Emulating present and future simulations of melt rates at the base of Antarctic ice shelves with neural networks

JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2023MS003829

Special Section:
Machine learning application to Earth system modeling

Emulating Present and Future Simulations of Melt Rates at the Base of Antarctic Ice Shelves With Neural Networks

C. Burgard[1,2], N. C. Jourdain[1], P. Mathiot[1], R. S. Smith[3], R. Schäfer[4], J. Caillet[1], T. S. Finn[5], and J. E. Johnson[1]

[1]University Grenoble Alpes, IRD, CNRS, INRAE, Grenoble INP, IGE, Grenoble, France, [2]Now at Laboratoire d'Océanographie et du Climat Expérimentations et Approches Numériques (LOCEAN), Sorbonne Université, CNRS/IRD/MNHN, Paris, France, [3]NCAS/Department of Meteorology, University of Reading, Reading, UK, [4]Physikalisch-Technische Bundesanstalt, Braunschweig, Germany, [5]École des Ponts and EDF R&D, Île-de-France CEREA, France

Abstract Melt rates at the base of Antarctic ice shelves are needed to drive projections of the Antarctic ice sheet mass loss. Current basal melt parameterizations struggle to link open ocean properties to ice-shelf basal melt rates for the range of current sub-shelf cavity geometries around Antarctica. We present a proof of concept exploring the potential of simple deep learning techniques to parameterize basal melt. We train a simple feedforward neural network, or multilayer perceptron, acting on each grid cell separately, to emulate the behavior of circum-Antarctic cavity-resolving ocean simulations. We find that this kind of emulator produces reasonable basal melt rates for our training ensemble, at least as close as or closer to the reference than traditional parameterizations. On an independent ensemble of simulations that was produced with the same ocean model but with different model parameters, cavity geometries and forcing, the neural network yields similar results to traditional parameterizations on present conditions. In much warmer conditions, both traditional parameterizations and neural network struggle, but the neural network tends to produce basal melt rates closer to the reference than a majority of traditional parameterizations. While this shows that such a neural network is at least as suitable for century-scale Antarctic ice-sheet projections as traditional parameterizations, it also highlights that tuning any parameterization on present-like conditions can introduce biases and should be used with care. Nevertheless, this proof of concept is promising and provides a basis for further development of a deep learning basal melt parameterization.

Plain Language Summary A warmer ocean around Antarctica leads to higher melting of the floating ice shelves, which influence the ice loss from the Antarctic ice sheet and therefore sea-level rise. In computer simulations of the ocean, these ice shelves are often not represented. For simulations of the ice sheet, so-called parameterizations are used to link the oceanic properties in front of the shelf and the melt at their base. We show that this link can be emulated with a simple neural network, which performs at least as well as traditional physical parameterizations both for present and much warmer conditions. This study also proposes several potential ways of further improving the use of deep learning to parameterize basal melt.

Key Points:

- We show that simple neural networks produce reasonable basal melt rates by emulating circum-Antarctic cavity-resolving ocean simulations
- Predicted melt rates for present and warmer conditions are similar or closer to the reference simulation than traditional parameterizations
- We show that neural networks are suited to be used as basal melt parameterizations for century-scale ice-sheet projections

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

C. Burgard,
clara.burgard@locean.ipsl.fr

Citation:

Burgard, C., Jourdain, N. C., Mathiot, P., Smith, R. S., Schäfer, R., Caillet, J., et al. (2023). Emulating present and future simulations of melt rates at the base of Antarctic ice shelves with neural networks. *Journal of Advances in Modeling Earth Systems*, *15*, e2023MS003829. https://doi.org/10.1029/2023MS003829

Received 17 MAY 2023
Accepted 29 NOV 2023

1. Introduction

The contribution of the Antarctic Ice Sheet to sea-level rise has been increasing in past decades and this increase is projected to continue with increasing greenhouse gas emissions (Fox-Kemper et al., 2021). Most of the mass loss is occurring at the margins of the ice sheet through faster ice flow from the grounded ice sheet to the ocean, mainly in West Antarctica (Khazendar et al., 2016; Mouginot et al., 2014; Rignot et al., 2014; Scheuchl et al., 2016; Shen et al., 2018; The IMBIE Team, 2018). This is because the floating ice shelves at the margins of the ice sheet, which usually buttress the ice flow, are rapidly thinning and retreating due to ocean-induced melt at their base (Adusumilli et al., 2020; Paolo et al., 2015; Rignot et al., 2013). In some bedrock configurations, increased ocean-induced melt can even trigger marine ice sheet instabilities (Gudmundsson et al., 2012; Schoof, 2007; Weertman, 1974), which have the potential to strongly increase Antarctic mass loss, on timescales below a century (Fox-Kemper et al., 2021). This makes ocean-induced sub-shelf melt, or *basal melt*, one of the main sources of uncertainty for future projections of sea-level rise.

Basal melt is a result of warm ocean water coming into contact with the base of the ice shelf. Which water masses reach the ice-ocean interface depends on the circulation of the water, not only in front of the ice shelf, but also

after entering the ice-shelf cavity (Dinniman et al., 2016). As a consequence, to simulate the properties of the water at the ice-ocean interface accurately, both the ocean circulation around Antarctica and the circulation in the cavities below the ice shelves need to be simulated accurately. A few global or circum-Antarctic ocean models already include ice-shelf modules (Comeau et al., 2022; Dinniman et al., 2015; Losch, 2008; Mathiot et al., 2017; Timmermann et al., 2012), but such ocean models are expensive to run on long timescales or for large ensembles. Instead, a majority of the global climate models used until now in the Coupled (CMIP) or Paleoclimate (PMIP) Model Intercomparison Projects still poorly represent the ocean dynamics along the Antarctic margins and do not include ice-shelf cavities (Beadling et al., 2020; Heuzé, 2021). Getting the right water masses in the right place around Antarctica is a matter for global and regional ocean modeling and will not be the focus of this study. In this study, we focus on the circulation within the ice-shelf cavities and the resulting melt.

To infer the basal melt forcing for projections of the Antarctic contribution to sea-level rise, ice-sheet models commonly rely on parameterizations linking hydrographic properties in front of the ice shelves, given by observations or oceanic output from global climate models, and the basal melt (Jourdain et al., 2020). Due to different assumptions and simplifications concerning the circulation in the cavities, the range of existing basal melt parameterizations leads to widely differing melt patterns and associated contributions to sea-level rise (Burgard et al., 2022; Favier et al., 2019). The magnitude of the resulting uncertainty contribution is similar, or even larger, than the choice of emission scenario used to force the projections (Seroussi et al., 2020; Edwards & the ISMIP6 Team, 2021).

Mimicking the ocean circulation within the cavity in simplified physical parameterizations is challenging and calls for exploring alternative approaches. We suggest that deep learning can be one tool to tackle this challenge. In recent years, the amount of ocean simulation output including ice-shelf cavities has increased and tools that make the application of deep learning techniques easily accessible have been developed, opening up the possibility of developing a neural network parameterization for basal melt. If trained with high-resolution model output, a neural network parameterization could implicitly include more intrinsic information about the system than a traditional physical parameterization. This approach has been applied promisingly in several areas of Earth System Sciences in the form of multilayer perceptrons applied on the grid-cell level (e.g., Gentine et al., 2018; Rasp et al., 2018), convolutional neural networks applied on multidimensional fields (e.g., Bolton & Zanna, 2019; Rosier et al., 2023) or random forests (e.g., Yuval & O'Gorman, 2020).

Deep learning has also been explored for basal melt parameterizations. Rosier et al. (2023) performed promising experiments that showed that a cavity-resolving ocean model can be emulated with a convolutional neural network in a variety of idealized ice-shelf geometries. In the present study, we choose a different deep learning approach to develop such a *deep emulator*, or *surrogate model*, which differs on two fundamental points. First, we train on the circum-Antarctic cavity-resolving ocean simulations with realistic geometries used in Burgard et al. (2022). Second, we use a multilayer perceptron architecture applied to each grid cell, as preliminarily explored in Bouissou et al. (2022). In the following, we present a proof of concept for a multilayer perceptron, which takes in hydrographic properties in front of the ice shelf and the geometric information at each grid point. In Section 2, we present the training and testing data, the neural network architecture, and the evaluation procedure. In Section 3, we explore the performance of the neural network using cross-validation techniques, while in Section 4 its explore the applicability to an independent testing data set. Finally, in Section 5, we discuss lessons learned from our study and give an outlook on possible directions to explore further in the future.

2. Data and Methods

The goal of this study is to explore if and how a neural network, in the form of a multilayer perceptron, can emulate the link between hydrographic properties in front of an ice shelf, geometric characteristics of the cavity, and the melt rates at its base as simulated by a cavity-resolving ocean model. In the following, we present the ocean model used and the set of simulations used for training, validation and testing the neural network; the neural network, its architecture, and its input variables; and the training and testing procedure.

2.1. Data

We choose to emulate a cavity-resolving version of the 3-D primitive-equation coupled ocean–sea-ice model NEMO (Nucleus for European Modeling of the Ocean, NEMO Team, 2019) run on the eORCA025 horizontal

grid (Storkey et al., 2018). This grid has a resolution of 0.25° in longitude on average, that is, a resolution of 4–14 km in the Antarctic seas and below the ice shelves, which is sufficient to capture the basic ocean circulation below multiple Antarctic ice shelves (Bull et al., 2021; Mathiot et al., 2017). Basal melt in the ice-shelf cavities is computed following Mathiot et al. (2017): a $z^\star$ coordinate is used for depth and the three equations (as proposed by Asay-Davis et al., 2016; D. Holland & Jenkins, 1999) are used to parameterize the ice-shelf melt in the ice-ocean boundary layer.

For the training phase, we use the same ensemble of simulations as used for the assessment of traditional basal melt parameterizations in Burgard et al. (2022). The ensemble is composed of four ocean simulations spanning 30–40 years, depending on the simulation, between 1979 and 2018. They were run with a standalone version of NEMO and forced with atmospheric forcing from JRA55-do version 1.4 (Tsujino et al., 2018). The Antarctic continental shelf bathymetry and ice shelf draft are constant and based on Bedmachine Antarctica version 2 (Morlighem et al., 2020). The simulations in the ensemble differ in a small number of parameters which are not directly related to the physics driving the ocean circulation and melt within the ice-shelf cavities but rather lead to a variety of hydrographic properties all around Antarctica. A more detailed description of the exact model configuration, differences in parameters and evaluation against observational estimates can be found in Burgard et al. (2022).

For the testing phase, we use two simulations independent from the ensemble used for training. In this case, NEMO was run in coupled mode as the oceanic component of the Earth System Model UKESM1.0-ice (Smith et al., 2021), which couples the UK Earth System Model (UKESM1, Sellar et al., 2019) to an adapted version of the ice-sheet model BISICLES (Cornford et al., 2013). In this coupled configuration, the cavities below the ice shelves are open and the ice-shelf melt is computed with the same approach as in the training ensemble. Due to the coupled setup, the ice-shelf draft evolves according to the simulated evolution of the ice sheet. Note that the position of the ice front at the surface remains fixed by ice-sheet model design. More details about the configuration of NEMO in this model setup can be found in Smith et al. (2021). The two test simulations differ in their atmospheric forcing. In the first one, which we will call "REPEAT1970," UKESM1.0-ice was run for several decades under constant 1970 greenhouse gas and other forcings. In the second one, which we will call "4xCO2," UKESM1.0-ice was run for several decades under instantaneously quadrupled 1970 CO_2 concentrations. In our study, we use 60 years of simulation, from year 10 to year 70, for both runs.

The training and the testing data set result from NEMO simulations. Nevertheless, next to differences in forcing from the atmosphere and the ice and bed geometry, the training and testing ensembles also differ in several technical aspects of NEMO. The training simulations were run with the version of 4.0.4. of NEMO (NEMO Team, 2019), including the sea-ice model SI^3, while the test simulations were run with the version 3.6 of NEMO (Madec & NEMO Team, 2017) and version 5.1 of the Community Ice CodE (CICE, Hunke et al., 2015). In addition, a few different parameter choices may affect the link between hydrographic properties in front of the ice shelf and the melt at the base of the ice shelf. The training ensemble was computed on 121 vertical levels (representing 20 m at 600 m depth), while the testing ensemble was computed on 75 vertical levels (representing 60 m at 600 m depth). In both ensembles, the thickness of the top boundary layer is bound at 20 m but can differ locally due to the different vertical resolutions. In the training ensemble, the thermal Stanton number is set to 7×10^{-4} while in the testing ensemble the thermal Stanton number is set to 1.45×10^{-3}. In the training ensemble, the top tidal velocity varies locally based on the CATS2008 data set (Howard et al., 2019; Padman et al., 2008), while it is fixed to 5 cm/s in the testing ensemble. In conclusion, this means that the testing ensemble is a slightly different model than the model which the neural network is trained to emulate and therefore represents a demanding testing experiment.

The training and testing ensembles cover a range of states that do not necessarily match observational estimates of hydrographic properties and basal melt rates. In both standalone and coupled mode, eORCA025 configurations are prone to biases in the ocean circulation around Antarctica (Smith et al., 2021). Nevertheless, in Burgard et al. (2022), we showed that, if the forcing and parameters were carefully chosen to reproduce realistic ocean conditions in the Southern Ocean, the resulting basal melt rates were in agreement with observational estimates from Rignot et al. (2013). The physical link between the hydrographic properties in front of the ice shelves and the basal melt rates is therefore reasonable. Based on this assumption, biases in the input properties should not affect the credibility of the training and evaluation procedure and the resulting neural network. On the contrary, a large variety of states is even beneficial because it provides more cases for our neural network to train on than only using the very limited sample of observations.

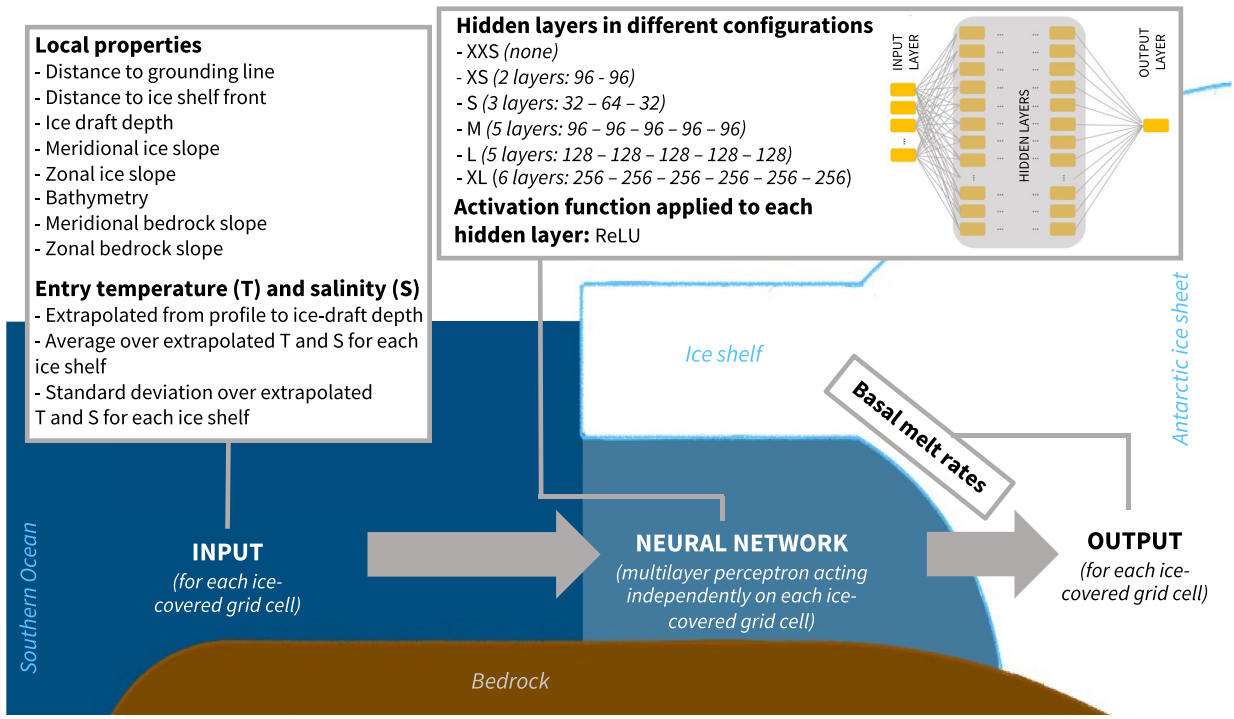

Figure 1. Schematic of the workflow around our neural network and description of the different neural network sizes.

On a more technical note, for this study, the NEMO output was interpolated bilinearly to a stereographic grid of 5 km spacing, as ice-sheet models and basal melt parameterizations are commonly run on a stereographic grid. All pre-processing, training, testing, and analysis is conducted using this regridded data. From this regridded data, we cut out the different ice shelves according to latitude and longitude limits defined on the present geometry (details found in Burgard, 2022) and then apply a routine to adapt this mask to slightly different geometries, like the ones resulting from the fully coupled UKESM1.0-ice runs. Of these ice shelves, we only keep the ice shelves under which we expect a reasonable representation of the ocean circulation. The effective resolution of physical ocean models, that is, the resolution below which the circulation might not be resolved well, is typically 5 to 10 times the grid spacing (Bricaud et al., 2020). We empirically choose a cutoff at an area of 2,500 km^2 (i.e., 6.25 Δx) to be in this range while keeping a sufficiently large number of ice shelves. Due to different geometries in the training and testing ensemble, this results into a slightly different ensemble of resolved ice shelves in these two ensembles (as listed in the Figures S1–S4 in Supporting Information S1).

2.2. Neural Network

We design our neural network to predict the basal melt rates based on information about the ocean temperature and salinity in front of the ice shelf and about the ice-shelf geometry (Figure 1). The scope of this study is to provide a proof of concept for the application of neural networks as a basal melt parameterization. We only investigate a small sample of neural network sizes for exploration in this study and do not claim that the best performing neural network in this study is the best performing neural network for the problem in general. This study is rather a proof of concept to encourage further research in this direction.

To link the input to the prediction, we use a multilayer perceptron, which is applied to each grid cell independently. A multilayer perceptron is the simplest form of a neural network and is a composition of functions (also called hidden layers), which takes an input array containing any number of variables and outputs a prediction. Specifying its number of neurons, each hidden layer is characterized by its parameters—the weights and biases, that connect each layer to its previous layer and shift the values in the hidden layer, respectively. An activation function in the hidden layer introduces non-linearities in the relationship between input and output. In this study, we explore different numbers of layers and numbers of neurons per layer. As activation function, we use the rectified

linear unit (ReLU, Fukushima, 1975; Nair & Hinton, 2010). The multilayer perceptron is implemented in Python with the package Keras (Chollet et al., 2015).

The strength of a neural network, and supervised machine learning techniques in general, is that it can reproduce complex non-linear relationships without being given the driving equations behind the data. Instead, its performance is driven by the supervised training phase, which determines the weights and biases of each neuron in the network. During training, the loss, describing the averaged distance of the network predictions to a given target output, is backpropagated to the weights of the network. The weights are then optimized with stochastic gradient descent. The training data set is randomly split up into batches, over which the optimization is looped. A complete pass through the batches defines an epoch, and the weights and biases are optimized over several such epochs. In parallel to the training, the neural network is applied to a validation data set to monitor its performance on data that has not been used for the training. After training, the final performance of the neural network is estimated by applying it to a previously unseen testing data set.

In this study, to train the neural network, the loss which we reduce is the mean-squared-error (MSE) over all ice-covered points between the predicted (m_{NN}) and target (m_{ref}) basal melt rates,

$$MSE = \frac{\sum_{i}^{N_{\mathrm{pts}}} \sum_{t}^{N_{\mathrm{years}}} (m_{\mathrm{NN}}[i,t] - m_{\mathrm{ref}}[i,t])^2}{N_{\mathrm{pts}} N_{\mathrm{years}}} \tag{1}$$

where the subscript NN stands for neural network, N_{pts} is the number of ice-covered grid points and N_{years} is the number of years used in the training. In Burgard et al. (2022), we argued that tuning on the grid-cell level would give too much weight to the larger ice shelves, as they cover a larger area. We still agree with this statement for traditional parameterizations because they already intrinsically contain assumptions about the physics of the circulation and the melt before tuning and have only one or two tuneable parameters. In the case of our neural network, the relationship between the properties in front of the ice shelf and the melt is learned from scratch, and it contains a larger number of parameters to adjust. We therefore argue that training on the grid-cell level is more sensible.

The neural network is optimized with Adam (Kingma & Ba, 2014), an initial learning rate of 0.001, an exponential decay rate for the first moment estimates (β_1) of 0.9 and an exponential decay rate for the second moment estimates (β_2) of 0.999. We split the training data set in batches with a size of 512 samples and optimize the neural network for at most 100 epochs. If the validation loss is not improved for 5 epochs, we reduce the learning rate by a factor of 2. If the validation loss is not improved for 10 epochs, we stop the training early. After early stopping, the model weights with the lowest validation loss are restored. More information about the choice of hyperparameters can be found in Supporting Information S1.

2.3. Input Variables

The multilayer perceptron takes an array of variables as input for each grid cell independently. In our case, the input array contains information about the geometrical properties of the grid cell and the hydrographic forcing (Figure 1).

For the geometrical properties, the input contains the following information: the ice draft depth, the local meridional and zonal slopes of the ice draft, the bathymetry, the local meridional and zonal slopes of the bedrock, and the distance of the grid cell to the nearest grounding line cell and the distance to the nearest ice front cell. All these variables are defined on the same horizontal plane and domain as the output array, the basal melt rates.

For the hydrographic forcing, more pre-processing is needed. To map the hydrographic forcing to the same grid cells as the other input variables, we proceed in the same manner as for traditional simple parameterizations in Burgard et al. (2022). First, we convert the conservative temperature and absolute salinity given by NEMO into potential temperature and practical salinity with the Gibbs SeaWater oceanographic toolbox (Firing et al., 2021). Second, we horizontally average the potential temperature and practical salinity, respectively, for each depth layer situated above the continental shelf within 50 km of the front of each ice shelf. The continental shelf is defined as grid cells where the depth of the bathymetry is shallower than 1500 m. The 50 km criterion imitates CMIP-type global ocean models that have resolutions around 1° (Heuzé, 2021), corresponding to a distance of between

38 km (70°S) and 56 km (60°S) in longitude. Third, we extrapolate the temperature and salinity from these mean vertical profiles in front of the ice shelf to the local ice-draft depth, resulting in one local temperature and local salinity value per grid cell in the ice-shelf domain. Fourth, we also compute, for each time step, the average and standard deviation of these extrapolated temperature and salinity fields and use them as additional input variables for each grid cell.

2.4. Training, Validation and Testing Methodology

In a first step, we explore different neural network sizes using the method of cross validation on our training ensemble. In a second step, we choose one of these neural networks to explore their performance on the testing data set.

We conduct two variations of leave-one-block-out cross validation to estimate the validation loss (MSE as defined in Equation 1), one on the ice shelf dimension and one on the time dimension, like in Burgard et al. (2022). This approach consists of dividing the data set into N blocks, training the neural network to minimize the training loss on $N-1$ blocks and using the left-out block to compute the validation loss (Roberts et al., 2017; Wilks, 2006). The procedure is re-iterated N times, leaving out each of the N blocks successively, so that, in the end, each Nth block has been left out of training once. All predictions for the left-out blocks, using the separately trained neural networks, are then concatenated to form a "synthetically independent" evaluation data set. Applying an evaluation metric on this evaluation data set, we assess how well the neural network generalizes to data "unseen" during training. We use $N = 35$ for the cross validation over ice shelves. For the cross validation over time, we divide the years into blocks of approximately 10 years (ten 10-year blocks and three 9-year blocks) to reduce the effect of autocorrelation, which is typically 2–3 years in our input temperatures. This results in $N = 13$ for the cross validation over time.

Before training, we normalize the training sample to put each of the 14 input variables (listed in Figure 1) as well as the output variable on a similar order of magnitude and avoid potential problems of gradient explosion. We do so by subtracting the mean and dividing by the standard deviation of the training sample. To avoid that validation data leaks into the training, this normalization is reiterated for each iteration of the cross validation.

We use the framework of cross validation to evaluate not only one but several neural networks to estimate the effect of their size on their performance. We sample different sizes ranging from an extra-extra small (XXS) neural network, with no hidden layer, and thus corresponding to a linear regression, to an extra-large (XL) neural network, with six hidden layers, each containing 256 neurons. The different sizes are listed in Figure 1.

To evaluate the resulting basal melt rates, we use the same metrics as in Burgard et al. (2022), namely: (a) the root-mean-squared error (RMSE) of the yearly integrated melt on the ice-shelf level and (b) the RMSE of the mean melt near the grounding line (GL) for each ice shelf. For the former, we compute the RMSE between the simulated and emulated yearly integrated melt (M) of the individual ice shelves [in Gt/yr] as follows:

$$RMSE_{\text{int}} = \sqrt{\frac{\sum_{k}^{N_{\text{isf}}} \sum_{t}^{N_{\text{years}}} (M_{\text{NN}}[k,t] - M_{\text{ref}}[k,t])^2}{N_{\text{isf}} N_{\text{years}}}} \tag{2}$$

where the subscript NN stands for neural network, N_{isf} is the number of ice shelves and N_{years} the number of simulated years, and the integrated melt M of ice shelf k [in Gt/yr] is:

$$M[k] = \rho_i \times 10^{-12} \sum_{j}^{N_{\text{grid cells in k}}} m_j a_j \tag{3}$$

where ρ_i is the ice density, m_j is the melt [in m ice per year] in grid cell j, and a_j is the area of grid cell j. For the latter, we compute the RMSE between the simulated and emulated yearly mean melt rate near the grounding line [in m ice per year]:

$$RMSE_{\text{GL}} = \sqrt{\frac{\sum_{k}^{N_{\text{isf}}} \sum_{n}^{N_{\text{simu}}} (m_{\text{GL,NN}}[k,n] - m_{\text{GL,ref}}[k,n])^2}{N_{\text{isf}} N_{\text{simu}}}} \tag{4}$$

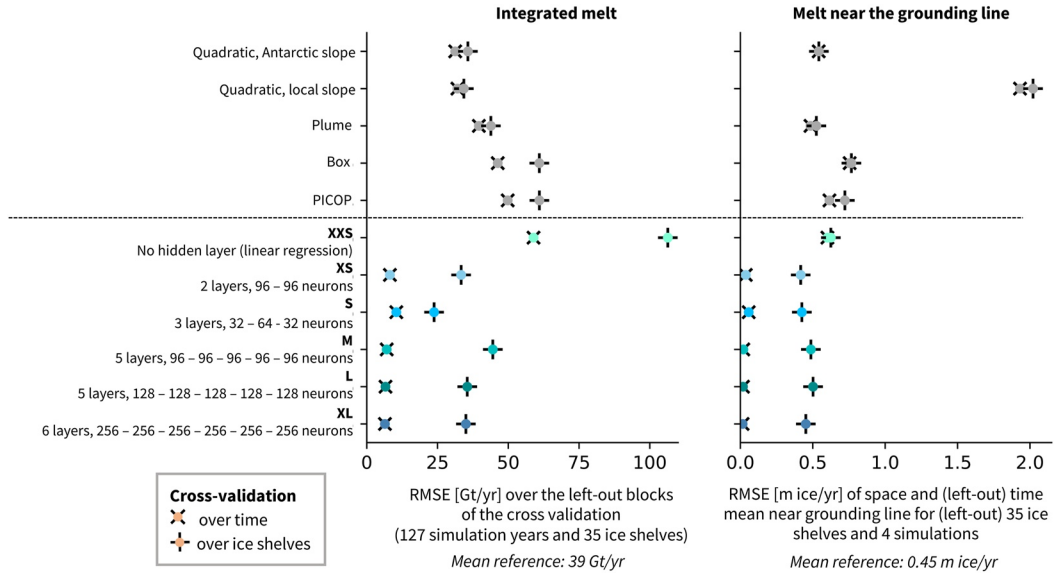

Figure 2. Summary of the root-mean-squared error (RMSE) of the integrated melt ($RMSE_{int}$) for the cross validation over time (×) and for the cross validation over ice shelves (+) [in Gt/yr] (left) and summary of the RMSE of the melt rate averaged over time and space near the grounding line ($RMSE_{GL}$) [in m ice/yr] (right). Shades of blue indicate the ensemble of neural network sizes and gray indicates a selection of traditional parameterizations (as shown in Burgard et al., 2022). The RMSE is computed following Equation 2, left panel, and Equation 4, right panel, on the synthetically independent evaluation data set.

where N_{simu} is the number of simulations in the ensemble and where m_{GL} for ice shelf k and simulation n is:

$$m_{GL}[k,n] = \frac{1}{N_{\text{years in n}}} \sum_{t}^{N_{\text{years in n}}} \frac{\sum_{j}^{N_{\text{grid cells near GL in k}}} (m_j a_j)}{\sum_{j}^{N_{\text{grid cells near GL in k}}} a_j} \tag{5}$$

The domain "near the grounding line" is the area covered by the first box prepared for the box parameterization, when considering a maximum amount of five boxes, and is equivalent to approximately 10% of the shelf area.

After cross validation, we choose the neural network producing the most satisfying results to do further evaluation on a completely independent data set. To do so, we reiterate the training of the subsample of neural networks over the whole training data set and choose to work with a deep ensemble (Lakshminarayanan et al., 2017). The final weights and biases of neural networks depend on the initialization of the weights before the first training iteration (Goodfellow et al., 2016). To account for this uncertainty and gain a more robust performance from the chosen neural network, we reiterate the training of the neural network 10 times with 10 different random initializations. We then apply this deep ensemble of 10 neural networks to the independent testing input and compute an ensemble mean over the 10 resulting melt rates.

3. Training and Cross Validation

3.1. Integrated Melt and Mean Melt Near the Grounding Line

The two evaluation metrics for the cross validation of the different neural network sizes are shown in Figure 2. To compare the performance to traditional parameterizations, we show the evaluation metrics for a subset of existing parameterizations: the quadratic local parameterization using a constant Antarctic slope (e.g., P. Holland et al., 2008) and using a local slope (e.g., Favier et al., 2019; Jourdain et al., 2020), the plume parameterization proposed by Lazeroms et al. (2019), the box parameterization with the same box amount as in Reese et al. (2018), and the PICOP parameterization from Pelle et al. (2019). These parameterizations are taken as formulated and tuned in Burgard et al. (2022). This means that they are tuned on the same training data set as the neural networks.

Corresponding to a linear regression, the XXS neural network leads to a RMSE of a similar order as traditional parameterizations in the cross validation over time and, for the melt near the grounding line, in the cross validation

over ice shelves as well. For the integrated melt, the cross validation over ice shelves leads to a comparably high RMSE. In the further course of this study, we therefore focus on neural networks that include hidden layers.

For both metrics, the RMSE for the cross validation over time is considerably reduced when using a neural network with hidden layers compared to traditional parameterizations and the XXS neural network. For the S-sized network, for example, the RMSE is reduced by 67%–79% for the integrated melt compared to traditional parameterizations and by 88%–97% for the melt near the grounding line. The RMSE for the cross validation over ice shelves is higher than for the cross validation over time but remains on the lower end of the range of RMSEs given by traditional parameterizations. For the S-sized network, for example, the RMSE is reduced by 31%–61% for the integrated melt and by 19%–78% for the melt near the grounding line.

The $RMSE_{int}$ of the cross validation over time is very similar between neural network sizes and spans between 6 Gt/yr (XL) and 11 Gt/yr (S). It remains well below the mean reference integrated melt on the ice-shelf level of 39 Gt/yr. The $RMSE_{int}$ of the cross validation over ice shelves varies more and is higher, between 24 (S) and 45 Gt/yr (M). The performance does not correlate with the neural network size. On the contrary, the lowest $RMSE_{int}$ of the cross validation over ice shelves is found for a comparably small neural network (S).

For the melt near the grounding line, the $RMSE_{GL}$ does not vary much in both cross validations between neural network sizes. The cross validation over time leads to a very low RMSE, varying from 0.02 m/yr (M, L, XL) to 0.06 m/yr (S). The cross validation over ice shelves leads to a RMSE between 0.42 m/yr (XS, S) and 0.50 m/yr (L), on the same order as the mean reference melt near the grounding line on the ice-shelf level, which is 0.45 m ice/yr.

The neural networks have more difficulties generalizing to unseen ice shelves than generalizing to unseen time periods. This means that one of the obstacles for the neural networks' performance is the application to unknown cavity geometries. Some of the cavity geometries are so different from the rest of the ensemble that they force the neural networks to extrapolate far from their training domain. However, if they have seen a given geometry at least once during training, they perform well on this geometry for another time step. This aspect is encouraging, as this means that the neural networks adapt well to temperature and salinity variations across the training ensemble.

3.2. Spatial Patterns

To add on the metrics at the ice-shelf level, we analyze the spatial patterns resulting from the S-sized neural network for the training ensemble member closest to realistic conditions (called REALISTIC in Burgard et al., 2022). We choose the S size because it represents the best compromise in the integrated metrics, having comparably low RMSE for both cross validations. For the cross validation over time, the pattern is nearly indistinguishable from the reference for Filchner-Ronne, Pine Island, and Fimbul ice shelves (Figure 3), while there is a slight overestimation by less than 1 m/yr over large parts of Ross ice shelf. For all ice shelves, the magnitude of the mean and standard deviation of the parameterized melt is nearly identical to the reference.

For the cross validation over ice shelves, the patterns are not matching in as much detail as in the cross validation over time. In particular, for the two largest ice shelves, Filchner-Ronne and Ross, it becomes clear that if the neural network has been trained without one of them, it will mimic the spatial pattern of the other because they are the only ones to share given ranges in the input variables, such as for example, large distances to the ice front and grounding line. For the other ice shelves, the parameterized patterns match the reference, but the magnitude of the melt deviates more from the reference than in the cross validation over time. For Pine Island, the neural network overestimates the average melt by 4 m/yr and underestimates the standard deviation by 3 m/yr. For Amery ice shelf, the high melt region in the South displays melt rates about 5 m/yr lower than the reference, whereas for Fimbul ice shelf the widespread melt is overestimated by less than 1 m/yr.

The low RMSE in the cross validation over time suggests an overfit on the geometry, which is fixed over time in the training data set. The melt patterns and magnitudes very close to the reference in the cross validation over time show that, even if our neural network is applied on each grid-cell separately, the location of the grid cell is more or less encoded in one or more input variables. However, as our problem is not necessarily well constrained with the input variables given, we suggest that this overfit can be used to our advantage. Our hypothesis is that, if the neural network has seen each ice shelf once, it has captured the variety of geometries and will be able to generalize to future changes in these "known" ice shelves. We do not expect new and completely different ice

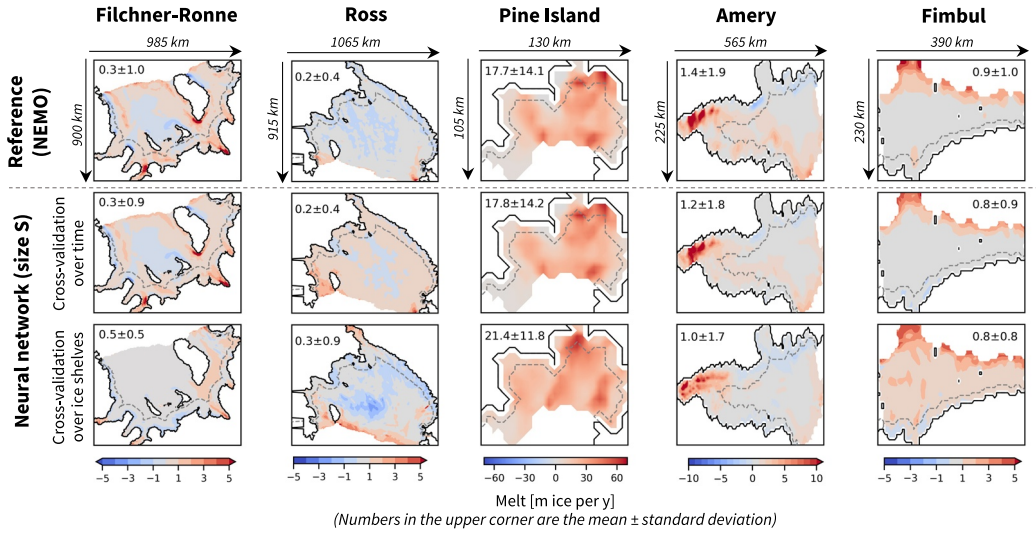

Figure 3. Subset of ice shelves for a visual evaluation of the melt patterns. Time average for the training ensemble member closest to real conditions (39 years) where the melt for each timestep has been computed with the neural network trained on the data set leaving out that timestep (cross validation over time, second row) and where the melt of each ice shelf has been computed with the neural network trained on the data set leaving out that ice shelf (cross validation over ice shelves, third row). Mean ± standard deviation are shown. The dashed line indicates the region used to evaluate the melt rate near the grounding line.

shelves to appear in the next centuries. To assess this idea, we need to investigate how well the neural network will perform on a geometry which is similar to but not identical to the training, and on hydrographic properties outside of the training range.

4. Testing on Independent Simulations

We apply our S-sized neural network on two independent data sets, one representing 60 years of constant 1970-forcing (REPEAT1970), and one representing warmer conditions, that is, 60 years of abrupt 4xCO_2 forcing (4xCO_2), from Smith et al. (2021). The REPEAT1970 simulation has a relatively steady ice-sheet geometry, similar (but not identical) to the training geometry and is useful to assess the sensitivity of the neural network to different near-present-day atmospheric conditions (from the UKESM atmosphere component), to different parameters used in NEMO, and to slightly different geometries. The 4xCO_2 simulation experiences larger changes in ice-sheet geometry and much warmer conditions, which is useful to test the neural networks far outside of their training range. As a consequence, this evaluation is demanding and permits to evaluate the limits of the neural network.

For evaluation, we divide the 4xCO_2 run into two 30-year blocks to capture potential differences with warming in time. As explained in Section 2.4, we train the neural network 10 times, with 10 different random initializations. In the following, the results shown are averages over the predictions of the 10 ensemble members, which represent one deep ensemble.

4.1. Integrated Melt and Melt Near the Grounding Line

The neural network reproduces well the REPEAT1970 melt rates integrated over individual ice shelves, with a $RMSE_{int}$ of 19 Gt/yr (Figure 4a, left). This error is slightly larger than in the cross validation over time (see Figure 2), and becomes similar to the quadratic and plume parameterizations. It should be noted that the $RMSE_{int}$ of these parameterizations is lower than in the cross validation, likely because of the overall lower melt rates in this simulation (24 Gt/yr compared to 39 Gt/yr in the training ensemble). The neural network still clearly outperforms the box and PICOP parameterizations ($RMSE_{int} \simeq 35$ Gt/yr).

For the melt near the grounding line, all parameterizations are uncertain, with $RMSE_{GL}$ close to the reference mean melt near the grounding line of 0.34 m/yr (Figure 4a, right). The neural network and the traditional parameterizations yield similar $RMSE_{GL}$, between 0.36 and 0.42 m/yr, except the quadratic using a local slope, which leads to a slightly lower RMSE, on the order of 0.22 m/yr.

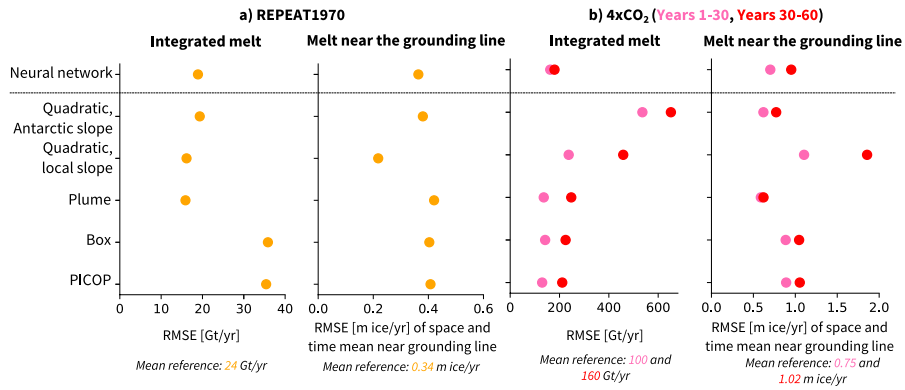

Figure 4. Summary of the root-mean-squared error (RMSE) of the integrated melt ($RMSE_{int}$) [in Gt/yr] and of the RMSE of the melt rate averaged over time and space near the grounding line ($RMSE_{GL}$) [in m ice/yr] for the application of the S-sized deep ensemble and a selection of traditional parameterizations on REPEAT1970 (a) and 4xCO_2 (b). Note the change in *x*-axis between the (a) and (b) panels.

For the warmer conditions (4xCO_2), all parameterizations struggle to reproduce the integrated melt on the ice-shelf level, with high spread in performance between the parameterizations (Figure 4b, left). The $RMSE_{int}$ is multiplied by more than 10 for the neural network and reaches nearly 650 Gt/yr for the quadratic parameterization using an Antarctic slope in the second period. While this jump in RMSE can be explained by a higher mean reference integrated melt (100 Gt/yr for the first period and 159 Gt/yr for the second period), it is probably also a result of forcing unseen during training such as much warmer and less saline ocean conditions (Figures S1 and S2 in Supporting Information S1). Over both periods, the neural network remains at the lower range of the difference to the reference melt rates. While neural network, plume, box and PICOP parameterization have comparable RMSEs for the first warm period (between 103 and 163 Gt/yr), the RMSE increases more for the plume, box and PICOP parameterization (between 211 and 248 Gt/yr) than for the neural network (180 Gt/yr) in the even warmer second period.

For the melt near the grounding line, the parameterizations perform differently than for the integrated melt, pointing to potential challenges outside the domain near the grounding line, along the path of the meltwater plume. These could be, for example, the effect of the ocean circulation in the wider cavity, interactions between the melt plume and the ambient ocean, the circulation of the melt plume or irregularities in the ice draft. The neural network performs in a similar uncertain manner as in the REPEAT1970 case (Figure 4b, right). Its $RMSE_{GL}$, 0.70 m/yr in the first period and 0.95 m/yr in the second period, is close to the reference mean melt near the grounding line (0.75 m/yr for the first period and 1.02 m/yr for the second period). In the first period, only the quadratic local parameterization using an Antarctic slope and the plume parameterization have lower $RMSE_{GL}$ (0.62 and 0.59 m/yr respectively), while in the second period only the quadratic parameterization using a local slope performs clearly worse than the other parameterizations. For all, the RMSE increases with warmer conditions but the gap between the periods depends on the parameterization, ranging from a difference of 0.04 m/yr for the plume parameterization to a difference of 0.76 m/yr for the quadratic parameterization using a local slope.

From this demanding application on an independent testing data set, several conclusions can be drawn. First, the neural network applies reasonably well to data that deviates from the training data set but represents near-present conditions. This means that, if it has seen all geometries of the main circum-Antarctic ice shelves, it can adapt to slightly different geometries. This is even more encouraging as the testing simulations were conducted with a slightly different version of NEMO than the neural networks were trained on. Second, the RMSE of the neural network is higher when applied to warmer conditions, but, in comparison with the traditional parameterizations, it performs at least as well or even better.

4.2. Spatial Patterns

Looking at the spatial patterns averaged over the last 10 years of the 4xCO_2 run, it becomes clear that all parameterizations, both neural network and traditional ones, struggle with warmer conditions and slightly different geometries to the training ensemble (Figure 5). The parameterization which struggles the most is the box parameterization, which widely underestimates the melt for all ice shelves, completely missing regions of strong melt.

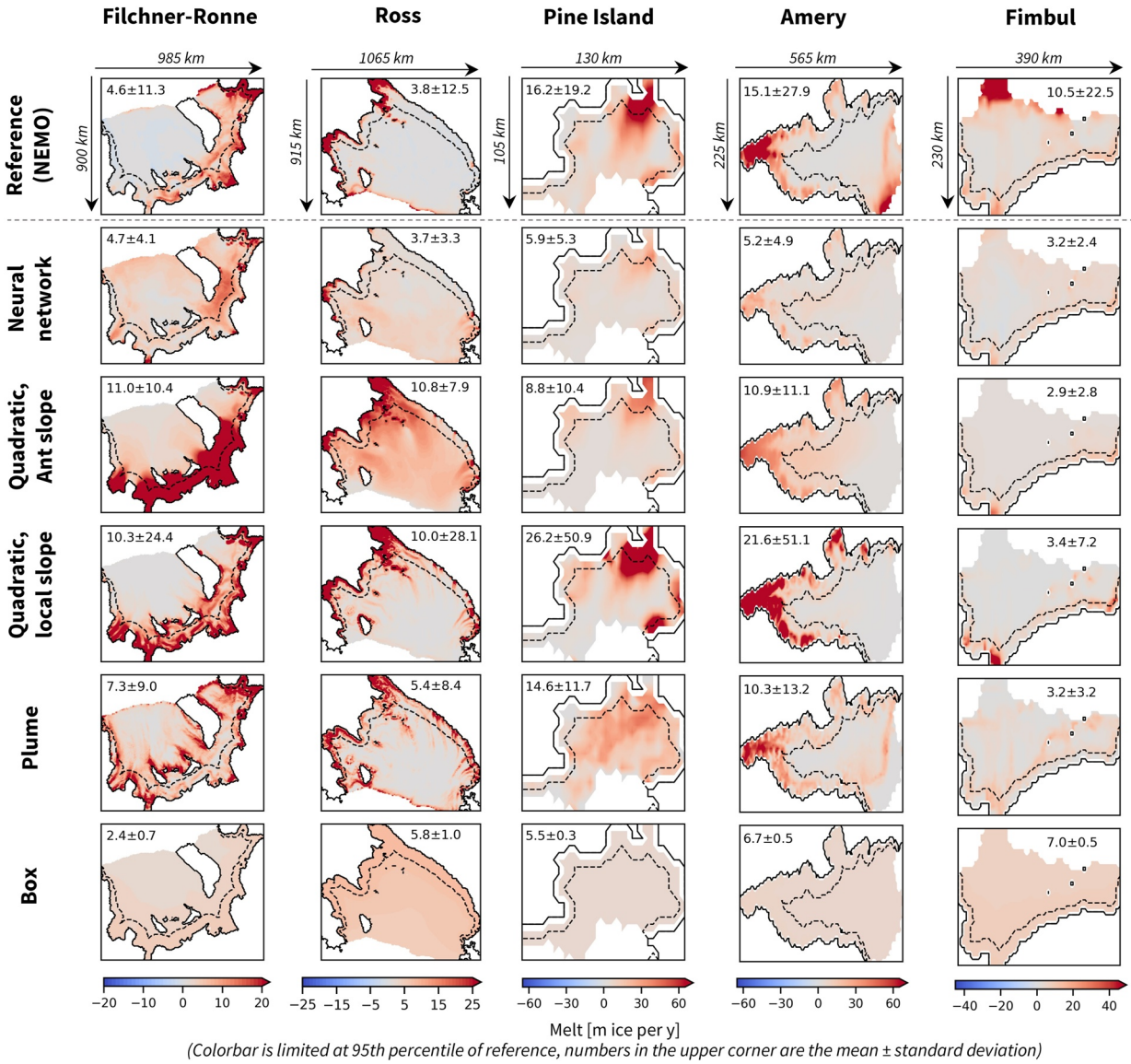

Figure 5. Subset of ice shelves for a visual evaluation of the melt patterns. Time average for the last 10 years of the 4xCO_2 run. The colorbar is limited to the 95th percentile of the NEMO reference. Mean ± standard deviation are shown. The dashed line indicates the region used to evaluate the melt rate near the grounding line.

For the large ice shelves of Filchner-Ronne and Ross, the neural network predicts similar mean melt rates as the reference but the distribution of the melt is mostly homogeneous over the ice shelves, with a few regions of comparably high but still underestimated melt. Except the box parameterization, all traditional parameterizations result in a more divided melt pattern, with higher melt near the grounding line for the quadratic parameterizations and higher melt near most coastlines for the plume parameterization. They overestimate the average melt by 150%–200% compared to the reference, a difference mostly introduced by an overestimation by about 20 m/yr in the regions of high parameterized melt.

For Pine Island and Amery ice shelves, a slight pattern similar to the reference can be seen in the melt predicted by the neural network but it is on average about three times lower than the average reference melt. The quadratic parameterizations both exhibit a similar pattern to the reference, but on average too low by 4–6 m/yr for the quadratic parameterization using the Antarctic slope and on average too high by 6–10 m/yr for the quadratic parameterization using the local slope. The plume parameterization has a more scattered melt pattern for Pine Island but a similar pattern to the reference with slightly too low melt for Amery. All parameterizations underestimate the melt for Fimbul ice shelf.

This spatial evaluation shows that the neural network has difficulties with input temperatures, salinities and melt rates well outside the training range. For the smaller ice shelves, the melt pattern is comparable to the reference but the melt rate is underestimated. We suggest that this is because it did not learn to compute melt rates above the range represented in the training data set. Also, some of this underestimation could be a result of the higher Stanton number, and therefore potentially slightly higher melt for same input properties, in our testing data set compared to our training data set. For the larger ice shelves, the neural network is struggling both with the melt rate and the pattern. We conjecture that this is a limitation of the overfit of the neural network and the neural network therefore extrapolates freely. We expect the overfitting effect to be largest for the large ice shelves because some ranges of input variables, such as large distances to ice front and grounding line as well as very deep ice-draft depth, are only found in these ice shelves, and these particular properties were not occurring in combination with warm conditions in the training.

5. Discussion

In this study, we showed that a simple multilayer perceptron can emulate melt rates as simulated by the cavity-resolving ocean model NEMO. This result is encouraging for further development because, as it is applied on a grid-cell level, this architecture is independent of the domain size and is therefore directly applicable to any ice shelf around Antarctica. It is also promising because the neural network's architecture is very simple and the hyperparameter tuning was mainly done empirically. In the following, we discuss insights from this study and possible further improvements to this approach.

5.1. Main Drivers of the Neural Network

One argument that is often made against the use of neural networks is that they remain statistical emulators of the training data and do not contain any physical constraints. The performance when applied to a slightly different model and to different conditions (see Section 4) already gives us a sense that the neural networks can reasonably adapt to conditions outside of training if they remain close to training conditions. In addition, we now evaluate which variables affect most the parameterized melt. First, this allows us to learn about the drivers of the neural network. Second, this could help future development of deep learning parameterizations as well as physical parameterizations to focus on these variables.

To assess the influence of the different variables on the predicted melt, we apply two variations of the permute-and-predict approach. In the permute-and-predict approach, one of the variables is shuffled randomly and used as input for the neural network alongside the other variables that remain in the original order. In the first variation, we shuffle the input variables within the REPEAT1970 sample to evaluate the influence of the different variables on the predicted melt in a situation close to the training conditions. In the second variation, we use a random sample from the $4xCO_2$ input for the shuffled variable and run the neural network using all other original input variables from the REPEAT1970 run to evaluate the effect of the variables in much warmer conditions. To avoid the potential effect of cross-correlation between some variables, we shuffle the variables by blocks. The shuffling is reiterated for each block separately. In the block *Position* we group the distance to the grounding line and to the ice front, in the block *Water column* we group the ice-draft depth and the bathymetry, in the block *Slopes bed* and *Slopes ice draft* we group the meridional and zonal slope of the bedrock and ice respectively, in the block *Temperature* and *Salinity* we group the local value, the average over the cavity and the standard deviation of temperature and salinity respectively.

For the shuffling within the REPEAT1970 run, the geometric properties, that is, the position of the grid cell and the water column height, are the variable groups which affect the predicted melt for Ross, Filchner-Ronne, Amery and Fimbul ice shelves the most (Figure 6a), with a few patches dominated by salinity. For Pine Island, the shuffling of temperature has the strongest effect in the central part but geometric variables are the most important in large areas near the grounding line. On the circum-Antarctic scale, the effect of shuffling position and water column height both lead to an increase of the RMSE of the integrated melt by 100%, while shuffling the temperature leads to an increase by 69% (Table 1). Near the grounding line, the shuffling of position and water column height has a lower effect on the RMSE. The temperature, ice slopes and salinity dominate, leading to an increase of the RMSE by 50%, 19% and 17% respectively.

When inserting random samples of $4xCO_2$ input, the patterns show that shuffling the temperature leads to the largest deviation to the original parameterized melt for most parts of all five ice shelves shown in Figure 6b. The

Table 1
Increase in RMSE of the Integrated Melt and of the Melt Near the Grounding Line (Near GL) Due To the Shuffling of the Different Variable Groups [in % of Original RMSE]

Difference in RMSE compared to original…		Position	Water col. height	Slopes bedrock	Slopes ice draft	Temperature	Salinity
…after shuffling within REPEAT1970	Integrated	100	100	1	6	69	27
	Near GL	3	3	0	19	50	17
…after inserting random sample from 4xCO_2 into REPEAT1970	Integrated	98	86	1	6	1624	507
	Near GL	6	0	0	19	39	0

most notable other features are the front of Ronne ice shelf, which is most affected by the shuffling of salinity and of the position, and some parts of Pine Island ice shelf and the grounding line of Amery ice shelf, which are most affected by the shuffling of geometrical properties. Again, this spatial evaluation is reflected in the circum-Antarctic evaluation metrics. The shuffling of temperature variables leads to an increase of the RMSE of the integrated melt by 1624%, followed by salinity with 507% and, further behind, position and water column height, with an increase by 98% and 86% respectively. For the melt near the grounding line, the increase in RMSE is not as high and remains of the same order of magnitude as using shuffled variables from REPEAT1970.

Several conclusions can be drawn from this experiment. First, this experiment shows that the geometry, in particular the position of the grid cell and the water column height, are key variables for the neural network to infer reasonable melt when applied on variables close to the training range, closely followed by the temperature. Water column depth, via the ice-draft depth, and temperature already are an integral part of existing parameterizations (Burgard et al., 2022). However, the position is currently only partly considered, and only in the more complex parameterizations such as the plume and box parameterizations (Lazeroms et al., 2019; Reese et al., 2018).

Second, when applied to much warmer conditions, the temperature and salinity, well outside the training range, clearly affect the resulting melt. This suggests that training the neural network on simulations of warmer

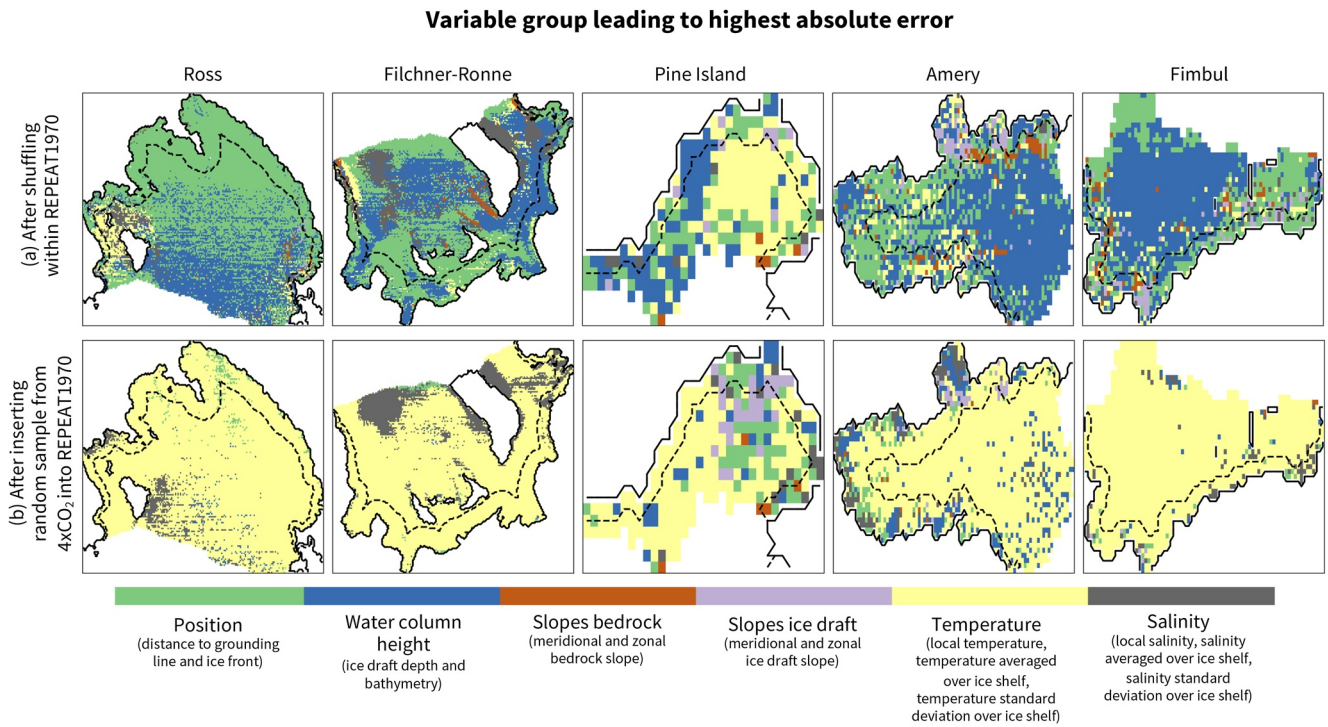

Figure 6. Variable group leading to highest absolute error in the time-mean pattern between originally parameterized melt and parameterized melt when using shuffled REPEAT1970 input (a) and using shuffled 4xCO_2 input (b) inserted in the REPEAT1970 input. This is conducted with the S-sized deep ensemble. The variables were shuffled by groups. The dashed line indicates the region used to evaluate the melt rate near the grounding line.

conditions could already improve its performance. Even more promising, the comparably low effect of geometry changes on integrated melt in the warm conditions presented here suggests that coupled ice-ocean simulations of warmer conditions are not necessarily needed for training and that cavity-resolving ocean simulations with fixed geometry could already be sufficient for projections of the near-future centuries.

Third, for the melt near the grounding line, the position of the grid cell is (maybe surprisingly) less important than for the integrated melt and the key variable is the temperature information, both near the training range and in warmer conditions. While the ice slope does not affect the integrated melt, it has some effect on the melt near the grounding line. This suggests that including ice slopes is necessary for a good performance near the grounding line. However, the way it is currently included in simple parameterizations is not successful as we showed in Burgard et al. (2022) that it leads to a clear overestimation of the melt in this region.

5.2. Possible Improvements

While the results of our neural network are encouraging, a variety of further improvements can be conducted in the future. The most obvious conclusion from this study is that predicting warmer conditions, similar to climate change conditions, is challenging for the neural network. To avoid extrapolation problems, we suggest, when possible, to introduce a set of simulations containing high-end future scenarios in the training data set to make the neural network more robust for future projections. At the same time, we saw that the traditional parameterizations struggle to represent future conditions as well. How to tune melt parameterizations to be applicable in both present and future conditions is therefore a problem that is not limited to deep learning approaches.

Another possible improvement is the treatment of the largest ice shelves. When looking at the cross-validation results into more detail, that is, at the scale of each ice shelf (not shown), the total RMSE over all ice shelves is strongly influenced by the high RMSE for the Ross ice shelf and, to a smaller extent, by the relatively high RMSE for the Filchner-Ronne ice shelves. These two ice shelves have an area which is much larger than the other ice shelves around Antarctica. Their cavities develop their own internal circulation (e.g., Gerdes et al., 1999; Naughten et al., 2021) and the residence time of water masses reaches several years (Michel et al., 1979; Nicholls & Østerhus, 2004). It is therefore not too surprising that parameterizations, which use input temperature and salinity averaged over thousands of kilometers at the front of the ice shelves and do not represent horizontal circulation explicitly, struggle with the representation of melt in these cavities. If we remove these two from the RMSE in the $4xCO_2$ case for example, we find that the RMSE is clearly reduced, below 100 Gt/yr, for both neural network and traditional parameterizations, compared to an original RMSE of several hundreds of Gt/yr. This shows that these rather simple parameterizations are not necessarily appropriate for the application on the Ross and Filchner-Ronne ice shelves. However, currently, these cavities are only resolved by ocean models on rare occasions. We advocate to strongly push efforts toward resolving these two cavities in ocean models by default, when possible, even at the lower resolution of 1°, as was already done for NEMO in Smith et al. (2021) or Hutchinson et al. (2023).

There is also space for improvement in the definition of input temperatures and salinities. Like in Burgard et al. (2022), the input profiles of temperature and salinity are here averaged over a given domain in front of the ice shelf. Then, we extrapolate the properties to the ice-draft depth. In addition, we computed the mean and standard deviation of these extrapolated temperature and salinity. However, machine learning gives us the opportunity to think bigger than traditional statistics when representing information about a given domain. One direction that could be explored in further development is the encoding of the important information about the water masses in front of the ice shelf using a machine learning technique. Ideally, this technique would take in a three-dimensional (horizontal plane and depth), or even a four-dimensional (taking also time as input to account for lags and residence time), field of temperature and salinity in front of the ice shelf and encode information about this field in a format to be given to the neural network. Such encoding might contain more information about the spatial distribution of the properties in front of the ice shelf and therefore potentially encode changes in the ocean circulation which might change the circulation within the cavities, as expected to happen in warmer conditions for the Filchner-Ronne ice shelf (Naughten et al., 2021).

A further source of improvement is the hyperparameter tuning. The hyperparameter choices that we made for this study, such as the number of hidden layers, the number of neurons per hidden layer, the activation function, the optimization method, the batch size, and the learning rate, lead to satisfying results. Further tuning or different

choices in these hyperparameters as well as introducing regularization methods could further improve our neural network. Using a different number of neural networks in the deep ensemble might also affect the parameterized melt.

Another aspect that can be further explored is the choice of architecture. Rosier et al. (2023) showed that a convolutional architecture can also be used to infer basal melt rates from hydrographic and geometric properties. A convolutional architecture, often U-Nets, is the preferred choice in many current studies exploring the application of machine learning to Earth System Sciences (e.g., Andersson et al., 2021; Ebert-Uphoff & Hilburn, 2020; Finn et al., 2023). In the case of basal melt and the ocean circulation in the cavity, such architectures clearly make sense as they can capture spatial patterns and correlations. Up to now, Rosier et al. (2023) demonstrate the performance of their MELTNET in a fixed domain and have not yet shown how to apply it to larger ice shelves than this domain. MELTNET remains however a promising approach and we are looking forward to its further development.

Finally, this study has focussed on the emulation of one ocean model at a given resolution. We acknowledge that NEMO's simulation of basal melt rates is not a perfect reflection of reality. An interesting further direction to follow would be to train a neural network to emulate NEMO at other resolutions and also to emulate other cavity-resolving ocean models. In this context, to ensure that the relationship remains sensible, we suggest training separate emulators and using them as an ensemble. This would provide an ensemble of emulators to be used as a variety of basal melt parameterizations, in addition to physics-based parameterizations. In a context where basal melt remains one of the main sources of uncertainty in projections of the Antarctic contribution to sea-level rise, a wide sample of this uncertainty in the form of a higher variety of parameterizations is welcome.

6. Conclusions

In conclusion, we show that a rather simple neural network architecture can be used to emulate a cavity-resolving ocean model. Our multilayer perceptron is designed to be rather simply useable as a basal melt parametrization for ice-sheet modelers. It uses input properties needed for the traditional parameterizations already and can be applied on the grid-cell level, similarly to most traditional parameterizations. While it struggles nearly as much as traditional parameterizations to generalize to ice shelves unseen during tuning, the neural network generalizes much better on time blocks unseen during training and the patterns are clearly better represented. In the demanding testing phase, on a data set produced with different NEMO parameters, geometry perturbations and forcing from the training, it still performs at least as well or even better than traditional parameterizations, both in historical and much warmer conditions. Nevertheless, for more robust applications on warmer conditions, we suggest including cavity-resolving ocean simulation output, or even coupled ocean-ice–sheet simulation output with projected geometry changes, in training data when possible, as more of these are planned to become available in coming years. In the present configuration, we suggest that, when possible, this neural network be used as part of a larger ensemble of parameterizations to cover this uncertainty.

Neural networks have been gaining lots of traction lately and efforts are done in many disciplines of the Earth System Sciences to explore their application. In this promising study, we provide guiding thoughts for further exploration and refinement of this approach, while this first proof of concept can already be used as an additional parameterization in the ice-sheet modeling landscape.

Data Availability Statement

The simulation data from Burgard et al. (2022) used for the training ensemble can be found on Zenodo: https://doi.org/10.5281/zenodo.7308352. The simulation data from Smith et al. (2021) used for the testing ensemble is available on Zenodo: https://doi.org/10.5281/zenodo.7886986. All other processed data, and code to train the neural networks and produce the figures can be found on Zenodo: https://doi.org/10.5281/zenodo.10149918. The Bedmachine Data is openly accessible (Morlighem, 2020).

Acknowledgments
We thank the three anonymous reviewers for their positive feedback and constructive comments. We also thank D. Jones, P. Holland, T. Andersson, A. Bradley, S. Thomas, A. Vaughan, and many others at the British Antarctic Survey for interesting discussions and exchange on ocean, ice, and machine learning. Most of the computations presented in this paper were performed using the GRICAD infrastructure (https://gricad.univ-grenoble-alpes.fr), which is supported by Grenoble research communities. The NEMO simulations were performed using high performance computing resources from GENCI–CINES (MISOCS project, allocations A0080106035 and A0100106035). This research was mainly conducted through the DEEP-MELT project (IRGA Pack IA 2021–2022), which is supported by MIAI @ Grenoble Alpes (ANR-19-P3IA-0003). This research was also supported by the European Union's Horizon 2020 research and innovation programme under grant agreements no. 869304 (PROTECT), 820575 (TiPACCs) and 101003536 (ESM2025), as well as by the French National Research Agency through the AIAI project (ANR-22-CE01-0014). It is PROTECT contribution number 79. CB and NCJ developed the original idea of this paper. CB carried out all analyses and wrote the manuscript. PM carried out the NEMO simulations used for training and RSS carried out the UKESM simulations. NCJ, RS and JC provided valuable help and code for the definition of the ice-shelf masks when the ice shelves evolve over time. TSF provided methodological input on the training of neural networks and JEJ provided useful input about how to think about machine learning. CB, NCJ, PM, RS, RSS, JC, TSF, JEJ contributed to discussions.

References

Adusumilli, S., Fricker, H., Medley, B., Padman, L., & Siegfried, M. (2020). Interannual variations in meltwater input to the Southern Ocean from Antarctic ice shelves. *Nature Geoscience*, *13*(9), 616–620. https://doi.org/10.1038/s41561-020-0616-z

Andersson, T., Hosking, J., Pérez-Ortiz, M., Paige, B., Elliott, A., Russell, C., et al. (2021). Seasonal Arctic sea ice forecasting with probabilistic deep learning. *Nature Communications*, *12*(1), 5124. https://doi.org/10.1038/s41467-021-25257-4

Asay-Davis, X., Cornford, S., Durand, G., Galton-Fenzi, B., Gladstone, R., Gudmundsson, G., et al. (2016). Experimental design for three interrelated marine ice sheet and ocean model intercomparison projects: Mismip v. 3 (mismip +), isomip v. 2 (isomip +) and misomip v. 1 (misomip1). *Geoscientific Model Development*, *9*(7), 2471–2497. https://doi.org/10.5194/gmd-9-2471-2016

Beadling, R., Russell, J., Stouffer, R., Mazloff, M., Talley, L., Goodman, P., et al. (2020). Representation of Southern Ocean properties across coupled model intercomparison project generations: CMIP3 to CMIP6. *Journal of Climate*, *33*(15), 6555–6581. https://doi.org/10.1175/JCLI-D-19-0970.1

Bolton, T., & Zanna, L. (2019). Applications of deep learning to ocean data inference and subgrid parameterization. *Journal of Advances in Modeling Earth Systems*, *11*(1), 376–399. https://doi.org/10.1029/2018MS001472

Bouissou, B., Burgard, C., & Jourdain, N. (2022). Parameterising ocean-induced melt of an idealised Antarctic ice shelf using deep learning. *ECCOMAS22 Conference proceedings*. https://doi.org/10.23967/eccomas.2022.216

Bricaud, C., Le Sommer, J., Madec, G., Calone, C., Deshayes, J., Ethe, C., et al. (2020). Multi-grid algorithm for passive tracer transport in the NEMO ocean circulation model: A case study with the NEMO OGCM (version 3.6). *Geoscientific Model Development*, *13*(11), 5465–5483. https://doi.org/10.5194/gmd-13-5465-2020

Bull, C., Jenkins, A., Jourdain, N., Vaňková, I., Holland, P., Mathiot, P., et al. (2021). Remote control of filchner-ronne ice shelf melt rates by the Antarctic slope current. *Journal of Geophysical Research: Oceans*, *126*(2). https://doi.org/10.1029/2020JC016550

Burgard, C. (2022). Multimelt, a python framework to apply existing basal melt parameterisation. Python Package Index - PyPI. https://pypi.org/project/multimelt/

Burgard, C., Jourdain, N., Reese, R., Jenkins, A., & Mathiot, P. (2022). An assessment of basal melt parameterisations for Antarctic ice shelves. *The Cryosphere*, *16*(12), 4931–4975. https://doi.org/10.5194/tc-16-4931-2022

Chollet, F., et al. (2015). Keras. Retrieved from https://keras.io

Comeau, D., Asay-Davis, X., Begeman, C., Hoffman, M., Lin, W., Petersen, M., et al. (2022). The DOE E3SM v1.2 cryosphere configuration: Description and simulated antarctic ice-shelf basal melting. *Journal of Advances in Modeling Earth Systems*, *14*(2), e2021MS002468. https://doi.org/10.1029/2021MS002468

Cornford, S., Martin, D., Graves, D., Ranken, D., Le Brocq, A., Gladstone, R., et al. (2013). Adaptive mesh, finite volume modeling of marine ice sheets. *Journal of Computational Physics*, *232*(1), 529–549. https://doi.org/10.1016/j.jcp.2012.08.037

Dinniman, M., Asay-Davis, X., Galton-Fenzi, B., Holland, P., Jenkins, A., & Timmermann, R. (2016). Modeling ice shelf/ocean interaction in Antarctica: A review. *Oceanography*, *29*(4), 144–153. https://doi.org/10.5670/oceanog.2016.106

Dinniman, M., Klinck, J., Bai, L.-S., Bromwich, D., Hines, K., & Holland, D. (2015). The effect of atmospheric forcing resolution on delivery of ocean heat to the antarctic floating ice shelves. *Journal of Climate*, *28*(15), 6067–6085. https://doi.org/10.1175/JCLI-D-14-00374.1

Ebert-Uphoff, I., & Hilburn, K. (2020). Evaluation, tuning, and interpretation of neural networks for working with images in meteorological applications. *Bulletin of the American Meteorological Society*, *101*(12), E2149–E2170. https://doi.org/10.1175/BAMS-D-20-0097.1

Edwards, T., Nowicki, S., Marzeion, B., Hock, R., Goelzer, H., Seroussi, H., et al. (2021). Projected land ice contributions to twenty-first-century sea level rise. *Nature*, *593*(7857), 74–82. https://doi.org/10.1038/s41586-021-03302-y

Favier, L., Jourdain, N., Jenkins, A., Merino, N., Durand, G., Gagliardini, O., et al. (2019). Assessment of sub-shelf melting parameterisations using the ocean–ice-sheet coupled model NEMO(v3.6)–Elmer/Ice(v8.3). *Geoscientific Model Development*, *12*(6), 2255–2283. https://doi.org/10.5194/gmd-12-2255-2019

Finn, T., Durand, C., Farchi, A., Bocquet, M., Chen, Y., Carrassi, A., & Dansereau, V. (2023). Deep learning of subgrid-scale parametrisations for short-term forecasting of sea-ice dynamics with a Maxwell-Elasto-Brittle rheology. *EGUsphere*. https://doi.org/10.5194/egusphere-2022-1342

Firing, E., Fernandes, F., Barna, A., & Abernathey, R. (2021). Teos-10/gsw-python: v3.4.1.post0. *Zenodo*. [used version 3.6.16]. https://doi.org/10.5281/zenodo.5214122

Fox-Kemper, B., Hewitt, H., Xiao, C., Adalgeirsdóttir, G., Drijfhout, S., Edwards, T., et al. (2021). Ocean, cryosphere and sea level change. [Chapter 9]. In V. Masson-Delmotte et al. (Eds.), *Climate change 2021: The physical science basis. Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change*. Cambridge University Press.

Fukushima, K. (1975). Cognitron: A self-organizing multilayered neural network. *Biological Cybernetics*, *20*(3), 121–136. https://doi.org/10.1007/BF00342633

Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G., & Yacalis, G. (2018). Could machine learning break the convection parameterization deadlock? *Geophysical Research Letters*, *45*(11), 5742–5751. https://doi.org/10.1029/2018GL078202

Gerdes, R., Determann, J., & Grosfeld, K. (1999). Ocean circulation beneath Filchner-Ronne Ice Shelf from three-dimensional model results. *Journal of Geophysical Research*, *104*(C7), 15827–15842. https://doi.org/10.1029/1999JC900053

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. Retrieved from http://www.deeplearningbook.org

Gudmundsson, G., Krug, J., Durand, G., Favier, L., & Gagliardini, O. (2012). The stability of grounding lines on retrograde slopes. *The Cryosphere*, *6*(6), 1497–1505. https://doi.org/10.5194/tc-6-1497-2012

Heuzé, C. (2021). Antarctic bottom water and North Atlantic deep water in CMIP6 models. *Ocean Science*, *17*(1), 59–90. https://doi.org/10.5194/os-17-59-2021

Holland, D., & Jenkins, A. (1999). Modeling thermodynamic ice–ocean interactions at the base of an ice shelf. *Journal of Physical Oceanography*, *29*(8), 1787–1800. https://doi.org/10.1175/1520-0485(1999)029⟨1787:MTIOIA⟩2.0.CO;2

Holland, P., Jenkins, A., & Holland, D. (2008). The response of ice shelf basal melting to variations in ocean temperature. *Journal of Climate*, *21*(11), 2558–2572. https://doi.org/10.1175/2007JCLI1909.1

Howard, S. L., Padman, L., & Erofeeva, S. (2019). Cats2008: Circum-antarctic tidal simulation version 2008. Retrieved from https://www.usap-dc.org/view/dataset/601235

Hunke, E., Lipscomb, W., Turner, A., Jeffery, N., & Elliott, S. (2015). CICE: The Los Alamos sea ice model, documentation and software, version 5.1 la-cc-06-012 (Computer software manual No. LA-CC-06-012).

Hutchinson, K., Deshayes, J., Ethé, C., Rousset, C., de Lavergne, C., Vancoppenolle, M., et al. (2023). Improving Antarctic Bottom Water precursors in NEMO for climate applications. *EGUsphere*. https://doi.org/10.5194/egusphere-2023-99

Jourdain, N., Asay-Davis, X., Hattermann, T., Straneo, F., Seroussi, H., Little, C., & Nowicki, S. (2020). A protocol for calculating basal melt rates in the ISMIP6 Antarctic Ice Sheet projections. *The Cryosphere*, *14*(9), 3111–3134. https://doi.org/10.5194/tc-14-3111-2020

Khazendar, A., Rignot, E., Schroeder, D., Seroussi, H., Schodlok, M., Scheuchl, J., et al. (2016). Rapid submarine ice melting in the grounding zones of ice shelves in West Antarctica. *Nature Communications*, *7*(1), 13243. https://doi.org/10.1038/ncomms13243

Kingma, D., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint. https://doi.org/10.48550/ARXIV.1412.6980

Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. In *Proceedings of the 31st international conference on neural information processing systems* (pp. 6405–6416). Curran Associates Inc.

Lazeroms, W., Jenkins, A., Rienstra, S., & vande Wal, R. (2019). An analytical derivation of ice-shelf basal melt based on the dynamics of melt-water plumes. *Journal of Physical Oceanography*, *49*(4), 917–939. https://doi.org/10.1175/JPO-D-18-0131.1

Losch, M. (2008). Modeling ice shelf cavities in a z coordinate ocean general circulation model. *Journal of Geophysical Research*, *113*(C8), C08043. https://doi.org/10.1029/2007JC004368

Madec, G., & NEMO Team. (2017). Nemo ocean engine (v3.6-patch). *Notes du Pôle de modélisation de l'Institut Pierre-Simon Laplace (IPSL), Zenodo*, *27*. https://doi.org/10.5281/zenodo.3248739

Mathiot, P., Jenkins, A., Harris, C., & Madec, G. (2017). Explicit representation and parametrised impacts of under ice shelf seas in the $z^{\star}$ coordinate ocean model nemo 3.6. *Geoscientific Model Development*, *10*(7), 2849–2874. https://doi.org/10.5194/gmd-10-2849-2017

Michel, R., Linick, T., & Williams, P. (1979). Tritium and carbon-14 distributions in seawater from under the Ross Ice Shelf Project ice hole. *Science*, *203*(4379), 445–446. https://doi.org/10.1126/science.203.4379.445

Morlighem, M. (2020). *MEaSUREs BedMachine Antarctica*. Version 2. NASA National Snow and Ice Data Center Distributed Active Archive Center. https://doi.org/10.5067/E1QL9HFQ7A8M

Morlighem, M., Rignot, E., Binder, T., Blankenship, D., Drews, R., Eagles, G., et al. (2020). Deep glacial troughs and stabilizing ridges unveiled beneath the margins of the Antarctic Ice Sheet. *Nature Geoscience*, *13*(2), 132–137. https://doi.org/10.1038/s41561-019-0510-8

Mouginot, J., Rignot, E., & Scheuchl, B. (2014). Sustained increase in ice discharge from the Amundsen sea Embayment, West Antarctica, from 1973 to 2013. *Geophysical Research Letters*, *41*(5), 1576–1584. https://doi.org/10.1002/2013GL059069

Nair, V., & Hinton, G. (2010). Rectified linear units improve restricted Boltzmann machines. In *Proceedings of the 27th international conference on international conference on machine learning* (pp. 807–814). Omnipress. https://doi.org/10.5555/3104322.3104425

Naughten, K., De Rydt, J., Rosier, S., Jenkins, A., Holland, P., & Ridley, J. (2021). Two-timescale response of a large Antarctic ice shelf to climate change. *Nature Communication*, *12*(1), 1991. https://doi.org/10.1038/s41467-021-22259-0

NEMO Team. (2019). Nemo ocean engine. *Scientific Notes of Climate Modelling Center*, *27*. https://doi.org/10.5281/zenodo.1464816

Nicholls, K. W., & Østerhus, S. (2004). Interannual variability and ventilation timescales in the ocean cavity beneath Filchner-Ronne Ice Shelf, Antarctica. *Journal of Geophysical Research*, *109*(C4), C04014. https://doi.org/10.1029/2003JC002149

Padman, L., Erofeeva, S., & Fricker, H. (2008). Improving antarctic tide models by assimilation of icesat laser altimetry over ice shelves. *Geophysical Research Letters*, *35*(22), L22504. https://doi.org/10.1029/2008GL035592

Paolo, F., Fricker, H., & Padman, L. (2015). Volume loss from Antarctic ice shelves is accelerating. *Science*, *348*(6232), 327–331. https://doi.org/10.1126/science.aaa0940

Pelle, T., Morlighem, M., & Bondzio, J. (2019). Brief communication: PICOP, a new ocean melt parameterization under ice shelves combining PICO and a plume model. *The Cryosphere*, *13*(3), 1043–1049. https://doi.org/10.5194/tc-13-1043-2019

Rasp, S., Pritchard, M., & Gentine, P. (2018). Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences*, *115*(39), 9684–9689. https://doi.org/10.1073/pnas.1810286115

Reese, R., Albrecht, T., Mengel, M., Asay-Davis, X., & Winkelmann, R. (2018). Antarctic sub-shelf melt rates via PICO. *The Cryosphere*, *12*(6), 1969–1985. https://doi.org/10.5194/tc-12-1969-2018

Rignot, E., Jacobs, S., Mouginot, J., & Scheuchl, B. (2013). Ice-shelf melting around Antarctica. *Science*, *341*(6143), 266–270. https://doi.org/10.1126/science.1235798

Rignot, E., Mouginot, J., Morlighem, M., Seroussi, H., & Scheuchl, B. (2014). Widespread, rapid grounding line retreat of Pine Island, Thwaites, Smith, and Kohler glaciers, West Antarctica, from 1992 to 2011. *Geophysical Research Letters*, *41*(10), 3502–3509. https://doi.org/10.1002/2014GL060140

Roberts, D., Bahn, V., Ciuti, S., Boyce, M., Elith, J., Guillera-Arroita, G., et al. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, *40*(8), 913–929. https://doi.org/10.1111/ecog.02881

Rosier, S., Bull, C., Woo, W., & Gudmundsson, G. (2023). Predicting ocean-induced ice-shelf melt rates using deep learning. *The Cryosphere*, *17*(2), 499–518. https://doi.org/10.5194/tc-17-499-2023

Scheuchl, B., Mouginot, J., Rignot, E., Morlighem, M., & Khazendar, A. (2016). Grounding line retreat of Pope, Smith, and Kohler Glaciers, West Antarctica, measured with Sentinel-1a radar interferometry data. *Geophysical Research Letters*, *43*(16), 8572–8579. https://doi.org/10.1002/2016GL069287

Schoof, C. (2007). Ice sheet grounding line dynamics: Steady states, stability, and hysteresis. *Journal of Geophysical Research*, *112*(F3), F03S28. https://doi.org/10.1029/2006JF000664

Sellar, A., Jones, C., Mulcahy, J., Tang, Y., Yool, A., Wiltshire, A., et al. (2019). UKESM1: Description and evaluation of the U.K. Earth system model. *Journal of Advances in Modeling Earth Systems*, *11*(12), 4513–4558. https://doi.org/10.1029/2019MS001739

Seroussi, H., Nowicki, S., Payne, A., Goelzer, H., Lipscomb, W., Abe-Ouchi, A., et al. (2020). ISMIP6 Antarctica: A multi-model ensemble of the Antarctic Ice Sheet evolution over the 21st century. *The Cryosphere*, *14*(9), 3033–3070. https://doi.org/10.5194/tc-14-3033-2020

Shen, Q., Wang, K., Shum, C., Jiang, L., Hsu, H., & Dong, J. (2018). Recent high-resolution Antarctic ice velocity maps reveal increased mass loss in Wilkes Land, East Antarctica. *Scientific Reports*, *8*(1), 4477. https://doi.org/10.1038/s41598-018-22765-0

Smith, R., Mathiot, P., Siahaan, A., Lee, V., Cornford, S., Gregory, J., et al. (2021). Coupling the U.K. Earth system model to dynamic models of the Greenland and Antarctic Ice Sheets. *Journal of Advances in Modeling Earth Systems*, *13*(10), e2021MS002520. https://doi.org/10.1029/2021MS002520

Storkey, D., Blaker, A., Mathiot, P., Megann, A., Aksenov, Y., Blockley, E., et al. (2018). UK global ocean GO6 and GO7: A traceable hierarchy of model resolutions. *Geoscientific Model Development*, *11*(8), 3187–3213. https://doi.org/10.5194/gmd-11-3187-2018

The IMBIE Team. (2018). Mass balance of the Antarctic Ice Sheet from 1992 to 2017. *Nature*, *558*(7709), 219–222. https://doi.org/10.1038/s41586-018-0179-y

Timmermann, R., Wang, Q., & Hellmer, H. (2012). Ice-shelf basal melting in a global finite-element sea-ice/ice-shelf/ocean model. *Geoscientific Model Development*, *53*(60), 303–314. https://doi.org/10.3189/2012AoG60A156

Tsujino, H., Urakawa, S., Nakano, H., Small, R., Kim, W., Yeager, S., et al. (2018). Jra-55 based surface dataset for driving ocean–sea-ice models (jra55-do). *Ocean Modelling*, *130*, 79–139. https://doi.org/10.1016/j.ocemod.2018.07.002

Weertman, J. (1974). Stability of the junction of an ice sheet and an ice shelf. *Journal of Glaciology*, *13*(67), 3–11. https://doi.org/10.3189/S0022143000023327

Wilks, D. (2006). *Statistical methods in the atmospheric sciences* (2nd ed.). Elsevier.

Yuval, J., & O'Gorman, P. (2020). Stable machine-learning parameterization of subgrid processes for climate modeling at a range of resolutions. *Nature Communications*, *11*(1), 1–10. https://doi.org/10.1038/s41467-020-17142-3