Review of ”*Brief communication: Nonlinear sensitivity of glacier-mass balance attested by temperature-index models*” by *Vincent and Thibert*

The Cryosphere

1 General comments

Vincent and Thibert present a brief communication with an experiment based on two French alpine glaciers, which shows the response of point glacier mass balance at different glacier altitudes and glacier-wide mass balance to temperature and precipitation changes. This study is designed as a reply to the Nature Communications paper ”Nonlinear sensitivity of glacier mass balance to future climate change unveiled by deep learning” by Bolibar et al. (2022)[1]. They claim that Bolibar et al. (2022) suggest that temperature-index models cannot capture nonlinear responses with respect to temperature and precipitation changes, and they aim at demonstrating the opposite.

This study serves as an extension to Bolibar et al. (2022), performing some additional analysis with temperature-index models that were not covered in that study. In that aspect, it serves to shed some additional light into the topic of glacier mass balance response to climatic changes. However, the scope of the study is very limited, and one is left feeling that only a few elements are analysed, often via cherry picking. My main concern regarding the study are the methods and the absence of objectivity in some of their claims. There is a lack of consistency in the way the information is presented and with which the different analyses are carried out. I will cover more in detail each one of these aspects in the following subsections of the global comments.

1.1 GC1: Cherry picking of sentences out of context

The first concern regarding this paper is the deliberate attempt to cherry pick sentences out of context in order to drive a point home. The most notorious of these is the widely repeated one in this study of ”temperature-index models can only provide a linear relationship between positive degree-days (PDDs), solid precipitation and mass balance (MB)”. While it is true that such a sentence is written in the article, many nuances are added around it. Bolibar et al. (2022) mention twice (pages 5 and 8), that the linear response to temperature is related to each individual degree-day factor (DDF), and that a temperature-index model with two DDFs (like the one in this study) virtually acts as a piece-wise function, able to partially account for some of the nonlinearities.

The authors seem very fixated with that idea, and seem to neglect this information altogether, showing a lack of objectivity. In that sense, this study serves to corroborate this hypothesis presented in Bolibar et al. (2022). Fig. 4c and Fig. 5 clearly display the piecewise behaviour of a temperature-index model with two DDFs.

In that sense, I believe it is important to nuance the message presented in this article, acknowledging the fact that this was already mentioned in Bolibar et al. (2022). This study is presented as an opposition to the message of Bolibar et al. (2022), whereas in fact it is building on top of it and corroborating a message evocated in that study. I would ask the authors to update all references to this sentence and to incorporate the elements described in this section into their study.

1.2 GC2: Model calibration and validation

Perhaps the most striking aspect of the methods is the lack of details regarding model calibration and validation. The authors present an equation used to model the mass balance, but they give no clues on how the two free parameters of the model (i.e. the DDFs of snow and ice) were obtained. The values are presented in the study, but one cannot know if these come from literature values or if these were calibrated somehow.

Seeing the model fit from Fig. 1, I am inclined to believe that these two parameters were manually calibrated, but it is unclear how that was performed.

- Is the model calibrated in an out-of-sample manner? Has the dataset been divided into a calibration/validation one and a test one? The model performance cannot be evaluated with the same data used for parameter calibration, otherwise one is ovefitting the model and reporting wrong metrics[2]. Please clearly explain how the model parameters have been calibrated, and if these have not been calibrated in an out-of-sample manner.
- What is the actual out-of-sample performance of the model for these two glaciers? Please report standard metrics (e.g. RMSE, bias and r^2)
- Have you taken into account ice dynamics in this model? How do you account for glacier geometry changes? How is the topographical feedback taken into account? Please specify.
- In Fig.1: why are there only MB simulations from 1990 onwards?

1.3 GC3: Interpretation of the glacier mass balance nonlinearities

An important aspect regarding this study is the interpretation of the nonlinear response of glacier mass balance to different climatic drivers (air temperature and winter snowfall in this case). The authors correctly point out that the reason behind the nonlinearities captured by their temperature-index model are the changes in duration of the accumulation and ablation season, which impact the snow/ice coverage ratio. While this is indeed one of the multiple nonlinear effects present in the response of glacier MB to climatic drivers, it is not the only one. The global picture is much more complex than that, with a complex combination of multiple feedbacks. These nonlinear effects are linked to the non stationarity of model parameters (i.e. DDFs for snow, firn and ice) in both the temporal and spatial dimensions[3]. These can vary in magnitude, and depending on the topographical and climatic setup of each glacier, one might be more important than the other. From our current understanding of these processes, these are the main ones:

1. **The influence of variations of the surface energy budget components under climate change**: This was the main topic of discussion and the most important result in Bolibar et al. (2022). Since the role of shortwave radiation in the energy budget in the past (i.e. the calibration period) is higher than in the future under climate change, its importance is bound to decrease in the future[4]. This results in a REDUCED sensitivity of DDFs (particularly of ice, due to its lower albedo) to future warming. This corroborates many studies in the literature that also encountered an overestimation of DDFs sensitivity to future warming[3,5,6]. For the whole region of the French Alps, Bolibar et al. (2022) found that this was the main nonlinear effect, driving differences in projected mass balance changes. Nonetheless, Bolibar et al. (2022) found that this was true only for glaciers with long response times or flat glaciers, due to the reduced effect of topographical adjustment. This nonlinearity affects parameters in the TEMPORAL dimension, resulting in a decrease in sensitivity over time, as air temperature rises.

2. **The influence of different surface types and therefore different DDFs in the temperature-index model**: The use of multiple DDFs for snow, firn and ice results in a nonlinearity in the SPATIAL domain. This nonlinear response will be affected by the spatial distribution of snow, firn and ice over the glacier. This spatial distribution will indeed also change through time, which will determine the switch between DDFs in the ablation season. Nonetheless, it is highly tied to glacier hypsometry. As reported in this study, in a warming climate, this lengthening of the ablation season exposes more ice surface linked to higher DDFs and therefore INCREASES the sensitivity.

3. **Surface albedo**: Changes in surface albedo through time also introduce a nonlinear response to warming in the TEMPORAL domain. These are also linked to 1, but they produce an opposite effect. Generally, in a warmer climate, surfaces tend to darken, thus further INCREASING the sensitivity of DDFs[7]. As mentioned, this process works in opposition to 1, so depending on the different topo-climatic setups, one might become more important than the other one.

4. **Glacier hypsometry**: This one affects both the previous processes, and it serves to display how complex are the interactions between all these feedbacks. As explained in Bolibar et al. (2022), flatter glaciers or glaciers with a long response time will display less topographical adjustment, thus enduring more extreme air temperatures. This will result in more climatic extremes and therefore increased nonlinear effects due to the reduced influence of shortwave radiation. Therefore, flatter glaciers will tend to display REDUCED senstivities to warming, whereas steep glaciers will not see many differences.

   The results of this study help shed light on the above-metioned point 2, but one should not jump too quickly to conclusions just because a model does display nonlinearities. As I just tried to argue, these nonlinear responses are combined in complex manners, and they are not straightforward to disentangle. While Bolibar et al. (2022) found that the above-mentioned point 1 seemed to be the most important nonlinear effect for the French Alps, this will most likely vary depending on the region and climate scenarios. More studies are needed to try to disentangle these nonlinear effects and to better understand their importance and effects for a wide range of topo-climatic setups.

In that sense, I believe it is important to mention and take into account this global picture in the conclusions of this study. Therefore, I think the results related to these nonlinear response should be presented as one of the multiple nonlinear responses of MB to climate change. Temperature-index models can indeed partially capture as a piece-wise function nonlinear effects linked to the spatial domain, but it remains unclear which is the most important nonlinear effect for multiple glaciological regions. Framing the results in this wider context will help place the scientific contributions of this study into the big picture.

1.4 GC4: Summer snowfall anomalies and plotting of nonlinear response

Bolibar et al. (2022) encountered that the strongest nonlinear response (from a statistical point of view) came from summer snowfall anomalies. The authors argued that it was the combination of both air temperature and precipitation during summer that determined wide changes in MB sensitivity. As explained above, Bolibar et al. (2022) argue that this is due to a reduced role of short-wave radiation in future climate scenarios, resulting in a reduced sensitivity of DDFs. Summer snowfall anomalies are tightly linked to summer air temperatures and also the ratio between snow and ice coverage on a glacier. These two are closely linked to processes mentioned in point 1 and 2 above, and were found to be the clearest drivers of nonlinearities. The statistical methods of Bolibar et al. (2022) served to shed some additional light on the subject, and open the door to exploring new ways to disentangle these processes. However, they did not allow a clear separation and understanding of how these processes operate.

Another important aspect in the comparison between the nonlinear sensitivities of Fig. 3 in Bolibar et al. (2022) with respect to Fig. 2 of this study, is the use of equivalent axis and ranges of values. Right now, both figures do not share the same range of values, and as it was displayed in Fig. 3 of Bolibar et al. (2022), there is a reduced range of values that will be encountered by French Alpine glaciers in future climate scenarios for different RCPs. This is particularly problematic for the case of winter snowfall anomalies. In Fig. 2 of this study, slight nonlinearities are displayed below -1.2 m.w.e. and above +1.7 m.w.e. These values are way beyond anything that will be seen in the 21st century for French alpine glaciers, as displayed in the vertical dashed lines in Fig. 3 of Bolibar et al. (2022). The most extreme values that French alpine glaciers will see until 2100 will range between -0.7 m.w.e. to +1.2 m.w.e. Anything beyond these limits makes no sense from a physical point of view for this sort of analyses, and will have no impact in projections for this century. At the very least, Vincent and Thibert should admit that nonlinearities of MB shown by Bolibar et al. (2022) for very extreme anomaly values out of the range of future likely encountered values, must simply not be taken into account in their analyses.

Additionally, one aspect that is not mentioned in this study is the fact that they are comparing the response of two glaciers with that of 660 glaciers. Bolibar et al. (2022) reported a strong variability in terms of mass balance sensitivity response to climatic forcings along different types of glaciers. The very reduced sampling used by Vincent and Thibert shows just a partial picture of all glaciers in the region. This should be specifically mentioned when presenting the comparisons.

In order to better understand these effects and to better compare both methods, I believe it is necessary to add the response of summer snowfall anomalies to Fig. 2. This would allow a comparison with the most meaningful response of the methods of Bolibar et al. (2022). Moreover, the future ranges of extreme values encountered by these glaciers under future climate scenarios (e.g. using the ADAMONT[8] product which is compatible with the SAFRAN[9] product used in this study), should be added to Figs. 2 and 3. This would clearly indicate where the nonlinearities actually will come into play and where they will be just model extrapolations beyond physically plausible values. This would also show that the nonlinearities linked to winter snowfall anomalies illustrated in Fig. 3 of Bolibar et al.(2022) will never occur during the 21st century in the French Alps, as they are out of the range of the values simulated by climate models.

1.5 GC5: Code and data availability

Another aspect that makes it hard to understand the methods is the fact that the source code used for this study is not shared. Following the principles of open science from The Cryosphere journal, I would strongly encourage the authors to share their code and data in an open repository (e.g. GitHub). This would make the study reproducible, and it would make it easier for reviewers and readers to understand what has been done.

If the authors strongly oppose to this, I would still ask them to privately share their code for this review in order to correctly understand what has been done.

2 Specific comments

- **L1** The current title does not give much information on what the sensitivity is linked to. I believe a correct title should be something like "Nonlinear sensitivity of glacier mass balance to future climate change attested by temperature-index models".

- **L18** The aspects regarding GC3 should be added here in the abstract.

- **L33-35** This is one of the cherry picking instances mentioned in GC1. To be adjusted accordingly.

- **L53** How has the temperature been downscaled to be used in the temperature-index model? Two versions of SAFRAN exist: one divided by massifs and altitudinal bands, and another one in a grid. Which one of the two has been used?

- **L61** As per the comments on GC2: how have been these two DDFs been obtained?

- **L79** This is one of the cherry picking instances mentioned in GC1. To be adjusted accordingly.

- **L97-98** This sentence is lacking solid arguments to back it. Could you please elaborate? Which parts of the model calibration might have issues? All the models in Bolibar et al. (2022) were cross-validated, ensuring a correct out-of-sample validation and a good generalization outside the seen dataset. This study so far does not provide any information regarding parameter calibration. In order to correctly compare both models and draw conclusions, a good understanding of both model calibration strategies is necessary.

- **L99-101** This is indeed true, and has already been reported in other studies. However, as argued in Bolibar et al. (2022) and as I explained in GC3, this is only part of the picture. This should be adjusted to mention that this is one of the multiple nonlinear processes in glacier mass balance sensitivity to climatic forcing, and that this process is a different one that the one reported in Bolibar et al. (2022).

- **L104-106** This is again a case of cherry picking. Bolibar et al. (2022) never claimed that ALL models in GlacierMIP 2[10] (not GlacierMIP 1, Hock et al. (2019), as stated by the authors) have linear relationships to PDDs and precipitation. To begin with, some of them use SEB. Bolibar et al.(2022) make a point that temperature-index models with a single DDF clearly behave like the Lasso; and even temperature-index models with 2 DDFs, can only partially account for the nonlinearities (and cannot capture the ones they show in their study). This is further corroborated by the comparisons made in that study between the Lasso MB model and the temperature-index MB model from GloGEM in the Supplementary material of Bolibar et al. (2022). To be modified accordingly.

- **L107-108** This is not accurate. The Open Global Glacier Model (OGGM), which was used in the paper[1] as an example of this behaviour, also has a single DDF.

  In order to avoid further cherry picking, I would ask the authors to be precise about their claims. Out of the of the 11 models in Marzeion et al. (2020)[10], 7 are using temperature-index models (2 of them with a single DDF), 1 is using a simple parametrizations relating MB to air temperature, 1 is using a mass balance gradient based on temperature indices, and 2 are using surface energy balance models. This means that at least 3 (potentially 4 if we take into account Kraaijenbrink et al. (2017)) models have direct simple linear relationships between PDDs and MB. The other 4 have 2 DDFs, which can partially account for nonlinearities (but not the ones from the above-mentioned point 1 in the temporal dimension).

- **L117-119** This is one of the cherry picking instances mentioned in GC1. To be adjusted accordingly.

- **L124-126** This was already done in that study. The results were shown in the Supplementary material. The exact same plots were not produced due to the difficulty of implementing that scheme on GloGEM. But the evolution of the MB for future scenarios was compared, yielding very similar results and responses to those of the LASSO. Therefore, the comparison between the LASSO and the TI model from GloGEM in terms of projected cumulative MB is not unfounded. Vincent and Thibert must point this aspect in an objective manner.

- **L129-131** As previously discussed in GC3, this is because the TI model used in this study does not account for DDF evolution over time. To be mentioned here in order to clarify the bigger picture.

References

1. Bolibar, J., Rabatel, A., Gouttevin, I., Zekollari, H. & Galiez, C. Nonlinear sensitivity of glacier mass balance to future climate change unveiled by deep learning. en. *Nature Communications* **13,** 409. ISSN: 2041-1723. https://www.nature.com/articles/s41467-022-28033-0 (2022) (Dec. 2022).
2. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning* ISBN: 978-0-387-84857-0 978-0-387-84858-7. http://link.springer.com/10.1007/978-0-387-84858-7 (2019) (Springer New York, New York, NY, 2009).
3. Ismail, M. F., Bogacki, W., Disse, M., Schäfer, M. & Kirschbauer, L. Estimating degree-day factors based on energy flux components. *The Cryosphere Discussions* **2022,** 1–40. https://tc.copernicus.org/preprints/tc-2022-64/ (2022).
4. Huss, M., Funk, M. & Ohmura, A. Strong Alpine glacier melt in the 1940s due to enhanced solar radiation. en. *Geophysical Research Letters* **36,** L23501. ISSN: 0094-8276. http://doi.wiley.com/10.1029/2009GL040789 (2021) (Dec. 2009).
5. Braithwaite, R. J. Positive degree-day factors for ablation on the Greenland ice sheet studied by energy-balance modelling. en. *Journal of Glaciology* **41,** 153–160. ISSN: 0022-1430, 1727-5652. https://www.cambridge.org/core/product/identifier/S0022143000017846/type/journal_article (2021) (1995).
6. Matthews, T., Perry, L., Guy, H., Wilby, R. & Edwards, T. *Modelling the impact of atmospheric heat accumulation on glacier mass balance* preprint (In Review, Oct. 2022). https://www.researchsquare.com/article/rs-2166876/v1 (2022).
7. Naegeli, K. & Huss, M. Sensitivity of mountain glacier mass balance to changes in bare-ice albedo. en. *Annals of Glaciology* **58,** 119–129. ISSN: 0260-3055, 1727-5644. https://www.cambridge.org/core/product/identifier/S0260305517000258/type/journal_article (2021) (July 2017).
8. Verfaillie, D., Déqué, M., Morin, S. & Lafaysse, M. The method ADAMONT v1.0 for statistical adjustment of climate projections applicable to energy balance land surface models. en. *Geoscientific Model Development* **10,** 4257–4283. ISSN: 1991-9603. https://www.geosci-model-dev.net/10/4257/2017/ (2018) (Nov. 2017).
9. Durand, Y. *et al.* Reanalysis of 44 Yr of Climate in the French Alps (1958–2002): Methodology, Model Validation, Climatology, and Trends for Air Temperature and Precipitation. en. *Journal of Applied Meteorology and Climatology* **48,** 429–449. ISSN: 1558-8424, 1558-8432. http://journals.ametsoc.org/doi/abs/10.1175/2008JAMC1808.1 (2019) (Mar. 2009).

10. Marzeion, B. *et al.* Partitioning the Uncertainty of Ensemble Projections of Global Glacier Mass Change. en. *Earth's Future.* ISSN: 2328-4277, 2328-4277. https://onlinelibrary.wiley.com/doi/abs/10.1029/2019EF001470 (2020) (Apr. 2020).