Ice-flow model emulator based on physics-informed deep learning

Guillaume Jouvet, Guillaume Cordonnier

Ice flow model emulator based on physics-informed deep learning

Guillaume Jouvet[1], Guillaume Cordonnier[2]

[1] *Université de Lausanne, IDYST, 1015 Lausanne, Switzerland*
<guillaume.jouvet@unil.ch>
[2] *Inria, Université Côte d'Azur, Sophia-Antipolis, France*

ABSTRACT. Convolutional Neural Networks (CNN) trained from high-order ice flow model realizations have proven to be outstanding emulators in terms of fidelity and computational performance. However, the dependence on an ensemble of realizations of an instructor model renders this strategy difficult to generalize to a variety of ice flow regimes found in the nature. To overcome this issue, we adopt the approach of physics-informed deep learning, which fuses traditional numerical solutions by finite differences/elements and deep learning approaches. Here, we train a CNN to minimise the energy associated with high-order ice flow equations within the time iterations of a glacier evolution model. As a result, our emulator is a promising alternative to traditional solvers thanks to its high computational efficiency (especially on GPU), its high fidelity to the original model, its simplified training (without requiring any data), its capability to handle a variety of ice flow regimes and memorize previous solutions, and its relatively simple implementation. Embedded into the "Instructed Glacier Model" (IGM) framework, the potential of the emulator is illustrated with three applications including a large-scale high-resolution (2400x4000) forward glacier evolution model, an inverse modelling case for data assimilation, and an ice shelf.

INTRODUCTION

In glacier and ice sheet models, ice is commonly described as a viscous non-Newtonian [Glen, 1953] fluid whose motion is governed by the 3D nonlinear Glen-Stokes equations [Greve and Blatter, 2009]. Solving these equations usually remains very costly compared to other glacial underlying processes. To reduce the costs, the ice flow equations are often simplified by neglecting higher-order terms in the aspect ratio of the ice domain ϵ (thickness versus length) considering it to be usually small. The truncation of the second-order terms in ϵ yields the First-Order Approximation (FOA) model [Blatter, 1995], which consists of a 3D non-linear elliptic equation [Colinge and Rappaz, 1999] for the horizontal velocity and remains expensive. Going one step further, the Shallow Ice Approximation [Hutter, 1983] (SIA) is obtained after dropping the first-order terms in ϵ in the FOA model. As a result, the analytical solution of SIA is computationally inexpensive to implement. The SIA remains a reference model for many applications [e.g., Maussion et al., 2019], despite strongly-simplifying mechanical assumptions and applicability limited to areas where ice flow is dominated by vertical shearing [Greve and Blatter, 2009]. The transfer of numerical methods from Central Processing Units (CPU) on Graphics Processing Units (GPU) architectures is currently a promising approach to bypass the computational bottleneck associated with high-order modelling [Brædstrup et al., 2014], however, massive parallelisation of solvers on GPU remains a complex task [Räss et al., 2020].

As an alternative to traditional solvers, deep learning surrogate models (or emulators) have been found very promising in reducing computational costs with minimal loss of accuracy [Brinkerhoff et al., 2021, Jouvet et al., 2022, He et al., 2023]. Deep learning is based on Artificial Neural Networks (ANNs), which are trained to capture the most essential relationship between the input and the output of an instructor model. The ANN is intended to be an efficient substitute for the original model within the range defined by the training dataset. Following this strategy, the computationally expensive Glen-Stokes model could be emulated by a simple Convolutional Neural Network (CNN) by Jouvet et al. [2022] with a speedup of several orders of magnitude and high fidelity levels in the case of mountain glaciers, and major benefits for inverse modelling purposes [Jouvet, 2023]. Another key asset of ANNs is that they run very efficiently on GPUs, permitting additional significant speed-ups, especially when

modelling high spatial resolution domains. However, the dependence on an instructor model makes the training of such an emulator technically difficult, not very flexible, and therefore limits its ability to generalize its validity range beyond the training data and its given spatial resolution.

In recent years, Physics-informed neural networks (PINNs) have emerged as a powerful approach in surrogate modelling to directly enforce physical laws (such as partial differential equations) in the learning process instead of matching datasets generated from physical models [e.g., Raissi et al., 2019]. Basic PINNs are trained to minimise the residual associated with the equations and the boundary conditions [Markidis, 2021]. In contrast, Variational PINNs (VPINNs) exploit the minimization form (or equivalently the variational form) of the problem as a loss function [Kharazmi et al., 2019], which has the advantage of involving derivatives of lower orders compared to residuals. An important aspect of VPINNs is their connections with traditional Finite Element Methods (FEM). For example, a standard FEM solver applied to an elliptic problem represents the solution in a finite element approximation space spanned by mesh-defined basis functions and seeks the function that minimises the associated energy in the approximation space [Ern and Guermond, 2004]. On the contrary, the Deep-Ritz method proposed by Yu et al. [2018] (which belongs to the category of VPINN) represents the solution as a neural network in an approximation space generated by the parameters of a neural network.

In ice flow modelling, PINNs have been used by Riel et al. [2021] to learn the time evolution of drag in glacier beds from observations of ice velocity and elevation and by Riel and Minchew [2022] to calibrate ice flow law parameters and perform uncertainty quantification. Recently, Cui et al. [2022] proposed a mesh-free method to solve Glen-Stokes equations using an approach inspired by the Deep-Ritz method.

In this paper, we propose two different methods to compute First-Order Approximation (FOA) ice-flow efficiently on GPU by exploiting the minimisation form associated with the FOA model and using optimisation techniques based on automatic differentiation and stochastic gradient. The first one is a conventional numerical solver, which is used mostly here as a reference to evaluate the second one. The second one on which the paper focuses is an emulator based on deep-learning. In more detail, we take the CNN ice flow emulator introduced previously by Jouvet et al. [2022] and propose a new training strategy inspired by VPINN to remove the dependence on an instructor model and obtain a more generic emulator that is easier to implement and faster to train. Here we train our CNN ice flow emulator at minimising directly the energy instead of minimizing the misfit with solutions from an instructor model as done previously (Fig. 1). A similar approach was used by Cordonnier et al. [2023] for modelling terrain formation by glacial erosion. Their target was to generate realistic images in computer graphics, whereas we propose a thorough evaluation of the method and its potential for glaciological applications.

This paper is structured as follows: First, we introduce the physical model which includes the ice flow FOA model and its minimisation formulation. Second, we describe the numerical model which includes the spatial discretization, the energy-based FOA solver and deep learning emulator. Last, we present and discuss our assessment results and examples of modelling applications.

PHYSICAL MODEL

Let Ω be a rectangular horizontal domain supporting a glacier / volume of ice V at time t. Glacier bedrock and surface interfaces are defined by functions $b(x, y)$ and $s(x, y, t)$ where $(x, y) \in \Omega$. According to these definitions, the ice thickness h is defined as being the difference between the two: $h(x, y, t) = s(x, y, t) - b(x, y)$, and the three-dimensional volume of ice V is defined as

$$V = \{(x, y, z), b(x, y) \leq z \leq s(x, y, t), \quad (x, y) \in \Omega\},$$

which has two boundaries: the bedrock

$$\Gamma_b = \{(x, y, z), z = b(x, y), \quad (x, y) \in \Omega\}$$

and the surface

$$\Gamma_s = \{(x, y, z), z = s(x, y, t), \quad (x, y) \in \Omega\}$$

interfaces, see Figure 2. The two interfaces coincide in ice-free areas.

Given an initial glacier geometry, the time evolution in ice thickness $h(x, y, t)$ is determined by the mass conservation equation, which couples ice dynamics and surface mass balance through:

$$\frac{\partial h}{\partial t} + \nabla \cdot (\bar{\mathbf{u}} h) = \text{SMB}, \tag{1}$$

where $\nabla\cdot$ denotes the divergence operator with respect to horizontal variables (x, y), $\bar{\mathbf{u}} = (\bar{u}, \bar{v})$ is the vertically-averaged horizontal ice velocity field, and SMB the Surface Mass Balance function, which consists of the integration of ice accumulation and ablation over one year. Eq. (1) is generic and can be applied to model glacier evolution in number of applications provided adequate SMB and ice flow model components. In the following, we mostly focus on developing an efficient numerical method to compute the ice flow $\bar{\mathbf{u}}$ considering it is often the most computationally expensive component in glacier evolution model [Jouvet et al., 2022].

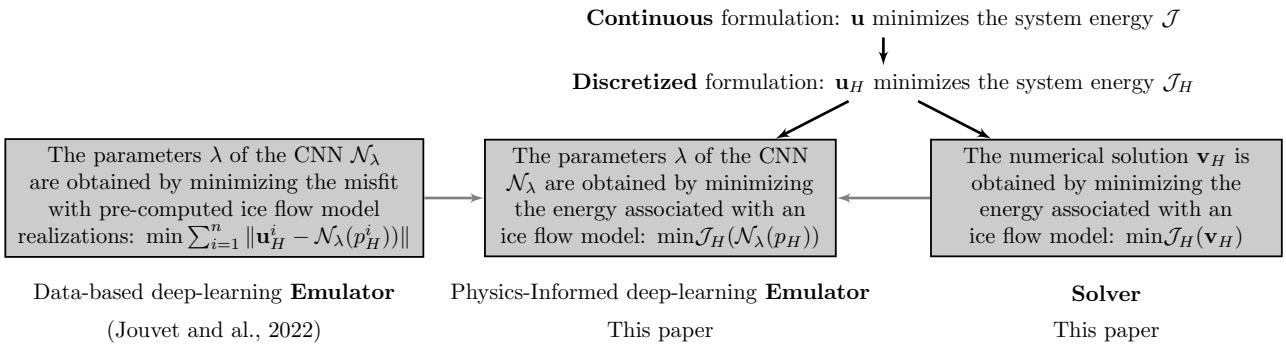

Fig. 1. Our Physics-Informed deep-learning emulator can be seen as a fusion of data-driven deep learning and traditional numerical solving strategies.

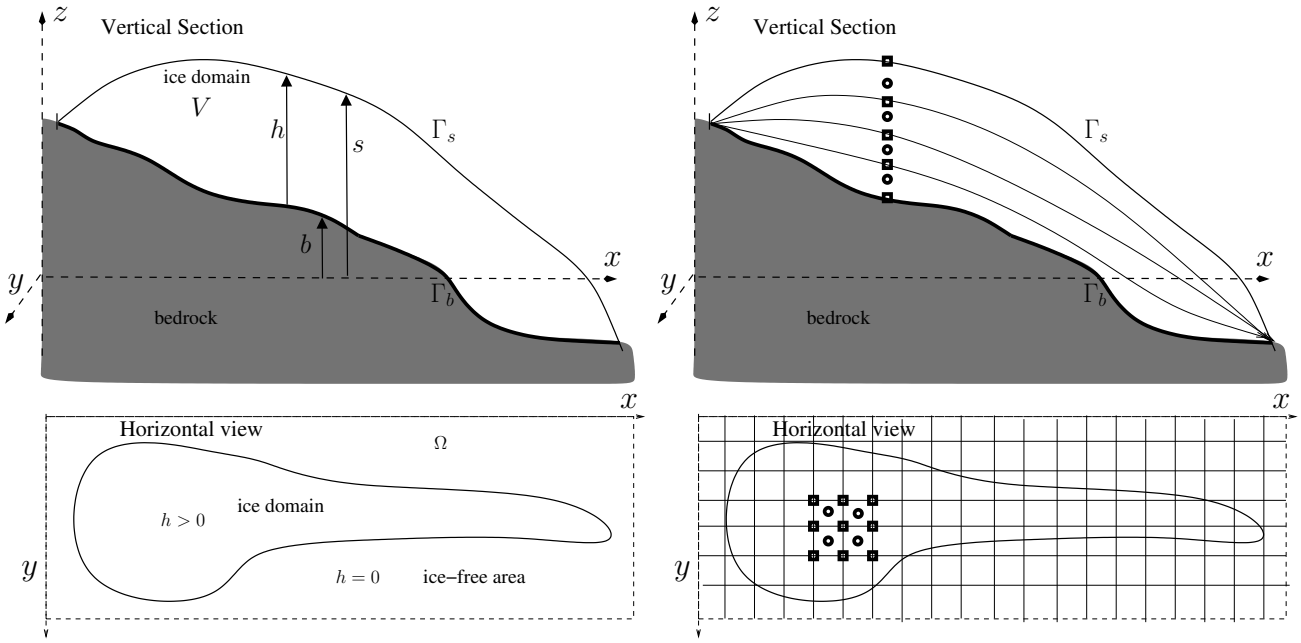

Fig. 2. Cross-section and horizontal view of a glacier with notations (left panel) and its spatial discretization (right panel), which is obtained using a regular horizontal grid and by subdividing the glacier into a pile of layers. All modelled variables (e.g. ice thickness) are computed at the corners of each cell of the 2D horizontal grid (materialised with squares) except the ice flow velocities, which are computed on the 3D corresponding grid. In contrast, the strain rate is computed on the staggered grid at the centre of each cells and layers (vizualized with circles).

Glen-Stokes model

The Stokes model consists of the momentum conservation equation when inertial terms are ignored, together with the incompressibility condition:

$$-\nabla \cdot \sigma = \rho \mathbf{g}, \qquad \text{in } V, \tag{2}$$

$$\nabla \cdot \mathbf{u} = 0, \qquad \text{in } V, \tag{3}$$

where σ is the Cauchy stress tensor, $\mathbf{g} = (0, 0, -g)$, g is the gravitational constant and $\mathbf{u} = (u_x, u_y, u_z)$ is the 3D velocity field. Let τ be the deviatoric stress tensor defined by

$$\sigma = \tau - PI, \tag{4}$$

where I is the identity tensor, P is the pressure field, with the requirement that $\mathrm{tr}(\tau) = 0$ so that $P = -(1/3)\mathrm{tr}(\sigma)$.

Glen's flow law [Glen, 1953], which describes the mechanical behaviour of ice, consists of the following nonlinear relation:

$$\tau = 2\mu D(\mathbf{u}), \tag{5}$$

where $D(\mathbf{u})$ denotes the strain rate tensor defined by

$$D(\mathbf{u}) = \frac{1}{2}(\nabla \mathbf{u} + \nabla \mathbf{u}^T), \tag{6}$$

μ is the viscosity defined by

$$\mu = \frac{1}{2} A^{-\frac{1}{n}} |D(\mathbf{u})|^{\frac{1}{n}-1}, \tag{7}$$

where $|Y| := \sqrt{(Y : Y)/2}$ denotes the norm associated with the scalar product $(:)$ (the sum of the element-wise product), $A = A(x, y) > 0$ is the Arrhenius factor and $n > 1$ is the Glen's exponent (here we take the most standard value $n = 3$). Note that A depend on the temperature of the ice [Paterson, 1994]. For simplicity, this paper assumes vertically constant ice temperature, however, this assumption could be released without further difficulties.

Boundary conditions

The boundary conditions that supplement (2), (3) are the following. Stress free force applies to the ice-air interface,

$$\sigma \cdot \mathbf{n} = 0, \quad P = 0, \qquad \text{on } \Gamma_s, \tag{8}$$

where $\mathbf{n}$ is an outer normal vector along Γ_s. Along the lower surface interface, the nonlinear Weertman friction condition reads [Hutter, 1983, Schoof and Hewitt, 2013]

$$\mathbf{u} \cdot \mathbf{n} = 0, \tag{9}$$

$$[(I - \mathbf{n}\mathbf{n}^T)\tau] \cdot \mathbf{n} = -c^{-m} |(I - \mathbf{n}\mathbf{n}^T) \cdot \mathbf{u}|^{m-1} (I - \mathbf{n}\mathbf{n}^T) \cdot \mathbf{u}, \tag{10}$$

on Γ_b for $k \in \{x, y\}$, where $m > 0$, $c = c(x, y) > 0$, and $\mathbf{n}$ is the outward normal unit vector to Γ_b. The relation (10) relates the basal shear stress $[(I - \mathbf{n}\mathbf{n}^T)\tau] \cdot \mathbf{n}$ to the sliding velocity $(I - \mathbf{n}\mathbf{n}^T) \cdot \mathbf{u}$, both of them projected onto the tangential plane. Note that $c = 0$ in case of no-sliding.

Minimization formulation

The above mentioned Glen-Stokes problem can be reformulated into variational and minimisation problems. We follow the derivation made by Jouvet [2016]. For that, we consider the following divergence free velocity field space [Girault and Raviart, 1986]:

$$\mathcal{X} := \{\mathbf{v} \in [W^{1,1+\frac{1}{n}}(V)]^3, \quad \nabla \cdot \mathbf{v} = 0, \quad \mathbf{v} \cdot \mathbf{n} = 0 \text{ on } \Gamma_b\},$$

where $W^{1,p}$ is the appropriate Sobolev space [Adams and Fournier, 2003]. The variational formulation associated with the Glen-Stokes problem writes: Find $\mathbf{u} \in \mathcal{X}$ such that for all $\mathbf{v} \in \mathcal{X}$ we have:

$$\int_V A^{-\frac{1}{n}} |D(\mathbf{u})|^{\frac{1}{n}-1} (D(\mathbf{u}), D(\mathbf{v})) dV \tag{11}$$

$$+ \int_{\Gamma_b} c^{-m} |\mathbf{u}|_M^{m-1} (\mathbf{u}, \mathbf{v})_M dS + \rho g \int_V (\nabla s \cdot \mathbf{v}) dV = 0, \tag{12}$$

where the bilinear form $(\mathbf{a}, \mathbf{b})_M := (M\mathbf{a}) \cdot \mathbf{b}$, and its associated norm $|\mathbf{a}|_M := \sqrt{(\mathbf{a}, \mathbf{a})_M}$ have for matrix

$$M = \begin{pmatrix} I + (\nabla_{\mathbf{x}} b)(\nabla_{\mathbf{x}} b)^T & 0 \\ 0 & 0 \end{pmatrix}. \tag{13}$$

The above problem is equivalent to seeking for $\mathbf{u} \in \mathcal{X}$ such that

$$\mathcal{J}(\mathbf{u}) = \min\{\mathcal{J}(\mathbf{v}), \mathbf{v} \in \mathcal{X}\}, \tag{14}$$

where the functional to be minimised is

$$\mathcal{J}(\mathbf{v}) = \int_V 2\frac{A^{-\frac{1}{n}}}{1+\frac{1}{n}} |D(\mathbf{v})|^{1+\frac{1}{n}} dV + \int_{\Gamma_b} \frac{c^{-m}}{1+m} |\mathbf{v}|_M^{1+m} dS$$

$$+ \rho g \int_V (\nabla s \cdot \mathbf{v}) dV. \tag{15}$$

It must be stressed that only the first term still depends on the vertical velocity in both formulations (12) and (15).

First-Order Approximation (FOA)

We introduce the aspect ratio $\epsilon = [h]/[\mathbf{x}]$ of the ice geometry V, where $[h]$ and $[\mathbf{x}]$ denote its typical height and length. It is easy to verify that in that the strain rate tensor $D(\mathbf{v})$ contains terms scaling with ϵ^{-1}, ϵ^0, and ϵ^1. As glaciers are usually thin objects with a small aspect ratio ϵ, it is a common practise to omit the highest order term. By doing so and invoking the incompressibility equation, the vertical velocity components ($\partial_x u_z$ and $\partial_y u_z$) of the strain rate tensor can be eliminated:

$$D(\mathbf{u}) = \tag{16}$$

$$\begin{pmatrix} \partial_x u_x & \frac{1}{2}\left(\partial_y u_x + \partial_x u_y\right), & \frac{1}{2}\left(\partial_z u_x\right) \\ \frac{1}{2}\left(\partial_y u_x + \partial_x u_y\right) & \partial_y u_y & \frac{1}{2}\left(\partial_z u_y\right) \\ \frac{1}{2}\left(\partial_z u_x\right) & \frac{1}{2}\left(\partial_z u_y\right) & -\partial_x u_x - \partial_y u_y \end{pmatrix}.$$

In turn, this eliminates the vertical velocity component u_z from the ice flow model. The resulting model (so-called First-Order Approximation, FOA, or Blatter-Pattyn model [Blatter, 1995]) is obtained by minimising the functional $\mathcal{J}$ defined

in (15) with $D(\mathbf{u})$ defined by (16). Advantageously, the constraints of the functional space $\mathcal{X}$ disappear when removing the vertical component of the velocity. As a result, the FOA model consists of a three-dimensional, non-linear, elliptic, and unconstrained problem, which is therefore simpler than the original Glen-Stokes problem. Provided enough friction at the bedrock (i.e., the coefficient c is not too high) and other suitable assumptions, one can show [Colinge and Rappaz, 1999, Schoof, 2006] that the functional $\mathcal{J}$ is continuous, strictly convex and coercive in the functional space $[W^{1,1+\frac{1}{n}}(V)]^2$, therefore, the FOA problem admits a unique solution.

NUMERICAL MODEL

The glacier evolution model (equipped with both the ice flow solver and the physics-informed deep learning emulator) is implemented in the "Instructed Glacier Model" framework (IGM, `https://github.com/jouvetg/igm`), which can simulate glacier time evolution (on CPU or GPU) given an initial glacier geometry and SMB forcing [Jouvet et al., 2022]. IGM is written in Python and relies on operations of the TensorFlow library to allow vector/parallel operations (such as used in neural networks) between large arrays that are computationally efficient on GPU. At each time step, IGM updates in turn the SMB (e.g., ELA-based parameterisation or climate-driven PDD model), the ice flow (solved or emulated according to the user's choice), and the ice thickness by solving conservation equation (1) using a first-order upwind finite-volume scheme on a regular 2D grid. Thus, the time step is computed adaptively to satisfy the CFL condition, make sure that the ice is never transported over more than one cell distance in one time step, and therefore to ensure numerical stability. We refer to Jouvet et al. [2022] for more details on the transport numerical scheme. In what follows, we focus on the computation of the ice flow by numerical solving and deep-learning emulation.

Spatial discretization

First, the horizontal rectangular domain Ω is discretised with a regular raster/structured grid of size $N_x \times N_y$ with constant cell spacing H in the x and y direction (Fig. 2, right panel). Variables such as the ice thickness h, the surface topography s, the rate factor A, and the sliding coefficient c are defined at the corners of each grid cell of the horizontal grid. In the following, we use subscript H to denote these discrete quantities such as $\mathbf{u}_H$, h_H, s_H, A_H, c_H defined on the horizontal grid. Note that our choice of a structured grid (instead of any other type of discretization) is essential to represent variables as 2D arrays and therefore to use Convolution Neural Networks (CNN) for emulating the ice flow mechanics later on. On the other hand, the ice thickness is discretised vertically using a fixed number of points N_z (in this paper we use $N_z = 10$). Layers are distributed according to a quadratic rule such that discretisation is fine close to the ice-bedrock interface (where the strongest gradients are expected) and coarse close to the ice-surface interface following the strategy given by Khroulev and the PISM Authors [PISM, 2020]. Subsequently, the approximation space X_H for velocities consists of piecewise linear functions defined at the corners of each grid cell in the horizontal direction and at the intersection of each layer in the vertical discretisation.

In finite elements, solving the nonlinear elliptic FOA problem requires minimizing the associated functional $\mathcal{J}$ in a finite-dimension approximation space X_H spanned by shape functions defined in the discretised domain instead of the full continuous solution space X. We follow a similar strategy here: Given $p_H = (h_H, s_H, A_H, c_H)$, we seek for $\mathbf{u}_H \in X_H$ such that

$$\mathbf{u}_H = \text{argmin}\{\mathcal{J}_{p_H}(\mathbf{v}_H), \mathbf{v}_H \in X_H\} \tag{17}$$

where

$$\begin{aligned}\mathcal{J}_{p_H}(\mathbf{v}_H) = \int_\Omega \Bigg(& \frac{2A_H^{-\frac{1}{n}}}{1+\frac{1}{n}} \int_{s_H-h_H}^{s_H} |D_H(\mathbf{v}_H)|^{1+\frac{1}{n}} dz \\ & + \frac{c_H^{-m}}{1+m} |\mathbf{v}_H|_M^{1+m} dS \\ & + \rho g \int_{s_H-h_H}^{s_H} (\nabla s_H \cdot \mathbf{v}_H) dz \Bigg) d\Omega. \end{aligned} \tag{18}$$

For simplicity, D is approximated by a finite difference scheme on a 3D staggered grid (Fig. 2, right panel). As D involves derivatives in the three dimensions, we apply either a finite difference or cell averaging to ensure that all derivatives in (16) are approximated consistently on the same 3D staggered grid (i.e., at the centre of cells horizontally and at the middle of layers vertically). The two other terms (sliding and gravity force related) are also computed on the staggered grid (otherwise, this would cause numerical artefacts, typically chessboard modes). Due to the layer-wise vertical discretisation, we first compute the horizontal derivatives of D_H in a layer-dependent system of coordinate $(x, y, \tilde{z})$ where $\tilde{z} = z - l$ and l is the layer elevation, and transfer them in the reference system of coordinate (x, y, z) using a simple rule of derivative: e.g., $\frac{\partial f}{\partial x} = \frac{\partial \tilde{f}}{\partial x} - \frac{\partial \tilde{f}}{\partial z}\frac{\partial l}{\partial x}$ for any quantity f (resp. $\tilde{f}$) defined in (x, y, z) (resp. $(x, y, \tilde{z})$). Lastly, the integration of (18) is done numerically using the rectangle method. Note that ice margins must be treated carefully

to prevent singular vertical derivatives of D_H as the vertical step size tends to zero. To overcome this issue, we assume a minimum ice thickness of one metre.

Solver

Our solver solves the convex optimisation problem (17) using a stochastic gradient descent method, namely the Adam optimiser [Kingma and Ba, 2014] with a step size of 1. Using the Keras [Chollet et al., 2015] and Tensorflow [Abadi and others, 2015] libraries, the derivatives of $\mathcal{J}_{p_H}$ with respect to $\mathbf{v}_H$ are obtained by automatic differentiation. When used for computing a single snapshot ice-flow, the optimisation scheme is initialised with zero ice velocity. When used multiple times in a transient glacier evolution run, the gradient scheme uses the ice flow from the previous time step as initialisation to predict the next one. In the following, we refer to the "solved" solution (in contrast to the "emulated" solution defined in the next section), the result of the solver at convergence. Note that we found in Appendix B a very good agreement between the "solved" and the reference solutions of the ISMIP-HOM [Pattyn and others, 2008] experiments. This test validates the numerical solver, as well as the implementation of the system energy in IGM, which is used for both the solver and the emulator.

Emulator

As an alternative to the previously-introduced solver, we now propose an ice flow emulator, which predicts horizontal ice flow $(\mathbf{u}_H, \mathbf{v}_H)$ from the input field p_H.

$$\begin{array}{rccc} \mathcal{N}_\lambda : & \{h_H, s_H, A_H, c_H, H_H\} & \longrightarrow & \{\mathbf{u}_H, \mathbf{v}_H\} \\ & \mathbb{R}^{N_X \times N_Y \times 5} & \longrightarrow & \mathbb{R}^{N_X \times N_Y \times N_Z \times 2} \end{array} \tag{19}$$

where input and output can be seen as two- and three-dimensional multichannel fields, which are defined on the regular horizontal grid (Fig. 3). Having selected these input parameters allows us to develop a generic ice flow emulator that can handle a large variety of glacier shapes, types of ice flow (from shearing to sliding dominant), and spatial resolutions. As the spatial resolution H_H is fixed in modelling application, including it as an input of the emulator is in fact not necessary. Here we added H_H for convenience such that one can take advantage of an initial pre-trained emulator (Appendix A) irrespective of the spatial resolution.

As an emulator, we choose an Artificial Neural Network (ANN), which maps input to output variables by a sequential composition of linear and nonlinear functions (or a sequence of network layers). Linear operations have weights $\lambda = \{\lambda_i, i = 1, ..., N\}$, which are optimised in the training stage. Here, we use a Convolutional Neural Network [CNN; Long et al., 2015], which is a special type of ANN that additionally includes local convolution operations to learn spatially variable relationships [LeCun et al., 2015] and proved to be capable of learning high-order ice flow models [Jouvet et al., 2022]. Here we retain the hyper-parameters found by Jouvet et al. [2022] as they provide a good trade-off between model fidelity and complexity: our CNN consists of 16 two-dimensional convolutional layers between input and output data (Fig. 3). Convolutional operations have a kernel matrix (or feature map) of size 3×3. A padding is used to conserve the frame size through the convolution operation. Convolutional operations are repeated using a sliding window with one stride across the input frame and 32 feature maps. As a non-linear activation function, we use leaky Rectified Linear Units. As a result, our CNN has about 140'000 trainable parameters.

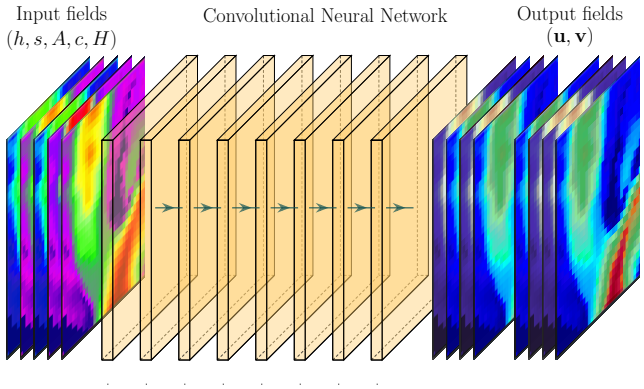

Fig. 3. Our emulator consists of a CNN that maps geometrical (thickness and surface topography), ice flow parameters (shearing and basal sliding), and spatial resolution inputs to 3D ice flow fields.

While Jouvet et al. [2022] proposed to train (19) by fitting to external ice-flow model realizations, we take here another strategy inspired from Physics-Informed Neural Networks (PINNs). We differ from traditional PINNs in two ways: first PINNs usually map the coordinate of the sampling points to the physical output, which forces them to retrain the network for different settings, while our inputs are essential model parameters (the coordinates at each pixel are not explicitly passed). Second, PINNs usually minimise the residual of the equation and/or boundary conditions involved in the physical model [e.g., Markidis, 2021]. Instead, we adopt the different variational PINN strategy [Kharazmi et al., 2019] by minimising the energy associated with the FOA model instead of the residual (Fig. 1). In more detail, the training consists of finding the weights of the CNN $\lambda = \{\lambda_i, i = 1, ..., N\}$

that minimise:

$$\lambda = \text{argmin}\left(\mathcal{J}_{p_H}(\mathcal{N}_\lambda(p_H))\right), \tag{20}$$

given geometrical and glaciological input data p_H. The optimisation problem (20) is solved again using the Adam optimiser [Kingma and Ba, 2014] – the derivatives of $\mathcal{J}_H$ with respect to λ being obtained from automatic differentiation. At first view, the minimisation problem (20) is expected to be more difficult to solve than that in (17), as there is no guarantee that $\mathcal{J}_{p_H}$ is convex with respect to the training parameters λ. On the other hand, problem (20) is expected to have much fewer control parameters (the number of training parameters is on the order of 10^5) than problem (17), which may have much more control parameters ($2 \times N_z \times N_y \times N_x$) when treating a large scale array.

While there exist different strategies for initializing the weights of CNNs, we found in Appendix A that using a CNN pretrained over a large glacier catalogue (Fig. 12) facilitates the convergence of the emulator, presumably because the diversity of the catalogue prevents against falling into local minima. The optimisation of the CNN is therefore always initialised with pre-trained weights (Appendix A). When used for computing a single snapshot ice-flow, we use an adaptive learning strategy including an exponential decay to launch the training aggressively ($\sim 10^{-4}$) for efficiency and to end it gently for fine-tuning ($\sim 10^{-5}$). When used in a transient glacier evolution run, one performs a single step of gentle ($\sim 10^{-5}$) training per iteration (or each X iteration to vary the degree of training) starting from the lastly-trained emulator.

RESULTS

In this section, we present fidelity and computational performance results of the "emulated" solution toward the "solved" solution simulations with different strengths of emulator training. For that purpose, we consider two glaciers of different sizes i) the present-day Aletsch Glacier, Switzerland, which the current largest glacier of the European Alps ii) the former Valais Glacier, Switzerland, which covered a large part of Switzerland during the last glacial maximum. The experiments for these two glaciers cover different applications from individual glaciers on a small grid (244x179 at 100 m of resolution for Aletsch) relevant for the modelling of today's glaciers to large ice fields on large grid (700x700 at 200 m of resolution for Valais) more relevant for paleo glacier modelling. We conduct two kind of experiments in turn: i) the computation of a snapshot solution to assess the best accuracy we should expect from the emulator without consideration for the computational price, and ii) the computation of a transient solution to assess both fidelity and computational performance in a modelling application.

Fidelity of snapshot solutions

First we consider the topography and ice thickness of Aletsch and Valais glaciers at a given time and fix the ice flow parameters (A, c) to constant physical values ($A = 78$ MPa^{-3} a^{-1}, $c = 10$ km MPa^{-3}). Based on these geometries, we computed two numerical solutions: i) a "solved" one $\mathbf{u}_H$ obtained by minimizing (18) within the space of solutions X_h, ii) an "emulated" one $\mathcal{N}_\lambda(p_H)$ obtained by minimizing (20) in the space of parameters of the CNN. Figures 4 and 5 present the results in terms of "solved" solution at convergence (panel A), "emulated" solution at emulator convergence (panel B), difference between "solved" and "emulated" solutions (panel C), decrease of the system energy during solving and training (panels D and E), and L_1 error of the emulated $\mathbf{u}_E$ toward the solved $\mathbf{u}_S$ solution through training iterations (panel F) defined by:

$$E_{L_1} = \int_\Omega \int_b^{b+h} |\mathbf{u}_E - \mathbf{u}_S|. \tag{21}$$

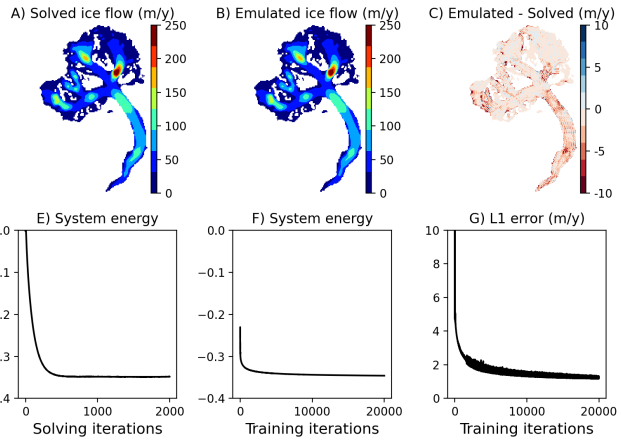

Fig. 4. Results of the solver and the emulator for the snapshot experiment related to Aletsch glacier. Panels A and B show the magnitude of the ice flow velocities obtained by solving and emulation, respectively. Panel C shows the difference between the two. Panels D and E show the decrease of the system energy through iterations. Panel F shows L_1 error of the emulated toward the solved solution through training iterations.

In general, the Adam optimiser succeeds at minimising the energy and capturing the "solved" and the "emulated" solutions. Indeed, the energy associated with the "emulated"

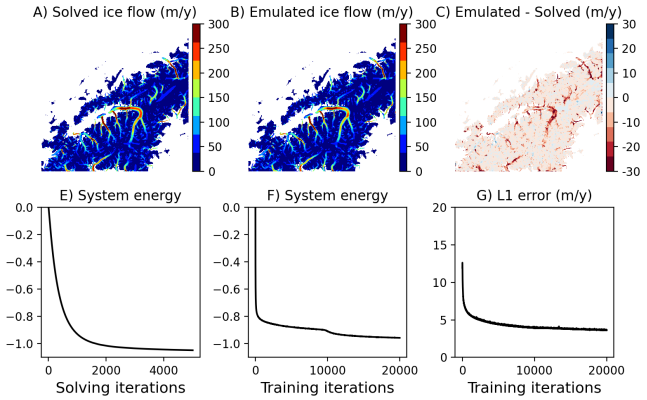

Fig. 5. Results of the solver and the emulator for the snapshot experiment related to Valais glacier. The meaning of panels is similar to Fig. 4.

solution decreases towards a value (~ -0.34) that is relatively close to the value obtained when solving (~ -0.35), demonstrating that our CNN has learnt well to minimise the energy for the Aletsch case (Fig. 4 E). Most importantly, the L1 error is small (~1.2 m/y, Fig. 4 F), showing that the "emulated" solution is very similar to the reference "solved". An in-depth comparison of the spatial pattern of the two solutions (Fig. 4 C) reveals minor and unevenly distributed discrepancies. Interestingly, the Valais glacier case shows a larger energy gap (~ 10%, Fig. 5 E), and a larger L1 error norm (~3.6 m/y, Fig. 5 F), which remains small considering the velocity scale (0-300 m/y). Comparing the spatial pattern of the two solutions (Fig. 5 C) shows that the error is mostly concentrated on the most prominent glacier tongue. The slight deterioration of the accuracy from the Aletsch to the Valais case can be explained as follows: An emulator trained to a single glacier in a small region is naturally expected to be more accurate (as more customized) than an emulator of the same complexity trained to a larger glacier network. Note that increasing the size of the CNN (increasing the number of layers, feature maps, or kernel size) have shown slight but not significant improvements of fidelity.

For computing a single snapshot ice-flow field at a given time, the solver was found to be more efficient than the emulator in terms of convergence and then in terms of computational performance, presumably due to different convexity properties. In the next section, we show that the opposite is true when we consider the evolution of a glacier over time.

Fidelity of transient solutions

For each glacier (Aletsch and Valais), we now perform two kinds of transient experiments: i) the first (referred to as "ELA-varying") assumes fixed ice flow parameters (A and c), and forces the Surface Mass Balance (SMB) with time-varying Equilibrium Line Altitudes (ELA) ii) ; the second (referred as "A/c-varying") assumes fixed ELA and force time-varying ice flow parameters (A and c). The goal of these two experiments is to test the memory capacity of the deep-learning emulator. As SMB, we use a simple parameterisation based on a given ELA z_{ELA}, vertical gradients of accumulation and ablation, and maximum accumulation rate:

$$SMB(z) = \begin{cases} \min(0.003 \times (z - z_{ELA}), 1), & \text{if } z \geq z_{ELA} \\ 0.006 \times (z - z_{ELA}), & \text{otherwise.} \end{cases}$$

Prior to running experiments, we collected the bedrock topography of the two regions [Grab, 2020], initialised the model with ice-free conditions and ran it with ice flow parameters $c = 10$ km MPa^{-3} a^{-1} and $A = 78$ MPa^{-3} a^{-1} and mass balance parameters $z_{ELA} = 2800$ m asl, and $z_{ELA} = 2200$ m asl for Aletsch and Valais, respectively. The goal of this preliminary phase is to simulate the build-up of glaciers until they reach a steady state shape. Then, the ELA-varying transient experiment consists of modelling 2000 years (starting from the obtained steady-state shape, and keeping the parameters constant) with the following ELA parametrisation:

$$z_{ELA} = 2800 + 200 \times \sin(\pi t/500) \text{ m},$$
$$z_{ELA} = 2200 + 300 \times \sin(\pi t/500) \text{ m},$$

for the Aletsch and Valais glaciers, respectively. On the other hand, the A/c-varying transient experiment consists of running the model for 2000 years (starting from the obtained steady-state shape and keeping the parameters constant) with the following ice flow parameters:

$$A = 78 + 22 \times \sin(\pi t/500) \text{ MPa}^{-3}\text{a}^{-1},$$
$$c = 10 + 5 \times \sin(\pi t/500) \text{ km MPa}^{-3}\text{a}^{-1},$$

to induce glacier variations (retreat-advance-retreat), and explore a variety of configurations for assessment.

The experiments were performed using the solver (our reference run) and the emulator (pretrained on a glacier catalogue, Appendix A) with different retraining strategies to compute the ice dynamics: i) with no retraining at all (Experiment "0%-0%"), ii) with 100% retraining during the first 1000 years (i.e. one step of retraining per iteration) and then no retraining (Experiment "100%-0%"), iii) with 100% retraining during the first 1000 years and then 10% retraining (i.e. on step of retraining each 10 iteration) during the second 1000

years (Experiment "100%-10%"). Table 1 summaries the design and the outcomes in terms of fidelity of all experiments. Figures 6 and 7 show the results of the ELA and A/c-varying experiments for Aletsch and Valais Glacier, respectively, in terms of fidelity (L1 error) of the "emulated" solution $\mathbf{u}_E$ to the reference "solved" one $\mathbf{u}_S$, and overall ice volume.

Exp.	Training level		Resulting fidelity	
name	[0,1] ky	[1,2] ky	[0,1] ky	[1,2] ky
0%-0%	0%	0%	Med./Low	Med./Low
100%-0%	100%	0%	High	Med./Low
100%-10%	100%	10%	High	High

Table 1. Design and results of ELA and A/c-varying experiments with various retraining strategies.

As a result, the pretrained emulator without further retraining (0%-0%) captures roughly the ice-flow in the ELA-varying and A/c-varying experiments of Aletsch Glacier when ice flow parameters are fixed with an L1 error of ∼5 m/y (Fig. 6), which is fairly small compared to the velocity scale (0-200 m/y). This shows that the shape of the Aletsch Glacier is relatively well represented in the pretraining glacier catalogue (Fig. 12). Therefore, the emulator has acquired a fair knowledge to predict a solution in line with the "solved" one. However, emulator-induced cumulative errors lead to an increasing bias in ice volume (Fig. 6). In contrast, the pretrained emulator performs very poorly with the Valais Glacier (very high L1 error in Fig. 7). This is likely due to the fact that the glaciers of this experiment go well beyond the glaciers in the catalogue (Fig. 12) in terms of shape, size, and ice flow behaviour.

In contrast, our results reveal that adaptive retraining of the emulator (100%-0%) shows largely improved accuracy with respect to the "solved" reference solution during the first 1000 years. Indeed, retraining damps the L1 error to small values: ∼1 m/y and ∼4 m/y in the Aletsch and the Valais Glacier experiments, respectively (Fig. 6 and 7) in the first 1000 years when one retraining step is applied to each time step. These errors as well as the spatial patterns of the error (not shown) are very similar to the ones found in the snapshot experiments (Figs. 4 and 5, panel A and B) with discrepancies, mostly in the trunk of Valais Glacier. As a result, the modelled volumes agree very well with the "solved" solution when systematic retraining is used (Figs. 6 and 7). It must be stressed that using more than one training iteration per time step did not show significant reduction of the L1 error.

As systematic online retraining during the first 1000 years is a relatively costly task (next section), we analyse the effect of releasing the retraining to assess the capability of the emulator to retain the ice flow solutions accurately (Fig. 6 and 7). As a result, switching off the retraining after 1000 years of simulation and repeating the experiments with the same forcing for another 1000 years (100%-0%) reveal different outcomes. Indeed, the emulator "retains" some of the relevant training in ELA-varying experiments, but deteriorates very quickly in the A/c-varying experiments, leading to notable biases in ice volume (Figs. 6 and 7). In contrast, the emulator remains as accurate as in the first phase when lightly retrained each 10 time steps (100%-10%) in the second phase. This means that the emulator has mostly retained the geometry-ice flow relationship during the first pass and that the accuracy can be maintained with a light computationally effective retraining provided an initial systematic training.

An important parameter for online retraining is the learning rate. A too low parameter (gently learning) will result in inefficient learning and solution biases, while a too high parameter (aggressive learning) will result in erratic/non-smooth accuracy curve and deteriorated memory of the emulator (not shown). As a trade-off between the two cases, we found that a learning rate of 2×10^{-5} is optimal in all our transient experiments.

Computational performance of transient solutions

We now compare the computational performance of the 3 solutions: "solved", "emulated without online retraining" and "emulated with online retraining" to lead the ELA and A/c-varying experiments presented in the previous section. Comparing the emulator and the solver is a challenge, as the first requires only one emulation step (the retraining does not require to be performed more than once per time iteration), while the solver may require several iterations per time step to converge. For this reason, we first discuss the costs associated with one individual step (i.e., one iteration of retraining or solving of the optimization algorithm) before analysing the overall costs.

Table 2 gathers together the computational times needed to achieve one step of i) solving, ii) emulating, and iii) retraining for modelling domains of various sizes, and on both CPU and GPU architectures of the same desktop computer (equipped with a 10-core Intel CPU i9-10900K and a 10'000 cores Nvidia GPU RTX 3090). As a result, the GPU (which has 1000 times more cores) systematically out-performs the CPU. While the CPU may be interesting for small-scale array domains, Table 2 shows that it is not a viable option to treat large-scale arrays. Therefore, we focus our performance analysis on the GPU only. We find that the emulation step is the most affordable task, followed by the solv-

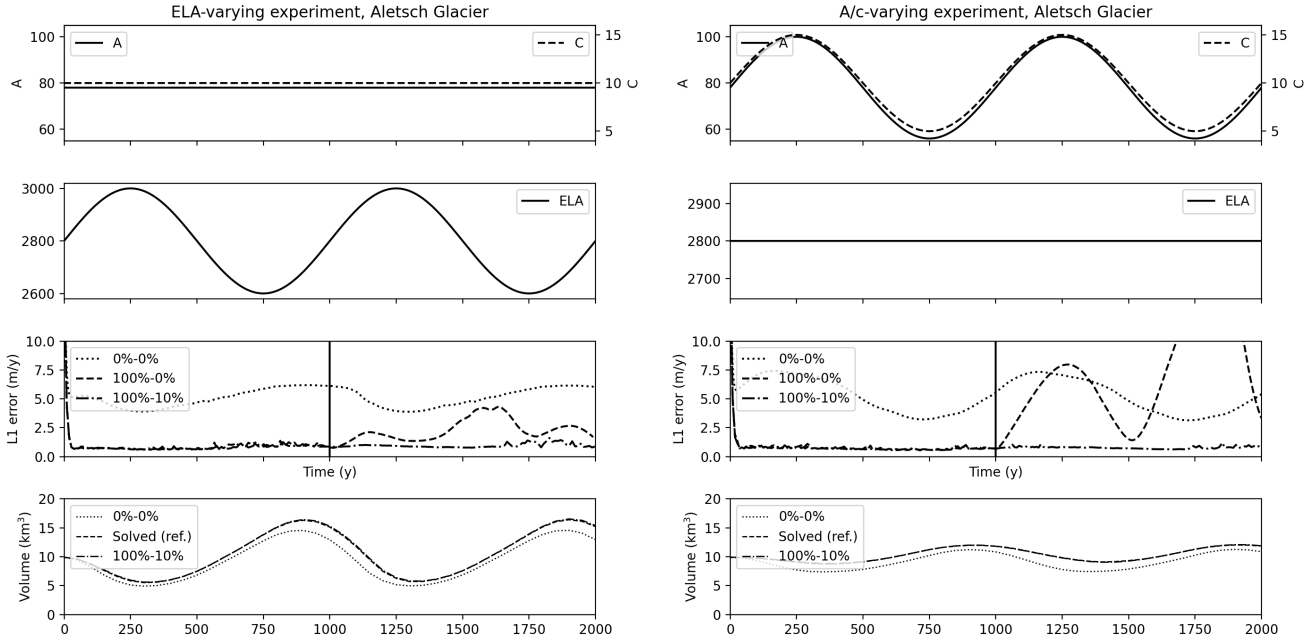

Fig. 6. Transient results of the ELA-varying (left panels) and A/c-varying (right panels) transient modelling experiments for Aletsch Glacier. The panels indicate the time evolution of input parameters (ice flow parameters and ELA), the resulting ice flow L1 error between all "emulated" solutions (with and without retraining) and the "solved" one, and the output ice volume obtained with the three modelling methods ("solved", "emulated" with and without retraining).

ing step, which is slightly (about 30%) more expensive, and the retraining step, which is about 3 times more expensive than emulation regardless of the domain size. This can be explained as follows. The emulation step is inexpensive as it only requires a single pass of the CNN. On the other hand, the solving step consists of a forward evaluation of the system energy followed by the computation of the energy gradients and an update of the ice flow. Last, the retraining step is naturally expected to be more costly than the "emulation + solving", as it combines the tasks of the two: one CNN evaluation, one system energy evaluation, the computation of the two gradients and an update of the weights of the CNN.

Since a CNN is evaluated sequentially layer by layer, the emulation step is memory efficient. Therefore, emulation step can be performed on large arrays (i.e. we achieved 2400x4000 with our 24 Gb GPU, Table 2), while the solving and retraining steps are more memory-demanding and therefore more limited by the GPU available memory. For example, none of the solving and retraining steps for the 2400x4000 domain were achievable with our GPU (we found that a maximum grid of about 2000x2000). Hopefully, this limitation can be overcome for the retraining (and not for the solving step, Table 2) by splitting the domain into smaller patches and sequentially retraining the emulator patch-wise.

Exp	Step	CPU	GPU
Aletsch	solver	125 ms	15 ms
244x179	emulator	39 ms	11 ms
	retrain	533 ms	29 ms
Valais	solver	1538 ms	51 ms
700x700	emulator	468 ms	38 ms
	retrain	5592 ms	110 ms
Entire Alps	solver	X	X
2400x4000	emulator	X	360 ms
	retrain	X	1465 ms

Table 2. Computational time required (in average) to perform one emulation, retraining, solving iteration step in modelling experiments for Aletsch, Valais, and the entire Alps. In the latter case, we reported "X" when the computation was not possible, or prohibitively too expensive. The CPU (i9-10900K) has 10 3.70 GHz cores with 64 Gb RAM while the GPU (RTX 3090) has about 10'000 1.70 GHz cores with 24 Gb RAM.

As the other modules (ice thickness and mass balance

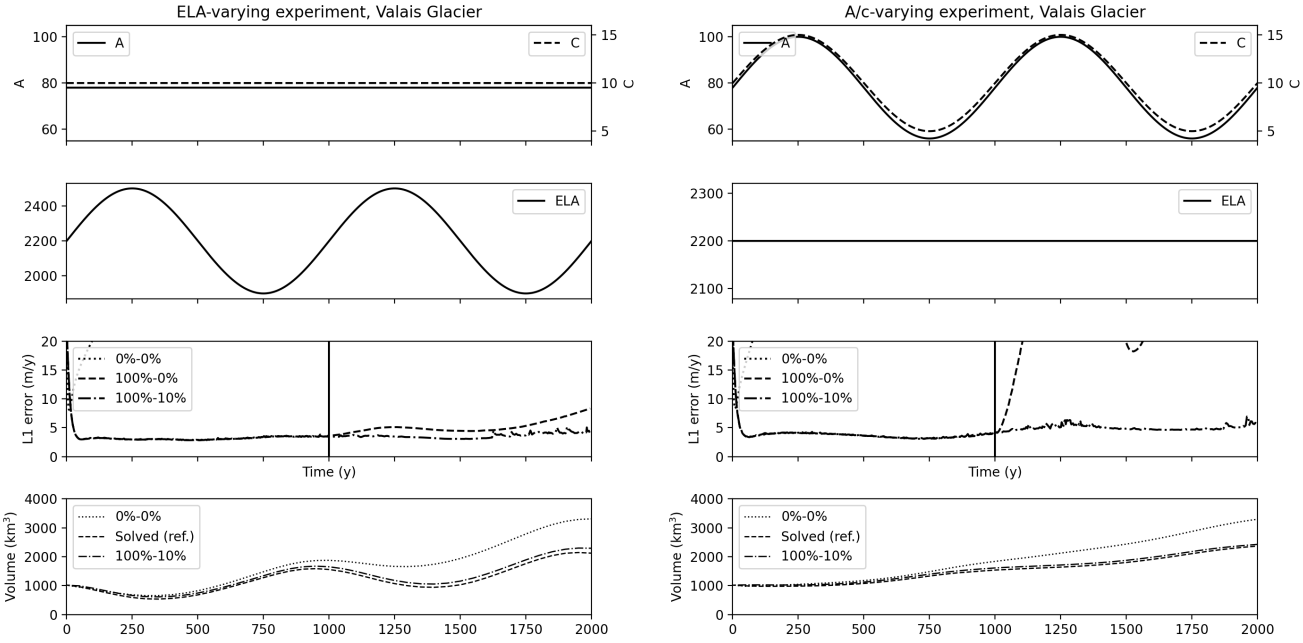

Fig. 7. Transient results of the ELA-varying (left panels) and A/c-varying (right panels) transient modelling experiments for Valais Glacier. This is similar to the caption of Figure 6.

updates) are computationally inexpensive compared to the ice flow model, the overall cost is mainly the number of time iterations times the costs of individual emulation (with or without retraining) or solver steps. In the ELA and A/c-varying experiments related to Aletsch and Valais glaciers, the time step was on the order of 0.1 y to maintain numerical stability meaning that ∼10'000 time iterations were needed per millennium of modelling irrespective of the chosen method (solver or emulation). When using the solver, several iterations were required to reach convergence at a given time step, however, this number is case-dependent: ∼3-4 and more than 10 iterations in the case of Aletsch and Valais glaciers, respectively. These numbers should be taken with care, as a more efficient optimizer (e.g. Newton-like) may reduce the number of required iterations. In contrast with the solver, the emulation only requires one step, while the retraining can be applied infrequently while remaining effective. In our case, the best trade-off in terms of accuracy to computational performance was found using light retraining (each 10 iterations) as it maintained accuracy (Figs. 6 and 7) at the cost of one cheap emulation per time step plus more expensive but infrequent retraining steps (Table 2).

APPLICATIONS

In this section, we illustrate the potential of our physics-informed ice-flow emulator for glaciological applications.

Paleo glacier modelling in the European Alps

Modelling paleo-glacier evolution is an important tool for understanding the history of glaciations. However, the long time scales and the size of the domain may render this exercise computationally very demanding. For example, the 120'000-year-long simulation of alpine glacier evolution in the Alps of Jouvet et al. [2023] at 2 km with the Parallel Ice Sheet Model [PISM, Khroulev and the PISM Authors, 2020] would take several weeks of computational time on a 10 core i9-10900K running at 3.70 GHz. It is, therefore, prohibitively expensive to explore subkilometre resolutions that would be required to resolve the complex topography of the Alps in the highest reaches. Therefore, the ice flow emulator with online retraining is a promising approach to overcome the computational bottleneck, especially on GPU, which allows large array computations. Here, we test its capability to simulate the paleo evolution of glaciers in the entire European Alps in very high resolution (200 m) over 10'000 years encompassing the Last Glacial Maximum (LGM, about 24'000

years ago).

To this end, we took over the model setting of Jouvet et al. [2023]. Initialising with ice-free conditions and today's topography of the Alps as bedrock, IGM was forced with a coupled modelled paleoclimate data and PDD surface mass balance model [Hock, 1999] from 28'000 years BP to 18'000 years BP. As a result, the 200 m IGM simulation at 21'000 years BP shows highly detailed glacier extents resolving small valleys and Nunataks (Fig. 8), and took about 2 days of computations on a ~10'000-core RTX 3090 1.70 Ghz GPU. Here, the GPU has 24 GB memory, which is key to treating very large arrays. The horizontal grid covers the entire Alps at 200 meters yielding a resolution of 2400x4000. This exercise illustrates the capability of our approach to achieving very high resolutions at affordable computational costs. For comparison, PISM at a much lower resolution (2km resolution, 240x400) would take about the same time to carry a similar simulation on a 10-core 3.70 GHz CPU. Of course, this comparison must be tempered by the fact that IGM does not include all the many physical components of PISM, especially the thermodynamics of ice, which is known to add substantial computational time.

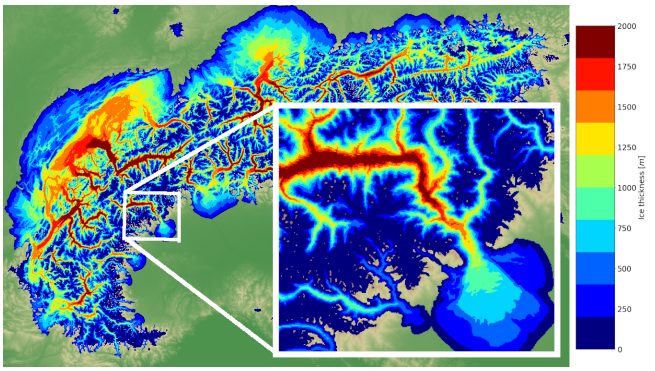

Fig. 8. Ice thickness of the alpine ice field obtained at 21'000 years BP modelled with IGM at 200 meters of resolution.

Ice flow model inversion/data assimilation

Inverse modelling is an essential step to initialise present-day glacier models, i.e., estimate unknown variables (such as ice thickness and/or ice flow parameters) such that the model matches at best observations (surface ice flow velocities or pointwise ice thickness profiles). Substituting the ice flow equations with a CNN emulator allows solving the inverse model (or the underlying optimisation problem) very efficiently by utilising automatic differentiation and stochastic gradient methods [Jouvet, 2023]. Therefore, the CNN emulator trained by physics-informed deep learning can also be used in a similar way. Most importantly, one can now simultaneously optimise the CNN parameters to fit the ice physics by minimising the system energy and the CNN inputs to match observations by minimising the misfit to the data. The coupled optimisation allows to perform the inversion with an accurate and customised-to-the-glacier CNN at the same time.

As an illustration, we solve the inversion problem for Aletsch Glacier proposed by Jouvet [2023] with this new strategy. Given present-day pointwise ice thickness measurements and surface ice velocity measurements, we use the CNN trained offline over the glacier catalogue, and seek alternatively for the CNN weights λ, the ice thickness distribution h and the distributed sliding parameter c, such that both the system energy (Eq. (20)) and the mismatch between the observed and modelled quantities (Eq. (5) in Jouvet [2023]) are minimised. Note that the regularisation terms for h and c are added to enforce smoothness and ensure a unique solution. As a result, Fig. 9 shows the convergence of the fields towards an optimal state and the reduction of the corresponding misfit values in terms of STandard Deviations (STD). Here, the quality of data assimilation is comparable to that obtained by Jouvet [2023]. However, the simultaneous emulator training/optimisation has a major benefit with respect to the former method (based on offline training): the online retraining permits to account for spatial variations of the sliding coefficient (Fig. 9, top-right panel) and makes the emulator nearly as accurate as the solver (Fig. 10). In contrast, the former emulator, which met only the glacier catalogue and spatially constant sliding coefficient at training, suffers from larger biases as observed in Appendix A.

Ice shelf

Ice shelves behave very differently to mountain glacier ice flow as modelled in the two previous applications. Indeed, they can be very fast due to the absence of friction under floating ice, and are therefore dominated by basal sliding. By contrast, friction under grounded glaciers usually induces an important vertical shearing component. Yet, modelling accurately the dynamics of ice shelves is essential to predict the evolution of the Antarctic ice sheet under climate change and the resulting sea level rise [Seroussi et al., 2020]. Here we demonstrate that IGM equipped with the new physics-informed deep-learning emulator has an important potential for modelling ice sheet/shelf systems by performing a simple experiment inspired by the Marine Ice Sheet Model Inter-comparison Project [MISMIP Pattyn et al., 2012]. The goal here is not to run all exercise simulations, but only to

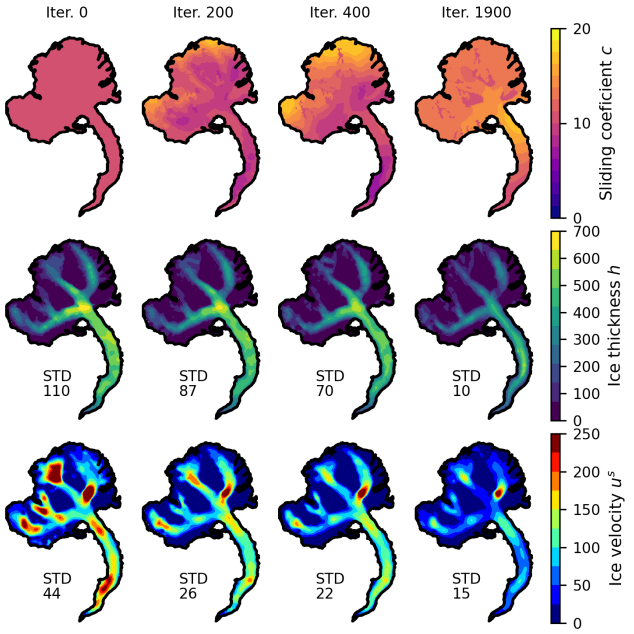

Fig. 9. Evolution of the sliding distribution c (unit: km MPa^{-3} a^{-1}), the ice thickness distribution h (unit: m), as well as resulting surface ice flow velocity field $\mathbf{u}^s$ (unit: m y^{-1}) through the iterations of the optimisation problem for Aletsch glacier. The STandard Deviation (STD) between the modeled and observed fields is reported at each step.

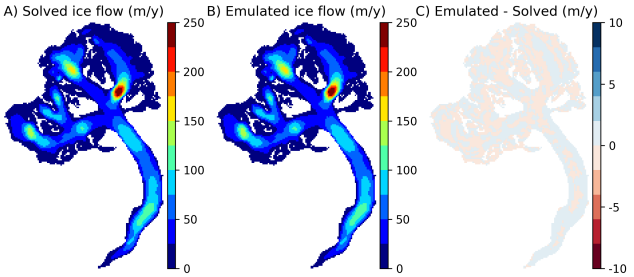

Fig. 10. Surface ice flow field of Aletsch Glacier with the parameters found after performing the simultaneous inversion and emulator training: A) using the solver B) using the retrained emulator. Panel C) shows the spatial difference between the two.

compute the ice dynamics associated with one state to prove the capacity of the emulator to handle sliding-dominant ice flow of ice shelves.

For that purpose, we consider an idealized ice sheet-shelf geometry lying on a ramp of constant slope in the x-direction over a distance of $L_x = 1100$ km (Fig. 11). All geometrical variables are constant in the y-direction to mimic the 2D MISMIP experiment 1 [Pattyn et al., 2012]. In that configuration, we distinguish the ice sheet ($x < x_{GL}$) and the ice shelf ($x > x_{GL}$) from the grounding location $x_{GL} \sim 966.5$ km (Fig. 11). The lower surface elevation l is either the bedrock when the ice is grounded or determined by Archimedes's principle when the ice is floating: $l = \max\{b, -(\rho_i/\rho_w)h\}$, where $\rho_i = 910$ kg m^{-3} and $\rho_w = 1000$ kg m^{-3} denote the densities of ice and water, respectively. Here, we use the following parameters: $A = 146.5$ MPa^{-3} a^{-1}, $m = 1/3$, $c = 71.2$ km MPa^{-3} a^{-1} where the ice is grounded and $c^{-1} = 0$ km MPa^{-3} a^{-1} where the ice is floating (no friction). In addition, we use the "Shallow Shelf Approximation" (SSA) model [Morland, 1987] instead of the FOA by simply setting a single layer in the vertical discretization (Fig. 2, right panel), which is equivalent to assuming vertically-constant ice flow velocities. Lastly, the function $\mathcal{J}$ defined by (15) is augmented with an additional term to account for balance stress conditions between ice and water columns at the Calving Front (CF) on the extreme right of the modelled domain (Fig. 11):

$$-\int_{CF} \frac{1}{2}\left(1 - \frac{\rho_i}{\rho_w}\right)\rho_i g h^2 v \cdot \mathbf{n}, \tag{22}$$

where $\mathbf{n}$ is an outer normal vector along CF [Schoof, 2006]. The above condition was implemented along the other terms of the system energy, and a 2D field (namely (22) along the calving front and zero elsewhere) was added to the emulator inputs (Eq. (19)) to control this boundary condition.

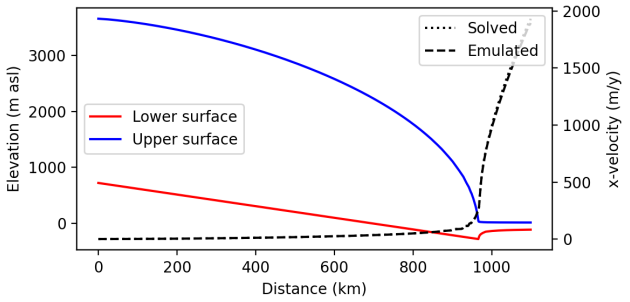

Fig. 11. MISMIP-inspired ice geometry of the ice shelf experiment along the x-axis, and resulting ice flow velocities modelled from the solver and the emulator with custom training on the specific geometry.

As a result, we find that after training the emulator on the specific geometry, the "Solved" and "Emulated" ice flow fields along the x-axis are nearly identical (Fig. 11). This experiment demonstrates that the approach of the paper is

not limited to grounded glacier flow, but is capable to handle the sliding-dominant flow of ice shelves.

DISCUSSION AND CONCLUSIONS

In this paper, we have introduced both a solver and a physics-informed deep learning emulator for modelling high-order ice flow on a regular grid that are designed to run efficiently on GPU. The solver relies on a stochastic gradient method and automatic differentiation tools to efficiently minimise the energy associated with the underlying ice-flow equations discretized by finite differences, similarly to Ritz-Galerkin methods in the finite element framework. On the other hand, the emulator relies on a CNN, which is trained to minimise the same energy. Therefore, our method (which belongs to the category of Deep-Ritz) can be seen as a fusion of finite element and deep learning approaches. Here, our approximation space for the ice flow is induced by the training parameters of our CNN instead of being spanned by finite element basis functions. As a result, we have shown that our emulator can reproduce the solutions of the solver with high fidelity. Unlike the former emulator [Jouvet et al., 2022], the new one does not require any data from an external ice flow model, as it enforces the ice flow physics directly in learning. Here, we used a glacier catalogue to pre-train the emulator and obtain a good initial guess that facilitates convergence. However, adaptive online training within the time-stepping of a glacier evolution model does not require any data and has proven to significantly improve the emulator accuracy. This strategy makes the new emulator generic, as it allows exploration of any parameters, types of ice flow, spatial resolutions, and glacier shapes, while the validity of the former emulator could not be ensured beyond the "hull" defined by the data and its associated spatial resolution used for training. In addition, CNN training is therefore significantly easier and cheaper as no data is required. Last, our new emulator models the full 3D ice flow field (instead of the vertical average horizontal speeds with the former version), which can be advantageous for some applications (e.g., Lagrangian 3D particle tracking).

The computational benefits of using a CNN emulator [Jouvet et al., 2022] remain unchanged. Indeed, one CNN forward evaluation can be done very efficiently, especially on GPU. In contrast, the solving and training steps are computationally more expensive (by a factor of 3 in our experiments). Therefore, to obtain the best computational performances, we mitigate the amount of training by limiting the frequency of retraining. Indeed, the memory capability of the CNN revealed in our experiments allows us to reduce the training costs for a given application. For instance, we found that a light cost-effective online retraining following a first systematic training is sufficient to maintain accuracy, as the CNN conserves most of the previously learnt solutions. Therefore, training costs can be strongly reduced in some modelling applications that meet several times similar glacier configurations (e.g., in paleo glacier modelling with repeated glacial cycle, or in parameter sensitivity analysis), yielding low overall computational costs.

There are a number of aspects that may be improved in the method presented in this paper. First, we used here the simplest finite-difference scheme to discretise the spatial derivatives in the strain rate on a staggered grid for simplicity. A more elaborated finite-element-like discretization is expected to yield a more accurate solution, possibly slightly increasing the training costs but without affecting the emulation costs. Second, we used here the Adam optimiser as it proved to be robust and simple to implement, however, other optimisers may improve the convergence. For example, the (deterministic) L-BFGS-B optimiser has proven to be efficient at fine-optimising physics informed neural networks after an initial coarse pass with Adam to avoid local minima [Taylor et al., 2022]. Similarly, the solver can be improved, and hybrid solver/emulation strategies that take advantage of two should be further investigated (e.g., using the emulator for preconditioning purpose or to help finding an initial guess). Third, here we investigated retraining strategies (to get the best accuracy while minimizing the amount of retraining) in an empirical way by quantifying *a posteriori* the error between the emulated and solved solutions. Future research should investigate more effective and generic retraining strategies, e.g., seeking for an *a priory* error estimate of the neural network approximation [e.g. Minakowski and Richter, 2023] as done FEM for estimating the numerical error [Ern and Guermond, 2004]. Lastly, the loss of accuracy with increasing domain size is another aspect of the emulator that should be improved, e.g., by using multiple region-specific emulators. It must be stressed that our CNN emulator (computationally-efficient on GPU) strongly relies on the structured discretisation grid assumption. Therefore, emulating ice flow on more complex mesh (e.g. with local refinements) would require to follow a different strategy (e.g., PINNs).

Our modelling experiments have shown that the new emulator embedded in a glacier evolution model can handle very efficiently large-scale and/or high-resolution domain arrays and/or very long time scales. Therefore, our method has a high potential for paleo-glacier simulations. Additionally, we found that the emulator is suitable for both inverse and forward modelling. Therefore, the method can be very beneficial to assimilate data and run prognostic models of

present-day glaciers on a global scale. Lastly, we have shown that our approach can be extended to fast-flowing ice as found in tidewater glaciers, opening promising perspectives for modelling the Antarctica and Greenland ice sheets in high spatial resolution. The code to run any solver-based or emulator-based glacier evolution simulations is open-source, relatively simple and publicly available with the "Instructed Glacier Model" (IGM, `https://github.com/jouvetg/igm`).

ACKNOWLEDGEMENTS

We acknowledge Jacob Downs and an anonymous referee for their valuable comments on the original manuscript. This project was sponsored by the Agence Nationale de la Recherche project Invterra ANR-22-CE33-0012-01 to Cordonnier.

AUTHOR CONTRIBUTIONS

Guillauem Jouvet conceived the study, wrote the code, performed the simulations, and wrote the article. Guillaume Cordonnier developed simultaneously a similar approach, provided valuable feedback on the method and the results, and helped to improve the manuscript.

REFERENCES

M. Abadi and . others. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL `https://www.tensorflow.org/`. Software available from tensorflow.org.

R. Adams and J. Fournier. *Sobolev Spaces*. Pure and Applied Mathematics. Elsevier Science, 2003.

H. Blatter. Velocity and stress fields in grounded glaciers: a simple algorithm for including deviatoric stress gradients. *Journal of Glaciology*, 41(138):333–344, 1995. doi: 10.3189/S002214300001621X.

C. F. Brædstrup, A. Damsgaard, and D. L. Egholm. Ice-sheet modelling accelerated by graphics cards. *Computers & Geosciences*, 72:210–220, 2014. doi: 10.1016/j.cageo.2014.07.019.

D. Brinkerhoff, A. Aschwanden, and M. Fahnestock. Constraining subglacial processes from surface velocity observations using surrogate-based bayesian inference. *Journal of Glaciology*, pages 1–19, 2021. doi: 10.1017/jog.2020.112.

F. Chollet et al. Keras, 2015. URL `https://github.com/fchollet/keras`.

J. Colinge and J. Rappaz. A strongly nonlinear problem arising in glaciology. *M2AN Mathematical Modelling and Numerical Analysis*, 33(2):395–406, 1999. doi: 10.1051/m2an:1999122.

G. Cordonnier, G. Jouvet, A. Peytavie, J. Braun, M.-P. Cani, B. Benes, E. Galin, E. Guérin, and J. Gain. Forming terrains by glacial erosion. *ACM Transactions on Graphics (SIGGRAPH Conference Proceedings)*, 42(4), Aug 2023. doi: 10.1145/3592422.

T. Cui, Z. Wang, and Z. Zhang. A variational neural network approach for glacier modelling with nonlinear rheology. *arXiv preprint arXiv:2209.02088*, 2022.

A. Ern and J.-L. Guermond. *Theory and practice of finite elements*, volume 159. Springer, 2004.

V. Girault and P. Raviart. *Finite Element Methods for Navier-Stokes Equations: Theory and Algorithms*. Springer Series in Computational Mathematics, 1986.

J. W. Glen. Rate of Flow of Polycrystalline Ice. *Nature*, 172:721–722, 1953. doi: 10.1038/172721a0.

M. Grab. Swiss glacier thickness – release 2020, 2020.

R. Greve and H. Blatter. *Dynamics of Ice Sheets and Glaciers*. Springer Verlag, 2009.

Q. He, M. Perego, A. A. Howard, G. E. Karniadakis, and P. Stinis. A Hybrid Deep Neural Operator/Finite Element Method for Ice-Sheet Modeling, 2023. arXiv.

R. Hock. A distributed temperature-index ice- and snowmelt model including potential direct solar radiation. *Journal of Glaciology*, 45(149):101–111, 1999. doi: 10.3189/S0022143000003087.

K. Hutter. *Theoretical Glaciology*. Reidel, 1983.

G. Jouvet. Mechanical error estimators for shallow ice flow models. *Journal of Fluid Mechanics*, 807:40–61, 2016. doi: 10.1017/jfm.2016.593.

G. Jouvet. Inversion of a Stokes glacier flow model emulated by deep learning. *Journal of Glaciology*, 69(273):13–26, 2023. doi: 10.1017/jog.2022.41.

G. Jouvet, G. Cordonnier, B. Kim, M. Lüthi, A. Vieli, and A. Aschwanden. Deep learning speeds up ice flow modelling by several orders of magnitude. *Journal of Glaciology*, 68(270):651–664, 2022. doi: 10.1017/jog.2021.120.

G. Jouvet, D. Cohen, E. Russo, J. Buzan, C. Raible, W. Haeberli, S. Kamleitner, S. Ivy-Ochs, M. Imhof, J. Becker, A. Landgraf, and U. Fischer. Coupled climate-glacier modelling of the last glaciation in the Alps. *Journal of Glaciology, in press*, 2023.

E. Kharazmi, Z. Zhang, and G. E. Karniadakis. Variational physics-informed neural networks for solving partial differential equations. *arXiv preprint arXiv:1912.00873*, 2019.

C. Khroulev and the PISM Authors. PISM, a Parallel Ice Sheet Model v1.2: User's Manual. 2020. URL `www.pism-docs.org`.

D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521 (7553):436–444, 2015. doi: 10.1038/nature14539.

J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

S. Markidis. The old and the new: Can physics-informed deep-learning replace traditional linear solvers? *Frontiers in big Data*, page 92, 2021. doi: 10.3389/fdata.2021.669097.

F. Maussion, A. Butenko, N. Champollion, M. Dusch, J. Eis, K. Fourteau, P. Gregor, A. H. Jarosch, J. M. Landmann, F. Oesterle, et al. The Open Global Glacier Model (OGGM) v1. 1. *Geoscientific Model Development*, 12(3):909–931, 2019. doi: 10.5194/gmd-12-909-2019.

P. Minakowski and T. Richter. A priori and a posteriori error estimates for the Deep Ritz method applied to the Laplace and Stokes problem. *Journal of Computational and Applied Mathematics*, 421:114845, 2023. doi: 10.1016/j.cam.2022.114845.

L. Morland. Unconfined ice-shelf flow. In C. Veen and J. Oerlemans, editors, *Dynamics of the West Antarctic Ice Sheet*, volume 4 of *Glaciology and Quaternary Geology*, pages 99–116. Springer Netherlands, 1987. doi: 10.1007/978-94-009-3745-16.

W. S. B. Paterson. *The Physics of Glaciers*. Pergamon, New York, third edition, 1994.

F. Pattyn and . others. Benchmark experiments for higher-order and full-Stokes ice sheet models (ISMIP-HOM). *The Cryosphere*, 2(2):95–108, 2008. doi: 10.5194/tc-2-95-2008.

F. Pattyn, C. Schoof, L. Perichon, R. C. A. Hindmarsh, E. Bueler, B. de Fleurian, G. Durand, O. Gagliardini, R. Gladstone, D. Goldberg, G. H. Gudmundsson, P. Huybrechts, V. Lee, F. M. Nick, A. J. Payne, D. Pollard, O. Rybak, F. Saito, and A. Vieli. Results of the marine ice sheet model intercomparison project, mismip. *The Cryosphere*, 6(3):573–588, 2012. doi: 10.5194/tc-6-573-2012.

M. Raissi, P. Perdikaris, and G. E. Karniadakis. 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, 2019. doi: 10.1016/j.jcp.2018.10.045.

L. Räss, A. Licul, F. Herman, Y. Y. Podladchikov, and J. Suckale. Modelling thermomechanical ice deformation using an implicit pseudo-transient method (FastICE v1. 0) based on graphical processing units (GPUs). *Geoscientific Model Development*, 13(3): 955–976, 2020. doi: 10.5194/gmd-13-955-2020.

B. Riel and B. Minchew. High-dimensional flow law parameter calibration and uncertainty quantification over Antarctic ice shelves: a variational Bayesian approach using deep learning. *Authorea Preprints*, 2022. doi: 10.1002/essoar.10509116.1.

B. Riel, B. Minchew, and T. Bischoff. Data-driven inference of the mechanics of slip along glacier beds using physics-informed neural networks: Case study on rutford ice stream, antarctica. *Journal of Advances in Modeling Earth Systems*, 13(11), 2021. doi: 10.1029/2021MS002621.

C. Schoof. Variational methods for glacier flow over plastic till. *J. Fluid Mech.*, 555:299–320, 2006. doi: 10.1017/S0022112006009104.

C. Schoof and I. Hewitt. Ice-sheet dynamics. *Annual Review of Fluid Mechanics*, 45(1):217–239, 2013. doi: 10.1146/annurev-fluid-011212-140632.

H. Seroussi, S. Nowicki, A. J. Payne, H. Goelzer, W. H. Lipscomb, A. Abe-Ouchi, C. Agosta, T. Albrecht, X. Asay-Davis, A. Barthel, et al. ISMIP6 Antarctica: a multi-model ensemble of the Antarctic ice sheet evolution over the 21st century. *The Cryosphere*, 14 (9):3033–3070, 2020. doi: 10.5194/tc-14-3033-2020.

J. Taylor, W. Wang, B. Bala, and T. Bednarz. Optimizing the optimizer for data driven deep neural networks and physics informed neural networks, 2022.

B. Yu et al. The deep Ritz method: a deep learning-based numerical algorithm for solving variational problems. *Communications in Mathematics and Statistics*, 6(1):1–12, 2018. doi: 10.1007/s40304-018-0127-z.

APPENDIX A: PRE-TRAINING OVER A GLACIER CATALOGUE

Pre-training of the CNN emulator over a glacier catalogue was found beneficial, especially to avoid local minima during online training, and improve the memory capability of the emulator. Here we describe the implementation of the pre-training, and assess the accuracy of the pre-trained emulator with respect to the solver. To generate glacier shape inputs in an offline training process of the CNN, we use a glacier catalogue of 36 mountain glaciers at 8 different times and 100 m resolution (covering advancing and retreating stages) obtained by Jouvet et al. [2022] by glacier evolution simulations (Fig. 12). Further details about the construction of this catalogue are given in Appendix C of Jouvet et al. [2022]. The catalogue consists of a heterogeneous dataset with a large variety of possible glacier shapes (large/narrow, thin/thick, flat/steep, long/small, straight/curved glaciers, ...).

First, we fix the ice flow parameters (A, c) and the spatial resolution H to constant standard values ($A = 78$ MPa^{-3}

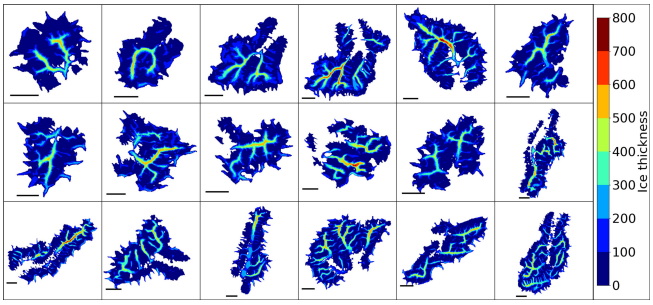

Fig. 12. Ice thickness at their maximum extent of half of the glacier catalogue (18 of the 36). Each glacier shape is a snapshot of a simulation initialised with ice-free conditions, and forced with a surface mass balance that permits building and retreat in successive phases over a total of 200 years. The horizontal bar represents 5 km to give the scale of each glacier.

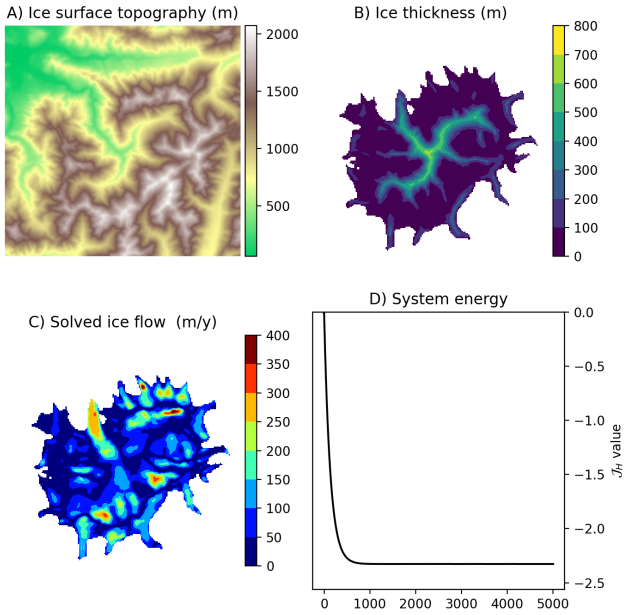

Fig. 13. Results of the solver on the "test" glacier: A) Ice surface topography and B) ice thickness of the "test" glacier C) "solved" surface ice flow solution at convergence D) evolution of the system energy through the iterations of the Adam optimiser.

a^{-1}, $c = 10$ km MPa^{-3}, $H = 100$ m) for simplicity. In a second experiment, we will vary these parameters at training.

A test glacier is selected in addition to the glacier catalogue, and a "solved" ice flow solution is obtained for this glacier by minimising the associated energy with the Adam optimiser. Figure 13 presents the results in terms of input data (panels A and B), "solved" solution (panel C), and a decrease in system energy (panel D).

Aside from the solver, we have trained a CNN emulator to minimise the system energy (solving the optimisation problem (20)) over the entire glacier catalogue (excluding the test glacier, Fig. 12), and evaluated its performance to reproduce the previously "solved" solution on a test glacier. As the size of the dataset is considerable, one used batches (a batch size of 8 was used here) to facilitate convergence (previously only a single glacier sample was used for online training at each iteration). In addition, we used an adaptive learning rate including an exponential decay to launch the training aggressively (10^{-4}) for efficiency and to end it gently (10^{-6}) for fine-tuning. Lastly, we have re-initialized the learning-rate each 5000 training iterations to prevent falling in local minima.

Figure 14 presents the results in terms of "emulated" solution when the training has converged (panel A), the difference between "solved" and "emulated" solutions (panel B), the L_1 error (panel C), and the decrease in the system energy through training iterations (panel D). As a result, the evolution of the L_1 error (panel C, Fig. 14) shows that the emulator captures well the ice flow after about 3000 iterations (the L1 error drops to ∼10 m/y). The effect of the adaptive learning rate (initially fixed at 10^{-4}, with exponential decay) is clearly visible: The first stage of training (iterations 0 to 1000) shows the largest decays and oscillations, while the last stage (iterations 4000 to 5000) is characterised by a smoother but slower decay. Interestingly, the energy associated with the "emulated" solution decreases towards a value (∼ -2.2) that is relatively close to the value obtained when solving (∼ -2.3), demonstrating that our CNN has learnt well to minimise the energy. Although the "emulated" and "solved" solutions show a fair degree of similarity (compare panel C of Fig. 13 with panel A of Fig. 14), the spatial pattern of the difference between the two (Fig. 14, panel B) reveals that the error is unevenly distributed, the highest discrepancy being found on the most prominent glacier tongue. This is presumably due to the relatively poor representation of large, fast-flowing glacier tongues in the glacier catalogue compared to a smaller one [Jouvet et al., 2022].

In a second experiment, we take over the emulator trained with fixed values of A, c, and H, and augment the training data by sampling additional values (but spatially constant) for $A \in [20, 100]$ MPa^{-3} a^{-1}, $c \in [0, 20]$ km MPa^{-3} a^{-1}, and training at a different resolution $H = 100, 200$ m. The ice flow parameters (A, c) were sampled with a uniform distribution within their ranges, while the spatial resolution H_H (initially 100 m) was randomly changed to 200 m by simple

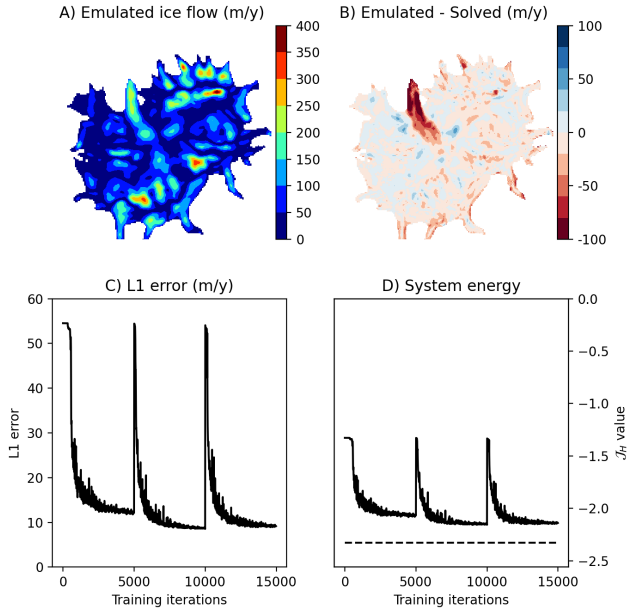

Fig. 14. Results of the emulator on the "test" glacier: A) "Emulated" surface ice flow at the surface of the test glacier (Fig. 13) at convergence of the offline training over the catalogue, B) difference between the "emulated" and "solved" solutions C) evolution of the L1 error between the two solutions and D) of the system energy through the training epochs. The jumps each 5000 iterations are due to the re-initialization of the learning rate.

data upscaling. As a result, the CNN meets a large set of input parameters in terms of glacier shape (sampling into the catalogue as before) and other parameters. To assess the performance of the emulator, we compare "emulated" and "solved" solutions obtained with 5 sets of parameters (A, c, H) for the test glacier in Figure 15. As a result, the emulator generally captures roughly the ice flow for various parameter sets (compare the first and second rows of Figure 15). However, we find relatively high spatial discrepancies when displaying the difference between the two (third row of Figure 15), with L1 errors between 10 and 20 m/y. Such a deteriorated accuracy is not surprising: the storage capacity of our CNN model emulator has reached its limit, and one cannot expect a model of a given size (about 140'000 parameters) to store more realisations with similar accuracy.

APPENDIX B: ISMIP-HOM VALIDATION SOLUTIONS

ISMIP-HOM [Pattyn and others, 2008] experiments consist of modelling exercises based on various synthetic ice geometries and boundary conditions to produce different types of ice flow, which can be met in real glacier modelling. Here, we focus on ISMIP-HOM experiments A and C, which represent a wide panel of various 3D ice flow scenarios (from shearing to sliding-dominant flows) over a square horizontal domain of length $L > 0$: $\Omega = [0, L] \times [0, L]$. In experiment A, the ice geometry is defined by

$$s(x, y) = -x \tan(0.5^\circ),$$
$$b(x, y) = s(x) - 1000 + 500 \sin(2\pi x/L) \sin(2\pi y/L),$$

and a no-slip condition is prescribed on the bedrock, while, in experiment C, the geometry is defined by

$$s(x, y) = -x \tan(0.1^\circ),$$
$$b(x, y) = s(x, y) - 1000,$$

and a slip condition is prescribed everywhere on the bedrock defined by $m = 1$ and

$$c(x, y) = [1000 \times (1 + \sin(2\pi x/L) \sin(2\pi y/L))]^{-1}.$$

In both experiments, we use $A = 100$ MPa^{-3} a^{-1} as Arrhenius factor in Glen flow law, and horizontal periodic boundary conditions connect the four horizontal sides of Ω, see Pattyn and others [2008] for further details. The squared horizontal domain Ω was divided into 100 cells in both horizontal directions to generate a regular grid, while the ice thickness is divided into 20 layers. To obtain a wide range of aspect ratios, we performed both experiments for several values of domain length L = 10, 20, 40, 80, and 160 km. Figure 16 compares the "solved" solutions at convergence with the reference 'oga1' solution obtained from Pattyn and others [2008] for all experiments.

As a result, we generally find a very good agreement between the two solutions. In line with model intercomparisons [Pattyn and others, 2008], there are small discrepancies in the experiments that have the smallest domain length L, which are known to be more sensitive to numerical parameters and schemes. This validates our numerical solver and verifies that the system energy (18) – which is used for solving and training the CNN – is correctly implemented.

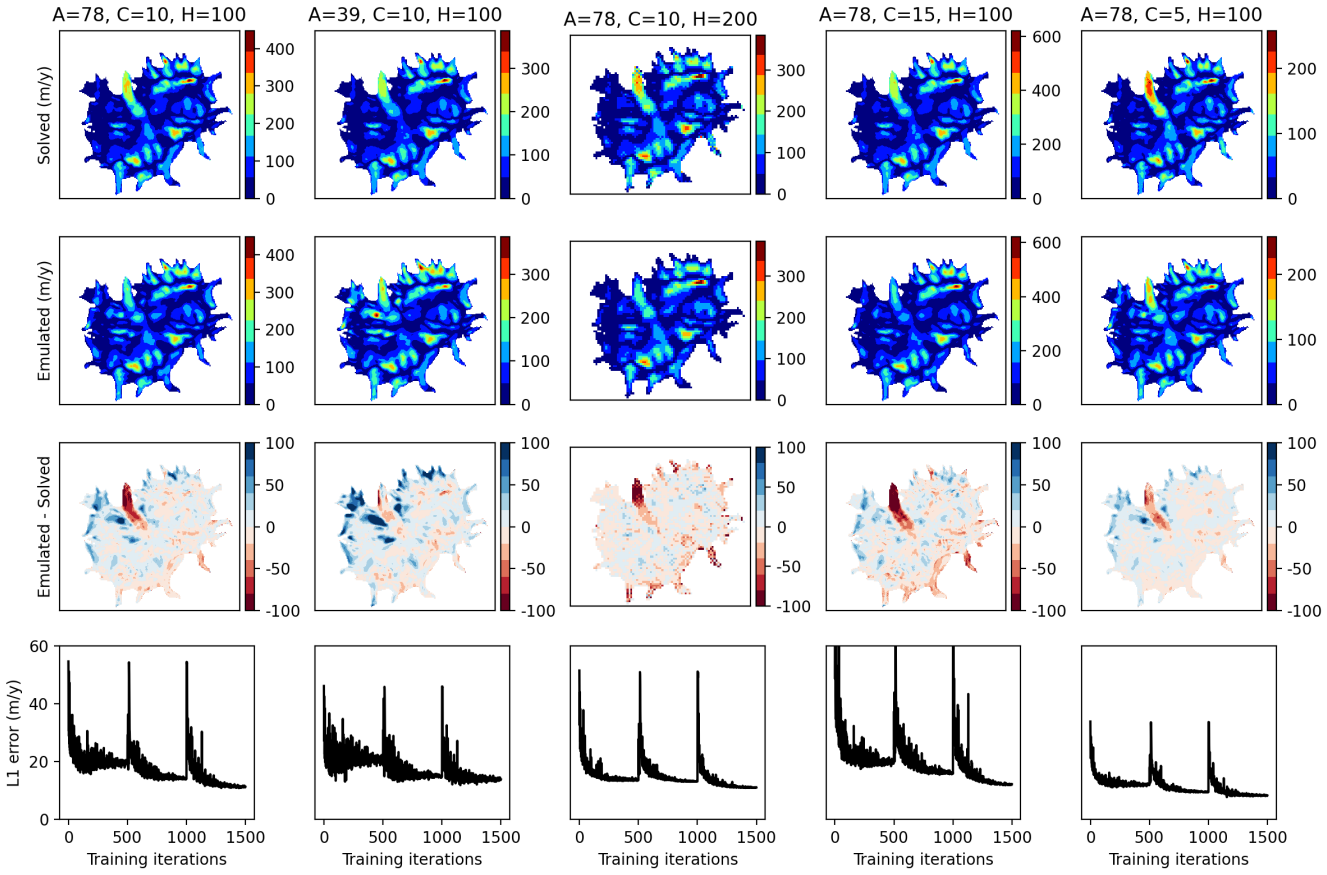

Fig. 15. Results of the emulator on the "test" glacier with varying values of A, c, and H. Each column corresponds to one parameter set (A, c, H) (the first column shows the default original parameters). The first row displays the "solved" surface ice flow solution. The second row displays the "emulated" solution after training over the glacier catalogue, while the third shows the difference between this solution and the "solved" one. The last raw shows the L1 error through the training. The jumps each 5000 iterations are due to the re-initialization of the learning rate.

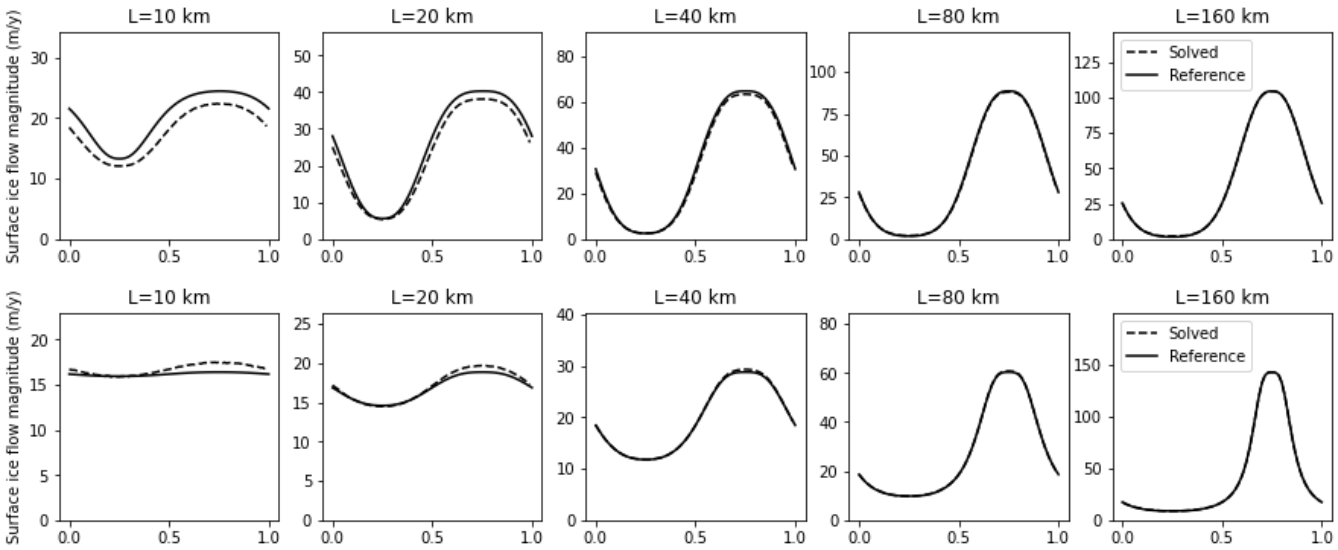

Fig. 16. Surface ice flow magnitude along the $y = L/4$ horizontal line for different length scales L = 10, 20, 40, 80, and 160 km in the ISMIP-HOM experiments A and C: comparison between "solved" with reference solution 'oga1' obtained from Pattyn and others [2008]. For simplicity, the x-axis was scaled with L.