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 AC (Author Comment).

**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 RTS 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 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 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:

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."

In the Discussion-Section 5.1 we added:

---

**Reply on RC1**
Philipp Bernhard et al.

Author comment on "Assessing volumetric change distributions and scaling relations of retrogressive thaw slumps across the Arctic" by Philipp Bernhard et al., The Cryosphere Discuss., https://doi.org/10.5194/tc-2021-137-AC1, 2021

[Further content added here...]

---
"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 attributes:

"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" (In 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. In 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 the new wording contribute to a better understanding.

**RC1 paragraph 5:** While the TanDEM-X elevation dataset has broad statistical characterization of the vertical accuracy (In 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 is needed. 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 statement 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 1.5 m to 1.9 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 500 m$^2$ to 1000 m$^2$. If the size of the erosion features approaches the pixel resolution also the accuracy of the estimated volume loss increases. Here InSAR related processes play 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 et al. 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 typos. We are thankfull for your advice, comments and suggestions to improve the paper.


**Corrected Typos:**
1. 18: suggest 'underlain' by permafrost
2. 29: nutrition => nutrients
3. 35: insides => insights
4. 39: It's not clear to me that a frequency distribution is a scaling law.
5. 40: 'disturbed area' => more explicitly, the erosion site (not deposition)
6. 48 were => where
7. 49 vertical - I think you mean horizontal, here.
8. 50 were => where
9. 56 only became available in the last few years
10. 59 ..and have observed the global land mass two to three times, now.
11. 72 ..modelling and will further improve..
12. 83 Due to the large extent of some areas
13. 95 The incident angles (?)
14. 97 We only studied winter acquisitions due to the low..
15. 144, 147 - beware affected vs effected - meaning is not clear, here.
16. 150 volume
17. 167 ..is to relate the area..
18. 177 violine => violine. Use either comma or thin space between thousands in presentation of numbers.
19. 184 exponential
20. 185 potentially
21. 199 By contrast, the PDF based on..
22. 204 exponential decay coefficients
23. 248 ..this universal scaling also applies to permafrost landscapes..
24. 258 ..order of magnitude higher growth rates..
25. 259 relict ice
l.313 data availability only allowed us to compute elevation changes.
l. 325 insides => insights