Response to Reviewer Comment by Anonymous Reviewer on "Changing Characteristics of Runoff and Freshwater Export From Watersheds Draining Northern Alaska" by M. A. Rawlins et al.

We are very grateful to the reviewer for the comments on this manuscript. We provide responses in blue below. Line numbers refer to the updated manuscript.

Review #1

This paper presents interesting results from a hydrological modeling study examining how runoff partitioning from arctic catchments is changing. The authors present an updated version of the Pan-Arctic Water Balance Model to better represent soil freeze-thaw processes and have renamed it the Permafrost Water Balance Model version 3 (PWBM v3). In general, the authors use the model to demonstrate that cold season discharge and groundwater flows are increasing in four arctic basins underlain by continuous permafrost. The authors do a very nice job of characterising how runoff and terrestrial water storage is changing in arctic catchments. This study is limited to basins underlain by continuous permafrost and differs from other work in that they do not attempt to generalise findings from large northern regions spanning different permafrost distributions (which is a good thing). The results and discussion are limited explicitly to model outputs, which are supported by only basic model validation from observed measurements. Without a better understanding of how the model performs, it is difficult to determine how valid the model outputs are, as well as potential errors associated with the outputs. Additionally, the novelty of this study is questionable as the main conclusion of this paper (as is stated many times in the discussion) is that arctic catchments are exporting increased runoff via subsurface pathways, which has previously been demonstrated in the literature. I think that this modeling study could been important contribution; however there are several significant revisions and additions that are required.

Major Points: A major weakness of this manuscript is the lack of model validation and performance evaluation. At this point it is impossible to understand how well the model performs, and consequently impossible to comment on whether the outputs area realistic interpretation of the physical system. By only discussing the outputs of the model there is potential for a large disconnect between what is being presented and the system for which the authors are trying to represent. Why is only one basin (Kuparuk)used for validation? There are other suitable gauged basins by the United States Geological Survey and the Water Survey of Canada that could be used as validation. This component is crucial to the success of the paper. The only validation presented in the results section states that freshet volume was similar, yet even on a monthly time step the model performance is weak (~30% error in both May and June). If the authors want to describe how the partitioning of runoff is changing by exclusively examining model outputs then it is imperative to prove that the model can simulate observations. To do this, it is necessary to use a finer resolution than monthly time-steps.

We appreciate the review of our manuscript. We have revised the draft to include additional validation comparisons with observed data, and the manuscript now includes a model validation section. Line 253. We have added a validation against river discharge for the Colville River. In the validation section we show that average active layer thickness closely matches estimates from another model (GIPL) developed at the Permafrost Laboratory, Geophysical Institute at the University of Alaska-Fairbanks. The PWBM captures the expected north-south spatial gradient, as does GIPL. In the validation section we also show and describe a comparison with SWE data across the Kuparuk basin. Model simulated end of season SWE is correlated ($r = 0.78$, $p < 0.01$) with the observations. The model captures interannual variability. For validation we then show a significant correlation ($r = 0.74$, $p < 0.001$) with measured Kuparuk River discharge. The time series plot confirms that the model well represents the correct magnitude and interannual variability based on measured data. The error in May and June arises due to peak discharge in the model simulation that is approximately 8 days early compared to the observations. The total simulated discharge over the freshet period May and June has low error of just +0.3%. The freshet period is also well resolved for the Colville River, with error of 10%. The model is run at a daily time step. We disagree that accurate daily resolution in the evaluations is required. On the contrary, with a goal to quantify seasonal export of constituents such as dissolved organic carbon and other nutrients, reasonably well constrained monthly climatologies and well correlated interannual variability is sufficient. My coauthors and other colleagues have discussed this issue at great length. The processes leading to the changes we describe in this paper arise largely due to long-term warming, which was substantial over the region, some 4.5 F warming over the 30 year period 1981-2010. How well the model simulates runoff on a daily basis has little bearing on its ability to simulate the processes fundamental to the myriad changes observed by other researchers and simulated via the PWBM. We will add that the study domain extends only a short distance into Canada, and we are aware of no observed discharge data for the small rivers in that area. We feel that we have a robust model validation given the paucity of spatially extensive data available in this region. Paragraph at lines 328-340 details the available long-term data for the largest rivers.

Why is modeled cold season discharge not evaluated against observations? Surely USGS publishes this data.

No consistent observations exist for discharge during the Nov-Apr period for any North Slope river, with the exception of the Kuparuk. Our goal in this work is to quantify and understand the freshwater export for the North Slope region. While we always seek more data with which to evaluate and better understand shortcomings in the approaches, in the end we believe that a numerical model must be used to obtain regional estimates for cold season discharge. Validation at that scale is obviously quite limited.

Why is there no model performance evaluation? There are many different evaluation techniques (e.g. Nash-Sutcliffe Efficiency, Root Mean Square Error, Percent Bias, Kling-Gupta Efficiency), but none are presented in the paper, nor is the reader referenced to other papers where they may be presented.

The model performance evaluation is based on the average error, percentage error, and correlation. Line 241. Model evaluation metrics based on squared values, like the RMSE, are known to be biased. We cite Willmott et al., 2005 and Willmott et al., 2015. Line 244.

Willmott, C.J. and Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research*, *30*(1), pp.79-82.

Willmott, C.J., Robeson, S.M., Matsuura, K. and Ficklin, D.L., 2015. Assessment of three dimensionless measures of model performance. Environmental Modelling & Software, 73, pp.167-174.

Is there any model calibration? Are there any empirical factors used? More information is needed.

Empirical factors have been described in prior published studies: Rawlins et al., 2003, 2013 and Yi et al., 2015. Two parameters we adjusted in this study as calibration. They involve runoff and evaporation from the surface pool. We have added language that the model calibration involved the surface storage pool and the river flow routing. Lines 213-217. Please also see our responses below.

This manuscript describes intensification of the hydrological cycle and is supported through re-analysis data and modeling efforts. The manuscript would benefit from supporting data from observations. It would be useful to plot precipitation from climate stations across Alaska and northern Canada to prove this, as well as using or referencing snow survey data. Modeling these changes is important, however these modeled changes need to be supported by observations.

Intensification is first mentioned in the Introduction, as it is an important element of climate change. We include snow survey data. Shown in Figures S3 and S4. The study domain extends mere kilometers into Northern Canada, near the coast, and there are no weather stations in that small area just west of the Mackenzie delta. Our introductory information regarding hydrological cycle intensification reflects the findings from earlier published studies. We state that no significant change occurred over the North Slope study region over the 30 years period 1981-2010. Line 320. Interannual variability renders the small time changes as insignificant. Our results are independent of time changes in precipitation.

The authors use much of the discussion to suggest that the proportion of groundwater runoff is increasing, yet there is very little discussion of how the structure of these flowpaths is changing. As the study sites located exclusively in continuous permafrost. I would assume that these changes would be through supra-permafrost groundwater flow, but this is not explicitly stated.Is the ice-rich transient layer (Shur et al., 2005, Permafrost and Periglacial Processes,10.1002/ppp.518) accurately represented in the model? This ice-rich layer retards active layer thickening due to the high latent heat requirements for thaw, and would also provide an additional water source once thawed.

Yes. The model captures the saturated ice-rich conditions at the top of the permafrost. In the PWBM there are 10 layers spanning the upper 3 m of the soil model soil column. Low hydraulic conductivity results in high water content in the uppermost permafrost. In fact, the ability of the PWBM to capture the zero curtain effect, the processes of phase change resulting in a long period time where soil stays near 0 C during thaw and freeze, was described recently by Yi et al. (2019) which we cite. Results shown in Figure 9 and 10 and described at lines 394-405 suggest a connection between a deepening of the soil active layer and increasing subsurface flow. This is likely due to increased storage of water into fall that allows for runoff generation. As we point out, losses in soil ice also contribute to runoff. Further study is required to test the first process. The paper includes an extensive discussion of study results in the context of other recent work.

Does changing seasonality of precipitation affect runoff generation? Some sentences are taken directly from other papers. These sentences should be changed in an attempt to synthesize other literature. For example, lines 436-39: “St. Jacques and Sauchyn concluded that increases in winter baseflow and mean annual streamflow in the NWT were caused predominantly by climate warming via permafrost thawing that enhances infiltration and deeper flowpaths and hydrological cycle intensification (Frey and McClelland, 2009; Bring etal., 2016)”. This text appears almost exactly word-for-word in the abstract of that paper. I also find it odd that a sentence from another paper has two additional references after it. Actually, St. Jacques and Sauchyn (2009) propose reactivation of deep groundwater flowpaths by making linkages between streamflow and climate. Also, many of the basins in this study are underlain by discontinuous permafrost, which would promote recharge of sub-permafrost groundwater aquifers that provide baseflow to rivers, a process not applicable in thick, continuous permafrost. Again, the changing physical processes need to be explored.

Characterization of seasonal precipitation change is beyond the scope of our study. We mention in the Discussion that changes in seasonality may play a role in the trends documented in our study. Line 555-557. We agree that permafrost thawing may be enhancing infiltration and promoting deeper flowpaths. We have re-worded the statement where St. Jacques and Sauchyn (2009) study was cited. Line 479. The information is appropriate. For a region of largely discontinuous permafrost these important processes would be occurring across a land unit defined by the presence of permafrost, such as a north facing slope. Observations (eg Jorgenson et al., 2008 http://permafrost.gi.alaska.edu/sites/default/files/AlaskaPermafrostMap_Front_Dec2008_Jorgenson_etal_2008.pdf) show that the entire North Slope domain is underlain by continuous permafrost. Our results point to a deepening of the soil active layer which is leading to increased flow in the thawed zone, contributing to enhanced subsurface runoff generation. Losses in soil ice which outweigh gains in liquid storage also contribute to the increasing fraction of subsurface runoff as a proportion of total annual runoff. The Discussion section includes perspective on changing physical processes. Line 501-514. Also line 521 and line 529.

If the authors are going to validate and calibrate model, why only use it for a period in the past? Analysis of past data can be conducted reasonably well with measured data. The authors may be better served to also use the model as a predictive tool to demonstrate how a changing climate may affect the streamflow regime of arctic rivers.

We appreciate the suggestion. However, the comment is invalid. There is an extreme lack of measured data in this region. Our study focus is on characterizing the baseline hydrology for the area of northern Alaska draining to the Beaufort Sea coast, for the 30 year period 1981-2010, and on understanding changes that are occurring. It is not possible to do this from the few observations. For example, river discharge has been measured at the Kuparuk River near the coast for several decades. The Colville River has been monitored since only 2002, but not in every month. Aside from those two rivers, long term records, to our knowledge, do not exist. Measurements in the cold season, when low flows exist under river ice cover, are virtually non existent. Information on data for these rivers has been added at paragraph starting at line 328.

The figures need substantial revision and improvement. They are not suitable for publication in their current form. The authors should provide a study site map delineating all four watersheds, as well as a layer identifying each underlying permafrost zone.

New map of the study domain has been added. Results maps include outlines of the Colville, Kuparuk, and Sagavanirktok rivers to aid in interpretation of results. The region is one single zone of continuous permafrost.

Line 108 states that the study area is underlain by continuous permafrost. Is this the case for the entire study site?

Yes. All prior studies published by other researchers for this part of Alaska suggest continuous permafrost is present over the entirety of our study domain.

Figure 1 is a very important figure and does not suffice as model validation. For example, the figure should be presented on a daily time-step (not aggregated into monthly intervals) to demonstrate how the model captures individual events. For example, there are substantial differences between May and June runoff, suggesting that the hydrological behaviour of the basin may not be captured.

We have added a new section on Model Validation including daily average discharge for the Kuparuk and Colville rivers. However, to quantify seasonal export of constituents such as dissolved organic carbon and other nutrients, reasonably well constrained monthly climatologies and well correlated interannual variability are sufficient. We contend that the results shown in the new section clearly demonstrate that the model simulations are valid.

Also, all time series plots should include each data point instead of a continuous line-graph. The dashed-line in the simulation makes it difficult to observe performance. The formatting of all figures should be improved in this manner.

Figures have been modified accordingly.

Figure 4 should present discharge normalised over basin area. As a result, the North Slope shows disproportionately more discharge due to the much larger basin area. I am not sure why the authors decided to present the data this way, considering that Figure 1 presents normalised runoff. Also, the current format-ting makes it next to impossible to discern runoff trends for the three smaller basins.

We disagree. Our intent for this figure, in part, is to help illustrate the differences in discharge volume flux for those rivers, and show them in relation to total discharge for the full North Slope domain. This has relevance for the export of river-borne constituents. The volumes are not so different as to require displaying in unit depth. Average values are listed in Table 2. The trends and their statistical significance are described in the text. We do not feel that a separate figure panel is needed.

Figure 5 is slightly misleading as the plot only shows the grid cells with significant changes.

We disagree that illustrating the magnitude of change for grid cells bearing significant change is misleading. We have re-drawn the plot to include all grid cells, and it simply shows many dots overlapping one another near zero change. The analysis and significance are clear. To our knowledge it is not uncommon to present information in this manner.

Figure 6 shows that many grid cells do not have significant change – but Figure 5 suggests that there is an increasing proportion of subsurface runoff in June and decreasing in July, when in fact these proportions may be relatively constant if the whole dataset it included.

We’ve made no statement that our results suggest that the proportion of subsurface runoff has increased for averages across the entire North Slope. Figure 6 is for annual runoff (subsurface and total). Figure 5 is for months May to September, and annual. The figures are fundamentally different, and complementary.

Minor Points: Line 21: Can you better define region based on watersheds?

New map (Figure 1) shows the study region, which we define as all land areas draining to the Beaufort Sea coast, not including the Mackenzie River basin.

Line 21: Do not need the word ‘annually”, this is given in your units.

Word ‘annually’ removed.

Line 22: Is this volume derived from modeled results or gauges? If the former, this needs to be stated, if the latter, these gauges should be used for validation

The baseline river discharge estimates now include both measured and model simulated data. Phrase “A synthesis of measurements and model simulations …” added. Line 22.

Line 24: The authors need a better preface for their results. At this point it is unknown what the results are describing.

“Our results...” changed to “The simulations...”. Line 24.

Line 34: I am not convinced that this shift is representative of the physical system, given section 3.3 states errors in freshet timing. Again, displaying data on a daily time step for all basins would be beneficial.

New figure 3 shows simulated and measured at a daily time step for the two rivers where evaluation is possible. NOAA data (Climate at a Glance Tool (https://www.ncdc.noaa.gov/cag/divisional/time-series/5001/tavg/2/5/1981-2010?trend=true&trend_base=10&firsttrendyear=1981&lasttrendyear=2010) shows that air temperature averaged across the North Slope has warmed in April-May (average) by 5.4 F. We believe that warming in late spring is resulting in earlier snowmelt and, in turn, the timing of peak discharge. Proving this is beyond the scope of the present study. Timing for the Colville River is well captured. The bias in simulated time of peak discharge is assumed based on comparison with Kuparuk River gauge data is inherently part of the uncertainty in our reported trend in timing of peak discharge for the region as a whole. We agree that the uncertainty in timing is considerable. But we are convinced that the shift is real. That said, we are prepared to drop that result from the paper. We are in the process of improving the snowmelt sub-model which should improve timing based on comparisons with measured daily discharge.

Line 47: Provide references for “mean freshwater budgets across the land”.

That statement is backed by the cited Serreze et al. (2006) paper. We are unaware of any other studies that examine the mean freshwater budgets across the land, atmosphere and ocean domains.

Line 52: This sentence is redundant given the previous sentence.

Sentences combined: A warming climate is expected to lead to intensification of the hydrological cycle, including increases in net precipitation (P) at high latitudes, and evidence is emerging (Peterson et al., 2002, 2006; Rawlins et al., 2010; Zhang et al., 2013; Bring et al., 2016). Line 51.

Lines 53-55: What about shorter ice duration on lakes and longer seasons for evaporation?

We now mention these processes in a subsequent sentence. Line 56.

Lines 69-71: This areal loss of permafrost applies to sporadic and discontinuous permafrost. The study site described by the authors indicates very thick continuous permafrost. This discussion should be limited to continuous permafrost environments so that the physical processes occurring in different environments are not confused.

That citation is early in the Introduction section and speaks broadly to permafrost loss in general, so we feel it would be helpful to the reader. We are not opposed to removing it. In areas where permafrost is discontinuous, the relevant hydrological changes to which we refer are taking place locally where permafrost is present. For example, on north-facing slopes, or where soil carbon amounts are high.

Lines 75-77: Similar comment to above, most of the rivers described in the cited studies are either subarctic or underlain by discontinuous permafrost. Runoff generation is very different between the two environments and this needs to be stated if there is extensive discussion about these systems.

In areas of discontinuous permafrost, where land units contain permafrost, the runoff generating processes would be similar. In areas where much of the landscape is defined by the absence of permafrost, runoff generation processes can be much different from areas where permafrost is nearly continuous. Sentence on runoff generation and discontinuous permafrost areas added at line 83.

Line 95: Why do you need to leverage a modeling framework to investigate changes in peak daily discharge? Would observational daily data not be a better method for this?

There is an extreme paucity of river discharge measurements at the mouths of North Slope rivers. It is clear that a better understanding of changes in the timing of peak discharge, at the coast, for this 196,000 km2 region, can only be obtained via advanced numerical modeling.

Lines 108-110: The study area is underlain by thick, continuous permafrost. This context needs to be explored in more depth in the discussion. The authors should describe how the flowpaths in this environment would differ from other studies in the literature. This has the potential to be a novel contribution and differentiate this work from other studies that it cites.

We have added detail and depth to the paragraph in the Summary and Discussion. Line 501. Additional language has been added through that section. Our focus is on mechanisms operating in regions of largely continuous permafrost.

Line 112: Provide a table of all observational data, agency responsible for collecting the data, locations of data collection, and period of data record.

We used observational data for SWE and river discharge. It is not clear that a Table would be helpful for just two data sets. We point the reader to the USGS data online. Details on the SWE data has been added in section 2 on Study Area, Data and Modeling. Starting at line 104.

Lines 157-159: I am not sure I understand this sentence. How do you compare modeled SWE against observed river discharge? These are very different parameters. Storage exerts a large control over how much snowmelt water is delivered to the stream network.

In that study end of season basin average SWE simulated by the PWBM was compared against discharge following snowmelt. In Arctic regions spring (or in general the ‘freshet’ period) discharge is largely controlled by the amount of snowpack water storage. In that study basin-averaged PWBM SWE prior to snowmelt explained a statistically significant fraction of interannual variability in spring (April – June) river discharge. We agree that storage potential plays an important role.

Lines 161-164: The authors either need to provide more information on how the model was parameterised and how it performs, or provide references to previous publications that have previously done this.

We cite four key papers in the Hydrological Modeling section. Lines 175-186. These are Rawlins et al. (2003); Rawlins et al. (2013); Yi et al. (2015); Yi et al. (2019). We also detail the new model updates in that section. Starting at line 187.

Lines 218-19: Can you provide more justification for why effective velocity was set to v= 0.175? This appears to be an important parameterisation of the model but there is very little justification given.

We selected the effective velocity based on the relatively flat topography of the North Slope. We find that the model is relatively insensitive to the choice of flow velocity in comparing with gauged data for the Kuparuk River. Indeed, applying the default flow velocity results in a bias in timing of peak discharge by -7.8 days early compared with gauge observations. In two additional simulations using a velocity 33% lower and 33% higher results in a bias of -5.4 and -9.0 days respectively. Many of the rivers in this region are shorter than the Kuparuk, so travel times are relatively short on the North Slope. It is no surprise that altering the flow velocity by 33% results in the timing of peak discharge shifting by only 1-2 days. The parametrization of flow velocity would have a much greater influence for long Arctic rivers like the Yukon, Mackenzie, and large Russian rivers. Accordingly we have added language at lines 295-299 with the result of sensitivity simulations.

Line 233: Are there any CALM sites or other field based observations from which the authors could compare their modeling results?

Simulated ALT is compared to estimates from a model developed at the Geophysical Institute. We previously described model validation for the sol thermal regime (Rawlins et al., 2013). Other recent studies using the PWBM (Yi et al., 2019) have compared estimates from the soil thermal model with observations. Point to grid cell comparisons for a few sparse locations should be viewed cautiously. We show that ALT calculated in the PWBM simulation with adjusted MERRA precipitation forcing closely matches the distribution simulated by the GIPL model, and that simulated ALT captures the expected spatial gradient across the region. Figure S2 and discussion starting at line 255.

Line 255: Why is only one basin used for validation?

Year-round discharge data at the coast is only available for the Kuparuk River. We have added a comparison with discharge for the Colville River for several months with observations. The data, however, are only available for the years 2002 onward, providing a nine-year climatology 2002-2010.

Line 263: Typo, “this occurs despite”

Word ‘occurs’ added.

Lines 267-268: Again, please display on daily timesteps and provide model performance evaluation.

The model is intrinsically daily time step. We feel that analysis of monthly runoff is sufficient for characterizing the hydrology at this time. Daily is simply expecting too much. See prior information in this review response.

Line 296: Please provide observational data to validate the modeled data.

The observational data for river discharge and SWE are provided freely to the research community. We have added detail of the SWE data. Line 123.

Line 309: Is surface runoff defined as overland flow?

Yes.

How are surface organics handled in the model?

Surface organics are parametrized using the Northern Circumpolar Soil Carbon Database (NCSCD). Lines 197-199 .

Many sites in the tundra have surface organics or peat layers where the porosity of near-surface soil is very close to 1, effectively eliminating overland flow due to the lack of resistance to flow exerted by the soil. In these situations would all runoff be subsurface? A better description of soil layers and modeling structure is needed to allow the reader to conceptualize the processes that are being explained.

Yes. The soil layers with high near-surface organic content have a poposity of 90%. This results in relatively high infiltration rates, and would, in most instances, lead to relatively higher amounts of subsurface runoff. Overland flow could still occur if surface (ponded) water is present and/or the infiltration capacity has been exceeded. Section 2.3 on the hydrological model is fairly detailed. Runoff occurs when water in a soil layer goes above field capacity. Line 172. The model is described in more detail in Rawlins et al. 2003, 2013 and Yi et al., 2015.

Lines 361-363: Provide references.

Several references added.

Lines 362-363: “materials exports to coastal zones ”typo

Corrected.

Lines 371-372: Why not test this and include in the current model?

A suite of model upgrades are currently being designed, tested, and implemented. Incorporating new upgrades is not feasible for the current study.

We look forward to describing upcoming model improvements in subsequent publications.

Lines 395-396: Which processes? The authors should be explicit about how hydrological processes are changing and cite field-based research to do so. For example, there have been quite a few relevant papers published from studies in northern Canada that are not referenced.

We appreciate the comment. Our manuscript is very explicit about how hydrological processes are changing, and we have cited field- and modeling-based research. We have modified the statement to indicate the changes in the Colville basin are greatest foothills regions. Line 473-474. We also added additional detail throughout the Summary and Discussion including in the paragraph starting at line 501. We now feel that the most relevant studies are cited.