Background Reading Notes

Goal: Estimating topography (beneath the ice) of greenland ice beds by:

* Predicting missing data in topographic maps using:
  + satellite/radar imagery
    - Ice flow speed (surf\_vx, surf\_vy)
    - Surface height (surf\_elv)
    - Ice thinning rates (surf\_dhdt)
    - Snow accumulation (surf\_SMB)

Why: Control the flow of ice, subglacial drainage, and impact of climate change as ice sheets melt and sea level rises (& help inform decisions on policy to protect ice sheets

What happening:

1. Warm water intruding fjords under the ice and melting it from below
   1. Fjord: U-shaped valleys with steep cliffs.
   2. Ice retreat (slow - bumps and ridges in the way vs unstoppable - retrograde bed slope) controlled by shape of land

How Computation: Combining all features of ice bed surface and using dense and RNN layers to predict future patterns

* RNN
* CNN

Similar work:

* Morlinghem (2014). Shallow ice approx and mass conservation (physics) to determine topography of Greenland Ice Sheets
* Fretwell (2013), topographic map of antarctic ice bed
* NASA, MacGregor (2021), improved ice thickness data
* Le Brocq (2013), applied ML to forecast antarctica bed with AI NN’s

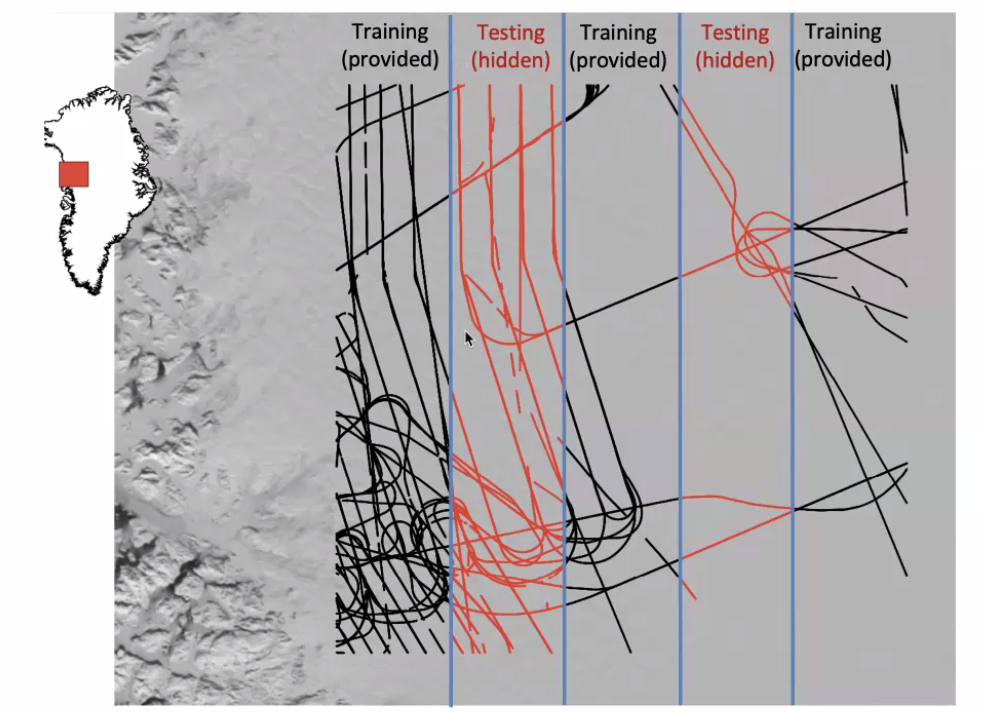
Data Collection: plane flying over land and collecting data immediately below.

Data Described:

* 10 features; 1201x1201=1,442,401 points
  + 632,706 (43.86%) data known
    - 396,734 training (62.7%)
    - 235,972 testing (37.3%)
  + Are we required to use this split for modeling?
    - Discussion of method says Adam method & mean squared error for optimization applied to each epoch shows 60-40 split
    - This can be changed.

.

* Only 5 independent vars in each grid location to use for prediction
  + Complex interaction? How?
    - relationship between changes - physics
* Target: track\_bed\_target
* Each data grid separated by 150m



Preprocessing:(Create data set & fill in scarcity through deriving) - data merging/integration

So the data is separated by 150m so we must derive features.

* (p,q)
  + (track\_bed\_x - surf\_x[0,1])/150 = first element of the index (p)
  + (surf\_y[-1,0] - track\_bed\_y)/150 = second element of the index (q)
* (surf\_smb, surf\_vx, surf\_vy, etc)

Do we have track\_bed\_x, track\_bed\_y or are these derived from track\_bed\_target?

What exactly is each variable?

* Ice flow speed (surf\_vx, surf\_vy)
* Surface height (surf\_elv)
* Ice thinning rates (surf\_dhdt)
* Snow accumulation (surf\_SMB)
* Coords cell centers: surf\_x, surf\_y

Select x and y points (150m apart). Get p,q which are row and column value.

Methodology:

7 derived features and target track\_bed\_target

Tried:

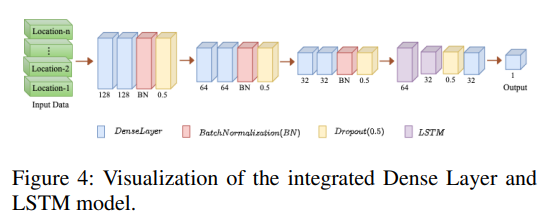
…

Best Result:

LSTM + Dense Layers

To extract the sequence

* Sigmoid activation used in each layer
* Batch size 20,000
* Epochs 100



Evaluation of Performance

* Root mean squared error – minimize
* Mean absolute error – minimize
* Coefficient of correlation (R^2) – maximize

Results and Discussion

* XGBoost performed better than others….
* LSTM (RNN) + Dense layers is the best.
* What is the accuracy?

Considerations for our work:

* Removing duplicate data points with different interpolation methods

Next Steps: Identify alternative approaches to improve models.