Change Log

Line references refer to the original submitted manuscript and may be on new lines in the revised version.

Overview

- Reviewer #2 comment: "L170: Are the section numbers correct here?"
    - Removed improper subsections that were coded with "subsubsection", fixing multiple section titles.

Abstract

- Reviewer #2 comment: "L6: Consider naming the method here."
    - L6: Named XGBoost method as suggested.

Data and Methods

- Reviewer #1 comment: "L125: Figures 1 and 2 in Jensen et al. (2023) show that the station network is densest in South and Southwest Greenland. It is often very far from the glaciers in your study to the nearest weather station and the observations might not be representative of a particular glacier even when choosing the nearest one. For instance, the weather station is located much closer to the open ocean than the glacier terminus in question. While direct observations are great, would it not be better to apply re-analysis data from e.g. CARRA or similar?"
    - L89: Added additional information on our choice of direct observation data over model products, where applicable.

- Reviewer #1 comment: "L125: Should be Danish Meteorological Institute (DMI) not Danish Meteorological Society"
    - L125: changed to "Danish Meteorological Institute"

- Reviewer #1 comment: "L131: Just to clarify, is the timeseries of ocean thermal forcing near the glacier fronts for the entire column of water or only at a specific depth or interval?"
    - L131-134: Added additional definition of ocean thermal forcing and the depth-averaging technique used to calculate it.

- Reviewer #1 comment: "I am not convinced that SST is a reliable proxy for mélange conditions. SST does not necessarily capture mélange strength or thickness, as you also note (L406). I understand the motivation, given the lack of a Greenland-wide mélange

product, but perhaps it would be more accurate to label this variable "fjord surface temperature," acknowledging that it reflects mixed influences from open water, fjord ice, and mélange. Alternatively, additional examples establishing its validity as a mélange proxy would strengthen the argument."

  - L140-148: Added further detail and exceptions to our use of sea surface temperature as a proxy for mélange presence, along with further justification for the use of sea surface temperature (SST) as a mélange proxy.
  - Figure 4 caption: Added a distinction regarding mélange presence referring to our use of sea surface temperature.

- Reviewer #1 comment: "Before your timeseries of data enter the machine learning setup, you remove the multi-annual component from the data leaving the seasonal component as input. You do this for every input variable (L153 - 154). Does this detrending apply to geometric variables such as bed slope and thickness, and to dynamic variables such as strain rate? If yes, what is the physical interpretation of removing the long-term component of bed slope or strain rate near the calving front (which changes position)? For instance, the velocity pattern and thus the strain rates at a given location are highly dependent on the geometry of the glacier (e.g. is it in a narrow or wider spot). Is it actually necessary to detrend the geometric data? Perhaps you can show the reader examples of the full and detrended time series of input variables like you have in Figure 2 for terminus advance. Also adding some more details of the method and on the resulting seasonal component would be useful to the reader, like does the timeseries have a zero mean?"
  - L153-155: Added additional details about the singular spectrum analysis.
  - Supplementary materials: Added 9 plots, similar to Figure 2, of the preparation plots for input variables used in the study.

- Reviewer #1 comment: "Because glaciers are not in steady state, more clarification is also needed on how the "magnitude of seasonal terminus change" is defined. Since termini do not necessarily return to the same position at the end of summer, how is this accounted for? In other words, do the long term changes still enter your prepared data through this? Are these situations where the algorithm struggles to make accurate predictions (L394)? Similarly, in Figure 2b some years show double peaks (e.g., 2013) — how are such cases treated?"
  - L155: Added additional details on the definition of "magnitude of seasonal terminus change" and other clarifications.
  - L155 Added that, not only do we expect some data leakage from the removed long term trends in our data, we expect this.

- Reviewer #1 comment: "L162: '…heterogeneity in this population (Table 4).' Should be Table 3, I think."
  - L162: Changed to Table 3.

- Reviewer #2 comment: "I very much appreciate that this article focuses on the results and the glaciological discussion. However, I think it would be useful for the reader to at least understand the basics how this method works. Section 2.1.1 need more information and

clarifications. For me, it currently raises a lot of questions. What are the 7 hyperparameters mentioned in L183? What learning rate was used? What about initialization? What's the stopping criteria used for the training? Also, it would be nice if it were briefly explained how the temporal aspects of the input time series are captured (e.g. were lag features or rolling statistics used?)."
    - L179-185: Added additional details about the machine learning model setup, including: naming the hyperparameters used, emphasizing that no lag features were used, and pointing readers towards the *Code and data availability* section.

- Reviewer #2 comment: "L188: To avoid confusion with the image processing technique, it would be better to use the terms 'input feature tracking' or 'feature importance tracking'."
    - L188: updated wording to be less confusing.

- Reviewer #1 comment: "L319: You have a mean temporal offset of 6.7 weeks between the predicted and observed peak. This is ~1.5 months so quite a large difference in timing. Is it a systematic bias or is it random?"
    - L207: Specified "the absolute difference" in time…

- Reviewer #2 comment: "…Would it make sense to specify an average distance between the actual and the model prediction, measured in metres?"
    - Line 210: Added a reference to the RMSE reported in Table 4.

Results

- Reviewer #2 comment: "L225: ... in Zhang et al. 2023."
    - L225-226: Changed to "with terminus position data from Zhang et al. (2023)."

- Reviewer #2 comment: "L227: This makes it sound like these glaciers are not so important. However, when it comes to discharge, the story is different. The NEGIS ice stream alone drains around 12% of the GIS. Also south and southwest glaciers are missing."
    - L227: Added mention of missing glaciers from the southern and southwestern coasts.

- Reviewer #2 comment: "…Would it make sense to specify an average distance between the actual and the model prediction, measured in metres?"
    - L242: Added the mean RMSE, now reported on Table 4.

- Reviewer #2 comment: "…Furthermore, I would like to see a handful of time series plots from the supplement (the '(b)' plots) included in the main paper. These would be very useful for understanding the text in section 3.2. So that there is at least one example for each of the different cases explained in the text."

    - Referred to the new Figure 5 in section 3.2. (L249,263-265)
-

Discussion

- Reviewer #1 comment: “L328 Perhaps change the subsection from ‘Feature Importance for Seasonal Terminus Prediction’ to ‘Feature Importance for Magnitude of Seasonal Terminus Prediction’?”
    - L328: Changed subsection title to “Feature Importance for Magnitude of Seasonal Terminus Prediction”.

- Reviewer #1 comment: “L357 forward: The studies in L357-359 you cite for findings on mélange impact on timing of retreat of the while you’re the focus of your study is the magnitude and not the timing (L338-340) of the seasonality. I am just wondering if you are comparing the right things.”
    - L360: Specified that our model was trained to predict the magnitude of seasonal terminus position.

- Reviewer #1 comment: “L394: You write that the best performing models are for glaciers that have experienced little overall retreat during the period. Does this mean, that geometric parameters are the main controls for outlet glaciers close to a steady state?”
    - Added detail to L394 to specify our model is predictive and that they perform best on glaciers close to a steady state.

- Reviewer #2 comment: “SHAP values assume independence of input features and can be misleading when they are correlated. I don't see any problem with experiments where the inputs are grouped into geometric, dynamic, and climatic categories. However, it could be problematic when the inputs are assessed independently (e.g. for Fig 4 and 5b). For example, if there is a significant correlation between strain rate and velocity, or OTF and ice mélange, the corresponding SHAP values may become unstable, with small changes in the data or model leading to significant fluctuations in the SHAP value. Perhaps XGBoost is more robust in this respect than other methods, but I would like to see this mentioned or addressed in some way in the paper. A table showing the mean correlation between the inputs would already help.”
    - Added Fig. 7, a heatmap of mean correlations between input variables, as suggested. Added a reference to this figure and a new source citation in our discussion, L431-434.

- Reviewer #1 comment: “In the abstract you write ‘We find that glacier geometry is important for accurate predictions of the magnitude of terminus seasonality and that environmental variables (mélange, ocean thermal forcing, runoff, and air temperature) are important for determining the onset of seasonal terminus change.’ I think these are the most important findings in your study, although perhaps not surprising? However, this is not really emphasized in your conclusions and outlook, where the focus is mostly on the potential of machine learning methods to improve our understanding e.g. seasonal

terminus changes from a data driven perspective and how they can be implemented with numerical ice flow models. While these are important points, they are more general/outlook, more weight should be given to the glaciological insights gained—specifically, what your study reveals about the processes governing seasonal terminus changes."

- Added a new paragraph to the Outlook section (now the 2nd paragraph of the section) that expands on our glaciological findings, as suggested.

Conclusions

- Reviewer #1 comment: "L455: suggests -> shows"
  - L455: Changed "suggests" to "shows".

Figures, Tables, & Supplementary Information

- Figure 1: Added GIDs to (a)--- caption now also refers to GIDs.

- Reviewer #1 comment: "Figure 2: A Figure 2c is mentioned in the caption but is not present. Are both trends now in Figure 2b?"
  - Figure 2 caption: removed reference to "2c", which has been consolidated into 2b.

- Reviewer #1 comment: mélange proxy
  - Figure 4 caption: Added a distinction regarding mélange presence referring to our use of sea surface temperature.

- Reviewer #2 comment: "…Furthermore, I would like to see a handful of time series plots from the supplement (the '(b)' plots) included in the main paper. These would be very useful for understanding the text in section 3.2. So that there is at least one example for each of the different cases explained in the text."
  - Added a new figure, new Figure 5, with four samples of additional '(b)' plots.
  - Referred to the new Figure 5 in section 3.2. (L249,263-265)

- Reviewer #2 comment: "Table 1: According to the paper (Zhang, E. 2023), the uncertainty of AutoTerm is 79 m. It's also not directly comparable to the 109 m of TermPicks. As far as I know AutoTerm users an average minimal distance, whereas TermPicks uses a median Hausdorff distance. Perhaps it would be better to make these uncertainty statements more general (e.g. ~100 m)."
  - Table 1: Changed uncertainty for AutoTerm and TermPicks to "~100 m".

- Reviewer #2 comment: "The model is evaluated using three accuracy metrics, which I think provide a great foundation for the subsequent experiments. That said, I don't think they currently do a particularly good job of giving one-time readers an idea of how well

this model actually predicts. Although the metrics are briefly described in Section 2.2, it may still be difficult to understand what the sentence in L241 actually means in terms of the quality of the predictions. Would it make sense to specify an average distance between the actual and the model prediction, measured in metres? Perhaps even an absolute and a relative (to the seasonal terminus variation) average?"

  - Table 4: Added a root mean square error (RMSE) column, in meters, as suggested.
  - Line 210: Added a reference to the RMSE reported in Table 4.

- Reviewer #1 comment: "… Perhaps you can show the reader examples of the full and detrended time series of input variables like you have in Figure 2 for terminus advance. Also adding some more details of the method and on the resulting seasonal component would be useful to the reader, like does the timeseries have a zero mean?"
  - Supplementary materials: Added 9 plots, similar to Figure 2, of the preparation plots for input variables used in the study.

References

- Reviewer #2 comment: "References: There seem to be many entries without a DOI. Consider adding them."
  - Added DOI to multiple entries.

- Reviewer #2 comment: "L673: This citation refers to the preprint."
  - Removed preprint citation from L673.
  - Fixed citation to correct source on L121-122.