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.

Reviewer 3

The paper aims at the analysis of baseline conditions and changes of hydrological elements at 42 catchments over the period 1981-2010. For this purpose, an updated version of the Pan-Arctic Water Balance Model (PWBM) was applied. The presented results indicate statistically significant increases in cold season discharge. A significant increase in the ratio of subsurface runoff to total runoff was found for 24 of 42 studied catchments. These changes correspond well to the increase of the active layer thickness due to higher air temperature and general climate warming.

The topic is potentially interesting for the hydrological society, especially the analyses of the non-stationarity of hydrologic processes in cold climate due to climate change.

However, I have many concerns and comments on the methodology that should be addressed. The most important is the lack of model validation. The presented results are mostly simulation-based and cannot be analysed without appropriate model validation. The results of validation for the Kuparuk catchment are not promising. I would not use this model for the assessment of changes in the timing of maximum flows. The error of maximum flows was estimated to 9 days while shifting in peak spring occur around 4.5 days earlier. The error is higher than the simulated changes.

Additional comparisons have been added. A new model validation section has been added. Line 253. We do not agree that validation for the Kuparuk catchment is not promising. Model simulated SWE exhibits a statistically significant correlation with measured SWE, and simulated runoff exhibits a statistically significant correlation with measured runoff (estimated via discharge). The error in runoff over the freshet period of May-June is a mere +0.3%. Yes, the error in timing of peak discharge is 8 days, nearly twice the report change in timing of maximum daily discharge for all watershed. There is uncertainty in that timing shift. That said, we believe the shift to earlier maximum discharge is being forced by spring warming that is approximately 5.4 F over the 30 year period examined. We are prepared to drop this aspect of the study. See also the responses to reviewer 1 and 2 for further information on model validation. Also the new statements and prior papers using PWBM cited at lines 175-186.

In my opinion, the model description is not sufficient. There is no information regarding solved equations, water balance, thermal balance. Is the energy balance included in the model? Is soil temperature modelled separately or is it included in the PWBM? There is no information regarding the model parameters (number of parameters, their meaning, how the model parameters were determined? By optimisation? Or just assumed? There is also lack of information on applied optimisation method). Some of the parameters were selected in a strange way without any explanation (for example changes in f parameter that is described in lines 207-209, the assumption of the effective velocity v=0.35m/s). Are the values of this parameter constant for the entire domain? Why is evaporation reduced to 1/3 of the potential ET rate?

The PWBM has been described in detail in several earlier peer-reviewed publications. These include Rawlins et al. (2003), Rawlins et al. (2013), Yi et al. (2015), and Yi et al. (2019). The first three publications each contain an Appendix which details the model processes and parameterizations. It is through these prior efforts that we have built a legacy of model descriptions. The soil thermal model was described in detail in Appendix A of Rawlins et al. (2013). We have added a sentence at line pointing the interested reader to the 4 key publications.

The results are largely insensitive to specification of flow velocity. Sensitivity test described at line 296. The PWBM is best described as an intermediate complexity model. For this study two new submodels were designed, tested, and implemented, involving a surface water layer and river flow routing. Three new parameterization were implemented: evaporation from the surface layer, runoff from the surface layer, and effective river flow velocity. These three parameters were established in these new model runs, in part, based on visually comparing with river discharge data. We find that the model is relatively insensitive to the choice of flow velocity 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. We are in the process of making additional improvements to the flow routing network and submodel. The evaporation rate of ⅓ of PET is set only for the ponded surface layer. Water in lakes and ponds evaporates at the PET rate. The value for surface ponding was chosen assuming the limiting effects of water sitting in local storage depressions. Lines 296 is the statement on the result of sensitivity simulations.

There is a lack of map and description of the study area. It is stated that 42 catchments are analysed, but only results for one or three catchments are presented. Are the results the same? Are there any differences in the results between catchments? How are these results summarised?

We have added a map of the study domain as new Figure 1. The domain is introduced beginning with the start of the paragraph at line 104. We have included a Table showing several defining elements of the 42 river basins of the simulated topological network. In the paper we described results for the three largest river basins which have some monitoring, and the full study regional. Results are not the same for all basins. We have characterized differences for the largest basins for select geophysical quantities, for example, Table 2 and Figure 6. We do not present summary results for all 42 watersheds and see no additional value in including analysis for smaller basins. The regional perspective, we feel, is more valuable. Runoff estimates along with the analyzed geophysical quantities will be made available to the research community for further analysis.

SWE simulations were evaluated using average values from observations collected at a 200x300 km domain. The PWBM was run at 25 km resolution. These are completely different scales. Large differences in SWE especially for 2004.

This concern would be valid if simulated SWE was compared against single ground-based SWE measurement. This is not the case. Ground-based SWE measurements were collected at multiple snow survey sites distributed across the entire Kuparuk River watershed so that snow measurements capture orographic effect from coastline to the mountains, Brooks Range. In this paper, SWE averaged from all snow survey sites in the Kuparuk River watershed was compared with the simulated SWE averaged over the same watershed. We clarified the compatibility of modeled/measured SWE in the text, see lines 122-126.

Why was a linear trend analysed? I suggest using a modified Mann-Kendall trend test for autocorrelated data for this purpose. The test should be applied separately for each catchment and then the results should be analysed I don't have any great advice for tidying up the manuscript, but basically I think it somehow needs to be streamlined, made easier to read and corrected. Some of the conclusions should be reconsidered, better highlighted and more concisely presented. The authors should be more clear about the meaning of statistical significance of their results and more careful when drawing conclusions from non-significant results. There are major errors or gaps in the paper but it could still become significant with major changes, revisions, and/or additional data.

In this study we evaluated time changes for end of season active-layer thickness, annual discharge, the fraction of subsurface runoff, cold season discharge, total terrestrial water storage (TWS), and its component storage amounts. For all but TWS, the trends are calculated from once-a-year values. We perform no statistical analysis on daily data. As described in Hirsh et al. "Statistical Treatment of Hydrologic Data" in Maidment (1993), while daily river discharge often exhibits auto-correlation, annual discharge does not. For the annual quantities free of autocorrelation we applied the standard Mann-Kendall non-parametric test to assess statistical significance through the associated p values and corrected Z values. The quantities comprising TWS have some memory, so we used the modified Mann-Kendall test for TWS and its component storages. Revised text at lines 244-252.

Maidment, D.R., 1993. *Handbook of hydrology* (Vol. 9780070, p. 397323). New York: McGraw-Hill.