Reply to RC1

"Quantifying Retrogressive Thaw Slump Mass Wasting and Carbon Mobilisation on the Qinghai-Tibet Plateau Using Multi-Modal Remote Sensing" by Maier et al. is the first study to quantify the impact of RTS on erosion, soil carbon mobilization, and ground ice loss across the entire QTP. Overall, the methods and analyses are well designed, and the results represent a meaningful contribution to our current understanding of RTS. However, significant improvement in the clarity of the figures/maps and language is needed before I can recommend publication.

We thank the reviewer for their thorough review and many helpful suggestions to improve the quality of our manuscript. We have re-worked the entire manuscript based on the reviewer's comments, adapted it accordingly, and replied in the following on a comment-per-comment basis. We hope that the answers to the original comments and the updated manuscript have improved our work to be considered for publication.

Major comments:

1. In general, I found the organization of the manuscript to need some adjustment, as there was quite a bit of material that seemed to be in the wrong section. For example, the final paragraph of the introduction read more like methods, while the second paragraph of section 2.1 read more like the introduction with numerous citations of other studies. Additionally, paragraphs should be arranged by topic with strong topic sentences. Currently, a lot of paragraphs feel like they are an amalgamation of various topics, many of which are addressed multiple times in multiple paragraphs.

We agree that the manuscript could use a certain level of reorganisation. We followed the specific comments in the following sections of the review and answered the reviewer's comments where fitting. In general, we have followed the reviewer's suggestions and re-organised parts of the introduction, methodology, results, and discussion. We reworked the topic sentences to be clearer and strongly highlight the subsequent sections. Please see the answers for the specific comments below for examples (answer to comment for L77-94).

2. All of the figures could use some refinement to improve clarity and visibility, both in the figures themselves and in the associated captions. Details are provided below.

We followed the reviewer's suggestions for the specific comments about the improvement of the manuscript's visualisation. Please find the answers at the respective specific comment below and the updated figures in the revised manuscript.

3. I found the distinction between the methods used with the delineations from Xia et al. 2024 and the manual delineations from the DEM difference map difficult to understand quickly. Although (I think) I eventually figured it out after bouncing back and forth between methods, figures, and results text, this point is central to the manuscript and should be clarified.

We thank the reviewer for pointing out that this part of the method description lacks clarity. We tried to simplify our terminology and reworked the respective section (Section 2.2.2 and 2.2.3) in the manuscript:

L171-178: "Xia et al. (2024) created a high-quality regional inventory of annual RTS delineations between 2016 and 2022 based on a semi-automated deep learning approach with optical high-resolution PlanetScope imagery (Xia et al., 2022) (Fig. 2 b). Spectral information in optical images distinguishes undisturbed from disturbed terrain, using differences in vegetation cover. RTS delineations derived from optical imagery often encompass a broader area than the active ablation zone, including zones of recent activity and depositional sections of the slump floor (i.e., features not directly involved in ongoing material loss), while excluding stable zones of past disturbance masked by lush vegetation growth. Therefore, we calculated the volume of eroded material δV for each RTS by summing only the negative elevation changes $\delta h_{< 0\,m}$ multiplied by the area of the DEM pixel (Fig. 2 c)."

L192-196: "In particular, δA stands for the RTS area that undergoes a negative elevation change or ablation and, therefore, is actively eroding within the monitoring period of T1 - T2. To test the influence of different definitions of the RTS area between delineations on optical images and on elevation change maps, we also performed area-volume scaling with the entire RTS delineation area A_{Xia}, which does not only include zones of ablation δA but potentially also bare soils disturbed by mud flows and the deposition of thawed material in the slump floor (Fig. 2 b)."

L239-244: "1) We manually delineated the RTS ablation area in the elevation change maps, where a distinct pattern of negative elevation change is visible to the human eye. At all validation sites, the elevation change maps consisted of TanDEM-X DEMs from 2011 and 2019. We statistically compared the delineations of the RTS inventory (Xia et al., 2024) from the same year (2019), the year before (2018) and the year after (2020) to the manually delineated ablation zones to investigate the agreement between the datasets in terms of RTS quantities, (ablation) area, and material erosion volume (Fig. 2 d)."

4. It seems like an ALT of 1 m was assumed in the ground ice loss calculations, despite the fact that the stated average ALT of the QTP is over 2 m. It was unclear to me whether there may be a reason for this discrepancy or if the analysis should be re-run with a different assumed ALT depth.

We have not been clear enough in manuscript regarding the use of the ALT dataset from Ran et al. (2022). Generally, we dynamically sampled and rounded (to the next integer) the ALT thickness from the gridded dataset from Ran et al. (2022) for each RTS. Out of 3,613 RTS, the majority, 2711 RTS, had an ALT = 2 m, 862 of 3 m, 34 of 1 m and 6 of 4 m. We assume that the value of 1 m ALT comes from the schematic in Fig. 2 and the related description of the calculation of SOC mobilisation from ground ice content and ALT in the different depth layers. The ground ice content is sampled from the gridded dataset of Zou et al. (2024) which is provided in three depth layers 2 to 3 m, 3 to 5 m, and 5 to 10 m as the authors of this study assume a constant AL without massive ground ice with an *average depth of 2 m in the entire QTP. We, however, decided to dynamically sample the ALT per RTS and round the sampled value to the nearest integer. Since we calculate the SOC mobilisation in 1-m depth steps (0 - 10 m), we scale with ground ice content accordingly. We assume that there is no ground ice present in the AL and that there is no SOC present in the massive ground ice. We adapted the manuscript to accommodate for improved clarity*

L216-219: "To illustrate, if we sample at an exemplary RTS location an ALT = 1.8 m and $GI_{2–3\ m}$ *= 31 % we round the ALT to 2 m. For 0 - 2 m depth, we compute the SOC mobilisation without scaling for ground ice since we assume no presence of massive ice in the active layer. For the depth layer 2 - 3 m we reduce the SOC mobilisation by 31 %."*

Section 2.1, paragraph 2: This reads like the introduction section. I would suggest combining and streamlining this information with the information already included about the QTP in the introduction.

We followed the reviewer's suggestion and reworked the introduction:

L38-49: "The largest high-altitude permafrost zone is the Qinghai-Tibet Plateau (QTP) with a total extent of 1.06 × 10^6 km^2 *at mean elevations greater than 4000 m (Wang and French, 1994; Liu and Chen, 2000; Zou et al., 2017). Similarly to high-latitude permafrost regions, the QTP is one of the most climate-sensitive regions on Earth (Liu and Chen, 2000) and has experienced a pronounced warming trend in recent decades, with an average increase in air temperatures of 0.035° C* a^{-1} *(Yao et al., 2019). The warming trend affects the thermal state of the permafrost: both ALT and ground temperatures have increased, as the regional permafrost extent declines (Cheng and Wu, 2007; Wu and Zhang, 2008; Zhao et al., 2021; Ran et al., 2022). Hence, the QTP is susceptible to permafrost thaw processes that have a substantial impact on the environment and communities, including threats to local transport and energy infrastructure, ecosystems and hydrology, as well as regional carbon budgets and water storage capacities (Luo et al., 2019; Li et al., 2022; Mu et al., 2017; Zhao et al., 2020; Kokelj and Jorgenson, 2013; Yi et al., 2025; Chen et al., 2024b). Permafrost thaw, together with the large soil organic carbon (SOC) stocks, makes the QTP a potentially considerable carbon source and an important region to monitor permafrost thaw processes (Ran et al., 2022; Chen et al., 2024a; Yi et al., 2025)."*

L59-64: "...In recent years, the plateau experienced strong expansion and initiation rates of RTS mainly on gentle north-facing slopes with fine-grained soils and high ground ice content. More than 30 % of all RTS activity is observed in a part of the central QTP, namely the Beiluhe River Basin where most activity started after 2010 (Luo et al., 2019, 2022; Huang et al., 2020; Xia et al., 2022, 2024). RTS retreat rates are relatively high, with mean rates up to 25 m a^{-1} *(2017 - 2019) (Huang et al., 2021), though like other highly active RTS sites in Alaska, northwest Canada, the Canadian High Arctic, and Siberia (Hall et al., in review)."*

We also reworked Section 2.1:

L110-121: "Our study region, the QTP, is located between 26° N and 38° N in the south-west of China at average elevations higher than 4000 m above sea level (Fig. 1 a). Permafrost covers 40 % of the plateau (Wang and French, 1994; Liu and Chen, 2000; Zou et al., 2017). The permafrost ground ice content averages around 30 %, decreasing spatially from north to south and west to east (Zou et al., 2024) (Fig. 1 b). Compared to the Arctic, the thickness of the active layer (ALT) is high (ALT = 2.34 m) (Ran et al., 2022), while permafrost thickness is relatively low (<60 - 350 m) (Zhao et al., 2020). A dry and cold climate in the northwest transitions to a warmer and wetter climate in the southeast of the plateau (Chen et al., 2015). The QTP permafrost also stores large amounts of soil organic carbon (SOC) with a median estimate of 1.41×10^{13} kg C (or 14.1 Pg C) for the top 3 m and 4.92×10^{13} kg C (or 49.2 Pg C) for the upper 25 m of soils (Wang et al., 2020; Chen et al., 2024a). The SOC stocks increase from west to east and from north to south (Wang et al., 2021; Chen et al., 2024a) (Fig. 1 c). Based on the spatial clustering of RTS identified by Xia et al. (2024), we divided the study area into five subregions, which include the West, West-Central, Central, East, and Northeast (Fig. 1a) to analyse spatial patterns of material erosion and mobilisation of SOC induced by RTS activity on the QTP. To validate the estimated volumes of RTS …"

Section 2.2: There is already an overview of the methods at the end of the introduction which reads better than this section, so I would highly recommend replacing everything before section 2.2.1 with that section. I found this section too vague as it left me with more questions than answers.

We followed the reviewer's suggestion and reorganised the introduction and Section 2.2. For details, please see the answer to comment for L77-94.

Section 3 (before 3.1): I don't think the commentary on how the results are structured is necessary, as it is only two sections and it follows the structure of the methods.

We followed the reviewer's suggestion and removed this introductory section.

Section 4.1: The organization of the last two paragraphs in this section could use some work. It is unclear to me which one topic is being discussed in each. Instead it feels like a lot of different topics are mentioned briefly and the comparison between the Arctic and QTP shows up in both paragraphs. Please reorganize and ensure that the topic sentences adequately describe the overarching topic of each paragraph.

We agree with the reviewer's comment and have rethought the organisation of the discussion section. We moved the part of Section 4.1 to the introduction where we introduce the results of Arctic and QTP studies on area-volume scaling. We changed the structure of the discussion: Section 4.1 "RTS activity, material erosion, and area-volume scaling across permafrost regions" and 4.2 "Setting into context: Magnitude of RTS SOC mobilisation and ground ice loss". In general, we simplified and shortened both sections.

Figures/Tables:

Fig. 1.a Although the reader should be able to infer that any pixel with RTS has permafrost, this is not possible to see in this map. Using a hexagonal grid with both color and size for RTS density would allow readers to see both the ground type and RTS density layers across the entire map. For example: https://www.esri.com/arcgis-blog/products/arcgis-pro/mapping/how-to-turn-a-ton-of-overlapping-data-into-a-hexagon-map. I think this would be a great use of this technique and would really make this map pop! Validation sites show up in the legend but not on the map. Please make sure they are visible. I would recommend including this information here and removing it from Fig. 5, which is really busy. The West-Central and Central clusters look like one cluster. Are they really different? More information needed in the methods.

We thank the reviewer for the thoughtful comment and the suggestions. We have missed the point vectors for the validation sites and have added them to Fig. 1 (and removed them in Fig. 5). We are sure that 3D hexagonal grids can be one option to improve this figure, however, due to limitations in GIS software access, we propose a different approach to improve the information content and visualisation of Fig. 1. We aim to give a visual overview of the study site and the spatial patterns of the identified RTS from the published dataset (Xia et al., 2024) that we combine with the generated TanDEM-X DEMs to estimate RTS material erosion volumes and SOC mobilisation. All RTS are within the permafrost zone of the QTP and we added this layer only to give readers unfamiliar with the study region a general overview of the spatial structure of permafrost on the QTP. We changed the color map and simplified the permafrost layer to ensure better visibility even below the RTS density grids (Permafrost / No permafrost). The grid used in this figure (as well as all remaining figures with gridded maps) has been adopted from Xia et al. (2024) where the original data comes from. Since in any case, grid or density visualisations only give a visual indication and aggregation of the spatial distribution of the variable in question, we decided for visualization purposes to use a grid scale of 50 x 50 km. We follow the visualisation decision from Xia et al. (2024), however, this value is arbitrary. We adapted the caption of Fig. 1 to ensure clear reasoning for the visualisation choices. West-Central and West are two cluster regions that we have adapted from Xia et al. (2024) where 22 individual RTS clusters were combined to cluster regions. We used the high-level cluster regions to ensure large enough sample sizes for the area-volume scaling. We adopted the figure (dotted circles to indicate rough location of clusters). We hope that this version of Fig. 1 improves the information transfer to the reader and addresses the reviewer's concerns.

1.b. why 2-3 m only? Is this to approximate layers below the active layer?

We only show the most upper layer available in the dataset of Zou et al. (2024) which is 2 - 3 m. In the study, we use all data available until 10 m depth (3 - 5 and 5 - 10 m). In this figure, we aim to give the reader an impression of the spatial pattern of ground ice distribution across the QTP. The general pattern transfers well across the different depth layers, so due to space limitations we only show ground ice content between 2 - 3 m, as *well as SOC content (based on Wang et al., 2021) for the uppermost three meters (0 - 3 m) of the datasets. We adapted the figure caption to address this thought better: "... (b) The ground ice content on the QTP for an exemplary depth layer (2 - 3 m) from Zou et al. (2024) is highest in the central part of the plateau. …"*

Fig. 2. "delta A" and "delta V" in captions cannot stand alone. Please spell out the variables rather than using their acronyms.

We changed the caption of Fig. 2 accordingly: "... (d) Validation at selected sites by (1) evaluating how accurately we can estimate the active erosion area δ A and the material erosion volume δ V based on the optical labels and …"

2a. Not being an expert in radar remote sensing, it is unclear what "geocoding" is and why it shows up twice here.

We thank the reviewer for highlighting this aspect as we do not use clear and consistent terminology for readers outside the SAR community. We have therefore adapted the manuscript including the schematics in Fig. 2 by removing the first geocoding box, as it misleads the reader, and rephrasing the second geocoding to "Height geocoding". In the standard bistatic InSAR processing we (1) generate a look-up table that enables us to transform between SAR slant-range and map coordinates of the reference DEM. We create an interferogram from the TanDEM-X observations as well as a simulated interferogram from the reference DEM (that we geocoded to SAR coordinates for this purpose). To obtain the residual phase we subtract the simulated interferogram from the observed interferogram. To reduce errors, we eliminated unreliable phase data by masking low coherence areas and areas affected by radar layover and shadow effects. After phase unwrapping with the minimum cost flow algorithm, the residual phase only contains the elevation difference (and error contributions) between the TanDEM-X observation and the reference DEM. We removed potential phase trends and converted the phase map into a height difference map to finally update the reference DEM by this elevation difference measurement. The newly generated DEM and associated products are finally geocoded to map coordinates using the look-up table. We follow a standard bistatic interferometric processing pipeline which is explained in more detail in earlier publications by Bernhard et al. (2020) and Maier et al. (2025). We have added a sentence to clearly reference these articles for further information:

L 149-154: "Using the global 12 m spatial resolution TanDEM-X DEM as a reference, pairs of bistatic SAR observations were processed with the GAMMA Remote Sensing software (Werner et al., 2000) following a standard InSAR processing workflow to generate a time series of DEM products (Fig. 2 a). Key steps include the generation of a differential interferogram, phase unwrapping, phase-to-height conversion, update of the reference DEM with the computed height difference, and the orthorectification or geocoding of the resulting DEM to map coordinates. Further information about the DEM generation pipeline can be found in Bernhard et al. (2020) and Maier et al. (2025)."

Fig. 3. I'd also recommend the variable size and color hexagonal grid here. Panels b. and c. are quite small.

For the first part of the comment, we refer to our answer to the reviewer's comment regarding Fig. 1. We have re-sized and simplified the figure to ensure better readability.

Fig. 4.

- "First row displays a box plot of the quantity's distribution per RTS, the second row the total quantity per subregion, and the third row additional data." This is a bit unclear. I would try something like, "The first row displays a box plot for values associated with individual RTS, the second row displays the total quantity across subregions, and the third row displays additional data that vary between columns."

We changed the caption of Fig.4 accordingly: "RTS mass wasting quantities between 2011 and 2020 on the QTP subregions: In the first column we report material erosion volumes, in the second ground ice loss, and in the third SOC mobilisation while the first row displays box plots for values associated with individual RTS, the second row bar charts for the total quantities across the subregions, and the third row additional data that vary between the columns. For all plots, the computed uncertainty is reported with error bars."

- I assume the color fill is based on region, not some underlying numerical value associated with those regions? If so, I would suggest retaining the different colors across columns to reflect the variable being displayed, but removing the different shades associated with the regions. Using a similar shading scheme to Fig. 3 made me think at first that it might reflect different average values across regions.

The base color is linked to the quantity according to Fig. 3 (red = material erosion, blue = ground ice, green = SOC) while the shades (from light to dark) are linked to the regions. We can see that this color scheme can be misleading for some readers. Therefore, we removed the base colors and assigned one grey scale value per subregion for all rows of the plot.

- I assume the dashed lines reflect mean (or median?) values across all the regions? Please add this information to the caption.

The dashed lines represent the mean value for the entire QTP for GI (b) and SOC (c), and the scaling coefficient for all RTS in QTP (a). We added a description in a separate legend.

Fig. 5. I would recommend removing panel a from this figure by moving the table to its own table and only including the validation sites in Fig. 1. It might also make sense to change the layout to allow panel b to be larger. I imagine panel b on the left with a row of panels c-f on the right would improve the visibility considerably. Captions c-e provide commentary on the meaning of the data without clearly describing what the figures show.

We followed the reviewer's suggestion and moved panel a from Fig. 5 to a separate table in the supplement (Table S3a) and added the validation sites to Fig. 1. We described the results present in Table S3 a now in the manuscript:

L298-305: "For the validation sites, we compared the delineations (2018 - 2020) of the RTS inventory (Xia et al., 2024) to the RTS ablation zones we manually identified on the *elevation change maps (2011 - 2019) based on TanDEM-X DEMs. Out of a total of 445 RTS in the RTS inventory, we identified 290 also in the lower-resolution elevation change maps, which accounts for an F1 score of 0.63. 17 RTS were missed in the RTS inventory, while 155 RTS present in the RTS inventory were not distinguishable from background noise in the elevation change maps. Most of the validation sites showed a good agreement between the two datasets. However, in Western Kunlun (A in Fig. 1 a) only 55 % of the RTS in the RTS inventory were detectable in the elevation change maps. More details on individual results at the test sites can be found in the Supplement S3 (Table S3 a and b)."*

We also increased the size of plot b. For subfigures c to e, we included more information in the figures and adapted the caption to better describe the content of the visualisations. Please see specific answers below. We changed the caption:

"Assessing compatibility and accuracy of multimodal remote sensing for RTS monitoring on the QTP aggregated for all validation sites: (a) RTS in the Beiluhe River Basin in an optical high-resolution image from 2018 (ESRI) and on a TanDEM-X elevation change map (2011 - 2019) with delineations (2018 - 2020) from the RTS inventory (Xia et al., 2024) (solid lines). The manually delineated active erosion area (only negative elevation change) is visualised by a dashed line. The RTS has grown over the course of the three years and its headwall extended upslope. The delineations from the RTS inventory based on optical images and disturbances of the vegetation cover include not only ablation zones but also of material accumulation further downslope (positive elevation change). (b) Sum of ablation and accumulation area based on the elevation loss / gain pixels for the delineations of the RTS inventory. Optical RTS delineations tend to cover a larger area than actual ablation area distinguishable on the DEM. (c) Distribution of the RTS ablation area. Due to the higher resolution of the optical images compared to the TanDEM-X DEMs, smaller RTS can be distinguished from the image background. Only small differences can be observed between the years. (d) Sum of material erosion volume based on the delineations of the RTS inventory and the negative elevation change. The volume computed from the 2019 delineation was closest to the actual erosion volume while 2018 under- and 2020 overestimates the actual erosion volume. (e) For six RTS in the Beiluhe River Basin site (D in Fig. 1a), we compared the maximum elevation loss δh_{max} of the TanDEM-X elevation change within the optical RTS delineation of 2020 to the average headwall height derived from drone-based single-time-step VHR DEMs (summer 2020). We assume that the maximum elevation loss can be used as an approximation of the headwall height of an RTS."

5.b. 2018-2020 delineations are not easy to see at this size. Different colors or linetypes may be necessary, although increasing the size may be sufficient. Personally, I would also flip the order of the years in the legend so that they are ascending. Red delineations on top of green imagery is not great for people with color blindness. I have found yellow to be quite visible on top of imagery, although I'm not sure it would work on the DEM layer.

We followed the reviewer's suggestion and changed the line style and color to improve visibility (green, yellow and purple have a better contrast to the background). We also flipped the order of the years in the legend.

5.c. I cannot tell what the difference is between the bars and the red-dotted line from the caption, including why there is a single dotted line across all years, but three different bars across the years.

We acknowledge the confusion of the reviewer and thank them for raising this issue. The dotted lines were the quantities based on the manually delineated RTS ablation zones (which we consider the "true" value to compare the optical inventory to). The dashed lines (see Fig. 5d) represent the minimum and maximum of the distribution of the quantity based on the manually delineated RTS ablation zone. Following the reviewer's comment regarding Fig. 4, we removed the shades for the different years to reduce confusion of the color choice and included the statistics based on the manual delineations as a fourth bar instead of the dashed and dotted lines to avoid confusion. We included a legend with improved information content.

5.d. It is unclear what the mean, minimum and maximum values represent.

Please see comment(s) above.

5.e. Same as c and d

Please see comment(s) above.

5.f. "We compare average headwall height of 6 RTS in the Beiluhe River Basin (Central QTP) from VHR DEM (2020) to the maximum elevation loss δh_{max} of the TanDEM-X DEM and the 2019 optical RTS delineation." The way the caption is phrased it makes it sound like the VHR DEM values are being compared to both the TanDEM-X DEM and the optical delineations rather than being compared to a value derived from both of those products. This could be fixed with "We compare average headwall height of 6 RTS in the Beiluhe River Basin (Central QTP) derived from VHR DEM (2020) to those derived from the maximum elevation loss δh_{max} of the TanDEM-X DEM and the 2019 optical RTS delineation."

We thank the reviewer for their observation and follow their suggestion: "For six RTS in the Beiluhe River Basin (D in Fig. 1a), we compared the maximum elevation loss δh_{max} of TanDEM-X DEMs within the optical RTS delineation of 2020 to the average headwall height derived from drone-based 2020 VHR DEMs. We assume that the maximum elevation loss can be used as a rough estimate of the headwall height of an RTS."

Fig. 6.a. Try a less transparent background on the legend to make it more readable. Make sure the lat and lon graticules are drawn underneath the points to improve visibility of the points. Solid points may also be a better choice for visibility. What is the background raster layer and is it necessary? It also contributes to the difficulty in seeing the points. What do the letters represent? Are they different study sites within the listed publications?

We moved the legend to the bottom of the figure, the points are now filled with a solid color, and the lat / lon graticules are removed. The background layer consisted of the *permafrost zones (Obu et al., 2019) but misses its reference. However, we agree with the reviewer that visibility can be improved by removing this information layer since it does not add any additional value here. The numbers (former letters) represent the study sites that are part of the studies we compared our results with. We clarified this in the caption as well as added it in the common figure legend. We removed the study Bernhard et al. (2022b) since we do not mention it further in the plots rather later in the manuscript. We changed the caption accordingly: "Comparison between RTS on the QTP and in Arctic permafrost regions with respect to RTS material erosion and area-volume scaling. (a) Map of Arctic sites investigated by previous studies with respect to RTS mass wasting: Kokelj et al. (2021) and Van Der Sluijs et al. (2023) estimated erosion volume and area-volume scaling based on OLS for RTS landforms based on high-resolution airborne DEMs and pre-disturbance reconstruction in northwestern Canada. Bernhard et al. (2022a) studied RTS area-volume relations based on TanDEM-X DEMs and ODR at eight sites in North America (Canada and Alaska) and Siberia between 2010 and 2016. (b) RTS density and material erosion volume per unit area for all sites reported in Bernhard et al. (2022a). RTS density and erosion volume are consistently higher on the QTP compared to Siberia and in a similar magnitude as the North American sites. (c) The area-volume scaling coefficients based on ODR reported by Bernhard et al. (2022a) are in a similar magnitude as the estimated* α_{ODR} *of QTP. The Siberian sites have generally lower* α_{OLS} *-values while Banks Island (2) and Peel Plateau (4) are closest to the QTP's* α_{ODR}*. Both Canadian sites are dominated by hillslope RTS compared to many lakeshore RTS in the Siberian sites (6 - 8). The coefficients based on OLS reported by Kokelj et al. (2021) and Van Der Sluijs et al. (2023) are distinctly higher than* α_{OLS} *on the QTP."*

Fig. 6.b. The caption provides commentary without fully describing the figure. The caption contains a colon, the meaning of which is unclear to me. The caption mentions differences based on continent, but it takes a lot of effort to compare between panels a and b to figure out which letters correspond to which continents. I would suggest adding an annotation to the x-axis or somehow grouping the bars to show the continent. If it's not too busy, it would also be useful to add an annotation to the x-axis showing which sites come from which publications.

We agree with the reviewer and changed the caption and the figure according to the reviewer's suggestion. In the new setting, we changed the map to include a legend with all locations mentioned including the numbers (change from letters), as well as to include the continent and the study in both plots b and c. We also changed the colors for the scatters to be consistent with the points of the locations on the map.

Table 1. Both SOC mobilization columns are rates, so I'm not sure the names make sense. Maybe "areal SOC mobilization rate" and "individual" or "feature-level" SOC mobilisation rate?

We agree with the need for improvement. We kept the column name "SOC mob. density" but changed the column name "SOC mob. rate" to "SOC mob. per RTS" and describing the column names further in the caption.

Table S2. What is the difference between OC and OC flux? I didn't think that any fluxes were calculated in this study, just mobilization.

As stated in common literature in the field of permafrost OC dynamics, e.g. Ramage et al. (2017), a flux is considered a yearly OC rate (normalisation by time). For the sake of simplicity, we have, however, in the manuscript avoided the term OC flux to reduce confusion. In the supplement, we have missed the occurrence. We thank the reviewer for raising this question and we adapted the table column name to OC_{yearly} and caption accordingly.

Table S3. Why isn't error reported for the change in area when it is reported for the change in volume?

The uncertainty estimates or error in the volume quantification is based on the height error estimation based on information of the satellite acquisition setting and the InSAR processing (coherence, Height of Ambiguity (HoA), number of looks during multi-looking) and can be propagated to volume change error. The error that is present in the delineations (manual or automatic, on a DEM or on optical imagery) is however much harder to quantify and not part of this study (e.g. Nitze et al. (2024)). Here, we assume that our manual delineations of the ablation zone on the DEM elevation change maps do not include any error, as well as the semi-manual RTS delineations from the dataset of Xia et al. (2024) does not include any error. However, this assumption is a simplification and certainly there is error involved in both manual delineations (e.g. subjectivity) and in automatically generated delineations by machine learning models (e.g. biases). We added some description in the methods:

L236-238: "We assume the RTS area A_{Xia} is error-free, despite biases and subjective influences in both automated and manual RTS segmentation, which are difficult to measure (Nitze et al., 2024b; Maier et al., 2025)."

Line edits:

L26: "air temperatures rise" would be better as "air temperature increases"

We changed the manuscript accordingly (L27): "The permafrost thaw is caused by a long-term increase in air temperatures…"

Paragraph 1-2 transition is abrupt. A final sentence in P1 that connects the two would be helpful.

We followed the reviewer's comment and rewrote the end of paragraph 1 (L35-37): "Existing Earth system models exhibit significant limitations in the accounting of soil organic carbon (SOC) and in the prediction of future changes for global permafrost regions (Turetsky et al., 2020; Virkkala et al., 2021; Nitzbon et al., 2020).

The largest high-altitude permafrost zone is the Qinghai-Tibet Plateau (QTP) with a total extent of 1.06 × 10^6 km^2 at mean elevations greater than 4000 m (Wang and French, 1994; Liu and Chen, 2000; Zou et al., 2017).”

L39 and elsewhere: "QTP" should always be preceded by "the"

We changed the manuscript accordingly (e.g. L40, L44, L104).

L42-44: "The QTP is susceptible to permafrost thaw processes that cast strong impacts from permafrost degradation, including threats to local transport and energy infrastructure and ecosystems, as well as to regional climate and water storage". Unclear, specifically the use of "that cast".

We changed the manuscript to accommodate for the comment (L44-47): “Hence, the QTP is susceptible to permafrost thaw processes that have a substantial impact on the environment and communities, including threats to local transport and energy infrastructure, ecosystems and hydrology, as well as regional carbon budgets and water storage capacities (Luo et al., 2019; Li et al., 2022; Mu et al., 2017; Zhao et al., 2020; Yi et al., 2025; Chen et al., 2024b).”

P2-3 transition is also abrupt.

We followed the reviewer’s comment and rewrote the start of paragraph 3 (L49-51): “Retrogressive Thaw Slumps (RTS) are permafrost landforms that occur in ice-rich permafrost terrain when ground ice is exposed, allowing for rapid thaw and downslope movement of the resulting debris (Burn and Lewkowicz, 1990; Kokelj and Jorgenson, 2013; Nesterova et al., 2024; Harris, 1988; CPA et al., 2024). The landform can expand …”

L46-50: The first two sentences of this paragraph feel pretty wordy and a bit repetitive. Maybe try something like "RTS are a form of abrupt permafrost thaw that occur when ground ice is exposed, allowing the rapid thaw and downslope movement of debris."

Please refer to the answer to the comment above.

L57: "manifesting itself with increased numbers, sizes, and faster retreat rates" would be better as "manifesting through increased numbers, sizes, and faster retreat rates"

We changed that part of the manuscript (L56-58): “Climate warming and human disturbance have intensified RTS activity not only in the Arctic (Lantz and Kokelj, 2008; Bernhard et al., 2022a; Van Der Sluijs et al., 2023; Young et al., in review), but also on the QTP.”

L58: I would break the paragraph at "However, due to their complex spatiotemporal dynamics..." as it's a very long paragraph and this seems like a natural transition.

We included a paragraph break (L64).

L59: "assessing its impact on regional carbon cycling" should be "assessing their impact on regional carbon cycling"

We changed the manuscript accordingly (L64).

L65-66: "Nevertheless, to quantify RTS-induced mass wasting and evaluate the potential implications on permafrost carbon emissions, additional datasets are still required - particularly those capturing lateral and vertical change and soil properties." This sounds to me like the data required for estimates of permafrost carbon mobilization and not carbon emissions, which would be much more difficult to estimate. As this manuscript does not focus on carbon emissions but rather carbon mobilization, I would change the wording to reflect that.

We changed the manuscript accordingly (L70-72): "Nevertheless, to quantify RTS-induced mass wasting and evaluate the potential implications on permafrost carbon mobilisation, additional datasets are required - particularly those capturing lateral and vertical change and soil properties."

L77-94: The final paragraph in the introduction is quite long and provides a more detailed explanation of the methods than is required. I'd recommend keeping only a few sentences about the methods here and replacing most of the overview text in 2.2 with this, as this section reads more smoothly clearly than what is currently in 2.2. Of course, there are a few pieces of information in 2.2 which are not included here, that should be kept.

We thank the reviewer for this suggestion. We have reworked the end of the introduction (L101-106): "In this study, we present the first regional empirical analysis on RTS mass wasting due to RTS activity on the QTP during the last decade. The high-quality RTS delineations of Xia et al. (2024), in combination with soil property datasets for the QTP, allowed us to estimate the volume of eroded material, associated ground ice loss, and mobilisation of SOC. By examining the allometric scaling of area and volume of RTS-induced material erosion on the QTP we provide a basis for (1) the (sub)regional comparison of RTS dynamics and (2) providing empirical scaling relationships to potentially constrain regional-scale estimates on material erosion and carbon mobilisation. We aim to show that the combination of multimodal and multitemporal datasets allows for a more detailed analysis of RTS mass wasting dynamics and further increases our understanding of the regional carbon budget impacts."

And we re-wrote the section text of 2.2 to accommodate the suggestion (L127-136): "We used bistatic TanDEM-X SAR observations to generate multi-temporal DEMs that span all RTS locations in the RTS inventory of Xia et al. (2024). By differencing DEMs from 2011 and 2020 and combining the resulting elevation change with the annual high-resolution RTS inventory, we estimated the volume change of the eroded material induced by RTS activity on the QTP (Fig. 2 a, b). Based on recently published datasets on the permafrost state and soil conditions including active layer thickness (ALT), volumetric water / ground ice (GI) content, and SOC stocks, we modelled the material erosion volumes into mobilised SOC fluxes for all RTS present on the QTP until 2020 (Fig. 2 c). We evaluated the uncertainty in the estimated erosion volume and the derived properties, examining *how spatial resolution affects errors in material erosion estimates (Fig. 2 d). Similarly to temperate landslides, scaling laws between the planimetric area and the erosion volume have been used to improve our understanding of the variability in geomorphology, process dynamics, and the drivers and controls of RTS."*

L100: "holding a total water volume of 3330 km3 in the top 10 m" This seems like unnecessary information to me.

We have removed this information following the reviewer's suggestion.

L116: "focussing" is typically spelled "focusing", although I just learned that both are apparently acceptable

We changed "focussing" to "focusing".

L124-125: "we established several validation sites that were spatially distributed throughout the QTP" More detail is needed about how these sites were chosen. Were they a stratified random sample or manually selected or something else?

We agree that this information is important and is missing currently in the manuscript. We manually selected the focus or validation sites following a set of criteria that we stated now in the manuscript (L120-125): "To validate the estimated volumes of RTS material erosion derived from integrated optical and elevation remote sensing data, we established the following validation sites covering each 400 km^2 (Fig. 1 a): Western Kunlun, Gaize, Southern Nima, Beiluhe River Basin, and Qilian Mountain. We selected the sites based on 1) presence of RTS activity and different cluster regions representing diverse geographic and terrain conditions based on the results of Xia et al. (2024), 2) expert knowledge on the availability of field data, and 3) the existence of TanDEM-X observations of the same year to ensure consistent data quality for validation purposes."

L136: "The German Synthetic Aperture Radar (SAR) satellite mission TanDEM-X allows us to generate temporally resolved digital elevation models based on bistatic SAR interferometry (InSAR)" I think you should be more direct. "The German Synthetic Aperture Radar (SAR) satellite mission TanDEM-X was used to generate temporally resolved digital elevation models using bistatic SAR interferometry (InSAR)."

We changed the manuscript accordingly (L 138-139): "We used satellite observations from the German Synthetic Aperture Radar (SAR) mission TanDEM-X to generate temporally resolved digital elevation models based on bistatic SAR interferometry (InSAR) (Krieger et al., 2007; Bojarski et al., 2021)."

L139-152: A lot of this reads like an introduction and is provided in more detail than is probably required. I would suggest reframing it to focus on what you did with concise explanations of why you did some things differently than previous studies. For example, "We used year round TanDEM-X observations to achieve full spatial coverage, although previous studies in the Arctic have used only winter data to avoid errors caused by dense and wet tundra vegetation during the growing season (sources). The errors introduced by vegetation characteristics are likely to be negligible in this study due to the low canopy heights in the alpine meadows, steppes, arid desert, and bare ground of the QTP." One more sentence could describe that the snow conditions required for winter data to be used were met.

We changed the manuscript accordingly (L141-147): "We used TanDEM-X observations throughout the year to achieve full spatial coverage on the QTP, although previous studies in the Arctic have used only winter data to avoid errors caused by dense and wet tundra vegetation during the growing season or melting snowpacks (Bernhard et al., 2020, 2022a; Maier et al., 2025). The potential errors introduced by vegetation characteristics and snow cover are likely to be negligible in this study due to the commonly low canopy heights of alpine meadows, arid desert, and bare ground (Wang et al., 2016; Xia et al., 2024) and shallow average snow depths well below the height sensitivity of TanDEM-X DEMs (Che et al., 2008; Yang et al., 2020)."

L167: "The resulting elevation change maps are normally distributed around zero with a SD representing the achievable vertical accuracy of the DEM pair." I think this is only true if you assume there is not widespread subsidence occurring. I'm not aware of any studies that have looked into this off the top of my head, but I wouldn't assume there is no subsidence occurring more broadly as the ALT is deepening and temperatures are rising in the region (I'm sure it would be smaller than the detection threshold for the data, in any case).

Yes, potential widespread permafrost subsidence (due to active layer deepening) would be a phenomenon potentially not visible in our DEM differencing method as we have a height sensitivity of TanDEM-X DEMs of several meters compared to gradual thaw subsidence of only several millimeters or centimeters per year (e.g., Chen et al., RSE 2022, https://doi.org/10.1016/j.rse.2021.112778). There could be widespread subsidence (or even uplift) at small magnitudes that could be physically present that would however not be visible in our data. For the QTP, we see a normal distribution of elevation change values around zero in our data with varying standard deviations depending on the terrain and the acquisition geometry for terrain not affected by abrupt thaw. We added a small note to the sentence to explain this shortly (L166-168): "For stable terrain, the resulting elevation change maps are normally distributed around zero with a SD representing the achievable vertical accuracy of the DEM pair."

L176-177: "we assigned the optical RTS shape that matched the last observation year to ensure a conservative estimation of the eroded volume". This sounds backwards to me. Using the latest observation year should ensure the largest RTS area, which would provide the largest estimate of the eroded volume, unless I'm missing something important.

We acknowledge that the phrasing is misleading. The reviewer is correct that using the latest (and mostly largest) RTS delineation will potentially include more pixels in the elevation change map than using a delineation from an earlier year which is likely smaller. Hence, it should say that we do not have a conservative approach here, rather the *opposite. However, what we tried to express here, is that in averaging over DEMs from several years we potentially conservatively estimate or underestimate the eroded volume change if the RTS has grown significantly during the time span of the averaged DEMs. We adapted the manuscript accordingly to clarify this misleading sentence and account for the reviewer's comment (L181-183): "If observations from several years contributed to the DEM of T2, we assigned the optical RTS annotation that matched the latest DEM. Hence, we capture the full planimetric extent of the RTS might, however, underestimate the actual erosion volume by averaging over different erosion states across the years."*

L198-199: "Therefore, we integrate existing datasets for QTP that define (1) ALT, (2) GI content between 2 and 10 m, and (3) SOC stocks between 0 and 3 m depth." I would suggest including the citations here, even though they are described in the following sentences.

We changed the manuscript accordingly (L197-198): "Therefore, we integrate existing datasets for the QTP that define (1) ALT (Ran et al., 2022), (2) GI content between 2 and 10 m (Zou et al., 2024), and (3) SOC stocks between 0 and 3 m depth (Wang et al., 2021)."

Eq. 3: Are the SOC (below ALT) and SOC (above ALT) switched here?

We thank the reviewer to mention this as we see that we are not fully clear here. We always describe depth as absolute value. This means that SOC($d < ALT$) means SOC above ALT value (within the AL) and SOC($d > ALT$) SOC below the active layer (depth is larger than AL depth). We changed the manuscript and the formula accordingly (L212-215): "... with the number of pixels n, the RTS ablation area δA [m], the depth of the active layer ALT (> 0 m), ground ice content GI [%] and soil depth d ($0 < d < 10$ m). If the depth d is higher than the ALT, then only the part of the eroded material that is not massive ground ice is added to the total SOC mobilisation. Similarly, we estimate the volume of RTS-induced ground ice loss across the depth layers by scaling the eroded material by 100 − GI."

L270: "corresponding to a relative error in elevation change of ~ 35%" the phrasing here confused me - I'd either remove it since the magnitude of the error is already listed (and it's pretty clear that it's about 1/3 of the total without complicated head math) or change the wording. Also, please be consistent about including or excluding a space between the number and "%".

We changed the manuscript accordingly and removed that sentence following the reviewer's comment. We also checked the manuscript for formatting issues but could not find any other occurrences.

L270-271: "The median elevation loss that roughly indicates the RTS headwall height was 1.2 ± 0.4 m." This is a fragment and I'm uncertain how the elevation loss is related to the headwall height. Is this coming from the allometric scaling? If so, what are the assumptions about RTS shape that are going into the relationship between the headwall height and the volume?

We agree with the reviewer that this result is reported as a fragment and not part of a meaningful analysis. We have therefore removed this sentence. The median or maximum elevation change within the ablation zone can be seen as proxies for the slump's concavity depth or headwall height. However, we do not think that this aspect should be included in this part of the study. We refer to Fig. 5 f and the results from the validation with the UAV DEM where we introduce the term headwall height. Since we have a mismatch in temporal data availability, i.e., no UAV DEMs from an earlier time (e.g., 2011), we cannot compute volume change of eroded material based on the high-resolution data for the RTS in question. Therefore, we want to compare a quantity that gives an idea if the TanDEM-X method gives reliable volume change estimates even though we cannot validate it directly. Here, we propose to compare the average elevation difference between the top of the headwall and the RTS floor computed from 100 randomly distributed points in an area defined by the transect in the UAV DEMs to the maximum negative elevation change in the ablation zone of the RTS retrieved from the TanDEM-X derived elevation change maps. Both values are a proxy of the concavity depth or headwall height of an RTS. However, the headwall is often heterogeneous and the slump floor rough making even in-situ measurements of headwall heights difficult to compare between slumps or between several measurements in time (Nesterova et al., 2024). Here, this comparison is even more difficult since we compare data from different spatial resolutions and different temporal viewpoints. However, since no other validation data is available, we decided to report this analysis to give the reader an (limited) impression of the method's accuracy. We have adapted Fig.2 d and the manuscript in the methodology and results:

L243-255: "(2) At the Beiluhe River Basin site, we compared the TanDEM-X-derived elevation change with photogrammetric very high-resolution (VHR) DEMs from an in-situ drone campaign covering in total six RTSs in August 2020. A DJI P4 Multispectral was used to obtain the multispectral drone images. The resulting DEMs have a spatial resolution of < 1 m and a georeferencing accuracy of 0.2 m RMSE. Since no VHR DEM was available for T1, we could not perform DEM differencing and directly validate our volume change estimates. Based on the hillshade VHR DEMs, we manually delineated the approximate location of the headwall with the help of transect profiles (Fig. 2 d). We defined small buffer zones (~ 5 m) and randomly distributed points (n = 100 per RTS) on both sides of the headwalls that represent the elevation of stable ground h_{stable} and the RTS slump floor h_{RTS}, respectively. We computed the average headwall height h_{VHR} per RTS as the median difference between h_{stable} and h_{RTS}. The monitored RTS were relatively small (< 10^4 m^2) and shallow (h_{VHR} < 4 m). Defining a headwall position and applying the same methodology with 10 m-resolution TanDEM-X DEMs is difficult. Therefore, we chose to compare the estimated headwall heights based on the VHR DEMs with the maximum negative elevation change δh_{max} that we estimated based on TanDEM-X elevation change maps, if the largest height loss aligns with the largest material ablation and is located at the headwall."

L321-325: "To further validate our results, we used very high-resolution (VHR) DEMs generated based on drone stereo photogrammetry from August 2020. Fig. 5 f shows an acceptable fit between the maximum elevation changes δh_{max} computed with TanDEM-X DEMs and the delineations of the RTS inventory and the average headwall height calculated from the VHR DEMs at six RTS locations. Details can be found in the Supplement in Table S4. However, the small sample size does not allow meaningful statistical analysis and, therefore, only allows for a qualitative comparison."

L275-276: "An α value of 1.30 indicates that RTS in QTP during the last decade followed a relationship between linear growth (α = 1.0) and exponential growth (α > 1.5)". Since the exponent does not change across values of x, this is a power relationship not an exponential relationship. The wording here also makes it sound like growth in RTS over time rather than the relationship between area and volume. Please revise the wording.

We agree with the reviewer and have adapted the wording here (L274-277): "An α value between 1.11 and 1.30 indicates that RTS on the QTP during the last decade followed a relationship between a growing scar zone with constant depth (α = 1.0) and growth with a constant width-depth ratio (α > 1.5) and falls in the range of soil landslides (1.1 - 1.4) based on the investigated scaling relations of landslides in temperate locations (Van Der Sluijs et al., 2023; Jaboyedoff et al., 2020)."

L280-281: "We calculated that ~ 64 % of the thawed ground ice was located in the first metre under the active layer (2 - 3 m)". Are you assuming an ALT of 1 m? This doesn't make sense to me, given that the average ALT of QTP cited in the methods section is 2.34 m.

We would like to refer also to the answer to the comment in the general section about active layer thickness and the use of the dataset. We think there has been a misunderstanding of the general process of sampling ALT values as well as the scaling of SOC mobilisation and ground ice loss computation due to lacking explanation in the manuscript. It is accurate to state the largest amount of lost ground ice comes from 2 - 3 m depth since most of the RTS have an ALT value of 2 m (rounded from the dataset from Ran et al., 2022) which means that d < 2 m (0 - 2 m) are massive ice free and the ground ice starts in our computations at 2 m depth. We rephrased here (L279): "We calculated that ~ 64 % of the thawed ground ice was located between 2 - 3 m depth, 32 % between 3 and 5 m…").

L287-288: "We normalised the total quantities per subregion by RTS count and analysed the distributions to ensure better comparability." This reads like methods and is unnecessary here.

We removed this statement following the reviewer's suggestion.

L291-293: "This points to a greater elastic distortion in the degree to which the concavity increases volumetrically with changing area for the mountainous northeast QTP (see Van

Der Sluijs et al. (2023) for coefficient interpretation, Fig. S1 and Fig. S3)." This should be in the discussion. The wording is unclear to me.

We moved this statement in the discussion (L369-372) and rephrased for improved readability: “However, the northeastern region has a substantially higher α coefficient, indicating larger headwalls and concavity depth per unit area growth. The mountainous northeast QTP potentially favours the development of relatively larger and deeper RTS (see Van Der Sluijs et al. (2023) for coefficient interpretation, Fig. S1 f and Fig. S6 f). ”

L295-296: "but the total amount is relatively low for the entire QTP compared to the eroded volume". It is unclear to me what point is trying to be made here, as the ice content is necessarily less than the total volume.

We agree with the reviewer that the ground ice loss is obviously lower than the erosion volume. We rephrased to account for the reviewer’s comment (L290-292): “Based on Zou et al. (2024), the GI content is highest in the central subregions (median of 32.8 % in central and 32.6 % in the west-central QTP) where we also found the highest total and average ground ice loss (Fig. 4 b). The lowest GI content was present in the northeast (median of 0 %, mean of 10.7 %).”

L298-299: "However, the distribution of SOC mobilisation follows the spatial trend present in the SOC stocks from Wang et al. (2021)". This belongs in the discussion.

We removed this statement and moved it in the discussion (L415-416): “We only quantified the magnitude of SOC mobilised by RTS activity and its spatial pattern closely following the spatial trend present in SOC stocks of Wang et al. (2021) on the QTP over the last decade. Exploring the complex pathways of mobilised SOC is beyond the scope of this study. ”

L302-304: "We manually delineated the ablation zones for 307 RTS scars on elevation change maps (2011 - 2019) at all validation sites and statistically analysed the potential differences from the RTS labels (2018, 2019, 2020) in the optical inventory of Xia et al. (2022)." This does not need to be repeated in the results section. The topic sentence here would better be an overarching statement about the agreement of the two methods.

We rewrote this part of the manuscript to account for this comment as well as to ensure that it is clear how we perform the validation based on the manually delineated ablation zones on the elevation change maps (L297-304): “For the validation sites, we compared the delineations (2018 - 2020) of the RTS inventory (Xia et al., 2024) to the RTS ablation zones we manually identified on the elevation change maps (2011 - 2019) based on TanDEM-X DEMs. Out of a total of 445 RTS in the RTS inventory, we identified 290 also in the lower-resolution elevation change maps, which accounts for an F1 score of 0.63. 17 RTS were missed in the RTS inventory, while 155 RTS present in the RTS inventory were not distinguishable from background noise in the elevation change maps. Most of the validation sites showed a good agreement between the two datasets. However, in Western Kunlun (A in Fig. 1 a) only 55 % of the RTS in the RTS inventory were detectable *in the elevation change maps. More details on individual results at the test sites can be found in the Supplement S3 (Table S3 a and b).”*

L309-311: "Figure 5 b shows an example RTS in the Beiluhe River Basin in Central QTP including the delineations of the optical inventory (Xia et al., 2022) of the summers one year before, the same year and one year after the SAR observation used for the DEM generation, as well as the manually delineated ablation zone visible on the elevation change map." This is unnecessary.

We removed this statement following the reviewer’s suggestion.

L312-316: "A RTS delineation based on multispectral imagery defines the boundary of the thaw feature by the difference in the spectral signal between disturbed and intact vegetation, whereas elevation maps showcase RTS activity by elevation loss (ablation at the headwall) and gain (accumulation at the floor). Typically, a larger planimetric area is present in RTS delineations based on optical images that often comprise most of the recently active slump floor and accumulation zone compared to the DEM-based boundaries." This is methods / discussion and does not belong here.

We removed this statement and summarised this aspect in the methods section (L171-175): “Spectral information in optical images distinguishes undisturbed from disturbed terrain, using differences in vegetation cover. RTS delineations derived from optical imagery often encompass a broader area than the active ablation zone, including zones of recent activity and depositional sections of the slump floor (i.e., features not directly involved in ongoing material loss), while excluding stable zones of past disturbance masked by lush vegetation growth.”

L316-320: "Comparing the RTS delineation area A_{Xia} with the active erosion area visible on the elevation change map, we see that the delineations for 2019 and 2020 include substantially more ablation area than what is distinguishable as RTS-induced erosion on the elevation change maps (Fig. 5 b, c)." It is unclear to me why these numbers are so different if both are derived from the DEM difference map and one is just a hand-delineation of more or less contiguous ablation area while the other includes all pixels showing ablation. I also had a really hard time keeping track of what the difference between these two methods is, so I think it is worth trying to find a concise label for the two methods that can help remind the reader.

We would like to refer to our answer on the reviewer’s comment #3 in the general section. We tried to improve the description in the method section concerning the validation effort with manually delineated ablation zones in DEMs that should help the reader to follow the presentation of the results. We have also adapted Fig. 5 according to the reviewer’s comments. We have also rephrased parts of the result section:

L297-320: “For the validation sites, we compared the delineations (2018 - 2020) of the RTS inventory (Xia et al., 2024) to the RTS ablation zones we manually identified on the *elevation change maps (2011 - 2019) based on TanDEM-X DEMs. Out of a total of 445 RTS in the RTS inventory, we identified 290 also in the lower-resolution elevation change maps, which accounts for an F1 score of 0.63. 17 RTS were missed in the RTS inventory, while 155 RTS present in the optical inventory were not distinguishable from background noise in the elevation change maps. Most of the validation sites showed a good agreement between the two datasets. However, in Western Kunlun (A in Fig. 1 a) only 55 % of the RTS in the RTS inventory were detectable in the elevation change maps. More details on individual results at the test sites can be found in the Supplement S3 (Table S3 a and b).*

Visually, the delineations from the RTS inventory fit well with the spatial patterns of the TanDEM-X elevation change map, with most of the ablation zone being part of the optical delineation. Due to the lower spatial resolution of the DEM, the differences between the optical delineations of 2018, 2019, and 2020 are rather small in this example (Fig 5 b). The total RTS area A_{Xia} consists of the material erosion zone (overlapping negative values in the elevation change map) which was growing between 2018 and 2020, and the material accumulation zone (positive elevation change) which stayed relatively constant across time. For all analysed years, the total RTS area was distinctly larger than the sum of the manually delineated erosion areas. When only considering the ablation zone within the RTS delineation, the difference was smaller (Fig. 5 b). The average RTS ablation area based on the manual delineation was 12,978 m^2, which is relatively close to the RTS inventory: 9,748 m^2 in 2018, 11,325 m^2 in 2019, and 13,512 m^2 in 2020, accounting for 62 to 64 % of the total RTS area A_{Xia}. However, in the high-resolution PlanetScope images used as the basis for the RTS inventory, substantially smaller RTS can be identified compared to the TanDEM-X DEMs. This leads to a lower limit of RTS ablation areas of approximately < 103 m^2 for TanDEM-X based RTS monitoring (Fig 5 d). However, in our analysis of the entire QTP, only 6 % of 3,613 RTS have ablation areas below this value. The total volume of material erosion computed with the RTS inventory differs only slightly from the manual delineations (Fig. 5 e). We estimated a total volume δV of 8.47, 9.98, and 10.97 × 10^6 m^3 for the RTS inventory delineations of 2018, 2019, and 2020, respectively, compared to an erosion volume of 9.94 × 10^7 m^3 defined by the manual delineations. The uncertainty of volume change is, however, slightly greater for optical RTS labels compared to the validation (Table S3 b).”

L321-322: what is the difference in observation time? I thought both methods used the same DEM difference map which was fairly well matched to the dates of the optical RTS labels?

The difference in observation time is 2018 - 2020, i.e. three years. One could have assumed that distribution of ablation area within the RTS delineations for the three consecutive years might increase more strongly from 2018 to 2020 (due to growing RTS). However, the distribution does not change strongly. We removed the statement as it does not add any value.

L324-325: "The lower threshold for distinguishing an RTS from background noise is substantially higher in the elevation change maps". Unclear language. What is the lower threshold?

We rephrased the sentence (L313-316): “However, in the high-resolution PlanetScope images used as the basis for the RTS inventory, substantially smaller RTS can be identified compared to the TanDEM-X DEMs. This leads to a lower limit of RTS ablation areas of approximately < 103 m2 for TanDEM-X based RTS monitoring (Fig 5 d). “

L331: "the optical label of 2019 shows the best fit." I don't think this provides much value, since they're not measuring the same thing.

We have adapted the manuscript to ensure clarity (L318-313): “We estimated a total volume δV of 8.47, 9.98, and 10.97 × 10^6 m^3 for the RTS inventory delineations of 2018, 2019, and 2020, respectively, compared to an erosion volume of 9.94 × 10^7 m^3 defined by the manual delineations.“

L332-334: "For all regions that did not undergo any erosion in the RTS scar or in the vicinity, we assume that the elevation change is normally distributed around zero with a standard deviation of approximately 2 - 3 m in flat terrain for TanDEM-X-derived DEMs." Couldn't it also be that the average elevation value was different than zero, but by an amount smaller than could be detected?

This is a valid question, and it is in theory possible. We would like to refer to our answer to the comment of L167. Since we added further explanation to the method section, we removed this sentence in the results section.

L334-337: "For volume estimation based on optical delineations, stable areas containing noise are likely included since most optical RTS boundaries are larger than the actual active erosion area. Although this minimally affects the total volume change estimate due to the low magnitude in negative elevation change, it adds additional random errors, thus contributing to the overall uncertainty budget." This is discussion.

We moved this statement in the discussion (L458).

L358: "Normalised by study area". It is unclear what has been normalized here.

We removed this part of the sentence since we agree that it is repetitive since we report the material erosion volume already per unit area (L338-339): “The RTS on the QTP relocated 101.8 m^3 a^{-1} km^2 material during the last decade, ...”

L362-363: "Most of these sites are mainly characterised by smaller and shallower RTS located on lake shores in relatively flat terrain". Is "these sites" referring to the remaining sites investigated by Bernhard et al. 2022? Please clarify.

We changed the manuscript accordingly (L340-342): "The remaining sites investigated by Bernhard et al. (2022a) in the Canadian (C) and Siberian (F - H) Arctic are characterised by smaller and shallower RTS located on lake shores in relatively flat terrain and showed less than half the magnitude of the volume change rates observed on the QTP."

L368: "Based on the fitting of a Ordinary Least Square model". "A" should be "an".

We changed the manuscript accordingly.

L376: "scaling coefficients below α < 1.3". This is repetitive. Either spell it out or use the symbols.

We changed the manuscript accordingly.

L385: "shallowed" should be "shallow"; "whereby" doesn't seem like quite the right word here

We changed the manuscript accordingly (L375-377): "The areas affected by RTS in this study are generally smaller and shallower than in the Arctic (Liu et al., 2024). It is therefore possible that the coarse resolution of the DEM might not correctly capture the area and volume change for these small areas skewing the scaling models."

L394-395: "Therefore, we should only see these results and comparisons as another indicator that RTS development in the QTP is on a level similar to the known hotspots in the Arctic permafrost region." I would not recommend starting a paragraph with "therefore" as it implies a direct relationship with the previous sentence. This statement is also very weak sounding and undercuts the importance of this study. Please rephrase.

We changed the manuscript accordingly (L387-389): "The magnitude of newly formed RTS on the QTP during the last decade potentially offsets the relatively low concavity depths, so that the total mass-wasting activity and material erosion volume during the last decade show a magnitude comparable to the thermokarst landscapes in the high Arctic."

L427: "the hinterlands". This isn't the right word. Maybe try "adjacent areas"?

We changed this part of the manuscript according to the reviewer's general comment regarding restructuring the discussion. Hence, the original sentence does not exist any longer.

L463-465: "In situ observations from northeast QTP showed a 23 to 37 % loss in surface (> 20 cm) (Mu et al., 2017, 2020; Liu et al., 2018) and slightly less (~ 20 - 28 %) for SOC stocks in deeper soils (Wu et al., 2018; Yi et al., 2025)." Loss of C? To the atmosphere? In an incubation or field measurement? During what time period? From an RTS? More detail is needed here to understand how this fits in.

We have changed the manuscript to account for the reviewer's comment and to be more precise (L435-437): "However, several studies that conducted soil sampling within disturbed permafrost areas for several consecutive years found that up to one-third of the surface SOC content (< 40 cm) has potential to be lost due to rapid permafrost thaw (Mu et al., 2017; Liu et al., 2018; Wu et al., 2018)."

L467-469: "Wang et al. (2020) estimated for QTP 0.19 to 0.38 × 10^{13} kg C (or 1.9 - 3.8 Pg C) from formerly frozen SOC will be subject to decomposition upon permafrost thaw until 2100 under moderate and high-emission climate scenarios that could switch QTP from a carbon sink to a source." The wording is a bit clunky here. Please rephrase.

We changed this part of the manuscript according to the reviewer's general comment regarding restructuring the discussion. Hence, the original sentence does not exist any longer.

L476-477: "However, we found that only 2 % of SOC mobilised by RTS activity contributed to the total loss of SOC stocks at soil depths below 3 m." Unclear, please rephrase.

We changed the manuscript accordingly (L409): "However, only 2 % of the total estimated mobilised SOC came from soils below 3 m depth."

L479-483: "Parts of mobilised SOC remain on the slump floor and are available for microbial decomposition and release as greenhouse gases (Wang et al., 2024) or deposited and even stabilised (Thomas et al., 2023; Liu et al., 2021, 2018; Mu et al., 2017), while other parts, together with thawed material, are laterally transported downslope into adjacent river and lake systems and SOC can undergo complex water chemistry processes such as dissolution or sedimentation (Lewkowicz and Way, 2019)." This is pretty long. I'd recommend breaking it up into two sentences.

We changed the manuscript accordingly (L411-415): "Parts of mobilised SOC remain on the slump floor and are available for microbial decomposition and release as greenhouse gases (Wang et al., 2024) or deposited and even stabilised (Thomas et al., 2023; Liu et al., 2021, 2018; Mu et al., 2017). Other parts, together with the thawed material, are laterally transported downslope into adjacent river and lake systems, and the mobilised SOC can undergo complex water chemistry processes such as dissolution or sedimentation (Lewkowicz and Way, 2019)."

L498-499: "Some permafrost regions in the Arctic have been focus areas and DEMs of two to three time steps in the last decade exist." This feels unnecessary to me.

We removed this statement following the reviewer's suggestion.

L519: "On the QTP, no study, to our best knowledge," would be better as "To our knowledge, no study on the QTP"

We removed this sentence.

L520-522: "These limitations highlight that applying such a scaling law to optical RTS inventories should be done carefully and rather to obtain regional estimates on material erosion volume and mass wasting derivatives such as ground ice loss or SOC mobilisation." Unclear, particularly "and rather". Please rephrase.

We changed the manuscript accordingly (L476-477): "Applying area-volume scaling to multimodal RTS datasets should be done carefully to obtain regional estimates on material erosion volume and mass wasting derivatives such as ground ice loss or SOC mobilisation."

L526-528: "We could show that remote sensing data with a spatial resolution of ~ 10 m misses ~ 35 % of RTS features that could be found in higher resolution PlanetScope images which, however, only accounted for a difference of < 1 % of the eroded material volume (Fig. 5 c)." It is unclear what "could show" means here? Does your work show this? If so, how?

We rephrased this sentence to account for the reviewer's comment (L485-487): "We showed that RTS monitoring on elevation change maps based on DEMs with a 10 m spatial resolution miss ~ 35 % of the present features compared to monitoring RTS with 3-m optical multispectral images. However, the difference in estimated material erosion volume from the two datasets is < 1 % (Fig. 5 c)."