Authors' response to the Reviewer 3

Black text: Reviewer's comment

Blue text: Authors' response

I commend the authors on a large amount of work, a well-written manuscript, and a thoughtful response to previous reviewers. I acknowledge how challenging it is to have a new reviewer come in mid-review and empathize with the authors on this. After reviewing the original and revised manuscripts as well as the response to reviewers. With that I've got some minor comments and some more major comments on figures.

Thank you for the review of our manuscript. We appreciate your constructive comments and suggestions. Please find our point-by-point response below. The line numbers in our responses are related to the revised version.

Abstract/Introduction:
I find these sections very well-written. In general, I think a clearer statement of why we need to know about runoff processes in the Antarctic Peninsula would increase readability. Specifically, after the first sentence in the abstract and around line 30. I think there's a degree of the importance that's implicit, but I think clearly stating why runoff processes matter would help your reader. I'm more familiar with why runoff processes matter in populated/agricultural settings in terms of food production, water supply, infrastructure, and hazards (drought/floods) so having a clear statement would be helpful to reorient myself.

We agree, and we added some more text both to the abstract (L11-12) and introduction (L40-43) sections highlighting the need to study hydrological processes in such remote areas, which, at the same time, belong to fragile environments that undergo major changes as a response to climate changes.

Methods:

I have concerns about the limitations in data availability for calibration, but recognize the limitations of field work and also commend the authors on a comprehensive response to previous questions. I have a few minor questions:

1. On line 153 is the single threshold temperature 0 deg C?

As stated on line 158 of the revised manuscript, "The TT is one of the calibrated model parameters." The median value resulting from 100 calibration runs was -0.21. We have added this information on line 466.

2. I'm still unclear about the role of PET as it appears to be estimated in the results, but I don't see any mention of how those data are estimated in methods. What are the key components of water vapor losses (i.e., is it sublimation given the relative aridity and temperature in the region)?

Thank you for the comment. The temperature-based method defined by Oudin et al. (2005) was used for the calculation of daily PET using observed air temperature measured at the Johann Gregor Mendel Station. The model then calculates actual evaporation from the input PET as a function of the simulated soil moisture. The snow routine does not account for either sublimation or evaporation from snow cover. We have reformulated the existing text or added a new text to the respective 2.3 Section (L151-152, 154 and 162-163).

Results

1. Are you assuming that any KGE value > 0 is “good” consistent with Knoben (line 195)? https://hess.copernicus.org/articles/23/4323/2019/ If so, I’d state that as I think it’s also somewhat common to assume that >0.5 is “good” for KGE.

We agree and have reworded lines 201-202 accordingly.

2. Since you define the WY as beginning on June 1, I would suggest just adopting a single year to describe each WY consistent with common practice outside the Antarctic to simplify labeling on figures.

We are aware that an indication of WY by a single year would make the text and figure captions easier to read (actually, we have considered it already while writing the original manuscript). However, we have chosen the two-year indication mostly because it is consistent with other glaciological studies. Besides, a single-year naming might raise questions about which period is meant since 1st June as a starting date is quite close to the middle of the calendar year (opposite to the typical WY starting date e.g. in the northern hemisphere.) For these reasons, we prefer to keep two-year naming to avoid confusion, although it results in longer text and figure captions.

3. Most units are presented in mm, but SWE is sometimes in cm. Not clear if there’s a reason for that, but flagging just in case.

We checked the text and found that the unit of cm is only used for snow depth. All SWE are always reported in mm.

4. I find all figures in the results to be very challenging to read/interpret given the overlap between daily and cumulative reporting and labeling. I know figures are such a challenge and that this was flagged in the previous review, but I think the readability of these figures is my largest issue with the manuscript.

We agree that some of our figures may require more time to look at, as they contain a lot of information. However, we feel that the information displayed in the figures is important to correctly interpret our results. Nevertheless, according to your suggestions, we reconsider the figures again and test several alternatives to improve the readability (see our detailed responses below).

a. Figure 4: Cyan is a very challenging color to read and it took me a very long time to understand how to interpret the columns for panel a versus the line. There is so much going on in this figure that grasping the relationships between WE and SWE and T is really hard. Specific thoughts:
i. at a minimum that there is greater space between panel a and b as I don’t think that will disrupt the readers’ ability to view relationships, but may make the figure less overwhelming.

We agree and have increased the space between the panels and changed the y-axis label in panel a) to make it clearer.

ii. Do you need to report annual WE since all other variables are daily or 10-day?

We believe that the annual change of WE is important because it represents essential information on glacier mass balance which is frequently used in other glaciological studies enabling the direct comparison of our results. In addition, the glaciological data used for model calibration were measured annually, so the change in annual WE also provides a direct reference to evaluate the model performance.

iii. It looks as though the figure has two grey bars on the right and left.

We were not able to find these bars in our original tiff figures. Perhaps, there might be some errors in the exported pdf of the original manuscript.

b. Figure 5: Again, there's simply too much packed into this figure for me to grasp the key points and the presentation of both daily and cumulative results is sort of challenging. Specific thoughts:
i. Could you separate the cumulative curves out and construct a panel b presenting all cumulative curves, including P with T plotted on top as is?

We tested your suggestion of creating the additional panel for the cumulative lines, but we found it less informative. In our opinion, the information from the daily values (areas) and the cumulative values (lines) complement each other well, so we prefer to keep them together.

ii. I think the filled in area graphs are really tough to read since Q_snow and Q_glacier overlap so much at the beginning of the water year.

The areas are stacked, so they do not overlap. We have added this important information to the figure caption. At the beginning of the water year, the runoff (if even occurs) is caused by snowmelt.

1. I don't fully understand how the cumulative curve for Q_glacier is so close to Q_rain when the area plots seem to imply that it's closer to Q_snow unless they're stacked in which case I still find the plots confusing.

As mentioned in our previous response, the areas are stacked (opposite to the lines). We have added this information to the figure caption to make it clearer.

c. Figure 6: I appreciate the attempt to include the contributing factors to Q within the Q_total, but I find this to be a challenging figure to read as well. This may be the best way to present this figure and I'm sure it took the authors a tremendous amount of work to get here, but it's still really hard to read. Specific thoughts:

i. think part of it relates to color and in many cases the lower bound of the error bars is very hard to determine.

We are aware, that the lower bounds of the whiskers may be difficult to see. This is because the related values are often small, even close to zero. In our opinion, it is not important to determine the exact value from the figure, as the overall variability of monthly runoff is evident. The colours were chosen to make them consistent with other figures (we tested different alternatives).

Nevertheless, we removed the horizontal lines from the figure which we hope makes the figure more readable.

ii. I would recommend adding discussion of uncertainty to your discussion around lines 255 since based on my read of this figure, uncertainty around total Q for peaks is close to 2x the estimated Q in some cases (e.g., Dec).

We agree, and we added some more text related to the monthly variability of runoff (lines 267-270).

iii. Also explaining why uncertainties in Q_snow are so large relative to other components would be helpful.

The whiskers in Fig. 6 do not represent uncertainty, but the inter-annual variability. Nevertheless, we agree that the variability in absolute numbers is largest for Qsnow. The possible explanation is provided in newly added lines 265-270.

d. Table 3: Can you explain bolded values in the caption. Also I don't know if this is an EGU formatting point, but usually I see table captions above tables.

The values in bold are the minimum and maximum values. We have added this information to the table caption. We did not find a specific rule regarding table captions placement for TC. However, we moved the caption above the table according to common praxis.

e. Figure 7: Same comments about cumulative and daily values and area graphs as above. It seems like daily values are most important give the text? Could you create a third panel comparing cumulative curves to highlight the role of Q_glacier?

We tested different alternatives but found the existing one as most informative. We also refer to our responses regarding Figure 5.

f. Figure 9: Why doesn't panel a have all the same boxes as panel b? Can you explain this in the text.

Please note that panel a) shows correlations of the annual values of the selected variables, while b) shows correlations for daily values of the selected variables. Therefore, the sum of positive temperatures (Tpositive) cannot be included in panel b) because, by definition, it represents a seasonal variable. As for global radiation (GR), the data were available only for the period from 1 January 2015 to 10 March 2018 thus there is not enough data to investigate annual correlations with other variables available for a longer period (see line 332). We have included this latter information in the figure caption. Additionally, we somewhat reworded the text related in lines 319-325 to provide more context (reflecting also your comment below).

i. I'm confused on Line 310 (the text implies three pairs, but the reporting is for only two? I'm not sure which pairing the reported stats correspond to. I would also say that it's slightly confusing to report a spearmans of 0.68 as 'strong,' but a spearmans of 0.63 as weaker. I would argue that both indicate moderate correlation with 0.82 indicating strong correlation.

Thank you, you are right there was a mistake in the text. Additionally, we have reformulated the whole text in lines 319-325 for clarity.

g. Table 4: Could this be combined with previous tables/figures somehow. It's quite a lot of have 4 tables and 10 figures.

We agree. We deleted the table and added its content directly to the text instead (lines 347-350).

i. I also don't follow line 321-323 completely. If Q_snow is the dominant driver of runoff (76% by my read) and Q_glacier drops as low as 1% per Section 3.4 couldn't it also be that smaller changes in Q_snow drive variability. That seems like at least one plausible explanation. I don't see the direct line between inter-annual variability and runoff based on results and it seems like a big leap to me to state that that's the primary driver based on coefficients of variation when the proportional contributions of those components are so different.

We have changed Fig. 10 and related interpretation in abstract (lines 20-22) results (lines 340-350), discussion (lines 413-418) and conclusions (line 485). Please see our response to your next comment for details.

h. Figure 10: I think this figure is really interesting, but also challenging to read. I'd suggest renaming WY to a single year to reduce text here. I'm also not immediately sure what the quadrants are supposed to represent other than deviations from x=0, y=0? I think I understand that circles are meant to emphasize positive anomalies in T in recent years, but honestly I find it confusing. I'd rather have the color scale be consistent and the y-axis vary since it really takes a while to see how these points diverge. Also I think relative anomaly needs to be more clearly explained as it is in the text above Figure 10. This figure takes a lot of thinking to digest and I think anything you can do to help your reader get to the main point sooner would be very helpful

Thank you for this detailed comment. Regarding the WY renaming we refer to our respective response above.

We agree that the original Fig. 10 was not fully clear since it mixed different types of information in individual panels (relative Q anomaly in panel (a) compared to relative anomalies of Q fractions in

other panels, which, by definition, already represent relative values). Therefore, we decided to show anomalies of runoff components in panels b), c), and d) to be consistent with panel a). Additionally, we have substantially rewritten the related text on lines 340-350 (results section) and 413-418 (discussion section).

5. I sincerely appreciate the author's transparency in their discussion of limitations. It makes me trust your science and results to be so upfront. With that said, I do think that it distracts from your key findings to have it take up such a prominent role in your discussion (e.g., that's what I get to first, so I immediately am thinking about your limitations rather than what you're showing). Could you move this lower in your discussion (at the end)?

We like your suggestion, thank you. We moved the data and modelling uncertainty subsection to the end of the discussion section.