In this document, we respond to the comments of reviewer 3 one by one. Whenever some entirely new text has been added to the manuscript, it has been added in italics and in red.

The proposed revised with and without track changes is added as a supplementary .pdf file.

Reviewer 3

Review of "Modelling the historical and future evolution of multiple ice masses in the western Tien Shan, Central Asia, using a 3D ice-flow model", Lander Van Tricht, Philippe Huybrechts The paper provides past and future glacier mass balance projections for 6 Central Asian ice masses, as well as estimates of changes in discharge. The methodology is appropriate and the results are useful for understanding the response of glaciers in the region to climate change. I believe that the paper is worth publishing after taking into account some comments listed below. As I am not an expert on dynamic flow models and climate model data, I rely on the other reviewers to review this section and to address the presented results on mass balance and data presentation in my review.
We would like to thank the reviewer for the insightful comments, which significantly contributed to enhancing the overall quality of the manuscript.

General comments

[RGC1.1] In general referencing could be improved. Certain relevant literature for this region has not been referenced in the introduction, study area and description of the climatic setting and drainage. For example, p.130 Kutuzov et al 2009 have already published annual precipitation totals for the region. For example, Sorg et al, 2012 is a review going back 10 years. There is much more recent literature showing the heterogeneous mass balance response for the region that is worth citing: Hoelzle et al., 2019; Brun et al., 2017 or 2019, Shean et al. 2020; Barandun et al., 2020 or 2021. Or, for example, l.490 also suitable is Kronenberg et al., 2016. In general, self-citation is quite prominent in the article.
Thank you for the suggestions. We added some of the references in the text.

[RGC1.2] Introduction paragraph 2, concerning the heterogeneity of the mass balance in the region: This paragraph is too superficial and not well substantiated by literature or own research. Many buzzwords are used without proper context and references are missing. This paragraph needs to be improved. In terms of the results of the paper, this paragraph should be restated in the discussion and better contextualised: do the results of this paper reflect the heterogeneity of glacier response or not? Are the ice masses chosen representative of the heterogeneity? And what role plays the uncertainty --> see also comments #6 & 7.
We agree with the reviewer, and because this was also mentioned by reviewer 2, we incorporated paragraph 2 into paragraph 1. We removed some sentences and added three references as suggested in RGC1.1. We think this has improved the readability of the introduction.

[RGC1.3] Location and climate: please restructure to show the geographical order (heterogeneous settings): West to East, lower elevation to higher elevation, increasing continentality etc. I would also prefer a more structured and systematic description and suggest a separate data section --> see comment #4
We concur with the reviewer's suggestion, and accordingly, we have restructured section 2.1 to provide a more systematic description of the six chosen ice masses.

[RGC1.4] For me a data section is missing. I would like to see a more structured description of the data. At the moment it is difficult to understand the origin of the different data sets used.

- It is also not entirely clear which data have been used for calibration, validation or both.
- There should also be a better description of what is measured data and what is modelled / interpolated data. As far as I understand, ice thickness measurements are only/mostly done on the frontal part of these glaciers?
- Where are the meteorological stations located? For what time period was the meteorological / precipitation data available for each station? Are the data filtered and gap filled and how representative are the data for each glacier? See also comment #5.

Golubin Glacier is located in the Western Tien Shan and has a significantly different regime (temperate vs polythermal), the glacier is located in a different climatological setting, much more precipitation than in the Central Tien Shan, lower elevation etc. The author announces to consider different climatological setting, but in my point of view, the authors do not pay sufficient attention to these differences in their calibration and interpretation in the discussion. For example, I am surprised that the basal sliding parameters are similar for Golubin (temperate glacier) and glaciers in the central Tien Shan (e.g. Bordu, Sary Tor, Gregoriev: polythermal glaciers). I am not convinced by the authors' suggestion that the sliding behaviour is regionally uniform because of the calibration results of similar basal sliding parameters. This needs some clarification. See also comment 6.

We agree with the reviewer that some of the issues that are mentioned were missing in the manuscript. We have made the following additions:

- Table 2: We have included an overview table that provides detailed information on the various meteorological and climatological datasets used in this study, including their respective time periods.

- In sections 4.3 and 4.4, we have improved the description of the calibration and validation procedures by explicitly specifying the use of ice thickness measurements and velocities for calibration, as well as geodetic mass balance and ice thickness in 2000 for validation.

- We have incorporated two sentences to clarify that the data underwent debiasing using overlapping periods. For further details regarding the locations of the meteorological stations used for Kumtor-Tien Shan and Chon-Kyzyl-Suu, we refer the readers to Van Tricht et al. (2021).

Regarding the Golubin glacier and its distinct climatic conditions, we have introduced a new paragraph in the discussion section. Furthermore, we have removed the sentence that previously mentioned regional uniformity.

[RGC1.5] Climatological settings: there is a large west to east gradient in continentality for the Tien Shan. Glacier regimes and the importance of summer precipitation events is changing from west to east (central part). This could be better described here.

We agree with the reviewer, and we reordered this section putting more emphasis on the precipitation gradient across the Tien Shan.

[RGC1.6] Meteorological input data: Precipitation data for the Tien Shan are critical and there is little high altitude precipitation data available. While the Tien Shan station may be representative of the glacier in the central Tien Shan, Ala Archa and Chon Kyzyl Suu are located much lower than Golubin and Kara-batkak glacier respectively. While a relatively detailed description of the extrapolation of the Ala Archa data is given for Golubin, it is completely lacking for Chon Kyzyl Su. Considering the less good result for Kara-batkak in calibration and validation, the representativeness and uncertainty of the precipitation data deserve at least a bit more discussion see also comment 8

The extrapolation procedure for the Kara-Batkak glacier is explained in more detail in Van Tricht et al. (2021b) and is not reiterated in the current manuscript. We acknowledge that this information was not clearly communicated, and thus we have reorganized the sentence order to address this. Additionally, we have introduced a new section in the discussion to provide further insights.

Regarding the representativeness of the Chon-Kyzyl-Suu (CKS) station, we hold a different perspective. Given the proximity of Kara-Batkak glacier to the CKS station and the strong correlation observed during a comparative analysis with an Automatic Weather Station (AWS) on the Kara-Batkak glacier over a short period, we consider the CKS station to be a suitable candidate for describing the precipitation regime on the Kara-Batkak glacier. We believe it offers superior representation compared to (regional) model outputs.

[RGC1.7] Discussion of limitation: I believe that the representativeness of the meteorological forcing, in particular the precipitation input over the past 100 and future 100 years, is not sufficient. Looking at Figure 2, the variability of the precipitation data, especially for Golubin and Chon Kyzyl Su, does not seem to be well represented. How does this affect the model results? Unfortunately, it is also not clear for which period measurements are available, but it also seems that the precipitation variability changes between 1960-1980 (depending on the station). Any explanation for this? Could it be due to the availability of direct measurements? How does this affect the results?

See our response in RGC1.4 We are of the opinion that the variability in precipitation data between 1960-1980 does not exert a significant influence on our results, especially not for the future projections. Indeed, the smaller variations appear to be inherent to the data utilised.

[RGC1.8] Surface mass balance model parameters: What does the hcrit parameter mean? It is not explained in the paper. If it is the height at which precipitation decreases, I think it is relatively low. Why is there no value for Ashu Tor and Gregoriev (and no second precipitation gradient)? Not all glaciers were modelled in Van Tricht et al 2021b. In addition to hcrit, the two precipitation gradients should also be explained.

The critical height, hcrit, corresponds to the elevation at which precipitation or accumulation diminishes due to moisture depletion and wind erosion. We have included this information in the text at the specified lines (308-310). Regarding Ashu-Tor and Grigoriev, we do not have observations leading us to conclude that two distinct gradients are necessary.

[RGC1.9] Comparison with observed ice thickness and surface velocity: It is not clear what was used for validation and calibration. I don't think the mismatch of velocity data for Kara-batkak is explained by the higher velocities in general. Golubin has similar velocities, but the results are much better. This glacier is not easy to model, very steep topography and a very thin glacier. Perhaps the model simply does not work as well for this glacier? Also looking at Figure 8, I am not sure how plausible the AAR evolution shown for Kara-batkak is. In general, the AAR seems to be overestimated for the present, later remaining fairly constant for the next 30 years and suddenly dropping to almost zero, I would expect a more gradual decrease? Any explanation for this?

See our answer in RGC1.4.

The larger RMSE for the Kara-Batkak glacier is primarily attributed to two data points that deviate significantly from the 1:1 line. These points are situated in the lower region of the accumulation area and closer to the glacier's margin. It is possible that our model underestimates the velocity in this particular area, maybe due to an underestimation of the ice thickness. However, we don't have observations to prove this.

The temporal evolution of the accumulation area ratio (AAR) for the Kara-Batkak glacier is highly influenced by its unique topography, characterized by distinct plateaus that are separated by ice falls. Consequently, the AAR can persist at a consistent level for an extended duration, even as the equilibrium line altitude (ELA) gradually increases. We also added this info in the text at lines 523-526:

"*The temporal evolution of the accumulation area ratio (AAR) for the Kara-Batkak glacier is highly influenced by its unique topography, characterised by distinct plateaus that are separated by ice falls. Consequently, the AAR can persist at a consistent level for an extended duration, even as the equilibrium line altitude (ELA) gradually rises.*"

[RGC1.10] Following comment 9, for Kara-Batkak, the AAR basically does not change at all from roughly today to 2100. Only the ablation area decreases. I am not sure how plausible this is, given the high uncertainty and reduced performance of the model for this glacier. Considering the geodetic mass balance shown later in the manuscript that is unfortunately highly inhomogeneous does not help to evaluate the model performance very well. Looking at the measured data provided by WGMS, the -0.5m w.e. seems more likely, which would then also challenge the large AAR. I would appreciate a discussion showing the possible limitations of the model, especially in the case of Kara-batkak and the results provided.

We refer to our answer in RGC1.9 for your concern about the AAR. With regard to your suggestion of a discussion, we added a part on lines X in the discussion.

[RGC1.11] Discussion, uncertainties: l 559: I am not sure how meaningful it is to compare precipitation importance for future modelling in comparison to uncertainties in calibration for Iceland and the Tien Shan. Tien Shan is a very large mountain range with heterogeneous topography and heterogeneous climatological settings that is largely under-measured at high altitude. You might consider De Kok et al., 2020 or Barandun and Pohl, 2023, which highlight the uncertainties in precipitation datasets for the region. I agree that the uncertainties introduced by model calibration, especially for region-wide applications, especially are large for the Tien Shan due to the lack of good validation datasets. However, without a proper uncertainty analysis, it is hard to say which is more important, the uncertainty introduced by the input or the calibration datasets. I would rather point out the unknowns that make it difficult to quantify the uncertainties in such a study. This is also true for the runoff estimate, which is highly dependent on the available precipitation data.

We added an entire section in 6.1 describing uncertainties and limitations of the meteorological data, as well as of the mass balance and ice flow model.

[RGC1.12] Discussion, uncertainties: Englacial temperature might be very different for Golubin and probably also for Kara-batkak in comparison to Gregoriev. This should also be discussed here
We added a paragraph in which we elaborate on the (sensitivity) of the englacial temperatures.

[RGC1.13] Comparison with geodetic mass balance: This chapter is the only independent validation, if I am not mistaken. l576: I believe that the problem is not only with the geodetic mass balance, but might also relate to the model performance for Kara-batkak. L.580: Large discrepancies in the region-wide geodetic estimates do not necessarily make it easy to use these values for validation. I would have more confidence in the glacier-specific, high-resolution estimates. For Kara-Batkak: I can only strongly recommend the authors to produce a high-resolution based geod. mass balance, which is so far missing for this glacier. I can understand if it is not in the scope of the study but it is unfortunate that this is not available yet, especially since this glacier was the focus of the authors' previous work and the lack of validation data might have been a limitation before. A number of high resolution sensors are now freely available (Pleiades? SPOT?) and I am convinced that this would be of great value as an independent validation dataset for various modelling studies carried out by the authors (maybe also in the future?).
We have reorganised this section, integrating the comparison with the geodetic mass balance as an independent validation, within section 4.4 (see our response in RGC1.2). Additionally, we concur with the reviewer's suggestion regarding the computation of a high-resolution geodetic mass balance for the Kara-Batkak glacier, as it would be highly valuable. However, the scope of this study does not allow for such calculations. Nonetheless, this should be considered for future investigations.

[RGC1.14] I miss a more detailed discussion about the heterogeneous response of the different glaciers in the different climatological settings as promised in the abstract and introduction.
A section is added (see 6.3).