Anonymous Referee #1

This manuscript represents an important step in our understanding of permafrost thaw dynamics from the Canadian Arctic. As the authors point out, the permafrost regions of the Arctic are not all similar, and region-specific work such as this are critical to our understanding of the impact of climate change on the Arctic as a whole. The key finding of the manuscript, that permafrost thaw slumps may in fact reduce DOC delivery to Arctic streams, is timely and should be of interest to permafrost biogeochemistry researchers in general. I have some minor concerns detailed below, but I don't think they will greatly impede publication of this manuscript. The manuscript is well written, and the tables and figures are well presented and generally very clear.
We thank Reviewer 1 for these positive comments, and for the detailed review. We have incorporated almost all of the reviewer suggestions, as detailed below.

Note that the 'current line numbers' listed below refer to line numbers in the uploaded, amended manuscript document. Because of the tracked changes in this document, there are fairly significant gaps in line numbers between some pages.

One general comment is, and this could be addressed in the discussion for example without necessarily the need for extra data, what is the significance of this main finding to the overall carbon cycle/budget for such landscape features? The authors measure total suspended sediment (TSS) loss from the study features, but give no indication of the carbon content (I guess it wasn't measured). If DOC export is reduced due to adsorption to fine-grained sediments, these sediments are also mobilised and exported, and must carry some carbon. In Figure 3A we see that TSS increases downstream of these thaw features (unlike DOC), so can anything be said about the fate of the carbon that is locked up in this flux?
Thanks for this good suggestion. We do see significant particulate organic carbon mobilization from these features; documenting this mobilization and the fate of this particle-attached carbon is currently underway, and being studied by another graduate student in our research group. While we have not added POC data to the current manuscript, we have added several lines of text to acknowledge the importance of particle-associated carbon flux (current L635-683).

Specific comments:

L. 17 (and/or introduction L. 66-69) – I recommend defining retrogressive thaw slumps specifically early on, do they differ from a normal thaw slump (active layer detachment, slide for example), and is the "retrogressive" characteristic of this type of slump especially different from thermokarst processes in general?
Agreed. We have added a clearer definition of retrogressive thaw slumps in the Introduction (current L86-87, including a reference to the image in Fig. 1), but not in the abstract, given space constraints.

L. 49-51 – some resilience in the region is also possible in response to gradual permafrost thaw (e.g. Dean et al. 2016 doi: 10.1007/s10533-016-0252-2).
Agreed. We have softened our wording in this sentence (current L54-56).

L. 53-56 – is DOC the primary substrate in soils or during aquatic transport/storage? I don't think there is a clear consensus on this point, and it's not clear exactly what your point is in this sentence. Yes, DOC can degrade to produce CO_2 (and CH_4) in Arctic aquatic environments, and this has been highlighted by recent studies (e.g. Spencer et al. 2015 GRL; Drake et al. 2015 PNAS; Mann et al. 2015 Nat Comms). But that doesn't mean DOC in the aquatic zone is the primary source of CO_2 released from streams and

lakes. Most of the CO2 released from streams is generated in the soil zone and transported laterally (Hotchkiss et al. 2015 Nat Geosci; Marx et al. 2017 doi: 10.1002/2016RG000547. So, I think it's important to not be too throw-away with this point, and be a bit clearer about how this aspect of Arctic aquatic carbon cycling relates to the study presented here.
Fair comment! The point we are trying to make here is a simple one – namely that DOC is a useful focal point when thinking about permafrost thaw and carbon, because it is the primary substrate for microbial mineralization. We've modified the sentence to be more inclusive of soil pore waters and the full aquatic continuum, by changing the text and removing the reference to Spencer (current L59-63).

L. 86-89 – would be good to compare this to pan-Arctic thermokarst estimates (e.g. Olefeldt et al. 2016 doi: 10.1038/ncomms13043).
At this point in the text, we are referring to individual thaw slumps ("individual thaw slumps commonly impact tens of hectares of terrain …"). We have modified our text slightly to clarify this point (L105-106), and have also made modifications towards the end of this paragraph (current L122-123) to state that "this till-associated, RTS-susceptible landscape type is found across the Laurentide and Barents-Kara glacial margins of Canada, Alaska, and Siberia".

We do not reference the Olefeldt paper here, because it provides a broad estimate for "hillslope thermokarst", which includes not only retrogressive thaw slump features, but also active layer detachments and thermal erosion gullies. However, we do now include this reference towards the end of our Discussion section.

L. 110-112 – please provide a reference to support this, or make it clear that this statement is hypothesis at this stage.
We have moved the Kothawala reference up from the sentence below to provide a clear citation for this statement.

L. 126 – please be quantitative and give an area estimate, rather than saying "large portions" of the Arctic. Earlier the authors emphasise that the Peel Plateau is different, hence the uniqueness of this study. Please clarify the aims and intent regarding this point.
We do not provide an area estimate (there isn't one available! See response above to L86-89 comment). However, we have modified the text to be much more specific, and read: "glacial margin landscapes throughout Canada, Alaska, and Siberia" (current L176). Based on this comment, we have also made slight modifications at other points in the introduction, to convey our meaning that the Peel Plateau is different from places where permafrost carbon release <u>has been studied previously</u>, rather than unique from a pan-Arctic perspective (see, for example, current L161-164). Thanks for helping us to clarify this important point.

L. 180 – so these were exceptionally wet years? ~ 100 mm greater than the monthly averages? What is the significance of this enhanced precipitation to the DOC and TSS dynamics described in this study compared to other years?
2014 was a wet year, although the recent trend on the Peel Plateau has been for exceptionally wet weather over the past ~decade, when compared to long-term norms. However, in double-checking this data, we noted that the comparison between 2014 and long-term average data was made between a (recently established) station on the Peel Plateau (for 2014), and a longer-term station at Fort McPherson (for the long-term average). We have now clarified this in the text, and added 2014 data for the Fort McPherson station. 2014 is still wetter than average, but the difference is not as substantial as it originally appeared (current L265-271).

L. 193 – how were the sites chosen to be representative? Was this done with remote sensing, or based on the experience of the authors? Either is fine, just to clarify.
Sites were chosen via an initial aerial (helicopter) survey, and previous knowledge of this landscape. We have modified the text to clarify this point (current line 275-276).

L. 207 – suggest change to "...geomorphology were previously described by..." Or is the Malone et al. data actually used here?
Modified as suggested. We do not use the Malone data in this paper; this sentence is merely intended to acknowledge this prior work.

L. 334-335 – would be good to clarify here and in the discussion, that only DOC concentrations are considered here, not fluxes. Dilution from the high rainfall reported for the study period could play a part. That said, the flux values from the FM3 site in Fig. 6 support the findings as presented.
We certainly recognize the importance of differentiating between DOC concentration and flux, and have made significant efforts to ensure that we are clear about when we are discussing one vs. the other. Towards this end, we have specified "DOC concentration" in the header for this section (4.1), and specify in this sentence that "DOC concentrations tended to be lower …". To address this point in the discussion, we have added some text to our first Discussion section that specifically reiterates the flux (vs. concentration) results (current L587-588), and worked to clearly present (and demarcate) flux and concentration results in section 5.4.

Section 4.2 – might be useful for compare the geochemistry to pristine waters from the NWT (e.g. Dean et al. 2016 doi: 10.1007/s10533-016-0252-2).
Thanks for pointing us to this useful reference. We have chosen not to include it here, because the Trail Valley Creek region is somewhat different from the Peel Plateau (more organic-rich soils; greater prevalence of ice-wedge polygons; north of treeline), as evidenced by the significantly greater DOC concentrations across streams sampled around Trail Valley Creek, when compared to on the Peel Plateau. Because our 'upstream' sites present values for pristine waters in the Peel Plateau region, we thought it was best to keep the focus of this section on differences in DOC concentration between thaw-affected and pristine sites on the Peel Plateau, specifically.

L. 346 – I wouldn't consider Ca2+ or Mg2+ to be conservative in general, it would be good to justify whether they are conservative in the current study system.
Fair comment! We have changed "conservative ions" to "major ions" here and elsewhere. Our intent for measuring Ca and Mg in this system was to track the impact of slumping, because we knew (at the outset of the study) that RTS features in this region had been shown to have significant concentrations of ions in their outflow waters.

L. 387-392 – with a rough mass balance, would it be possible to estimate the downstream DO14C values?
We don't undertake this type of analysis because samples for DO^{14}C were collected two years after the main study, to allow us to include this type of information in the paper. While it is unfortunate that our DO^{14}C samples are not true 'paired' samples with the DOC concentration data that is the focus of the paper, we feel they still contribute important information that aids with our understanding of the system. As a result of the offset in collection date, we keep our analysis of the ^{14}C data fairly simple.

L. 399-401 – this is a very important point, showing that the fluxes show the same pattern as concentrations (see comments on L. 334-335). Maybe emphasize this linkage of the FM3 findings to all the study sites.
Thanks for this comment – we agree that this is an important finding. We haven't added text to explicitly link the FM3 finding to all there study sites here (in the Results section), but do add some text towards this end in the Discussion (current L768-769), and in the methods, where we now outline the fact that FM3 was chosen for the more intensive flux work because it is roughly "representative of active slumps on the Peel Plateau with headwalls that erode Holocene- to Pleistocene-aged sediments" (current L362-363).

L. 421 – the Drake et al. and Mann et al. 2015 citations don't really fit here. These studies focus on degradation dynamics not mobilisation. Also, the Drake study used permafrost soils from Alaska. Please check the other references in this sentence.
Agreed that the Drake and Mann references are misplaced here. We have removed Drake, replaced the Mann reference with a reference to Spencer et al. (2015), re-worded this sentence slightly for clarity, and also removed a couple other references that were not entirely appropriate here (current L581-582).

L. 433-435 – DOC:ion ratios are key here to support the DOC flux vs. concentration issue I pointed out earlier. Why not emphasize these ratios more? (note comment regarding L. 346).
Thanks for this suggestion. We have added text to the results (current L504-508) and at this point in the Discussion (current L596-597) to further emphasize DOC: ion (and also DOC: TSS) ratios. We add these as inverse ratios (ie, ion:DOC) to more clearly demonstrate differences between upstream and downstream sites.

L. 439 – what field evidence specifically?
This statement refers to the evidence presented in the following few sentences. Text has been changed to "our results", for clarity (current L602).

L. 447 – why not data from the other sites too?
Our 'environmental controls' work at site FM3 provided us with the discharge data necessary for these calculations. The calculations specified in this section of the text were not planned – they were undertaken as a post hoc *assessment to help us understand the patterns we were seeing a bit better. As a result, we're only able to do these calculations at site FM3, and not elsewhere. This is now clarified in the Methods section (current L458-459).*

L. 449-454 – much of this should really go in the methods.
We have added details on these calculations to the methods (bottom of Section 3.4), and have deleted the text specifically dealing with the methodological component of this text from this section (current L 623-626).

L. 452 – TSS shouldn't be considered conservative in my opinion: entrainment and deposition processes would be occurring, so how can conservative behaviour be justified? Or is that the point here, that it isn't conservative? Hard to follow the reasoning here. Perhaps if the approach used here were more clearly outlined in the methods section, and the results presented more clearly (e.g. in a table), this might help clarify this point.
Thanks for this comment. We have changed our text here (current L624), and added text to the methods (bottom of Sn 3.4) to clarify our approach, which was to calculate a mass balance for DOC upstream and downstream of slump FM3, and couple this with a similar mass balance for TSS. Our calculations show

that the TSS load balances nicely (ie, downstream = upstream + inputs from slumping), while the DOC load decreases downstream of the slump, despite considerable DOC input from the slump feature. We've changed our wording to acknowledge that TSS is a "rough" (rather than conservative) tracer of slump inputs, given that entrainment and deposition are certainly occurring. We also add text to clarify that these calculations occur "over a < 1km span between upstream and downstream locations" (current L 462-463).

L. 469 – where is it sequestered? Within the exported TSS load, or in the depositional environments within the study systems? If sorbed to mineral complexes, is the carbon sill bioavailable? What does this process mean for the overall fate of the carbon released from the thaw slump features? This is an important aspect of the discussion that is currently missing.
Our hypothesis at this time is that the dissolved OC becomes sequestered on mineral surfaces. Although we do not know whether this carbon is still bioavailable (there is very little in the literature discussing decomposition of POC!) this is something other members of our research group are actively working on. We've clarified our text to explain that we expect that this sequestration is occurring onto mineral surfaces, and have also added text to discuss POC fluxes in general from these slump features (see also our response to the general comments by this referee, above; current L635 - 682).

L. 491 – what do these values actually say about the condition of the DOM? This could also be added to the results section.
We have added some text to describe what these values indicate for DOM character (current L699-700). This text is already present in the results section (Section 4.3, current L511 and 516), so we have not modified the text there.

L. 504-511 – would it be possible to quantify the relative contributions from these end members (e.g. Winterfeld et al. 2015 doi: 10.5194/bg-12-3769-2015)? Would make a nice addition to the manuscript.
We agree that this would have been a wonderful addition to the manuscript. However, we're not really able to undertake this kind of analysis because the samples are not paired. See also our response to the comment for lines 387-392.

L. 565 – "across multiple measurement points" – not all points were used?
We did use all of the data here: "across the multiple measurement points that we considered". To increase clarity, we've removed this component of the sentence (current L802).

L. 579-585 – this sentence begins with "This result clearly highlights ... " but the rest of the sentence is long and unclear. Please rewrite to clarify the sentence's clarity.
Fair comment! This has been split into two sentences and reworded for clarity (current L817-822).

L. 582-583 – Where is this evidenced in the present study?
We think the reviewer is referring to the statement that slump features "drain contemporary active layers". Because thaw slump headwalls expose the contemporary active layer, in addition to Holocene and Pleistocene layers (see Figure 1c and d), drainage from these features necessarily incorporates active layer materials. Hopefully our re-working of this long sentence (see comment and response for L579-585 above) also helps to clarify our text.

L. 597 – inter-regional or slump type? If referring to the results of the present study vs. previous/pan-Arctic findings, then state this more definitively.

We have re-worked the text in this section somewhat for clarity (current L851-856 and onwards). In this sentence, and the sentences that follow, we are making the point that we should expect permafrost thaw (and, it's biogeochemical consequences) to play out differently in different regions of the Arctic. We have changed our wording slightly to link this sentence more clearly to the one that follows it, in which we specifically compare our results to findings from other regions.

L. 602-606 – how important is thaw slump derived DOC to the Arctic carbon budget? Is it large enough to justify its inclusion in ESMs at this stage? Or is its inclusion into smaller-scale process models more justified? This information could also be incorporated into the introduction for context.
Our sentence here was considering more the fact that DOC release as a result of thermokarst (or, permafrost thaw generally) could be predicted somewhat with a consideration of factors such as Quaternary history, soil composition, and the nature of permafrost thaw. This, in-turn, could lend itself to modelling efforts to understand variability in the effects of thermokarst at the pan-Arctic scale. To address the reviewer's comment, we have modified our text in this sentence slightly, including changing the beginning of the sentence from "modelling efforts" to simply "efforts" (current L856-859).

L. 610 – is this explored in the present study?
I think this question comes from a change in terminology: we use "paleo active layer" here to refer to the Holocene-aged layer we discuss throughout the manuscript. Our text has been modified to "relative depth of the <u>Holocene-aged</u> paleo active layer" (current L864).

Technical corrections:

L. 42-43 – it's good practice to cite specific chapters in the IPCC reports rather than the whole report or the summaries as used here. *Done.*

L. 49 – remove the "s" from "thaws". *Done.*

L. 93 – Should this be "thaw-" or "thermokarst-" affected regions, rather than "permafrost-affected regions"? *"Permafrost" has been changed to "thaw".*

L. 141 – how brief is brief? Please give numbers. *We now specify "a maximum of 2,000-3,000 years" in the text.*

L. 225 – Helicopter! Cool. *Thanks! We definitely enjoyed this part of the study …*

L. 232-240 – how long was there between sampling and analysis, generally? *This detail has been added (current L355-357).*

L. 257-265 – please provide precision/sensitivity estimates for each analysis. *Precision or error estimates have been added as applicable throughout this paragraph (current L383 and onwards).*

L. 300 – define AIC acronym when used for the first time. *Done.*

L. 423 – missing space between comma and permafrost. *Space has been inserted.*

L. 457-460 – reference? *The reference had been provided at the end of the following sentence; it has been moved up for clarity.*

L. 483 – I would add the following references from the region near the current study: Street et al. 2016 doi: 10.1002/2016JG003387; Quinton and Pomeroy 2006 doi: 10.1002/hyp.6083). *We have added the Street reference here, but have not added Quinton and Pomeroy because it focuses on major ion chemistry, rather than organic constituents.*

L. 586 – remove semi-colon. *Done.*

L. 612 – missing an “as”. *Inserted.*

L. 612 – should be Vonk et al. 2015b? *The Watanabe reference at the end of this sentence is as intended. The Watanabe study looks at within-region variations in lake chemistry in response to thermokarst. We like this specific study for the argument we are presenting in this sentence, rather than the more general review presented in Vonk et al. 2015.*

L. 619 – suggest changing “effects” to “impacts” otherwise too many effects/affects in this sentence. *We avoid this use of ‘impacts’ because it has a specific meaning in the geophysical sciences, but have re-worded this sentence to avoid the (substantial!) repetition (current L896 onwards).*

Table 2. Is the “t” from the Wilcoxon output? *The t is from the mixed effects model. This is now clarified in the Figure caption.*

Please add the figure captions to the figures next time (personal preference). *Figure captions are included with figures for this re-submission.*

Figure S1. These pictures are impressive, why not include in the main manuscript? Is it possible to include a rough scale in the photos also? *We put a fair bit of consideration into this request, but have decided to leave the images in the Appendix because (1) they are somewhat repetitive with panels c-e in Figure 1 and (2) to avoid an extra page of manuscript length. We haven’t added a scale bar because we didn’t take measurements that would allow us to do this when we took the photographs. However, the figure caption does point the reader to Table 1, which provides aerial estimates for each of the slumps that we studied.*