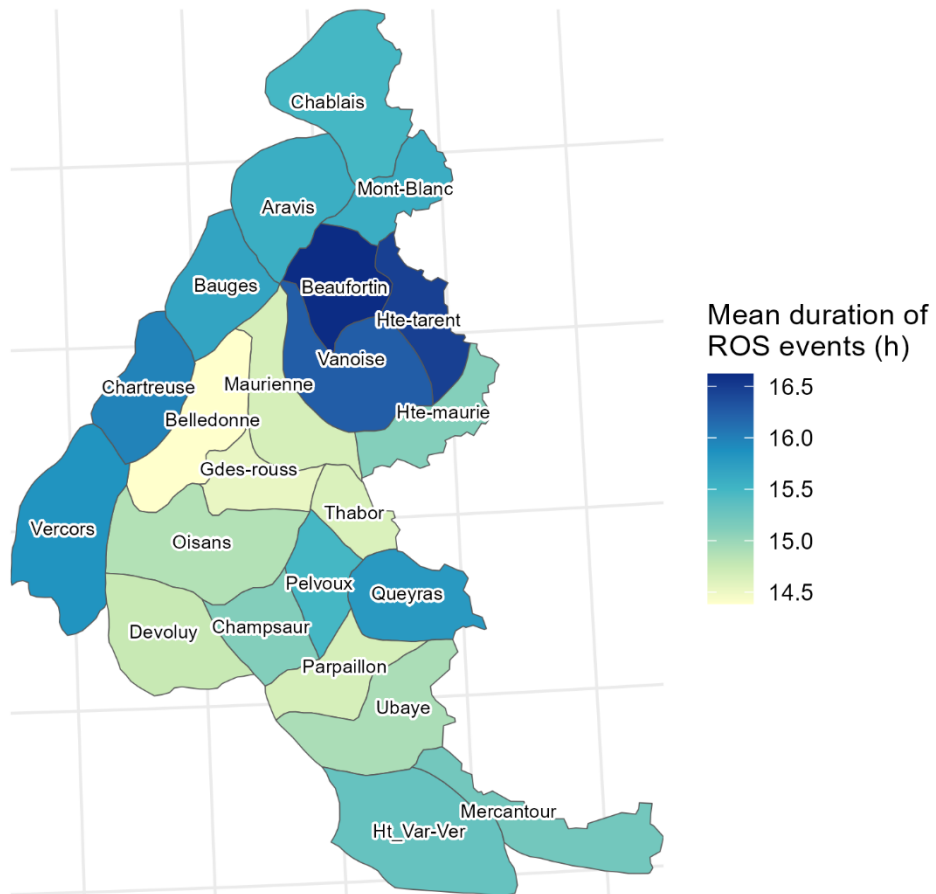


Figure VI : Mean duration of ROS events averaged over the 1958-2024 period.