Ice Breaker

Team Members:

Chhaya Kulkarni

Atefeh Jebeli

Mathieu Morlighem

Omar Faruque

Homayra Alam


Can we use AI to map
the landscape hidden under the ice sheet
based on what we see at the surface?

Team Goal

Prediction of Greenland Ice Bed Topography

Questions for the hackathon:

1) How do you process the input data to train the model?

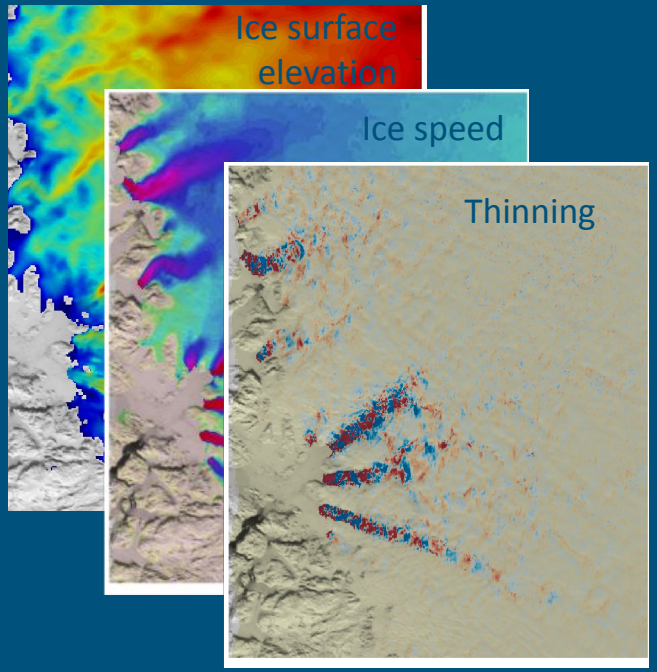
2) What model did you use?

3) Did you partition the training set to test the performance of your model?

Project Challenges

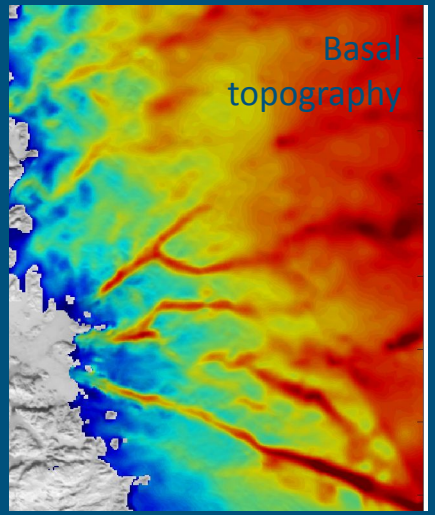
- Understanding the domain
- Data Mapping
- Finding a formula to extract features of each track bed point

Satellite derived observation of the surface

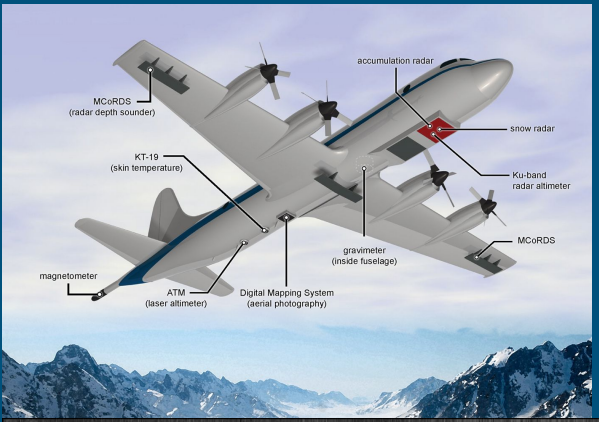


AI

Infer shape of the bedrock underneath (bed height)

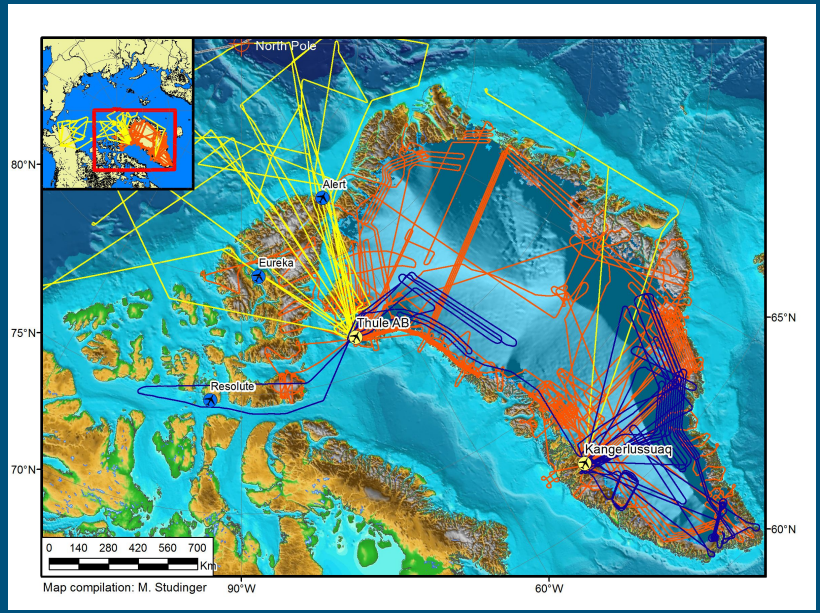


Training set: Ice penetrating radar data

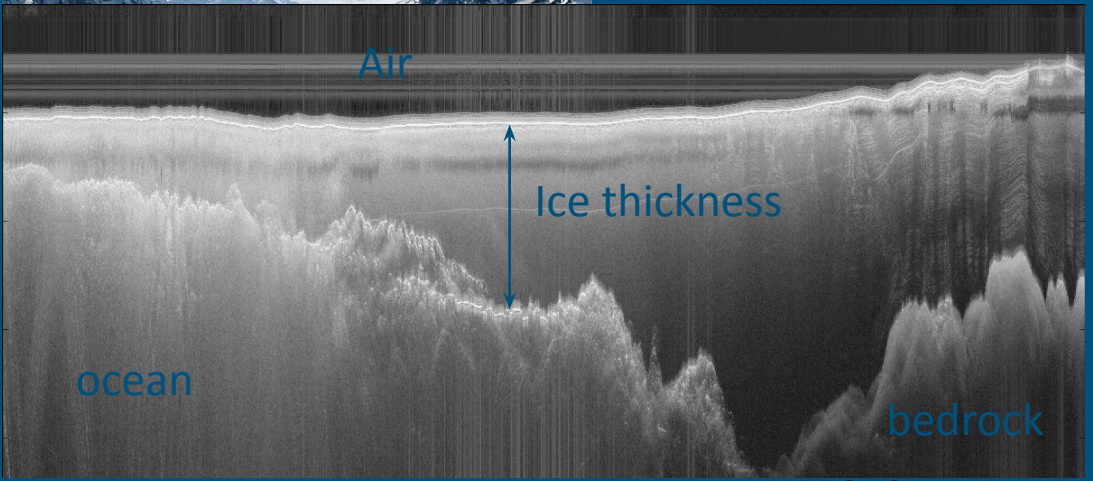


NASA P-3

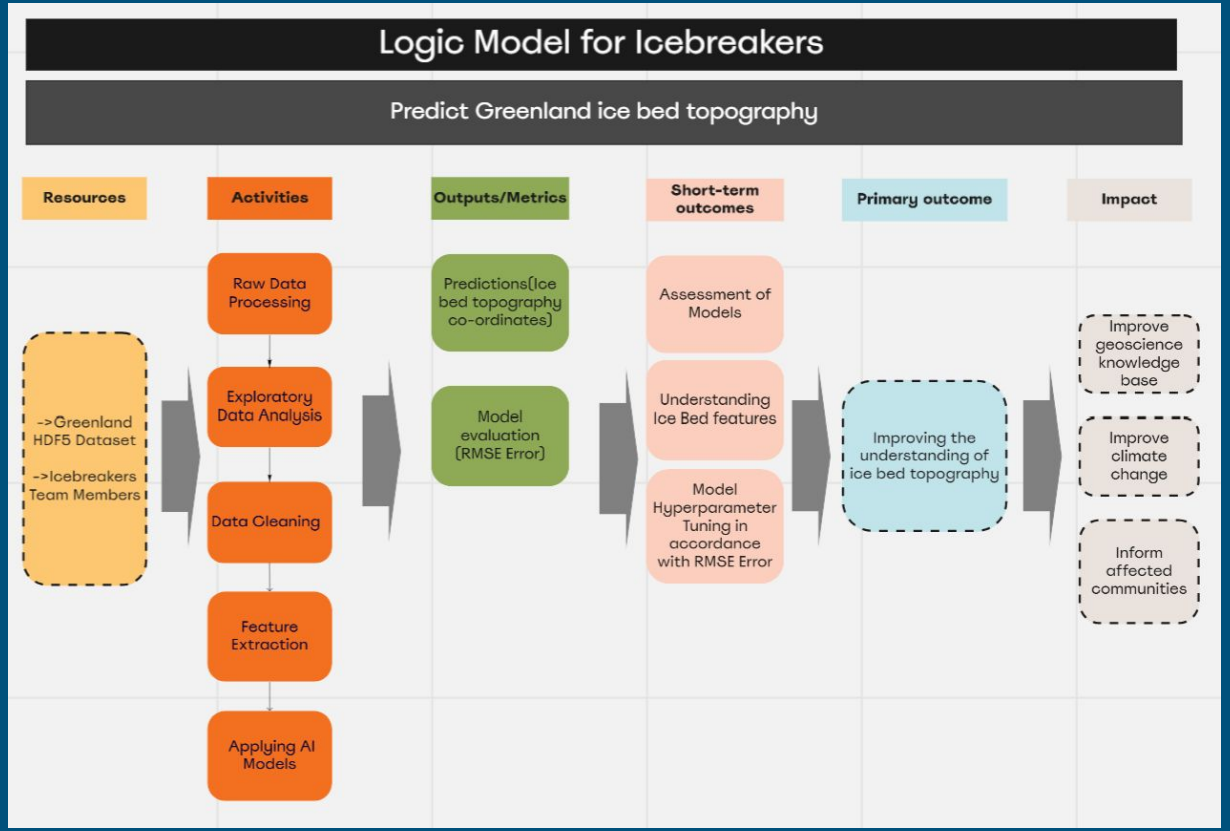
Flight lines OIB Arctic 2012



Radar echogram

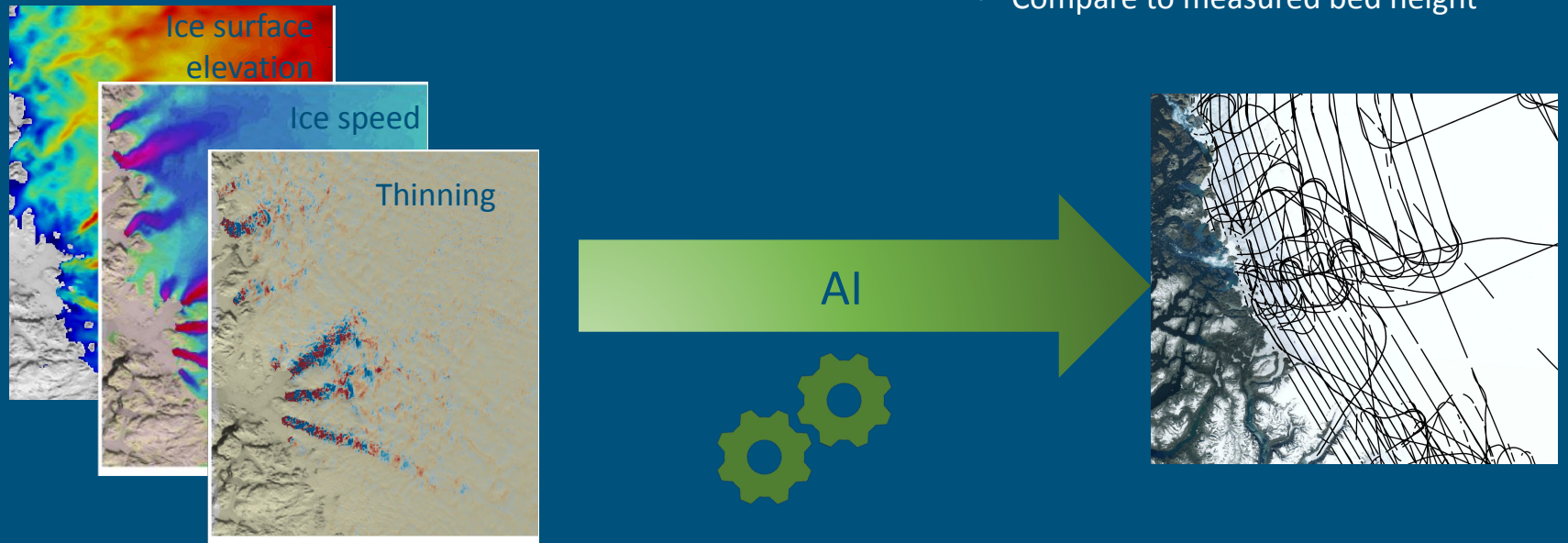


Logic Model


Logic Model for Icebreakers
Predict Greenland ice bed topography
Resources
Activities
Outputs/Metrics
Short-term outcomes
Primary outcome
Impact
->Greenland HDF5 Dataset
->Icebreakers Team Members
Raw Data Processing
Exploratory Data Analysis
Data Cleaning
Feature Extraction
Applying AI Models
Predictions(Ice bed topography co-ordinates)
Model evaluation (RMSE Error)
Assessment of Models
Understanding Ice Bed features
Model Hyperparameter Tuning in accordance with RMSE Error
Improving the understanding of ice bed topography
Improve geoscience knowledge base
Improve climate change
Inform affected communities

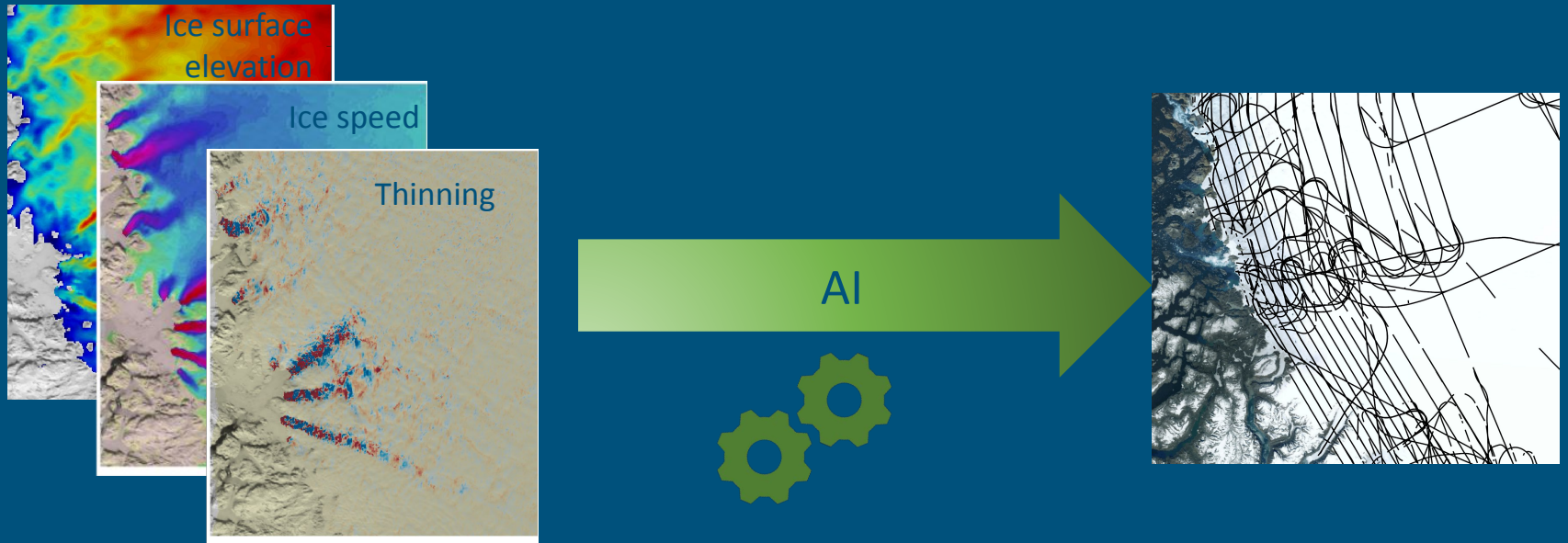
- Input data: 2D grids (150 m resolution, HDF5 format)
  - Ice surface elevation [m] (`surf_elv`)
  - Ice flow velocity vectors [m/yr] (`surf_vx`, `surf_vy`)
  - Ice thinning rates [m/yr] (`surf_dhdt`)
  - Snow accumulation [m/yr] (`surf_SMB`)
- Coordinates of cell centers [m]: `surf_x`, `surf_y`

- Training set: 1D flight line data (along lines)
  - `track_bed_training`
    - 1st column: x [m]
    - 2nd column: y [m]
    - 3rd column: bed height [m]
- For each point (x,y)
  - Find corresponding pixel in input grids
  - Prepare input data
  - Infer bed height
  - Compare to measured bed height



- Testing set: 1D flight line data not used in training set
  - `track_bed_testing`
    - 1st column: x [m]
    - 2nd column: y [m]

- The RMSE of the model is computed to the measured bed height



Data Preprocessing

1. Read the raw HDF5 dataset
2. Find the closest pixel index of track bed points from surf_x and surf_y
3. Find the values of surf_dhdt, surf_elv, surf_SMB, surf_vx, surf_vy corresponding to the pixel index
4. Create the training dataset in CSV “Hackathon_train”
5. Repeat the procedure for creating the test dataset

Machine Learning Model

- The train test split is 70%(training), 30%(testing)
- We applied Grid Search for Random Forest Regressor
- Applying the best model on the dataset

Results

First Approach: Linear Regression

MAE train: 107.08
MAE test: 111.09

MSE train: 21451.62
MSE test: 21749.83

RMSE train: 146.46
RMSE test: 147.48

R^2 train: 0.40
R^2 test: 0.41

mean(y_real)= 360
mean(y_predic)=175

Results(Cont.)

Second Approach: Decision Trees

MAE train: 107.08
MAE test: 111.09

MSE train: 21451.62
MSE test: 21749.83

RMSE train: 146.46
RMSE test: 147.48

R^2 train: 0.40
R^2 test: 0.41

mean(y_real)= 360
mean(y_predic)=175

Questions