Second round of review of *"Annual mass changes for each glacier in the world from 1976 to 2023"* by *Ines Dussaillant et al.*

Earth System Science Data: essd-2024-323

Dear Ines Dussaillant and co-authors, dear Editor,

The manuscript has improved significantly since the previous submission. I appreciate the considerable 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

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?

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).

1.2 Interpolating with kriging

L185–L215 (including Fig. 3): This is a very interesting new way to select the glacier annual MB anomalies 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?
- 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.
- **L196–201**: You write that the observed 5-year anomaly "Hugonnet et al. (2021) spatial correlation 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 minimally, 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?
- 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, GloGEM'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 precipitation 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.

- **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?
- **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.

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.
- **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.
- **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?
- **Fig. 9d y-label**: missing ")" bracket
- **caption**: mass-changeestimate -> mass-change estimate

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

- **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.
- **L320**. There are two dots after Zemp et al. (2019)
- **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..."
- **L415–417**, **L438–439**: "We remind here that, by construction, nearby glaciers share a large fraction 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... "
- **L450, Figure 7 caption**: You write sometimes 73 glaciers and sometimes 74 glaciers. I guess 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.
- **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.
- **L564**: is there a word missing? Should it be "35% smaller than the XXX predicted"?