Machine learning for glacier modelling

Jordi Bolibar — CNRS, IGE Institut des géosciences de l'environnement, UGA Université Grenoble Alpes

Facundo Sapienza — Stanford University

Kamilla Hauknes Sjursen — Western Norway University of Applied Sciences

Overview

1. Modelling the glacier system

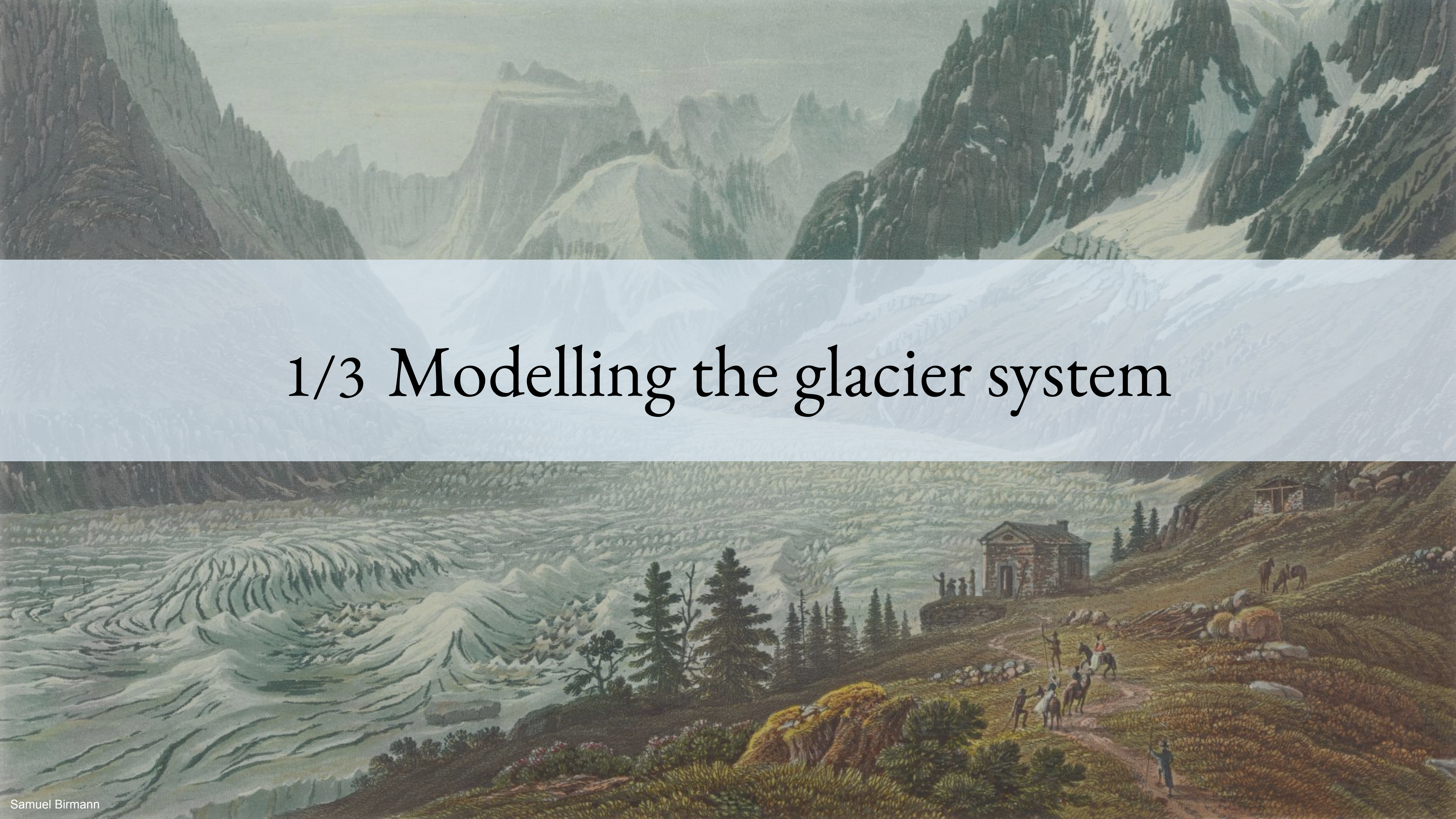
- What is a glacier and how do we model them?

2. Machine learning to model glacier physical processes

- What is machine learning for regression?
- How can it be useful in glacier modelling?
- How can we be sure that we are actually respecting physics?
- What are the main pitfalls and limitations when training ML models?

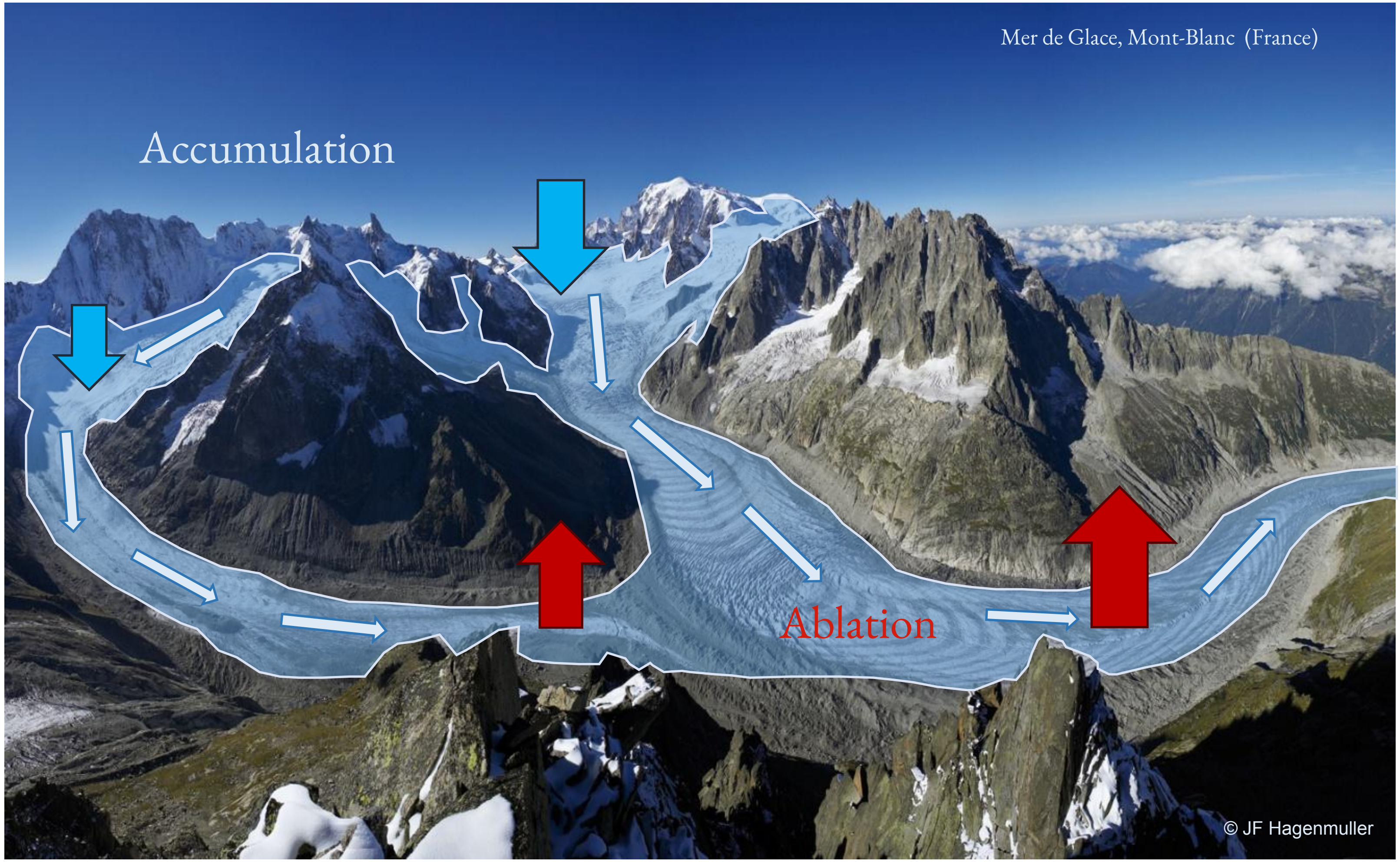
3. Project description

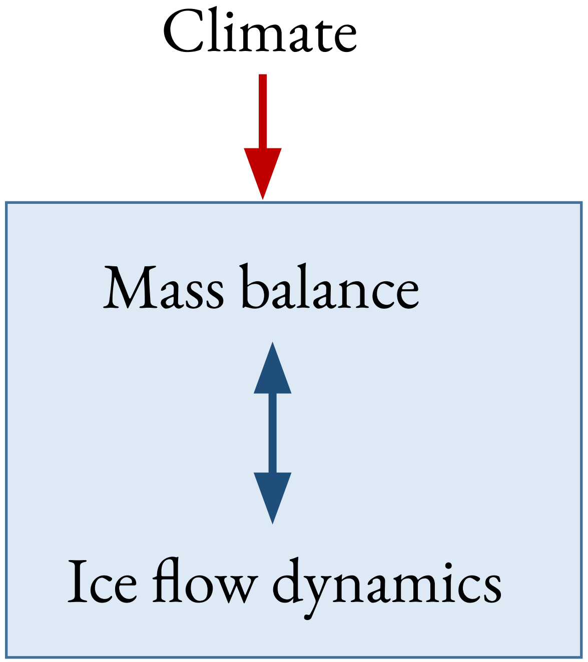
1/3 Modelling the glacier system



Samuel Birmann

The glacier system





Glacier evolution model

Glacier evolution models

Mass balance component

- Ablation
- Accumulation
- Refreezing
- Debris cover
- Calving

…

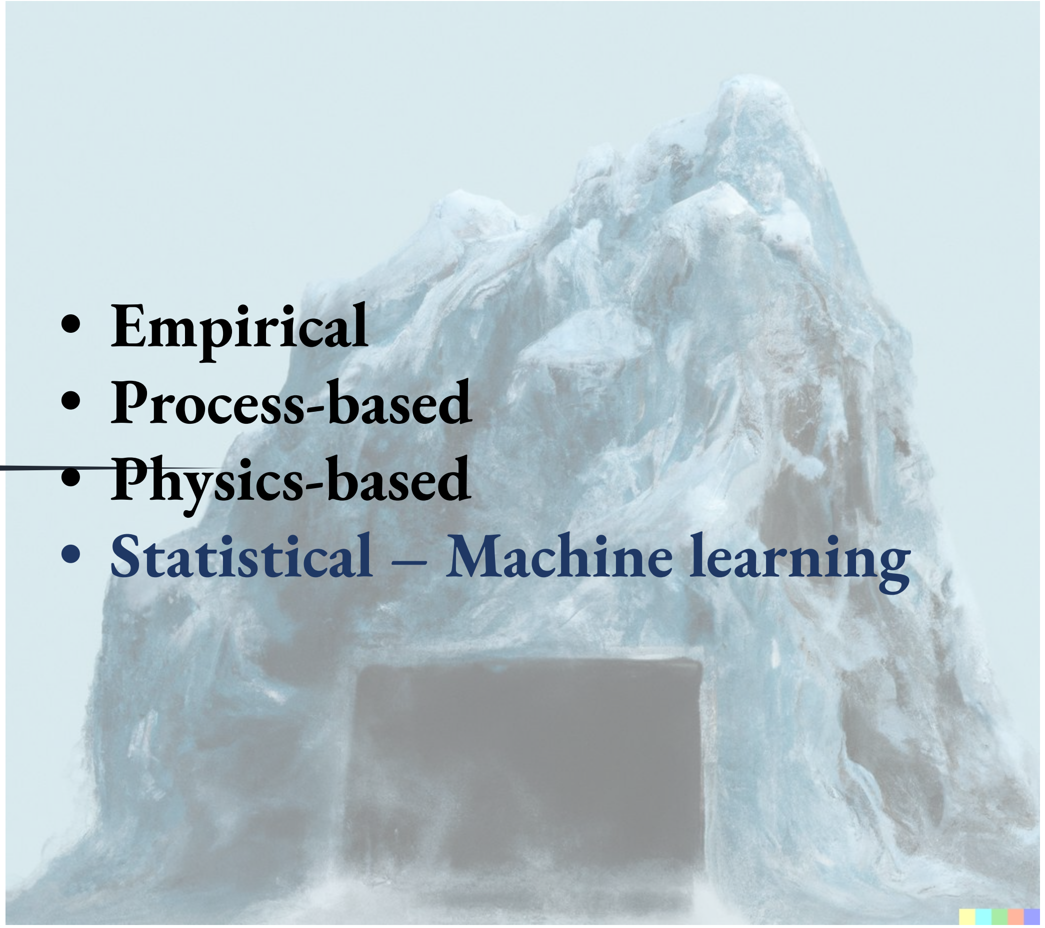
Analytical form

Ice dynamics component

- Creep
- Basal sliding

…

Differential equations



- **Empirical**
- **Process-based**
- **Physics-based**
- **Statistical – Machine learning**

Local vs Global glacier modelling

Local modelling (1-10 glaciers)

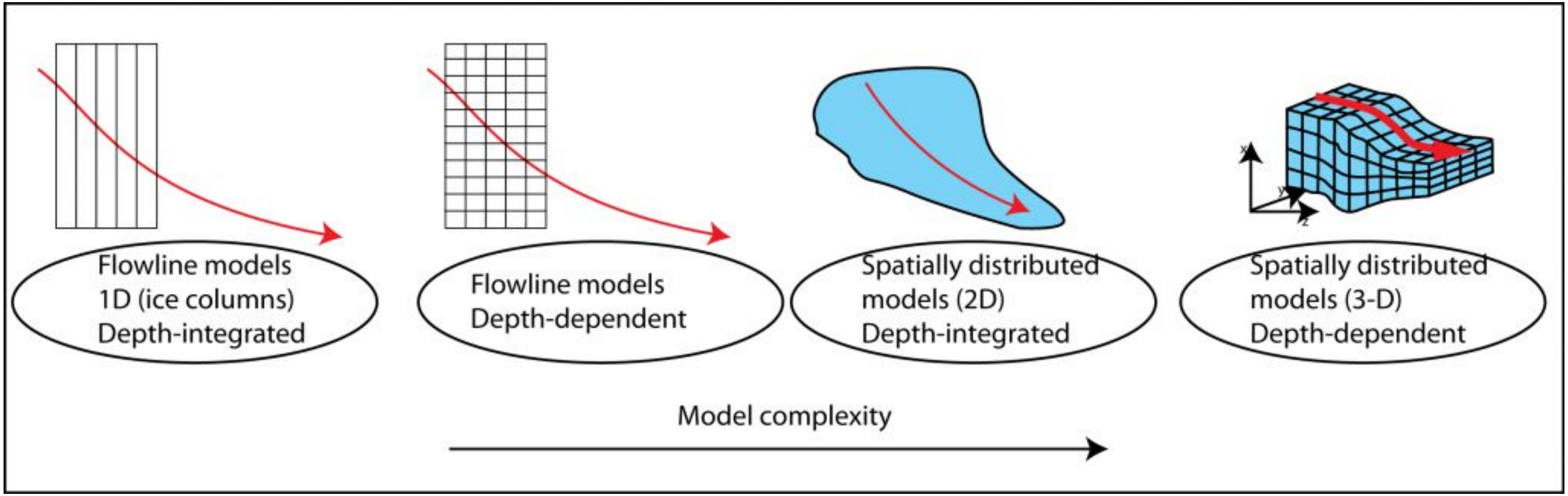
- More complex models can be used (e.g. Full Stokes)
- Model parameters can be fine tuned manually

Regional to Global modelling (100s-200,000 glaciers, ice sheets)

- Computational cost needs to be taken into account by using simplifications (e.g. Δh, SIA, flowline...)
- Model parameters need automation based on optimization

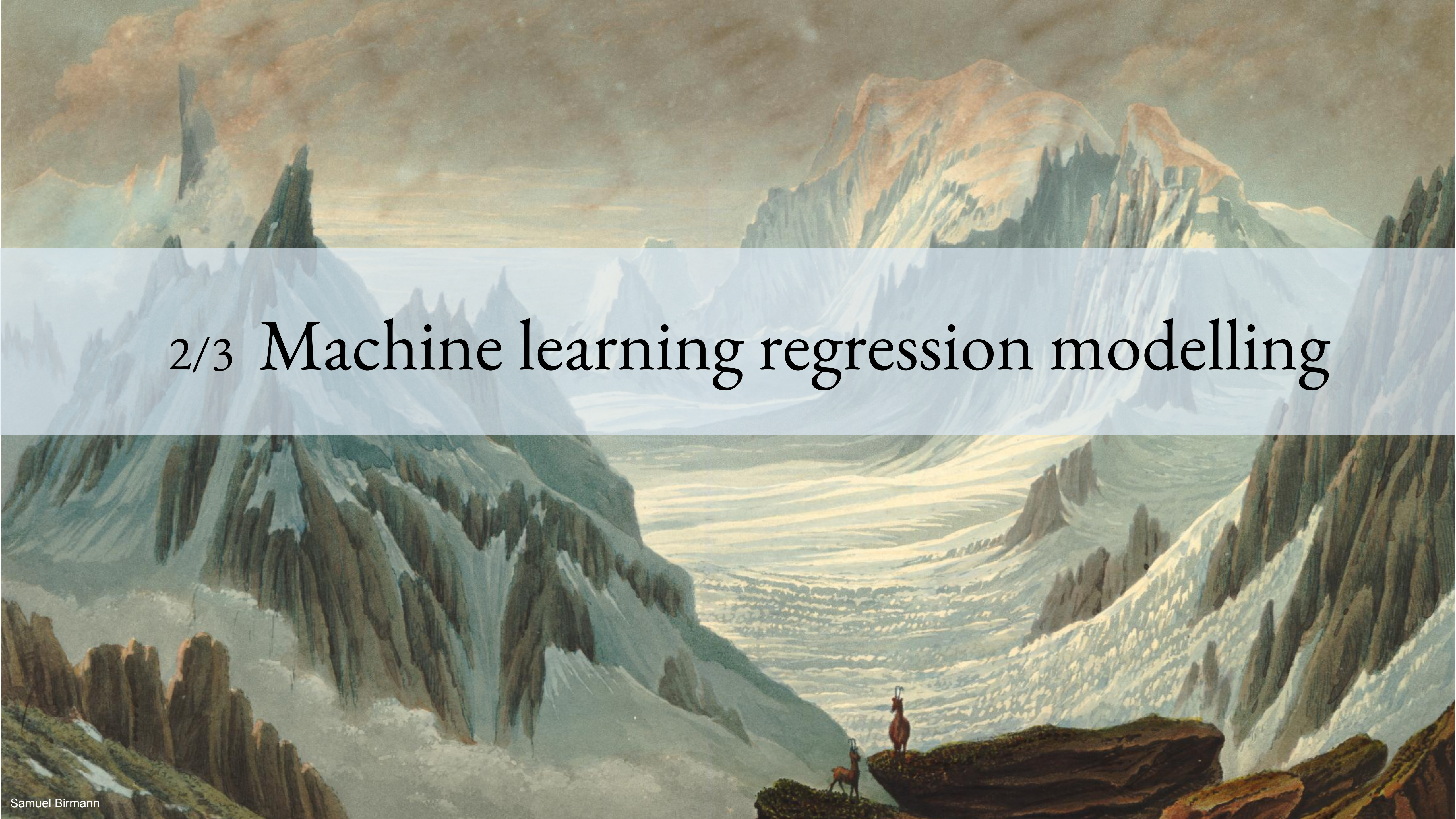
Model complexity vs capacity (flexibility)

Local vs Global glacier modelling



antarcticglaciers.org

2/3 Machine learning regression modelling



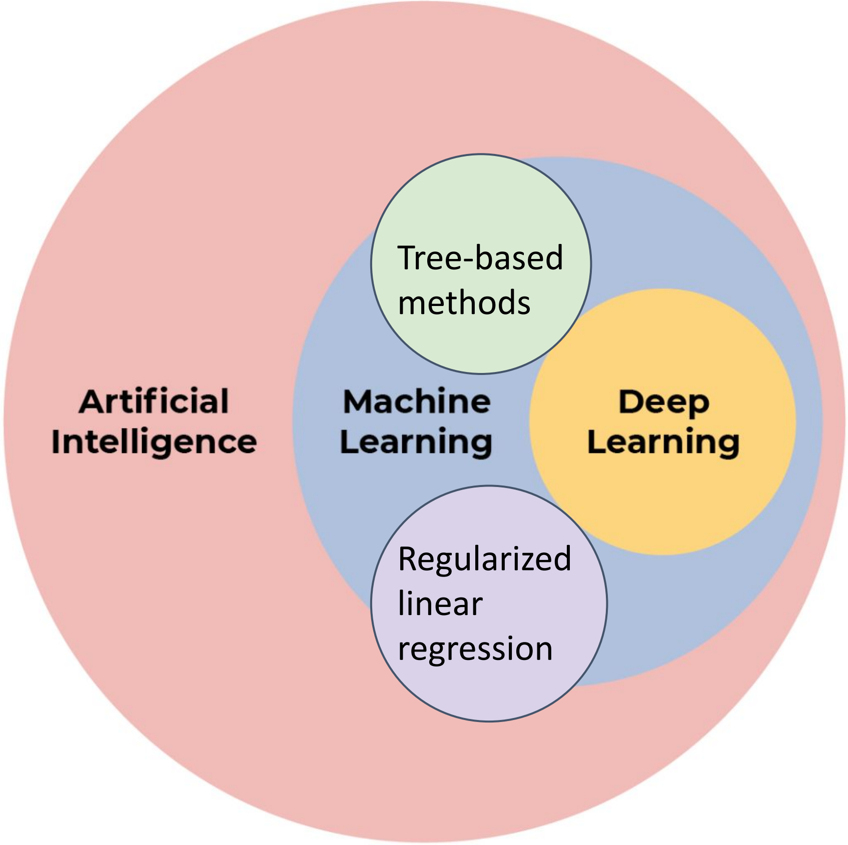
Samuel Birmann

Machine learning models

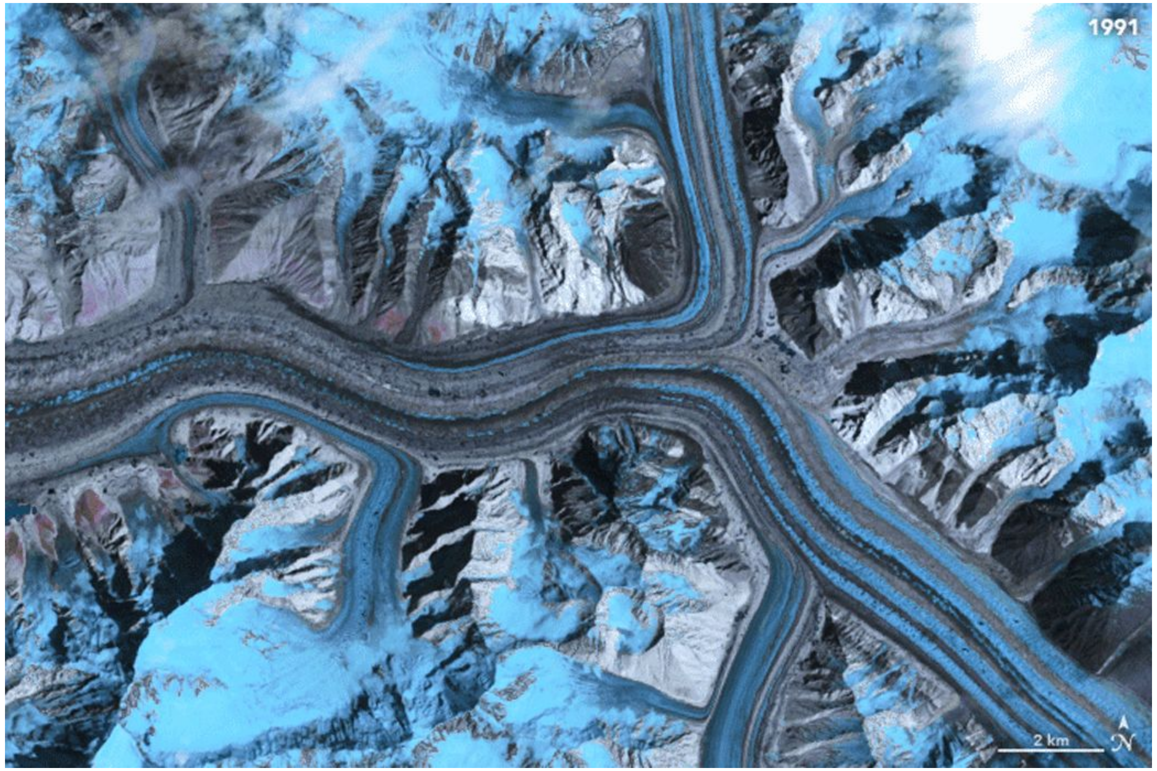
Statistical
data-driven models

Regression

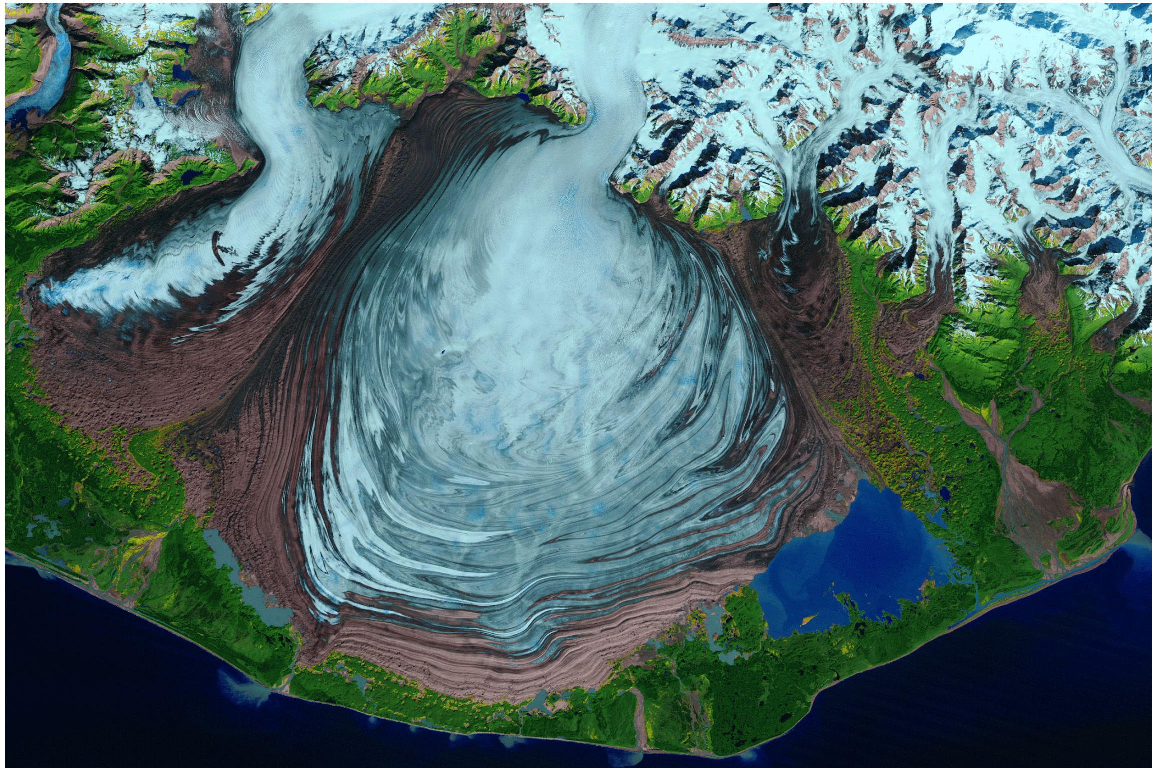
Approximating functions



A new era of remote sensing observations



Baltoro glacier, Karakoram



Malaspina glacier, Alaska

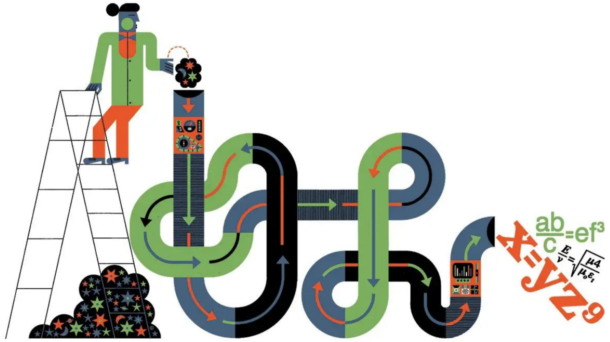
How can ML be useful in glacier modelling?

Machine learning modelling can help in many directions:

- **Speed**: Statistical emulators

- **Prediction:** Data-driven machine learning

- **Discovering hidden physics:** Universal Differential Equations, Physics-Informed Neural Networks



Raymond Biesinger

How can ML be useful in glacier modelling?

Let’s see a few examples…

Speeding up scientific models with emulators

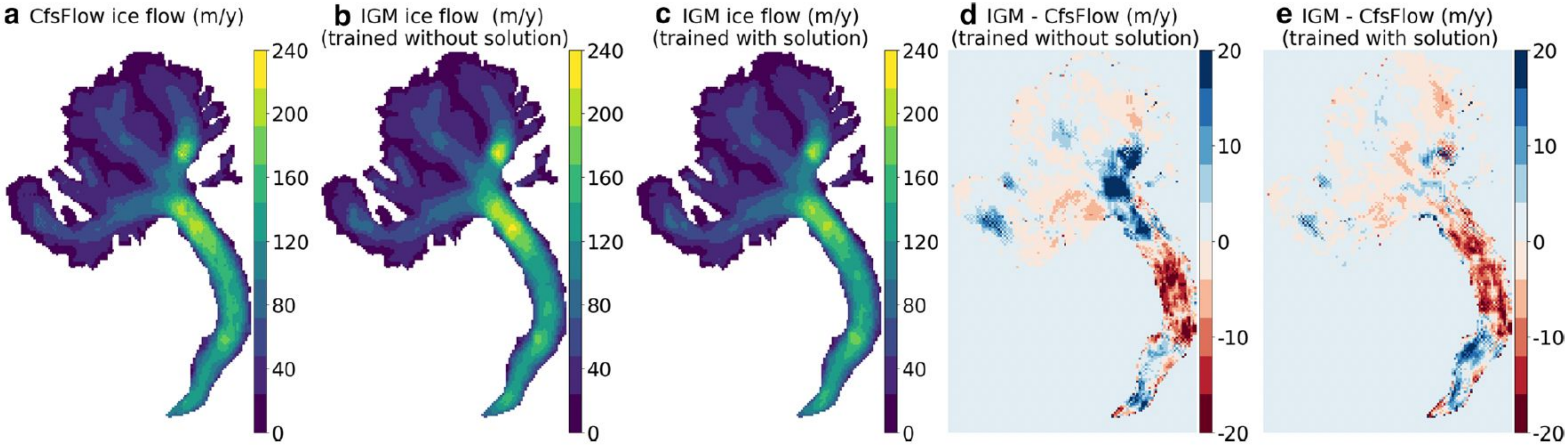
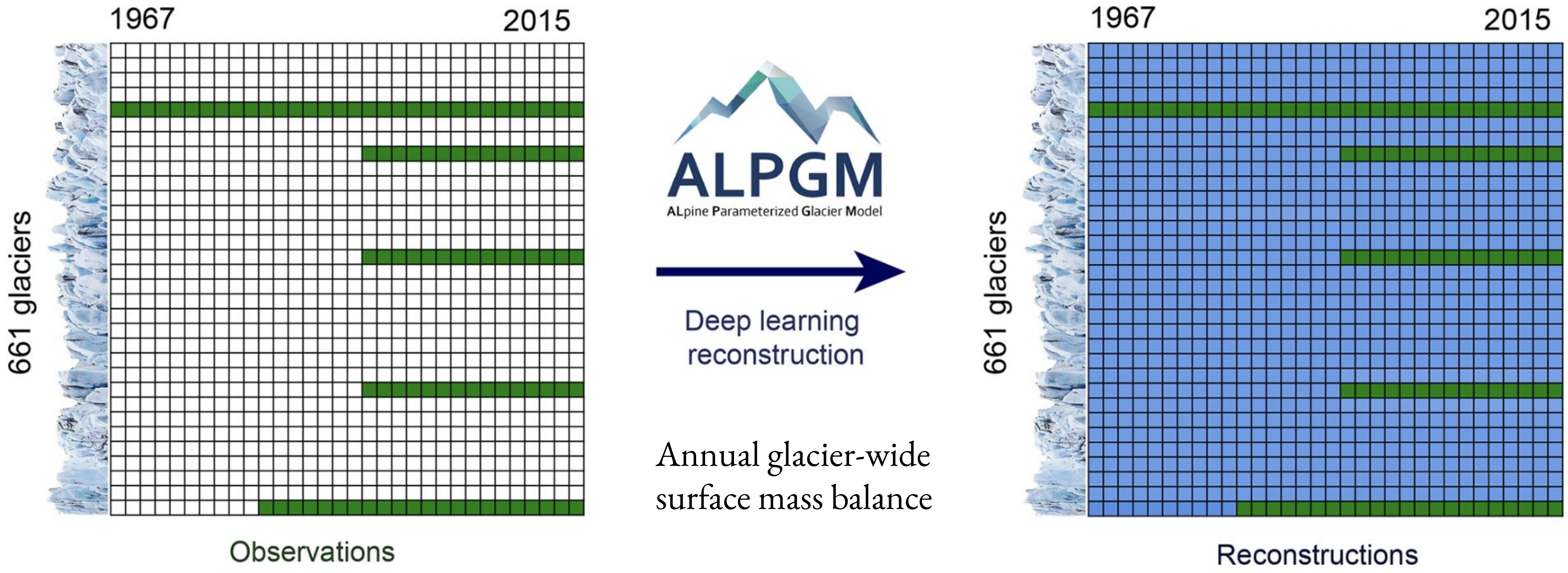


Fig. 10. Vertically-averaged ice flow magnitude of the Aletsch glacier at its maximum state: CfsFlow reference solution (a), the IGM solution trained without (b) and with (c) the solution, and the difference between IGM and CfsFlow solutions (d and e).

Fig. 3. The function we aim to emulate by learning from hybrid SIA + SSA or Stokes realizations maps geometrical fields (thickness and surface slopes) and basal sliding parametrization to ice flow fields.

Jouvet et al. (2021)

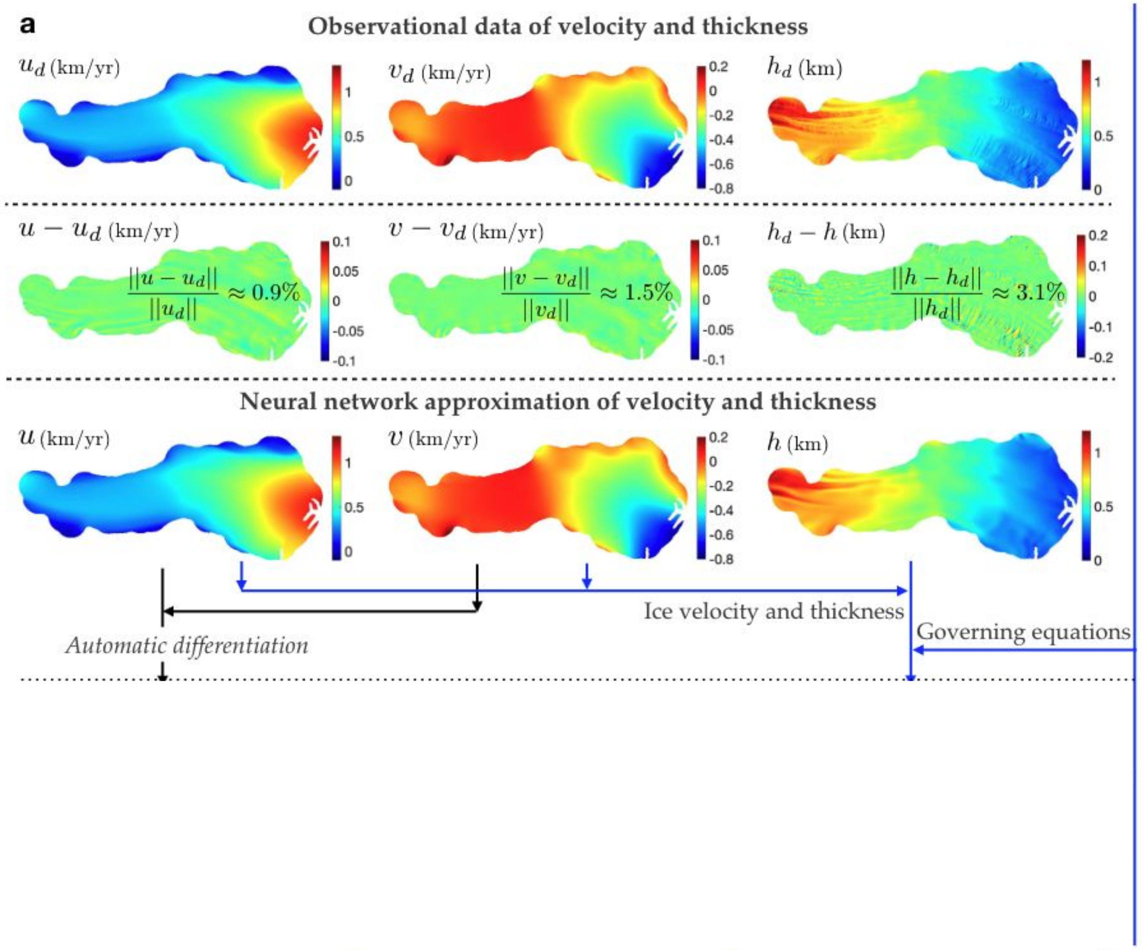
Making predictions with ML models



Annual glacier-wide
surface mass balance

Bolibar et al. (2020b)

Discovering hidden physics



Discovering the rheology of Antarctic Ice Shelves via physics-informed deep learning. Wang et al., *in review*

How do we actually train ML models?

Regression machine learning pipelines

- Split X and Y into train, validation and test datasets, where X is the feature matrix and Y is the target variable
- Train parameters of model M to learn a function f, such as $Y = f(X) + \epsilon$
- Tune hyperparameters which determine model M using the train and validation datasets
- Assess final model performance on independent test dataset

Regression for physical processes

Modelling physical processes implies a series of challenges:

1. How can we be sure that we are respecting basic physical properties?
   Feature selection, Physical constraints
2. How can we be sure that we are modelling the actual physical process?
   Testing and validation
3. What are the limitations in the model design and its predictions?
   Training strategies, Spatiotemporal resolution

Regression for physical processes

Modelling physical processes implies a series of challenges:

1. How can we be sure that we are respecting basic physical properties?

   Feature selection, Physical constraints

2. How can we be sure that we are modelling the actual physical process?

   Testing and validation

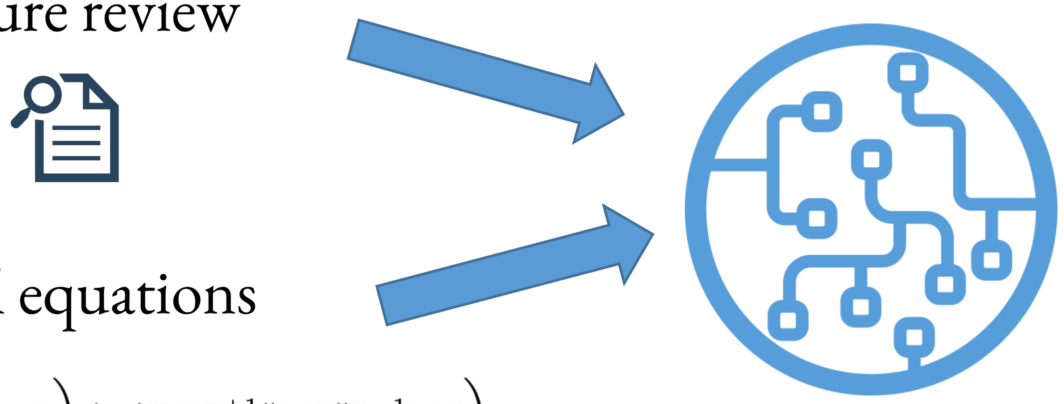
3. What are the limitations in the model design and its predictions?

   Training strategies, Spatiotemporal resolution

1. Respecting physics

- Good domain knowledge in the physics involved. Make sure the chosen features are relevant to the process.

Literature review



Model design

Review physical equations

$$\frac{\partial H}{\partial t} = \dot{b} + \nabla \cdot \left(\left(C + \frac{2A}{n+2} H \right) (\rho g)^n H^{n+1} \|\nabla S\|^{n-1} \nabla S \right)$$

1. Respecting physics

- Add physical constraints in the model design (e.g. PINNs, UDEs)
- Train model on physical model outputs (i.e. an emulator)

Want to learn more?

Attend bonus lecture on Physics-informed Machine Learning!

Regression for physical processes

Modelling physical processes implies a series of challenges:

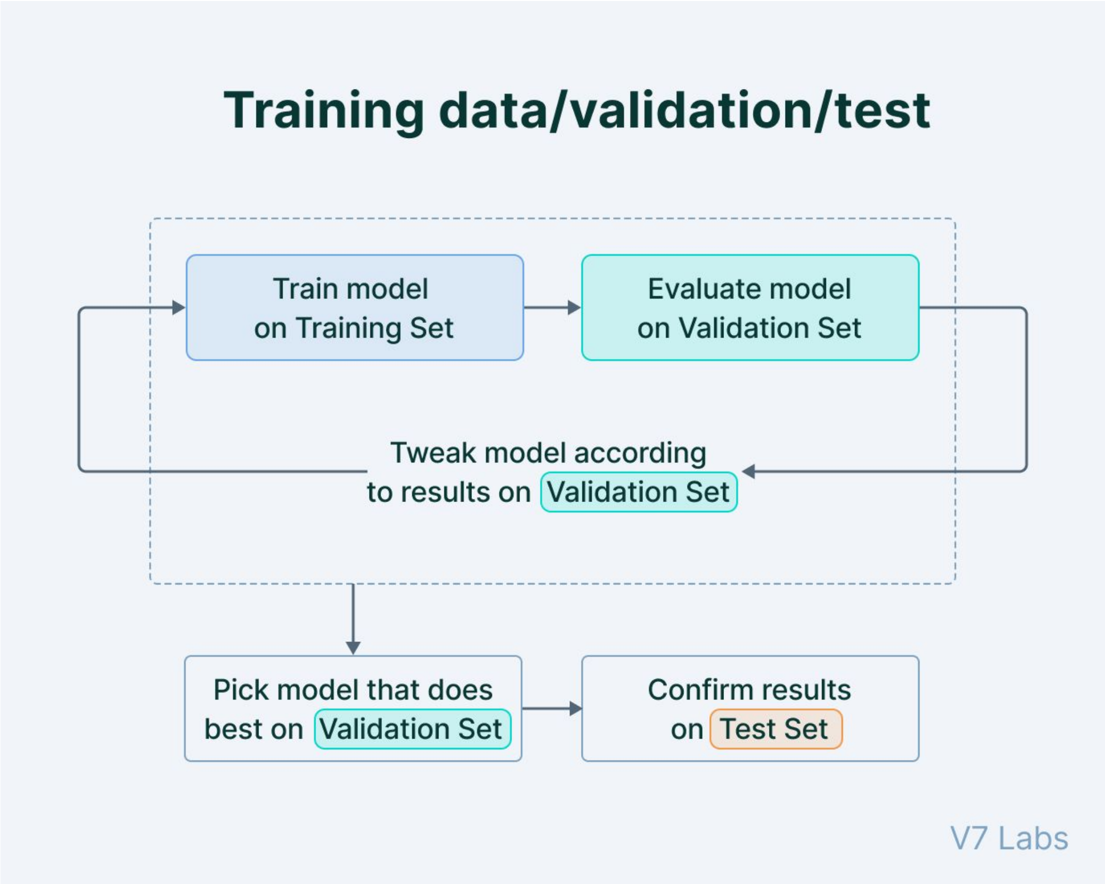
1. How can we be sure that we are respecting basic physical properties?
   Feature selection, Physical constraints

2. How can we be sure that we are modelling the actual physical process?
   Testing and validation

3. What are the limitations in the model design and its predictions?
   Training strategies, Spatiotemporal resolution

2. Trustworthy models

- A robust validation and test pipeline is the best way to make sure we are learning useful information.
- Respect spatiotemporal structures in dataset
- Using an independent **well-chosen** test dataset will help identify if the model has any pathologies.



Regression for physical processes

Modelling physical processes implies a series of challenges:

1. How can we be sure that we are respecting basic physical properties?
   Feature selection, Physical constraints

2. How can we be sure that we are modelling the actual physical process?
   Testing and validation

3. What are the limitations in the model design and its predictions?
   Training strategies, Spatiotemporal resolution

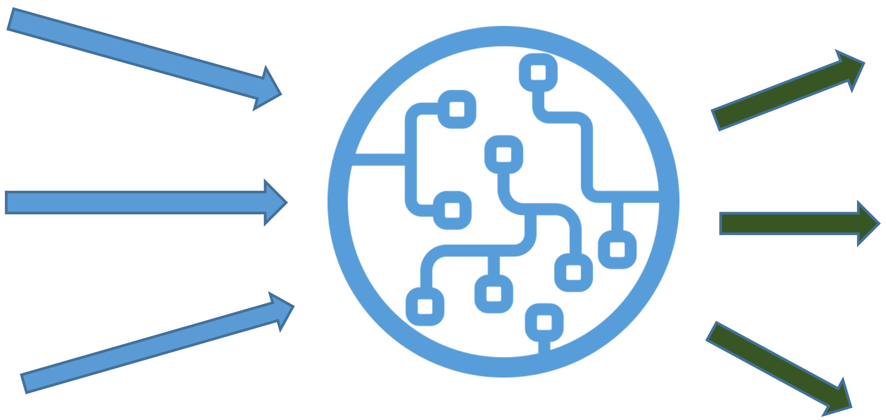
3. Being mindful about model limitations

- It is crucial to understand how the model has been designed and what its strengths and weaknesses are

Spatiotemporal data resolution

Validation/test and cross-validation strategy

Data quality and coverage



Acceptable applications

Spatial or temporal domain generalization

Uncertainty and error interpretation

3/3 Project description

Samuel Birmann

Machine learning for glacier modelling

Main project:

Explore different regression methods and learn the fundamentals of a machine learning pipeline, applied to glacier mass balance modelling.

Bonus projects:

1. Predict distributed (2D) geodetic mass balance rates
2. Infer distributed (2D) glacier ice thickness

Machine learning for mass balance modelling

Follow the tutorials from the following Jupyter notebooks:

https://github.com/Machine-Learning-in-Glaciology-Workshop/Project_MB_Regression

1_Preprocessing.ipynb	grammar check	5 days ago
2_Data_exploration.ipynb	grammar check	5 days ago
3_Training.ipynb	grammar check	5 days ago
4_Validation.ipynb	grammar check	5 days ago
5_Distributed_preprocessin...	grammar check	5 days ago
6_Bonus_Projects.ipynb	grammar check	5 days ago


Thank you for your attention
© Tom Bernardo
© NASA