We thank the reviewer for the detailed and constructive comments. We considered each comment carefully and address them point by point below. We hope that our modifications will make the manuscript clearer and more complete.

In the following we abbreviate: RC1 (Reviewer 1 Comment), RC2 (Reviewer 2 Comment) and
5 AC (Author Comment)

Reviewer 1

RC1 paragraph 2

The analysis focuses first on thaw slump volume, the average volume change among the three change models potentially available (ln 122-123). It is important to note in this paper that thaw
10 slumps are chronic, multi-year (often multi-decade) features that produce variable eroded volume over time, and the erosional intensity and morphological complexity tends to change with the age of the feature. This is a critical point of distinction from landslide studies that commonly examine the scaling of the total erosion from a scar zone. Both approaches have important outcomes: from an annual yield perspective, the area-volume scaling relationships presented here agree well with
15 established power law parameters, and the resultant regression will likely be helpful for estimating annual yield from mapped RTS polygon areas. From the perspective of the full scar erosion depth, measures of the time-integrated changes in morphology can yield a regression that might tell us more about the longer-term trajectory of the landscape. However, generating a 'pre-disturbance' surface can be time consuming, and there is the prospect of erroneous reconstruction, particularly
20 in the case of larger slumps in complex topography. I'm not suggesting the latter analysis be incorporated, but it is important to highlight this distinction.

AC

We agree that this distinction is an important point to make and we edited the manuscript at two points to highlight this distinctions:
25 In Methods-Section 3.3 we added:
"It is to note that RTS are multi-year features with a strong variability in the erosional intensity as well as a potential change of their morphology over time. In the interpretation of the results and specifically the comparison to landslide studies the use of the integrated change over several years needs to be considered."
30 in the Discussion-Section 5.1 we added:
"It should be emphasized here that another difference of our analyses to common landslides studies is that RTSs are a multi-year phenomena with variable yearly erosion rates. Some variability in the exact form of the distributions should therefore be expected if different time periods are chosen."

RC1 paragraph 3

The authors should perhaps clarify that the "area" term denotes primary scar zones (only) - not including spoil zone or other reworking, for clearer comparison with other datasets. There is invariably some detritus that fills the primary erosion site, particularly in older and larger RTS features on more subdued slopes, so the precise volume of most recent erosion is not always accessible.

AC

To clarify the term "area" is this context we added to the Section 3.3 RTS attibutes the sentences: "For all calculations we used the area outlined by the polygon indicating the areas showing an elevation change and thus a net volume loss. It is to note that this area can also be a zone of deposition, especially for small and low-relief RTSs or if the time between observations increases. Areas such as the debris tongues or zones of alluvial deposits can not be accurately detected by the DEM difference data and are not included."

RC1 paragraph 4

The term "volumetric change rate density" (ln 127) is clarified as "volumetric change per unit area", but the statement goes on to say this is calculated "by dividing the study region size by the total volumetric change rates", which seems to be rather the reciprocal – and a "change rate" (e.g. ln 275) is different from volumetric change. I'm perhaps misunderstanding your intent here, but some clarification of this specific yield term is needed.

AC

We introduced the term "volumetric change rate density" to investigate how much volume is eroded in a specific area (analogous to a "RTS density" - the number of RTS found in an area e.g. a square kilometre). We made a mistake in saying that the "volumetric change rate density (volumetric change per unit area)" was computed by dividing the study region size by the total volumetric changes per year. We added some clarifications and the correct sentence is: "To quantify the volumetric change rate density (volumetric change rate per unit area) we first use a simple approach by dividing the summed total of all RTS volumetric changes per year by the study region size." We hope that with this mistake fixed is is clear what we are trying to do.

RC1 paragraph 5

While the TanDEM-X elevation dataset has broad statistical characterization of the vertical accuracy (ln 101-102), the problem of volumetric change in landslides, gullies and other mass-wasting zones present a more specific problem: how well is the the scar zone volume characterized by the

grid of elevation values interpolated in and around it? Given the focus on allometric relationships, it is important to assess the propagation of various errors, some that are likely to vary with scale. As the scale of erosion features approaches the pixel resolution, the estimated volume will be increasingly approximative. Admittedly the problem of error characterization and propagation in landslide inventories has not advanced very far generally, but given that this work could be a stepping stone to even further extrapolations of sediment and carbon export, it would be quite helpful to establish a list of factors that contribute to error and some estimation of the overall precision that can be achieved with this methodology. Some calibration with finer-scale elevation datasets could help with this problem, as well.

AC

We agree with the reviewer that a detailed quantification of errors would be very helpful, unfortunately quantifying the error in a rigorous way is very challenging due to the combination of a spatial, a vertical as well as a time component that is necessary to characterize a RTS. To investigate this in more detail a reference dataset with a high resolution change estimates with a sufficiently closely matching temporal period. This is currently not available. We addressed this problem in greater depth in your last publication ((Bernhard et al., 2020)) To include a more detailed error estimation and uncertainties we added the following to the Method section 3.2:
"The lower limit for a RTSs to be detectable in terms of headwall height and retreat is very hard to quantify due to the limited amount of available high resolution, three dimensional RTS inventories. Here also the timescales on which the RTSs are monitored plays an important role. The 90th percentile in terms of elevation changes of the 10 smallest detected RTSs is in the range of $1.6\,\mathrm{m}$ to $2.1\,\mathrm{m}$ and can be seen as an approximation for the smallest RTS headwall heights that are detectable. Similarly, the smallest total area changes of detected RTSs are on the order of $1000\,\mathrm{m}^2$ to $1500\,\mathrm{m}^2$. If the size of the erosion features approaches the pixel resolution also the accuracy of the estimated volume loss increases. InSAR related processes play here the biggest role like the about 45 degree right looking viewing geometry in an ascending orbit and inaccuracies in the estimated coherence. These error sources and increased uncertainties especially for small RTSs, both in terms of spatial and vertical changes, should be considered in the interpretation and future use of the dataset."

RC1 paragraph 6

The results show some noise in the scaling relationship, which is certainly not unexpected given the diversity of drivers and physiographic factors that govern thaw slump development. The capacity to explain this variability based on remotely-sensed landscape factors is limited, but as stated, with further refinement of methods and proxy measures of ground conditions (ice content, soil thickness, base-level controls), there is great potential to advance our understanding of the transformations of the landscape that are underway. It would be good to see some further

speculation on the reasons for variation in the scaling exponents in different regions - what do they signify? Section 5.2 is conspicuously brief on this. In the Banks Island dataset, for instance, smaller erosion features tend to be shallow surficial failures, resulting in proportionately smaller volumes in that part of the size spectrum, and thus a steeper regression curve. In the Peel Plateau setting, there is very little confining topography to arrest headwall development, and thus the relatively larger features can get very large, again contributing to a steeper relation. Other sites may see less topographic variation across scales, which might contribute to a shallower slope on the regression curve. Glacial legacy plays a very important role in moderating this relation, as well. Some further, fairly general, geomorphic terrain interpretation could yield insights into how conditions change with scale.

AC

We agree that a more thorough interpretation of the observed differences in the scaling coefficient of the regression line is warranted. We added some interpretation and furthermore discuss the age component of RTSs. We added to the Discussion section 5.2:
"On the other hand, for RTSs in the Peel Plateau there is only little confining topography and deep layers of ice-rich tills which allows the headwall to grow to large sizes and consequently a steeper regression curve (Lacelle et al., 2015). The diversity in landform characteristics also contributes to the scaling relationship. In the study areas Banks Island or Noatak, shallow detachments are dominant in the small-area range. They may promote larger scaling coefficients when combined with older, deeper thaw slumps (Lewkowicz, 1987). Furthermore, most RTSs initiate as shallow active layer detachments. The gradual transition following an extreme initiation event could lead to a temporal change in the scaling coefficient. Further investigations relating the scaling coefficients to additional RTS and area characteristics (e.g. soil properties, climatic history, age of the RTSs) are needed. "

RC1 paragraph 7

The prospects for broad scale repeat monitoring thaw slump evolution is appealing - the work presented here shows that with a good supporting dataset of landscape information there a good possibility of achieving this, and advancing models for periglacial landscape evolution in the Anthropocene. The paper is well structured, and the charts and graphics are nicely rendered, but there are quite a few typos and grammatical issues in the text; this should be carefully reviewed before resubmission. There is some confusion regarding the numbering of figures in Section 4 and 4.1 (Figs 3-5) that require some attention. A few points are listed below. With the resolution of these minor points and a few points addressing error/uncertainty and some interpretation of the regression slopes, I recommend advancing this paper to publication.

AC

We carefully reviewed the manuscript for typos and grammatical errors and corrected the confusion related to the figure labelling. We corrected all listed points.

140 **Typos**

RC1

l. 18: suggest 'underlain' by permafrost

RC1

l. 29: nutrition => nutrients

145 **RC1**

l. 35: insides => insights

RC1

l.39: It's not clear to me that a frequency distribution is a scaling law.

RC1

150 l.40: 'disturbed area' => more explicitly, the erosion site (not deposition)

RC1

l.48 were => where

RC1

l.49 vertical - I think you mean horizontal, here.

155 **RC1**

l.50 were => where

RC1

l.56 only became available in the last few years

RC1

160   l.59 ..and have observed the global land mass two to three times, now.

RC1

l.72 ..modelling and will further improve..

RC1

l.83 Due to the large extent of some areas

165   **RC1**

l.95 The incident angles (?)

RC1

l.97 We only studied winter acquisitions due to the low..

RC1

170   l.144, 147 - beware affected vs effected - meaning is not clear, here.

RC1

l.150 volume

RC1

l.167 ..is to relate the area..

175   **RC1**

l.177 violine => violine. Use either comma or thin space between thousands in presentation of numbers.

RC1

l.184 exponential

180 **RC1**

l.185 potentially

RC1

l.199 By contrast, the PDF based on..

RC1

185 l.204 exponential decay coefficients

RC1

l.248 ..this universal scaling also applies to permafrost landscapes..

RC1

l.258 ..order of magnitude higher growth rates..

190 **RC1**

l.259 relict ice

RC1

l.313 ..data availability only allowed us to compute elevation changes..

RC1

195 l. 325 insides => insights

Reviewer 2

R2

Line 18: Only 15% of the Northern Hemisphere is underlain by permafrost. Please correct accordingly. Explanation for why one quarter is not the correct value can be found in this publication: https://doi.org/10.1029/2021JF006123 . I would also suggest adding it as a reference.

AC

We implemented the correction and added the reference.

RC2

Line 54: "only remote sensing techniques are feasible": I would argue that other techniques are not impossible to perform. Consider changing to "more feasible" or similar.

AC

We changed this part to "remote sensing techniques are the most feasible".

RC2

Line 55: The sentence is not clear.

AC

We omitted the term "supra-regional" and changed the sentence to: "Digital elevation models (DEMs) that cover the pan-Arctic permafrost terrain with a high enough resolution to study RTSs became only available in the last few years."

RC2

Lines 110-112: "It is difficult to assess RTS delineation procedure based only on this text. I suggest adding examples of RTS delineation from each site to the supplement."

AC

We added several examples of RTS delination with aditional false-color optical Sentinel-2 images to the supplement.

RC2

Lines 117-118: It is not clear for reader at this point what "shoreline" and "hillslope" relate to.

AC

We added some clarifications: "Additionally, the location of the RTS in terms of "shoreline" (located close to a waterbody) or "hillslope" (located at trenches or riverbeds) was noted."

RC2

Line 120: From the text, one could assume that you only delineated RTS according to the active elevation change occurring at headwall. Substantial area of RTSs is a zone where material is transported and no significant elevation change occurs.

AC

Yes we only delineated RTSs according to the active elevation change. In the DEM difference images only this part is clearly visible. To make this distinction more clear we specify the zone we delineate as primary scare zone and explicitly exclude the debris tongues. We added to Section 3.3 RTS attibutes the sentence:
"For all calculations we used the area outlined by the polygon indicating the areas showing an elevation change and thus a net volume loss. It is to note that this area can also be a zone of deposition, especially for small and low-relief RTSs or if the time between observations increases. Areas such as the debris tongues or zones of alluvial deposits can not be accurately detected by the DEM difference data and are not included."

RC2

Lines 168-177: This part would be more suitable for the methods section. Consider moving it.

AC

We assume that the reviewer refers to the sentence: "Due to the low density of RTSs in Yamal and Gydan and the two study regions in Taymyr we combined these to one study region (in the following "Yamal/Gydan" and "Taymyr") according to their geographical and geophysical proximity." We think that there is no clear section in the methods were we could add it and think that the decision to merge them are motivated by the results and could stay there.

RC2

Line 189 and other sentences describing figures before and after: It would be more appropriate to refer to the figure inside the text, instead of writing separate sentences just for that purpose. I would suggest extending these sentences to describe the main results shown in these figures.

AC

We changed the manuscript to the reviewers suggestions and rewrote and removed several sentences.

RC2

Line 204: Is there an established methodology on how to quantitatively discriminate megaslumps from slumps? Otherwise I would suggest avoiding this term.

AC

The term "mega slump" is relatively new but has been used in several recent publications (e.g. (Lacelle et al., 2015; Kokelj et al., 2015; Jones et al., 2019). The exact definition varies slightly but generally includes RTSs with an area larger then $20\,ha$ $(2 \cdot 10^5 \mathrm{m}^2)$.

RC2

Line 254: The correct would be "relict ice". Since not all of the relict ice is necessarily excess or massive ice, using "massive" or "excess" would be more appropriate.

AC

We changed the term to massive ice.

RC2

Lines 296-297: This statement might be too general. Differences between the aspects according to solar radiation might be relatively small in the high Arctic, where also north-facing slopes receive quite some solar radiation during the Arctic summer. Given that RTS headwalls are close to vertical, relatively low sun angles might still be efficient in melting ice.

AC

We think that the aspect factor does indeed plays an important role in RTS activity. We agree with the reviewer that in the case of vertical headwall the impact reduces, but considering the our

study areas are located at 60-75 degree latitude and that RTSs initiate and grow at slope angles of
15-25 degree the difference in the energy availability due to solar irradiation plays an important
role. See for example Figure 3 in (Ohmura, 2012) for the hourly ration measured during summer
at Summit (Greenland) at 72.6 degree latitude.

Typos

RC2

Line 20: Missing comma after "Furthermore".

RC2

Line 29: You probably meant nutrient cycles.

RC2

Line 54: "Arctic RTSs": Are there any RTSs outside the Arctic? Consider omitting "Arctic"

RC2

Line 132: You probably meant "than" instead of "then".

RC2

Line 137: Missing comma after "For this computation"

References

290 Bernhard, P., Zwieback, S., Leinss, S., and Hajnsek, I.: Mapping Retrogressive Thaw Slumps Using Single-Pass TanDEM-X Observations, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13, 3263–3280, https://doi.org/10.1109/JSTARS.2020.3000648, 2020.

Jones, M. K. W., Pollard, W. H., and Jones, B. M.: Rapid initialization of retrogressive thaw slumps in the Canadian high Arctic and their response to climate and terrain factors, Environmental Research Letters, 295 14, 055 006, 2019.

Kokelj, S., Tunnicliffe, J., Lacelle, D., Lantz, T., Chin, K., and Fraser, R.: Increased precipitation drives mega slump development and destabilization of ice-rich permafrost terrain, northwestern Canada, Global and Planetary Change, 129, 56–68, https://doi.org/https://doi.org/10.1016/j.gloplacha.2015.02.008, 2015.

Lacelle, D., Brooker, A., Fraser, R. H., and Kokelj, S. V.: Distribution and growth of thaw slumps in 300 the Richardson Mountains-Peel Plateau region, northwestern Canada, Geomorphology, 235, 40–51, https://doi.org/10.1016/j.geomorph.2015.01.024, 2015.

Lewkowicz, A. G.: Nature and Importance of Thermokarst Processes, Sand Hills Moraine, Banks Island, Canada, Geografiska Annaler: Series A, Physical Geography, 69, 321–327, https://doi.org/10.1080/04353676.1987.11880218, 1987.

305 Ohmura, A.: Present status and variations in the Arctic energy balance, Polar Science, 6, 5–13, https://doi.org/https://doi.org/10.1016/j.polar.2012.03.003, special Issue: The Second International Symposium on the Arctic Research (ISAR - 2), 2012.