Author Comments

Central Asia's spatiotemporal glacier response ambiguity due to data inconsistencies and regional simplifications.

Martina Barandun and Eric Pohl

We would like to thank the two reviewers for the positive feedback and the constructive comments on the submitted manuscript. We agree on most of the critique and provide in the following our suggestions on how to implement the changes to the manuscript for all the issues pointed out. As several issues were raised by both reviewers we post this combined response identically to both RCs. In cases where we disagree, we present our reasoning and arguments, which we would then implement in an edited version. In order to further improve the clarity, we would also make use of a professional proofreader if the proposed changes should satisfactorily resolve the current issues with the manuscript.

Below, we respond to all comments, and state how we plan to account for them in the revised version of the manuscript. The responses (normal font style) to the reviewers' comments are written directly after the reviewer comments (displayed in italic font style).

1 Reviewer #1

1.1 Major revision

Direct comparison of datasets: *Looking at how the different reanalysis datasets correlate with the glacier mass balance datasets is interesting but I feel that it adds quite some complexity in the interpretation. I am missing in this manuscript a separate comparison of all mass balance products together and all climatic data together. This would likely help understand the multiple linear regressions better.*

We will add a comparison between the different mass balance and reanalysis products. See also answer to comment of Reviewer 2. We follow the design by de Kok et al. (2020) for the comparison of reanalysis products, and add a map in the style of the distributed pie chart maps for the mass balance comparison.

Regarding a comparison with meteorological station data, we do not think this is helpful for the present manuscript. There are multiple studies that have investigated new data products for precipitation. There is, however, no real baseline to compare any product against because most of the few available meteorological stations are located in valleys and thus there is limited to no ground truthing at higher elevation. We can see the motivation for the reviewers' request. Reviewer two has pointed also out that the stated objectives should be adjusted to better match the presented results and discussion. We agree on this and adjust the objective description to more accurately match the presented work. Regarding the presentation of meteorological data, we think that the overview in the style of de Kok et al. (2020) is providing the important information on spatially different estimates by the products. The issue of missing ground truthing data is referenced in the text. We are only planning to provide the comparisons of the annual data for temperature and precipitation, instead of producing two (variables) times five multi-panel figures as in de Kok et al. (2020). For the main text, we are planning to include the precipitation comparison and refer to the Annex/Supplementary for the temperature comparison. That is because these figures occupy almost a page.

Regarding the mass balance data, we show an overview map in the same style as for all other presented maps with pie charts per 25 km by 25 km regions to make differences easier to spot with the eye. The strongest differences can be seen at the short period and we would include this figure in the main text. We would provide the map for the long period in the Supplementary.

Barandun et al. (2021) mass balance: *It strikes me that this mass balance model is based on ERA-interim data, which is also a reanalysis product similar to ERA5. It has been calibrated with snowlines and geodetic mass balance, but I suspect the results are likely influenced by the climate data as well. Have the authors compared the ERA-interim with the ERA5 to look for possible changes? This is likely to influence the regressions and result in some circularity – do the regressions agree or not with the results of Barandun et al. (2021)?*

This point has been raised by both reviewers and we understand their concern. However, we believe the use of the two datasets is justified in the way they have been used. The approach presented in Barandun et al. (2021) fortunately does not rely on precise meteorological input (which simply does not exist extensively in the high mountains of Central Asia). Barandun et al. (2018) showed that using transient snowlines for model calibration reduces the sensitivity to the meteorological input data. The sensitivity analysis in Barandun et al. (2018) uses an average temperature and precipitation dataset without any year-to-year variability. Obtained mass balances using the transient snowline approach still provided results with an RMSE of less than 0.2 m w.e. yr^{-1} in comparison with the direct measurements. This highlights that the snowline observations used for calibration, for each glacier and year individually, are primarily responsible for the modelled mass balances. For more details, we believe it is sufficient to refer to the previous work Barandun et al. (2018) that provides an in-depth analysis of the model sensitivity which the current manuscript builds upon.

In Barandun et al. (2018) ERA-interim was used becauseOrsolini et al. (2019) showed the superior performance of ERA-interim for mountainous regions like High Mountain Asia over ERA5, mainly due to the assimilation of *in situ* station data in ERA-interim that is omitted in ERA5. The products have thus significant difference in the generated precipitation fields that are independent from each other. This provides the opportunity for further analysis of possible correlations between the mass balance time series and the climatic drivers with ERA5 or other independent reanalysis datasets. Currently ERA5 is one of the most used reanalysis products due to its superior spatial resolution. However, often the above outlined shortcomings are not considered in correlation analyses. Despite the low model sensitivity we did not to use ERA-interim for the correlation analysis. We are planning to add a justification of the dataset in the manuscript, explaining also the insensitivity of the mass balance time series to the meteorological input data.

Uncertainties in mass balance data: *I am concerned about the use of yearly glacier mass balance, especially as it is not clear to me how representative of the actual mass balance. This is relatively well explained for the Barandun et al. (2021) dataset, but less for the Hugonnet et al. (2021) – are these actual geodetic measurements made on a yearly basis or are they extracted from the general trend? In either case, I expect the uncertainties on this data to be quite high relative to the glacier mass balance values, especially for glaciers in Central Asia, which tend to not lose mass very quickly. This seems to be confirmed by the fact that the Western regions have higher significant correlation frequency (also the ones with clearer mass balance signal). I am therefore wondering how valid it is to take yearly data and whether taking decadal trends would not be a better avenue to analyze the spatio-temporal patterns. At the very least a discussion about these uncertainties would be necessary to include in the manuscript.*

This is completely true. The idea behind using more then one mass balance time series is to show the uncertainties not just in the reanalysis datasets but also in the mass balance time series currently available. Hugonnet et al., also point out that the data should not really be used in an annual study – even though they provide the values. The annual values are mass conserving with respect to the long-term calculated geodetic values and thus the provided mean annual values are usable. Our reasoning for using the annual estimates anyways is that it is so far the only other mass balance dataset at annual resolution available in the region. We will point this out more clearly in the reworked manuscript.

Both mass balance time series use geodetic estimates calculated from the same dataset (ASTER) but use two different processing approaches, leading to very different glacier-specific mass balance estimates. This highlights the uncertainties in geodetic methods and how these relate into uncertainties further down the processing chain.

We believe the case study in Barandun et al. (2018) provides an extensive assessment of the model sensitivity and uncertainty and eventually provides a very conservative estimate. Barandun et al. (2021) adopted the mean uncertainties provided in Barandun et al. (2018) ($\pm 0.32\,\mathrm{m\,w.e.\,yr^{-1}}$) associated with the snowline-constrained mass balance model and combined it with the error estimate from the geodetic surveys. The estimated uncertainty of $\pm 0.37\,\mathrm{m\,w.e.\,yr^{-1}}$ does not assume independence of the errors from year to year. We believe that this is a fair estimate of uncertainties and openly show the limitation of such approaches for regional mass balance estimates. At the current stage, this is what is available for Central Asia.

We already present the standard deviation map to showcase that the two mass balance datasets are very different at an annual resolution. We additionally will provide a difference map of the mean annual mass balance to show also the inherent difference of the two mass balance times series at decadal scale. We do not want to judge the performance of the two mass balance estimates but just show their disagreements. Finally, we will highlight the uncertainty of the annual mass balance time series in the data section and add a statement in the discussion.

Downscaling of the reanalysis data: *I was surprised to see that the different reanalysis datasets had not been downscaled, especially considering that their respective resolution varies a lot. Is there not a risk that this will introduce elevation biases between subregions? Why has this not been considered in the study?*

We use the datasets intentionally in their original spatial resolution. Any downscaling method will be subject to another level of uncertainty and subjectivity. In this work, we do not want to answer the question of "How should precipitation (temperature) data be downscaled?". We agree that there should be an elevation bias. However, we believe that our two types of analyses and our choice of derived statistics can account for this. First, our temporal analysis investigates the correlation and thus any systematic bias in form of a simple offset (fixed lapse rate) does not affect the outcome. Second, in the spatial analysis, we incorporate statistics like trend, and standard deviation as explanatory variables, which both are in the same way not dependent on the systematic offset. A simple downscaling that we could potentially performed with our expertise would also not help to resolve this problem. A downscaling using a linear function from a coarse to a finer pixel size would keep the trend and variability of the original coarse pixel time-series.

By incorporating the CHELSA data, we provide a means to see how a state-of-the-art downscaling affects the correlation analysis. CHELSA is maybe the most advanced downscaling of ERA5, incorporating various meteorological fields to cover assumptions about precipitation distributions in complex high mountains. The most important message from this exercise is finally that a downscaling can have a severe impact on our interpretation.

Following the critical tenor of this paper, this shows the "danger" of applying downscaling without ground truth data as we create a reality that might fit "better" our story.

We will add a paragraph at the beginning of the data section to condense these thoughts and provide a justification for our data choice and our choice to not downscale the data.

Dependence of snow cover on temperature and precipitation: *Does the fact that snow cover is dependent on temperature and precipitation not affect the regressions?*

We strongly assume this, and we think that this is also reflected in our Figures 5 and 6, where we show that the incorporation of snow cover does not increase the number of significant correlations overall. Instead, a shift of important variables from precipitation and temperature to snow cover occurs. And we discuss this using the same argumentation (see Section 5.1, L.355 f.; L.399 ff.). We also mention the collinearity that results from this and is maybe the biggest concern of the comment. While not filtering the results for collinear occasions, the fact that there is this distinct shift from precipitation and temperature to snow cover seems a good proof that the collinearity is not biasing the overall outcome. The referenced paper by Vatcheva et al. (2016) (L. 233) states that the anticipated effect of collinearity results in masking out (removing) significant variables, rather than including more. The results can thus be understood as a more conservative identification of significant variables.

We will directly mention this example in the relevant paragraph.

Summarizing the scenarios in the discussion: *These potential scenarios are interesting but lengthy and difficult to follow for readers not familiar with the particularities of the region. Could these be synthetized in a figure and streamlined? A short summary of the main differences would also be welcome.*

We will provide a figure summarizing the two analyses in the form of a map. The map will show the most important drivers identified for the specific region for each analysis. We will shorten the text accordingly and synthesize instead what the figure shows more accessibly.

Moving forward: *This is very briefly mentioned in the abstract only (as far as I can tell). I was a bit frustrated that there were not more discussions on this – how should one then proceed to interpret the mass balance patterns? What possible other tools could be used, what additional data should be collected?*

We will add a more elaborated part on possible solutions to improve the current situation. This will cover in more detail instrumentation, advanced techniques for downscaling, holistic energy and mass conserving modelling approaches, as well as remote sensing validation data from independent data sources.

1.2 Minor revision

1.2.1 Abstract

Overall, I find that the abstract could be streamlined and the main message made clearer.

We will change the abstract to provide a clearer and better streamlined main message.

L3-4: 'Meteorological analysis, remote sensing products and novel approaches . . . all provide . . . '

We will adjust this accordingly.

L9: 'only . . . do we find'

We will adjust this accordingly.

L13-14: This feels like a repeat from above

We will shorten the statement.

L16-18: this part is barely mentioned in the discussion and could be developed more.

We agree with this comment. Please see our answer and suggested modification outlined in the major comment above.

Introduction

In general, the introduction is interesting and well written but I think it would benefit from some reorganization efforts and some streamlining to make the message clearer.

We will restructure the introduction as shown below in the specific comments.

L25-26: This sentence does not bring much and could be removed. Are the two references to Gerlitz et al., 2019, 2020 really needed here?

We will restructure the first paragraph of the introduction.

L32-34: This feels out of place.

We will rephrase the statement (see comment above).

L20-36: There are lots of ideas in this first paragraph but the logical links are missing. These need to be reorganized/structured.

We will reorganize this part (see suggestion two comments earlier).

L46: 'Barandun et al. (2021) have applied'

We will adjust this accordingly.

L44-48: I am not sure that many details are necessary here, especially as these are described in length in the methods.

We will shorten this part.

L59-60: Are these details really necessary here? A simple reference to Hugonnet et al. (2021) is likely enough.

We will shorten this as suggested.

L61: something wrong with the English at the end of this sentence

This will be rephrased.

L63: There are actually some region-wide debris thicknesses assessments. See Rounce et al. (2021) and McCarthy et al. (2022)

We will rephrase the statement and add the suggested reference.

A scale-dependent debris-cover mass balance relationship and limited observation-based distributed debris thickness estimates hamper the explanatory power of debris cover for region-wide glacier mass balance patterns in Central Asia

---

L67: references missing. A recent one could be the work by Glasser et al. (2022) - 10.1016/j.geomorph.2022.10829

We will add the reference.

L67-69: Not sure these details are necessary here.

We will shorten the statement.

L69: Have the authors considered avalanching as a possible morphological control? It feels like for some of the steeper ranges of the region that could actually play a significant role (Brun et al., 2019)

This is a good point. Unfortunately, it is not straightforward to accurately quantify the avalanche contribution to mass balance and hence its importance as morphological driver. We will add this point to the sections where we introduce the potential morphological drivers and point out in the method section that we do not integrate it in our analysis.

Data

L87: Study site should be plural. No need to capitalize the nouns in the titles.

We will adjust this accordingly.

L121: I don't think this acronym has been defined before in the main text.

We will introduce the acronym in the text.

L124: Simple reference to RGI v6.0 is enough here (text can be shortened)

We will adjust this accordingly.

L134: The Barandun et al. (2021) dataset also uses geodetic mass balance products to calibrate their model (second order calibration).

We will rephrase this statement.

L135: Suggest adding reference to Figure 1 here.

We will adjust this accordingly.

L134-159: Are all the details provided here really needed considering that these are already published approaches?

We agree and will shorten this paragraph and add a few lines on the uncertainties.

L141: (Dee et al., 2021) should come after 'data'.

We will adjust this accordingly.

L148: 'observational' is not really true for Barandun et al. (2021) MB, as it comes from modeling.

We will remove this statement due to the shortening of the paragraph.

L155: can you explain a bit better the sentence 'local and regional scale biases can persist'? A reference might be needed here.

We will reword this. The sentence is supposed to say that the approach is sometimes performing badly at the single glacier scale (Hugonnet et al., 2021; e.g. Extended Data Fig. 5d) and shows varying uncertainties at the regional scale (e.g. Hugonnet et al., 2021; e.g. Extended Data Fig. 5d).

L181: Could you also provide the resolution in km for consistency with the other datasets?

We will include this.

L205-209: I would recommend putting this in a separate subsection.

We will add a subsection.

L240-241: I struggle a bit with this k-mean clustering. Could you give a few more details.

We will add some more details on the k-mean clustering method with a standard reference.

Results

L276: A3 should come before A4 in the text.

Figure order will be changed.

Discussion

L357: remove the comma

We will adjust this accordingly.

L359-360: this makes sense as they are likely related

Yes, we agree. This is also discussed later in the manuscript, as mentioned in one of the earlier comments.

L434: snow cover decrease reported by

We will adjust this accordingly.

L489-490: missing parenthesis

We will adjust this accordingly.

Conclusion

L543-544: Use lower cases for Barandunetal and Hugonnetetal

We will adjust this accordingly.

Figures

The panels of the different figures need to be numbered.

We will adjust this accordingly.

Figure 1

- *Instead of a globe, a map of HMA would be sufficient and more informative here.*
- *Shouldn't the mass balance should be in m.w.e? Also for consistency with the text.*
- *In general I don't really like the term 'surface' mass balance here – at least for Hugonnet et al., 2021, these are geodetic mass balance measurements. I would suggest sticking to 'glacier' mass balance.*

We agree on all points and adjust labels in figures, and all occurrences in the text accordingly.

Figure 2

- *Why are the spatial and temporal analysis linked in the figure? Aren't they done independently from one another?*
- *It would be good to distinguish the data boxes and the methods boxes. Having the seasonal aggregation in the same box as the monthly meteorological time-series is confusing.*

They are independent and we are thankful for the suggestions and will change the figure according to the suggestions.

Figure 3

- *Could you specify what each cluster corresponds to in the figure? There are likely many ways of clustering this data, which criteria were used here and why?*

As mentioned briefly in the text this is based on a k-means clustering algorithm. It is an unsupervised classification, meaning that there is not a precise association of a class with e.g. a process. The k-means clustering is based on the standard deviation of the mass balance time series of Barandun et al. Therefore, classes represent glaciers with different variability. We will clarify this in the figure caption and in the text (see also comment above).

Figure 4,5 & 6
Most of the comments here hold for the next figures as well:

- *Suggest writing out SC, P, T*
- *Do these meteorological correlations relate then to trend, mean, STD?*
- *Which variables do the seasons refer to? This would need to be specified in the caption, and maybe even in a supplementary figure?*
- *Specify in the caption that 'short' and 'long' refer to the 2000-2014 and 2000-2018 periods.*
- *It needs to be specified also that this is an aggregation of all glaciers in the region.*

We will write out the abbreviations and increase their size in the figure. The correlations are indeed incorporating all derived statistics. As explained also in the downscaling section, this allows to find relationships where there is e.g. an altitudinal bias. It also allows us to use the meteorological data in both the temporal and in the spatial analysis. The latter requires a single value for any given glacier, whereas the former utilizes the time-series as it is. We will clarify this and the other points in the text and also in the figure captions.

Figure 7 & 8
These comments also hold for figure 8:

- *Stay consistent with acronyms in figures*
- *The explanatory diagram does not need to be repeated in every subplot if it stays the same*
- *Where are the subplots with and without snow cover? I cannot find the legend.*

We will remove all explanatory pie diagrams but one and change the labels to be consistent with the other figures. The subplot without the snow cover was moved to the Appendix and we missed changing the caption accordingly. The figures are in A6. We will adjust the figure caption.

Figure A4

- *Would there not be a way to weight the results of the right panel by the number of glaciers?*

The only thing we want to show with this figure is that we did not identify any different relationship between surge- and non-surge-type glaciers with mass balance among the different subregions. It would certainly be possible to weigh the mass balance estimates but the main aim is really to show that there is not a systematical higher or lower value for surge-type glaciers. We will clarify this in the caption.

2 Reviewer #2

2.1 Major revision

I think the authors need to restructure their objectives to be something they more clearly address:

Objective 1 seems to encompass two separate research questions about the drivers and physical processes themselves, and then the degree to which this is simply a result of which data are used to explore the problem.

Objective 2 is related to the latter part of the first objective in my opinion and additionally implies that the authors explicitly compared ground truth data for correlation analysis. I'm not convinced that the authors really test enough the limitations of the gridded data in this manuscript (so objective 1, part 1), or at least not from the data/figures presented.

We agree that the description of the objectives was not well chosen. As it was written, one would indeed expect a more detailed analysis, and more importantly, conclusive results on the interplay between climate and mass balance. Instead, we finally provided what the reviewer correctly points out as second part of Objective 1. The motivation of our work is, nevertheless, a better understanding of the climate-mass balance relationship. This is why we chose the unfortunate wording in L.70 ff. We will reword in particular this section to differentiate motivation and resulting objective. The objective will more clearly state that rather than revealing dominant factors, we will explore the relationships using the newly available datasets. This will then address the question if we can identify consistent relationships between mass balance and climatologic/morphologic drivers.

We will address the limitations critique by providing a pre-analysis of the datasets used. This will be done following the suggested graphs in the style of de Kok et al. (2020), and by comparing the mass balance datasets using the already introduced aggregated pie chart maps.

It is clear that different reanalysis datasets will produce a different correlation with the same mass balance product. While the authors perform a nice assessment of these expected differences, I'm *left wondering how comparable the forcings really are in the first place. The authors apply ERA5 (approx.. 36km), HAR30 (30km) and CHELSA (1km, using ERA5 to downscale). While the first two are closer in terms of spatial resolution and the processes they can represent at the surface, CHELSA represents a major effort to include finer scale topographical features, such as windward/leeward precipitation adjustments, even though it likely remains with significant seasonal biases. Accordingly, I would not expect to find a similar relationship to glacier mass balance (spatially or temporally) when comparing to ERA5 or HAR30 (e.g. L286-7, L343-4), which cannot represent precipitation dynamics at the scales relevant to glacier processes (e.g. L401).*

The authors have given some consideration to these limitations in their discussion section, but I'm not convinced that it is a fair test and fully supports the conclusions about an ambiguity to glacier response in the region. I think that redoing analysis with additional datasets would also not necessarily advance the findings of this manuscript and perhaps cloud the main messages. However, I do think that the authors need to address these discrepancies and perhaps even remove CHELSA from the analysis. More importantly, I would like to see some level of comparison between the different products for air temperature and precipitation estimates (especially if CHELSA remains included). What explains the presence or lack of correlations in given subregions, for example? Some supplementary figure(s) would be useful to that end.

A more detailed answer to this valid and important point regarding the spatial resolution can be found in an answer to reviewer 1 about downscaling. We argue that using the relatively coarse datasets alongside the fine resolution CHELSA dataset is a valuable and valid approach for what we want to show. Reviewer 1 was concerned about the elevation bias that results from not downscaling the data. The coarse data provide ultimately the same data for multiple glaciers. Here we argue twofold: 1) a systematic offset in the independent (meteorological) variable will have no impact on the outcome of the temporal analysis (significant relationship or not) because we perform a correlation analysis, where a determined significance is independent of a linearly scaled variable; 2) this holds true for the spatial analysis because the statistics derived from the meteorological time-series that serve as independent variable reflect temporal components (standard deviation, trend) and thus the same applies as for point (1).

However, this does not eliminate the problem that there are the same time-series or the same derived statistic for multiple glaciers encompassed by a single grid cell. Finding a relationship might be prevented if the glaciers within the extent of a grid cell show a stronger variability in mass balance than between different grid cells. This is, however, not the case for the entire study region, or the larger regional subsets, as we show in our study (e.g. Fig. 4,5,6). For smaller regional subsets, as shown particularly in Fig.5, this might have actually been the reason for the lack of significant correlations. However, because we also use CHELSA, we can see that this is not necessarily a spatial resolution problem. We argue that this can be related to changing variable importance that we do not cover in our study, e.g. a transition into more radiation-dependent processes (Sec. 5.1, L.367 f.). This assessment is supported by having CHELSA as an additional fine spatial resolution dataset in our analysis.

We do agree that having a better idea of how the different datasets compare to each other would provide a much-appreciated basis for any reader. The additional figure about meteorological datasets will show the differences and correlations now.

We will add the outlined considerations in the Data and Discussion sections.

Building on the previous point, I think the authors also need to address any potential circular issues related to the use of ERA-Interim for the derivation of the $MB_{Barandunetal}$ mass balance results and its comparison to the next generation of the ECMWF reanalysis.

This point has been raised by both reviewers and we understand their concern. However we believe the use of the two datasets is justified in the way they have been used. The approach presented in Barandun et al. (2021) fortunately does not rely on precise meteorological input (which simply do not exist extensively in the high mountains of Central Asia). Barandun et al. (2018) showed that using transient snowlines for model calibration reduces the sensitivity to the meteorological input data. The sensitivity analysis in Barandun et al. (2018) uses an average temperature and precipitation dataset without any year-to-year variability still provided results with an RMSE of less than 0.2 m w.e. yr^{-1} in comparison with the direct measurements. This highlights that the snowline observations used for calibration, for each glacier and year individually, are primarily responsible for the modelled mass balances. For more details, we believe it is sufficient to refer to the previous work Barandun et al. (2018) that provides an in-depth analysis of the model sensitivity which the current manuscript builds upon.

In Barandun et al. (2018) ERA-interim was used becauseOrsolini et al. (2019) showed the superior performance of ERA-interim for mountainous regions like High Mountain Asia over ERA5, mainly due to the assimilation of *in situ* station data in ERA-interim that is omitted in ERA5. The products have thus significant difference in the generated precipitation fields that are independent from each other. This provides the opportunity for further analysis of possible correlations between the mass balance time series and the climatic drivers with ERA5 or other independent reanalysis datasets. Currently ERA5 is one of the most used reanalysis products due to its finer spatial resolution and often shortcomings as outlines above are not considered in correlation analysis. Despite the low model sensitivity we did not use ERA-interim for the correlation analysis. We will add a justification of the dataset in the manuscript, explaining also the insensitivity of the mass balance time series to the meteorological input data.

The temporal analyses produce clearer and more interpretable plots for this manuscript (which I liked), but I have concerns about the robustness of these results given their short temporal scale (14 years for correlations using HAR30) and high uncertainties for mass balances given annual data series. While the ideal minimum sample size required can vary based on the type of observations, it has generally been considered that sample sizes less than 30 can produce imprecise correlation estimates. I'm curious why the authors did not consider a few longer term reanalysis/gridded products of similar spatial resolution (e.g. HARv2, ERA5Land,WRF9km) that all provide data since 1980 at least. I guess there are also temporal limitations for the mass balance data period, but this should at least be discussed more clearly.

This is a good point and we would have liked to use longer time-series. We agree fully that it would certainly make the analyses more robust. Our work is unfortunately restricted by the length of the mass balance estimates rather than the meteorological datasets. But, we were also keen to use the HAR version 1 dataset as it is one of the best performing datasets in the region that has been shown to actually resolve the discharge variability over the greater Pamir domain (Pohl et al., 2017), and also over the Tibetan plateau (Biskop et al., 2016).

Comparison between HAR version 1 and version 2 (not shown here) show that these datasets are very different. This is not surprising as version 2 uses ERA5 as input data. We certainly intend to raise some awareness about what looks like a tendency to more and more use ERA5 and ERA5-land directly - or as basis for further downscaling (HAR v2, CHELSA)- because it is so convenient with its long temporal coverage and good spatial resolution. With our choice of datasets we were hoping to provide simultaneously two critical views on the climate-glacier interaction topic. As the comparison between CHELSA and ERA5 shows, the downscaling has a significant impact on the obtained results. By incorporating HAR version 1, with a similar spatial resolution as ERA5, we also have a comparison between different forcing datasets for the downscaling. Without that, the main deductible conclusion would be that downscaling affects the outcome.

We will add an additional paragraph on our justification for using the indeed rather short HAR version 1.

Plotting such large datasets can be challenging and I think the authors have done a great job in many respects. I think there are several areas where they might still be improved for clarity and to help the message of the manuscript. I have given some specific comments below.

Thank you for the comment. We will address the specific comments below and refer to the answers given to each specific comment.

As mentioned, the work is generally well written. However, there are several instances where the wording is unclear or with grammatical errors. I have tried to provide some examples of this below, but the authors should carefully check the entire manuscript at the next submission.

We went through the specific comments and will include the changes in the updated manuscript. We have also decided to ask for a professional proofreader to eliminate unclear wording. ———

2.2 Specific Comments

L90+ : I think that the authors provide some very valuable and interesting information about the climatic setting in this paragraph, but I do not feel that the discussion reflects enough upon these interesting variations. Building upon my main point about comparisons of the gridded datasets, the authors should present spatial maps of mean (summed) temperature (precipitation) from different products, or better yet, the spatial trends of some of these variables (where are they statistically significant for ERA5/HAR30?). Exact comparisons can be challenging due to differences in spatial scale, but the authors could consider spatially binning the grids as presented, for example, in de Kok et al. (2020).

As mentioned in other comments, we plan to provide such a figure.

L185-188: The authors state that snow depth over-estimation in ERA5 renders use of radiation variables "problematic", and rely upon only temperature and precipitation data. However, there are well reported cold biases for air temperature at high elevation regions (such as those in this manuscript) which also result from this albedo-effect (e.g. Wang et al. 2020).

For any type of correlation analysis, a systematic and linearly scaled over-/underestimation does not pose a problem. Our main reason for the choice of the two main variables temperature and precipitation is, however, that they are widely available. For reanalysis with meteorological station data assimilation these two variables should also be better constrained than others, which are rarely or not at all measured at meteorological stations or not retrieved by other means.

We will clarify this.

———

L192: Given that the authors test the latest products from ECMWF (ERA5) and CHELSA, I'm surprised that they did not also consider the HARv2 product which is openly accessible and available back to 1980 (thus eliminating the shorter temporal focus and affecting the strength of correlations). I think a line should be added somewhere to justify the use of these three datasets specifically and why they are considered comparable, especially given the points from my main comment above.

This point is largely answered in a previous comments. We will address this in the data section with more rigour.

———

L205: I'm not convinced that the MOD10CM product at a 5km resolution does much to aid the analysis presented here. Given its spatial resolution, can it reasonably represent anything meaningful at the scale of the glaciers? Please add some additional information

Even though the resolution of MOD10CM is not matching the individual glacier scale it provides a robust small (10km) scale snow cover evolution time-series. We chose this product for the robustness (averaging out falsely classified pixels and cloud cover issues) with the intention to have a general overview of snow patterns. Similar to the coarse (30km) ERA5 and HAR products, we use either the time-series directly in the temporal analysis or the statistics that provide information derived from the temporal domain (standard deviation, trend, and also snow cover change per time step) for the spatial analysis. Many times the high spatial resolution MODIS snow cover products provide either time-invariant signals at the highest resolution (250 m and 500 m), or slightly varying values suggesting changes in winter where these changes are more likely to reflect measurement or classification uncertainties than actual changes.

In our analysis we use snow cover mainly to check the results for plausibility, i.e. if we can substitute an interplay of precipitation and temperature (leading to snow cover) with an actual estimate for snow cover. As such, the resolution closer to the coarse meteorological datasets is likely better suitable than a fine resolution dataset that would capture small scale processes that certainly are not represented in the meteorological datasets.

An advantage of the coarse resolution is certainly that changes over time are more clearly captured. This is because the averaging over a larger area will detect a decline or increase in snow cover due to the more variable evolution at lower elevations. With other words, we obtain a better signal-to-variability ratio. In the end this is a trade-off. But we think that the use is justified for the presented work and our idea of testing the obtained variable importance for plausibility.

We will incorporate these thoughts into the Data and Discussion sections.

L215-216 (and 243-245): I do not understand clearly how the authors relate the spatial means of mass balance with trends of meteorological variables. From the trend you will have just one value per gridded pixel, correct? So the authors combine all glaciers in a given binned region (circles in Figure 1) or just the entire region, and regress all the trends in a given variable (say temperature) at the corresponding pixel of each glacier against the mean or STD of mass balance at that glacier? I believe the authors need to more clearly explain this. Unfortunately, Figure 2 does not help much with this interpretation.

We will explain this in more detail in addition to the proposed changes from the previous comments about the coarse reanalysis resolution. We do not calculate a mass balance mean but instead use each glacier as one data point. This will - explained in the other comments - indeed lead to often multiple glaciers sharing the same meteorological data point. Regarding Figure 2, we do not have a direct solution for how to incorporate these ideas in the figure but we think that with our argumentation of why we think that our approach is sound (explanations in previous comments), we can provide a much clearer discussion on the this issue.

L221-222: Please re-write this sentence to improve syntax. It makes sense, but is not well written.

We will rewrite the statement.

L225-226: This is also a little unclear from the writing. Do the authors mean that they do not mix variables from different products in their analyses (e.g. temperature from ERA5 and precipitation from CHELSA)?

Yes, exactly. We will clarify this in the text.

L242: Please redefine here what is meant by these 'hotspots'. I note that this comes from the author's previous published work, but it would be better to clarify explicitly for the reader what is meant by a hotspot of mass balance variability. I interpret that as the areas in Figure 1 where STD is greatest, for example.

We will introduce what we mean with "hot spots" in the introduction.

L253: The authors consider the ambiguity of the glacier response here as being largely due to different meteorological or mass balance datasets considered, but what about the uncertainties stemming from the annual-series of geodetic mass balances themselves, which are, to the authors own admission (e.g. L156), with less accuracy, and I assume, less precise? Again, it would be ideal to be able to assess the comparison of the different gridded meteorological and mass balance products independently to identify where the largest differences in the datasets lie, not just the correlations/significance.

In accordance with the comment of reviewer #1, we will outline the differences between the annual time series and highlight the usefulness of using the two different mass balance time series. We add also a statement on the uncertainties related to both mass balance time series. For a detailed answer on the uncertainties and proposed changes in the manuscript please also refer to the answer to reviewer #1 (major revision comment #2).

L264: I think the authors should attempt to signpost some of the key findings a little better in the figures and simplify some of the main figures where possible (see specific figure comments below). I find it a little hard to navigate from text to figure to follow the story-line.

We will improve the figures of the paper and refer here to the answer of to specific comments.

L344: The authors do not consider all available reanalysis datasets or atmospheric data for their analysis. Please rephrase this.

We will rephrase the statement.

L346: "at the local scale..." Please check the text for many of these small grammatical issues or minor mistakes.

We will adjust this accordingly.

L346-8: This sentence does not make sense to me. Please rephrase to make your point clear.

We will adjust the entire paragraph to account also for the next three comments.

L350: but they do not prevent this analysis here, the authors have an annual mass balance value for temporal analysis. I think this first paragraph needs to be restructured slightly to highlight what the authors are able to show, which is normally not considered/available.

This comment is not fully clear to us. The temporal analysis provides us with the means to avoid any arbitrary aggregation and instead look at each glacier individually. We will rephrase the entire paragraph, so that it becomes more clear to the reader (see comment above).

L369-379: I think the authors should rewrite this segment as their point is not particularly clear upon reading it. Do the authors mean that their HiMAP regions encompass multiple sub-climatic regions and associated atmospheric processes which therefore limits the usefulness of these defined regions for such analyses?

There are various points addressed that encompass spatial and temporal aspects. But yes, this is often the problem when using standard regional divisions for correlation analysis. Apart from this, the heterogeneity of glacier response to changing atmospheric settings that again leads to changes in sensitivity can create a very complex picture within one subregion. This is then difficult to relate to the reanalysis datasets. We will try to be clearer with our statement and rewrite and restructure the section.

L382: Please define what "rotation" of mass balance gradients means.

We will explain this better.

L385: reword "already slight uncertainties..." It does not make sense. Small uncertainties? What estimates? Which variables?

We will rephrase the statement

L387: How "invisible"? Are simply not resolved by the coarse reanalysis grid scale?

This is indeed what we meant to say. We will clarify.

L388-9: Not clear. Do the authors suggest that coarse grid scale products can capture changes in these regions because there is less sub-grid variability in topography and its associated meteorological complexity?

We meant that due to the different mass balance sensitivities of glaciers in continental or more maritime settings, their reaction to a change will be different in amplitude. For glacier at high altitudes and more continental climate regimes, with lower mass balance sensitivity, larger changes are needed for a glacier response. These more important changes are probably easier to capture in reanalysis datasets than small scale changes and might explain the higher number of significant correlations found for more continental settings.

However for glaciers in sub-continental regions, where sensitivity is highest, small variations at various spatial scales over a glacier might remain unresolved in the coarsely resolved reanalysis products, leading to the lower correlation. This adds to the heterogeneous mass balance response and renders finding direct correlations with climatic drivers difficult.

We will try to be more clear and better structured with the argumentation within the entire paragraph (see also changes above).

L392: What do the authors mean with "outer orogene margins"?

We will adjust this. The "outer" was not necessary.

L399: "remains unclear..."

We will adjust this accordingly.

L406: please provide examples of which physical snow properties that the authors refer to and what variables missing from the analysis might influence the results.

We will adjust this accordingly.

L410: "extent".

We will adjust this accordingly.

L434: Citation requires parenthesis, check formatting.

We will adjust this accordingly.

L489: Check formatting.

We will adjust this accordingly.

L497-8: Related to the general comments, 14 years does not produce an ideal correlation period. Some discussion on this is needed.

We will add more discussion on this as explained also in an earlier comment.

LL503: The authors cannot evaluate the biases and patterns of reanalysis datasets due to lack of ground data, but should still highlight how the products relate to each other. Although long term in situ observations are rare, and I understand that it is not the goal of the manuscript to identify the 'best' product to use, but I would be interested to see how the different products compare to even short term measurements from sporadic AWS measurements familiar to the authors from previous work (e.g. Abramov Glacier , Golubin Glacier).

We will include the dataset comparison. The comparison with AWS or meteorological station data in general comes with a lot of culprits that we do not want to address in this paper. At least for the HAR dataset there is a comparison available in Pohl et al. (2015). For Abramov there is already the problem that precipitation is not measured by the modern AWS (Kronenberg et al., 2021) and data gaps are very common at other sites as well. We think that trying to find a suitable method for a sound comparison would - although interesting - not change the conclusions in the present work. We hope this is acceptable for the reviewer. We will link the two references mentioned before and look for and include references from other sites where there has been an assessment made. Either way, a good or a bad correlation should only locally increase our trust in any of the products but not over the whole study domain.

Figures

Figure 1: The inset map in the upper right panel should ideally be of High Mountain Asia to provide a better spatial context as to where the study site is. If able, the authors should attempt to neaten the location and intersect of the legend numbering as there is overlap in places.

We will include a HMA inset map and neaten the legend.

Figure 4: Please clarify in the caption how variable importance is defined and why this is given as a relative freq. in the plot labels. Also ideally provide indices (a,b,c etc) to aid navigation for the reader from the main text. I don't find a 'short' analysis period as meaningfully different to the long one for the lower panels. Such morphological characteristics (e.g. debris) won't greatly change over 4 years. Consider removing it to streamline and simplify the figure. In general I find it very hard to interpret the information given in these figures (4,5,6, A1, A2). Any attempt to simplify it or remove parts would aid interpretation.

We use the term "Relative Importance" which at this point is only explained in the methods. We will add this information in the caption with a simpler wording and reference to the main text to aid understanding the graphs.

Figure 5: I think this figure has far too much information and becomes difficult to interpret. Please consider streamlining the most crucial information where possible, perhaps just showing

the short periods, or those with MOD10. For upper panels, one has to struggle to align the x axis label and interpret what is the take-home information. legends should be increased in size, but could easily be kept for just one panel for the whole left and right side.

This is a difficult point. We agree that this is a lot of information provided in the figure. However, we also think it is well suited to show the differences per sub-regions, the changing variable importance with and without snow cover, and between the two considered time periods. As we take reference to these figures, we would work on improving the readability by adding helping lines and by highlighting bars that we reference later on on the text.

Figure 6: Similar considerations to the above.

We will address this in the same way as pointed out for the previous comment.

Figure 7: These plots are more clear than the previous ones. Perhaps the authors can re-structure the panels to be 3 rows and two columns, so that the size on the page can be enhanced and the reader can more easily interpret the small circles. Why using 'M' in front of the meteorological datasets? I think it would be clearer without. I do not understand from the figure caption the "with or without snow cover..." as all show scf as a variable importance coloured on the map. Please clarify or correct this. Please use indices for the subplots to help the reader. For Figure 8 as well.

We will homogenize the different abbreviations and write out the data combinations (M for MODIS) to avoid this confusion. We will try to optimize the figure readability by the proposed 3 x 2 format to increase the individual panel sizes.

References

Barandun, M., Huss, M., Usubaliev, R., Azisov, E., Berthier, E., Kääb, A., Bolch, T., and Hoelzle, M.: Multi-decadal mass balance series of three Kyrgyz glaciers inferred from modelling constrained with repeated snow line observations, The Cryosphere, 12, 1899–1919, 2018.

Barandun, M., Pohl, E., Naegeli, K., McNabb, R., Huss, M., Berthier, E., Saks, T., and Hoelzle, M.: Hot spots of glacier mass balance variability in Central Asia, Geophysical research letters, 48, e2020GL092 084, 2021.

Biskop, S., Maussion, F., Krause, P., and Fink, M.: Differences in the water-balance components of four lakes in the southern-central Tibetan Plateau, Hydrology and Earth System Sciences, 20, 209–225, 2016.

de Kok, R. J., Kraaijenbrink, P. D., Tuinenburg, O. A., Bonekamp, P. N., and Immerzeel, W. W.: Towards understanding the pattern of glacier mass balances in High Mountain Asia using regional climatic modelling, The Cryosphere, 14, 3215–3234, 2020.

Kronenberg, M., Machguth, H., Eichler, A., Schwikowski, M., and Hoelzle, M.: Comparison of historical and recent accumulation rates on Abramov Glacier, Pamir Alay, Journal of Glaciology, 67, 253–268, 2021.

Orsolini, Y., Wegmann, M., Dutra, E., Liu, B., Balsamo, G., Yang, K., de Rosnay, P., Zhu, C., Wang, W., Senan, R., et al.: Evaluation of snow depth and snow cover over the Tibetan Plateau in global reanalyses using in situ and satellite remote sensing observations, The Cryosphere, 13, 2221–2239, 2019.

Pohl, E., Knoche, M., Gloaguen, R., Andermann, C., and Krause, P.: Sensitivity analysis and implications for surface processes from a hydrological modelling approach in the Gunt catchment, high Pamir Mountains, Earth Surface Dynamics, 3, 333–362, https://doi.org/10.5194/esurf-3-333-2015, 2015.

Pohl, E., Gloaguen, R., Andermann, C., and Knoche, M.: Glacier melt buffers river runoff in the Pamir Mountains, Water Resources Research, 53, 2467–2489, 2017.

Vatcheva, K. P., Lee, M., McCormick, J. B., and Rahbar, M. H.: Multicollinearity in regression analyses conducted in epidemiologic studies, Epidemiology (Sunnyvale, Calif.), 6, 2016.