Anonymous Referee #1

Model Validation:
Although improvements to this section have been made, there is still uncertainty as to how the model performs. Given the errors in timing of the spring freshet between observed and simulated runoff in the Kuparuk River, reviewers suggested including additional model validation. The authors have included the Colville River, but these data depict large errors in the volume of summer runoff (rationale is provided by the authors on lines 600-605). The authors have not included model evaluation metrics (although rationale for omitting them is given) leaving average error, correlation, and visual interpretation of the hydrographs as the only ways to assess model performance. I would like to see model performance (simulated vs observed) plotted for individual years instead of aggregated over large time periods (9 and 30 years for Colville and Kuparuk respectively). This would allow the reader to assess how the model performs in response to individual events. A suggestion would be a figure similar to Figure 6a and b in Krogh et al., (2017).

We have added a plot of simulated vs observed for each year with measured discharge data. New figure S6a-c.

Minor comments (note: line numbers are associated with the revised version that documents changes from previous version):

125-129: The additional information on observed snow survey data is appreciated. I suggest the authors include an additional sentence briefly outlining the basic methodology used in Stuefer et al., 2013 (e.g. SWE data were collected using depth-integrated density and snow depth measurements across 50 m snow survey transects at a ratio of 10:1 depth to density measurements.)

Sentences added: “The SWE data were collected using depth-integrated density and snow depth measurements across 50 m snow survey transects, with a 1 m sampling interval used along each L-shaped transect. Ten depth measurements were made for each snow depth core measurement.”

330-334: If simulated runoff is more conservative (lower variability), it does not follow that the model would over-estimate when runoff is high and underestimate when runoff is low. Would this not imply that the maximum simulated values should be higher than the observed maximum values and the lowest simulated values should be lower than the observed values?

Sentence corrected.

558-560: I would suggest including a sentence discussing possible subsurface flow through taliks in continuous permafrost as a possible routing mechanism arising from permafrost thaw.

Sentences added: “Permafrost thaw is enhancing deeper flowpaths and contributing to the development of taliks, unfrozen material formed by hydrothermal and thermal processes near and beneath the ground surface within permafrost which produce flowpaths that allow subsurface runoff to emerge as streamflow. The development of new taliks has been hypothesized as the primary mechanism contributing to increased groundwater storage across the Alaskan Arctic coastal plain (Muskett and Romanovsky, 2011).

Muskett, R. and Romanovsky, V., 2011. Alaskan permafrost groundwater storage changes derived from GRACE and ground measurements. Remote Sensing, 3(2), pp.378-397.

594-595: Can you briefly describe the characteristics of the basins showing the shifts to earlier discharge? If the date of maximum discharge across the entire basin is only marginally significant, and there are significant increases in some sub-basins, then it would follow that there would be no change in other sub-basins. It would also be worth including other relevant literature in this discussion (e.g. Shi et al., 2015 found a delay in spring discharge timing in the western Canadian Arctic). An expansion to this part of the discussion would better frame the results of this study within the context of the literature.

We find a significant trend ($p < 0.05$) to earlier maximum daily discharge for three of the 42 river basins. There are no common characteristics, but they are concentrated in the area from 148 ^{o}W to 152 ^{o}W longitude. We have added a new Figure S9 which shows the date of maximum discharge (Q) time series (1981-2010) for each of the three, which includes the large Sagavanirktok River.

Anonymous Referee #2

(1) I would ask the authors to consider putting in the calendar dates (e.g. 1 June) for Figure 3 instead of DOY. I think that DOY on the supplemental figures are fine but it would be clearer to use the calendar dates on the main figures of the paper.

Done. Updated figure 3a,b.

(2) The authors should make sure that their list of references follow the guidelines of The Cryosphere. Right now, some of the journal titles are short form and then other journals are written out in full titles.

All references now consistent with guidelines