Review of Koo et al. (2024) 'Calibrating calving parameterizations using graph neural network emulators: Application to Helheim Glacier, East Greenland'

Summary
I reviewed this paper as part of a third-round of review; I had previously reviewed it in its second round. The paper demonstrates the application of graph neural networks as emulators for ISSM unstructured-grid simulations of Helheim Glacier, showing that the GNNs are able to accurately emulate the model outputs. The authors then go on to show that the much quicker run time of the emulator allows them to easily determine the required sigma_max parameter in a von Mises calving law to reproduce observed calving-front positions at Helheim between 2007 and 2020.

I congratulate the authors on the substantial work they put in after the second round of reviews to address the concerns raised. I think the current iteration of the paper is much the better for it. On the machine-learning side, the paper is solid and the modifications the authors have made to the manuscript help the reader to better understand what the emulator can achieve. However, I still have some issues regarding the glaciological side of the paper that I don't think the authors dealt with as convincingly, chiefly around the interpretation of the results and the model validation. Improving the validation probably represents enough work to push this into major rather than minor revisions, but I think it has to be done before I can recommend publication.

Page and line numbers refer to those in the clean version of the submitted manuscript.

Major Comments

- Applicability of the emulator: the paper very convincingly shows that the emulator functions well on Helheim over the 2007-2020 period, even with different parameter values to those with which it was trained. However, it still does not provide any indication of whether it would function effectively, without retraining, at a different location (or even Helheim under different climatic and topographical conditions). As the authors say, even if the emulator needs retraining for other settings, it will require far fewer simulations than finding the optimal sigma_max by other means, but I think it's important to show whether the emulator could function more or less out of the box, or would need a decent chunk of work to be usable on another use case. If I wanted to use it somewhere else and it turned out I had to compile my own record of calving-front positions and retrain the emulator, that's clearly something I'd really want to know before deciding to use it. At the moment, given the validation is essentially on the setup used to train it, it doesn't persuade me that it has learned something that really exists outside that model domain
- Glaciological interpretation: This is certainly much better than it was before, but I'm still not convinced. Figure 6 shows that predicted sigma_max does mostly follow the expected behaviour (advancing front = higher sigma and vice versa), but it doesn't always (late 2017, for example), and it would really help if the authors can provide some explanation of this. The underlying data are ISSM simulations, so, if everything is physically meaningful, it should be possible for the authors to provide some explanation for these discrepancies that makes sense. Some possibilities are mentioned in the discussion, but I think the paper really needs to offer something more concrete than a few maybes here.

Minor Comments

- p.3, l.63-65: Maybe better rephrased as 'GNN emulators take direct advantage of the unstructured meshes of the Ice-sheet and Sea-level System Model (ISSM), allowing flexible spatial resolution and efficient allocation of computational resources'
- p.3, l.67: I think it would be very helpful to evaluate the model on a different site too
- p.3, l.69: 'advances', not 'advancements'
- Section 5.2 and Table 2: Could the authors provide the various timings in hours and minutes as well in brackets after the time in seconds? '948 seconds', for example, is a sufficiently large number of seconds that I have to work out what it means in minutes, which is a bit of a pain.
- p.16, l.357-360: This paragraph is a bit redundant. The entire premise of this paper is that sigma_max has a substantial impact on the terminus position and that its behaviour has some

physical basis. If it didn’t, that would be more noteworthy! This might instead be a good place to note the times when sigma_max and the terminus position don’t agree and discuss why (or have a sentence saying this will be discussed more in the next section)

- p.16, l.362: ‘migration’, not ‘mitigation’
- p.16, l.364: Better rephrased as ‘the calving front retreats too much such that the modeling results no longer match the observations’
- p.18, l.375: ‘how difficult it is for calving to happen’
- p.18, l.380: Delete ‘this’
- p.18, l.381: But, if the emulator is trained with spatially uniform sigma_max values, would it be able to meaningfully predict a spatially varying field? How much variation/resolution would be needed? Two, three, four flowlines? More? My point is that this is easy to say, but I’m not convinced it would be easy to implement. I might rephrase this less blithely.
- p.18, ;382-386: Might the authors be able to do some of that here in relation to the deviations between sigma_max and the calving front position? The two lines don’t always agree in Figure 6 and this would be the perfect place to explore why. Similarly, Figure 8 shows that the optimal-sigma_max run smooths out a lot of the seasonal variation in ice-front position, though it follows the trend well; discussing why this happens would also be interesting