https://doi.org/10.5194/egusphere-2024-1732
Preprint. Discussion started: 14 June 2024

Physics-aware Machine Learning for Glacier Ice Thickness Estimation: A Case Study for Svalbard

Viola Steidl[1], Jonathan L. Bamber[1,2], and Xiao Xiang Zhu[1,3]

[1]Chair of Data Science in Earth Observation, Department of Aerospace and Geodesy, Technical University of Munich, 80333 Munich, Germany
[2]Bristol Glaciology Centre, School of Geographical Sciences, University of Bristol, Bristol, BS8 1SS, UK
[3]Munich Center for Machine Learning, 80538 Munich, Germany

Correspondence: Viola Steidl (viola.steidl@tum.de)

Abstract. The ice thickness of the world's glaciers is mostly unmeasured, and physics-based models to reconstruct ice thickness can not always deliver accurate estimates. In this study, we use deep learning paired with physical knowledge to generate ice thickness estimates for all glaciers of Spitsbergen, Barentsøya, and Edgeøya in Svalbard. We incorporate mass conservation and other physically derived conditions into a neural network to predict plausible ice thicknesses even for glaciers without any in situ ice thickness measurements. With a glacier-wise cross-validation scheme, we evaluate the performance of the physics-informed neural network. The results of the experiments let us identify several challenges and opportunities that affect the model's performance in a real-world setting.

1 Introduction

Glacier ice thickness is a fundamental variable required for modelling the evolution of a glacier, as ice thickness and surface slope govern the ice flux at each point of the glacier (Cuffey and Paterson, 2010). However, direct measurements of ice thickness are scarce. In situ ice thickness measurements exist for only around 4 700 of 215 000 glaciers in the world (Millan et al., 2022).

Physics-based approaches aim to reconstruct glacier ice thicknesses from the limited in situ data. Farinotti et al. (2017) compared 17 models and found that their ice thickness estimates differ considerably on the test glaciers. Following these results, Farinotti et al. (2019) created an ensemble of five models to develop a *consensus* estimate of ice thicknesses for the world's glaciers in 2019. Later, Millan et al. (2022) derived ice thickness estimates for the world's glaciers using ice motion as the primary constraint. However, these results still differ from Farinotti et al. (2019) *consensus* estimate. It is evident, therefore, that significant uncertainty remains in ice thickness estimates.

Machine learning approaches are flexible and adapt well to complex structures and non-linear behaviour. They have already been employed to model glacier quantities like surface-mass-balance (Bolibar et al., 2020; Anilkumar et al., 2023), classify surge-type glaciers (Bouchayer et al., 2022), or model glacier flow (Jouvet, 2023; Min et al., 2019). One advantage of data-driven approaches is a significant speed-up compared to physics-based models (Jouvet et al., 2022). The disadvantages of

https://doi.org/10.5194/egusphere-2024-1732
Preprint. Discussion started: 14 June 2024

purely data-driven models are that they do not guarantee the physical correctness of the predicted quantities, and they often need huge amounts of training data to fully represent the system's complexity.

Recently, a new framework of data-driven but physically constrained models was described as physics-informed neural networks (PINNs) by Raissi et al. (2018). They exploit that neural networks can represent solutions to partial differential equations (PDE) if the squared residual of the governing PDE acts as the loss function of the neural network (Lagaris et al., 1998). Partial derivatives with respect to the model inputs are easy to calculate with the automatic differentiation algorithm that is used to train neural networks. PINNs do not require a discrete grid to be evaluated. Therefore, they are very data-efficient as the physics-based loss can be evaluated at any point within the training domain (Xu et al., 2023). Additional ground truth data can be used to compute a data loss that acts as a boundary condition to solve the PDE.

Teisberg et al. (2021) used a mass-conserving PINN to produce realistic ice thickness and depth-averaged ice flow maps for a single glacier in Antarctica. They showed that solving for mass conservation and additional constraints or regularizing terms is possible with a PINN.

This work extends the approach to predict ice thickness for all non-surging glaciers in Spitsbergen, Barentsøya, and Edgeøya in Svalbard. These regions include glaciers with various morphologies, from valley glaciers to ice caps. To better account for the glaciers' multiple geometries, sizes, and flow velocities, we include additional input features, e.g., slope and elevation. The challenge is to find a configuration of inputs and physical constraints that is general enough to describe the variety of glaciers in the study region. At the same time, the constraints and inputs should be strict enough to force the model to produce physically correct outputs.

Ice thickness measurements exist only for a fraction of the glaciers in the dataset, and there is no benchmark dataset to measure the model's performance. Therefore, we need a validation method that assesses the performance, although there is no ground truth. To this end, we estimate the expected drop in performance for glaciers without ice thickness measurements performing glacier-wise cross-validation. Also, we compare our ice thickness estimates to those of Millan et al. (2022), the *consensus* estimate of Farinotti et al. (2019) for our study region. These estimates are no benchmark datasets but are widely accepted in the community. Additionally, we compare to the recently published ice thickness estimate of van Pelt and Frank (2024) tailored to the region of Svalbard.

2 Physics-aware machine learning

2.1 Mass conservation

Assuming ice to be incompressible and integrating vertically along the depth of a glacier, we retrieve the two-dimensional form of the mass conservation

$$\frac{\partial H}{\partial t} + \nabla \cdot (\bar{\boldsymbol{v}} H) = \dot{b} \tag{1}$$

with H being the ice thickness and $\dot{b}$ denotes the mass balance of the glacier. $\bar{\boldsymbol{v}} = (\bar{v}_x, \bar{v}_y)$ is the velocity caused by the deformation of ice, averaged along the vertical axis. We will refer to $\bar{\boldsymbol{v}}$ as the depth-averaged velocity in the following. Equation

https://doi.org/10.5194/egusphere-2024-1732
Preprint. Discussion started: 14 June 2024

1 can be reformulated as

$$\nabla \cdot (\bar{\boldsymbol{v}} H) - \dot{a} = 0 \tag{2}$$

with $\dot{a} = \dot{b} - \partial_t H$ known as the apparent mass balance. In other words, the flux divergence on a glacier equals its apparent mass balance.

2.2 Depth-averaged velocity and basal sliding

The surface flow velocity $\boldsymbol{v}_\mathrm{s}$ of a glacier can be measured from space but is composed of the velocity from ice deformation and the ice sliding along the bedrock: $\boldsymbol{v}_\mathrm{s} = \boldsymbol{v}_\mathrm{d} + \boldsymbol{v}_\mathrm{b}$. We are only interested in the contributions to the glacier velocity caused by ice deformation because this is the only component directly related to ice thickness via Glen's flow law (Glen, 1955). Therefore, we estimate the contribution of basal sliding by introducing a factor β that is derived from the ratio of slope and observed surface velocity assuming $\boldsymbol{v}_\mathrm{d} = \boldsymbol{v}_\mathrm{s}(1 - \boldsymbol{\beta})$.

Now, the depth-averaged velocity is not equal to the surface velocity because, at its base, the glacier experiences drag from its bed. The basal drag balances the horizontal shear stress. We estimate the depth-averaged velocity to be within the bounds of

$$l_\mathrm{lower} \cdot v_\mathrm{s}(1 - \beta) < \bar{v} < v_\mathrm{s}(1 - \beta) \tag{3}$$

where l_lower should lie within $]0, 1]$. We fix $l_\mathrm{lower} = 0.7$ (see Appendix B for derivation). Depth-averaged velocities are estimated for the x- and y-direction and the velocity magnitude. Therefore, we calculate three separate β-values.

2.3 PINN Model

Figure 1 shows a schematic of the model with its input features, outputs, and loss components. The model is a neural network with eight fully connected layers and a preceding Fourier layer. Tancik et al. (2020) described the theory of Fourier layers. They map the input coordinates to a higher dimensional space, which enables the network to learn high-frequency functions in low-dimensional problem domains. In addition to the spatial coordinate of each training point, we also feed the neural network with auxiliary data like slope, elevation, distance to the border of the glacier, and the area of the glacier the point belongs to. Also, the network receives the three β values that were computed to estimate the contribution of basal sliding.

A non-linear activation function follows each layer of the neural network except the output layer:

$$\text{Softplus: } f(x) = \log(1 + \exp(x)) \tag{4}$$

Softplus is infinitely differentiable. This is important to avert vanishing gradients during optimization (Leng and Thiyagalingam, 2023).

The model outputs three quantities at each point of query: the ice thickness and depth-averaged deformation velocity in x- and y-direction. The predicted quantities must fulfil the mass conservation described in Equation 2. The squared deviation from

https://doi.org/10.5194/egusphere-2024-1732
Preprint. Discussion started: 14 June 2024

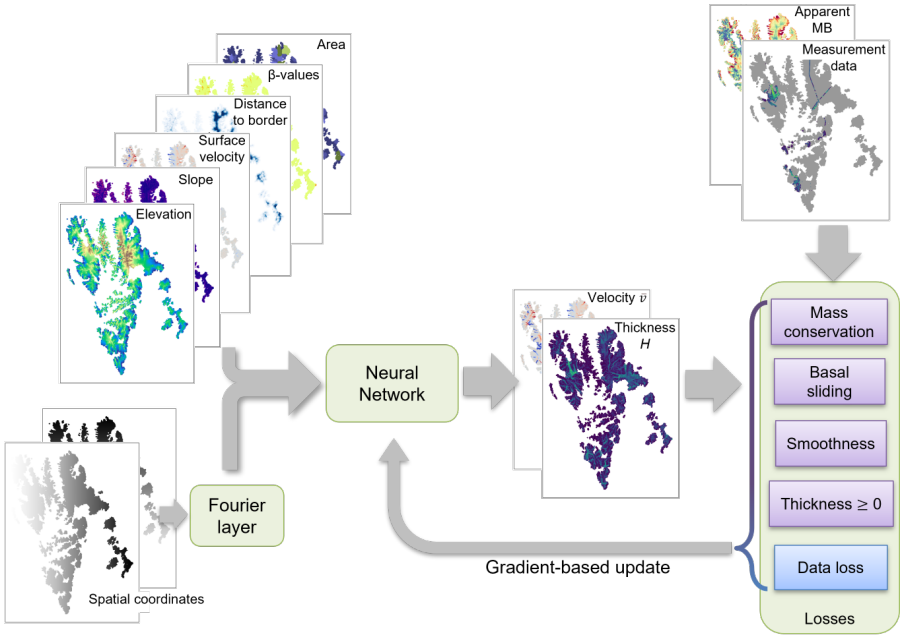

Figure 1. Physics-informed model with inputs, outputs, and components adding to the loss function.

this equation is the first component of the loss function:

$$\mathcal{L}_{mc} = (\nabla \cdot (\bar{\boldsymbol{v}} H) - \dot{a})^2 \tag{5}$$

The second component of the loss function is the amount by which the depth-averaged velocity estimates in the horizontal plane exceed the boundaries given in Eq. 3:

$$\mathcal{L}_{vel} = \begin{cases} (v_\mathrm{s}(1-\beta) - \bar{v})^2 & \text{if } \bar{v} < v_\mathrm{s}(1-\beta) \\ (l_\mathrm{lower} \cdot v_\mathrm{s}(1-\beta) - \bar{v})^2 & \text{if } \bar{v} > l_\mathrm{lower} \cdot v_\mathrm{s}(1-\beta) \\ 0 & \text{else} \end{cases} \qquad \text{with } \bar{v} \in \{\bar{v}_x, \bar{v}_y, \bar{v}_{magnitude}\} \tag{6}$$

$\mathcal{L}_{vel}$ is calculated separately for the x- and y-component and the magnitude of the depth-averaged velocity.

We include two more physics-aware constraints to improve the model performance: First, the ice thickness is assumed to be smooth, so the model will penalize large spatial derivatives within the ice thickness prediction:

$$\mathcal{L}_{smooth} = (\nabla H)^2 \tag{7}$$

https://doi.org/10.5194/egusphere-2024-1732
Preprint. Discussion started: 14 June 2024

Secondly, ice thickness cannot be smaller than 0. Therefore, we add a loss component that punishes negative ice thicknesses to the loss function.

$$\mathcal{L}_{H>0} = \begin{cases} H^2 & \text{if } H < 0 \\ 0 & \text{else} \end{cases} \tag{8}$$

The final loss component is the data loss. It penalizes the deviation from the in situ ice thickness measurements and acts as the boundary condition to solve the mass-conserving PDE. Each loss component will have a different scale, so we balance them with individual weights λ_i. Summing up all the loss components, we get the complete loss function as:

$$\mathcal{L} = \lambda_{mc}\mathcal{L}_{mc} + \lambda_{vel}\mathcal{L}_{vel} + \lambda_{smooth}\mathcal{L}_{smooth} + \lambda_{H>0}\mathcal{L}_{H>0} + \lambda_{data}\mathcal{L}_{data} \tag{9}$$

All the physics-aware losses are evaluated at any point in the study region. In contrast, the data loss is only evaluated wherever ice thickness measurements are available.

The training data is scaled to have a mean of 0 and a variance of 1. Before computing the physics-aware loss components, we scale the quantities back to their original units for physical consistency.

2.4 Validation

We evaluate the performance of the PINN model by calculating the root mean squared difference (RMSD) and the mean absolute percentage difference (MAPD) between the model prediction and the in situ ice thickness measurements. However, in situ measurements within a glacier are highly correlated due to their proximity. Therefore, a simple random split of the data into training and test datasets will not yield a realistic view of the model performance.

We employ a glacier-wise cross-validation (CV) approach as done by Bolibar et al. (2020) to better judge the model performance. This also allows us to make assumptions on how well the model will perform on a glacier without any measured ice thicknesses.

For the Leave-One-Glacier-Out (LOGO) CV, we chose seven glaciers that serve as test glaciers. In an alternating way, we train the model without one of those glaciers' ice thickness measurements.

It is important to note that only data labelled with ice thickness measurements of the test glacier is left out of the dataset. All the data needed to enforce the physical consistency for the test glacier stays in the training dataset. The mass-conserving PDE of Eq. 1 will still be solved at the test glacier but without enforcing boundary conditions with ice thickness measurements.

Upon validation, the RMSD and the MAPD are calculated for the test glacier. All the test glaciers are thoroughly mapped with ice thickness measurements and differ in size, mean measured ice thickness, and location in Svalbard.

3 Data

In this study, we focus on the glaciers on the islands of Spitsbergen, Barentsøya, and Edgeøya. Glaciers in an active surge phase during the data acquisition period for the surface velocity are not considered. The information on active surge phases is collected from Koch et al. (2023).

https://doi.org/10.5194/egusphere-2024-1732
Preprint. Discussion started: 14 June 2024

3.1 Data management

We processed all the data needed for the training of the PINN using the open global glacier model (OGGM) framework developed by Maussion et al. (2019). OGGM is an open-source framework to simulate glacier evolution. It provides models for mass balance, distributed ice thickness, and ice flow, as well as downloading tools for glacier outlines, digital elevation models (DEM), and climate data. The mass balance model is a temperature index melt model relying on climate data.

OGGM saves all the information for each glacier separately in Glacier Directories. The Randolph Glacier Inventory (RGI), Version 6.0 contains the outlines for the glaciers (RGI Consortium, 2017).

Using the outlines, OGGM defines a spatial grid for each glacier. The grid resolution is adapted individually according to the size of the glacier. In our study region, the grid resolution ranges from 12 m to 200 m. OGGM provides the data for this study, which is projected onto the glacier grids.

3.2 Auxiliary data

The elevation of each point comes from the global DEM from Copernicus DEM GLO-90, which was acquired from 2010 to 2015 (Copernicus). The slope is then computed by OGGM based on the glacier's smoothed topography and over the length of a grid cell on the glacier. The distance to the border of the glacier outline is computed for each point within a glacier. The glacier area is also retrieved from the RGI.

3.3 Surface Velocity Data

Millan et al. (2022) derived the surface flow velocity of the world's glaciers using image pairs acquired between 2017 and 2018 by Landsat 8, Sentinel-2, and Sentinel-1. They tracked glacier motion using a cross-correlation approach. The resolution of the velocity product is 50 m with an estimated accuracy of about 10 $\mathrm{m\,yr^{-1}}$. Using OGGM, velocity in x- and y-direction and velocity magnitude are projected onto the individual glacier grids and then smoothed with a two-dimensional Gaussian filter.

We introduce the aforementioned β value (see Sect. 2.2) to estimate the influence of basal sliding on the measured surface velocity $\boldsymbol{v}_\mathrm{d} = \boldsymbol{v}_\mathrm{s}(1 - \boldsymbol{\beta})$. Following Millan et al. (2022), we set β equal to 0.1 in areas where the ratio between slope and observed surface velocity is greater than $0.001\ \mathrm{yr^{-1}\,m}$ and modulate up to 0.9 for areas where the ratio is less than $0.001\ \mathrm{yr^{-1}\,m}$.

3.4 Apparent mass balance

The apparent mass balance is the difference between the point-wise mass balance and the thickness change rate dh/dt at each grid point. The mass balance at each point of a glacier grid is estimated using the *ConstantMassBalance* model from OGGM. It calculates the average mass balance during a chosen period from given climate data, calibrated with geodetic mass balance data from Hugonnet et al. (2021). To match the acquisition period of the surface velocity, we set the climate period for the mass balance model to 2016-2018.

https://doi.org/10.5194/egusphere-2024-1732
Preprint. Discussion started: 14 June 2024

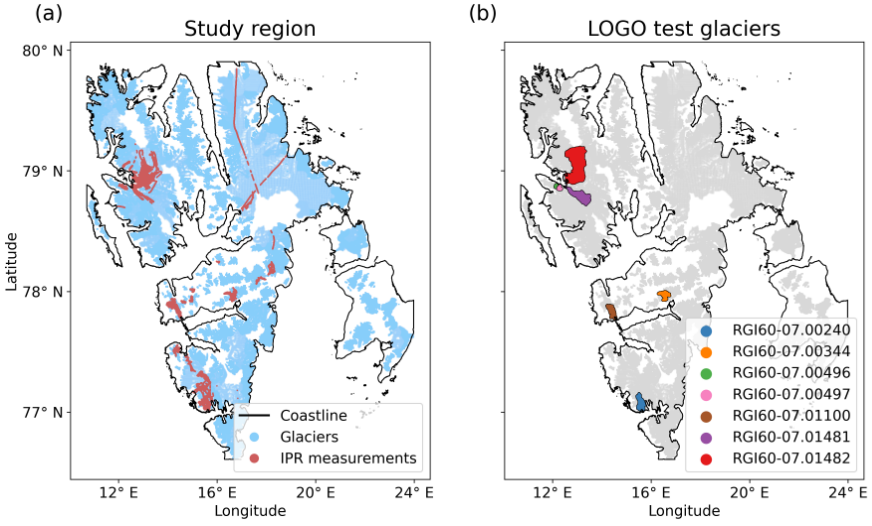

Figure 2. Glaciers in the training dataset. (a) The locations of in situ measurements are marked in red. (b) The locations of the test glaciers. The coastline is retrieved from Moholdt, G. et al. (2021).

The rate of thickness change dh/dt is retrieved from ASTER DEM differences between 2015 and 2019 (Hugonnet et al., 2021). The data is projected onto the glacier grids using OGGM and then smoothed with a two-dimensional Gaussian filter.

3.5 Thickness Data

The data-driven machine learning model needs ice thickness data as ground truth for its supervised training. The Glacier Thickness Database (GlaThiDa) is a comprehensive public database of in situ glacier thickness measurements collected from various studies (GlaThiDa Consortium, 2020). Version 3.1.0 was released in 2020 with nearly one million measurements from ice penetrating radar (IPR) on 207 glaciers or ice caps in Svalbard.

In situ ice thickness measurements are not error-free. GlaThiDa lists reported uncertainties of almost 80% of the measurements in Svalbard. The mean and standard deviation of the thickness uncertainty are 6.2 m and 4.4 m with a maximum uncertainty of 21 m.

During the preprocessing, the measurements are projected onto the OGGM glacier grids by aggregating and averaging them at their nearest point on the glacier grid. We only consider aggregated ice thicknesses where the average acquisition year is after 2000. That leaves us with $27\,554$ points labelled with ice thickness on 65 glaciers.

The full dataset of points with and without ice thickness labels consists of over 3 million data points from the grids of 1465 glaciers. Figure 2 (a) displays the considered glaciers in light blue and the acquisition lines of the IPR measurements in red.

https://doi.org/10.5194/egusphere-2024-1732
Preprint. Discussion started: 14 June 2024

RGI ID	Area [km^2]	IPR Mean [m]	IPR StD [m]
RGI60-07.00240	64.211	216.5	94.3
RGI60-07.00344	36.087	161.3	70.1
RGI60-07.00496	5.016	82.1	39.5
RGI60-07.00497	6.249	87.6	43.8
RGI60-07.01100	50.408	146.3	61.1
RGI60-07.01481	108.297	240.6	97.1
RGI60-07.01482	378.765	317.9	171.8

Table 1. Area of each test glacier together with the mean and standard deviation of the IPR ice thickness measurements.

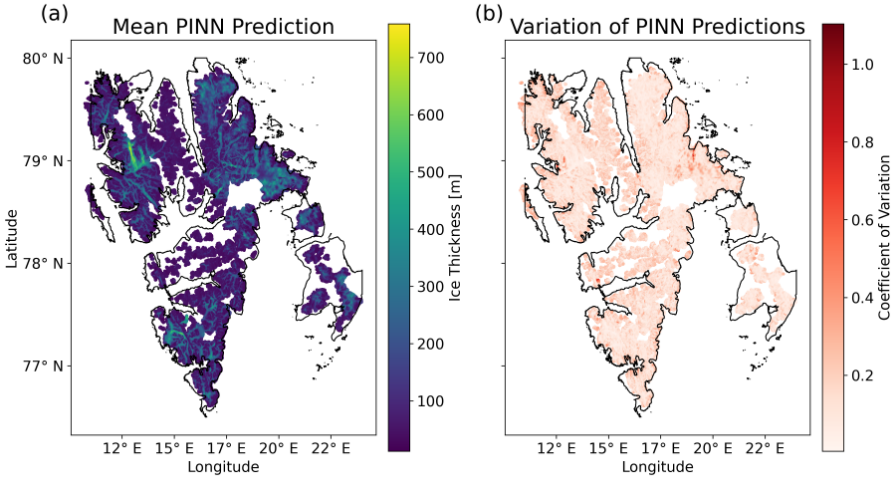

Figure 3. Mean (left) and coefficient of variation (right) of the ice thickness predictions from all seven models from CV.

3.6 Test glaciers

Seven glaciers with the most in situ measurements are chosen as test glaciers for the LOGO CV. They differ in size and mean thickness and are located in different areas of Spitsbergen. No glaciers on Barentsøya and Edgeøya are mapped well enough to use them as test glaciers.

Figure 2 (b) shows the location of the test glaciers. Table 1 lists the area, mean thickness, and standard deviation of the thickness measurements on those glaciers.

https://doi.org/10.5194/egusphere-2024-1732
Preprint. Discussion started: 14 June 2024

Test glacier	In-sample validation		LOGO validation	
ID	RMSD [m]	MAPD	RMSD [m]	MAPD
RGI60-07.00240	29	0.32	82	0.31
RGI60-07.00344	30	0.37	54	0.40
RGI60-07.00496	32	0.35	40	0.51
RGI60-07.00497	29	0.32	49	0.82
RGI60-07.01100	30	0.35	42	0.70
RGI60-07.01481	29	0.35	79	0.75
RGI60-07.01482	28	0.33	121	0.58
Mean	30	0.34	67	0.58

Table 2. Results of the LOGO CV.

4 Results

The LOGO CV produces seven models with the same architecture but different model weights. Each model was trained on all the unlabelled data to enforce the physical constraints at every point. After putting aside the test glacier's labelled data, the remaining glaciers' labelled data was randomly split into 60% training and 40% validation data.

The in-sample performance is measured based on the validation data the model did not see during the training. Table 2 lists the RMSD and MAPD for the in-sample validation data. The PINNs predict glacier ice thickness with a mean in-sample RMSD of 30 m corresponding to a MAPD of 34%.

Figure 3 shows the mean ice thickness prediction and the coefficient of variation over all seven LOGO models for the study region. The coefficient of variation measures the variability in relation to the mean ice thickness at each point of the grid. 90% of the points have a variability below 0.25. This is low compared to the variation between the three physics-based models (Farinotti et al., 2019; Millan et al., 2022; van Pelt and Frank, 2024), with more than 0.70 variability for 90% of the points. The PINN models agree with their predictions, although they were trained with different sets of ice thickness measurements as boundary conditions.

4.1 LOGO results

The model performance for the test glaciers delivers insights on the performance we can expect for glaciers where we do not have any in situ measurements. Table 2 shows the results of the LOGO validation for each of the test glaciers. As expected, the RMSD and MAPD are significantly higher than for the in-sample validation data. Figure 4 shows the difference between the model's ice thickness estimate and the in situ measurements along the IPR acquisition lines for the seven LOGO test glaciers that were excluded from the dataset during training. Overall, the models underestimated the ice thickness. However, the model

https://doi.org/10.5194/egusphere-2024-1732
Preprint. Discussion started: 14 June 2024

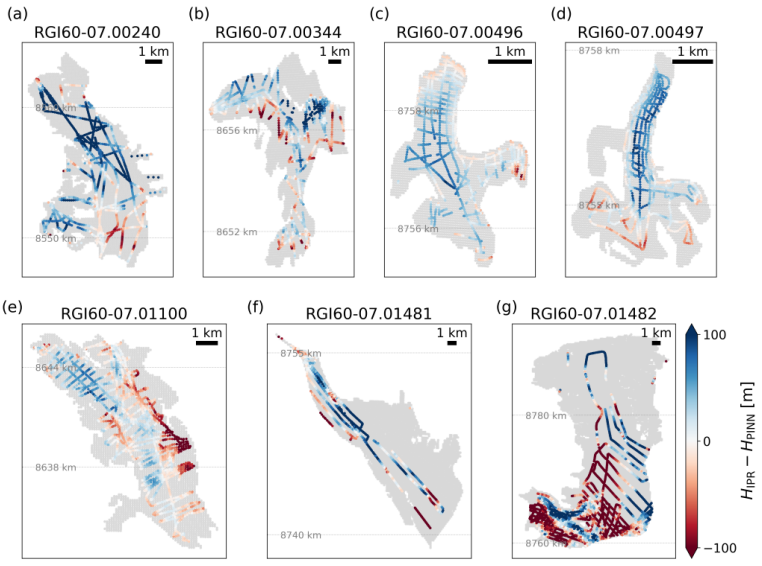

Figure 4. Difference between the predicted ice thickness and the IPR ice thickness measurements for the seven LOGO test glaciers.

trained without thickness data of glacier RGI60-07.01482 overestimates its ice thickness. The ice thickness estimates for the entire grids of the glaciers are displayed in Fig. A1.

The test glaciers differ significantly in mean ice thicknesses (see Table 1). For thinner glaciers like RGI60-07.00496 and RGI60-07.00497, the MAPD is very high, although their RMSD is comparable to the in-sample scores. The RMSD of glacier RGI60-07.01482 is four times as high as the mean RMSD score of the in-sample glaciers, but as, on average, in situ measurements are very thick, the MAPD is closer to the in-sample MAPD than the MAPD for a thinner test glacier. This makes it clear that considering both validation scores is necessary to view the model performance accurately.

Over all seven test glaciers, the mean RMSD and the mean MAPD are about 67 m and 0.58%, respectively, i.e. significantly worse than the in-sample metric. This indicates that the PINN is overfitting on glaciers with thickness measurements.

4.2 Comparison to other estimates

As we do not have full coverage with in situ measurements, the model scores only represent the model's performance at the acquisition lines of the IPR measurements. Therefore, we compare the ice thickness predictions to the estimates of Millan et al. (2022), Farinotti et al. (2019), and van Pelt and Frank (2024) to see how much the estimates differ. All of those ice-thickness products are derived from physics-based models. Farinotti et al. (2019) estimated ice thickness using an ensemble of up to five models; therefore, the name *consensus* estimate. Millan et al. (2022) rely on a single model that uses the shallow-ice

https://doi.org/10.5194/egusphere-2024-1732
Preprint. Discussion started: 14 June 2024

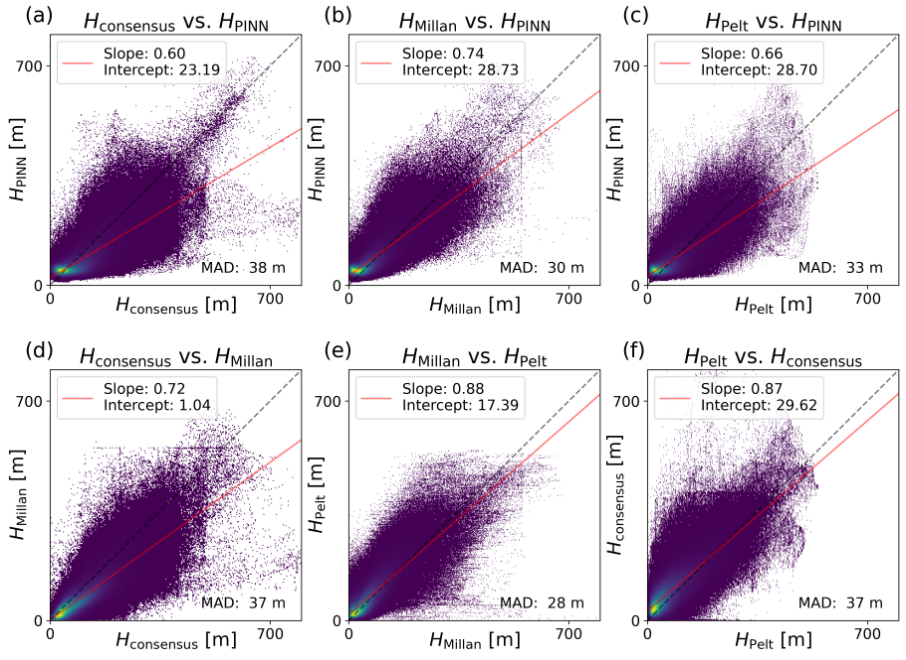

Figure 5. Comparisons of ice thickness estimates: a) Ice thickness prediction from the ensemble of PINNs versus ice thickness estimates of Millan et al. (2022), Farinotti et al. (2019), and van Pelt and Frank (2024). b) Ice thickness estimates from physics-based models against each other. The color indicates the point density (the brighter the denser). The solid red line shows the linear fit between the two ice thickness products. The black dashed line is the 1:1 line.

approximation and surface motion to compute ice thickness. van Pelt and Frank (2024) use two inverse methods, one for small and one for larger and surging glaciers, to create their ice thickness product.

The plots in Fig. 5 show the scatter plot of the ice thicknesses of the PINN ensemble estimates versus the three other estimates for each point in the study region. The solid red line shows the linear fit between the two ice thickness products, while the black dashed line is the 1:1 line. The values of slope and intersect for the linear fits indicate that the PINN estimates agree slightly less on the ice thickness at each grid point than the three physics-based models. Comparing the mean ice thickness estimate from the ensemble of PINNs to the estimates of physics-based models shows that the deviations from the other estimates are within the range of the differences between the physics-based models. Mean absolute differences (MAD) between the PINN and physics-based predictions are all close, with a mean of 33.7 m. The MAD between the physics-based models is in the same order of magnitude, and their mean MAD is at 34.3 m.

Since physics-based models also work with simplifications of ice dynamics, their ice thickness products can not be taken as definitive truth. Comparing the predictions of the PINN ensemble to those only serves to estimate the qualitative validity of the ice thickness predicted by the PINN ensemble. The overall correlation between the ice thickness estimates leads us to believe that the PINN ensemble produces valid ice thickness estimates.

https://doi.org/10.5194/egusphere-2024-1732
Preprint. Discussion started: 14 June 2024

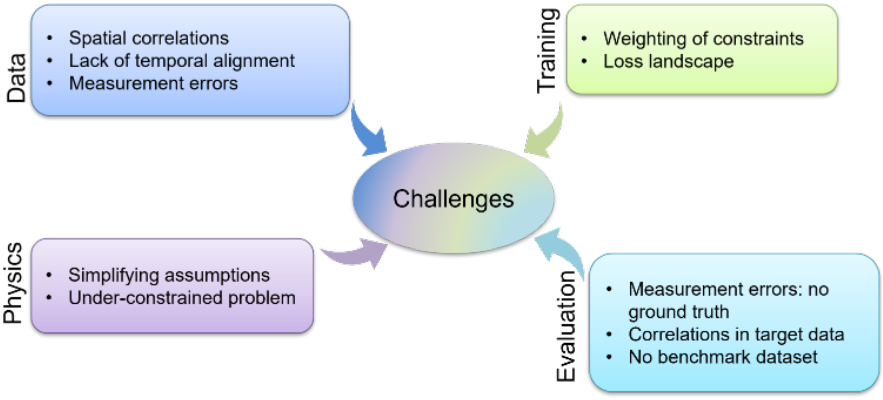

Figure 6. Challenges for PINNs in a real-world setting.

4.3 Depth-averaged velocity

The models estimate the depth-averaged velocities in x- and y-direction. There is no ground truth data for the depth-averaged velocities, so we can not evaluate the models's accuracy. However, during training, the loss of the predicted velocities is reduced significantly, showing that the constraints of Eq. (3) are enforced.

5 Discussion

Evaluating the PINN performance with the in-sample validation set and comparing predicted ice thicknesses to other products suggests that the PINN produces reasonable ice thickness estimates. However, testing the model with a glacier-wise CV scheme unveils the lack of generalizability to glaciers without any measurement data. The differences between the predicted ice thicknesses and the measurements are much higher for the test glaciers of the LOGO CV than for the in-sample validation. We identified several factors that may improve generalizability but are also challenging to address. The schematic of Fig. 6 shows an overview of the domains and the particular issues we judge as the most pressing to address.

5.1 Data

Although PINNs are generally relatively less dependent on training data than purely data-driven methods, their performance relies on the quantity and quality of input data. This study collected the thickness measurements from 65 glaciers. The individual measurements lie close together along the acquisition lines. As a result, most of the measurements have high spatial correlations with each other. The amount of independent training data to learn the physics of glaciers is, therefore, far less than the actual number of measurements. On the other hand, redundant data introduces a bias. To reduce the overfitting, we could

https://doi.org/10.5194/egusphere-2024-1732
Preprint. Discussion started: 14 June 2024

reduce the correlations in the training dataset by averaging or sub-sampling the observations, for example. This should improve the performance on glaciers without any labelled training data.

Secondly, the training data is not aligned temporally. In situ ice thickness measurements were collected between 2000 and 2017, while the surface velocity was derived from satellite data acquired between 2017 and 2018. Using surface velocity for the same year as ice thickness was measured would result in a better estimate of the ice flux for the labelled data and therefore improve physical consistency within the model. However, this is only important if either ice thickness or velocity changes significantly with time.

Lastly, we do not consider measurement errors in our dataset. Future work could account for measurement errors with standard uncertainty quantification methods or even introduce error margins to the loss components as Morlighem et al. (2011) did.

5.2 Model training

Training PINNs is difficult (Xu et al., 2023). One major challenge is to find the optimal balance between the multi-scale contributions of each loss component. We empirically set the loss weights to a fixed value for the entire training process. A more sophisticated approach would be to automatically update the weights of the loss components during the training. This introduces a minimal computational overhead but prevents the model from minimizing certain loss components faster than others (Wang et al., 2023).

The multiple loss components create another challenge: the loss landscape is highly complex, and finding its global minimum is difficult. Recently, Rathore et al. (2024) investigated different optimizers for PINNs and showed that combined first- and second-order optimizers lead to faster convergence. Implementing their newly introduced second-order optimizer could improve the PINN convergence. Optimizing the loss function is a rapidly evolving area of research, and we expect significant advances to be achieved soon.

5.3 Physical constraints

Ultimately, we want to achieve the physical consistency of our predictions. However, the physics-aware losses are based on simplifying assumptions to make the problem tractable. In our case, the two elements that should be revised are modelling the mass balance and estimating the basal sliding velocity.

In this study, we use the *ConstantMassBalance* model from OGGM out of the box without further calibration. Using a more sophisticated mass balance model to precisely calibrate it for our purpose could enhance physical consistency, leading to better results.

The estimate of basal sliding is a simplified approach using a threshold calibrated by Millan et al. (2022). We could, in principle, circumvent the need to estimate basal sliding contributions using surface velocity data acquired during winter months. In winter, due to the absence of extensive meltwater, basal sliding is inhibited. Therefore, the measurable surface velocity will have less contribution from basal sliding. In this way, we could avoid estimating the β-parameter and still have a reasonable approximation of the depth-averaged deformation velocity.

https://doi.org/10.5194/egusphere-2024-1732
Preprint. Discussion started: 14 June 2024

Simplifying the physics also means we pick physical constraints that are easy to incorporate. In return, we have a highly under-constrained system. This may cause a lack of generalizability for glaciers without any ice thickness data. It would surely be beneficial to incorporate additional constraints like momentum conservation. While this is technically easy to do, it further complicates the optimization of the model. Therefore, high-quality data is required to support the additional constraint. Depending on the data quality, we risk introducing more uncertainty than improving the physical consistency.

5.4 Evaluation

In geospatial machine learning, evaluation is generally challenging (Rolf, 2023). As mentioned in Sect. 2.4, the in situ measurements are heavily correlated since they are clustered on only a fraction of the glaciers. Therefore, we employ a spatially-aware evaluation method to estimate the true model performance. However, the CV procedure only includes seven of 65 glaciers with measurements. There is no guarantee that we will fully capture the model performance.

As mentioned, in situ ice thicknesses are subject to measurement errors, and some measurements might have higher errors than others. To be as precise as possible when evaluating the model performance, we should consider the trustworthiness of every ice thickness label.

Ultimately, our problem has no benchmark dataset, so it is impossible to know the model performance exactly. Although we compare our ice thickness estimates with others, these also have errors that are not well constrained and are, in no respect, benchmarks that can be used for uncertainty quantification. It is, therefore, difficult to state which estimation method considered here produces the most reliable ice thickness estimate.

Despite the above-mentioned limitations, we show that a relatively simple PINN can produce reasonable ice thickness estimates while treating an entire area and not only a single glacier at once. Although the lack of high quality data is an overarching challenge that can hardly be overcome, we expect that by implementing the proposed adjustments in data curation, model training, and physical constraints, the physical consistency and accuracy of the model will be improved. This may especially be the case for glaciers without measured ice thicknesses.

6 Conclusions

We have demonstrated that it is possible to train a physics-aware machine learning model to produce ice thickness estimates for multiple glaciers, including glaciers without any ground truth ice thickness: in other words, out of sample targets. We deploy a relatively simple physical constraint by imposing mass conservation in the loss function of the PINN. More complex approaches and physical constraints could be employed (Karniadakis et al., 2021) and would, we anticipate, improve the results further. Nonetheless, we demonstrate that physics-aware machine learning is a promising approach for tackling a geophysical problem where a physical law or condition provides a strong constraint for the solution. There are many other geophysical problems where, for example, conservation of mass, energy or momentum would provide a similarly effective constraint and would lead to a more scientifically meaningful result, as breaching such constraints is non-physical.

https://doi.org/10.5194/egusphere-2024-1732
Preprint. Discussion started: 14 June 2024

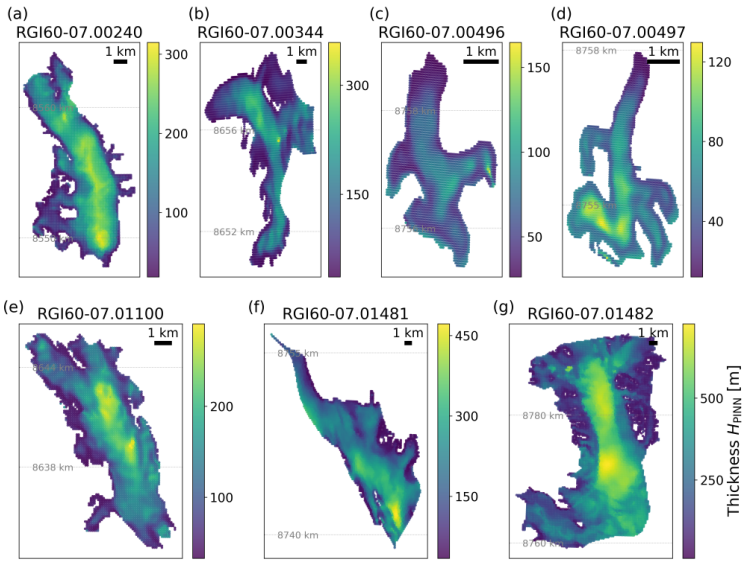

Figure A1. Ice thickness prediction for the seven LOGO test glaciers.

Code and data availability. The code and data that was used to train and evaluate the model as well as generating the figures are available at https://doi.org/10.5281/zenodo.11474955. Additionally, the code can be viewed at https://github.com/viola1593/glacier_pinn.

Appendix A: PINN ice thickness prediction on LOGO test glaciers

The ice thicknesses of the seven test glaciers were estimated by the model that was trained without the in situ measurements of the respective glaciers as ground truth data. The results are displayed in Fig. A1

Appendix B: Relation between surface and depth-averaged velocity to set l_{lower}

To derive a relation between the surface velocity and the depth-averaged velocity, we follow the analysis in Cuffey and Paterson (2010). Let u be the x-component of velocity and H be the ice thickness. Assuming parallel flow, the glacier deforms in simple shear, so the only nonzero deviatoric stress is τ_{xz}, and the z-component of the velocity is also zero. Therefore, the creep relation derived by Glen (1955) simplifies to

$$\frac{1}{2}\frac{du}{dz} = A\tau_{xz}^{n} \tag{B1}$$

https://doi.org/10.5194/egusphere-2024-1732
Preprint. Discussion started: 14 June 2024

where A is the creep parameter and n the creep exponent. We assume a linear increase of shear stress along the glacier depth

$$\tau_{xz} = \tau_b \left[1 - \frac{z}{H}\right] \tag{B2}$$

with τ_b being the shear stress at the bed of the glacier. If we now integrate B1 along the vertical direction up to z we get

$$u(z) = u_b + \frac{2A}{n+1}\tau_b^n H \left[1 - \left[1 - \frac{z}{H}\right]^{n+1}\right]. \tag{B3}$$

Accordingly, the velocity at the surface is given by

$$u_s = u_b + \frac{2A}{n+1}\tau_b^n H \tag{B4}$$

and integrating B3 along the vertical axis to derive the depth-averaged velocity we get

$$\bar{u} = u_b + \frac{2A}{n+2}\tau_b^n H. \tag{B5}$$

A is assumed to be constant although depends on temperature and other variables that change within a glacier profile. Temperate glaciers are nearly isothermal, whereas in cold-based glaciers, the temperature increases with a smaller distance to the bed. The highest values of A are found near the glacier bed. Therefore, the shear deformation is concentrated closer to the base than in a temperate glacier. The velocities within the bottom half of the glacier are sensitive to the value of A as it multiplies stress to the power of n. As stress decreases in the upper half of the glacier, the velocity is insensitive to the values of A there.

From B4 and B5, we can derive the relation between the surface velocity and the depth-averaged velocity in the case of parallel flow and if there is no basal sliding:

$$\frac{\bar{u}}{u} = \frac{n+1}{n+2} = 0.8 \tag{B6}$$

for $n = 3$. After all, parallel flow in a glacier is a strong assumption and $n = 3$ is not always the case. Therefore, we set l_{lower} to 0.7 to allow more flexibility in estimating the depth-averaged velocity.

Author contributions. VS conceived the study, conducted the analysis and developed the ML framework and wrote the manuscript with contributions from all authors. JLB provided advice on how to incorporate the physics into the model and model training setup. XZ conceived the study and contributed to the interpretation of the results.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work is jointly supported by the German Federal Ministry for Economic Affairs and Climate Action in the framework of the “National Center of Excellence ML4Earth” (grant number: 50EE2201C), the German Federal Ministry of Education and Research (BMBF) in the framework of the international future AI lab "AI4EO – Artificial Intelligence for Earth Observation: Reasoning,

https://doi.org/10.5194/egusphere-2024-1732
Preprint. Discussion started: 14 June 2024

Uncertainties, Ethics and Beyond" (grant number: 01DD20001), the Helmholtz Association under the joint research school “Munich School for Data Science - MUDS“, and the Munich Center for Machine Learning (MCML). Jonathan Bamber also received funding from the European Union’s Horizon 2020 research and innovation programme through the project Arctic PASSION (grant number: 101003472). We thank Dr. Ward van Pelt and Thomas Frank for providing us with their ice thickness dataset that was used for comparison in this study. We also want to thank Thomas O. Teisberg for the insightful suggestion on using Fourier feature embedding to improve the model performance.

https://doi.org/10.5194/egusphere-2024-1732
Preprint. Discussion started: 14 June 2024

References

Anilkumar, R., Bharti, R., Chutia, D., and Aggarwal, S. P.: Modelling point mass balance for the glaciers of the Central European Alps using machine learning techniques, The Cryosphere, 17, 2811–2828, https://doi.org/10.5194/tc-17-2811-2023, 2023.

Bolibar, J., Rabatel, A., Gouttevin, I., Galiez, C., Condom, T., and Sauquet, E.: Deep learning applied to glacier evolution modelling, The Cryosphere, 14, 565–584, https://doi.org/10.5194/tc-14-565-2020, 2020.

Bouchayer, C., Aiken, J. M., Thøgersen, K., Renard, F., and Schuler, T. V.: A Machine Learning Framework to Automate the Classification of Surge-Type Glaciers in Svalbard, Journal of Geophysical Research: Earth Surface, 127, e2022JF006 597, https://doi.org/10.1029/2022JF006597, 2022.

Copernicus: Copernicus DEM GLO-90, [data set], https://doi.org/https://doi.org/10.5270/ESA-c5d3d65.

Cuffey, K. and Paterson, W.: The Physics of Glaciers, Elsevier Science, 4. ed. edn., ISBN 978-0-08-091912-6, 2010.

Farinotti, D., Brinkerhoff, D. J., Clarke, G. K. C., Fürst, J. J., Frey, H., Gantayat, P., Gillet-Chaulet, F., Girard, C., Huss, M., Leclercq, P. W., Linsbauer, A., Machguth, H., Martin, C., Maussion, F., Morlighem, M., Mosbeux, C., Pandit, A., Portmann, A., Rabatel, A., Ramsankaran, R., Reerink, T. J., Sanchez, O., Stentoft, P. A., Singh Kumari, S., van Pelt, W. J. J., Anderson, B., Benham, T., Binder, D., Dowdeswell, J. A., Fischer, A., Helfricht, K., Kutuzov, S., Lavrentiev, I., McNabb, R., Gudmundsson, G. H., Li, H., and Andreassen, L. M.: How accurate are estimates of glacier ice thickness? Results from ITMIX, the Ice Thickness Models Intercomparison eXperiment, The Cryosphere, 11, 949–970, https://doi.org/10.5194/tc-11-949-2017, 2017.

Farinotti, D., Huss, M., Fürst, J. J., Landmann, J., Machguth, H., Maussion, F., and Pandit, A.: A consensus estimate for the ice thickness distribution of all glaciers on Earth, Nature Geoscience, 12, 168–173, https://doi.org/10.1038/s41561-019-0300-3, 2019.

GlaThiDa Consortium: Glacier Thickness Database 3.1.0, World Glacier Monitoring Service [data set], https://doi.org/10.5904/wgms-glathida-2020-10, 2020.

Glen, J. W.: The Creep of Polycrystalline Ice, Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences, 228, 519–538, https://doi.org/10.1098/rspa.1955.0066, 1955.

Hugonnet, R., McNabb, R., Berthier, E., Menounos, B., Nuth, C., Girod, L., Farinotti, D., Huss, M., Dussaillant, I., Brun, F., and Kääb, A.: Accelerated global glacier mass loss in the early twenty-first century, Nature, 592, 726–731, https://doi.org/10.1038/s41586-021-03436-z, 2021.

Jouvet, G.: Inversion of a Stokes glacier flow model emulated by deep learning, Journal of Glaciology, 69, 13–26, https://doi.org/10.1017/jog.2022.41, 2023.

Jouvet, G., Cordonnier, G., Kim, B., Lüthi, M., Vieli, A., and Aschwanden, A.: Deep learning speeds up ice flow modelling by several orders of magnitude, Journal of Glaciology, 68, 651–664, https://doi.org/10.1017/jog.2021.120, 2022.

Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., and Yang, L.: Physics-informed machine learning, Nature Reviews Physics, 3, 422–440, https://doi.org/10.1038/s42254-021-00314-5, 2021.

Koch, M., Seehaus, T., Friedl, P., and Braun, M.: Automated Detection of Glacier Surges from Sentinel-1 Surface Velocity Time Series—An Example from Svalbard, Remote Sensing, 15, 1545, https://doi.org/10.3390/rs15061545, 2023.

Lagaris, I., Likas, A., and Fotiadis, D.: Artificial neural networks for solving ordinary and partial differential equations, IEEE Transactions on Neural Networks, 9, 987–1000, https://doi.org/10.1109/72.712178, 1998.

Leng, K. and Thiyagalingam, J.: On the Compatibility between Neural Networks and Partial Differential Equations for Physics-informed Learning, arXiv [preprint], https://doi.org/10.2139/ssrn.4392241, 2023.

https://doi.org/10.5194/egusphere-2024-1732
Preprint. Discussion started: 14 June 2024

Maussion, F., Butenko, A., Champollion, N., Dusch, M., Eis, J., Fourteau, K., Gregor, P., Jarosch, A. H., Landmann, J., Oesterle, F., Recinos, B., Rothenpieler, T., Vlug, A., Wild, C. T., and Marzeion, B.: The Open Global Glacier Model (OGGM) v1.1, Geoscientific Model Development, 12, 909–931, https://doi.org/10.5194/gmd-12-909-2019, 2019.

Millan, R., Mouginot, J., Rabatel, A., and Morlighem, M.: Ice velocity and thickness of the world's glaciers, Nature Geoscience, 15, 124–129, https://doi.org/10.1038/s41561-021-00885-z, 2022.

Min, Y., Mukkavilli, S. K., and Bengio, Y.: Predicting ice flow using machine learning, arXiv [preprint], https://doi.org/10.48550/arXiv.1910.08922, 2019.

Moholdt, G., Maton, J., Majerska, M., and Kohler, J.: Annual coastlines for Svalbard, Norwegian Polar Institute. [data set], https://doi.org/https://doi.org/10.21334/npolar.2021.21565514, 2021.

Morlighem, M., Rignot, E., Seroussi, H., Larour, E., Ben Dhia, H., and Aubry, D.: A mass conservation approach for mapping glacier ice thickness, Geophysical Research Letters, 38, https://doi.org/10.1029/2011GL048659, 2011.

Raissi, M., Perdikaris, P., and Karniadakis, G.: Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, Journal of Computational Physics, 378, 686–707, https://doi.org/10.1016/j.jcp.2018.10.045, 2018.

Rathore, P., Lei, W., Frangella, Z., Lu, L., and Udell, M.: Challenges in Training PINNs: A Loss Landscape Perspective, arXiv [preprint], https://doi.org/10.48550/arXiv.2402.01868, 2024.

RGI Consortium: Randolph Glacier Inventory - A Dataset of Global Glacier Outlines, Version 6, Region 7, NSIDC: National Snow and Ice Data Center. [data set], https://doi.org/10.7265/4M1F-GD79, 2017.

Rolf, E.: Evaluation Challenges for Geospatial ML, arXiv [preprint], https://doi.org/10.48550/arXiv.2303.18087, 2023.

Tancik, M., Srinivasan, P. P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., Ramamoorthi, R., Barron, J. T., and Ng, R.: Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains, arXiv [preprint], https://doi.org/10.48550/arXiv.2006.10739, 2020.

Teisberg, T. O., Schroeder, D. M., and MacKie, E. J.: A Machine Learning Approach to Mass-Conserving Ice Thickness Interpolation, in: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, pp. 8664–8667, IEEE, Brussels, Belgium, https://doi.org/10.1109/IGARSS47720.2021.9555002, 2021.

van Pelt, W. and Frank, T.: A new glacier thickness and bed map for Svalbard, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2024-1525, 2024.

Wang, S., Sankaran, S., Wang, H., and Perdikaris, P.: An Expert's Guide to Training Physics-informed Neural Networks, arXiv [preprint], https://doi.org/10.48550/arXiv.2308.08468, 2023.

Xu, Q., Shi, Y., Bamber, J., Tuo, Y., Ludwig, R., and Zhu, X. X.: Physics-aware Machine Learning Revolutionizes Scientific Paradigm for Machine Learning and Process-based Hydrology, arXiv [preprint], https://doi.org/10.48550/arXiv.2310.05227, 2023.