We would like to thank both reviewers for their insightful comments, in particular to improve the explanation of machine learning for an audience which may not be familiar with some terminology, and to improve the legibility of the figures. We have responded to each comment individually below.

Referee 1

The study looks at ways of improving Antarctic ice shelf melt rate estimation in climate models with particular focus on the landward margins of the models. The authors describe a comprehensive study examining how fine resolution models might be used to improve coarser models to enable some form of cross talk between small scales and those that large/long scale models need to work at if they are to provide useful climate insights. This bridging of timescales is a major challenge for climate science and so studies like this are important in helping "speed up" simulations in ways that still brings the needed physics along. I find some of the technical terminology in machine learning articles as a little challenging but probably OK from a GMD perspective.

"Partially resolve" in the title is somewhat broad brush given that the actual advance is to examine the thin water column margins of cavities.
Previous emulators of sub-shelf melt have been designed to reproduce melt rates in the entire ice-shelf cavity, for use in stand-alone ice-sheet models, or ocean models which do not include sub-shelf cavities. This work is different because the emulator is designed for use with cavity-resolving ocean models, which include the full physics of the ocean behaviour in some parts of the cavity. However, as shown in Figure 1, even in a cavity-resolving ocean model, significant parts of the sub-shelf cavity are not included. These are the regions which the DeepMelt-GL emulator is used for. To distinguish this work from previous works, we therefore specify in the title of the manuscript that the DeepMelt-GL emulator is for use in ocean models which partially resolve ice-shelf cavities.
We have updated paragraph 5 of the introduction to make this point clearer.
"In both studies, neural networks were used to parametrise the whole of the ice-shelf cavity for use with stand-alone ice-sheet models, or ocean models which do not include sub-shelf cavities. Since some ocean models include ice-shelf cavities, here we explore the suitability of using neural networks to parametrise the parts of cavities which are not resolved in these ocean models. For small cavities, this is the entire cavity, whereas for large cavities this means just the part closest to the grounding line. This approach allows us to maintain the full physics of water flow in ice-shelf cavities where resolved, and to produce sub-shelf melt patterns for the entire cavity."

The frontal terminal face and major crevasses also likely are resolution induced challenges – can they be examined in a similar way?
A similar machine learning approach could be applied to melting at the terminal face of ice streams and glacier, and in major crevasses, but these are separate questions outside the scope of this manuscript. The resolution that would be required to resolve these processes is much higher than existing simulations, probably of the order of tens of meters horizontally. Situable simulations for neural network training which cover a variety of Antarctic conditions do not currently exist.

A critical point that stands out to me is the training is happening from model to model with no reference to any in situ evidence base. There *are* studies that describe processes

in these marginal regions - all that in red in Fig 2. As far as I could tell there was no appeal to the insight provided from this important evidence base (see references). They point to issues like stratification, basal crevasses, tidal variability, even tsunamis - as all being present (lead authors – Begeman, Davis, Lawrence, Schmidt, Stewart & Washam). It would seem useful to have a reality check as to if these observed scales are reflected at some level in the high-resolution modelling. Is there a way of walking through the various likely processes in this unique environment and identifying how each process might fare in the upscaling?

We have updated the final paragraph of the discussion (Section 4), to mention more of the processes which are not included in the NEMO 0.25 degree model, and to acknowledge that these are processes which have been observed to influence basal melt rates.

"Additionally, since we train with the NEMO 0.25◦ model as our "reference" melt rates, any processes that are neglected at this resolution in NEMO will also not be included in the emulator. This includes processes such as small-scale eddies, tidal fluctuations in the grounding line, melting in basal crevasses and boundary layer turbulence, which are known to influence melt rates beneath Antarctic Ice shelves from observations (Davis and Nicholls, 2019; Begeman et al., 2020; Washam et al., 2023) or model simulations (Richter et al., 2022; Patmore et al., 2023). However, since NEMO melt rates are tuned to be close to estimates from remote sensing, and temperatures and salinity are close to in-situ observations (Mathiot and Jourdain (2023), some of these missing processes are implicitly accounted for in NEMO's parameter values."

Martin et al provide a useful insight into the connection between modelling and observations and point to (i) the need for approaches such as this manuscript but also (ii) ways to better connect with observations. I am not suggesting including these direct observations in the training but I think a few sentences connecting between the challenge at hand and the in situ evidence – and what extra evidence might be gathered – would help the impact of the work.

The NEMO ocean model is forced using atmospheric reanalysis, which integrates observations, and tuned to be close to ocean and ice observations. We have added the following sentences in the Methods (Section 2.1) to explain this:

"We use simulated melt rates as our "reference" melt rates because there are currently insufficient simultaneous observations of melt-rate and water properties, and because this allows us to include future climatic conditions which have not yet been observed. We stress that model parameters in the NEMO ocean model are tuned using estimated sub-shelf melt rates and water properties from the last 40 years (Mathiot and Jourdain, 2023)."

I also thought the final paragraph on the NEMO application was somewhat abrupt as it is an important future aspect of the work if it is to help improve ocean models more generally. The authors are proposing a significant shift in how future ocean models might operate and it raises lots of questions around how this might play out in terms tracking the ML enhancements within the general ocean model evolution. Can this be expanded a little in terms of speculation or next steps?

We have extended this paragraph to discuss some of the steps which might need to be taken when implementing the neural network into the NEMO model, with reference to selected model studies which have also implemented machine learning downscalings for other components of climate and ocean models.

"However, it will be necessary to do further model validation once the emulator is implemented in the 1◦ model to assess if it continues to reproduce the melt rates observed in the high-resolution model, as has been done in other studies including for parameterisations of convection (Yuval and O'Gorman, 2020) and vertical mixing (Zhu et al., 2022)."

Referee 2

The work presented in this manuscript produces an emulator for sub shelf melt rates not resolved in 1-degree resolution models for use with the NEMO model. Unique to this paper, this emulator allows for ocean physics to be solved in parts of the ice shelf cavity where model resolution allows. Accompanying the manuscript is code for generating the emulator dataset (if wanting to train on other data) and their trained output. This work is well written and addresses a relevant gap in current modeling capability for ocean-induced ice shelf melt rates. I applaud the authors' methodology for being easily updated for when NEMO's model physics improves and for preserving model physics when possible. This manuscript will be a nice addition to the ice sheet/ocean modeling and will be useful for a wide variety of future research questions. My main concerns are with the manuscript's figures and some clarifying questions. Additionally, there are a few typos throughout the paper. Below, I have specific comments and suggestions.

Line 11: Please fix the word 'Tthe'.
We have corrected this to "The".

Figure 1: The colors should have higher contrast to increase readability. I would suggest making the ocean a lighter blue (or white or light grey) and making the ice shelf cavity resolved in the 0.25 degree only a darker blue. This should make it more color blind friendly as well.
We have updated the colours on this figure as suggested.

Line 95: For the thinned ice shelf draft scenario, ANTFGEOM, does the grounding line placement move? I find that in general, there is a missing discussion on how the grounding line position will affect the applicability of the trained data to future scenarios. Could you add a few sentences (here or in the discussion) on how you expect changing pinning points to affect the emulator's performance?
We have added a note in the methods to clarify that the grounding line moves (as can be seen in Figure A1) in the ANTFGEOM simulation.
"In […] ANTFGEOM, ice shelf draft is artificially thinned in the Amundsen and Weddell Sea regions, and the grounding line retreated, following the MISOMIP2 protocol (Figure A1)."
Additionally, we have added a few sentences in Section 5.
"We note that as we have trained this neural network with a wide range of temperature conditions, we would expect it to produce realistic melt rates for conditions similar to the present day and up to 2300. However, as we have trained limited variation in geometries and grounding line positions (2 x 155 basins), we would expect good performance in a model with a fixed grounding line, and would be more cautious about performance in an ocean model with an evolving boundary. To improve the generatlisation to changing geometries, we recommend adding simulations with different geometries and grounding line positions when available, and especially a larger range of ice-shelf slopes. Another approach could be to use synthetic ice-shelf geometries such as those used by Rosier et al. (2023) and upcoming high-resolution ocean-ice-sheet coupled simulations (e.g. De Rydt et al., 2024)."

Figure 2: This figure is not easy to read, and I would suggest a major edit. I would suggest moving the schematic of the ocean, ice shelf, and topography up and increasing the opacity to 100%. The topography should be brown or gray, really any color besides blue. Below the schematic, there could be an 'algorithm flow chart' with boxes for water

property inputs, and geometry inputs, each with arrows pointing to the emulator box. The emulator box would describe what it does and have an arrow pointed to the 'output' box containing melt rates. I do think a figure like this is needed as it helps with understanding methodology, but it should be cleaned up.
We have edited this figure to make it cleaner, primarily by moving the majority of the labels from the schematic into a separate flow chart beneath the main image.

Line 141 Please change '3' to three.
We have changed this to "three".

Line 141: What are hidden layers? I recognize that describing these might be outside the scope of this paper, but for someone not familiar with machine learning/emulators, this is unclear.
We have added the following sentence in this paragraph:
"Inputs are processed through three sets of computational layers of artificial neurons (hidden layers), which apply weightings and non-linear activation functions, allowing the neural network to represent complex non-linear relationships."

Line 146: What is a ReLU activation function? Why use this?
Activation functions are used in machine learning to allow non-linear relationships to be represented. The rectified linear unit activation function (ReLu) is the most commonly used activation function in machine learning as it is easy to compute, and reduces the problem of vanishing gradient when the gradient descent algorithm is applied to optimise weightings during the neural network training process.
We have added a citation here (*Ramachandran, P., Zoph, B., & Le, Q. V. (2017). Searching for activation functions. arXiv preprint arXiv:1710.05941.).*

Line 207: Please change '4' to 'four'.
We have changed this to "four".

Line 228: Please fix 'warmm'.
We have changed this to "warm".

Line 231: Please fix 'ANt2300'.
We have changed this to "ANT2300".

Line 233: Why specifically a temperature threshold and not a salinity one? You should expand on this and potentially reference Figure 5 which shows RMSE varies mostly across CT. You touch on this a bit in line 261, but it could be expanded upon!
We have modified this paragraph to expand on this:
"*This suggests that there is a threshold above which the neural network performance begins to degrade, which is controlled by the training data. Looking at RMSE as a function of conservative temperature and absolute salinity (Figure 5), we see that reducing salinity does not affect RMSE as much as increasing temperature, and so suggest that this threshold is likely to be a temperature threshold.* "

Figure 4: Ideally, each segment could have its own color. Please adjust Filchner Ronne and Amery's titles so they don't overlap with the coastline segment.
We have changed the colours of the segments to be alternating as you go around the

coastline – this increases contrast without being overwhelming. We have moved the labels so that they do not overlap with the coastline segment.

Figure 5: For panels a and b, why are there two outlined groups of training data? Should this outline not connect and be one group? Why isn't this outline on panel c? What is the purpose of panel f?
The middle line was meant to outline the edge of the region with the densest data within the overall training dataset– we have changed this to be dashed to make this clearer. We have additionally added an outline to the data for the testing simulation (ANT2300), in panel c), which is also present on panels d) and e). Panel f) shows the mean temperature for the last 10 years of data in the ANT2300 dataset, so that interested readers can cross reference with Figure 4.

Line 267: So, the geometry is fixed, except for in the ANTFGEOM case. In the ANTFGEOM case, does the grounding line retreat? Does the area of ice shelf cavities not resolved in the 1-degree model increase because of grounding line retreat? One or two sentences addressing the issue of grounding line migration would be helpful!
We have added a note in the methods to clarify that the grounding line moves (as can be seen in Figure A1) in the ANTFGEOM simulation.
"In […] ANTFGEOM, ice shelf draft is artificially thinned in the Amundsen and Weddell Sea regions, and the grounding line retreated, following the MISOMIP2 protocol (Figure A1)."

Line 273: Please fix 'FILL IN'.
We have updated this to say:
'*The ensemble spread is 5.31 Gt yr−1 demonstrating that the parameter space is well constrained and all ensemble members give similar results.*'

Line 287: Please fix 'Instead.' to be 'Instead,'.
We have changed this to "Instead,".

Figure 6: I'm unsure if the middle figure (panel 't') is providing much information. I believe this figure could be improved by instead using a center inset like figure A1 and only plotting Filchner-Ronne (J-Jpp, Jpp-K) and Amundsen (G-H). Then this circum-Antarctic figure can be in the appendix. Since Figure 6 mirrors Figure 4, I would suggest having the regional names also printed in text in the inset figure.
We have modified this figure to only show the regions E-Ep, G-H, Jpp-K and J-Jpp. Instead of one panel which shows the geometry changes for the whole of Antarctica, we have added regionally zoomed panels which show the geometry changes in each of our chosen regions. We have moved the old Figure 6 to the appendix.

Figure 7: Panel c is difficult to see since it is small. If Figure 6 is changed to highlight only the Weddell and Amundsen Sea regions, I would also limit panel c in Figure 7 to show only those areas.
We have modified panel c to only show the Weddell Sea region.

Figure 8: Please adjust the placement of the grey text in panel a so it is not covered by the plotted lines.
We have adjusted the placement of the grey text.

In the caption, please edit "… applied to the ANT2100 simulation is also shown in grey (67.5 Gt yr−1)…" to be: "… applied to the ANT2100 simulation is also shown in light grey (67.5 Gt yr−1)".
We have changed this to "light grey".

Panel b needs a legend or explanation of what the simulation clusters represent. Why is the same color used for different groupings? Are they a part of the same cluster? That is, each color has a cluster near 0,0 as well as in the middle of the figure.
The clusters represent each of the simulations shown in panel a), and were originally coloured based on the number of years of training data. We have adjusted the colours to make it clearer that the cluster near 0,0 is from different simulations to the clusters in the main area of the plot.

Lines 314-315: "Sometimes, a high ensemble spread can be assocaited with a low RMSE, but we do not see any examples of a low ensemble spread and high RMSE value." Please fix the spelling of associated.
We have changed this to "associated".