Second round of review of *"Annual mass changes for each glacier in the world from 1976 to 2024"* by *Ines Dussaillant et al.*

Earth System Science Data: essd-2024-323

We are very thankful for the second round of minor reviews. We hereby respond point by point to all specific concerns raised by both reviewers. We also profit to notify the editor and the two reviewers that we have recently ingested to the FoG database the latest annual mass balance observations for reference and benchmark glaciers during the hydrological year 2024. This means we were able to update all our results to include the year 2024 before publication. The tittle, total resulting numbers, all result figures and the discussion of the result have been updated accordingly.

Report reviewer #2:

This revised version has been significantly improved compared to the initial manuscript. It reads much better. Most of my comments on the initial version have been taken into account, providing the previously lacking information. It remains that sections 2.2.1, 2.2.2, and 2.2.3 are still hard to read for a non expert (like me, Reviewer 2). Several variables are still not defined. To improve understanding of the proposed methodology, I suggest that at the beginning of each of the three subsections, the authors summarize by a few sentences what they intend to do and what is the expected outcome.

I note that Reviewer 1 made substantial comments in the 'Method' section and that the authors responded amply. Thus being unable to judge the details of the calculations, I will rest on Reviewer 1's evaluation. Concerning the rest of the manuscript, I think that significant clarification has been made by the authors. Thus except for the method section, the revised manuscript looks to me now publishable.

We thank reviewer #2 for providing a wider perspective on our work and helping the revised manuscript to be better understood by a broader and non-specialized audience. The method sections 2.2.1, 2.2.2 and 2.2.3 are already introduced in the second paragraph of the methods section with a brief summary of what is done in each step and what is the main outcome. We note that these sentences were previously referring to the panels of Fig. 2 but not to the specific methodological sections to which they correspond. We now added clearer references to these specific sections, however we do not think that it is needed to repeat these summary sentences at the beginning of each section.

L151-158: Our processing algorithm is summarized in three key steps, described in the following sections and in Fig. 2. First, focusing on a specific glacier in the RGI-6.0 inventory, we estimate the detrended temporal variability of annual mass change for the glacier, referred here as the glacier mean annual mass-balance anomaly, using the interannual variability of nearby glaciological time series (Section 2.2.1 and Fig. 2a). Secondly, we calibrate the mean annual mass-balance anomaly to the long-term trends from the geodetic sample available for the respective glacier (Section 2.2.2 and Fig. 2b). Third, we integrate all these calibrated time series into a single, area-weighted average, *producing a data-fused annual mass change time series unique for every individual glacier (Section 2.2.3 and Fig. 2c). All given uncertainties in the tables, figures, main text and reported in the dataset files are at the one σ level (68% confidence interval), unless stated otherwise.*

Report reviewer #1:

Dear Ines Dussaillant and co-authors, dear Editor,

The manuscript has improved significantly since the previous submission. I appreciate the consider able effort the authors have invested in addressing my comments and refining the methods, manuscript structure, the story telling, and data description. I would like to thank you for your patience in thoroughly reviewing and carefully considering each of my suggestions. I am glad to see that some of my proposed ideas may have contributed to making the dataset more robust and to articulating further the strengths and limitations of it. I must also apologize for any misunderstandings or oversights on my part during the first review process. I have gained valuable insights into the topic through this exchange and appreciate the opportunity to engage with this work.

I am particularly pleased with the clarified use of "geostatistical modeling," the enhancements made to the methods description, and the updated mass balance (MB) anomaly selection via kriging spatial interpolation. The addition of a leave-block-out cross-validation scheme is another commendable improvement. However, I am no expert in "kriging spatial interpolation" and can not very well judge this new methodological part of the manuscript.

Your research addresses a complex topic, and I believe this paper lays a solid foundation for further advances in this direction. At this stage, I don't ask for any major changes. Nonetheless, I have identified a few aspects that may need to be addressed before the manuscript is ready for publication. I apologize for the inconvenience, but I trust these adjustments will further improve your work.

1 General Comments

1.1 Description of error bars/uncertainties

In the response to my comment you wrote:
"Reported uncertainties in the text correspond to 2σ = 95% confidence. Therefore, the term "uncertainty" corresponds to 1σ when describing equations and 2σ for reported values."

I truly believe it would be important to also add that information into the manuscript. If someone wants to use the total error that you provide in the dataset or e.g. in the abstract as input for their model, how should they interpret your provided error? Is it at "one or two s.e.m (standard error of the mean)" or is this impossible to say as you don't know about the uncertainty levels of your data contributors? Clarifying this would be very important for the data users. From what I have seen the uncertainties are explained only once at the very end at the caption of Fig. 8 but not in the text or any other figure, or did I miss something? I would suggest that you describe the meaning of the uncertainty in the manuscript's methods by saying 1 sth. like: "All given uncertainties in the tables, figures and main text are at the 2σ level (around 95% confidence interval)" [or, if this is true: "... 2 s.e.m."]. I am not sure if this is correct for Fig. 2, L373, Fig. 4, Fig. 10, Table 4, Table 5, Table 8? Though, I guess, Fig. 7 is in "1σ"? In that case, it may be better to describe it briefly at every caption and once in the text?

It seems that there is still some confusion with the use of sometimes one σ and sometimes two σ. For consistency and to avoid further confusions on users and readers, we finally decide to keep all uncertainties at the one σ level across all figures, tables and reported dataset values, unless stated otherwise. Users can decide to consider their preferred confidence intervals when using the dataset. We checked all figures and tables to make sure that uncertainties are at the one σ level.

We added the following statement at the beginning of the method section:

L156-157: All given uncertainties in the tables, figures, main text and reported in the dataset files are at the one σ level (68% confidence interval), unless stated otherwise.

I would also add the error description to the data description table where you mention the "error". I would even mention it already once in the abstract (L27, though this may be a question of "taste", so your choice).

Added to the data description table at the data file names section. Now files corresponding to uncertainties show that they express the one σ level. We do not think it is necessary to add it in the abstract. This has been properly clarified in the method section and across the manuscript.

1.2 Interpolating with kriging

L185–L215 (including Fig. 3): This is a very interesting new way to select the glacier annual MB anoma lies and to assess associated uncertainties.

I have the following comments on this new approach. Some of those are just "very minor comments", but I thought it is better to gather all "kriging-related" comments here.

- You write that the predicted kriging uncertainty grows with distance. How do the uncertainties increase with the distance? Is it somehow possible to visualise that within/beside Figure 3? Or possible to briefly explain further?

How uncertainties increase with distance was already described in the text:

L192-194: "the predicted kriging uncertainty $\sigma_{\overline{\beta}_{g,Y}}$ grows with distance, from the measurement error of the inputs σ_{B_a} at close distances from a measured glacier, to the signal variability (spread of β_Y) at distances far away from any measured glacier, where the prediction is more poorly constrained."

Further explanations of the method have been added in the revised manuscript as described in the following answers.

- I am no expert in kriging. I think it would be great if you can give some references to studies that use similar kriging approaches. From your code, I understood that you use "OrdinaryKriging" from the PyKrige package. I think it would be good to cite that package.

We have added the following statements to the text regarding kriging:

L195-203: Kriging is a core method of spatial statistics (Cressie, 1993), often coined 'best linear unbiased interpolator' due to its non-parametric nature and empirical variance minimization. It emerged in mining applications (Matheron, 1965), and has since become ubiquitous for spatial interpolation across many fields (Webster and Oliver, 2007). In glaciology, kriging has been for instance used to spatially interpolate sparse ablation measurements (Hock and Jensen, 1999) or ice thickness measurements (Fischer, 2009). Recently, the rise of machine learning methods has extended kriging concept to any kind of dimension through Gaussian Processes (Rasmussen and Williams, 2006), which have also found applications in glaciology, from remote sensing time series interpolation (Hugonnet et al., 2021) to model error emulation (Edwards et al., 2021).

L190: We have added a citation to PyKrige.

- **L196–201**: You write that the observed 5-year anomaly "Hugonnet et al. (2021) spatial correla tion patterns" validate the modelled annual MB anomalies. Can you write in your paper another sentence explaining that? They both have the same pattern of a decreasing correlation over the distance (what is expected). Though, apart from that, the two look to me, as non-expert, quite different. For example, the observed 5-year anomaly spatial correlation pattern starts at much lower correlation values (maybe also expected, but something to eventually describe?), and the correlation decreases first stronger and then decreases less with the spatial lag. In comparison, the correlation of the modelled yearly MB anomalies decreases at small spatial lags only mini mally, but then decrease at larger spatial lags stronger. What I want to say: the "shape of the curve" is different between the two, or not?

We agree with the referee that the validation of annual modelled anomalies using 5-year observed ones was not sufficiently explained. We have updated the statement as follow::

L204-2011: In order to estimate the spatial correlation of the annual mass-balance anomaly ρ_(β,Y)(d) to constrain the kriging, we sampled empirical variograms for both local-scale modelled annual mass balance anomalies (Huss and Hock, 2015) and for observational 5-year anomalies (Hugonnet et al., 2021), the latter validating the spatial correlation patterns observed in the modeled estimates (Fig. 3). The 5-year anomalies are used only to validate annual anomalies. As climatic patterns driving correlations in regional anomalies should have a size that is largely consistent in time, we expect 5-year anomalies to be spatially correlated at similar distances than annual anomalies but with a lesser amplitude due to the cancelling of positive and negative anomalies over time. We indeed identify that both anomalies have significant spatial correlation up to 5000 km, with a smaller amplitude for 5-year anomalies (Fig. 3).

Additionally, to address the referee's comment about the form of the correlation function and the sensitivity to modelling uncertainties (further detailed in next comment), we add the following statement:

L228-235: We note that, because kriging is a non-parametric interpolation method, its prediction primarily depends on the observations themselves, so uncertainties in the correlation function stemming from the modelled estimates of Huss and Hock (2015) have little influence on our results. Furthermore, because our correlations span multiple orders of magnitudes (from 10 km to 5000 km), the choice of functional form of the correlation has been shown to have minimal impact on the prediction (Hugonnet et al., 2022). To exemplify this, we compared kriging with inverse-distance weighting, a different interpolation method altogether, and found almost equal regional estimates as those are primarily driven by the input data. Differences between kriging and inverse-distance weighting only showed at the glacier-scale, where kriging allows to further refine anomalies and derive empirical uncertainties."

- Related to that, I am missing one sentence of the potential influence of using modelled glacier MB anomaly data to assess the correlations (to add here or in the discussion). If I understand it correctly, you use glacier MB anomalies from GloGEM (Huss and Hock, 2015) which is calibrated with regional geodetic MB data. Each individual glacier's specific mass balance was forced to match the average regional specific MB during the same multi-year time period. In addition, Glo GEM's modelled interannual mass-balance variability likely depends on the chosen calibration option / calibrated precipitation factor. So I am wondering, does the way how much the precipi tation factor varies from one glacier to the next influence the interannual MB variability and with that the Kriging results? I know analysing this is completely out of the scope of this study, but I think it would be really great to mention this potential model-biased issue very briefly. Or, if you don't think it is an issue, describe why.

See the previous answer.

- **Eq. 4**: Do I understand it correctly that $\rho_{\beta,y}$ describes the y-axis of Fig. 3 (blue line). In that case, I believe, something has to be wrong with the parameters or the equation. The current equation 4 will give correlations ρ with values above 0.23 for all real "d"-values. Fig. 3, however, shows, that the fitted ρ reaches correlations near to zero. Or do I misunderstand here something, and Eq. 4 shows another "unit/metric" than Fig. 3?

Good catch from the referee. This was an old version of the equation, not properly converted from variogram function to correlation function. The correct correlation function does not require a nugget n, and reads:

$$\rho_{\beta,Y}(d) = s_1 e^{-\frac{3d}{r_1}} + s_2 e^{-\frac{3d}{r_2}} \; if\ d > 0, else\ 1$$

where d is the distance between two glaciers, $s1$ = 0.37, $s2$ = 0.59 are the partial sills and $r1$ = 200 km and $r2$ = 5000 km are the correlation ranges

- **Fig. 3**: I was first a bit confused about the 23 crosses for the "empirical variogram". Can you maybe add in the caption one or two words to clarify that? I first thought that the "23" corresponds to the eventually 23 used "glacier sub-periods" (but then understood that this does not make any sense). If I understand it now correctly, the crosses describe the "Average empirical variogram". If yes, consider adding "average" to the label to make it easier understandable.

The confusion must have come from the fact that empirical variogram and correlation were used interchangeably. We have modified the legend and caption to consistently use “correlation” everywhere for describing the graph.

1.3 Leave-block-out cross validation

Thanks a lot for adding that additional analysis. To make the new analysis even more useful, I suggest to consider the following aspects:

- **Table 6**: You provide the ME and S residual. Do these values come from "the yearly results" (same metric as in Fig. 7a)? Would it be easy to add also the metrics for the "Ba variability vs leave-one out BA var. STD" (Fig. 7c)? I would find that interesting to understand whether the interannual variability underestimation increases with the leave-one-block-out estimate. If I understand it correctly, the metrics presented in Figure 9 do not directly describe how the interannual MB variability changes with the leave-block-out cross validation.

Yes, this is possible. We added the variability residuals of the leave-block-out experiment to a revised Table 6.

- **Figure 9a, b**: I am not completely sure if I understand correctly what is represented. Subplots a, b do have a "violet" color, do they also present the "1km" threshold? I guess no and they rather represent all threshold options. Please clarify.

We color-coded the results in panels 9a and 9b to avoid confusion.

- **L545–547**: "S is larger than σ only in some few cases with distances to the closest glacier >500km, b... .": I guess this estimate comes from a quantitative analysis of the data of Fig. 9d? Figure 9d looks like more than a "few" glaciers, but it is very difficult to check as the dots overlap. Can you add some kind of statistics to the end of that sentence?

We changed the statement to:
L 572: On average S starts to become larger than σ with distances to the closest glacier larger than 500km, but the large spread suggests this is coming from the randomness of the predictions.

- **Fig. 9d y-label**: missing ")" bracket

Corrected

- **caption**: mass-change estimate -> mass-change estimate

Corrected

2 Specific comments

I list these points in the order of their appearance in the manuscript, rather than by their significance.

- title: I don't have a strong opinion here, and I am ok with both titles

The editor has chosen to keep: Annual mass change of the world's glaciers from 1976 to 2023 by temporal downscaling of satellite data with in-situ observations

- **L61**: mountain ranges(Brun et al –> missing "space"

Corrected

- **L81**: "we use glaciological observations from approximately 500 glaciers" ... please add from how many glaciers you use the glacier MB anomaly. I know I mentioned that already in the last round and you answered that this number is visible for every region in Fig. 1. This is true, though I would really appreciate it if you add just behind this sentence in a bracket (15X glaciers used for the glacier MB anomalies...). Or do you use annual MB glacier observation data from the other around 350 glaciers? I understand L179 clearly in that way, that you use those glaciers with a "glacier MB anomaly", i.e those glaciers with 8 years of MB data within the 10-year reference period.

We added a statement of the total number of glaciological glaciers used in the method section, just after mentioning the rules of selection. We think the introduction is too early to talk about glacier anomalies since we explain that later, therefore there we use the number for the total amount of observation ingested to the processing as input data which is about 500.

L180: We use a total of 158 individual glacier anomalies for the assessment.

- **L320**. There are two dots after Zemp et al. (2019)

Corrected

- **L321**: You assume that all mass change occurs above sea-level. If I understood it correctly, your dataset anyway only describes mass change above sea-level and not the subaqueous mass loss. Maybe you can clarify that in such kind of a sentence, such as : "As our dataset does not capture subaqueous mass loss, we assume that all estimated glacier mass loss occurs above the sea level..."

This information was added in Table 9 for this study + Hugonnet et al., 2021 and Zemp et al., 2019

- **L415–417**, **L438–439**: "We remind here that, by construction, nearby glaciers share a large frac tion of the variance in mass balance variability and are thus not independent"... You write this sentence here at the beginning of the section. And repeat a similar sentence at the end of that section ".. annual mass-balance anomalies are extracted from a handful of glaciers in each region and thus , in each region, individual glaciers share a large fraction of these variabilities" (and in the conclusion). However, I still miss in this section a bit the "uncertainty/error" component. Would it be possible to add one half sentence or sentence on the end ... Something like: "The data user should carefully check the associated errors to decide if the dataset can be used for their specific use case... "

We think adding a phrase like this one is redundant, it stands more as a recommendation for users than a description or analysis of the dataset itself. We manifest already multiple times and in a transparent way that this aspect is the main limitation of our dataset, and it is also apparent in the large uncertainties at the individual glacier time series. We instead added:

L164: This limitation is, however, well evidenced by large uncertainties on under sampled regions and periods.

• **L450, Figure 7 caption**: You write sometimes 73 glaciers and sometimes 74 glaciers. I guess it should be 74 everywhere?

Corrected, it should be 74 everywhere

• **Fig. 8.1/8.2** There are a few glaciers and years, where the annual MB is not included within the uncertainties of the leave-one-out estimates mass-balance (e.g. Mittivakkat or Djankuat). This shows that the estimated annual MB can also be "completely" off even when considering the provided uncertainties/errors. Eventually consider mentioning that in the discussion.

We added to the discussion:

L510-521: Still large differences may occur between reference and predicted values for individual years (outliers in Fig. 7a), also evident on the few annual values where the reference annual mass balance is not included within the uncertainties of the leave-one-out annual estimates (e.g. Mittivakkat, Fig. 8j and Djankuat, Fig. 8n).

• **Table 7** Huss and Hugonnet (in prep): Is this the same as the "Huss (in preparation)" somewhere else in the manuscript? If yes, use consistent naming.

Done and made consistent throughout the text: density propagation based on Hugonnet et al, (2022), and Huss et al. (in prep.)

L564: is there a word missing? Should it be "35% smaller than the XXX predicted"?

Corrected