Reply to Reviewers :

We have complied with all the comments made by Reviewers (see the point-by-point response). The replies are in italics.

The manuscript has been revised by a professional native speaker (not this Reply).

Reply to Reviewer 2

General comments :

The submitted manuscript presents numerical experiments on glacier surface mass balance (SMB) based on a temperature index (TI) model. The authors aim on demonstrating that even with linear relationships between air temperature and snow/ice melt, such models show a non-linear sensitivity to climatic variations. The manuscript is a direct response to a publication by Bolibar et al. (2022), which presents comparisons between a deep ANN approach for estimating the glacier response to climate projections with simple TI models.

Vincent and Thibert criticise the proposition in Bolibar et al. (2022) that simple TI models do not show a non-linear sensitivity to climate variations in their experiments, in contrast to the deep ANN experiments. This is not the place to discuss details and validity of the Bolibar et al. (2022) paper, but rather to review the issues presented in the manuscript at hand. The authors raise an interesting question, which was also discussed earlier: what is the characteristic behaviour of TI models for changing climatic boundary conditions? It seems that one potential conflict is not as severe as it is presented. Even though Bolibar et al. (2022) state that TI models usually show a linear response to climate variations, they mainly compare a fully linear Lasso approach with their ANN model. They even claim that TI models with different degree day factors for snow and ice melt, show some non-linear behaviour, but that their response is not adequate compared to the ANN approach. Therefore, the response of Vincent and Thibert should be focussed on the validity of the TI model response, rather than on just demonstrating the non-linearity per se.

However, this manuscript provides rather interesting insights into the fundamental behaviour of TI models and this analysis could serve as a great study about the model characteristics, if some shortcomings could be fixed. The authors concentrate on demonstrating the non-linear response on a change in forcing, instead of discussing the fundamental interaction of the differences in snow and ice melt for the final glacier mass balance. There is a multitude of publications, which discuss the limitations of TI models due to their fixed relationship between air temperature and melt, while temperature is an indicator of energy availability, not energy transfer. But as long as the forcing stays within certain limits, TI models provide a robust and simple method for SMB estimates. Therefore, the critical investigation of the non-linearity characteristics within these limits would add high value to the discussion, also in the light of the application of AI approaches.

Thanks for your comments. We agree that this is not the place to discuss details and validity of the Bolibar et al. (2022) paper. The purpose of our study is to discuss the glacier response to climate projections with simple TI models only.

Given that temperature-index models have been widely used by the glaciological community for glacier projections in large-scale studies over the 21st century, we believe that it is crucial to clarify these issues.

Major concerns:

The data section does not provide the necessary information to evaluate the experiments. Only the two glaciers are described, but details neither about the mass balance data are given, nor about the necessary additions information, like DEMs etc. The methods section does not provide sufficient details. It is not clear how the model is applied to the glaciers. Is it a spatially distributed model, which cell resolution is used, is the glacier surface elevation static, or is there a dynamic response? What is the time step? How was the forcing parameterised across the elevations/aspects?

With regard to the general criticism of the Bolibar et al. (2022) paper, it needs to be highlighted that they estimate SMB for an entire region, while here two individual glaciers are considered. This allows a more detailed investigation of local SMB reaction, compared to general trends. There is no section about the determination of the DDF values. I would expect a section about calibration and validation with the available forcing data set, or at least information where to find these details.

Details about mass balance and DEM data are now given in the Data section.
The point mass balances are calculated for each elevation, for Argentière and Sarennes glaciers. In addition, we calculated the glacier-wide mass balance of Argentière glacier using the point mass balances for elevation range and geodetic mass balances (Vincent et al., 2009).
The degree-day factors for snow and ice are 0.0035 and 0.0055 m w.e. $K^{-1}d^{-1}$ *for Argentière glacier (Reveillet et al., 2017) and 0.0041 and 0.0068 m w.e.* $K^{-1}d^{-1}$ *for Sarennes glacier (Thibert et al., 2013). The calibration and validation of these factors have been done in (Reveillet et al., 2017) and (Thibert et al., 2013). It is not possible to provide more details in the present paper but further information can be found in these papers.*

In the Results section, you describe the non-linearity of the model response with an increase of sensitivity with respect to the anomaly (L.81). A major point would be to relate magnitude of these sensitivities to the sensitivities found by Bolibar et al. (2022) with their deep ANN approach and discuss the consequences within the bounds of potential future anomaly ranges.

It is not easy to compare the magnitude of sensitivities obtained in our study and found by Bolibar et al. (2022). Indeed, in the paper of Bolibar et al. (2022), the anomalies are calculated and averaged (i) from 660 glaciers, (ii) from glacier-wide mass blances of these glaciers, (iii) over the 1967-2015 period.
The purpose of our study is to show that responses in MB are not linear to temperature or precipitation changes even using a simple degree-day model. We used the point mass balances because they are free from dynamics impact (surface changes). The surface changes which influence the glacier-wide mass balance can lead to additional non-linearities. But here, we aim to discuss the response of mass balance using TI model only.
As seen in Figure 3, the response of annual mass balance to air temperature and to winter accumulation is different according to the elevation. The magnitude of these sensitivities are also different. It is also different from a glacier to another one. Thus, it is difficult to relate magnitude of these sensitivities to the sensitivities found by Bolibar et al. (2022) with their deep ANN approach using glacier-wide mass blances of 660 glaciers. It would require to apply the degree-day model on 660 glaciers, which is beyond the scope of our study.
Our topic is to show that the responses in MB calculated with a simple degree-day model are not linear to temperature or precipitation changes. Except the very high elevations where the glacier is covered by snow during the whole year, we demonstrate that the response of calculated annual mass balance using TI model to meteorological variables is not linear.

I wonder why you did not investigate summer snow fall in your experiments, as this is the major actor of non-linear response in the ANN approach. It should also be expected that summer snow fall has a strong non-linear response in the TI model, because of the difference in DDF values for snow and ice and the strong reduction of melt in the main ablation season.

According to this suugestion, we have done numerical experiments to analyse the summer snowfall anomalies.
The results are reported below:

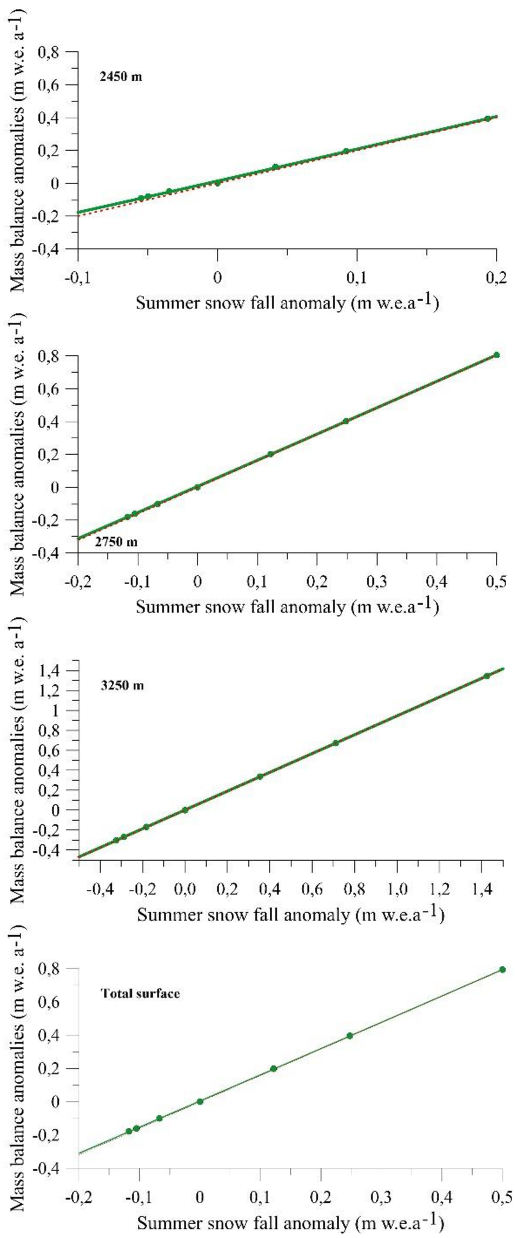


Figure: Surface mass balance anomalies against the summer snow fall anomalies, at Argentière glacier

We found an almost linear response of SMB to summer snow fall anomalies. These results have been checked using the synthetic data. Here below, the mass balance anomalies have been calculated from synthetic data we have used in our paper. The snowfall anomaly has been changed by increment of +/-100%. The sensitivity remains almost unchanged.

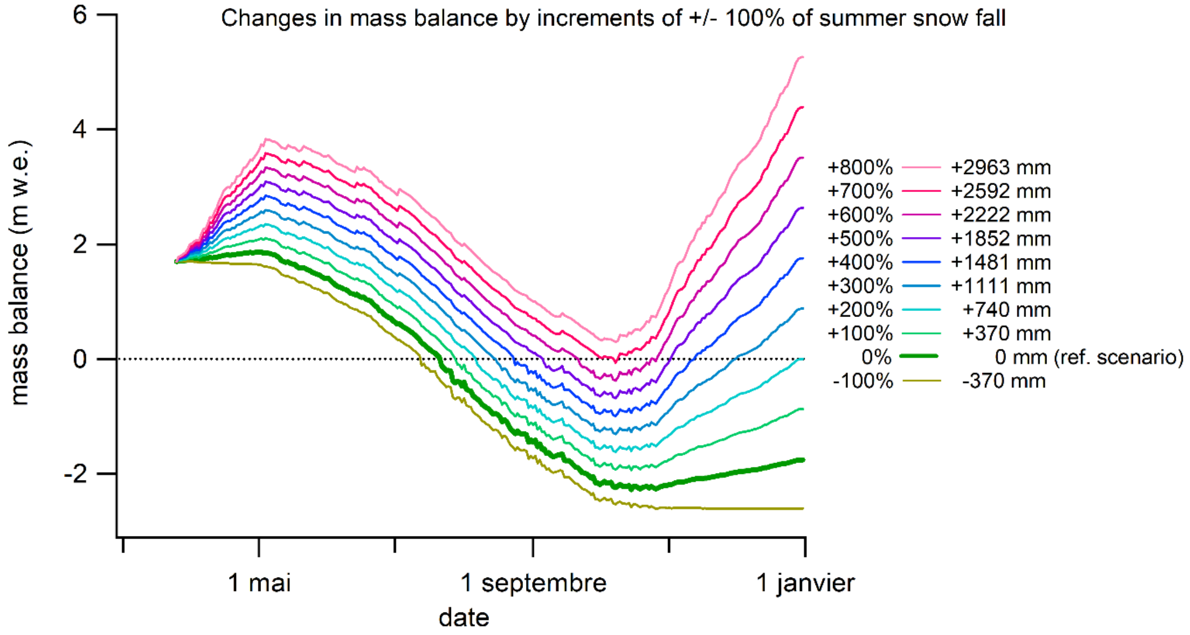


Figure: Surface mass balance calculated from synthetic data with different summer snow fall anomalies.

One can conclude that the response of SMB to summer snow fall anomalies, using degree day model is almost linear. The annual mass balance anomaly cannot be detected, probablybecause the summer snow falls are low and do not affect significantly the sensitivity. Although the summer snow fall can affect the summer mass balance, the response to summer mass balance anomaly is almost linear. It is consistent with in situ observations given the low quantity of summer snow fall on alpine glaciers. In addition, in the future, the summer snowfall will be increasingly lower. It is very surprising that Bolibar et al. (2022) encountered that the strongest nonlinear response (from a statistical point of view) came from summer snowfall anomalies.
However, this new discrepancy with Bolibar et al. study is beyond the scope of our paper given that the topic of our paper is to demonstrate that the responses of degree day model to temperature and winter accumulation are non linear.
Given the limited format of "Brief communication", we did not add anything about the response of annual mass balance to the summer snow fall anomalies.

Sensitivity to winter balance: in L.94-98 you describe the contrasting results of the TI model with respect to the ANN model with regards to a decreasing winter balance. However, there needs to be an explanation why the TI model explains reality and what are the consequences in the view of the ANN results. As ANN is more or less a black box, the results cannot be judged in the view of physical constraints, but just in respect to validation data sets. An investigation on the physical basis for the TI models' sensitivity would improve the discussion about the pros and cons of the two different approaches.

To explain the increased sensitivity of mass balance with decreasing winter accumulation, we performed runs of our PDD model on synthetic data under different conditions of winter balance. The results are reported in Figure 5.
The results show that the increase in sensitivity can be physically explained by the earlier disappearance of the winter snow cover. The earlier increase in the ablation rate under lower conditions of winter balance results in nonlinearity attested by the spread between MB plots in Figure 5b. For instance, with winter accumulation decreased by -1500 mm, the ice ablation starts very early (by the end of May) and the annual MB is close to -5.55 m w.e. a^{-1} in October. With winter

accumulation increased by +1500 mm, the ice ablation starts in mid-September and the annual MB is close to -0.21 m w.e. a^{-1} in October.

Our findings are consistent with in-situ observations (Six and Vincent, 2014).

It is also consistent with the results of Reveillet et al. (2018) using observations and energy balance modelling. It has been well illustrated in Figure 6b of this paper:

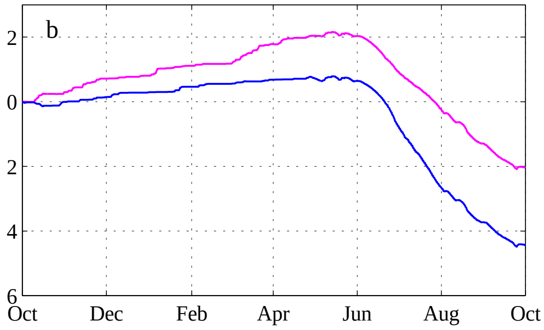


Figure 6b of Réveillet et al. (2018 Surface mass balance at stake 10 (2760 m a.s.l.), over one hydrological year, using averaged summer conditions (over 1996–2015), 2000–2001 winter conditions (pink) and 2008–2009 winter conditions (blue), representing the two extreme results.

In Reveillet et al. (2018), the tests of the annual mass balance sensitivity to seasonal mass balance using the Crocus model were performed at seven stakes in the ablation area, ranging between 2700 and 2870 m a.s.l. Only the results for stake 10 (located at 2760 m a.s.l.) are presented in Fig. 6b above, but conclusions are similar for all the stakes. The difference between winter accumulation of these 2 years (2001 and 2009) is 1.2 m w.e (Fig. 6b). Using the same summer conditions, the difference at the end of the hydrological year is 2.4 m w.e. (i.e. twice the difference at the end of the winter season). Thus, from these results, one can conclude that the sensitivity of annual mass balance to winter accumulation is close to 100 % when the winter accumulation is close or exceeds 2.4 m w.e. (because the glacier is covered by the snow over the whole melting season). In the contrary, with very low winter accumulation, as observed in 2008/2009, the sensitivity of annual mass balance to winter accumulation is close to 200 % (1.2 to 2.4 m w.e). These results from observations and energy balance modelling are very consistent with the results shown in Figure 3 of our paper at 2750 m (right panel): the sensitivity of annual mass balance to winter accumulation is about 110 % when the anomaly of winter accumulation exceeds 1 m w.e. On the contrary, the sensitivity of annual mass balance to winter accumulation is very close to 200 % when the anomaly of winter accumulation is lower -1 m w.e.
One can conclude that these results show (i) a strong non-linear effect (ii) and an obvious increase in MB sensitivity with low-winter accumulation.
Some explanations have been added in the manuscript.

Réveillet, M., Six, D., Vincent, C., Rabatel, A., Dumont, M., Lafaysse, M., Morin, S., Vionnet, V., and Litt, M.: Relative performance of empirical and physical models in assessing the seasonal and annual glacier surface mass balance of Saint-Sorlin Glacier (French Alps), The Cryosphere, 12, 1367–1386, https://doi.org/10.5194/tc-12-1367-2018, 2018.

Six, D. and C. Vincent (2014), Sensitivity of mass balance and equilibrium-line altitude to climate change in the French Alps. J. Glaciol. 60, 867–878. doi:10.3189/2014JoG14J014

Minor issues:

L.24: I would prefer "Surface mass balance" projections instead of "glacier mass projections", as the projections aim on SMB not on the full mass variations (which include basal melt and other processes).

It has been changed.

L.24-29: There should be at least a short characterisation of the differences in model approaches, in order to clarify the topic.

The temperature index model is described in Method Section. We added some explanations about the ANN approach: « A neural network is a collection of interconnected simple processing elements called neurons. These processing elements are connected with coefficients or weights, which constitute the neural network structure. Every connection of a neural network is assigned a weight that comes through training the ANN (Agatonovic-Kustrin and Beresford, 2000). »

Agatonovic-Kustrin, S. and Beresford, R. : Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research, Journal of Pharmaceutical and Biomedical Analysis, 22, 5, https:// doi.org./ 10.1016/s0731-7085(99)00272-1,2000.

L.37-38: Already here it would be helpful to shortly discuss the basic non-linearity of coupled linear relationships.

In Introduction, we present the questions and the topic of our paper. We do not think more explanations about non-linearity are required here in Introduction. It is thoroughly discussed later in the manuscript

L.54: The information about the reanalysis data needs to be described in the data section. It requires also some information on periods used, resolution etc.

It has been done.

L.64: Is k a function or just a two-value parameter?

Parameter k is depending on the site elevation to account for the precipitation gradient and is determined from winter balance measurements and precipitation data. Some information have been added in Method section.

L.74: how was the anomaly applied to the original data? I guess that you applied a constant anomaly to the daily values of the forcing series, in order to calculate a SMB anomaly.

The anomaly is a shift of the mean of the distribution of the original data in temperatures and winter balances. We kept the same distribution around the means to reproduce the year-to-year variability.
We added the following sentence in the manuscript. "The anomaly is generated as a shift (increment/decrement) of the mean of the distribution of the original data in temperatures and winter balances. The distribution around the means is unchanged (same year-to-year variability as found in the original data)."

L.77: It is not clear what you did here. I assume that you ran the model at specific points (where you presumably also have stake information) and as a distributed model across the entire glacier (which grid, etc.?).

We calculated the response of point mass balance at 2750 m and 3100 m on the Argentière and Sarennes glaciers (Fig. 2). In addition, we calculated the response of point mass balance at 2450 m, 2750 m, 3250 m a.s.l. and the response of glacier-wide mass balance on the Argentière glacier (Fig. 3). This information has been added in the manuscript.

L.79: Your statement with respect to Bolibar et al. (2022) is not exactly correct: Bolibar et al. (2022) write about piecewise linear relationships and implied non-linear response on page 5. However, their conclusion is that the TI model results are rather similar to the Lasso approach, which is clearly not confirmed by your investigations.

We agree that Bolibar et al. (2022) write about piecewise linear relationships and implied non-linear response on page 5. Indeed, in the Results section of Bolibar et al. (2022), on can read : « In that study, a temperature-index model with a separate degree-day factor (DDF) for snow and ice is used, resulting in piecewise linear functions able to partially reproduce nonlinear MB dynamics ».
However, they write in the following sentences: « Both the Lasso and the temperature-index MB model rely on linear relationships between PDDs, solid precipitation and MB. Therefore, their sensitivities to the projected 21st century increase in PDDs are linear. »

In addition, the statement is clear in the abstract of Bolibar et al. (2022), one can read : « Deep learning captures a nonlinear response of glaciers to air temperature and precipitation, improving the representation of extreme mass balance rates compared to linear statistical and temperature-index models. »

These statement is repeated at numerous places in the manuscript.

In the Introduction of Bolibar et al. (2022), on can read : « This type of model uses a calibrated linear relationship between positive degree-days (PDDs) and the melt of ice or snow. The main reason for their success comes from their suitability to large-scale studies with a low density of observations, in some cases displaying an even better performance than more complex models. However, both the climate and glacier systems are known to react non-linearly, even to pre-processed forcings like PDDs, implying that these models can only offer a linearized approximation of climate-glacier relationships. »

In the results section of Bolibar et al. (2022), on can read : « In that study, a temperature-index model with a separate degree-day factor (DDF) for snow and ice is used, resulting in piecewise linear functions able to partially reproduce nonlinear MB dynamics.

In Discussion of Bolibar et al. (2022), the authors justify tha analogy between Lasso model and Temperature index model. They write : « At this point, it is important to clarify the different ways of treating PDDs in the Lasso and the temperature-index MB models analysed in this study in order to justify analogies » and they conclude : « Nonetheless, since they are both linear, their calibrated parameters establishing the sensitivity of melt and glacier-wide MB to temperature variations remain constant over time. »

When they analyse the glacier models in GlacierMIP, they bring some nuances : « Despite the existence of a wide variety of different approaches to simulate glacier dynamics, all glacier models in GlacierMIP rely on MB models with linear relationships between PDDs and melt, and precipitation and accumulation. Some of these models use a single DDF, while others have separate DDFs for snow and ice, producing a piecewise function composed of two linear sub-functions that can partially account for nonlinear MB dynamics depending on the snowpack", but they finally conclude that « As we have previously shown, these models present a very similar behaviour to the linear statistical MB model from this study »

Finally, in the last paragraph of Bolibar et al. (2022), the authors claim that « By unravelling nonlinear relationships between climate and glacier MB, we have demonstrated the limitations of linear statistical MB models to represent extreme MB rates in long-term projections. Our analyses suggest that these limitations can also be translated to temperature-index MB models, as they share linear relationships between PDDs and melt, as well as precipitation and accumulation ».

Thus, we believe that Bolibar et al. (2022) clearly question the ability of temperature-index models to capture nonlinear responses of glacier surface-mass balance (SMB) to high deviations in air temperature and solid precipitation

L.82: Details about the synthetic input series are missing (how did you construct these series? Do they represent a certain realistic SMB range?).

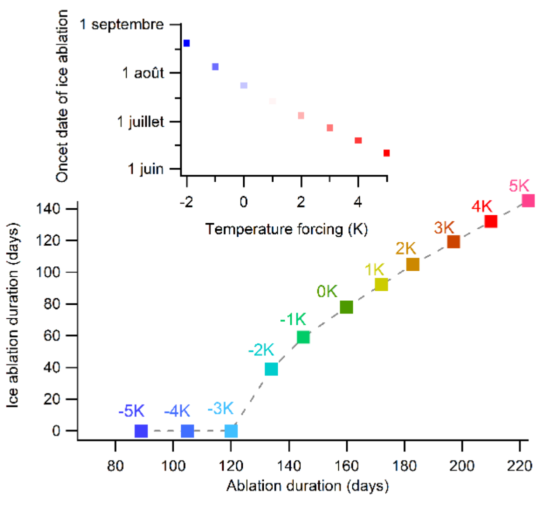
We added the following sentence in the manuscript:
"The reference scenario (unforced temperature and winter balance reference conditions) of synthetic data is typical for a location in the upper ablation area of an Alpine glacier. "

Further information are given in the following sentences:

"Runs of our PDD model on synthetic data under different conditions of winter balance (Fig. 5) used a reference scenario of 1,700 mm of winter balance changed by increments of ±300 mm in precipitation". Temperature data are shown in Figure 4a, typical for a course of atmospheric temperature from spring to autumn around 3000 m of elevation in the Alps. We use PDD factors for snow and ice from Thibert et al. (2013).
In addition, results from the PDD simulations on synthetic data are now accessible from the open data repository: 10.5281/zenodo.7603415.
It has been added in the Code and Data Avaibility of the new version of our manuscript.

L.85: It might be a good idea to show the length of ice ablation period vs total ablation period and the onset date of ice ablation. The onset of ice ablation is a measure of the nonlinear character.

According to your suggestion, we have done the following Figure:



It shows the ice ablation duration with respect to the total ablation duration for different temperature forcing. It shows the non linear response. However, we did not include this Figure in the manuscript in order to avoid to lengthen the manuscript which is a Brief Communication".

L.91: You describe an increase in sensitivity, but this should be quantified with respect to the disappearance of winter snow to judge the physical basis.

To explain the increased sensitivity of mass balance with decreasing winter accumulation, we performed runs of our PDD model on synthetic data under different conditions of winter balance. The results are reported in Figure 5.

The results show that the increase in sensitivity can be physically explained by the earlier disappearance of the winter snow cover. The earlier increase in the ablation rate under lower conditions of winter balance results in nonlinearity attested by the spread between MB plots in Figure 5b. For instance, with winter accumulation decreased by -1500 mm, the ice ablation starts very early (by the end of May) and the annual MB is -5.55 m w.e. a-1 in October. With winter accumulation increased by +1500 mm, the ice ablation starts in mid-September and the annual MB is -0.21 m w.e. a-1 in October. This asymmetry clearly shows that the response to winter accumulation is not linear.

L.110-111: There is a basic difference between the Lasso-models and the TI approach, as the second one uses a step function of the DDF parameter. Therefore, it cannot be expected that the two models provide the same response. This should be made clear.

The LASSO MB model is based on a regularized multi-linear regression. We added some explanations in the manuscript.

At numerous places in the manuscript of Bolibar et al. (2022), the authors claim that Lasso model and Temperature-index model both provide linear responses. In Results section, they wrote: « Both the Lasso and the temperature-index MB model rely on linear relationships between PDDs, solid precipitation and MB. Therefore, their sensitivities to the projected 21st century increase in PDDs are linear. »
In Discussion, the authors justify the analogy between Lasso model and Temperature index model. They write : « At this point, it is important to clarify the different ways of treating PDDs in the Lasso and the temperature-index MB models analysed in this study in order to justify analogies » and they conclude : « Nonetheless, since they are both linear, their calibrated parameters establishing the sensitivity of melt and glacier-wide MB to temperature variations remain constant over time. »

However, the discussion about the differences between LASSO model and temperature index model is beyond the scope of our study, which is focused on the non-linearity of Temperature Index models.

L.117: It might be a good idea to mention also earlier investigations who pointed out this basic behaviour.

As mentioned in the reply to general comments, our findings are consistent with in-situ observations (Six and Vincent, 2014). It is also in agreement with the results of Reveillet et al. (2018) using observations and energy balance modelling. Some explanations have been added in the manuscript.

Réveillet, M., Six, D., Vincent, C., Rabatel, A., Dumont, M., Lafaysse, M., Morin, S., Vionnet, V., and Litt, M.: Relative performance of empirical and physical models in assessing the seasonal and annual glacier surface mass balance of Saint-Sorlin Glacier (French Alps), The Cryosphere, 12, 1367–1386, https://doi.org/10.5194/tc-12-1367-2018, 2018.

Six, D. and C. Vincent (2014), Sensitivity of mass balance and equilibrium-line altitude to climate change in the French Alps. J. Glaciol. 60, 867–878. doi:10.3189/2014JoG14J014

L.121-126: This is a solid argumentation and provides a core conclusion. But is should be expanded by the major points, mentioned above.

Thank you for your comment.
We do not believe that we can expand our manuscript given that our paper is a Brief Communication.

L.129: It is only mentioned that the physical reasons are given for a higher sensitivity of TI models to lower winter MB, but I did not find a sound discussion.

It has been discussed in details earlier in the manuscript:

"Concerning the winter balance, we found a nonlinear response of MBs to winter precipitation with our PDD model and this is also inconsistent with the conclusions of Bolibar et al. (2022) relative to the sensitivity of temperature-index models. Runs of our PDD model on synthetic data under different conditions of winter balance (Fig. 5) used a reference scenario of 1,700 mm of winter balance changed by increments of ±300 mm in precipitation. Results show that the increase in sensitivity can be physically explained by the earlier disappearance of the winter snow cover. The earlier and abrupt increase in the ablation rate under lower conditions of winter balance (Fig.5a) results in nonlinearity attested by the spread between MB plots in Figure 5b. Surprisingly, we detect sensitivity to winter accumulation, contrary to the Bolibar et al. (2022) findings using their ANN (Fig. 2 and 3). Indeed, MB sensitivity increases with low winter-accumulation anomalies using our model, but decreases in the deep-learning model of Bolibar et al. (2022). The opposite results obtained from the deep-learning model are paradoxical and may be due to an issue in the calibration of the model."

Here, in the Conclusions, we summed up these results and this discussion.

L.132-136: The main concern is, that TI models applied far outside the calibration range of parameters might not be able to represent the energy exchange between the atmosphere and snow/ice correctly. However, this might also be true for the ANN approach, because it is unclear how good the performance is far beyond the training domain. Therefore, only a comparison of the different models in a large parameter space with a physical energy balance model would provide serious assessment of the model performance. This is out of scope of this manuscript, but could be mentioned as a valuable future step.

We agree. However, as you said, it is beyond the scope of our manuscript which is a Brief Communication with limited space.

Data availability requires at least the references to the data sets.

The Data are accessible from the Observatory of French glaciers through the project website at https://glacioclim.osug.fr. Results from the PDD simulations on synthetic data are now accessible from the open data repository: 10.5281/zenodo.7603415.
It has been added in the Code and Data Avaibility.