Accelerating subglacial hydrology for ice sheet models with deep learning methods

Vincent Verjans[1] and Alexander A Robel[1]

[1]Georgia Institute of Technology

July 9, 2023

Abstract

Subglacial drainage networks regulate the response of ice sheet flow to surface meltwater input to the subglacial environment. Simulating subglacial hydrology evolution is critical to projecting ice sheet sensitivity to climate, and contribution to sea-level change. However, current numerical subglacial hydrology models are computationally expensive, and, consequently, evolving subglacial hydrology is neglected in large-scale ice sheet simulations. We present a deep learning emulator of a state-of-the-art subglacial hydrology model, trained at multiple Greenland glaciers. Our emulator performs strongly in both temporal (R2>0.99) and spatial (R2>0.96) generalization, offers high computational savings, and can be used to force numerical ice sheet models. This will enable century- and large-scale ice sheet model simulations, including interactions between ice flow and increased meltwater input to the subglacial environment. Generally, our work demonstrates that machine learning can further improve ice sheet models, reduce computational bottlenecks, and exploit information from high-fidelity models and novel observational platforms.

Accelerating subglacial hydrology for ice sheet models with deep learning methods

Vincent Verjans[1], Alexander Robel[1]

[1]School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, GA, USA

Key Points:

- We develop a deep learning emulator to simulate evolving subglacial hydrology in response to meltwater input for ice sheet simulations.
- The emulator shows generalization capabilities, large computational savings, and can be used to force numerical ice sheet models.
- We demonstrate that machine learning has substantial potential in improving ice sheet models, through using information-rich data sets.

Corresponding author: Vincent Verjans, `vverjans3@gatech.edu`

Abstract

Subglacial drainage networks regulate the response of ice sheet flow to surface meltwater input to the subglacial environment. Simulating subglacial hydrology evolution is critical to projecting ice sheet sensitivity to climate, and contribution to sea-level change. However, current numerical subglacial hydrology models are computationally expensive, and, consequently, evolving subglacial hydrology is neglected in large-scale ice sheet simulations. We present a deep learning emulator of a state-of-the-art subglacial hydrology model, trained at multiple Greenland glaciers. Our emulator performs strongly in both temporal (R^2 >0.99) and spatial (R^2 >0.96) generalization, offers high computational savings, and can be used to force numerical ice sheet models. This will enable century- and large-scale ice sheet model simulations, including interactions between ice flow and increased meltwater input to the subglacial environment. Generally, our work demonstrates that machine learning can further improve ice sheet models, reduce computational bottlenecks, and exploit information from high-fidelity models and novel observational platforms.

Plain Language Summary

Meltwater at the surface of ice sheets can drain to the subglacial environment, lubricate the bed, and influence ice sheet flow. Complex numerical subglacial hydrology models represent the subglacial drainage system, but are too computationally expensive to be included in large-scale and long-term ice sheet simulations. Consequently, model predictions of future ice sheet contribution to sea-level rise ignore ice flow modulation by evolving subglacial hydrology. Here, we use deep learning to emulate a state-of-the-art subglacial hydrology model. The emulator can directly force large-scale ice sheet models to capture ice flow sensitivity to subglacial hydrology. The computational speed and accuracy of our emulator show the potential to use machine learning to efficiently incorporate previously neglected processes into ice sheet models.

1 Introduction

The Greenland ice sheet has experienced accelerating mass loss since the early 1990s (Otosaka et al., 2023). Ice loss has been driven by increasing surface melt (Fettweis et al., 2016) and accelerating ice flow into the ocean (King et al., 2020). These two processes are linked by surface meltwater drainage into the subglacial environment. The evolution of the subglacial drainage system in response to meltwater input determines the subglacial water pressure, which regulates the speed of ice sliding over the bed (Nienow et al., 2017). Observations have demonstrated a strong sensitivity of ice flow speed to meltwater supply to the bed on timescales ranging from days to months (Zwally et al., 2002; Shepherd et al., 2009; Smith et al., 2021). The subglacial drainage system modulates this sensitivity, and there is no simple relationship between meltwater forcing and ice flow speed, due to the complexities of subglacial hydrology (Bartholomew et al., 2011; van de Wal et al., 2015).

Subglacial hydrology models simulate the evolution of subglacial water pressure under different conditions of ice sheet geometry and meltwater input (Werder et al., 2013; de Fleurian et al., 2018). They represent the transient evolution of the subglacial drainage system, are highly complex, and require many parameters and substantial computational expense (Werder et al., 2013; Hoffman et al., 2016). When applied to individual Greenland glaciers, studies have shown that accurately representing the evolution of the subglacial hydrology system with such complex models is necessary to explain observed variability in ice flow (Hewitt, 2013; Hoffman et al., 2016; Ehrenfeucht et al., 2022). However, their computational expense prohibits long, ice sheet scale simulations. In contrast, simple first-order formulations of the growth and decay of subglacial water flux do exist, facilitating large-scale simulations (Kazmierczak et al., 2022). But these formulations do not represent the different drainage components and assume a constant transmissivity. As such, they have critical limitations when considering high meltwater input scenarios, complex topographies, or reproducing realistic variability in subglacial hydrology and ice flow (de Fleurian et al., 2018).

Most subglacial hydrology models divide subglacial drainage into distributed and channelized systems (Schoof, 2010; Werder et al., 2013; de Fleurian et al., 2018). The former is typical of the early melt season, when the subglacial hydrology system is inefficient and water pressure increases strongly with meltwater supply. Later in the melt season, the system develops into a channelized system, efficiently evacuating meltwater, and causes water pressure to decrease with increasing meltwater input (Schoof, 2010; Bartholomew et al., 2011; Cowton et al., 2013). Ice flow variability is sensitive to the representation of, and transitions between different forms of drainage. Nevertheless, because of their computational expense, subglacial hydrology models are not included in model simulations at ice sheet scale (Goelzer et al., 2020; Seroussi et al., 2020). Consequently, current sea-level projections ignore a critical process regulating ice flow.

The climate modeling community has used machine learning techniques successfully to parameterize processes unresolved in coarse resolution global models (e.g., Brenowitz & Bretherton, 2018; Rasp et al., 2018). In particular, artificial neural networks (ANNs) are particularly powerful tools for parameterizing complex relationships between input and output variables, as they are capable of approximating any continuous function (the universal approximation theorem, Hornik et al., 1989). Furthermore, recent improvements in computational hardware, software, and optimization techniques have led to important ANN developments in multiple fields, including the Earth sciences (LeCun et al., 2015; Reichstein et al., 2019). Once trained, ANN models are computationally efficient, and ANNs have been used previously to emulate glacier flow models (Brinkerhoff et al., 2021; Jouvet et al., 2021). In this study, we use deep learning to enable representation of subglacial hydrology in large-scale ice sheet model simulations. Specifically, we develop an ANN emulator of the Glacier Drainage System model (GlaDS, Werder et al., 2013), an advanced and computationally expensive subglacial hydrology model. More generally, our work is a proof of concept for an important advancement in ice sheet modeling: we demonstrate that deep learning techniques can replace computationally-demanding or poorly constrained processes in large-scale ice sheet models.

2 Methods

We run GlaDS for 40 years at eight major Greenland glaciers (Petermann, Jakobshavn, Helheim, Kangerlussuaq, Humboldt, Koge Bugt, Russell, and Upernavik, see Fig. S1 for locations). The glacier geometries and ice flow velocities are taken from present-day observations (Joughin et al., 2017; Morlighem et al., 2017), and the meltwater runoff forcing from the 1970-2009 output of the diurnal Energy Balance Model (Krebs-Kanzow et al., 2020) (see Supporting Information). GlaDS simulates the evolution of the subglacial hydraulic potential, ϕ, in time and space by representing both channelized and distributed drainage. Accurate representation of ϕ in ice sheet models is critical, as it directly determines the subglacial water pressure, p_w. In turn, p_w determines the effective pressure at the ice-bed interface, N:

$$\begin{cases} N = p_{ice} - p_w \\ p_w = \phi - g\rho_w B \\ p_{ice} = g\rho_{ice}H_{ice} \end{cases}, \tag{1}$$

where p_{ice} is ice pressure [Pa], g is gravitational acceleration [m s^{-2}], H_{ice} is thickness of the above-lying ice column [m], B is bed elevation [m], and ρ_w and ρ_{ice} are water and ice density [kg m^{-3}], respectively. Critically, N [Pa] is a key variable in basal sliding laws for ice flow (Budd et al., 1984; Hewitt, 2013). We run GlaDS at a 2-hourly time step to preserve numerical stability.

Dynamics of ϕ are governed by the amount of surface meltwater draining through the ice sheet to the bed, followed by the routing of the water through the subglacial system. The goal of our ANN is to predict ϕ based on ice sheet state and meltwater runoff forcing, such that our p_w emulation accounts for spatio-temporal evolution of subglacial hydrology. Specifically, our ANN uses as inputs ice thickness, ice velocities, bedrock topography, and meltwater runoff fields, as well as moulin locations where meltwater reaches the subglacial domain. All these variables are common variables, parameters, or inputs to typical ice sheet models. Similar to current subglacial hydrology models, our ANN is aimed at one-way coupling with ice sheet models, i.e., it needs to be run prior to the ice sheet model and the ANN output is subsequently used as a forcing to the ice sheet model. We discuss prospects for full two-way coupling in the *Discussion* section.

Our ANN is a convolutional neural network, based on the U-Net architecture (Ronneberger et al., 2015). Our ANN thus uses two-dimensional input fields, and outputs a two-dimensional ϕ field at any given time step (see Supporting Information). In total, the ANN has 259,953 trainable parameters that are optimized such that ANN predictions of ϕ match training targets with accuracy. We calibrate our ANN to the GlaDS 1975-2004 output at seven of the eight glaciers. We keep the last 5 years (2005-2009) of output at these seven glaciers as test data to evaluate the temporal generalization capabilities of the ANN. Furthermore, we keep all the output of the eighth glacier for test data in order to evaluate the ANN spatial generalization capabilities. All the GlaDS output test data were totally unseen by the deep learning algorithm or the authors during the calibration. The calibration data are separated into training data, used to optimize the ANN parameters, and validation data, used to optimize hyperparameters and to avoid overfitting (see Supporting Information). All the results presented in the next sections have been computed on the test data.

3 Results

3.1 Temporal Generalization Performance

To assess the ability of the ANN to reproduce GlaDS hydraulic potential (ϕ) fields, we start by comparing their respective outputs at the 7 calibration glaciers over the last 5 years of simulations (2005-2009), which have not been used for the ANN training. Figures 1a and 1b show for one of the calibration glaciers (Helheim glacier) that the mean ϕ field of the ANN over these 5 test years reproduces the spatial patterns of GlaDS output well. In particular, the ANN captures the radial patterns of high ϕ values and variability centered at moulin locations, where meltwater runoff drains to the subglacial environment. The ANN performs well throughout the domain, as the root-mean-square error (RMSE) is mostly lower than 0.5 MPa (Fig. 1c). The temporal dynamics are also captured well, as ϕ time series show close correspondence between ANN and GlaDS outputs in most of the domain (Fig. 1d). The ANN captures the different ranges in seasonality and inter-annual variability of ϕ. It also reproduces the asymmetry between early and late melt season, which is due to changing drainage efficiency (Nienow et al., 2017). In Figure 1d, the orange curve shows a time series selected within the high RMSE area (Fig. 1c). This example illustrates that the ANN still captures the temporal variability, but is biased low at this particular location.

Figure 1e shows the ANN performance over the 5 years of test data at all the seven calibration glaciers, demonstrating low bias and RMSE. Furthermore, the ANN explains >99% of the variance in GlaDS ϕ output, as quantified by the coefficient of determination (R^2). These results demonstrate that the ANN is able to predict ϕ with good accuracy over years, and thus meltwater input conditions, unseen during training.

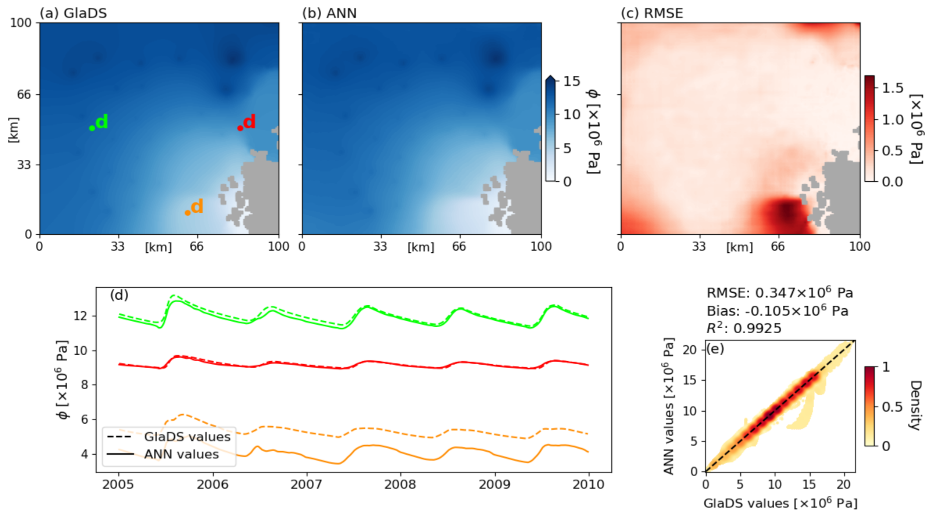

Figure 1. Temporal generalization performance of the ANN over the test period. Maps of mean 2005-2009 ϕ fields at one calibration glacier (Helheim, Fig. S1 for location) simulated by (a) GlaDS, and (b) the ANN. The Root-Mean-Square Error (RMSE) of the ANN with respect to GlaDS is shown in (c). Time series (d) of ϕ at specific grid points simulated by GlaDS (dashed lines) and the ANN (solid lines), with color-coded locations shown in (a). Performance statistics (e) of the ANN with respect to GlaDS evaluated at all the grid points of the 7 calibration glaciers for all the 2005-2009 time steps.

3.2 Spatial Generalization Performance

We now perform a similar evaluation, but at the test glacier (Upernavik), not included in the ANN calibration. This task presents a harder challenge, as both years of meltwater runoff and glaciological characteristics have not been used in calibration. The time-mean spatial patterns of ϕ at the test glacier are well-reproduced (Fig. 2a, 2b). The RMSE is low throughout the domain (Fig. 2c), except on a small portion near the glacier terminus. The time series shown in Figure 2d demonstrate that the ANN performance in reproducing temporal dynamics at seasonal and inter-annual scales are similar to its performance on calibration glaciers (compare with Fig. 1d). Here also, the orange lines in Fig. 2d show a time series corresponding to a low-performance location. Again, the ANN still captures temporal dynamics correctly, but has a consistent bias over the time series at this location. Finally, the metrics of performance for this test glacier show an RMSE lower than 0.56 MPa, a small positive bias of 0.24 MPa, and R^2 >0.96.

3.3 Sensitivity to number of calibration years

To calibrate the ANN, we have used 30 years (1975-2004) of GlaDS output at the calibration glaciers, and preserved 5 years (2005-2009) of output for testing (see Supporting Information). In this section, we investigate the sensitivity of ANN accuracy to the number of calibration years. Starting with all 30 years of available GlaDS output, we reduce the calibration data by 3-year increments, re-train the ANN at each increment, and evaluate performance metrics always on the 2005-2009 years.

Figure 3a shows the ratios in the R^2, RMSE and absolute bias metrics for each sensitivity experiment, with respect to the results computed when all the 30 calibration years are used in the ANN calibration. There is no decrease in accuracy for calibration data

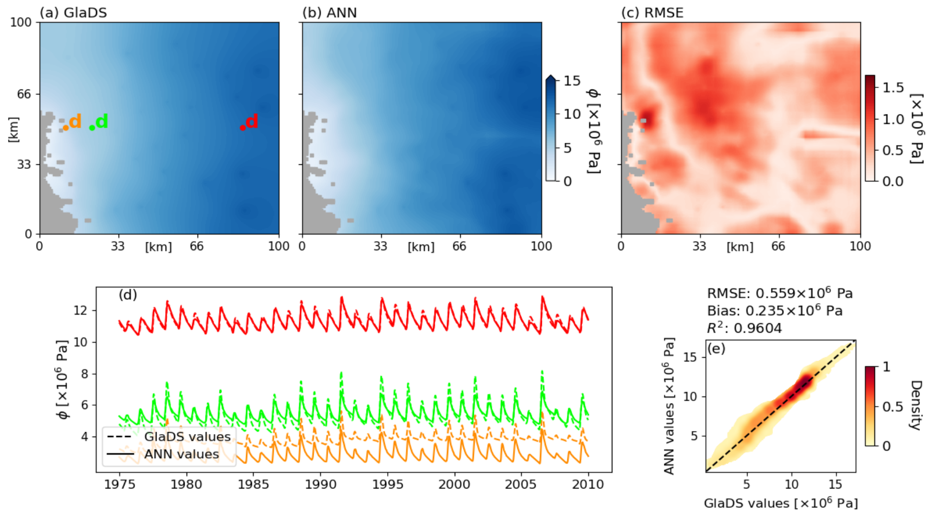

Figure 2. Spatial generalization performance of the ANN. Maps of mean 1975-2009 ϕ fields at the test glacier (Upernavik, Fig. S1 for location) simulated by (a) GlaDS, and (b) the ANN. The Root-Mean-Square Error (RMSE) of the ANN with respect to GlaDS is shown in (c). Time series (d) of ϕ at specific grid points simulated by GlaDS (dashed lines) and the ANN (solid lines), with color-coded locations shown in (a). Performance statistics (e) of the ANN with respect to GlaDS evaluated at all the grid points of the test glacier for all the 1975-2009 time steps.

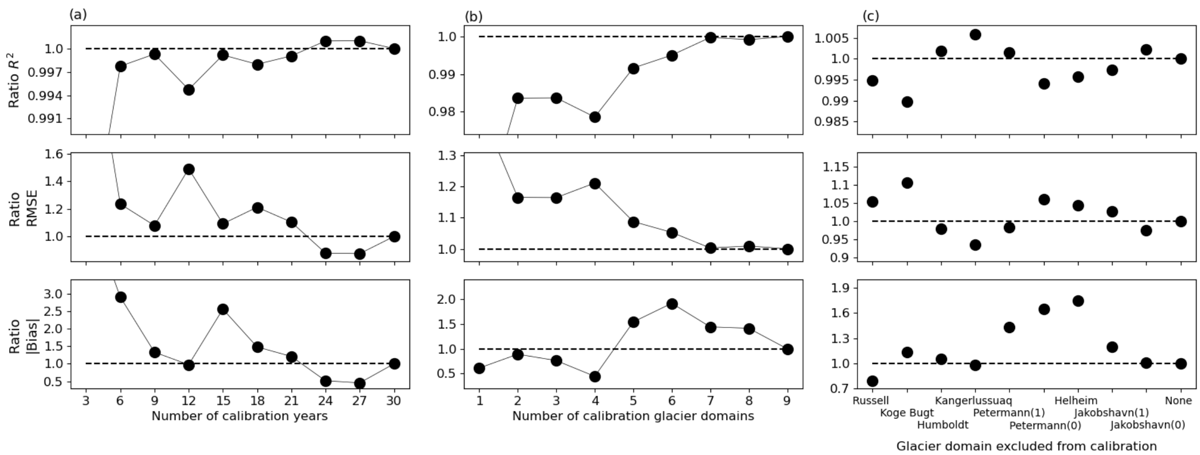

Figure 3. Sensitivity experiments (a) for number of years included in the calibration data, with evaluation on the 2005-2009 test years. Sensitivity experiments (b) for number of glacier domains included in the calibration data, with evaluation on the test Upernavik glacier domain. Leave-one-out experiments (c) in which each calibration glacier domain is separately excluded from the calibration data, with evaluation on the test Upernavik glacier domain. Dashed lines show no performance change with respect to the model calibrated with the default calibration data (30,9,'None' in a,b,c).

reduced down to 21 years. Small accuracy variations down to this limit are likely due to the inherent randomness in the training procedure of neural networks, through the parameter initialization method and the optimization algorithm. If calibration years are reduced to 18 years or less, performance metrics decrease and show more volatility. RMSE is increased by 21 % for 18 years, and by >250% for <6 years of calibration data.

3.4 Sensitivity to number of calibration glaciers

We also perform a sensitivity analysis to the number of glacier domains included in the calibration. Our initial calibration data set consists of seven glaciers, two of which have been split in two to match with the ANN domain input size, thus resulting in nine distinct glacier domains for calibration (see Supporting Information). Here, we sequentially drop one additional random glacier domain from the calibration data, and evaluate the ANN performance metrics always on the same test glacier (Upernavik).

Figure 3b shows the sensitivity of the performance metrics to the number of calibration glaciers. Accuracy decrease is minor when the calibration data are reduced to 8 and 7 domains, except for increases in the absolute bias. However, we observe a strong deterioration in accuracy for calibration data sets of 6 or fewer glacier domains. At 4 domains, the increase in RMSE reaches 21%. The ANN accuracy has levelled off for >6 glacier domains, showing that we have used sufficient calibration data.

Finally, we investigate if any single glacier domain is disproportionately important to the ANN accuracy at the test glacier. We repeat the ANN training with each one of the 9 calibration glacier domains left out of the calibration data, and then evaluate performance metrics on the test glacier (Fig. 3c). In agreement with the results from excluding only a single glacier domain shown above, changes in performance metrics are mostly small. For some of these leave-one-out experiments, performance metrics even improve slightly. The maximal increase in RMSE is 10.6 %, occurring when Koge Bugt glacier is left out (Fig. 3c). These results show that the ANN calibration is not excessively sensitive to any particular glacier. This verifies that the ANN does not predict at an out-of-sample glacier based only on the characteristics from the most similar glacier seen in training, but rather that it learns general relationships controlling ϕ patterns across different glaciological contexts.

3.5 Ice sheet model forcing

We now demonstrate that our ANN emulator can readily be used as forcing for an ice sheet model. We run the Ice-sheet and Sea-level System Model (Larour et al., 2012) at the test glacier of our data set: Upernavik (see Supporting Information for simulation details). We perform two 1975-2009 simulations: one forced with p_w from GlaDS output (GlaDS-forced run), and the other with p_w from the ANN (ANN-forced run). Except for the p_w forcing, the two simulations share identical initial conditions, climatic forcing, and other parameterizations, thus isolating differences in ice thickness and ice flow caused by discrepancies in the ANN emulation of GlaDS.

Figures 4a and 4b show the change in ice thickness (ΔH_{ice}) over the 35 years of simulations for the GlaDS-forced and ANN-forced runs, respectively. The patterns of ΔH_{ice} are very close between these two runs, which is confirmed at a grid point level (Fig. 4c, R^2=0.88). To quantify ice flow variability, we compute the temporal standard deviation in ice velocity ($\sigma(u_{ice})$) at each grid point. For this metric also, the GlaDS-forced and ANN-forced runs are in close agreement throughout the domain (Fig. 4d, 4e). However, $\sigma(u_{ice})$ is slightly underestimated by the ANN-forced run at the glacier terminus, especially at the northernmost branch where we observed the modest bias on ϕ of the ANN (Fig. 2c, 2d). At the two other branches, $\sigma(u_{ice})$ in our two simulations agree well. Throughout the domain, the ANN-forced run explains 78% of the variance in $\sigma(u_{ice})$ of the GlaDS-

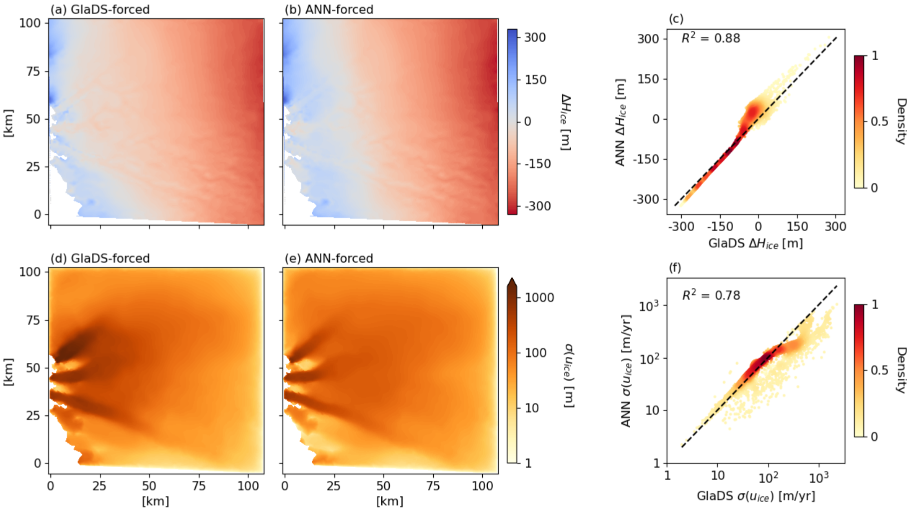

Figure 4. Results of 1975-2009 ice sheet model runs at the test glacier (Upernavik), with subglacial hydrology forcing from GlaDS (a,d) and from the ANN (b,e). Maps show ice thickness change (a,b, variable ΔH_{ice}), and standard deviation in ice velocities (d,e, variable $\sigma(u_{ice})$) over 1975-2009. Performance of the ANN-forced run with respect to the GlaDS-forced run in ΔH_{ice} (c) and $\sigma(u_{ice})$ (f). Note the logarithmic colorbar in (d,e) and axes in (f).

forced run (evaluated on logarithmic scale). The previous sections demonstrated the high accuracy of the ANN in reproducing ϕ spatio-temporal evolution as modeled by GlaDS. This section shows that this accuracy translates into dynamical ice sheet model results being only weakly sensitive to substituting our ANN for GlaDS to prescribe the p_w forcing.

In terms of computation, savings are large: simulating the 1975-2009 period in GlaDS over the Upernavik domain requires 859.9 CPU-hours, compared to 1.0 CPU-hour for predictions from our ANN on an identical core, i.e., close to $\mathcal{O}(10^3)$ faster. Finally, the 35-year GlaDS simulation required 268 times more CPU-hours than the ice sheet model simulation itself (3.2 CPU-hours), showing that subglacial hydrology models are a major computational bottleneck for large-scale ice sheet simulations.

4 Discussion

Our ANN produces realistic spatio-temporal patterns of subglacial hydraulic potential. It is skillful at temporal and spatial generalization on out-of-sample cases, when trained on as few as seven glacier domains and two decades of data. We do find small discrepancies between the ANN and GlaDS ϕ outputs, typically ranging between 0.2 and 1.5 MPa. Such values are smaller than discrepancies between subglacial hydrology models calculated in a recent intercomparison study (de Fleurian et al., 2018). Note that this comparison is not exact, because the intercomparisons used idealized configurations, whereas we use realistic Greenland glacier configurations. Still, because subglacial hydrology models are themselves an imperfect representation of real subglacial hydrology, the ANN output falling within typical inter-model spread reinforces our confidence that the ANN performs similarly to state-of-the-art numerical models.

Despite the demonstrated generalization capabilities of our ANN, we emphasize that deep learning models are prone to large errors, and possibly implausible behavior, when used to extrapolate beyond their range of training conditions (Rasp et al., 2018; Reichstein et al., 2019). Training data should encompass the range of meltwater runoff and glaciological conditions that will be targeted for predictions of the subglacial hydrology deep learning model. For future Greenland ice sheet projections, training should include high-runoff forcing, as surface melting is predicted to increase (Fettweis et al., 2013). We have verified the quality of our ANN training through sensitivity analyses, demonstrating that calibration data are sufficient, and that the ANN does not overfit but has learned general spatio-temporal relationships inherent to subglacial hydrology.

The ANN presented in this study, and machine learning techniques more generally, provide solutions to the extreme computational expense of running subglacial hydrology models in realistic ice sheet simulations. In addition to subglacial hydrology, machine learning techniques could also potentially replace other inaccurate parameterizations of ice sheet processes, where sufficient observations and/or high-fidelity model output exist to use as training data. For example, the physics of iceberg calving remain challenging to simulate, but capturing observed temporal dynamics of calving rates could be the target of machine learning parameterizations. As another example, such parameterizations can aim to represent ice sheet surface mass balance at fine scales without the need for expensive climate model downscaling, as has already been demonstrated for Alpine glaciers (Bolibar et al., 2020) and for the Antarctic Peninsula (van der Meer et al., 2023).

Observations of subglacial water pressure are scarce, especially when considering the large data requirements for deep learning. Thus, our emulator has been calibrated exclusively with output from high-fidelity models, which may themselves be biased. The value of observations could be exploited through pre-training on model output followed by fine-tuning on existing, spatio-temporally sparse observations (e.g., Rasp & Thuerey, 2020). In addition, there are other possible future avenues for improving this deep learning emulator. Associating the convolutional nature of our ANN with recurrent neural networks would allow to simulate temporal dependencies explicitly, in addition to spatial patterns. Temporal dependencies are here accounted for in an ad-hoc manner through our processing of inputs (see Supporting Information). Also, here the coupling of the ANN and the ice sheet model is one-way; the ANN is run first, and its output used as forcing to the ice sheet model. This approach allows the subglacial hydrology emulator to be used directly with any ice sheet model. Tight two-way coupling would capture feedback processes between subglacial hydrology and changes in ice sheet geometry and velocities, but requires implementation of the ANN within the source code of an ice sheet model. The lack of deep learning libraries in low-level languages, which are the basis of most modern ice sheet and climate model architectures, makes such implementation challenging (Partee et al., 2022). Recent development of new ice sheet models within high-level languages (e.g., Shapero et al., 2021) hold promise for better integration of machine learning directly into ice sheet models.

5 Conclusion

Our study demonstrates that deep learning techniques enable simulation of subglacial hydrology for ice sheet model forcing. Our emulator reproduces output of a state-of-the-art subglacial hydrology model with great fidelity, strong generalization skills, and $\mathcal{O}(10^3)$ savings in computation time. This advance has the potential to enable coupled simulations of ice sheet flow and evolving subglacial hydrology over entire ice sheets on centennial and longer time scales. Our work also demonstrates how machine learning techniques can be adopted in the ice sheet modeling community to resolve current issues related to knowledge gaps and computational bottlenecks. This general methodology is not limited to emulating subglacial hydrology models, but can potentially improve the representation of many other ice sheet model processes. Recent advances in computational capabilities and machine learning will, in parallel with traditional ice sheet model development, bring key improvements in predictions of ice sheet response to climate change.

6 Open Research

Model code is openly available at: https://doi.org/10.5281/zenodo.8006962

The code includes all scripts to run GlaDS, to process data, to train the ANN, and to predict with the ANN. The input files, hydrology model results, trained ANN parameter files, and final ANN predictions at the 8 glaciers of this study are included. Detailed data and code descriptions are provided.

Acknowledgments

VV was funded by the Heising-Simons Foundation (Grant 2020-1965). Computing resources were provided by the Partnership for an Advanced Computing Environment (PACE) at the Georgia Institute of Technology, Atlanta. The authors thank Mathieu Morlighem and Shivani Ehrenfeucht for advice about GlaDS.

References

Bartholomew, I. D., Nienow, P., Sole, A. J., Mair, D. W. F., Cowton, T. R., Palmer, S. J., & Wadham, J. L. (2011). Supraglacial forcing of subglacial drainage in the ablation zone of the greenland ice sheet. *Geophysical Research Letters*, *38*.

Bolibar, J., Rabatel, A., Gouttevin, I., Galiez, C., Condom, T., & Sauquet, E. (2020). Deep learning applied to glacier evolution modelling. *The Cryosphere*.

Brenowitz, N. D., & Bretherton, C. S. (2018). Prognostic validation of a neural network unified physics parameterization. *Geophysical Research Letters*, *45*, 6289 - 6298.

Brinkerhoff, D. J., Aschwanden, A., & Fahnestock, M. A. (2021). Constraining subglacial processes from surface velocity observations using surrogate-based bayesian inference. *Journal of Glaciology*, *67*, 385 - 403.

Budd, W. F., Jenssen, D., & Smith, I. N. (1984). A three-dimensional time-dependent model of the west antarctic ice sheet. *Annals of Glaciology*, *5*, 29 - 36.

Cowton, T. R., Nienow, P., Sole, A. J., Wadham, J. L., Lis, G., Bartholomew, I. D., ... Chandler, D. (2013). Evolution of drainage system morphology at a land-terminating greenlandic outlet glacier. *Journal of Geophysical Research: Earth Surface*, *118*, 29 - 41.

de Fleurian, B., Werder, M. A., Beyer, S., Brinkerhoff, D. J., Delaney, I., Dow, C. F., ... Sommers, A. N. (2018). Shmip the subglacial hydrology model intercomparison project. *Journal of Glaciology*, *64*, 897 - 916.

Ehrenfeucht, S., Morlighem, M., Rignot, E., Dow, C. F., & Mouginot, J. (2022). Seasonal acceleration of petermann glacier, greenland, from changes in subglacial hydrology. *Geophysical Research Letters*, *50*.

Fettweis, X., Box, J. E., Agosta, C., Amory, C., Kittel, C., Lang, C., ... Gallée, H. (2016). Reconstructions of the 1900–2015 greenland ice sheet surface mass balance using the regional climate mar model. *The Cryosphere*, *11*, 1015-1033.

Fettweis, X., Franco, B., Tedesco, M., van Angelen, J. H., Lenaerts, J. T. M., van den Broeke, M., & Gallée, H. (2013). Estimating the greenland ice sheet surface mass balance contribution to future sea level rise using the regional atmospheric climate model mar. *The Cryosphere*, *7*, 469-489.

Goelzer, H., Nowicki, S. M. J., Payne, A. J., Larour, E. Y., Seroussi, H., Lipscomb, W. H., ... van den Broeke, M. R. (2020). The future sea-level contribution of the greenland ice sheet: a multi-model ensemble study of ismip6. *The Cryosphere*.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *2015 IEEE International Conference on Computer Vision (ICCV)*, 1026-1034.

Hewitt, I. J. (2013). Seasonal changes in ice sheet motion due to melt water lubrication. *Earth and Planetary Science Letters*, *371*, 16-25.

Hoffman, M. J., Andrews, L. C., Price, S. A., Catania, G. A., Neumann, T. A., Lüthi, M. P., ... Morriss, B. F. (2016). Greenland subglacial drainage evolution regulated by weakly connected regions of the bed. *Nature Communications*, *7*.

Hornik, K., Stinchcombe, M. B., & White, H. L. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, *2*, 359-366.

Joughin, I. R., Smith, B. E., & Howat, I. M. (2017). A complete map of greenland ice velocity derived from satellite data collected over 20 years. *Journal of Glaciology*, *64*, 1 - 11.

Jouvet, G., Cordonnier, G., Kim, B., Lüthi, M. P., Vieli, A., & Aschwanden, A. (2021). Deep learning speeds up ice flow modelling by several orders of magnitude. *Journal of Glaciology*, *68*, 651 - 664.

Kazmierczak, E., Sun, S., Coulon, V., & Pattyn, F. (2022). Subglacial hydrology modulates basal sliding response of the antarctic ice sheet to climate forcing. *The Cryosphere*.

King, M. D., Howat, I. M., Candela, S. G., Noh, M.-J., Jeong, S., Noël, B. P. Y., ... Negrete, A. (2020). Dynamic ice loss from the greenland ice sheet driven by sustained glacier retreat. *Communications Earth & Environment*, *1*.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *CoRR*, *abs/1412.6980*.

Krebs-Kanzow, U., Gierz, P., Rodehacke, C. B., Xu, S., Yang, H., & Lohmann, G. (2020). The diurnal energy balance model (debm): a convenient surface mass balance solution for ice sheets in earth system modeling. *The Cryosphere*, *15*, 2295-2313.

Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2016). Simple and scalable predictive uncertainty estimation using deep ensembles. In *Nips.*

Larour, E. Y., Seroussi, H., Morlighem, M., & Rignot, E. (2012). Continental scale, high order, high spatial resolution, ice sheet modeling using the ice sheet system model (issm). *Journal of Geophysical Research*, *117*.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*, 436-444.

Lemley, J., Bazrafkan, S., & Corcoran, P. M. (2017). Smart augmentation learning an optimal data augmentation strategy. *IEEE Access*, *5*, 5858-5869.

Morlighem, M., Williams, C. N., Rignot, E., An, L., Arndt, J. E., Bamber, J. L., ... Zinglersen, K. B. (2017). Bedmachine v3: Complete bed topography and ocean bathymetry mapping of greenland from multibeam echo sounding combined with mass conservation. *Geophysical Research Letters*, *44*, 11051 - 11061.

Nienow, P., Sole, A. J., Slater, D. A., & Cowton, T. R. (2017). Recent advances in our understanding of the role of meltwater in the greenland ice sheet system. *Current Climate Change Reports*, *3*, 330-344.

Otosaka, I. N., Shepherd, A., Ivins, E. R., Schlegel, N., Amory, C., van den Broeke, M. R., ... Wouters, B. (2023). Mass balance of the greenland and antarctic ice sheets from 1992 to 2020. *Earth System Science Data*.

Partee, S., Ellis, M. S., Rigazzi, A., Shao, A. E., Bachman, S. D., Marques, G., & Robbins, B. (2022). Using machine learning at scale in numerical simulations with smartsim: An application to ocean climate modeling. *J. Comput. Sci.*, *62*, 101707.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. In *Neural information processing systems.*

Rasp, S., Pritchard, M. S., & Gentine, P. (2018). Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences of the United States of America*, *115*, 9684 - 9689.

Rasp, S., & Thuerey, N. (2020). Data-driven medium-range weather prediction with a resnet pretrained on climate simulations: A new model for weatherbench. *Journal of Advances in Modeling Earth Systems*, *13*.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven earth system science. *Nature*, *566*, 195-204.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *ArXiv*, *abs/1505.04597*.

Schoof, C. (2010). Ice-sheet acceleration driven by melt supply variability. *Nature*, *468*, 803-806.

Seroussi, H., Nowicki, S. M. J., Payne, A. J., Goelzer, H., Lipscomb, W. H., Abe-Ouchi, A., … Zwinger, T. (2020). Ismip6 antarctica: a multi-model ensemble of the antarctic ice sheet evolution over the 21st century. *The Cryosphere*.

Shapero, D., Badgeley, J. A., Hoffmann, A., & Joughin, I. R. (2021). icepack: a new glacier flow modeling package in python, version 1.0. *Geoscientific Model Development*.

Shepherd, A., Hubbard, A., Nienow, P., King, M. A., McMillan, M., & Joughin, I. R. (2009). Greenland ice sheet motion coupled with daily melting in late summer. *Geophysical Research Letters*, *36*.

Siu, K. (2022). Modelling subglacial hydrology under future climate scenarios in wilkes subglacial basin, antarctica..

Smith, L. C., Andrews, L. C., Pitcher, L. H., Overstreet, B. T., Rennermalm, Å. K., Cooper, M. G., … Simpson, C. E. (2021). Supraglacial river forcing of subglacial water storage and diurnal ice sheet motion. *Geophysical Research Letters*, *48*.

van der Meer, M., de Roda Husman, S., & Lhermitte, S. (2023). Deep learning regional climate model emulators: A comparison of two downscaling training frameworks. *Journal of Advances in Modeling Earth Systems*, *15*(6), e2022MS003593. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2022MS003593` (e2022MS003593 2022MS003593) doi: https://doi.org/10.1029/2022MS003593

van de Wal, R., Smeets, C. J. P. P., Boot, W., Stoffelen, M., van Kampen, R., Doyle, S. H., … Hubbard, A. (2015). Self-regulation of ice flow varies across the ablation area in south-west greenland. *The Cryosphere*, *9*, 603-611.

Werder, M. A., Hewitt, I. J., Schoof, C., & Flowers, G. E. (2013). Modeling channelized and distributed subglacial drainage in two dimensions. *Journal of Geophysical Research: Earth Surface*, *118*, 2140 - 2158.

Zwally, H. J., Abdalati, W., Herring, T. A., Larson, K. M., Saba, J. L., & Steffen, K. (2002). Surface melt-induced acceleration of greenland ice-sheet flow. *Science*, *297*, 218 - 222.

Accelerating subglacial hydrology for ice sheet models with deep learning methods

Vincent Verjans[1], Alexander Robel[1]

[1]School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, GA, USA

Key Points:

- We develop a deep learning emulator to simulate evolving subglacial hydrology in response to meltwater input for ice sheet simulations.
- The emulator shows generalization capabilities, large computational savings, and can be used to force numerical ice sheet models.
- We demonstrate that machine learning has substantial potential in improving ice sheet models, through using information-rich data sets.

Corresponding author: Vincent Verjans, vverjans3@gatech.edu

Abstract

Subglacial drainage networks regulate the response of ice sheet flow to surface meltwater input to the subglacial environment. Simulating subglacial hydrology evolution is critical to projecting ice sheet sensitivity to climate, and contribution to sea-level change. However, current numerical subglacial hydrology models are computationally expensive, and, consequently, evolving subglacial hydrology is neglected in large-scale ice sheet simulations. We present a deep learning emulator of a state-of-the-art subglacial hydrology model, trained at multiple Greenland glaciers. Our emulator performs strongly in both temporal (R^2 >0.99) and spatial (R^2 >0.96) generalization, offers high computational savings, and can be used to force numerical ice sheet models. This will enable century- and large-scale ice sheet model simulations, including interactions between ice flow and increased meltwater input to the subglacial environment. Generally, our work demonstrates that machine learning can further improve ice sheet models, reduce computational bottlenecks, and exploit information from high-fidelity models and novel observational platforms.

Plain Language Summary

Meltwater at the surface of ice sheets can drain to the subglacial environment, lubricate the bed, and influence ice sheet flow. Complex numerical subglacial hydrology models represent the subglacial drainage system, but are too computationally expensive to be included in large-scale and long-term ice sheet simulations. Consequently, model predictions of future ice sheet contribution to sea-level rise ignore ice flow modulation by evolving subglacial hydrology. Here, we use deep learning to emulate a state-of-the-art subglacial hydrology model. The emulator can directly force large-scale ice sheet models to capture ice flow sensitivity to subglacial hydrology. The computational speed and accuracy of our emulator show the potential to use machine learning to efficiently incorporate previously neglected processes into ice sheet models.

1 Introduction

The Greenland ice sheet has experienced accelerating mass loss since the early 1990s (Otosaka et al., 2023). Ice loss has been driven by increasing surface melt (Fettweis et al., 2016) and accelerating ice flow into the ocean (King et al., 2020). These two processes are linked by surface meltwater drainage into the subglacial environment. The evolution of the subglacial drainage system in response to meltwater input determines the subglacial water pressure, which regulates the speed of ice sliding over the bed (Nienow et al., 2017). Observations have demonstrated a strong sensitivity of ice flow speed to meltwater supply to the bed on timescales ranging from days to months (Zwally et al., 2002; Shepherd et al., 2009; Smith et al., 2021). The subglacial drainage system modulates this sensitivity, and there is no simple relationship between meltwater forcing and ice flow speed, due to the complexities of subglacial hydrology (Bartholomew et al., 2011; van de Wal et al., 2015).

Subglacial hydrology models simulate the evolution of subglacial water pressure under different conditions of ice sheet geometry and meltwater input (Werder et al., 2013; de Fleurian et al., 2018). They represent the transient evolution of the subglacial drainage system, are highly complex, and require many parameters and substantial computational expense (Werder et al., 2013; Hoffman et al., 2016). When applied to individual Greenland glaciers, studies have shown that accurately representing the evolution of the subglacial hydrology system with such complex models is necessary to explain observed variability in ice flow (Hewitt, 2013; Hoffman et al., 2016; Ehrenfeucht et al., 2022). However, their computational expense prohibits long, ice sheet scale simulations. In contrast, simple first-order formulations of the growth and decay of subglacial water flux do exist, facilitating large-scale simulations (Kazmierczak et al., 2022). But these formulations do not represent the different drainage components and assume a constant transmissivity. As such, they have critical limitations when considering high meltwater input scenarios, complex topographies, or reproducing realistic variability in subglacial hydrology and ice flow (de Fleurian et al., 2018).

Most subglacial hydrology models divide subglacial drainage into distributed and channelized systems (Schoof, 2010; Werder et al., 2013; de Fleurian et al., 2018). The former is typical of the early melt season, when the subglacial hydrology system is inefficient and water pressure increases strongly with meltwater supply. Later in the melt season, the system develops into a channelized system, efficiently evacuating meltwater, and causes water pressure to decrease with increasing meltwater input (Schoof, 2010; Bartholomew et al., 2011; Cowton et al., 2013). Ice flow variability is sensitive to the representation of, and transitions between different forms of drainage. Nevertheless, because of their computational expense, subglacial hydrology models are not included in model simulations at ice sheet scale (Goelzer et al., 2020; Seroussi et al., 2020). Consequently, current sea-level projections ignore a critical process regulating ice flow.

The climate modeling community has used machine learning techniques successfully to parameterize processes unresolved in coarse resolution global models (e.g., Brenowitz & Bretherton, 2018; Rasp et al., 2018). In particular, artificial neural networks (ANNs) are particularly powerful tools for parameterizing complex relationships between input and output variables, as they are capable of approximating any continuous function (the universal approximation theorem, Hornik et al., 1989). Furthermore, recent improvements in computational hardware, software, and optimization techniques have led to important ANN developments in multiple fields, including the Earth sciences (LeCun et al., 2015; Reichstein et al., 2019). Once trained, ANN models are computationally efficient, and ANNs have been used previously to emulate glacier flow models (Brinkerhoff et al., 2021; Jouvet et al., 2021). In this study, we use deep learning to enable representation of subglacial hydrology in large-scale ice sheet model simulations. Specifically, we develop an ANN emulator of the Glacier Drainage System model (GlaDS, Werder et al., 2013), an advanced and computationally expensive subglacial hydrology model. More generally, our work is a proof of concept for an important advancement in ice sheet modeling: we demonstrate that deep learning techniques can replace computationally-demanding or poorly constrained processes in large-scale ice sheet models.

2 Methods

We run GlaDS for 40 years at eight major Greenland glaciers (Petermann, Jakobshavn, Helheim, Kangerlussuaq, Humboldt, Koge Bugt, Russell, and Upernavik, see Fig. S1 for locations). The glacier geometries and ice flow velocities are taken from present-day observations (Joughin et al., 2017; Morlighem et al., 2017), and the meltwater runoff forcing from the 1970-2009 output of the diurnal Energy Balance Model (Krebs-Kanzow et al., 2020) (see Supporting Information). GlaDS simulates the evolution of the subglacial hydraulic potential, ϕ, in time and space by representing both channelized and distributed drainage. Accurate representation of ϕ in ice sheet models is critical, as it directly determines the subglacial water pressure, p_w. In turn, p_w determines the effective pressure at the ice-bed interface, N:

$$\begin{cases} N = p_{ice} - p_w \\ p_w = \phi - g\rho_w B \\ p_{ice} = g\rho_{ice} H_{ice} \end{cases}, \tag{1}$$

where p_{ice} is ice pressure [Pa], g is gravitational acceleration [m s^{-2}], H_{ice} is thickness of the above-lying ice column [m], B is bed elevation [m], and ρ_w and ρ_{ice} are water and ice density [kg m^{-3}], respectively. Critically, N [Pa] is a key variable in basal sliding laws for ice flow (Budd et al., 1984; Hewitt, 2013). We run GlaDS at a 2-hourly time step to preserve numerical stability.

Dynamics of ϕ are governed by the amount of surface meltwater draining through the ice sheet to the bed, followed by the routing of the water through the subglacial system. The goal of our ANN is to predict ϕ based on ice sheet state and meltwater runoff forcing, such that our p_w emulation accounts for spatio-temporal evolution of subglacial hydrology. Specifically, our ANN uses as inputs ice thickness, ice velocities, bedrock topography, and meltwater runoff fields, as well as moulin locations where meltwater reaches the subglacial domain. All these variables are common variables, parameters, or inputs to typical ice sheet models. Similar to current subglacial hydrology models, our ANN is aimed at one-way coupling with ice sheet models, i.e., it needs to be run prior to the ice sheet model and the ANN output is subsequently used as a forcing to the ice sheet model. We discuss prospects for full two-way coupling in the *Discussion* section.

Our ANN is a convolutional neural network, based on the U-Net architecture (Ronneberger et al., 2015). Our ANN thus uses two-dimensional input fields, and outputs a two-dimensional ϕ field at any given time step (see Supporting Information). In total, the ANN has 259,953 trainable parameters that are optimized such that ANN predictions of ϕ match training targets with accuracy. We calibrate our ANN to the GlaDS 1975-2004 output at seven of the eight glaciers. We keep the last 5 years (2005-2009) of output at these seven glaciers as test data to evaluate the temporal generalization capabilities of the ANN. Furthermore, we keep all the output of the eighth glacier for test data in order to evaluate the ANN spatial generalization capabilities. All the GlaDS output test data were totally unseen by the deep learning algorithm or the authors during the calibration. The calibration data are separated into training data, used to optimize the ANN parameters, and validation data, used to optimize hyperparameters and to avoid overfitting (see Supporting Information). All the results presented in the next sections have been computed on the test data.

3 Results

3.1 Temporal Generalization Performance

To assess the ability of the ANN to reproduce GlaDS hydraulic potential (ϕ) fields, we start by comparing their respective outputs at the 7 calibration glaciers over the last 5 years of simulations (2005-2009), which have not been used for the ANN training. Figures 1a and 1b show for one of the calibration glaciers (Helheim glacier) that the mean ϕ field of the ANN over these 5 test years reproduces the spatial patterns of GlaDS output well. In particular, the ANN captures the radial patterns of high ϕ values and variability centered at moulin locations, where meltwater runoff drains to the subglacial environment. The ANN performs well throughout the domain, as the root-mean-square error (RMSE) is mostly lower than 0.5 MPa (Fig. 1c). The temporal dynamics are also captured well, as ϕ time series show close correspondence between ANN and GlaDS outputs in most of the domain (Fig. 1d). The ANN captures the different ranges in seasonality and inter-annual variability of ϕ. It also reproduces the asymmetry between early and late melt season, which is due to changing drainage efficiency (Nienow et al., 2017). In Figure 1d, the orange curve shows a time series selected within the high RMSE area (Fig. 1c). This example illustrates that the ANN still captures the temporal variability, but is biased low at this particular location.

Figure 1e shows the ANN performance over the 5 years of test data at all the seven calibration glaciers, demonstrating low bias and RMSE. Furthermore, the ANN explains >99% of the variance in GlaDS ϕ output, as quantified by the coefficient of determination (R^2). These results demonstrate that the ANN is able to predict ϕ with good accuracy over years, and thus meltwater input conditions, unseen during training.

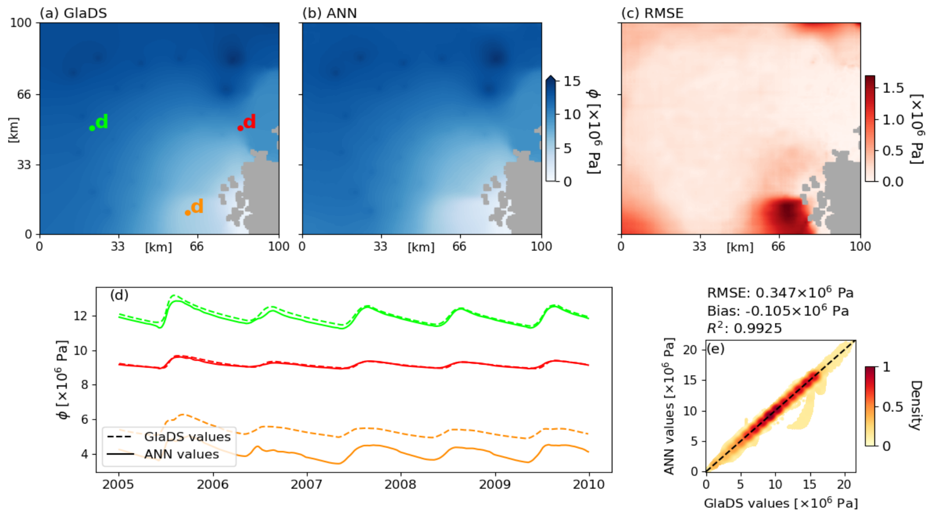

Figure 1. Temporal generalization performance of the ANN over the test period. Maps of mean 2005-2009 ϕ fields at one calibration glacier (Helheim, Fig. S1 for location) simulated by (a) GlaDS, and (b) the ANN. The Root-Mean-Square Error (RMSE) of the ANN with respect to GlaDS is shown in (c). Time series (d) of ϕ at specific grid points simulated by GlaDS (dashed lines) and the ANN (solid lines), with color-coded locations shown in (a). Performance statistics (e) of the ANN with respect to GlaDS evaluated at all the grid points of the 7 calibration glaciers for all the 2005-2009 time steps.

3.2 Spatial Generalization Performance

We now perform a similar evaluation, but at the test glacier (Upernavik), not included in the ANN calibration. This task presents a harder challenge, as both years of meltwater runoff and glaciological characteristics have not been used in calibration. The time-mean spatial patterns of ϕ at the test glacier are well-reproduced (Fig. 2a, 2b). The RMSE is low throughout the domain (Fig. 2c), except on a small portion near the glacier terminus. The time series shown in Figure 2d demonstrate that the ANN performance in reproducing temporal dynamics at seasonal and inter-annual scales are similar to its performance on calibration glaciers (compare with Fig. 1d). Here also, the orange lines in Fig. 2d show a time series corresponding to a low-performance location. Again, the ANN still captures temporal dynamics correctly, but has a consistent bias over the time series at this location. Finally, the metrics of performance for this test glacier show an RMSE lower than 0.56 MPa, a small positive bias of 0.24 MPa, and R^2 >0.96.

3.3 Sensitivity to number of calibration years

To calibrate the ANN, we have used 30 years (1975-2004) of GlaDS output at the calibration glaciers, and preserved 5 years (2005-2009) of output for testing (see Supporting Information). In this section, we investigate the sensitivity of ANN accuracy to the number of calibration years. Starting with all 30 years of available GlaDS output, we reduce the calibration data by 3-year increments, re-train the ANN at each increment, and evaluate performance metrics always on the 2005-2009 years.

Figure 3a shows the ratios in the R^2, RMSE and absolute bias metrics for each sensitivity experiment, with respect to the results computed when all the 30 calibration years are used in the ANN calibration. There is no decrease in accuracy for calibration data

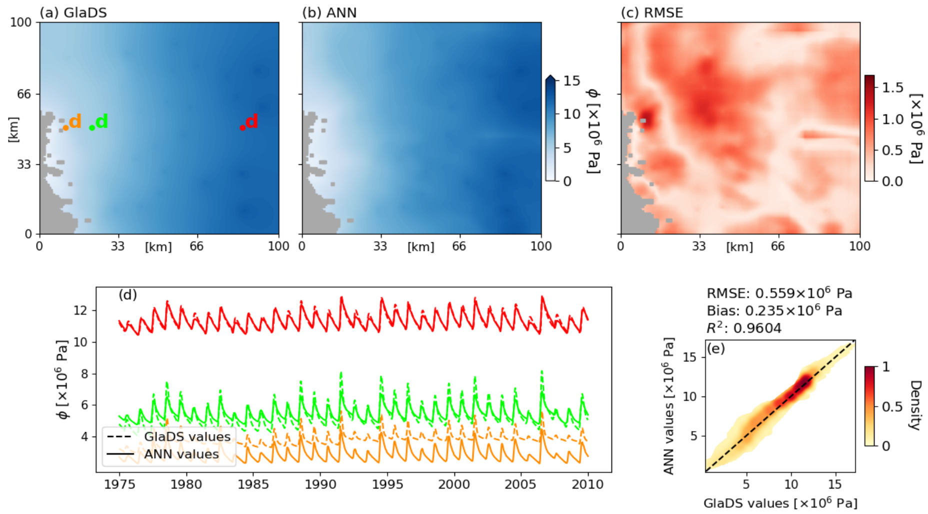

Figure 2. Spatial generalization performance of the ANN. Maps of mean 1975-2009 ϕ fields at the test glacier (Upernavik, Fig. S1 for location) simulated by (a) GlaDS, and (b) the ANN. The Root-Mean-Square Error (RMSE) of the ANN with respect to GlaDS is shown in (c). Time series (d) of ϕ at specific grid points simulated by GlaDS (dashed lines) and the ANN (solid lines), with color-coded locations shown in (a). Performance statistics (e) of the ANN with respect to GlaDS evaluated at all the grid points of the test glacier for all the 1975-2009 time steps.

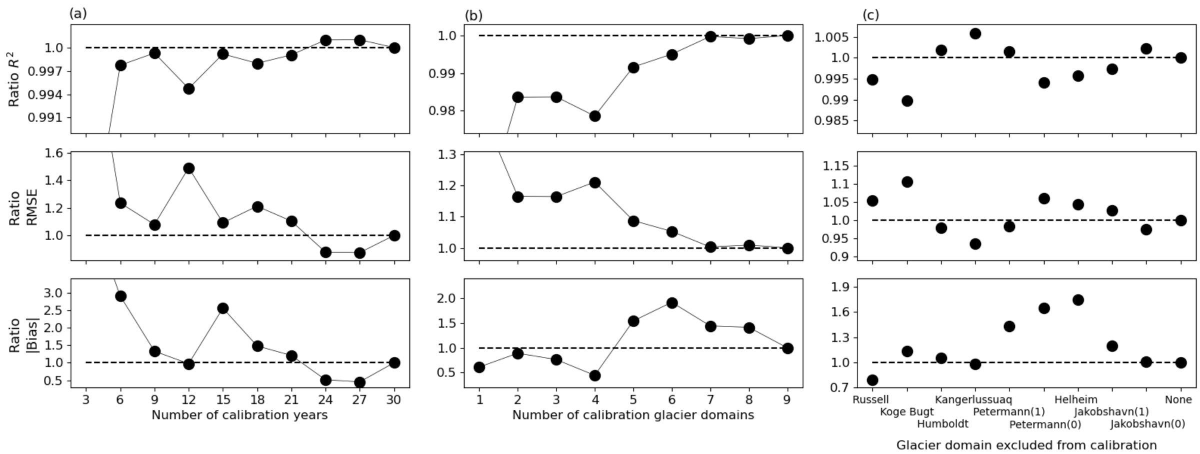

Figure 3. Sensitivity experiments (a) for number of years included in the calibration data, with evaluation on the 2005-2009 test years. Sensitivity experiments (b) for number of glacier domains included in the calibration data, with evaluation on the test Upernavik glacier domain. Leave-one-out experiments (c) in which each calibration glacier domain is separately excluded from the calibration data, with evaluation on the test Upernavik glacier domain. Dashed lines show no performance change with respect to the model calibrated with the default calibration data (30,9,'None' in a,b,c).

reduced down to 21 years. Small accuracy variations down to this limit are likely due to the inherent randomness in the training procedure of neural networks, through the parameter initialization method and the optimization algorithm. If calibration years are reduced to 18 years or less, performance metrics decrease and show more volatility. RMSE is increased by 21 % for 18 years, and by >250% for <6 years of calibration data.

3.4 Sensitivity to number of calibration glaciers

We also perform a sensitivity analysis to the number of glacier domains included in the calibration. Our initial calibration data set consists of seven glaciers, two of which have been split in two to match with the ANN domain input size, thus resulting in nine distinct glacier domains for calibration (see Supporting Information). Here, we sequentially drop one additional random glacier domain from the calibration data, and evaluate the ANN performance metrics always on the same test glacier (Upernavik).

Figure 3b shows the sensitivity of the performance metrics to the number of calibration glaciers. Accuracy decrease is minor when the calibration data are reduced to 8 and 7 domains, except for increases in the absolute bias. However, we observe a strong deterioration in accuracy for calibration data sets of 6 or fewer glacier domains. At 4 domains, the increase in RMSE reaches 21%. The ANN accuracy has levelled off for >6 glacier domains, showing that we have used sufficient calibration data.

Finally, we investigate if any single glacier domain is disproportionately important to the ANN accuracy at the test glacier. We repeat the ANN training with each one of the 9 calibration glacier domains left out of the calibration data, and then evaluate performance metrics on the test glacier (Fig. 3c). In agreement with the results from excluding only a single glacier domain shown above, changes in performance metrics are mostly small. For some of these leave-one-out experiments, performance metrics even improve slightly. The maximal increase in RMSE is 10.6 %, occurring when Koge Bugt glacier is left out (Fig. 3c). These results show that the ANN calibration is not excessively sensitive to any particular glacier. This verifies that the ANN does not predict at an out-of-sample glacier based only on the characteristics from the most similar glacier seen in training, but rather that it learns general relationships controlling ϕ patterns across different glaciological contexts.

3.5 Ice sheet model forcing

We now demonstrate that our ANN emulator can readily be used as forcing for an ice sheet model. We run the Ice-sheet and Sea-level System Model (Larour et al., 2012) at the test glacier of our data set: Upernavik (see Supporting Information for simulation details). We perform two 1975-2009 simulations: one forced with p_w from GlaDS output (GlaDS-forced run), and the other with p_w from the ANN (ANN-forced run). Except for the p_w forcing, the two simulations share identical initial conditions, climatic forcing, and other parameterizations, thus isolating differences in ice thickness and ice flow caused by discrepancies in the ANN emulation of GlaDS.

Figures 4a and 4b show the change in ice thickness (ΔH_{ice}) over the 35 years of simulations for the GlaDS-forced and ANN-forced runs, respectively. The patterns of ΔH_{ice} are very close between these two runs, which is confirmed at a grid point level (Fig. 4c, R^2=0.88). To quantify ice flow variability, we compute the temporal standard deviation in ice velocity ($\sigma(u_{ice})$) at each grid point. For this metric also, the GlaDS-forced and ANN-forced runs are in close agreement throughout the domain (Fig. 4d, 4e). However, $\sigma(u_{ice})$ is slightly underestimated by the ANN-forced run at the glacier terminus, especially at the northernmost branch where we observed the modest bias on ϕ of the ANN (Fig. 2c, 2d). At the two other branches, $\sigma(u_{ice})$ in our two simulations agree well. Throughout the domain, the ANN-forced run explains 78% of the variance in $\sigma(u_{ice})$ of the GlaDS-

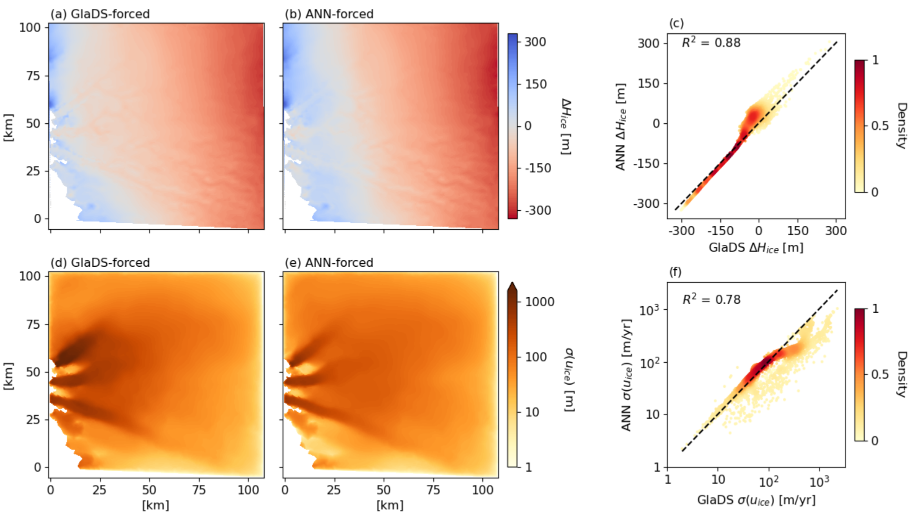

Figure 4. Results of 1975-2009 ice sheet model runs at the test glacier (Upernavik), with subglacial hydrology forcing from GlaDS (a,d) and from the ANN (b,e). Maps show ice thickness change (a,b, variable ΔH_{ice}), and standard deviation in ice velocities (d,e, variable $\sigma(u_{ice})$) over 1975-2009. Performance of the ANN-forced run with respect to the GlaDS-forced run in ΔH_{ice} (c) and $\sigma(u_{ice})$ (f). Note the logarithmic colorbar in (d,e) and axes in (f).

forced run (evaluated on logarithmic scale). The previous sections demonstrated the high accuracy of the ANN in reproducing ϕ spatio-temporal evolution as modeled by GlaDS. This section shows that this accuracy translates into dynamical ice sheet model results being only weakly sensitive to substituting our ANN for GlaDS to prescribe the p_w forcing.

In terms of computation, savings are large: simulating the 1975-2009 period in GlaDS over the Upernavik domain requires 859.9 CPU-hours, compared to 1.0 CPU-hour for predictions from our ANN on an identical core, i.e., close to $\mathcal{O}(10^3)$ faster. Finally, the 35-year GlaDS simulation required 268 times more CPU-hours than the ice sheet model simulation itself (3.2 CPU-hours), showing that subglacial hydrology models are a major computational bottleneck for large-scale ice sheet simulations.

4 Discussion

Our ANN produces realistic spatio-temporal patterns of subglacial hydraulic potential. It is skillful at temporal and spatial generalization on out-of-sample cases, when trained on as few as seven glacier domains and two decades of data. We do find small discrepancies between the ANN and GlaDS ϕ outputs, typically ranging between 0.2 and 1.5 MPa. Such values are smaller than discrepancies between subglacial hydrology models calculated in a recent intercomparison study (de Fleurian et al., 2018). Note that this comparison is not exact, because the intercomparisons used idealized configurations, whereas we use realistic Greenland glacier configurations. Still, because subglacial hydrology models are themselves an imperfect representation of real subglacial hydrology, the ANN output falling within typical inter-model spread reinforces our confidence that the ANN performs similarly to state-of-the-art numerical models.

Despite the demonstrated generalization capabilities of our ANN, we emphasize that deep learning models are prone to large errors, and possibly implausible behavior, when used to extrapolate beyond their range of training conditions (Rasp et al., 2018; Reichstein et al., 2019). Training data should encompass the range of meltwater runoff and glaciological conditions that will be targeted for predictions of the subglacial hydrology deep learning model. For future Greenland ice sheet projections, training should include high-runoff forcing, as surface melting is predicted to increase (Fettweis et al., 2013). We have verified the quality of our ANN training through sensitivity analyses, demonstrating that calibration data are sufficient, and that the ANN does not overfit but has learned general spatio-temporal relationships inherent to subglacial hydrology.

The ANN presented in this study, and machine learning techniques more generally, provide solutions to the extreme computational expense of running subglacial hydrology models in realistic ice sheet simulations. In addition to subglacial hydrology, machine learning techniques could also potentially replace other inaccurate parameterizations of ice sheet processes, where sufficient observations and/or high-fidelity model output exist to use as training data. For example, the physics of iceberg calving remain challenging to simulate, but capturing observed temporal dynamics of calving rates could be the target of machine learning parameterizations. As another example, such parameterizations can aim to represent ice sheet surface mass balance at fine scales without the need for expensive climate model downscaling, as has already been demonstrated for Alpine glaciers (Bolibar et al., 2020) and for the Antarctic Peninsula (van der Meer et al., 2023).

Observations of subglacial water pressure are scarce, especially when considering the large data requirements for deep learning. Thus, our emulator has been calibrated exclusively with output from high-fidelity models, which may themselves be biased. The value of observations could be exploited through pre-training on model output followed by fine-tuning on existing, spatio-temporally sparse observations (e.g., Rasp & Thuerey, 2020). In addition, there are other possible future avenues for improving this deep learning emulator. Associating the convolutional nature of our ANN with recurrent neural networks would allow to simulate temporal dependencies explicitly, in addition to spatial patterns. Temporal dependencies are here accounted for in an ad-hoc manner through our processing of inputs (see Supporting Information). Also, here the coupling of the ANN and the ice sheet model is one-way; the ANN is run first, and its output used as forcing to the ice sheet model. This approach allows the subglacial hydrology emulator to be used directly with any ice sheet model. Tight two-way coupling would capture feedback processes between subglacial hydrology and changes in ice sheet geometry and velocities, but requires implementation of the ANN within the source code of an ice sheet model. The lack of deep learning libraries in low-level languages, which are the basis of most modern ice sheet and climate model architectures, makes such implementation challenging (Partee et al., 2022). Recent development of new ice sheet models within high-level languages (e.g., Shapero et al., 2021) hold promise for better integration of machine learning directly into ice sheet models.

5 Conclusion

Our study demonstrates that deep learning techniques enable simulation of subglacial hydrology for ice sheet model forcing. Our emulator reproduces output of a state-of-the-art subglacial hydrology model with great fidelity, strong generalization skills, and $\mathcal{O}(10^3)$ savings in computation time. This advance has the potential to enable coupled simulations of ice sheet flow and evolving subglacial hydrology over entire ice sheets on centennial and longer time scales. Our work also demonstrates how machine learning techniques can be adopted in the ice sheet modeling community to resolve current issues related to knowledge gaps and computational bottlenecks. This general methodology is not limited to emulating subglacial hydrology models, but can potentially improve the representation of many other ice sheet model processes. Recent advances in computational capabilities and machine learning will, in parallel with traditional ice sheet model development, bring key improvements in predictions of ice sheet response to climate change.

6 Open Research

Model code is openly available at: https://doi.org/10.5281/zenodo.8006962

The code includes all scripts to run GlaDS, to process data, to train the ANN, and to predict with the ANN. The input files, hydrology model results, trained ANN parameter files, and final ANN predictions at the 8 glaciers of this study are included. Detailed data and code descriptions are provided.

Acknowledgments

VV was funded by the Heising-Simons Foundation (Grant 2020-1965). Computing resources were provided by the Partnership for an Advanced Computing Environment (PACE) at the Georgia Institute of Technology, Atlanta. The authors thank Mathieu Morlighem and Shivani Ehrenfeucht for advice about GlaDS.

References

Bartholomew, I. D., Nienow, P., Sole, A. J., Mair, D. W. F., Cowton, T. R., Palmer, S. J., & Wadham, J. L. (2011). Supraglacial forcing of subglacial drainage in the ablation zone of the greenland ice sheet. *Geophysical Research Letters*, *38*.

Bolibar, J., Rabatel, A., Gouttevin, I., Galiez, C., Condom, T., & Sauquet, E. (2020). Deep learning applied to glacier evolution modelling. *The Cryosphere*.

Brenowitz, N. D., & Bretherton, C. S. (2018). Prognostic validation of a neural network unified physics parameterization. *Geophysical Research Letters*, *45*, 6289 - 6298.

Brinkerhoff, D. J., Aschwanden, A., & Fahnestock, M. A. (2021). Constraining subglacial processes from surface velocity observations using surrogate-based bayesian inference. *Journal of Glaciology*, *67*, 385 - 403.

Budd, W. F., Jenssen, D., & Smith, I. N. (1984). A three-dimensional time-dependent model of the west antarctic ice sheet. *Annals of Glaciology*, *5*, 29 - 36.

Cowton, T. R., Nienow, P., Sole, A. J., Wadham, J. L., Lis, G., Bartholomew, I. D., ... Chandler, D. (2013). Evolution of drainage system morphology at a land-terminating greenlandic outlet glacier. *Journal of Geophysical Research: Earth Surface*, *118*, 29 - 41.

de Fleurian, B., Werder, M. A., Beyer, S., Brinkerhoff, D. J., Delaney, I., Dow, C. F., ... Sommers, A. N. (2018). Shmip the subglacial hydrology model intercomparison project. *Journal of Glaciology*, *64*, 897 - 916.

Ehrenfeucht, S., Morlighem, M., Rignot, E., Dow, C. F., & Mouginot, J. (2022). Seasonal acceleration of petermann glacier, greenland, from changes in subglacial hydrology. *Geophysical Research Letters*, *50*.

Fettweis, X., Box, J. E., Agosta, C., Amory, C., Kittel, C., Lang, C., ... Gallée, H. (2016). Reconstructions of the 1900–2015 greenland ice sheet surface mass balance using the regional climate mar model. *The Cryosphere*, *11*, 1015-1033.

Fettweis, X., Franco, B., Tedesco, M., van Angelen, J. H., Lenaerts, J. T. M., van den Broeke, M., & Gallée, H. (2013). Estimating the greenland ice sheet surface mass balance contribution to future sea level rise using the regional atmospheric climate model mar. *The Cryosphere*, *7*, 469-489.

Goelzer, H., Nowicki, S. M. J., Payne, A. J., Larour, E. Y., Seroussi, H., Lipscomb, W. H., ... van den Broeke, M. R. (2020). The future sea-level contribution of the greenland ice sheet: a multi-model ensemble study of ismip6. *The Cryosphere*.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *2015 IEEE International Conference on Computer Vision (ICCV)*, 1026-1034.

Hewitt, I. J. (2013). Seasonal changes in ice sheet motion due to melt water lubrication. *Earth and Planetary Science Letters*, *371*, 16-25.

Hoffman, M. J., Andrews, L. C., Price, S. A., Catania, G. A., Neumann, T. A., Lüthi, M. P., ... Morriss, B. F. (2016). Greenland subglacial drainage evolution regulated by weakly connected regions of the bed. *Nature Communications*, *7*.

Hornik, K., Stinchcombe, M. B., & White, H. L. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, *2*, 359-366.

Joughin, I. R., Smith, B. E., & Howat, I. M. (2017). A complete map of greenland ice velocity derived from satellite data collected over 20 years. *Journal of Glaciology*, *64*, 1 - 11.

Jouvet, G., Cordonnier, G., Kim, B., Lüthi, M. P., Vieli, A., & Aschwanden, A. (2021). Deep learning speeds up ice flow modelling by several orders of magnitude. *Journal of Glaciology*, *68*, 651 - 664.

Kazmierczak, E., Sun, S., Coulon, V., & Pattyn, F. (2022). Subglacial hydrology modulates basal sliding response of the antarctic ice sheet to climate forcing. *The Cryosphere*.

King, M. D., Howat, I. M., Candela, S. G., Noh, M.-J., Jeong, S., Noël, B. P. Y., ... Negrete, A. (2020). Dynamic ice loss from the greenland ice sheet driven by sustained glacier retreat. *Communications Earth & Environment*, *1*.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *CoRR*, *abs/1412.6980*.

Krebs-Kanzow, U., Gierz, P., Rodehacke, C. B., Xu, S., Yang, H., & Lohmann, G. (2020). The diurnal energy balance model (debm): a convenient surface mass balance solution for ice sheets in earth system modeling. *The Cryosphere*, *15*, 2295-2313.

Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2016). Simple and scalable predictive uncertainty estimation using deep ensembles. In *Nips*.

Larour, E. Y., Seroussi, H., Morlighem, M., & Rignot, E. (2012). Continental scale, high order, high spatial resolution, ice sheet modeling using the ice sheet system model (issm). *Journal of Geophysical Research*, *117*.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*, 436-444.

Lemley, J., Bazrafkan, S., & Corcoran, P. M. (2017). Smart augmentation learning an optimal data augmentation strategy. *IEEE Access*, *5*, 5858-5869.

Morlighem, M., Williams, C. N., Rignot, E., An, L., Arndt, J. E., Bamber, J. L., ... Zinglersen, K. B. (2017). Bedmachine v3: Complete bed topography and ocean bathymetry mapping of greenland from multibeam echo sounding combined with mass conservation. *Geophysical Research Letters*, *44*, 11051 - 11061.

Nienow, P., Sole, A. J., Slater, D. A., & Cowton, T. R. (2017). Recent advances in our understanding of the role of meltwater in the greenland ice sheet system. *Current Climate Change Reports*, *3*, 330-344.

Otosaka, I. N., Shepherd, A., Ivins, E. R., Schlegel, N., Amory, C., van den Broeke, M. R., ... Wouters, B. (2023). Mass balance of the greenland and antarctic ice sheets from 1992 to 2020. *Earth System Science Data*.

Partee, S., Ellis, M. S., Rigazzi, A., Shao, A. E., Bachman, S. D., Marques, G., & Robbins, B. (2022). Using machine learning at scale in numerical simulations with smartsim: An application to ocean climate modeling. *J. Comput. Sci.*, *62*, 101707.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. In *Neural information processing systems*.

Rasp, S., Pritchard, M. S., & Gentine, P. (2018). Deep learning to represent subgrid

processes in climate models. *Proceedings of the National Academy of Sciences of the United States of America*, *115*, 9684 - 9689.
Rasp, S., & Thuerey, N. (2020). Data-driven medium-range weather prediction with a resnet pretrained on climate simulations: A new model for weatherbench. *Journal of Advances in Modeling Earth Systems*, *13*.
Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven earth system science. *Nature*, *566*, 195-204.
Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *ArXiv*, *abs/1505.04597*.
Schoof, C. (2010). Ice-sheet acceleration driven by melt supply variability. *Nature*, *468*, 803-806.
Seroussi, H., Nowicki, S. M. J., Payne, A. J., Goelzer, H., Lipscomb, W. H., Abe-Ouchi, A., . . . Zwinger, T. (2020). Ismip6 antarctica: a multi-model ensemble of the antarctic ice sheet evolution over the 21st century. *The Cryosphere*.
Shapero, D., Badgeley, J. A., Hoffmann, A., & Joughin, I. R. (2021). icepack: a new glacier flow modeling package in python, version 1.0. *Geoscientific Model Development*.
Shepherd, A., Hubbard, A., Nienow, P., King, M. A., McMillan, M., & Joughin, I. R. (2009). Greenland ice sheet motion coupled with daily melting in late summer. *Geophysical Research Letters*, *36*.
Siu, K. (2022). Modelling subglacial hydrology under future climate scenarios in wilkes subglacial basin, antarctica..
Smith, L. C., Andrews, L. C., Pitcher, L. H., Overstreet, B. T., Rennermalm, Å. K., Cooper, M. G., . . . Simpson, C. E. (2021). Supraglacial river forcing of subglacial water storage and diurnal ice sheet motion. *Geophysical Research Letters*, *48*.
van der Meer, M., de Roda Husman, S., & Lhermitte, S. (2023). Deep learning regional climate model emulators: A comparison of two downscaling training frameworks. *Journal of Advances in Modeling Earth Systems*, *15*(6), e2022MS003593. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2022MS003593` (e2022MS003593 2022MS003593) doi: https://doi.org/10.1029/2022MS003593
van de Wal, R., Smeets, C. J. P. P., Boot, W., Stoffelen, M., van Kampen, R., Doyle, S. H., . . . Hubbard, A. (2015). Self-regulation of ice flow varies across the ablation area in south-west greenland. *The Cryosphere*, *9*, 603-611.
Werder, M. A., Hewitt, I. J., Schoof, C., & Flowers, G. E. (2013). Modeling channelized and distributed subglacial drainage in two dimensions. *Journal of Geophysical Research: Earth Surface*, *118*, 2140 - 2158.
Zwally, H. J., Abdalati, W., Herring, T. A., Larson, K. M., Saba, J. L., & Steffen, K. (2002). Surface melt-induced acceleration of greenland ice-sheet flow. *Science*, *297*, 218 - 222.

Supporting Information for "Accelerating subglacial hydrology for ice sheet models with deep learning methods"

Vincent Verjans[1], Alexander Robel[1]

[1]School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, GA, USA

Contents of this file

Introduction This Supporting Information provides all the methodological details of the study. Text S1 details the simulations performed with the Glacier Drainage System model. Text S2 details the architecture of the artificial neural network (ANN) developed in this study. Text S3 details our selection and processing of inputs for the ANN. Text S4 details how training, validation, and test data have been separated. Text S5 details the training procedure of the ANN. Text S6 details the configuration of the ice sheet model simulations, which are presented in section *Ice sheet model forcing* of the main text. Text S7 presents an additional sensitivity analysis to quantify the importance of each input in

---

Corresponding author: V. Verjans, School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, GA, USA (vverjans3@gatech.edu)

the performance of the ANN.

Table S1 provides the parameters used for the Glacier Drainage System model. Table S2 shows the architecture of the ANN. Table S3 shows the separation between training, validation, and test data. Figure S1 shows the configuration of the glaciers used for the subglacial hydrology model simulations. Figure S2 shows the results of the input importance sensitivity analysis.

All references are provided here, as well as in the reference list of the main text.

Text S1: Hydrology model simulations

We use the Glacier Drainage System model (GlaDS, Werder et al., 2013) implemented into the Ice-sheet and Sea-level System Model (ISSM, Larour et al., 2012) to generate data for this study. GlaDS is run separately over the seven calibration glaciers (Petermann, Jakobshavn, Helheim, Kangerlussuaq, Humboldt, Koge Bugt, and Russell, see Fig. S1) for 40 years with a two hour time step, saving outputs every three days. Outputs from these simulations are used to train, validate, and test our Artificial Neural Network (ANN) in mapping a set of inputs to a spatial field of hydraulic potential (ϕ). All domains have dimensions 100×100 km^2, except Petermann and Jakobshavn which have dimensions of 100×200 and 200×100 km^2, respectively (Fig. S1). We use a mesh resolution varying between 800 m in areas of fast ice flow and 2 500 m in areas of slow ice flow. We prescribe ice velocities from Joughin et al. (2017), and bedrock topography and ice thickness fields from Morlighem et al. (2017). We note that limited areas need bedrock smoothing to help with numerical stability of GlaDS for the Helheim, Petermann, and Jakobshavn domains (14%, 3%, and 4% of the domains, respectively). For our simulations, we integrate surface runoff over the glacier domains from the diurnal Energy Balance Model over the period 1970-2009 (Krebs-Kanzow et al., 2020). The surface runoff is directed to the bedrock at 30 locations representing moulins. The moulin locations are randomly distributed, under the conditions that ice thickness is greater than 500 m, ice velocity is greater than 25 m/yr, and that there is at least 10 km distance to the ocean, to the domain borders, and to any other moulin location (see Fig. S1). We refine the mesh resolution to 800 m around moulin locations to help with numerical stability. At any time step, runoff is equally partitioned between the 30 moulins.

We use an ice viscosity parameter corresponding to an ice temperature of 271.15 K in the parameterization of (Cuffey & Paterson, 2010). The spatial fields of the basal friction coefficient are obtained through an inversion method, based on the present-day geometry and ice velocities. GlaDS requires several parameters for the subglacial hydrology system. We take all the parameter values following the original implementation (Werder et al., 2013) and the default values of a recent intercomparison of subglacial hydrology models (de Fleurian et al., 2018). The parameter values of the subglacial hydrology system are listed in Table S1. Emulating GlaDS with other parameter values would require re-training the ANN. For the subglacial water system, we use zero-flux boundary conditions on the domain borders, and fixed hydrostatic ocean pressure at the grounding line. To preserve numerical stability, GlaDS is run with a 2 hour time step. Numerical instabilities still appeared in the simulations, in the form of infinite growth of the subglacial hydrological sheet thickness at some mesh elements on domain boundaries, close to the grounding line, or close to peripheral ice zones. This was caused by a negative cavity closing term, due to negative effective pressure values. At such mesh elements, we enforce a zero effective pressure, i.e, floatation.

In addition to these seven simulations, we perform a simulation at a test glacier (Upernavik, Fig. S1h) to generate additional test data. This simulation uses the same strategy for runoff generation and meshing as described above. The domain of Upernavik is of size 100×100 km^2, and all the GlaDS parameter values remain the same (Table S1).

We note here that GlaDS calculations can lead to unphysical negative water pressure values. Water pressure, p_w is defined as the hydraulic potential, ϕ, minus the elevation potential: $p_w = \phi - \phi_m$, where $\phi_m = \rho_w g B$ with ρ_w, g, B being water density, gravitational acceleration, and bedrock elevation, respectively. In the GlaDS routine for computing ϕ, there is no constraint on enforcing that $\phi \geq \phi_m$, and negative p_w values can arise in areas with high bed elevation and thin ice thickness (Siu, 2022). This only affects zones of peripheral ice.

Text S2: Architecture of the Artificial Neural Network

Our ANN is implemented with the Pytorch library (Paszke et al., 2019), and is a modified version of the U-Net architecture developed in Ronneberger et al. (2015). Our ANN architecture, detailed in Table S2, consists of an encoding and a decoding pathway. In the encoding stage, features are extracted from the two-dimensional input fields and spatial resolution is progressively reduced. Encoding is performed through a series of down-convolution blocks. Each down-convolution block consists of three operations: two convolution operations, each with a nonlinear activation function, and one pooling operation. The convolutions use a 3×3 kernel size, a stride of 1, zero-padding, and a bias term. We use the ReLU activation function after each convolution:

$$\mathrm{ReLU}(x) = \max(0, x). \tag{1}$$

The pooling operation is a 2×2 max-pooling, and thus reduces the horizontal resolution at each down-convolution block by a factor of 2 along each spatial dimension. The number of output features from the first down-convolution block is 24, and is then doubled for each subsequent down-convolution block. We experimented with different numbers of output features from the first down-convolution block, and found that 24 gives optimal model performance. The last down-convolution does not use pooling and has 94 output features instead of 96 to allow concatenation of two additional inputs at the end of the encoding stage. Specifically, we concatenate the time inputs to the 94 features. The time inputs are the cosine and sine of the time step (in units of years) multiplied by 2π. The decoding stage is symmetric to the encoding stage. It consists of up-convolution blocks. Each up-convolution block consists of four operations: one transpose convolution, one concatenation, and two convolutions. Every transpose convolution uses a 2×2 kernel size, halves the number of features, and enhances the horizontal resolution by a factor of 2 along each spatial dimension. The concatenation process allows to concatenate the features from the corresponding encoding level, allowing propagation of information from higher-resolution features. The convolution operations are similar to those from the encoding stage. We exclude the concatenation operation from the last up-convolution block, as we found that excluding it improves the spatial smoothness of the results from the ANN, in better agreement with the validation data. This is explained by not passing information at the high resolution of the initial data directly to the last up-convolution block, but rather forcing all features to undergo at least one pooling operation. The final layer of the ANN is a 1×1 convolution operation. No activation function is applied to this last convolution, of which the output is the standardized predicted ϕ field.

Text S3: Inputs to the Artificial Neural Network

Because our ANN is a convolutional neural network, its input consists of two-dimensional images, referred to as input features. We use as input features the bed topography, the ice thickness, and the ice velocity fields. These input features are fixed in time, and are therefore the same at any time step. However, they differ between the seven calibration glaciers, and are thus different for training samples corresponding to the different glaciers. In addition to these three input features, we use the spatial distribution of surface meltwater inflow. Surface meltwater inflow is non-zero only at moulin locations, over which the meltwater is distributed uniformly at any time step. We integrate the spatial meltwater inflow over different past time periods to be provided as input features to the ANN. To select these past time periods, we use a feature selection method. This consists of adding an increasing number of input features until no performance gain is achieved. In our feature selection method, our baseline case is considering only the instantaneous meltwater inflow, and meltwater inflow integrated over the previous 10 days. The first step is to add also meltwater inflow integrated over the previous month. The second step is to add meltwater inflow integrated over month-minus-1 to month-minus-2. We proceed iteratively, adding one month of meltwater inflow information at a time. We find that the optimal combination of features includes meltwater inflow (i) at the current time step, integrated over (ii) the previous 10 days, (iii) the previous month, (iv) month-minus-1 to month-minus-2, (v) month-minus-2 to month-minus-3, and (vi) month-minus-3 to month-minus-4. Including months beyond this time period does not improve model performance, when evaluated on the validation data. However, through our feature selection process, we find that adding (vii) the meltwater inflow integrated over the entire previous year further improves the ANN performance. This feature selection process results in a total of 12 inputs: 10 two-dimensional input features, and the cosine and sine of the time step. One input sample thus consists of the 12 inputs for a given glacier at a given time step. While GlaDS is run at a 2-hourly time step to ensure numerical stability, model outputs are saved every three days for storage reasons.

The two-dimensional input features have an inherent spatial scale, which our ANN is sensitive to. For this reason, we consistently train and evaluate our ANN over 100×100 km^2 windows. GlaDS runs on an irregular finite-element mesh, but results are bilinearly interpolated on a regular 128×128 mesh. It is important to preserve consistency in the spatial scaling, and using the ANN for predictions over a domain size different than the domain size used for training would be inappropriate. However, for larger domains of interest, it is straightforward to use the ANN multiple times over separate 100×100 km^2 parts of the domain, and concatenate the results. This is what has been done for the Jakobshavn and Petermann glaciers in this study, each being separated in two subdomains. As such, our data set of 8 glaciers corresponds to 10 domains. The ANN could also be trained, and thus used for predictions, on any other domain size.

As explained in Section *Architecture of the Artificial Neural Network*, we concatenate the time input at the end of the encoding stage as cosine and sine of the current time multiplied by 2π (Table S2). This improves the ANN performance due to the assymetry between early- and late-melt season behavior of the subglacial hydrology system. The time inputs are not passed in the first input layer because they are not spatial fields. Still, passing them at the end of the encoding stage allows the ANN to capture interactions between time of year and the other inputs through the decoding stage. Each time step is treated independently by the ANN, but our method of integrating past meltwater inflow provides, de facto, some temporal dependence. Future work can focus on associating the convolutional structure of our ANN with recurrent neural networks, which explicitly simulate temporal dependencies.

Text S4: Separation of training, validation, and test data

The first 5 years (1970-1974) of the GlaDS simulations are discarded, as GlaDS evolves transiently from an arbitrary initial state. For the seven calibration glaciers, years 1975-2004 are used as calibration years, and years 2005-2009 are preserved for test data. The data of the seven calibration glaciers consist of nine domains because two glaciers (Jakobshavn and Petermann) span the size of two domains, and are therefore provided separately to the ANN. The calibration years are further split between training and validation data, with 90% (years 1975-2001) and 10% (years 2002-2004) of the data, respectively. For the test glacier (Upernavik), all the data (years 1975-2009) are preserved for test data. Furthermore, we proceed to data augmentation by applying three transformations of each glacier domain and its input fields to use as additional training and validation data. Data augmentation improves performance of artificial neural networks by increasing the amount of data for calibration (Lemley et al., 2017). The transformations are a vertical axial symmetry, a horizontal axial symmetry, and a diagonal axial symmetry. For each training glacier, one of these transformations is used exclusively as validation data , and the two others are used as training (years 1975-2001) and validation (years 2002-2004) data. The splitting of the data between training, validation, and test data is detailed in Table S3.

Text S5: Training of the Artificial Neural Network

For training efficiency, we standardize every two-dimensional input feature and the ϕ output feature, such that our variables have zero mean and unit standard deviation. For predictions, our ANN thus requires inputs standardized accordingly, and predicted ϕ must be rescaled accordingly. Cosine and sine of time are not scaled, because they range between -1 and 1. We initialize the parameters of our ANN using the He Normal initialization method (He et al., 2015). We train our ANN with the training data such that parameter values are updated through backpropagation by minimizing a loss function measuring the misfit between ϕ fields predicted by the ANN and the GlaDS output. We use the L2 loss function, as it showed better results than when using alternative loss functions such as mean absolute error or the Huber loss. The L2 loss is defined as:

$$L(\boldsymbol{\phi}_{ANN}) = \sqrt{\frac{1}{N}\sum_i \left(\phi_{i,ANN} - \phi_{i,GlaDS}\right)^2}, \tag{2}$$

where $\phi_{i,GlaDS}$ denotes a ϕ value calculated by GlaDS, $\phi_{i,ANN}$ denotes the corresponding value of ϕ calculated by the ANN, and $\boldsymbol{\phi}_{ANN}$ denotes the full sample of ϕ values calculated by the ANN, with dimensions determined by the number of samples, and by the number of pixels in the two-dimensional spatial domain. In the loss calculation, we exclude all pixels with ice thickness less than 20 m or ice velocity less than 5 m/yr. Simulating ϕ in such regions is not necessary, as ice flow variability has minimal impact on ice sheet dynamics. And, because these outlier regions lead to different behaviors of subglacial hydrology models, we prefer to make the ANN calibration insensitive to these regions. During training, an epoch consists of passing the entire training data in sequences of randomly selected batches to the ANN. We use a batch size of 32 samples, as we found that it results in optimal model performance and training speed. The loss function is evaluated on the training batch, used to update parameter values via backpropagation, and on the validation data. Our backpropagation algorithm uses the Adam optimizer (Kingma & Ba, 2014) with an initial learning rate of 0.001. We use an adaptive learning rate, decreasing it by a factor of 2 after 5 consecutive epochs without improving the validation loss. This allows more localized search in the parameter space as the training procedure approaches a local minimum of the loss function. We stop the training after 10 consecutive epochs without improving the validation loss to avoid overfitting the training data. The final parameter values saved from the training procedure are those having led to the best validation loss score. As an additional tool to avoid overfitting, we use dropout. Similarly to (Ronneberger et al., 2015), we implement dropout only after the last convolution operation of the encoding stage (layer 8 in Table S2), and we use a dropout probability of 0.2. We train separately an ensemble of 20 networks. The training procedure is identical for these networks, and they only differ due to the random initialization of the parameters and the randomness of the optimization algorithm. Our final ANN is the ensemble mean output of these 20 members, as this averaging approach has been shown to improve deep neural network performance (Lakshminarayanan et al., 2016).

Text S6: Details on ice sheet model runs

The ice sheet model runs at Upernavik glacier, shown in section *Ice sheet model runs*, are performed using the Ice sheet and Sea-level System model (Larour et al., 2012). The ice rheology and basal friction coefficient parameters are kept identical as in the GlaDS simulations (see Supporting Information). As initial conditions, we use the ice geometry and ice velocity fields applied in the GlaDS runs. We prescribe a surface mass balance field that is constant in space and time, which is taken as the mean 1970-2009 surface mass balance averaged over the domain from the diurnal Energy Balance Model (Krebs-Kanzow et al., 2020). We start the simulations from 1975, to avoid impacts from the first 5 years of GlaDS run, during which GlaDS evolves from an arbitrary initial state. We perform a GlaDS-forced run, which applies the p_w field as predicted by GlaDS. Similarly, we perform an ANN-forced run, which applies the p_w field as predicted by the ANN. Ice flow dynamics are coupled to p_w through the basal sliding law, for which we use the Budd sliding law (Budd et al., 1984):

$$\boldsymbol{\tau_b} = -C^2 \boldsymbol{u_b} N, \tag{3}$$

where $\boldsymbol{\tau_b}$ is the basal stress [Pa], $\boldsymbol{u_b}$ is the basal ice velocity [m yr^{-1}], and C^2 is the basal friction coefficient, varying in space [m^{-1} yr]. We also perform a control run, in which the p_w assumes a simple hydrostatic connection to the ocean: $p_w = -g\rho_w B$. The control run captures the transient changes caused by the initial ice geometry not being in equilibrium. We subtract these transient changes from the GlaDS-forced and ANN-forced runs when analyzing the results in terms of ice thickness and ice velocities. As existing ice sheet sliding laws are not applicable at very low effective pressures, we follow (Ehrenfeucht et al., 2022) in applying a lower limit on N equal to 6% of the ice overburden pressure.

Text S7: Input importance

We evaluate the importance of each input feature to the quality of the ANN predictions. To this end, we perturb randomly each input feature individually. After each perturbation, we use the ANN to predict a $\widetilde{\boldsymbol{\phi}}_{ANN}$ field, which is of lower accuracy than the $\boldsymbol{\phi}_{ANN}$ field predicted without input perturbation. To perturb a given input feature, we add white Gaussian noise of standard deviation 1 to the input feature. We add the noise to the standardized features. In this way, each input feature is perturbed in a similar fashion, because all the standardized input features have mean 0 and standard deviation 1 by construction. The cosine and sine of time are not standardized. As such, we perturb these variables by their standard deviation, which is approximately 0.67. We evaluate $\widetilde{\boldsymbol{\phi}}_{ANN}$ on the validation data of the non-transformed glacier domains (see Table S3). Figure S2 shows the ratio in coefficient of determination (R^2), in Root Mean Square Error (RMSE), and in absolute bias of $\widetilde{\boldsymbol{\phi}}_{ANN}$ with respect to these metrics evaluated with $\boldsymbol{\phi}_{ANN}$. These ratios thus show the reduction in model accuracy caused by each perturbation. Results are shown for each perturbed input individually. Figure S2 shows that perturbing the ice thickness input has the most consequential impact on the ANN accuracy, as both the R^2 coefficient and the RMSE are more strongly impacted than when perturbing any other input. The different time periods over which we integrate the meltwater inflow show a similar importance on the ANN accuracy, except for the period over the entire previous year, which has a stronger impact on accuracy. This could be due to the strong correlation between each meltwater inflow input with its neighboring meltwater inflow input, for example month-minus-0 to month-minus-1 and month-minus-1 to month-minus-2. In contrast, the meltwater inflow integrated over the entire previous year has no neighboring time period with which it is strongly correlated. Finally, we find that perturbing any input causes a decrease in the ANN accuracy, which implies that all the inputs are to some extent useful for prediction.

Table S1. Parameters of the GlaDS simulations

Parameter	Value	Units
Englacial void ratio	10^{-3}	/
Pressure melt coefficient	7.5×10^{-8}	K Pa^{-1}
Latent heat of fusion	334×10^{3}	J kg^{-1}
Bedrock bump height	0.1	m
Cavity spacing	2.0	m
Sheet conductivity	0.01	$m^{7/4}$ $kg^{-1/2}$
Channel conductivity	0.1	$m^{3/2}$ $kg^{-1/2}$

Table S2. Architecture of the ANN

Layer number	Layer used as input	Layer type	Activation	Output shape
0	-	Input	-	128×128×10
1	0	Conv 3×3	ReLU	128×128×24
2	1	Conv 3×3	ReLU	128×128×24
3	2	MaxPool 2×2	-	64×64×24
4	3	Conv 3×3	ReLU	64×64×48
5	4	Conv 3×3	ReLU	64×64×48
6	5	MaxPool 2×2	-	32×32×48
7	6	Conv 3×3	ReLU	32×32×94
8	7	Conv 3×3	ReLU (dropout p=0.2)	32×32×94
9	8,cos(2πt),sin(2πt)	Concat	-	32×32×96
10	9	Trans-Conv 2×2	-	64×64×48
11	5,10	Concat	-	64×64×96
12	11	Conv 3×3	ReLU	64×64×48
13	12	Conv 3×3	ReLU	64×64×48
14	13	Trans-Conv 2×2	-	128×128×24
15	14	Conv 3×3	ReLU	128×128×24
16	15	Conv 3×3	ReLU	128×128×24
17	16	Conv 1×1	-	128×128×1

Table S3. Training, validation, and test data split

Glacier (subdomain)	Transformation	Years 0-5	Years 5-32	Years 32-35	Years 35-40
Jakobshavn (0)	None	Discarded	Train	Validation	Test
Jakobshavn (0)	Diagonal symmetry	Discarded	Validation	Validation	Discarded
Jakobshavn (0)	Vertical symmetry	Discarded	Train	Validation	Discarded
Jakobshavn (0)	Horizontal symmetry	Discarded	Train	Validation	Discarded
Jakobshavn (1)	None	Discarded	Train	Validation	Test
Jakobshavn (1)	Diagonal symmetry	Discarded	Train	Validation	Discarded
Jakobshavn (1)	Vertical symmetry	Discarded	Validation	Validation	Discarded
Jakobshavn (1)	Horizontal symmetry	Discarded	Train	Validation	Discarded
Helheim (0)	None	Discarded	Train	Validation	Test
Helheim (0)	Diagonal symmetry	Discarded	Train	Validation	Discarded
Helheim (0)	Vertical symmetry	Discarded	Train	Validation	Discarded
Helheim (0)	Horizontal symmetry	Discarded	Validation	Validation	Discarded
Petermann (0)	None	Discarded	Train	Validation	Test
Petermann (0)	Diagonal symmetry	Discarded	Validation	Validation	Discarded
Petermann (0)	Vertical symmetry	Discarded	Train	Validation	Discarded
Petermann (0)	Horizontal symmetry	Discarded	Train	Validation	Discarded
Petermann (1)	None	Discarded	Train	Validation	Test
Petermann (1)	Diagonal symmetry	Discarded	Train	Validation	Discarded
Petermann (1)	Vertical symmetry	Discarded	Validation	Validation	Discarded
Petermann (1)	Horizontal symmetry	Discarded	Train	Validation	Discarded
Kangerlussuaq (0)	None	Discarded	Train	Validation	Test
Kangerlussuaq (0)	Diagonal symmetry	Discarded	Train	Validation	Discarded
Kangerlussuaq (0)	Vertical symmetry	Discarded	Train	Validation	Discarded
Kangerlussuaq (0)	Horizontal symmetry	Discarded	Validation	Validation	Discarded
Humboldt (0)	None	Discarded	Train	Validation	Test
Humboldt (0)	Diagonal symmetry	Discarded	Validation	Validation	Discarded
Humboldt (0)	Vertical symmetry	Discarded	Train	Validation	Discarded
Humboldt (0)	Horizontal symmetry	Discarded	Train	Validation	Discarded
Koge Bugt (0)	None	Discarded	Train	Validation	Test
Koge Bugt (0)	Diagonal symmetry	Discarded	Train	Validation	Discarded
Koge Bugt (0)	Vertical symmetry	Discarded	Validation	Validation	Discarded
Koge Bugt (0)	Horizontal symmetry	Discarded	Train	Validation	Discarded
Russell (0)	None	Discarded	Train	Validation	Test
Russell (0)	Diagonal symmetry	Discarded	Train	Validation	Discarded
Russell (0)	Vertical symmetry	Discarded	Train	Validation	Discarded
Russell (0)	Horizontal symmetry	Discarded	Validation	Validation	Discarded
Upernavik (0)	None	Discarded	Test	Test	Test

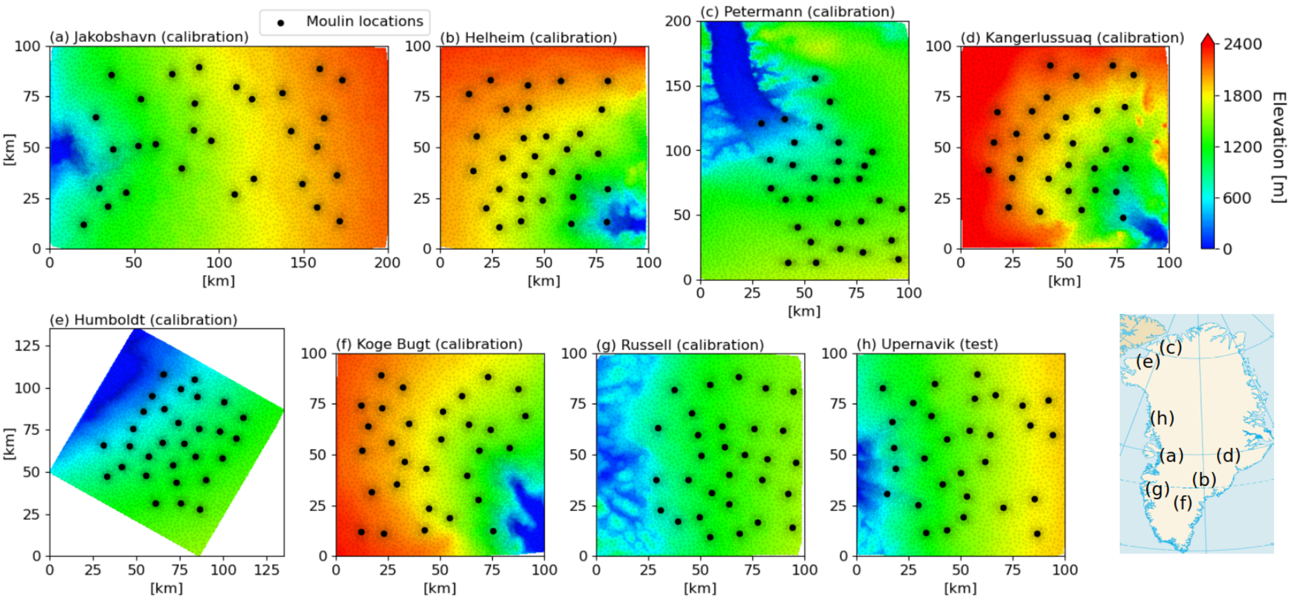

Figure S1. Model domains of the seven calibration glaciers ((a) Jakobshavn, (b) Helheim, (c) Petermann, (d) Kangerlussuaq, (e) Humboldt, (f) Koge Bugt, (g) Russell), and of the test glacier ((h) Upernavik). Map in the inset shows glacier locations. Light-grey points show mesh vertices.

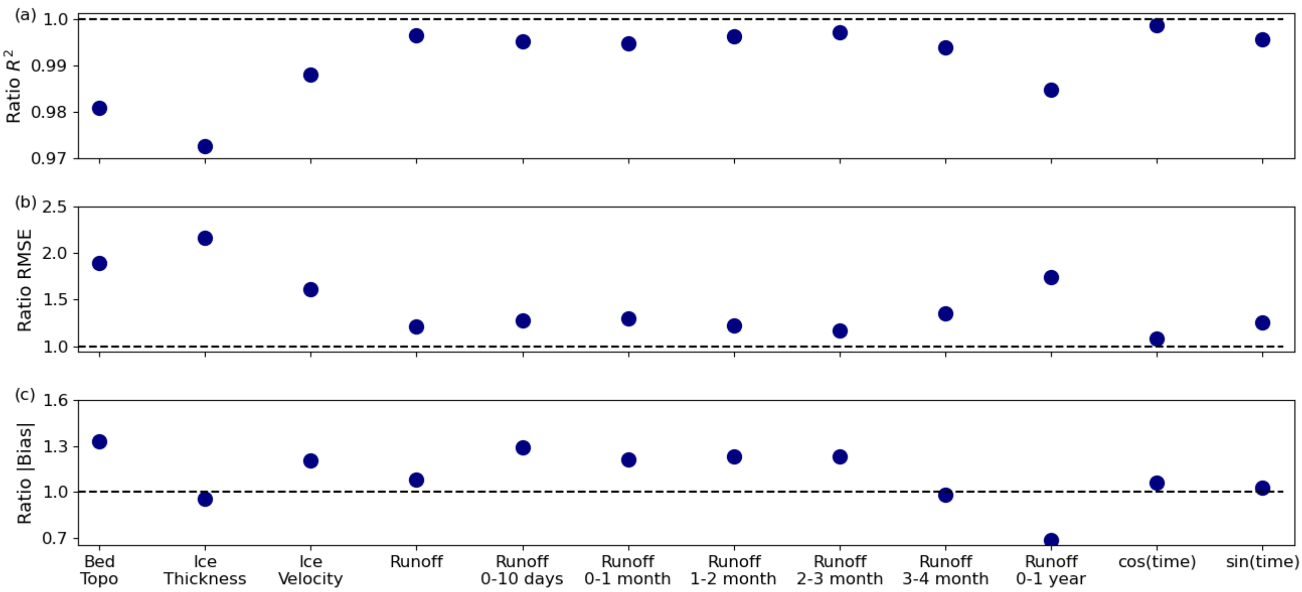

Figure S2. Ratio of performance metrics on the validation data after random white noise perturbation of input fields. The ratio is computed as the performance metric of the ANN with input perturbation with respect to the performance metric of the ANN without input perturbation. Metrics are (a) the coefficient of determination, (b) the Root Mean Square Error, and (c) the absolute bias. The black horizontal dashed line shows the value of 1, corresponding to no performance deterioration due to random input perturbation. Bed Topo is bed topography, Runoff denotes the meltwater inflow at the instantaneous time step, Runoff x-y period denotes the meltwater inflow integrated over the time interval between the past period x and period y.

References

Budd, W. F., Jenssen, D., & Smith, I. N. (1984). A three-dimensional time-dependent model of the west antarctic ice sheet. *Annals of Glaciology*, *5*, 29 - 36.

Cuffey, K. M., & Paterson, W. S. B. (2010). The physics of glaciers ed. 4..

de Fleurian, B., Werder, M. A., Beyer, S., Brinkerhoff, D. J., Delaney, I., Dow, C. F., ... Sommers, A. N. (2018). Shmip the subglacial hydrology model intercomparison project. *Journal of Glaciology*, *64*, 897 - 916.

Ehrenfeucht, S., Morlighem, M., Rignot, E., Dow, C. F., & Mouginot, J. (2022). Seasonal acceleration of petermann glacier, greenland, from changes in subglacial hydrology. *Geophysical Research Letters*, *50*.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *2015 IEEE International Conference on Computer Vision (ICCV)*, 1026-1034.

Joughin, I. R., Smith, B. E., & Howat, I. M. (2017). A complete map of greenland ice velocity derived from satellite data collected over 20 years. *Journal of Glaciology*, *64*, 1 - 11.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *CoRR*, *abs/1412.6980*.

Krebs-Kanzow, U., Gierz, P., Rodehacke, C. B., Xu, S., Yang, H., & Lohmann, G. (2020). The diurnal energy balance model (debm): a convenient surface mass balance solution for ice sheets in earth system modeling. *The Cryosphere*, *15*, 2295-2313.

Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2016). Simple and scalable predictive uncertainty estimation using deep ensembles. In *Nips.*

Larour, E. Y., Seroussi, H., Morlighem, M., & Rignot, E. (2012). Continental scale, high order, high spatial resolution, ice sheet modeling using the ice sheet system model (issm). *Journal of Geophysical Research*, *117*.

Lemley, J., Bazrafkan, S., & Corcoran, P. M. (2017). Smart augmentation learning an optimal data augmentation strategy. *IEEE Access*, *5*, 5858-5869.

Morlighem, M., Williams, C. N., Rignot, E., An, L., Arndt, J. E., Bamber, J. L., ... Zinglersen, K. B. (2017). Bedmachine v3: Complete bed topography and ocean bathymetry mapping of greenland from multibeam echo sounding combined with mass conservation. *Geophysical Research Letters*, *44*, 11051 - 11061.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. In *Neural information processing systems.*

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *ArXiv*, *abs/1505.04597*.

Siu, K. (2022). Modelling subglacial hydrology under future climate scenarios in wilkes subglacial basin, antarctica..

Werder, M. A., Hewitt, I. J., Schoof, C., & Flowers, G. E. (2013). Modeling channelized and distributed subglacial drainage in two dimensions. *Journal of Geophysical Research: Earth Surface*, *118*, 2140 - 2158.