

Review of “Towards more realistic calving simulations: the effects of subglacial hydrology on stochastic calving at Sermeq Kujalleq (Store Glacier), west Greenland.”

Summary:

This paper implements a stochastic calving parameterization into a realistic glacier model of Sermeq Kujalleq (Store Glacier) with the aim to test whether the parameterization can capture statistical calving patterns and the impact of coupling the parameterization to a subglacial hydrology model. The study varies the seasonal state (summer versus winter), a parameter of the calving law “ k ”, the coupling or not with subglacial hydrology, and the random seed of the stochastic calving law among five simulations. Some of the broader conclusions of the study have the potential to have a large impact on the field. In particular, the idea that including a more realistic seasonal basal water pressure in the calving parameterization helps the model capture seasonal patterns in terminus change in a way that including mélange and seasonal fjord temperatures does not. However, it is very difficult to be confident of many of the findings throughout the paper given that 4 variables change for 5 simulations, with one variable (the random seed) never being held constant. The changing random seed makes it impossible to confidently tease apart the impacts of any of the main three modeling choices this study addresses. I strongly suggest either rerunning the simulations with a) the same random seed or b) an ensemble approach. Overall, the paper is thorough, has some compelling explanations, and very interesting findings. It would greatly benefit from further explanations of choices and methods and quantification of some of the results, but the point above needs to be addressed first and foremost.

General Comments:

1) This study hinges on comparing five simulations with varying parameters for hydrology and calving. The random seed of the stochastic calving parameterization; however, appears to vary for all five of the simulation, such that confident attribution of a result to the model parameters is extremely difficult. Though it could be argued (and maybe shown) that the integrated impact over time is the same, the time series and spatial details discussed throughout the paper are likely to be dependent on the random seed. We can see that the random seed leads to a few very large events (which of course is also dictated by other things in the model that are also being varied, like hydrology) in Figure 5. If the time series of the probability, p , is plotted, I would expect to see a couple of large spikes corresponding to the larger circles in the righthand column of Figure 5. There are three large events in SK24U that appear to affect the results of Figure 7 panels g, h, and i. This leads me to question the interpretation of the results because I cannot be confident that they are not a result of different time series of p .

Lines 421-424 confirm that random, large events can have a large impact on the results, while lines 390-399 suggest that the results are in large part determined by the random seed which determine when the large events occur. As lines 553-555 of the paper appear to suggest, definitive causal links cannot be established in all cases due to the stochasticity and lack of a random seed. The authors suggest that many of the results are unlikely due to coincidence in the model behavior, but I would argue that, in many cases, it cannot be ruled out and therefore leads me to question the robustness of the results. I note these instances in the specific comments below, but give a couple examples here of results I think could easily be due to the different random seed rather than the stated or implied cause:

- a. Lines 315-320: “[...] uncoupled simulations have relatively fewer larger events [...] Fully coupling GlADS to the calving function [...] therefore seems to be a key ingredient in reproducing observed calving behaviours [...]” This could be a result of unlucky draws or one unlucky draw that easily skews the means across just 2-3 simulations. There is good rationale for why uncoupled simulations should have fewer larger events; however, k was adjusted in the coupled simulations to account for this, so it is not entirely clear by theory alone that uncoupled simulations will necessarily have fewer larger events.
- b. Lines 379-381: “The uncoupled equivalent - WK24U - advances a little too far, showing that explicit modelling of the basal water pressure [...] is necessary [...]” Could this not simply be a poor “luck of the draw” that WK24U didn’t get enough large p values to trigger larger calving events before the simulation was over whereas the coupled equivalent did get more large p values just due to the random seed?

The published companion paper, Benn et al. (2026) does a 3-member ensemble for an idealized Greenland-like glacier with the only changing variable being the random seed. They also use a similar k ($k=22$) value for their Greenland set up, which I expect means their set up is comparable to this study. They show that there are general similarities (each capture the overall seasonality), but there are huge differences for any given 3-month period (Benn et al., 2026 Figure 13). They show that for smaller calving events, these differences do not greatly impact the cumulative quantities, like total calving volume, for multi-annual timescales (Benn et al., 2026 Figure 15c); however, for shorter timescales and/or locations with larger, more infrequent events, even the cumulative quantities can be greatly impacted by the random seed (Benn et al., 2026 Figure 15c-d). These results suggest that the random seed cannot be confidently ruled out as a cause for differences between any of the 5 simulations in the present study.

To fix this issue, the simulations should be given the same random seed. An arguably more robust approach would be to use ensembles for each of the five parameter combinations, with the size of the ensemble being justified statistically. The paper acknowledges the ensemble approach as the preferred approach in lines 190-195, but cites computational expense as an issue (which could be circumvented by seeding each of the 5 simulations). The purpose of the paper is also cited for why an ensemble approach was not used: “[... 1] to demonstrate that the model is capable of accurately reproducing general calving behaviour at Store and [2] how this is affected by hydrology and k values [...]”. The first purpose could be achieved with most any combination of simulations, but, for the second purpose, I have not been convinced that this can be confidently achieved without seeding the random variable or taking an ensemble approach. An alternative to these approaches is to accept that less can be concluded from the given 5 simulations and to rewrite the paper accordingly. This option would unfortunately miss out on the opportunity to convincingly come to some interesting conclusions about subglacial hydrology and calving that could be transformative for future modeling studies.

2) Some choices need to be better justified, methods explained, and results quantified. The specific comments point out the cases I found, which were numerous enough to require an additional note here.

Justifying choices: Several choices made in the setup of the study, for example, the study region and the examination of the role of coupled subglacial hydrology, are stated but are not justified.

Explaining methods: In particular, an explanation is needed for how the value “ k ” was chosen for the coupled models. Currently a vague statement is supplied in the methods, and some rationale is given in the discussion, but it is unclear what observations were compared to, what range of “ k ” values were tested, and what optimization method was used.

Quantifying results: Most of the results reported are qualitative descriptions, which can be appropriate for identifying similarities in patterns (because trying to exactly match the observations is not the point here, which was well justified by the authors). The stochastic method, however, is touted as being able to match statistical properties of calving, yet very few explanations of statistical properties of the observations and model results are given in the results and discussion. There are some qualitative statements that are also just not obvious or do not seem to match the referenced figure (e.g., lines 352-354 explained more in the specific comments). Quantitative results would help make these statements clearer and more convincing.

Specific Comments:

Lines 50-52: This sentence only describes two statistical populations, please clarify.

Line 66: Why is Sermeq Kujalleq (Store Glacier) the focus of this study? This choice is addressed in part in the discussion (e.g., the ability to compare floating and grounded terminus regions with the same glacier), but it should be motivated in the introduction or study site sections.

Line 68: The case for including a coupled subglacial hydrology model becomes apparent later in the manuscript, but the idea is not adequately motivated in the introduction.

Lines 68-74: This seems like a summary of some of the findings of this study rather than a literature-based introduction to the study. Either add citations, clarify, or move these two sentences.

Figure 1 caption: Which MEaSURES dataset? Please specify and cite.

Line 92: Why was 2012 chosen?

Line 193-194: “how this is affected by hydrology and k values” will be difficult to confidently evaluate without a random seed or ensemble approach.

Lines 197-200: Describe how the fit was optimized – e.g., which observations were used, what method of comparison, and what values of “ k ” were tested. Also explain here (rather than in the discussion) why a different value of “ k ” is needed for the coupled model and why it is “wrong” to use $k=20$ for the uncoupled model.

Lines 204-208: Why was the 2012 model set up from Cook et al. (2021a) used instead of the 2017 model set up that would overlap both observational datasets compared to in section 4.1? I understand this modeling effort is not trying to perfectly reproduce observations, but the fact that there’s an opportunity to have the model set up and the observations for the same year seems like a missed opportunity. Please justify this choice.

Lines 228: This is a discussion point, not a result. This is also a conclusion based on the behavior of one stochastic simulation. How can we be confident that this is a result of the water pressure and isn’t an artifact of the random seed?

Section 4.1: The qualitative comparisons are important and interesting; however, the introduction explained how the stochastic calving parameterization reproduces the statistical properties of observations (Lines 50-60). Do these simulations reproduce the statistical properties of the observations?

Figure 2a: Add the $1e6$ unit to the y-axis or change both x- and y-axes to units of km for easier interpretation. Whatever is chosen, make this consistent across other figures (e.g., Figure 4).

Figure 2b: It would be helpful to see examples of the summer and winter data plotted on the same X scale (and x-axis) as the models. The way this is currently presented shows two interesting plots but does not let the reader really understand the similarities and differences between expected observed patterns and the patterns in the model.

Figure 2b: Please plot SK20U. It is informative to see all the results and helps the reader understand why it is “wrong”. As the manuscript states, the goal is not to match the observations but to understand the impact of k and subglacial hydrology. Therefore, it is more informative and important to show the results of ALL the simulations. Also, as I said before, the explanation for why the k value and basal water pressure do not match needs to be given in the methods section, otherwise this statement comes as a surprise and does not make sense until later in the manuscript.

Figure 3: All figures should show North towards the top of the page. Otherwise, it is extremely difficult for the reader to compare the different plots and types of observations. These types of plots also need a legend or explanation for the size of the circles.

Figure 4:

- It would be helpful to see similar plots showing examples from the Black et al. dataset for comparison.
- It is difficult to distinguish the black from the red lines. I also do not see the black lines discussed in the text. Are they necessary to include? Could they be better distinguished with dashed or dotted lines?
- Panel f: Impossible to read the color and scale bar labels. What does the purple indicate?

Figure 4f: Are all the GlaDS basal water pressures the same at the start of the simulation for all the models of the same season? If not, why not and what impact does this have? Why is it important to show the difference after one day? Do all uncoupled models have the same basal water pressure for the crevasses from the Todd et al. parameterization at the start of the simulation? If not, is it because of the spin up or something else and what impact does it have?

Lines 266-267: It would be helpful to remind the reader here that “coupled” refers to GlaDS output being used in the calving parameterization and that all models use the GlaDS output for the basal sliding law and the plume model.

Lines 271: This fits perfectly with expected behavior from what observation source? “Fits perfectly” seems like a generous interpretation given that the differences in these two simulations look qualitatively small. Figure 2 shows that the mean position of these two simulations follows a similar pattern and Figure 4 shows slight retreat in the North and larger

advance in the South with minor differences in the details. The small differences could easily be due to the different random seeds.

Line 273: The following sentence does not back up the statement that “similar behaviour” is observed.

Lines 274: What figure or observation shows that SK24U is “broadly within the stochastic range of summer behaviour”? Was this behavior quantitatively characterized? The cited Figure 4 doesn’t have any observations, so it does not back up this statement.

Line 276: “wrong” needs to be defined earlier and reiterated (as is done) here in lines 278-282. It also is a misleading word, so I would recommend finding another word to describe the pairing of $k=24$ and the coupled model. Perhaps “uncalibrated”?

Line 314-320: The conclusion here is interesting (and likely belongs in the discussion rather than results section), but it is based on simulations that were not seeded, so it is difficult to know whether this result is due to the coupling or the random seed. Even with a seed, the role that changing “ k ” has on these larger events should also be discussed.

Line 329: Cite and/or provide the “definition of a calving cascade”.

Figure 5: Flip the figures so that North is towards the top of the page consistently across the figures (and as is more intuitive for readers).

Figure 6a: It is difficult to distinguish the different blue scales. Though the details of this plot are not key to understanding the study, it is frustrating to not be able to distinguish the channels from the discharge. Also show the red dot in panel a for easy comparison.

Figure 6c: Text is too small to read. Also, there is a very large jump in the summer total calving volume around time step 70. Figures 5a and 7h make it difficult to believe a large calving event occurred at this time stamp. Figure 7i makes it seem possible, but uncertain. Are these the correct models or quantities being plotted?

Figure 7: Text is too small and the subplot labels overlap some of the x-axis numbers. Define relative abundance. Are total abundance and total calving volume the same? The figure label and caption use different terminology.

Figure 8: Here and elsewhere in the text, be clear it is an inference or hypothesis that the sustained melting led to the calving events.

Lines 352-354: I do not come to this same conclusion looking at Figure 5. Please quantify the results to help back up this conclusion.

Lines 371-373: It is not obvious that the clustering of calving events in the observations and model would be due to two different reasons. If so, then how can it be argued that the model is reproducing real-world calving behavior as is stated in the following sentence?

Lines 376: “match” is an odd choice of words. What about k can match a basal water pressure field? As before, how k was chosen for the coupled models needs to be explained more thoroughly, but I expect the process would be better described as “tuning”.

Lines 379-381: This result could simply be due to the luck of the draw with the random seed.

Lines 407-409: As stated above, this could have been 'luck of the draw'. There are compelling arguments why it is real, but this needs to be shown beyond any doubt or be qualified with some degree of uncertainty throughout the manuscript.

Lines 413-414: Without an ensemble or a set random seed, how can we definitively know this difference is due to the higher basal water pressure?

Lines 415-416: Could the need to lower the value of k imply GlaDS has too high of water pressure? What are the tradeoffs between changing k and changing the water pressure? This is another result that is very difficult to disentangle from the changing random seed.

Line 424: If the rare, large summer events drive the difference, then the random seed will be very important, especially for relatively short simulations.

Lines 442-444: Summer and winter are normally thought of as 3-month periods, which these simulations are also. Should that not be long enough to see the differences? Summer and winter could be thought of as 6 months, but they certainly are not 12 months. If this parameterization can capture calving statistics, then it should be able to show real summer-winter differences in 3 to 6-month long simulations.

Lines 444-446: It is possible this result could also be due to the random seed being different. Lines 449-451 are a reasonably compelling explanation; however, the explanation does not replace the need for the random seed being set or an ensemble approach.

Line 474: This “would be” or “is” in line with theoretical and observational considerations? Please cite any appropriate prior work here.

Lines 476-478: A statement about confidence in capturing a real feedback must be backed up by evidence.

Lines 497-498: Two examples are evidence for but do not definitively show k is not dependent on model domain or other small changes in the model setup. This is a very interesting result, however, if it is true.

Lines 499-502: How was it determined that model resolution is the important factor?

Lines 503-505: How is the 'calibration' approach here different than a standard 'tuning' approach? Is one based on statistical properties while the other matches positional observations? Do these differences cause the approach in this paper to better fit the definition of 'calibration' over 'tuning'? Please clarify.

Lines 514-516: What is the margin of error for the value, a ? Does a different study report this error margin, and can you show that 1.0 is within the margin? Otherwise, it is very difficult for the reader to evaluate whether this difference is "reasonable".

Lines 521-522: I do not understand how Figure 2a backs up this sentence. It is not a time series and does not show a particular location of a stable front.

Lines 527-528: The case for stochastic calving has also been published elsewhere, which needs to be acknowledged here or in the introduction (e.g., Bassis, 2011; Ambelorum and Robel, 2025).

Lines 553-555: This needs to be clearly stated at the beginning of the manuscript and given as a reminder to the reader throughout the paper, especially for certain findings as noted above. The coincidence is not impossible and is hard to ignore for certain findings laid out in this manuscript (see comments above).

Technical Corrections:

Line 115: "to that our" typo or an extra word?

Line 145: "d" is underlined instead of having an overbar

Line 219: Though it is the author's choice, using "whilst" rather than "while" throughout the manuscript is unnecessary and distracting.

Line 221: "minimal" should be "minimum"

Line 237: Figure 5 is referenced before Figure 4. The figures should be ordered according to the order they are referenced in the manuscript.

Line 286: Figure 6 and 7 need to be switched because Figure 7 is referenced first.

Line 297: "is through" is a typo or has the order of the words switched

Line 521: space needed in "functionartefact"

References:

Ambelorun AA, Robel AA. The integrated ice sheet response to stochastic iceberg calving. *Journal of Glaciology*. 2025;71:e79. doi:10.1017/jog.2025.10056

Bassis JN. The statistical physics of iceberg calving and the emergence of universal calving laws. *Journal of Glaciology*. 2011;57(201):3-16. doi:10.3189/002214311795306745