Author Response to reviewers: “Random forest parameterization of Antarctic subglacial hydrology for coupled ice-flow modelling”

Tim Hill, Matthew J. Hoffman, Gwenn E. Flowers, Derek Bingham,

Reviewer comments are in black and we provide our responses in blue.

Reviewer Comment 1

This paper sets up a random-forest algorithm to predict flotation fraction and thus effective pressure at an Antarctic-wide scale based on solved solutions obtained using the physics-based model, GlaDS. The authors undertake a very comprehensive parameter sweep to ensure their GlaDS training ensemble sufficiently samples the probable parameter space for Antarctic simulations, such that the random-forest model does not have to extrapolate outside its training hull. The authors also conduct a range of sensitivity tests and other evaluations to demonstrate the generalisability, applicability and limits of the model, and convincingly show that the trained model is behaving sensibly and robustly. The results show that the trained model is able to reproduce the GlaDS solutions with a good degree of accuracy both under current and future geometries, and performs considerably better than other simple parameterisations without increasing computational cost/meaningfully. The discussion is also exceptionally thorough.

Overall, I think this is an excellent paper. It is certainly long and complex, but it is well-structured and -written, and reflects the thoroughness of the authors in testing their model, rather than any deficiencies in communication. To be honest, I have very few points of meaningful criticism. To me, the authors demonstrate convincingly that the model performs well and comprehensively describe the set of conditions under which it can be considered reliable. The work presented in the paper is also a major advance that could be of great interest to the ice-sheet modelling community. I recommend minor revisions, as there’s a couple of very small points I think should be addressed (see below), but they really are very minor. Congratulations to the authors! Page and line numbers refer to the clean version of the revised manuscript.

Thank you for the positive reception of this manuscript. We appreciate your careful reading of the manuscript and your helpful suggestions. We have responded to your Minor Comments individually below.

Sincerely,
Tim Hill
On behalf of the authors

Major Comments
None

Minor Comments
p.1,l.20: ‘depend’

Corrected ‘depends’ to 'depends’ from the line: “[...] the rate and magnitude of future mass loss from the Antarctic Ice Sheet **depend~~s~~** on effective pressure and [...]” (L20).

p.3, l.71: ‘pressure’
Corrected ‘pressures’ to ‘pressure’ as suggested in: “We prohibit cavities from opening by creep when water **pressure~~s~~** exceeds ice overburden pressure [...]" (L74).

p.7, l.173-177: True, but maybe burying the lead a little here? Greenland is hydrologically much more complex than Antarctica because a lot depends on surface melt, which is far more variable than basal melt. So it might be that the extra complexity of an NN is needed to adequately capture the extra dimensions of variability there. Another sentence considering that here might be worthwhile to make it clear that the two ice sheets might require different choices, rather than it being purely dependent on what’s easier to integrate into an ice-sheet model. A similar sentence added to the relevant bit of the discussion (p.23, l.476-487) would also be useful.
This is a good point. The additional dimensions of variability introduced by time-varying surface melt are one reason that we have focused on Antarctica in the present work, whereas we have focused on Greenland in previous work. We have added a sentence pointing out that Greenland hydrology is strongly forced by time-varying surface melt inputs (L181):

> The choice of a random forest regression model, rather than other machine-learning models, is based on balancing the ability to learn and generalize from data against training cost and the number of parameters that need to be estimated. **Other models, such as convolutional or graph neural networks may also be appropriate, although a neural network model may be harder to fully integrate within ice-sheet model software than a random forest. For example,Verjans and Robel (2024) used a convolutional neural network to emulate spatiotemporally resolved hydraulic potential from GlaDS simulations in Greenland. We have chosen to focus on Antarctic hydrology in part to avoid the additional complexity of seasonal meltwater forcing that is characteristic of Greenland hydrology.**

We have also added a brief discussion of these differences to the corresponding discussion section (L574):

> Given the different characteristics between the Antarctic and Greenland ice sheets, it is challenging to make further conclusions about methodological differences between these studies. It is possible that different approaches might be best suited for Antarctica compared to Greenland considering the seasonal surface melt forcing of the latter.

Equation 1: rho_sw is in the equation, then rho_w in the text
Corrected rho_w to rho_sw in the text (L239, Eq. 2).

Figure 2 caption: Looks as if there’s something wrong with the formatting in the first line where‘language=en’ is interpolated.
We have removed ‘language=en’.

p.14, l.288: I think there’s something missing from ‘This should be expected since pressures are near flotation.’ As written, it doesn’t seem to make sense or be coherent? Unless the implication is that the

effective pressure everywhere is near 0 and therefore nearly everywhere is floating, which seems unlikely? I may have misunderstood something, but please take a look to check.
Thank you for pointing out that our meaning was not clear here. We mean to say that, since water pressures are typically high (e.g., >80% of flotation), errors strongly propagate into effective pressure. We have integrated the fragment ‘This should be expected since pressures are near flotation.’ with the following sentence to convey our intention:
“This should be expected since pressures are near flotation, such that a small change in flotation fraction translates into a large relative change in N” (L366)

p.15, l.300: ‘with these features’
Corrected by adding ‘with’: “[...] since errors associated **with** these features are weighted more heavily.” (L378)

p.15, l.311: There’s a full stop that shouldn’t be there after ‘despite’
Corrected by removing the full stop.

p.37, l.653-4: In Data availability, the final link (and the most interesting one!) to the code and model data for this study is broken
We apologize for the broken link. We have corrected the link address (https://doi.org/10.5281/zenodo.18381581) and will verify that the link works in the revised PDF.

Reviewer Comment 2

This manuscript (MS) presents and discusses a new, statistical parametrisation of subglacial drainage. The statistical model is a Random Forest (RF) model trained on flotation fraction outputs of an ensemble of Glads model runs of seven large Antarctic catchments. The model is evaluated in terms of its capability to reproduce flotation fraction, effective pressure as well as ice flow speeds calculated with a one-way-coupled SSA ice flow model (using a Budd sliding law). The evaluation is done both for present day and in future configurations of Antarctica (2050 for West Antarctica and 2300 for East Antarctica). The ice flow outputs are also compared against SSA model runs using a sliding relation fed by Glads and by a perfect ocean connection (POC) subglacial drainage model runs.

The tuned RF model performs well in reproducing flotation fraction as modelled by Glads, a bit less well for effective pressure, but well also when coupled to ice flow. The MS shows that for future ice flow predictions the choice of model for effective pressure matters and that this RF of model could be useful for such simulations.

Unfortunately, the MS suffers from a incomplete presentation of the methods, too many model runs and, hence, results which are difficult to follow. It thus needs to be streamlined considerably before publication. In particular, the methods do not fully describe the RF model nor how it is evaluated against other models. For instance, the RF model needs to be described fully (e.g. number of trees, levels, total number of parameters) and what cost function it aims to minimise. Many methods applied are not defined, for instance what "mean" in various context means, what "perturbed parameters" are, what "cross-validation" is, etc. Last, there was not enough selection of which of the many performed model runs are presented. In the Results section, in many subsections a new set of model runs are presented and discussed without ever being specified accurately.

Concluding, the manuscript presents an interesting and potentially very useful subglacial drainage parametrisation. However, the confusing presentation of both methods and results lead me to not fully understand what the authors present and to a quite difficult time reading the MS. I suggest that (I) the methods are updated to carefully define the RF model, all the model runs (Glads, RF, POC, SSA), the calibrations and the evaluations, (II) the numbers of presented results is reduced to what is needed to support the main results (which should probably be: RF works well and RF potentially useful for predictive ice flow modelling). Point I is necessary whereas II would be nice-to-have.

Thank you for your detailed review and helpful suggestions to improve the manuscript. For suggestion (I), we have added to the sections describing the random forest model. We have reorganized the methods and results sections to ensure that all model runs presented in the results are first described in appropriate methods sections. We hope that the improved organization of the methods and results sections allows the main results to stand out more clearly without reducing the number of results (suggestion II), since we use all the results at various times to support our conclusions. We provide more details in individual responses to Major and Minor Comments below.

Sincerely,
Tim Hill, on behalf of the authors

Major Comments

The issues mentioned above (and below) with the Methods and Results section.
See the changes described in our responses below.

The paper by Brinkerhoff et al 2021 is not cited nor discussed. This was the first paper doing subglacial drainage system emulation and ice flow, thus this is an important context which is missing.
Thank you for pointing out that we had not discussed Brinkerhoff et al. (2021). As listed below, we have added several references to this important work, which we have considered in detail in some of our previous work.

- L521: reference to using surface speed as a data source for model calibration
- L532: corroboration of the equifinality from friction/hydrology tradeoff
- L594: prospects for two-way coupled hydrology/dynamics emulation

What are the Glads parameters corresponding to the mean simulation? In fact, they don't exist as the mean of an ensemble of Glads runs is unlikely a result which could be (re)produced by a Glads-run with some suitable parameters. I don't think this is an issue as, likely, the mean would not be too far off a realisable Glads run with parameters close to some mean of the parameters. So, no need to change the methodology for this but it should be acknowledged.
You are correct that the mean of the ensemble is not the same as any one simulation. We have added a paragraph to Section 2.2.1 to explain our calculation and use of the ensemble mean and median (L155):

> From the 100-member ensemble, we compute the mean and median flotation fraction and effective pressure fields. The mean effective pressure is used in basal friction inversions (described in Section 2.5), sensitivity analyses, and as a reference for presenting results. By integrating out the effect of parameter variations, the mean effective pressure field also provides a simpler test case (i.e., to parameterize how effective pressure varies in space only) that is used to evaluate the importance of each prediction feature (described in Section 2.3) and the sufficiency of the training data. The median simulation is identified as that with median spatially averaged flotation fraction. For the ensemble design described above, the median simulation has parameter values listed in the caption for Fig. 1.

It could be interesting to plot the RF predicted values of floatation-fraction for that median set of Glads parameters (or the mean of all) versus thickness and bed elevation (the two most important parameters), maybe as a contour plot. This could give some insights what kind of function the RF encodes and how smooth it is.

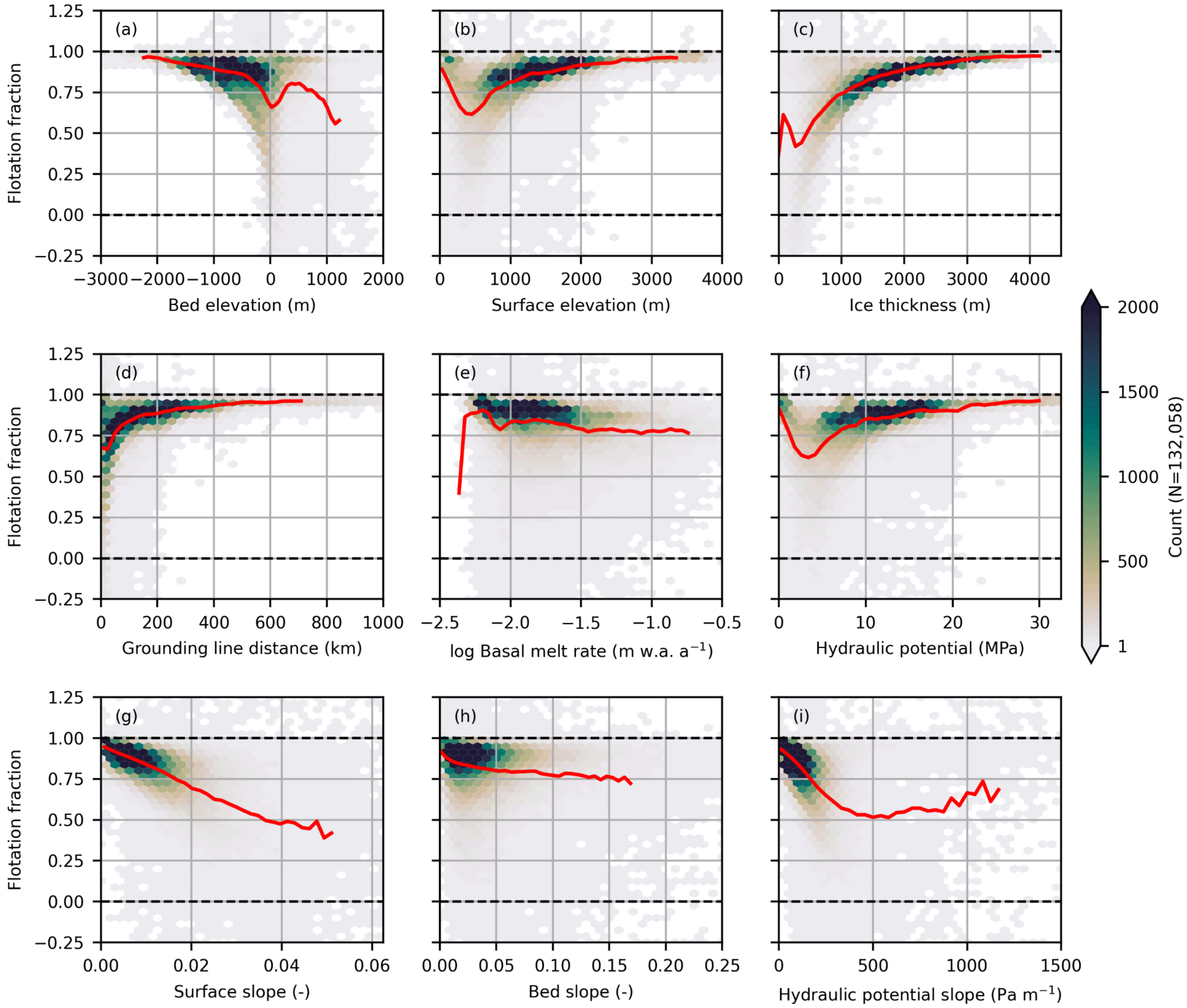

Figure D1. Heatmap of values of the nine geometric features and the ensemble-mean simulated flotation fraction from the present-day GlaDS ensemble mean (GlaDS-1). Simulated flotation fraction values < 0 and >1 are not used for training the random forest. Random forest predictions (solid lines) are binned and averaged according to each feature value to assess the mean trends captured by the model.

This is a good suggestion, we have added the figure copied above to Appendix D (Fig. D1).

Model/calibration runs mentioned in the Results

To aid to resolve my confusion I made a summary of the model runs encountered in the Results section. Many of them are not described in the Methods section (which would be the classic paper-structure). I indicate with "(...)": where model run was introduced and with "[...]" recommendation for update.

Thank you for the suggestions to improve the organization of the methods and results sections. In the following comments, we have summarized how we have incorporated these recommendations.

3.1 Glads present day ensembles (in Methods) []
(No suggestions made)

3.2 RF trained on mean (new) [is this really necessary?]
The RF trained on the ensemble mean is used for sensitivity analyses (Section 3.2). We have added a description and justification for the ensemble mean to Section 2.2.1 (L155, see text copied in response to Section 3.2 minor comment, page 5 of this document).

3.2.1 method-like description of RF geometry-parameter selection (new) [importance scoring should be described in the Methods]; new "production" RF model fitted on only four geometric variables (partially mentioned in methods)
Details about the feature importance metric have been moved to the methods and our intention to drop features with lower importance values has been more strongly worded (L192):

> Anticipating that only a subset of these features will be important for the random forest parameterization, we measure the importance of each feature using a permutation-based metric (Breiman, 2001) that quantifies the decrease in R2 for predictions where the values of one feature are randomly shuffled, and eliminate features that do not meaningfully contribute to predictions to define the final random forest model. This type of backwards elimination of features is commonly automated in order to find parsimonious models where the dataset consists of many features (e.g., >1000; Díaz-Uriarte and Alvarez de Andrés, 2006). Since we have considered only nine features, we eliminate features manually.

On line 341, we have added that "Unless otherwise indicated, results refer to the random forest with four geometric features (ice thickness, bed elevation, Shreve potential and surface slope) and five GlaDS parameters (Table 1)."

3.2.2 (1) RF retrained with leaving basins out (new) [Needs mention in Methods. "cross-validation exercise" is not defined anywhere]
We have added a subsection to explain the cross validation evaluation (Section 2.3.3). See the next response re: sensitivity analysis for the training data.

3.2.2 (2) RF retrained taking "subsets of the training data" (new) [not clear to me what is done exactly. Needs mention in Methods]

We have added a brief mention of these two sensitivity analyses to the Methods (L214): "To evaluate the sufficiency of the simulation data for training, we retrain the random forest model (1) after removing one basin at a time from the training data, to determine if the random forest is unreasonably sensitive to simulations from any one other basin, and (2) on simulations from subsets of 1--6 basins, to explore the number of simulations needed to train the model." We have kept a full description of these tests to the results (3.2.2) on the basis that the core methods, the random forest model and the cross-validation procedure, have now been described more thoroughly in the methods.

3.3 Here, I think, the retrained RF of 3.2.1 is used (previously defined) [be clear what is used]
You are correct, we have made this clear with the added text on line L192 (copied above in Section 3.2.1 comment).

3.4 Here RF is run at 2300 for all of Antarctica, which is slightly confusing as before it was said that West Antarctica is only run in 2050.
Correct, here the RF is run at 2300 for all of Antarctica for the purpose of having one map to represent possible changes in effective pressure from present day to a future date. The RF predictions for WAIS at 2300 are not used quantitatively elsewhere.

3.5.1 (1) describes inversion for ice flow modelling, which is at least partially described in the Methods already (previously defined) [only describe in one place]
3.5.1 (2) forward simulations with using different C and N
3.5.2 future simulations: unclear what glads parameters are used. Figure 7 suggests it's the ensemble of of all. In all it seems 2x2 simulations for the two ensembles and then 2 simulations for POC. Additionally there is the RF trained on mean simulation mentioned in the last sentence.
3.5.3 It seems to me that the simulations described in the first sentence are the same as in 3.5.2. If so this is a bit confusing as it sounds like a new set of simulations is done.
In response to these four suggestions about Section 3.5, we have moved the description of the ice-flow modelling to the methods (Section 2.5) and expanded upon the details that are suggested, including in subsections devoted to the present-day (2.5.2) and future (2.5.3) scenarios.

I recommend that the all the model runs and calibrations are at least mentioned in the Methods and that at least the ones which feature several times are given a label, say GE-pres for Glads-ensemble present day, RF-mean calibrated on the mean of Glads. A table could help too.
As summarized above, we have added text describing all of the GlaDS, RF, and ISSM model runs to the methods section. Where appropriate, we have removed duplicated descriptions to streamline the results section. We have added a table that summarizes the GlaDS and ISSM model runs (Table 2, copied below). This table includes identifiers in the first column ("GlaDS-1", "GlaDS-2", "ISSM-1", etc) to precisely link textual descriptions to simulations. We have not used descriptive labels (e.g., GE-pres as suggested) to avoid overloading labels with the many details that would be required for them to be easily interpretable.

Table 2. Complete list of GlaDS (top section) and ISSM (bottom section) simulations. See Fig. 1a map labelled with ice-flow basin letters. All GlaDS simulations use identical samples of parameter values. ISSM simulations list the effective pressure N used in the friction inversion and that used for the forward velocity solution."RF" effective pressure refers to predictions from the random forest that is trained on all present-day GlaDS simulations. "CV" refers to the random forest cross-validation predictions. "POC" refers to the perfect ocean connection effective pressure.

Label	Model	Ice-flow basins	Ice thickness	Parameter values	
GlaDS-1	GlaDS	B-C, C-Cp, Cp-D, Ep-F, G-H, J-Jpp, Jpp-K	Present day	100 samples of 5 parameters listed in Table 1	
GlaDS-2	GlaDS	G-H	2050	100 samples of 5 parameters listed in Table 1	
GlaDS-3	GlaDS	B-C, C-Cp, Cp-D	2300	100 samples of 5 parameters listed in Table 1	
				N for friction inversion	N for forward solution
ISSM-1	ISSM SSA	B-C, C-Cp, Cp-D, G-H	Present day	GlaDS present	GlaDS present
				RF present	RF present
				POC present	POC present
ISSM-2	ISSM SSA	B-C, C-Cp, Cp-D, G-H	Present day	GlaDS present	RF present
				GlaDS present	RF (CV) present
				GlaDS present	POC present
ISSM-3	ISSM SSA	G-H	2050	GlaDS present	GlaDS 2050
				RF present	RF 2050
				POC present	POC 2050
				GlaDS present	GlaDS present
				RF present	RF present
				POC present	POC present
ISSM-4	ISSM SSA	B-C, C-Cp, Cp-D	2300	GlaDS present	GlaDS 2300
				RF present	RF 2300
				POC present	POC 2300
				GlaDS present	GlaDS present
				RF present	RF present
				POC present	POC present

Minor Comments

Title: What is the difference between "parametrisation", "surrogate model" and "emulation"? The authors have another paper with a statistical model where they call it "emulation", why now "parametrisation"?
The use of "parameterization" throughout the present manuscript reflects how we conceptualize the contribution of this manuscript. The random forest model here is intended to be used as a "parameterization" of subglacial hydrology within an ice-sheet model, just as the perfect ocean connection parameterization is frequently used. Instead of having simple algebraic equations to parameterize a certain physical process (as in the POC model), we propose using the random forest regression model's predictions. Calling the random forest model an emulator or surrogate of GlaDS would also be appropriate in the context of standalone subglacial hydrology modelling. However, the usage of

the random forest subglacial hydrology model as part of ISSM solutions parallels the usage of machine learning “parameterizations” of subgrid-scale processes in climate models (e.g., Fig. 2 and Section 3.1 from Lai et al., 2025), hence the use of “parameterization” in the title.

L2: "in response" could refer to both frictional changes or eff.-pressure
We mean to say that friction (rather than effective pressure) is changing in response to the listed factors and have rewritten the sentence as “These frictional changes are modulated by the effective pressure in the subglacial drainage system and driven by changes changes in ice thickness, basal melt, and slip rates” (2)

L13: always use "perfect ocean connectivity" (likely an issue in other places)
Added “perfect” here and three instances in the Discussion. We have scanned the text and have not found any other instances missing “perfect” but would be happy to correct any specific examples that are found.

L24: Cite https://doi.org/10.5194/tc-12-3931-2018 instead of Fischler
We have changed the Fischler (2023) reference to the original Beyer (2018) manuscript describing the CUAS hydrology model (L24).

L25: Also cite Hewitt 2013
Thank you for pointing out, added (L25).

L43: needs context of emulation based subglacial drainage work. Brinkerhoff & al, Verjans&Robel, other?
We have added a reference to Verjans and Robel (2024) as one example of a subglacial drainage emulator (L39). Since Brinkerhoff et al. (2021) emulate ice flow speed rather than hydrology variables, we have pointed to their important work elsewhere as described in the first major comment.

L53: state clearly that it is a one-way coupling only
Added “(i.e., one-way coupled)”: “Finally, by solving the stress balance using the Ice-sheet and Sea-level System Model (ISSM) (Larour et al., 2012) forced with our effective pressure fields **(i.e., one-way coupled)**, we evaluate [...]” (L54)

L54: here could be a good location to give a bit more an overview over the many methods used in this paper
The preceding paragraph aimed to provide an overview of our desired outcomes and a very brief introduction to the methods (L45–55). Now that we have moved many details from the results section, particularly about the ice-sheet model solutions and created new subsections, we would be happy to consider more specific suggestions for what methods would be a good fit for this introductory section.

L68: a bit more context on what E_creep does would be good. Just one sentence.
We have added the following for context: “ The modulation of basal ice rheology by E_{creep} allows for anisotropic material properties and the concentration of water and debris near the bed. High E_{creep} values may compensate for certain shortcomings in the subglacial drainage model by producing higher water pressures closer to overburden (Hill et al., 2025b)” (L72).

L75: "Present day ice sheet geometry and meshing"
We have changed the section 2.1.1 title to “Present-day ice-sheet geometry” as suggested.

L107: "perturbed-parameter ensemble" is never defined. I find it helpful when papers define concepts like this clearly, name them and then strictly stick to the defined name. This could be done for this phrase but there are other concepts/phrases in this MS too.
You are correct that the perturbed-parameter ensemble is not defined until the following section. We have added a reference to that forthcoming section: “The perturbed-parameter ensemble **(described in Section 2.2.1)** pushes GlaDS near its numerical limit”

In that section (2.2.1), we have rephrased the opening sentence to provide a definition: “For each basin, we run a **perturbed-parameter** ensemble of 100 GlaDS simulations, varying the values of five GlaDS parameters (Table 1).” The remainder of the paragraph explains the ensemble.

L109: "numerical stability limit"
We have added “stability” as suggested: “The perturbed-parameter ensemble pushes GlaDS near its numerical **stability** limit [...]”.

L111: "five-year epochs" not defined.
What we mean to say is that we recalculate the convergence statistics every five years. Procedurally, this is accomplished by running simulations in five-year epochs, restarting from the previous final state. We have clarified what we mean as follows: “For the ensembles, we continue running all simulations, including those that are already converged, in five-year epochs, **recalculating the quantiles of sheet and channel rates of change at the end of the epoch,** until these criteria are met for 95% of the simulations.”

Figure 1: How is the median simulation picked? What median? This needs to be in the Methods. Note that using "Glads simulation" to refer to a simulation output is confusing, in particular in the context of many ensemble runs. This is often encountered throughout the MS.
The median simulation is picked as that with the median spatially averaged flotation fraction. We have added this detail to the paragraph, see text copied at line L155 in response to your major comment about the ensemble mean (page 5 of this document).

As detailed in our response to your major comments, we have added details about GlaDS and ISSM simulations to the methods. We hope that this clarity also helps to answer what simulations we are referring to throughout the manuscript. We would be happy to address specific instances of confusing language.

Figure 1: "cross-validation" is never defined. Note that it features both as "cross-validation evualuation" and "cross-validation prediction", unclear how those are related. There are a number of ways cross-validation is done in statistics, so this needs clarification. In fact, I think there are two types of cross validations done in the MS.
We have added a description of the cross-validation procedure to the random forest methods description, Section 2.3 (L219) that is copied in our response to L267 comment (page 16 of this document).

Figure 1: never defined what errors, absolute, relative?
Errors are the raw errors in units of MPa, we have added the units (MPa) to the caption (also indicated in the colourbar label for the effective pressure error) to clarify this important detail (Fig. 1 caption).

L135: "full-factorial sampling" is (maybe?) too much stats jargon for TC journal?
This is a good point, we have changed “full-factorial” to “grid-based” sampling (L140).

L135: "Samples are drawn from the logarithm of the parameter values" this is not clear. Maybe "Parameter values are log-uniformly sampled."
Changed as suggested.

L136: "and" -> "using"
We mean to say that parameter bounds are chosen based on previous modelling, including large ensembles such as those of Hill et al. (2025) but also other unpublished work. We have added “e.g.,” to clarify our meaning here: “with the lower and upper bounds for parameter values chosen based on previous modelling and large ensembles (e.g., Hill et al., 2025a)”.

L143: briefly state what l_c does
We have expanded as follows: “We vary the sheet-width beneath channels l_c**, which controls the portion of sheet flow contributing to channel melt,** widely around the commonly used value of 2 m.

L155: 100 samples again? What is the melt water input?
Correct that we use 100 samples again. We use the present-day basal melt rate to avoid circularity issues with using a projected future basal melt rate, which depends on the slip speed and therefore effective pressure. These details have been added: “This ice-sheet geometry is combined with present-day basal melt rates to run a future GlaDS hydrology ensemble. The ensemble consists of the same 100 samples as for present-day conditions and the four basins used for ice-flow modelling (Section 2.5).” (L173).

Section 2.3: this needs to be much more careful.
In addition to the additions described below, we have organized this section under the following headings:
2.3 Random forest subglacial drainage parameterization
2.3.1 Prediction features
2.3.2 Random forest architecture
2.3.3 Cross-validation
2.3.4 Sufficiency of training data

We would be happy to address other specific details that we may have overlooked.

L160: specify that it is a "spatial pointwise" mapping
We have specified that the random forest is a “spatially pointwise” mapping, as suggested.

L162: as far as I can tell, the RF trained on nine ice-sheet variables and the full ensemble is never used.

Correct that we do not use the RF that is trained on the full set of 9+5 features and that this is intentional. We have now made this intention with the added text copied in response to Section 3.2.1 Major Comment (p. 7 of this document).

L163: "forcing fields" undefined
We have changed "forcing fields of the subglacial hydrology model" to "basal melt rate used to force the subglacial hydrology model" (L190).

L166: "... importance values for the final parametrisation." but come up with a better name for the "final" parametrisation.
See L162 comment.

L175: Brinkerhoff & al 2021 needs to be discussed here
Brinkerhoff et al. (2021) do not emulate hydrology variables and we have added a discussion of their approach as one way to include two-way coupling among other references (see first major comment). The Verjans and Robel (2024) work is the only published neural network emulator of a subglacial hydrology variable that we are aware of.

L176: Not sure a neural net is harder to integrate into ISSM. It would be, for instance, differentiable and thus better for control-method inversions.
There are several ways to integrate the random forest model into ISSM. One simple approach is to automatically generate C code for all if/else conditions in each decision tree (or for one such tree). While a large volume of code, this is not intractable as generating similar code for a CNN would be. It is true that a differentiable model would be advantageous for control-method inversions. However, ISSM is not integrated with a compatible machine-learning library at the moment. This problem is surely surmountable, but these points lead us to believe that the random forest that requires less software machinery would be easier to integrate.

L178: This is a separate drainage model and warrants a separate section
As suggested, we have added the new section 2.4 "Perfect ocean connection effective-pressure parameterization" here.

L180: "physics-based model" What is this? Is it Glads or POC? Just say it.
"Physics-based model" has been changed to "GlaDS" (L237).

L204: There are 700 scenarios here for both Glads-runs and RF-runs, which one?
We use the ensemble mean for the friction inversions. We have added this detail to this paragraph, a few sentences after the highlighted line: "In the case of GlaDS and the random forest parameterization, we use the mean of the perturbed-parameter ensemble for friction inversions" (L262).

L205: "friction coefficient fields"
Changed as suggested.

L215: state briefly why also log-misfit is used

Log-misfit is used to match observed speeds that vary over orders of magnitude: “Inversions use **squared-error** and log-misfit cost functions to fit observed surface speeds **that span orders of magnitude** [...]” (L269)

L215: Eq B2 states abs-squared misfit
We have changed “absolute” to “squared-error” for precision (see text copied in previous response).

Section 3.1: here "ensemble" is used both in plural and singular. The authors need to come clear on how they want to name it and then be consistent throughout. (could be an ensemble for each catchment or just one ensemble)
We have checked the usage of ensemble(s) throughout and have decided to use “ensemble” to collectively refer to all 700 present-day simulations or all 400 future simulations. “Ensembles” is used to refer to all 700 + 400 simulations. This reflects the idea that the ensemble is defined as the 100 samples of GlaDS parameter values and that each row of the new Table 2 is an “ensemble”. We have changed the usage of ensemble(s) throughout.

Figure 2: I don't understand how flotation fraction is capped at 1 but eff.-pressure can go negative (most obvious in panel c vs f)
Thank you for highlighting this inconsistency. In producing this figure, we were accidentally capping the maximum flotation fraction at 1 after computing the effective pressure, whereas we should be plotting the simulated values with no cap. We have fixed this inconsistency by removing the flotation fraction cap and updated the figure (copied below) in the manuscript.

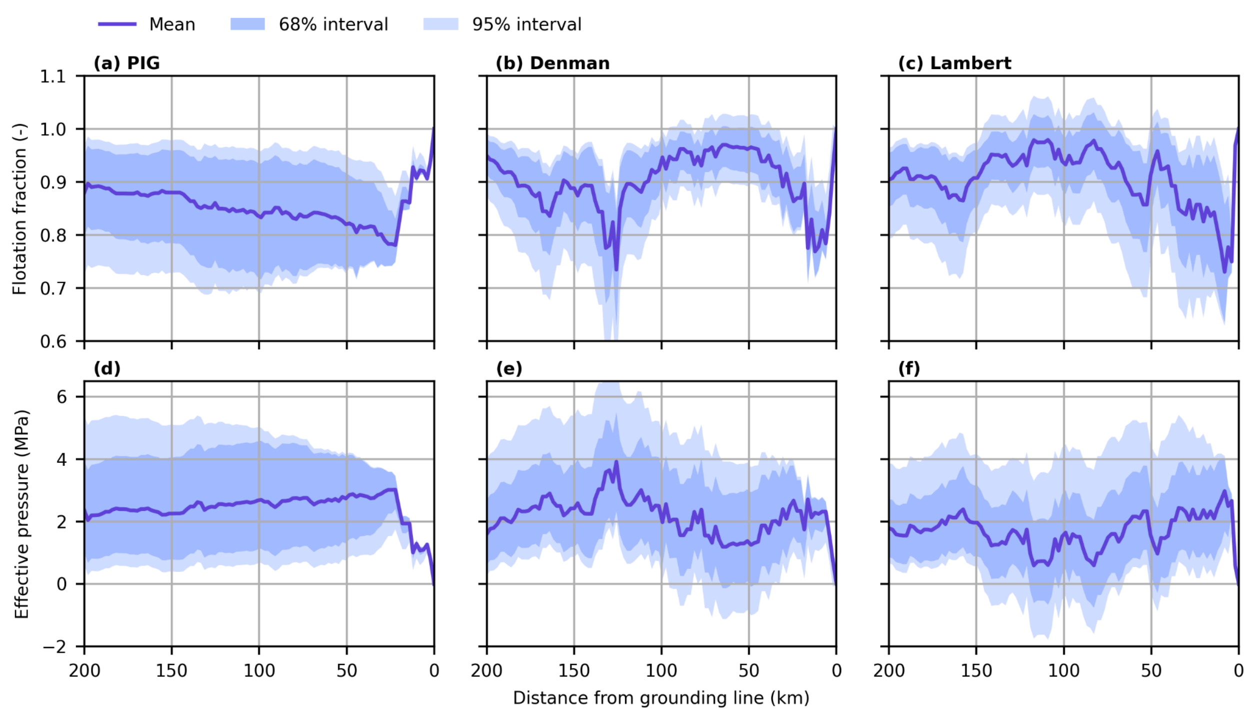

Figure 2: delete "language=en"
Deleted.

Section 3.2: a new model run setup is presented here ad-hoc (typically that is done in the methods). This would not be the end of the world but is really symptomatic of this MS: almost every section in the results introduces some new model setups and runs. I found this really hard to ingest.
Thank you for pointing out this example. We have added a paragraph to the methods (Section 2.2.1) where we introduce the ensemble mean and its purpose. See the text copied in our response to the Major Comment re: ensemble mean (p. 2 of this document).

L242: "mean" is never defined
See the text copied in our response to the Major Comment re: ensemble mean (p. 2 of this document).

L243: according to a later section (3.6), the training of the RF is quick, so this argument does not hold.
While we could use the complete RF here, the cross-validation evaluation including GlaDS parameters that you are suggesting (which involves training the RF 140 times as now described on line L326) would make this the most expensive application of the RF (according to our calculation, ~25 CPU-days). Since we don't believe that including GlaDS parameter variations would change the conclusions of this section, we have left the computational reasoning while foregrounding the reasoning to isolate geometry: "This sensitivity analysis uses a random forest model trained on only the mean of the 100 perturbed-parameter simulations to isolate the role of the geometric features and for computational reasons."

L255: this 1.8 and 18 must be a factor? If so state. But as R2 can go negative, an absolute change would be better to state. Same for Fig 3
These are both absolute changes in R2, not relative changes. We have added an explicit definition of R2 (Eq. 1) to the methods section 2.3 (Random forest subglacial drainage parameterization) and expanded the explanation of what R2<0, R2=0, and R2=1 mean to help the reader understand how we obtain such negative R2 values (L230). See text copied in response to L267 comment.

We agree that changes as large as -18 and the implication that R2=-17 are surprising at first. We have checked these results numerous times and they hold. Realizing that the random forest primarily uses ice thickness for predictions, consider what happens when the ice thickness values are shuffled. Small ice thicknesses, which usually indicate we are near the coast and are related to low flotation fraction, can be shuffled into the interior. For these areas, the effective pressures for the shuffled predictions are close to the ice overburden pressure, rather than being ~5–10% of ice overburden. Since the residuals for the shuffled effective pressure predictions (0–100% of overburden) are much larger than the range of simulated effective pressure (5–10% of overburden), we end up with R2 of about -17.

L258: odd indeed that the Shreve potential, which is the sum of bed and thickness, is also needed. Have the authors tried to use effective pressure in units of meter H2O instead of Pa? This would "scale" the quantity better, or does RF not care about scaling?
We agree that it is interesting that the Shreve potential, ice thickness, and bed elevation are all used, for the reasons that you point out. However, we caution that we do not say that the Shreve potential is needed, rather that the random forest in this configuration weakly uses it for prediction. As with the test where we exclude the ice thickness field (Fig. D4), it is possible that the Shreve potential could also be excluded

with minimal impact on the prediction quality. We have not exhaustively tested this possibility, and since the Shreve potential is easy to compute, it does not harm the model to retain this feature.

There are two reasons that changing the scaling from Pa to m would not change the results: (1) we train using flotation fraction (now described on line 212), and (2) the random forest is not sensitive to a scalar multiplication. In fact, we standardize the flotation fraction values to have standard deviation of 1 before the training.

L262: "... retrained reduced model compared to the previous model (Fig 3a versus 3b)"
We have changed the figure reference to "(comparing Fig. 3a, b)"

L267: "repeat", as far as I can tell no previous cross-validation exercise was done (apart form a mention if Fig 1)
You are correct that we had neglected to describe the cross-validation procedure. We have added a description to the random forest methods section (2.3.3, line 218), copied below. With this description in place, we hope it comes across that we "repeat" the same cross-validation exercise as described in these lines. We have also added the caveat that the cross-validation is only done "basin-wise" on L346.

> The random forest model is evaluated using cross-validation with respect to the ice-flow basins and GlaDS parameter values. Taking one of the seven basin at a time, we leave out simulations from that basin and perform 20-fold cross-validation with respect to the GlaDS parameters (i.e., removing 5 randomly selected simulations at a time from the training data). Predictions corresponding to the subset of data omitted from training (5 parameter values for the omitted basin) are stored for comparison against the GlaDS simulated values. In each of the 20 iterations for each basin, the random forest model is retrained on the 95 simulations in the six basins that have not been omitted. This procedure is repeated 140 times to cover all 100 GlaDS parameter values and 7 ice-flow basins.
>
> The cross-validation predictions are evaluated by the coefficient of determination R2 that measures the proportion of variance of the GlaDS model outputs that is captured by the random forest parameterization,
>
> $$R^2 = 1 - \frac{\mathrm{SS}_{\mathrm{res}}}{\mathrm{SS}_{\mathrm{tot}}} = 1 - \frac{\sum_{i=1}^{N} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{N} (\bar{y} - y_i)^2}$$
>
> for residual sum-of-squares $\textrm{SS}_{\textrm{res}}$ (i.e., the squared error that is minimized by fitting the random forest) and total sum-of-squares $\textrm{SS}_{\textrm{tot}}$. Random forest predictions are indicated by $\hat y_i$, GlaDS-simulated data by y_i, and $\bar y$ is the simulated mean. Under this definition, a perfect model would obtain $R^2 = 1$, and $R^2=0$ indicates a poor fit with none of the GlaDS variations captured by the predictions. Negative R2 values are possible and indicate that the predictions $\hat y_i$ have greater residuals than would be obtained by predicting the mean $\bar y$ of the simulations as a prediction.

L269: "each" -> "one"

We have changed “each” to “either”: “The strongest sensitivity is found for Recovery basin (Jpp-K), which has improvements of $\Delta R^2 = 0.12$ when excluding **either** the Amundsen sea (G-H) or western Filchner-Ronne (J-Jpp) basins [...]” (349)

L275: state what "quantity", is it the number of catchments?
The description of the basin sensitivity experiments has been rewritten and expanded in 2.3.2, 3.2.1 and 3.2.2.

Table 3: the "u" column is not described/used in the text, I think
The speed column (u) of Table 3 is referred to in 3.5.1 “Present-day ice flow” to quantify the similarity of ISSM ice-flow solutions that use different combinations of basal friction field and effective pressure. Thank you for checking that all the information we provide in tables is discussed in the text.

We have also moved considerable detail about the ice-sheet model solutions to the methods section. We hope that this column will be more accessible now that these methods have been more clearly defined.

Table 4: Not clear how the comparison is really done. Is RF run for all parameter combos for which Glads was run?
Yes. We have moved details about the future ice-sheet model solutions to methods, including that we compute RF predictions for all 100x GlaDS parameter values for comparison.

L298-308: isn't this one of the key results of the paper? Should this not be given more room? Would it not fit better into the previous section?
The listed line numbers span a few different aspects of the results. If we are understanding correctly, you are pointing out the performance of the random forest for the future ice-sheet geometry as a key result of the paper. We have more strongly highlighted this result by restructuring the headings in the section as follows:

3.3 Evaluation of the random forest parameterization
3.3.1 Accuracy for present-day geometry
3.3.2 Accuracy for future geometry

This fact has been moved under 3.3.2 (line 385).

L306: here "perturbed geometetry" term is defined. But instead it should be defined in section 2.2.2
We have added a definition for “perturbed geometry” to section 2.2.2: “The thickness changes are applied to the BedMachine Antarctica v3 initial condition (Section 2.1.1) to compute a **perturbed ice-sheet geometry** representative of plausible future conditions.” (L169)

L311: delete "."
Deleted as suggested.

L321: not clear what is done here. I think that this is already described in the methods, but as so many extra model runs & setups are described ad-hoc in the results, I am not sure. Is this using the mean?

Thank you for highlighting that our steps here were not clear. As detailed in our response to your major comment, we have moved the details about these runs to the methods section, removed duplication, and tabulated all numerical model simulations in Table 2.

You are correct, these results refer to the ensemble mean. Line 416 now says this explicitly: “Since the inferred friction coefficient compensates for differences in effective pressure when varying the parameters of the subglacial drainage model, the results in Fig. 6, presented for the ensemble mean, should be qualitatively similar for other ensemble members.”

L344: not clear what effective pressure fields are used. Mean, ensemble?
Yes, we are using the mean here and have added this detail: “Here we examine future ice flow for the ensemble-mean hydrology scenario” (L422)

L344: "using the corresponding friction coefficient inferred from present-day conditions and surface velocity observations" is a repetition of the previous sentence.
Thank you for pointing out our duplication of information. As detailed in our response to your major comment, we have moved these details into the methods in Section 2.5.3 “Future ice-sheet model solutions” in an effort to reduce duplication.

L363: "unconstrained friction coefficient", what is this?
We have added an explanation that this is a result of gaps in the surface speed data: “excluding Totten which has **incomplete surface speed observations** and therefore an unconstrained friction coefficient in the region of the retreated grounding line.” (L439)

L366: state what C is used
See our response to your major comments on pages 5–7 of this document. We have reorganized the ice-flow model description and results sections in order to reduce duplication while making it unambiguous what C and N fields are used.

L366: "grounding line speed" is unclear. ice speed at the grounding line? Also L431
We have now defined these instances more precisely as “speed at the grounding line” (lines 12, 427, 510, 675). We have kept other instances of “grounding-line speed” for brevity where they closely follow this longer definition, as these added definitions should provide context to make the meaning unambiguous.

L381: Do the Glads simulations run in parallel on the 32 cores of this CPU or single core?
Each GlaDS simulation runs in serial on a single CPU core. We have added this detail, noted that the ensembles are trivially parallelized, and clarified that random forest benchmarks refer to the same hardware: “GlaDS simulations take ~16–97 hours to run per five-year epoch on a single AMD EPYC 7532 CPU **core**, with total runtimes between 32–388 hours to reach steady state as defined in Section 2.1.3. **GlaDS simulation ensembles are trivially parallelized, with each simulation assigned to its own core, to minimize wall-clock time. On the same single CPU core**, [...]” (L460)

L382: Assuming Glads runs on a single core, the runtimes to steady state and number of runs suggests that this CPU would have needed to run for almost half a year to complete the Glads runs (11 ensembles x 100 samples x ~100h runtime / 32 cores ~ 140 days). Is this correct? Elsewhere 16days are stated.
The total wall-clock time is ~16 days since the 11x100 simulations are spread across an equal number of CPUs (as described in the added text above), hence the upper bound of 16 days for the timescale of GlaDS runs.

L388" "ice dynamics"
Changed as suggested (ice-flow → ice dynamics, L441)

Figure 6: for SSA "depth averaged" and surface flow speed are the same. I suggest to only use the latter as that is what is also measured.
Correct that there is only one meaning of "speed" for SSA since there is no vertical shear. "Depth averaged flow speed" has been changed here to "Ice flow speed" for consistency with Fig. 7 caption and text.

Figure 7: not clear what "sensitivity" means. This should be defined somewhere
To be more precise, we have changed the first sentence of this caption to read "Hydrologic changes relative to present day and ice flow speed for representative future ice-sheet geometries" (Fig. 7 caption). We have made the same change to the Figure C1 caption.

Figure 7: not sure what the bands refer to. Ensemble, something else?
Correct, the shaded areas indicate spread across the ensemble. We have added this detail to the caption: " Shaded areas indicate the central 68% of values **across the perturbed-parameter ensemble**" (Fig. 7 caption).

Figure 7: is there a difference between "perturbed-parameters samples" used here and "perturbed-parameters" used elsewhere?
This sentence is referring to the same set of 100 simulations. We have rephrased to ensure we use consistent language: "The vertical bars indicate the spread of the central 68% (thick line) and 95% (thin line) speeds from the GlaDS perturbed-parameter ensemble and the corresponding random forest predictions" (Fig. 7 caption).

L395: I suggest to use "momentum balance" everywhere
This is a good suggestion since the equations being solved arise from the conservation of momentum. We have changed instances of "stress balance" to "momentum balance". Note that we still refer to individual stresses (shear stress, longitudinal stress coupling) in L501.

L433: how can a glacier accelerate by 25 years?
We have clarified that we are referring to how hydrology–dynamics feedbacks accelerate the timing of retreat, not the glacier itself: "[...] accelerate the **timing of rapid retreat** of Totten Glacier by 25 years"

L441: also cite https://doi.org/10.1017/jog.2020.116 and https://doi.org/10.1029/2018JF004921

We have added a brief discussion of these and Brinkerhoff et al. (2021): “Calibration using surface observables (Brinkerhoff et al., 2021) and comprehensive datasets including tracer breakthrough data (Irarrazaval et al., 2019, 2021) may provide additional constraints but share the irreducible uncertainty in calibrated parameter values”

Section 4.2: discuss Brinkerhoff & al here. In particular, they fit scalar parameters and not fields like C here, which would be interesting to contrast.

Good suggestion. Following the discussion of the relationship between GlaDS parameters and the magnitude of dynamic changes in the future scenarios, we have added the following text describing similar findings by Brinkerhoff et al. (2021) for present day (L531):

> These tradeoffs between the friction coefficient and GlaDS model parameters are illustrated in the joint friction--parameter distributions inferred by Brinkerhoff et al. (2021) for a two-way coupled hydrology and dynamics model. In that work, negative correlations between the friction coefficient and the sheet and channel hydraulic conductivities show that similar present-day speed solutions can be obtained with different drainage system configurations because of the compensation of the friction coefficient.

L481: fast to train once the Glads simulations are run

That is correct. This sentence is contrasting the random forest with the CNN used by Verjans & Robel, which would share the upfront cost of GlaDS simulations.

L590: what about the outlook of doing combined ice-flow-drainage inversions? Would this be feasible with RF+ice-flow?

This is still challenging, but the methods that we illustrate could be extended for various inversion tasks. One main limitation would be the lack of two-way coupling here (see line 591).

If you mean joint inversions for the basal friction field to improve model initialization, this could perhaps best be solved with tightly two-way coupled physics based models such as the Brinkerhoff et al. (2021) implementation, where hydrology and ice dynamics equations are solved simultaneously to obtain the friction field. While this capability exists (D. Brinkerhoff, pers. communication, June 2024), it was not illustrated in that work. Or, perhaps this problem would be solved better by a neural network hydrology emulator with the built-in differentiability.

If you mean inversion for scalar model parameters as in Brinkerhoff et al. (2021) and Hill et al. (2025, hydrology only), then yes, we believe this would be feasible, as illustrated by those emulator-based works.

To avoid oversimplifying these problems, and since these topics are beginning to stray away from the evidence that we have to support arguments, we have not added these details to the text. We hope that questions such as these initiate further work in these areas.

Table A1: The caption does not define the shown quantities well enough

We have added that “The mean discharge Q, used to compare against GlaDS simulation ensembles, is the average over the 1--3 listed values reported in the literature.”

Table A1: https://doi.org/10.1016/j.scitotenv.2024.172144 gives 51m3/s for PIG

Thank you for providing this source that we missed. We have added this source to Table A1, recalculated the mean literature discharge (was 42.5 m3/s, now 45.3 m3/s) and correspondingly adjusted the PIG panel of Fig. A1.

Figure A1: "Histograms of ..." from what ensemble is this? Present day?

We have added that these are discharge values from the present-day GlaDS simulations (Fig. A1 caption).

L615: not clear what Glads parameters or if for ensemble

As explained in the previous sentence and in the caption, Table C3 is reporting results for the ensemble mean. To ensure consistency, we have changed all instances of “ensemble average” to “ensemble mean”, including here (L697).

Figure C1: caption suggests that column C should have 8 lines in each plot

There are six lines corresponding to three effective pressure models (GlaDS, RF, POC) in static/evolving configurations.

References

Brinkerhoff, D., Aschwanden, A., & Fahnestock, M. (2021). Constraining subglacial processes from surface velocity observations using surrogate-based Bayesian inference. *Journal of Glaciology*, *67*(263), 385-403. https://doi.org/10.1017/jog.2020.112

Lai, C. Y., Hassanzadeh, P., Sheshadri, A., Sonnewald, M., Ferrari, R., & Balaji, V. (2025). Machine learning for climate physics and simulations. *Annual Review of Condensed Matter Physics*, *16*(1), 343-365. https://doi.org/10.1146/annurev-conmatphys-043024-114758

Verjans, V., & Robel, A. (2024). Accelerating subglacial hydrology for ice sheet models with deep learning methods. *Geophysical Research Letters*, *51*(2), e2023GL105281. https://doi.org/10.1029/2023GL105281