The 8th European Congress on Computational Methods in Applied Sciences and Engineering
ECCOMAS Congress 2022
5-–9 June 2022, Oslo, Norway

PARAMETERISING OCEAN-INDUCED MELT OF AN IDEALISED ANTARCTIC ICE SHELF USING DEEP LEARNING

Benjamin Bouissou[1], Clara Burgard[1], and Nicolas C. Jourdain[1]

[1] Univ. Grenoble Alpes, CNRS, IRD, G-INP, IGE, 38000 Grenoble, France

Key words: Ice shelves, Climate science, Geophysical Fluid dynamics, Machine Learning, Neural Networks, Deep Emulators

Abstract. The largest uncertainty when projecting the Antarctic contribution to sea-level rise comes from the ocean-induced melt at the base of Antarctic ice shelves. Current parameterisations used to link the hydrographic properties in front of ice shelves to the melt at their base struggle to accurately simulate basal melt patterns. We suggest that deep learning can be used to tackle this issue. We train a deep feed-forward neural network to emulate basal melt rates simulated by highly-resolved ocean simulations in an idealised geometry. We explore the advantages and limitations of this new approach through sensitivity studies varying hyperparameters, input variables and training choices. We show that large neural networks perform better, that the input format of the temperature and salinity matters most, and that the neural network can be applied to conditions outside of its training range if trained appropriately. The results are promising and we make recommendations for further work with this approach.

1 INTRODUCTION

The Antarctic Ice Sheet has increasingly been losing mass in past decades and this increase in mass loss is projected to continue with increasing greenhouse gas emissions, leading to global sea-level rise [11]. Most of this mass loss is occurring in the form of faster ice flow from the grounded ice sheet to the floating ice shelves at its outskirts, mainly in West Antarctica [21, 26, 28, 17, 31, 32]. This is because the ice shelves, the floating ice tongues that usually buttress the ice flow, are rapidly thinning and retreating due to ocean-induced melt at their base [25, 22, 1]. In some bedrock configurations, increased melt can even trigger marine ice sheet instabilities [33, 29, 14], which have the potential to strongly increase Antarctic mass loss, on timescales below a century [11]. Taking ocean-induced sub-shelf melt into account is therefore key for future sea-level projections.

Sub-shelf melt, or *basal melt*, is a result of oceanic water above the local freezing point getting in contact with the lower side of the ice shelf (Fig. 1, left). 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 [8]. This circulation depends on the size, geometry, and location of the cavity. Therefore, ideally, sea-level projections would be done using a fully coupled ocean–ice-sheet model resolving the ocean circulation in the cavity below the ice shelf, as shown in [7, 30]. However, running coupled ocean–ice-sheet simulations on a circum-Antarctic scale at a resolution on the order of 5 to 10 km is computationally expensive and is therefore

not suitable for multi-centennial scales or for large ensembles, necessary to estimate the range of uncertainty around such simulations.

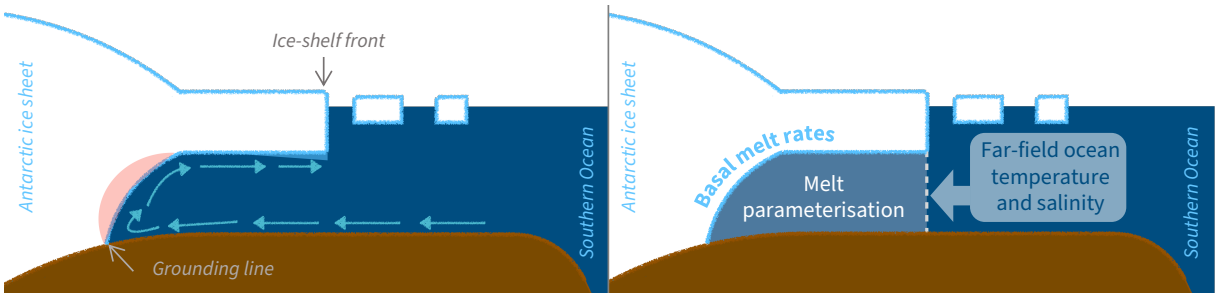


Figure 1: Schematic of a vertical section of an ice-shelf cavity (left) and how a melt parameterisation works (right).

Instead, the Antarctic contribution to sea-level rise is simulated using standalone ice-sheet models forced by oceanic output from global climate models [16], which often poorly represent ocean dynamics along the Antarctic margins and do not resolve ice-shelf cavities [3, 15]. Due to different assumptions and simplifications, existing basal melt parameterisations linking hydrographic properties in front of the ice shelf to basal melt rates (Fig. 1, right) lead to largely differing melt patterns and associated contribution to sea-level rise [10, 5]. The magnitude of the resulting uncertainty contribution is similar, or even larger, than the choice of emission scenario used to force the projections [9].

With the recent development of tools that make the application of deep learning techniques easily accessible, we suggest that deep neural networks are a promising direction to explore new ways of parameterising basal melt. Neural networks are able to reproduce complex non-linear relationships by solving a regression problem between a set of input and output variables. The increasing amount of coupled high-resolution ocean-ice-sheet simulation output that has been produced in diverse Model Intercomparison Projects (MIPs) around ice shelves [2] is a game-changer for the consideration of neural networks. If trained with high-resolution model output, a neural network parameterisation could implicitly include more intrinsic information about the system than a "classical" physical parameterisation. This idea has been applied promisingly in several areas of Earth System Sciences [13, 24, 34, 4].

In the context of ice-shelf basal melt, [27] is, to our knowledge, the only study that applies deep learning to parameterise basal melt. Their model architecture relies on a complex construction including a convolutional neural network and an autoencoder. We suggest that very simple architectures, such as a feed-forward neural network applied on the grid-cell level, can already reproduce the most important features of the relationship between ocean properties in front of the ice shelf and basal melt. In this study, we explore this hypothesis in an idealised case. We train a deep feed-forward neural network to emulate the behaviour of a high-resolution ocean model that resolves the cavities below ice shelves in an idealised geometry. The goal is to better understand the feasibility and limitations of such an approach. To do so, we conduct experiments to identify the sensitivity of the resulting basal melt to the variations in neural network hyperparameters, input variables and to the choice of training data.

2 METHODS

2.1 Neural network

Our neural network is designed to predict basal melt rates based on information about the ocean temperature and salinity in front of the ice shelf and about the ice-shelf geometry. The neural network is a composition $f : \mathbb{R}^n \to \mathbb{R}^p$ of k functions (also called hidden layers), which takes the input $\mathbf{x} \in \mathbb{R}^n$ and outputs the prediction $\hat{\mathbf{y}} \in \mathbb{R}^p$,

$$\hat{\mathbf{y}} = f(\mathbf{x}) = f_k \circ \cdots \circ f_2 \circ f_1(\mathbf{x}). \tag{1}$$

We use a feed-forward neural network architecture, where each of the functions $f_i : \mathbb{R}^m \to \mathbb{R}^o$ maps from the m-dimensional input $\mathbf{h}_{i-1}$ (output from the previous layer containing m neurons) to an o-dimensional output (with the current layer containing o neurons). The i-th layer then reads with its weight matrix $\mathbf{W}_i \in \mathbb{R}^{m \times o}$, its biases $\mathbf{b}_i \in \mathbb{R}^o$, and its activation function $g_i : \mathbb{R}^o \to \mathbb{R}^o$,

$$f_i(\mathbf{h}_{i-1}) = g_i(\mathbf{h}_{i-1}\mathbf{W}_i + \mathbf{b}_i). \tag{2}$$

For our regression task, the prediction of the network is inferred from the output from the last hidden layer f_k without applying the activation function. The implementation of the neural network in this study is done with the Python package Keras [6].

The particularity 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 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 optimised with stochastic gradient descent. An "optimizer algorithm" can be used to optimise the amplitude and direction of the gradient descent steps and rapidly reduce the loss. The learning rate scales the steps of the optimizer. The training takes place iteratively over several batches and several epochs. In parallel to the training, the neural network is applied to a validation dataset to monitor the loss on data that has not been used for the training. In this study, to train the neural network, we reduce the mean-squared-error over all ice-covered points between the predicted $melt_{\mathrm{NN}}$ and reference $melt_{\mathrm{ref}}$ basal melt rates,

$$MSE = \frac{\sum_{i=x_{\min}}^{x_{\max}} \sum_{j=y_{\min}}^{y_{\max}} (melt_{\mathrm{NN}}[i,j] - melt_{\mathrm{ref}}[i,j])^2}{(x_{\max} - x_{\min})(y_{\max} - y_{\min})} \tag{3}$$

One of the purposes of this study is to investigate the sensitivity of the results to the number of hidden layers and to the number of neurons used in the different layers. There is, however, a range of additional hyperparameters that we take as given from the beginning. These are the activation function applied to the hidden layers and the optimizer for the training. We choose to use the "swish" activation function as it was shown to produce as good if not better results compared to the more traditional "ReLU" (Rectified Linear Unit) [23, 27]. Swish is also known to scale better with respect to the neuron architecture, outperforming ReLu on deeper neural

networks. Therefore choosing this activation function will give as much headroom as possible for expansion later on. It is defined as

$$\text{Swish}(\mathbf{x}) = \frac{\mathbf{x}}{1 + e^{-\mathbf{x}}} = \mathbf{x} * \text{Sigmoid}(\mathbf{x}) \tag{4}$$

The optimization of the training is done with the "Adam" algorithm [18], where we use a learning rate equal to 10^{-3}, β_1=0.9 and β_2=0.999, the default values implemented in Keras, unless specified otherwise.

2.2 Experimental setup

In this study, we explore the feasibility of emulating a cavity-resolving ocean model with a neural network. We use an idealised setup to conduct sensitivity experiments to assess the sensitivity of the training of a neural network to its hyperparameters, to the input variables, and to the choice of training and validation data. The aim is to provide recommendations on which neural network emulates best the behaviour of the given ocean model. To do so, we use a range of simulation output provided through previous studies [2, 10].

3 DATA

As the neural network's performance depends on the training, a common saying is that a neural network is "just as good as the data it was trained with". Our goal is to emulate basal melt rates as would be simulated by a cavity-resolving ocean model. In the following, we present the different model configurations we use, the idealised geometry and scenarios, and the available input variables.

3.1 The ocean model

We use simulation output from the version 3.6 of the 3-D primitive-equation coupled ocean–sea-ice model NEMO [19] that was produced for the Ice-Sheet-Ocean Model Intercomparison Project (ISOMIP+) [2]. NEMO solves prognostic equations for the ocean temperature, salinity velocities and resolves ice-shelf cavities, as explained in [20]. Basal melt is computed through three equations describing: (1) the heat balance at the ice-ocean interface, (2) the salt balance at the ice-ocean interface, and (3) the pressure and salinity dependent freezing temperature, as described in [2].

3.2 The ice-sheet model and coupling

When exploring the sensitivity of the training to training and validation datasets in Sec. 4.3, we also use output from coupled ocean-ice simulations that were produced for [10]. The ocean circulation and melt rates are computed with NEMO and the ice response is simulated with finite-element ice-sheet model Elmer/Ice [12]. The detailed setup of Elmer/Ice is described in [10]. Melting is only applied to floating nodes, meaning that the first floating element (partially or not) may be affected by melting. No dynamic calving criterion was applied as opposed to the ocean-only runs, allowing the ice to be as thin as possible. The heat transfer parameter Γ_T used in the computation of the melt is 4×10^{-2} compared to 2.6×10^{-2} in the ocean-only experiments. More details about the setup used for coupling NEMO and Elmer/Ice is described in [10].

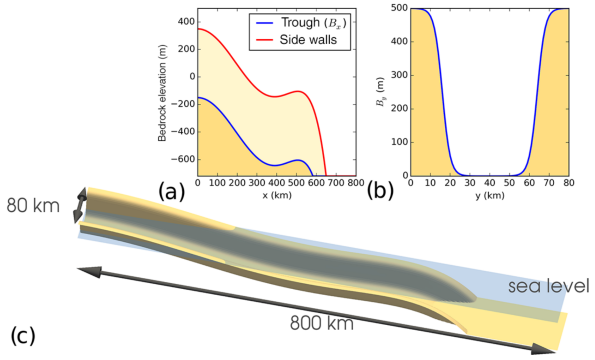


Figure 2: Bedrock topography for ISOMIP+ as shown in Fig. 1 of [2].

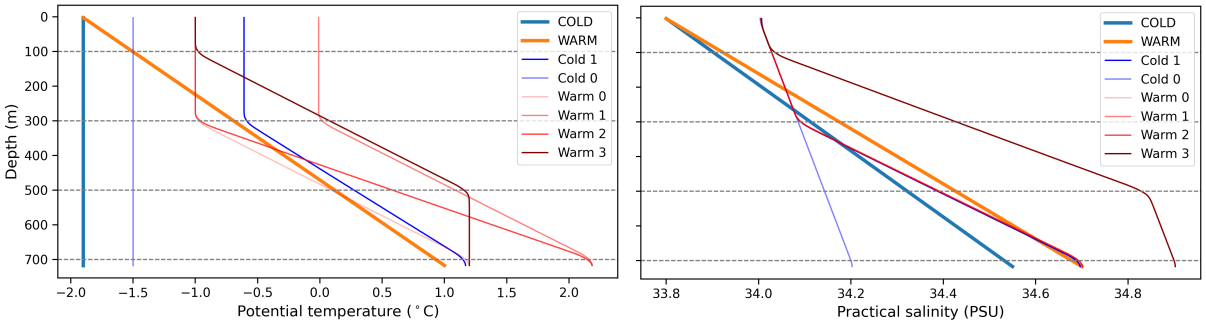
3.3 The idealised geometry and scenarios

The simulations used in the study were run in an idealised geometry defined for a range of model intercomparison projects, amongst others the Ice-Sheet-Ocean Model Intercomparison Project ISOMIP+ and the Marine Ice Sheet Ocean Model Intercomparison Project MISOMIP [2] (Fig. 2). The domain is a box bounded by $0 \leq x \leq 800$ km and $0 \leq y \leq 80$ km, where the x-axis spans along the ice-shelf and the y-axis spans across the ice shelf. The ice shelf initially covers the domain bounded by $459 \leq x \leq 640$ km. In the ocean-only simulations, this stays constant while in the coupled simulations, the ice shelf evolves with time. The z-axis spans 144 vertical levels. The horizontal resolution is 2 km and the vertical resolution is 5 m. More details about the exact configuration of the bathymetry and other prescribed parameters can be found in [2].

In Sec. 4.1 and 4.2, we use NEMO simulation output conducted for ISOMIP+ under four idealised scenarios forced with the WARM and COLD profiles shown in thick lines in Fig. 3. The forcing is applied on the far field at around $x = 800$ km (more details in [2]). Two scenarios are the COLDWARM and WARMCOLD, both run on 20 years with a fixed ice-sheet geometry, starting with COLD and WARM conditions respectively and forced with WARM and COLD conditions, respectively. The other two scenarios are the WARMWARM and COLDCOLD scenarios, both run on 100 years with a prescribed change in ice-sheet geometry, forced with the same conditions as their initialised state. We combine all these simulation outputs and randomly sample 80%, which become our training dataset, while the other 20% become our validation dataset. In Sec. 4.3, we add output from eight coupled experiments under six different forcings, produced for [10]. The profiles used for the different forcings are shown in Fig. 3.

3.4 Variables

The goal of a basal melt parameterisation is to link hydrographic properties to the melt happening at the ocean-ice interface in the cavity. To assess the potential of the neural network to emulate the processes at work, we need to use input variables that we could also gain from a model that does not resolve the cavities. As input to our neural network, we therefore use the temperature and salinity in front of the ice shelf given by the simulations and geometric information at each grid cell, such as the ice draft depth, the bathymetry, the slope of the



COLD: T (surface and bottom) = -1.9°C, S (surface) = 33.8 PSU, S (bottom) = 34.55 PSU
WARM: T (surface) = -1.9°C, T (bottom) = 1.0°C, S (surface) = 33.8 PSU, S (bottom) = 34.7 PSU

Cold0: Conditions resembling cold cavity from Filchner-Ronne, no temporal change

Cold1: Cold0 T => warms up to match the temperature of a warm cavity within a century, S increase
Warm0: Present-day Amundsen sea conditions, no temporal change
Warm1: Warm0 T => +1°C/century, no change in S
Warm2: Warm0 T => +0°C/century at surface to +1°C/century below thermocline
Warm3: Warm0 T and uplifts the thermocline and halocline by 200m

Figure 3: Forcing scenarios and associated temperature (T) and salinity (S) profiles for the set of simulations used in different parts of this study.

bathymetry and of the ice draft, and the distance to the grounding line and to the ice front. The output variable we aim to emulate is the basal melt rate at each grid point, as computed in the simulations.

We aim to assess the applicability of a feed-forward neural network structure, acting on the grid-cell level, to parameterise the link between hydrographic properties in front of the ice shelf and basal melt rates. The main difficulty is that this type of neural networks takes input at the grid-cell level, on the x-y-plane covering the ice-shelf domain. While the information about the ice and bedrock geometry is given on this x-y-plane, our input temperature and salinity are given across the front of the ice shelf over depth, on a y-z-plane. The information is therefore not yet on the same grid. One sensitivity study therefore focusses on the effect of the method used to extrapolate the input temperature and salinity onto the x-y-plane (see Sec. 4.2).

Before starting to train a neural network, normalising the training dataset is strongly advised as it puts every input and output variable on a similar order of magnitude and avoids potential gradient explosion problems. After trying out several normalisation approaches, we settle for the common approach of subtracting the mean μ and dividing by the standard deviation σ. This results in a normalised value $\mathbf{X}'$ for each input and output variable $\mathbf{X}$,

$$\mathbf{X}' = \frac{\mathbf{X} - \mu}{\sigma} \tag{5}$$

4 RESULTS

4.1 Sensitivity of the training to neural network hyperparameters

In this section, we explore the importance of the size of the neural network, in terms of number of layers and neurons, on the performance of the neural network. As a first step, we want to confirm that this problem calls for a deep neural network and cannot simply be solved with a linear regression. To do so, we construct two basic neural networks. On the

one hand, we construct a very simple neural network with an input and output layer, without any hidden layer, equivalent to a linear regression. On the other hand, we construct a deep neural network with three hidden layers with 32, 64, 32 neurons respectively. We train the two networks over eight epochs using 80% of the output from the four idealised NEMO-only simulations. As input variables, we use the temperature and salinity extrapolated to the ice draft depth (method EXTRAPz explained more in detail in Sec. 4.2), and the ice draft depth. To assess the performance after training, we apply the network to the whole dataset (training and validation) and we compute the RMSE of the integrated melt over all timesteps. The RMSE of the linear regression is 5.2 Gt/yr, while the RMSE of the deep neural network is 2.7 Gt/yr. This shows that a neural network brings a clear advantage to address this issue.

As a next step, we explore how large our neural network has to be. A general rule of thumb for neural networks is: The more complex the problem is, the more neurons are required. It becomes a problem of finding a balance between (1) having a large enough neural network to learn the complex relationship, (2) avoiding overfitting where the neural network learns too much specificities from the training data, and (3) avoiding too high computational cost during training. We train multiple neural networks using different neuron architectures, from very simple (1 layer containing 1 neuron) to deep (5 layers of 128 neurons each), and compute their associated error with respect to the reference melt. All the neural networks are trained using the same validation and training dataset (the same as before), on eight epochs and with a batch size of 128.

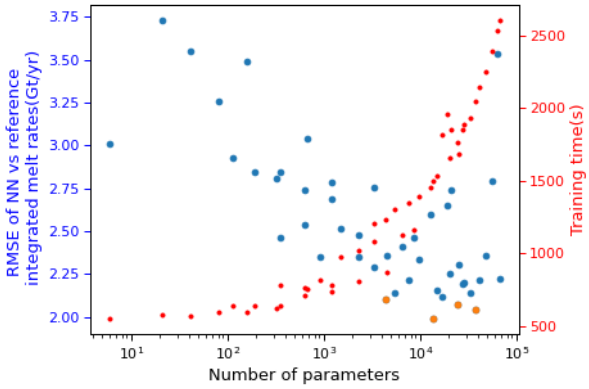


Figure 4: RMSE of the integrated melt [in Gt/yr] with respect to the reference (blue y-axis) and training time [in s] (red y-axis) as a function of the number of parameters of the different neural networks. The orange points represent the configurations with the lowest RMSE. Note that the x-axis is on a logarithmic scale.

The general trend is that a higher parameter number, i.e. more layers and/or more neurons per layers, results in a lower RMSE of the integrated melt (Fig. 4). The training time increases rapidly with the parameter number. We select the neural network with four layers and a neuron configuration of 32-32-96-96 for further testing as it leads to the lowest RMSE and is relatively lightweight, making it quicker to apply multiple variations of training.

4.2 Sensitivity of the training to input variables

As mentioned in Sec. 3.4, one of the challenges of using a feed-forward neural network in our case is that the input temperature and salinity profiles are not on the same plane as the geometric information (such as e.g. the ice draft depth) and as the output variable, the basal melt rate.

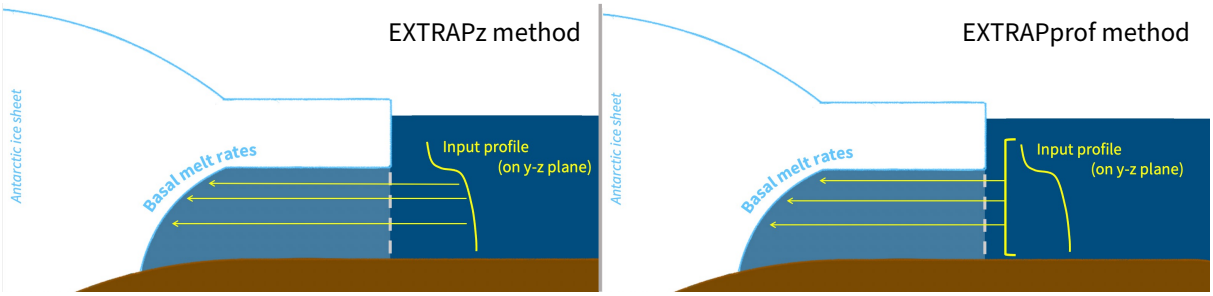


Figure 5: Schematic to explain the two approaches explored to extrapolate the temperature and salinity profiles from the y-z plane to the x-y plane.

We therefore explore two ways of extrapolating the input profiles from the y-z-plane in front of the ice shelf to the x-y-plane of the ice-ocean interface (Fig. 5). In the method EXTRAPz, we extrapolate the temperature and salinity to the ice draft depth of each ice-shelf point on the same y-coordinate, which means that $\mathrm{T}(x,y) = \mathrm{T}(y,z_{\mathrm{draft}})$ and $\mathrm{S}(x,y) = \mathrm{S}(y,z_{\mathrm{draft}})$, resulting in one temperature and salinity information per grid point. In the method EXTRAPprof, we feed the whole profile to each point. To reduce the amount of data, we choose to reduce the number of vertical levels from 144 to 40. This means that $\mathrm{T}(x,y) = \mathrm{T}(y,z)$ and $\mathrm{S}(x,y) = \mathrm{S}(y,z)$ are flattened into 40 temperature and 40 salinity input variables for each grid point respectively.

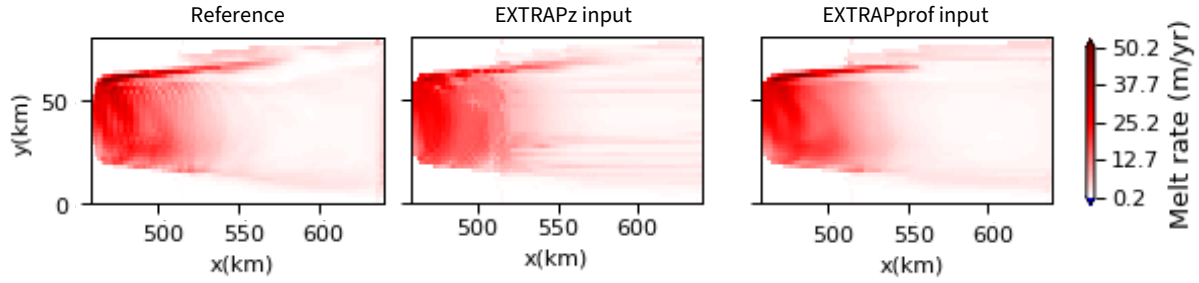


Figure 6: Melt rates averaged over time for the forcing scenario COLDWARM. Left: Reference, Middle: Parameterised using ice draft depth and the EXTRAPz method to extrapolate temperature and salinity to the ice-shelf domain. Right: Parameterised using ice draft depth and the EXTRAPprof method to extrapolate temperature and salinity to the ice-shelf domain.

Training our neural network (32-32-96-96) on 32 epochs using as input only the ice draft depth and the two variations of the input temperature and salinity shows that using the flattened profile as input is an important source of information as the RMSE of the integrated melt decreases

from 3.10 Gt/yr to 1.40 Gt/yr. A clear improvement can also be seen in the spatial patterns (Fig. 6). This means that giving the whole profile of temperature and salinity helps the neural network to learn to some extent that the ocean circulation in the cavity has a horizontal and vertical component and cannot simply be ignored, as is assumed in the EXTRAPz method.

As a next step, we add additional geometric information as input. Introducing the distance to the grounding line and to the ice front reduces the RMSE of the integrated melt to 0.97 Gt/yr. Introducing additionally the bathymetry and the norm of the slopes of the ice draft and bathymetry further reduces the RMSE to 0.77 Gt/yr. Finally, introducing the slope components in x- and y-direction instead of the norm, to give information about the sign of the slope, leads to a RMSE of 0.58 Gt/yr.

As a conclusion, we recommend using the extrapolated whole profiles of temperature and salinity and all available geometric information provided to refine the results of the neural network. The order of magnitude of the RMSE in the integrated melt (lower than 1 Gt/yr) is promising as the reference value of the integrated melt can reach up to 120 Gt/yr.

To accommodate the neural network for these findings and therefore larger input dataset, we increase the architecture to five layers, each containing 96 neurons and train it for 128 epochs in the following. Additionally, we apply an evolution to the learning rate, where the learning rate is divided by two if the RMSE on the validation dataset does not improve beyond 0.007 for a duration of ten epochs.

4.3 Sensitivity of the training to training and validation data

Finally, we explore the sensitivity of the training to the training and validation dataset. To do so, we introduce the subset of coupled simulations, as described in Sec. 3.3. We apply our final neural network structure (see end of Sec. 4.2) to two of these simulations, namely the Cold1 and the Warm2 (see Fig. 3 for the associated forcing). It does not perform well (Fig. 7, orange). We suggest that this is mainly due to the forcing temperature and salinity profiles being outside of the range used during training. Some of the discrepancy could also be because, in the newly introduced simulations, there is no calving criterion at 100 m and the parameter Γ_T used in the ocean model differs slightly between the coupled and uncoupled version (see Sec. 3.2). We therefore add a subset of the coupled scenarios (Warm0, Warm1, Warm3, Cold0) to the training dataset to investigate if this is enough information to improve the applicability of the neural network on other simulations. This "hybrid" neural network performs much better, even on scenarios that were not used in the training (Fig. 7, green).

5 DISCUSSION & CONCLUSION

Our results show that it is possible to train a feed-forward neural network to emulate the link between hydrographic properties in front of the ice shelf and melt rates at the base of the ice shelf in an idealised geometry. This is promising for further exploration of the application of this approach in more realistic geometries. We show that larger neural networks perform better and suggest to use an architecture with five layers and 96 neurons each. We also show that the format of the input temperature and salinity matters and that giving information about the whole profile leads to a better performance. Additional geometric information further refines the results. Finally, after including additional simulations in the training, the neural network

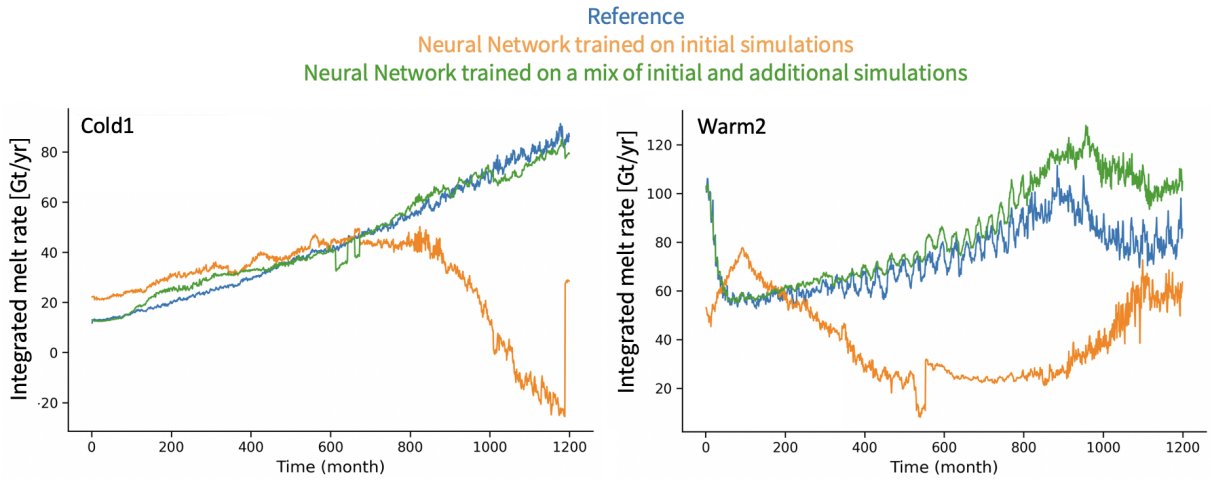


Figure 7: Integrated melt rates in Gt/yr in the scenarios Cold1 and Warm2 for the reference simulation (blue), a neural network trained on the ocean-only runs (orange) and a neural network trained on the ocean-only runs and the coupled scenarios Warm0, Warm1, Warm3, Cold0 (green).

was able to adapt to represent the melt in scenarios that were not used for training.

These findings are a first exploration and promising for further work on this approach. Still, a few limitations remain, which we have started exploring and want to explore further in future work. First, from a physical point of view, this simple type of neural networks remains a black box. We have started investigating the importance of the variables with a "variable shuffling" method, in which the different input variables are shuffled individually before being fed to the network. For the first set of simulations (forced by COLD and WARM profiles), shuffling the input temperature and the ice draft depth influences most the results, while for the second set of simulations, shuffling the input salinity and bathymetry has the most influence on the results of the neural network. These results and discrepancy between the two sets are not straightforward to interpret. Further work is planned to better understand these results.

Second, the training and validation dataset were selected randomly from the pool of simulation outputs. As the points are highly autocorrelated in time and space, we suggest that this could have biased the training and results of the neural network. We are currently exploring ways to sample the training and validation dataset more systematically, for example by dividing the data in coherent blocks in time and space, that might be less correlated between each other.

Finally, several ways forward can be explored. On the one hand, this study only focussed on the emulation of one single model. Through the Model Intercomparison Projects, several models have produced similar simulation output. It could be interesting to explore how neural networks emulating the different models compare and how a neural network trained on all of them performs. On the other hand, this study focussed on an idealised geometry. It is important to look forward and apply it to a realistic geometry and investigate the new challenges and opportunities that arise in that context.

Acknowledgments: The analysis 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 HPC resources from GENCI–CINES (MISOCS project, allocations A0080106035 and A0100106035). The authors were supported by the ANR project EIS (ANR-19-CE01-0015) and the H2020 project PROTECT. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869304.

REFERENCES

[1] S. Adusumilli, H.A. Fricker, B. Medley, L. Padman, and M.R. Siegfried. Interannual variations in meltwater input to the Southern Ocean from Antarctic ice shelves. *Nature Geoscience*, 13:616–620, 2020.

[2] X.S. Asay-Davis, S.L. Cornford, G. Durand, B.K. Galton-Fenzi, R.M. Gladstone, G.H. Gudmundsson, T. Hattermann, D.M. Holland, D. Holland, P.R. Holland, D.F. Martin, P. Mathiot, F. Pattyn, and H. Seroussi. 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). *Geosci. Model Dev.*, 9(7):2471–2497, 2016.

[3] R.L. Beadling, J.L. Russell, R.J. Stouffer, M. Mazloff, L.D. Talley, P.J. Goodman, J.B. Sallée, H.T. Hewitt, P. Hyder, and A. Pandde. Representation of Southern Ocean Properties across Coupled Model Intercomparison Project Generations: CMIP3 to CMIP6. *Journal of Climate*, 33(15):6555–6581, 2020.

[4] T. Bolton and L. Zanna. Applications of Deep Learning to Ocean Data Inference and Subgrid Parameterization. *Journal of Advances in Modeling Earth Systems*, 11(1):376–399, 2019.

[5] C. Burgard, N.C. Jourdain, R. Reese, A. Jenkins, and P. Mathiot. An assessment of basal melt parameterisations for Antarctic ice shelves. *The Cryosphere Discussions*, 2022.

[6] François Chollet et al. Keras. https://keras.io, 2015.

[7] J. De Rydt and G.H. Gudmundsson. Coupled ice shelf-ocean modeling and complex grounding line retreat from a seabed ridge. *Journal of Geophysical Research: Earth Surface*, 121(5):865–880, 2016.

[8] M. Dinniman, X. Asay-Davis, B. Galton-Fenzi, P. Holland, A. Jenkins, and R. Timmermann. Modeling Ice Shelf/Ocean Interaction in Antarctica: A Review. *Oceanography*, 29(4):144–153, 2016.

[9] T.L. Edwards and the ISMIP6 Team. Projected land ice contributions to twenty-first-century sea level rise. *Nature*, 593(7857):74–82, 2021.

[10] L. Favier, N.C. Jourdain, A. Jenkins, N. Merino, G. Durand, O. Gagliardini, F. Gillet-Chaulet, and P. Mathiot. 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, 2019.

[11] B. Fox-Kemper, H.T. Hewitt, C. Xiao, G. Adalgeirsdóttir, S.S. Drijfhout, T.L. Edwards, N.R. Golledge, M. Hemer, R.E. Kopp, G. Krinner, A. Mix, D. Notz, S. Nowicki, I.S. Nurhati, L. Ruiz, J-B. Sallée, A.B. Slangen, and Y. Yu. *Ocean, Cryosphere and Sea Level Change*, chapter 9. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2021.

[12] O. Gagliardini, T. Zwinger, F. Gillet-Chaulet, G. Durand, L. Favier, B. de Fleurian, R. Greve, M. Malinen, C. Martín, P. Råback, J. Ruokolainen, M. Sacchettini, M. Schäfer, H. Seddik, and J. Thies. Capabilities and performance of Elmer/Ice, a new-generation ice sheet model. *Geosci. Model Dev.*, 6:1299–1318, 2013.

[13] P. Gentine, M. Pritchard, S. Rasp, G. Reinaudi, and G. Yacalis. Could Machine Learning Break the Convection Parameterization Deadlock? *Geophysical Research Letters*, 45(11):5742–5751, 2018.

[14] G.H. Gudmundsson, J. Krug, G. Durand, L. Favier, and O. Gagliardini. The stability of grounding lines on retrograde slopes. *The Cryosphere*, 6(6):1497–1505, 2012.

[15] C. Heuzé. Antarctic Bottom Water and North Atlantic Deep Water in CMIP6 models. *Ocean Science*, 17(1):59–90, 2021.

[16] N.C. Jourdain, X. Asay-Davis, T. Hattermann, F. Straneo, H. Seroussi, C.M. Little, and S. Nowicki. A protocol for calculating basal melt rates in the ISMIP6 Antarctic ice sheet projections. *The Cryosphere*, 14(9):3111–3134, 2020.

[17] A. Khazendar, E. Rignot, D.M. Schroeder, H. Seroussi, M.P. Schodlok, J. Scheuchl, B.and Mouginot, T.C. Sutterley, and I. Velicogna. Rapid submarine ice melting in the grounding zones of ice shelves in West Antarctica. *Nature Communications*, 7(1):13243, 2016.

[18] D.P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv*, 2014.

[19] G. Madec and the NEMO Team. NEMO ocean engine, 2016. NEMO reference manual 3_6_STABLE, Note du Pôle de modélisation, Institut Pierre-Simon Laplace (IPSL) No 27. ISSN No 1288-1619.

[20] P. Mathiot, A. Jenkins, C. Harris, and G. Madec. 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, 2017.

[21] J. Mouginot, E. Rignot, and B. Scheuchl. Sustained increase in ice discharge from the Amundsen Sea Embayment, West Antarctica, from 1973 to 2013. *Geophysical Research Letters*, 41(5):1576–1584, 2014.

[22] F.S. Paolo, H.A. Fricker, and L. Padman. Volume loss from Antarctic ice shelves is accelerating. *Science*, 348(6232):327–331, 2015.

[23] P. Ramachandran, B. Zoph, and Q. Le. Searching for activation functions. *arXiv*, 2018.

[24] S. Rasp, M.S. Pritchard, and P. Gentine. Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences*, 115(39):9684–9689, 2018.

[25] E. Rignot, S. Jacobs, J. Mouginot, and B. Scheuchl. Ice-shelf melting around Antarctica. *Science*, 341(6143):266–270, 2013.

[26] E. Rignot, J. Mouginot, M. Morlighem, H. Seroussi, and B. Scheuchl. 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, 2014.

[27] S.H.R. Rosier, C.Y.S. Bull, and G.H. Gudmundsson. Predicting ocean-induced ice-shelf melt rates using a machine learning image segmentation approach. *The Cryosphere Discussions*, 2022.

[28] J. Scheuchl, B.and Mouginot, E. Rignot, M. Morlighem, and A. Khazendar. 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, 2016.

[29] C. Schoof. Ice sheet grounding line dynamics: Steady states, stability, and hysteresis. *J. Geophys. Res.*, 112(F3):F03S28, 2007.

[30] H. Seroussi, Y. Nakayama, E. Larour, D. Menemenlis, M. Morlighem, E. Rignot, and A. Khazendar. Continued retreat of Thwaites Glacier, West Antarctica, controlled by bed topography and ocean circulation. *Journal of Geophysical Research: Earth Surface*, 44(12):6191–6199, 2017.

[31] Q. Shen, K. Wang, C.K. Shum, L. Jiang, H.T. Hsu, and J. Dong. Recent high-resolution Antarctic ice velocity maps reveal increased mass loss in Wilkes Land, East Antarctica. *Scientific Reports*, 8(1):4477, 2018.

[32] The IMBIE Team. Mass balance of the Antarctic Ice Sheet from 1992 to 2017. *Nature*, 558(7709):219–222, 2018.

[33] J. Weertman. Stability of the Junction of an Ice Sheet and an Ice Shelf. *Journal of Glaciology*, 13(67):3–11, 1974.

[34] J. Yuval and P.A. O'Gorman. Stable machine-learning parameterization of subgrid processes for climate modeling at a range of resolutions. *Nature communications*, 11(1):1–10, 2020.