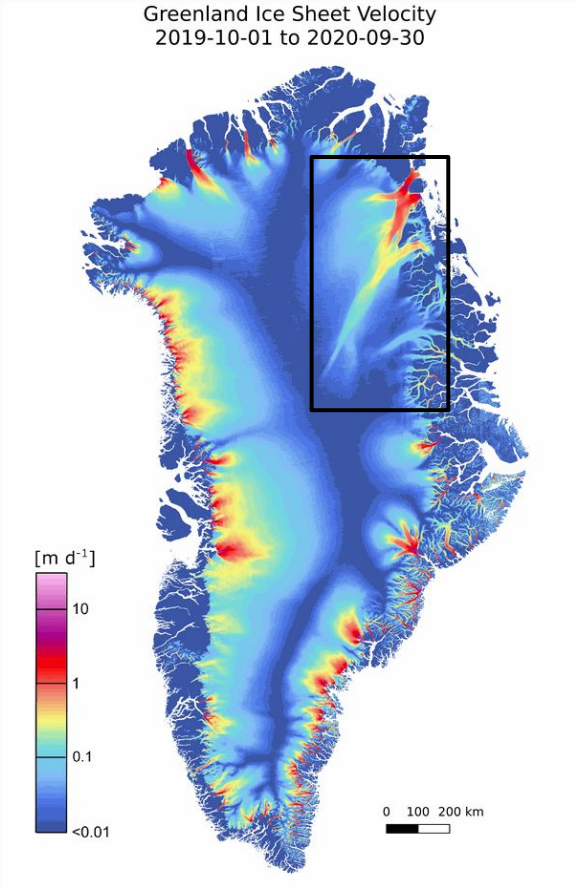
Predicting Seasonal Velocity Values at Zachariæ Isstrøm, Northeast Greenland

Claire Jensen

ESS-569 Final

Background

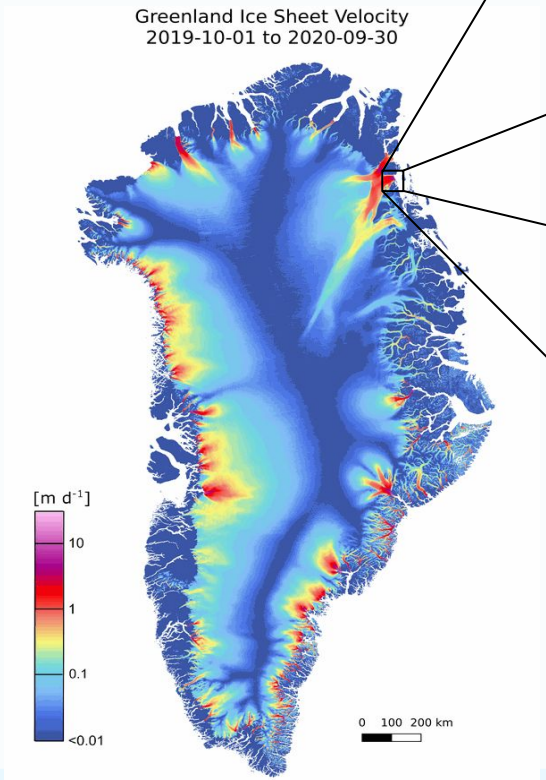
- NEGIS **drains ~12%** of the ice sheet
- 3 glaciers drain the NEGIS: (Zachariæ Isstrom (**ZI**), Nioghalvfjerdsfjorden (**79N**) and Storstrømmen (**Store**))

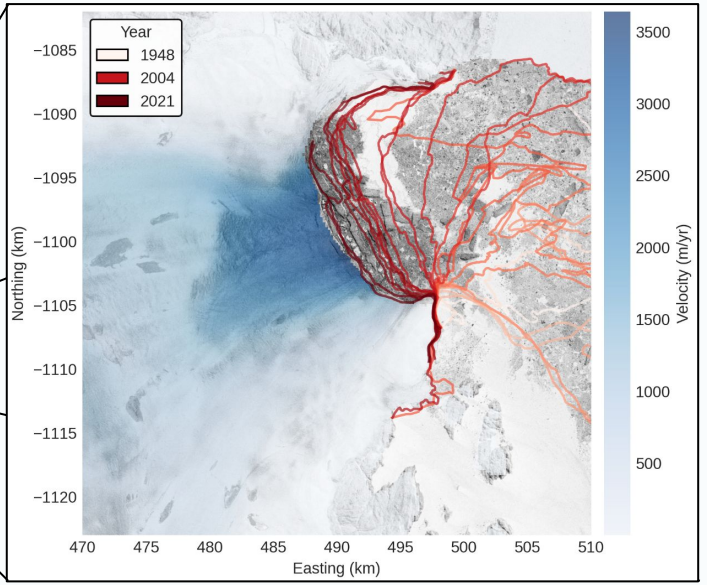


Copernicus Climate Change Service (2020)
Rignot & Mouginot (2012)

Background - ZI

- Ice tongue **collapsed in 2013**, leading to dramatic speedup



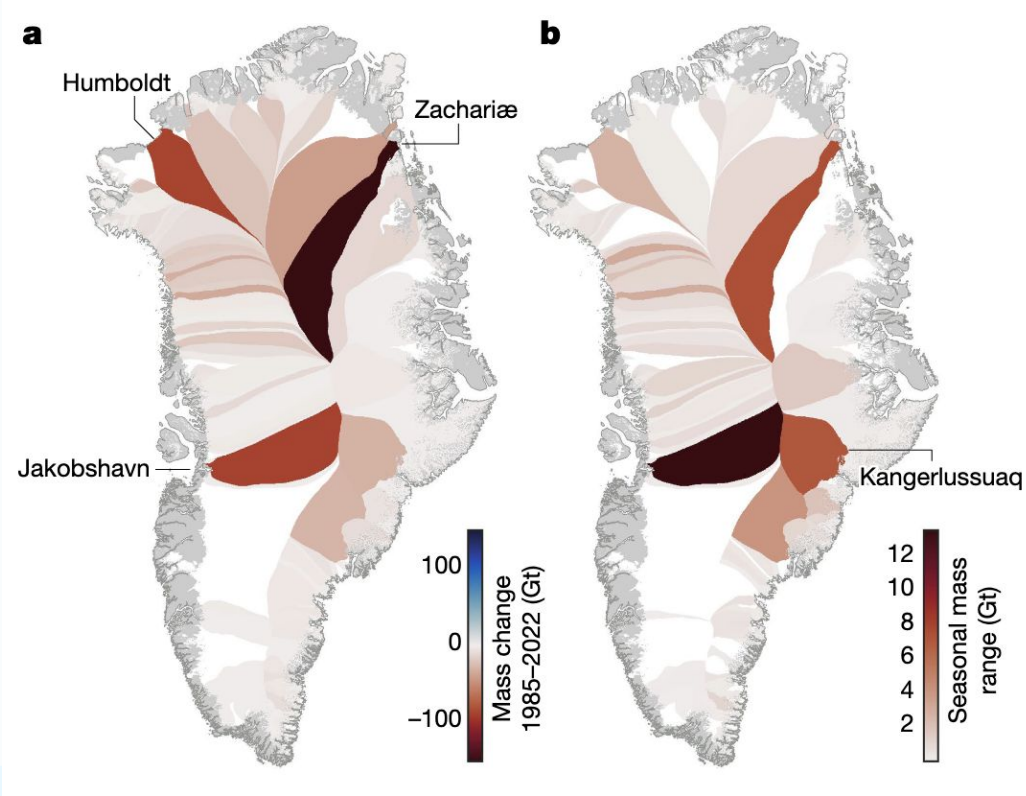


ZI's velocity (blue) and terminus over time (red)

Copernicus Climate Change Service (2020)

Seasonality at NEGIS

- Higher seasonality correlates with larger long-term mass loss
- NEGIS has **lost the most mass** and has the second most **seasonal mass change**



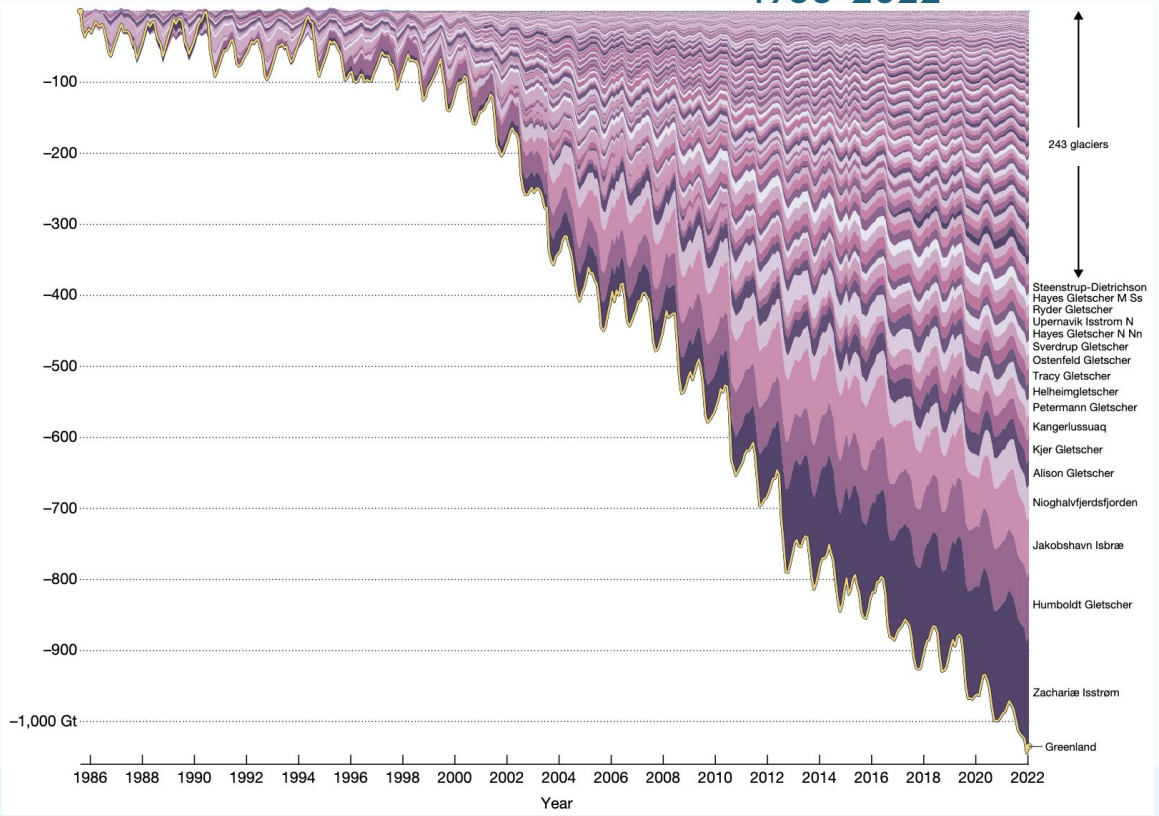
a. Cumulative mass change by catchment b. Seasonal mass change by catchment

Greene et al. (2024)

Seasonality at ZI

- ZI has the **highest cumulative mass loss**
- Summer **meltwater discharge** and **mélange** (sea ice) buildup might contribute to seasonality

Cumulative mass change
1986-2022



Greene et al. (2024)

Goals

Tier 1

Predict velocity values at 3 points on ZI

Tier 2

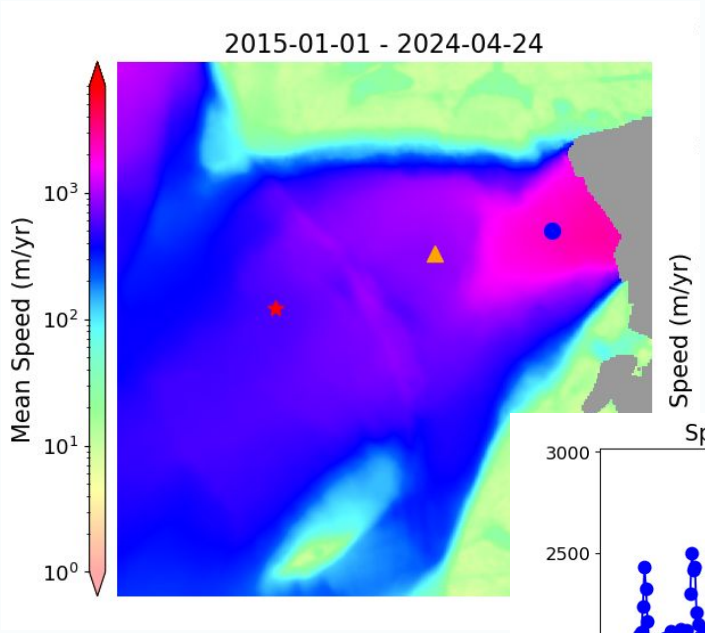
Predict velocity along a flowline

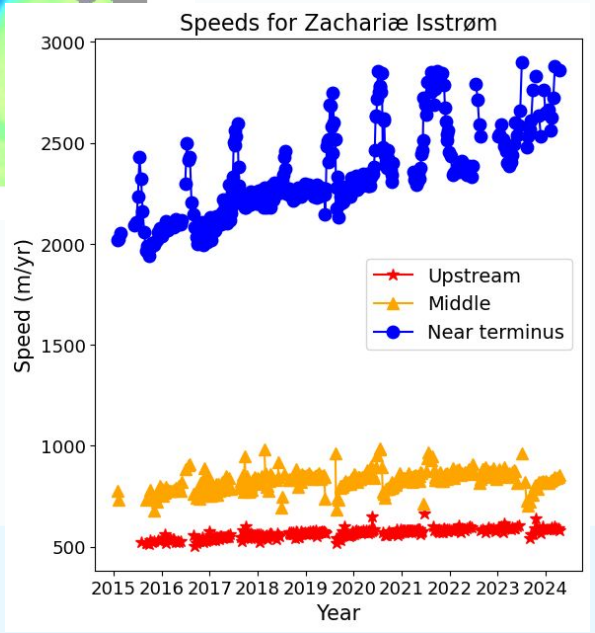
Tier 3

Predict velocity at entire glacier

Data

- Satellite-derived velocity estimates from GrIMP at 6 or 12 day frequency
- Picked 3 points: upstream, middle, terminus
  - GrIMP promotes open access!

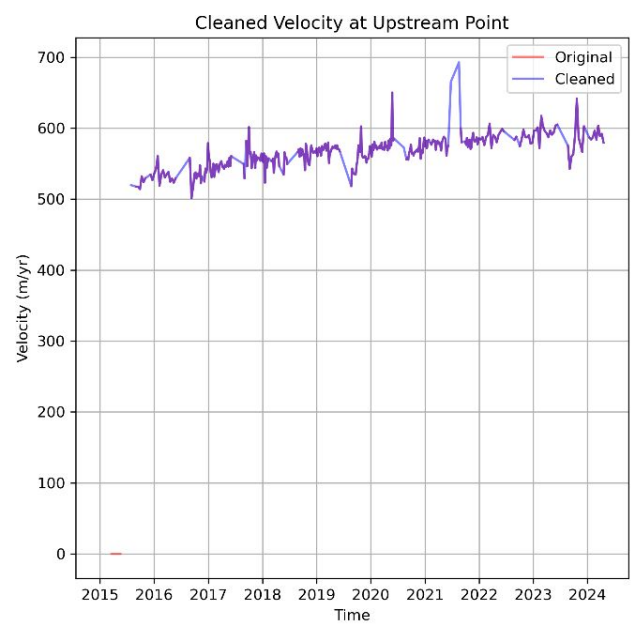


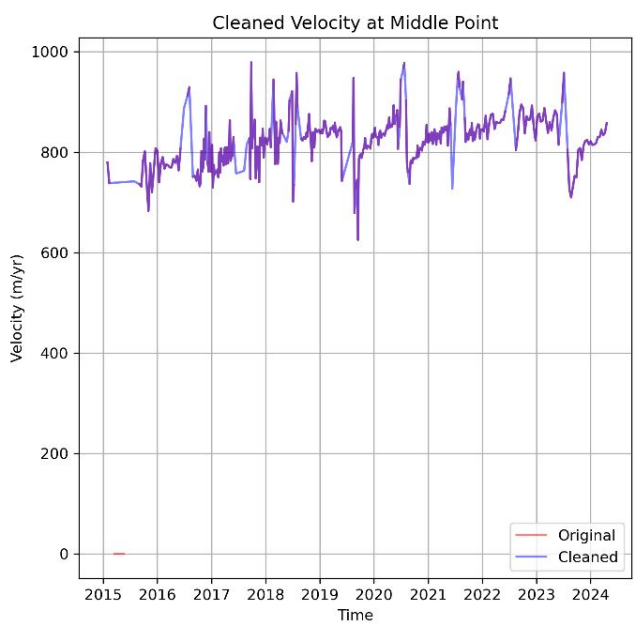


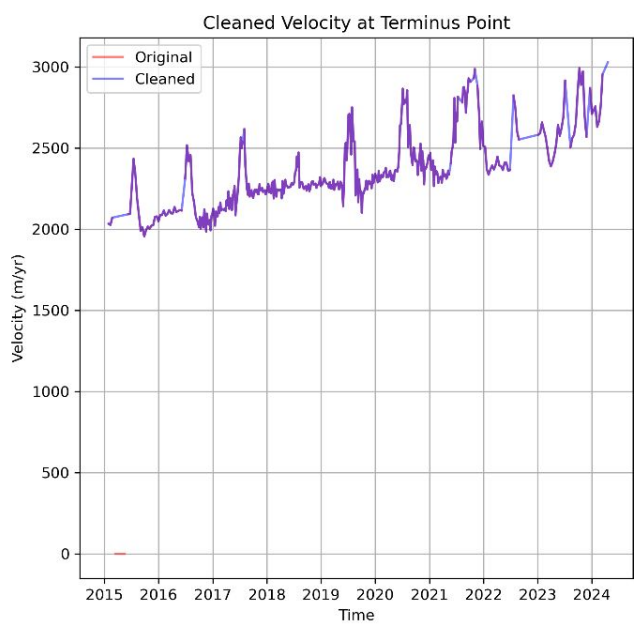
Preprocessing

- NaNs and zeroes were removed from each point

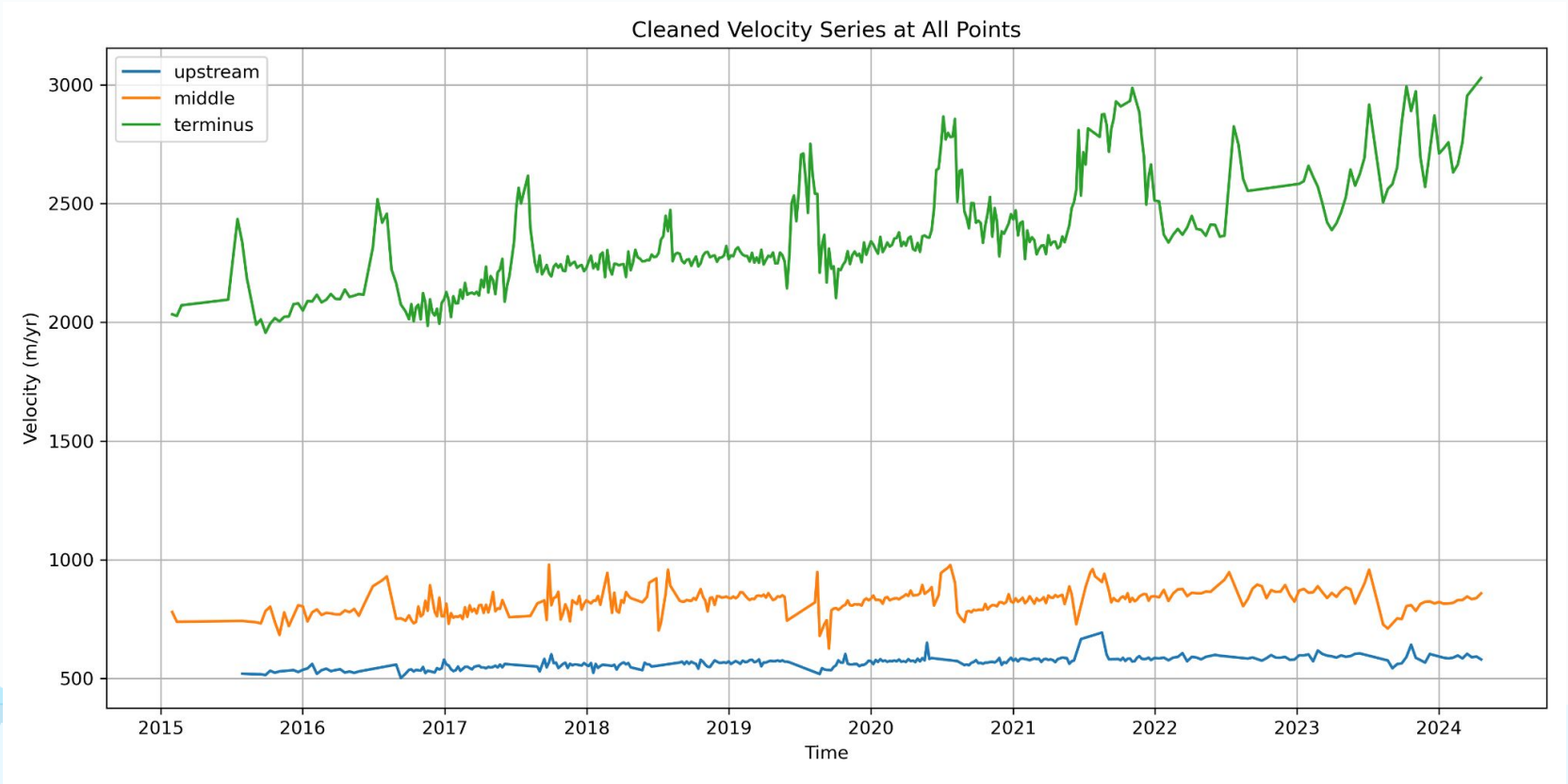
Look at the scale bars!!





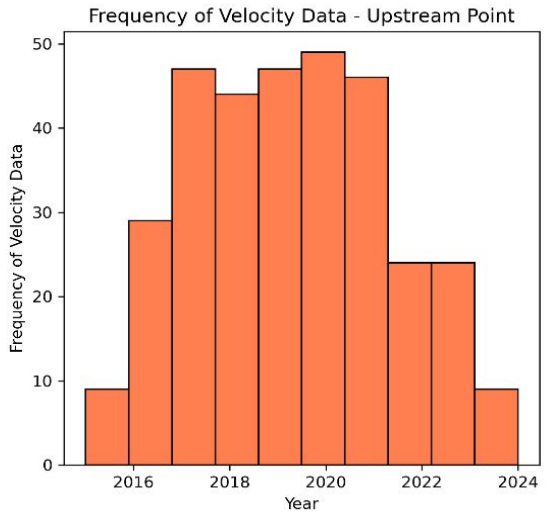


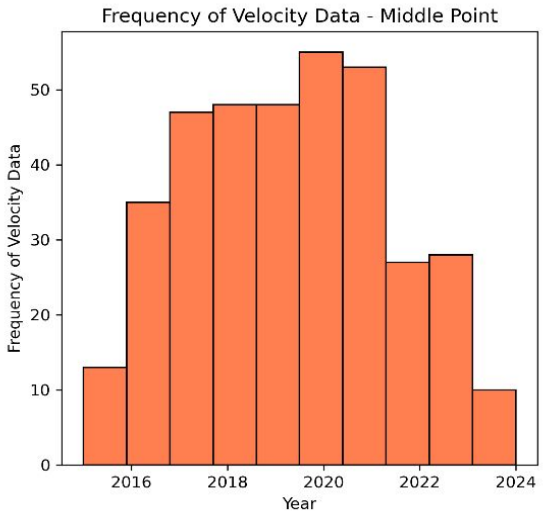
Preprocessing

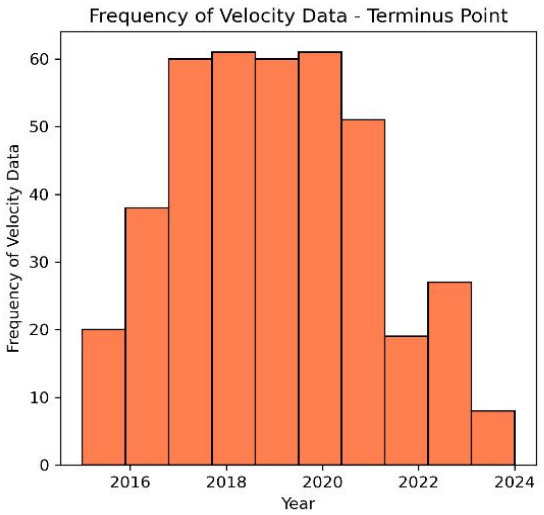


Data

- Most data where Sentinel-1A and Sentinel-1B overlap (~2016-2022)

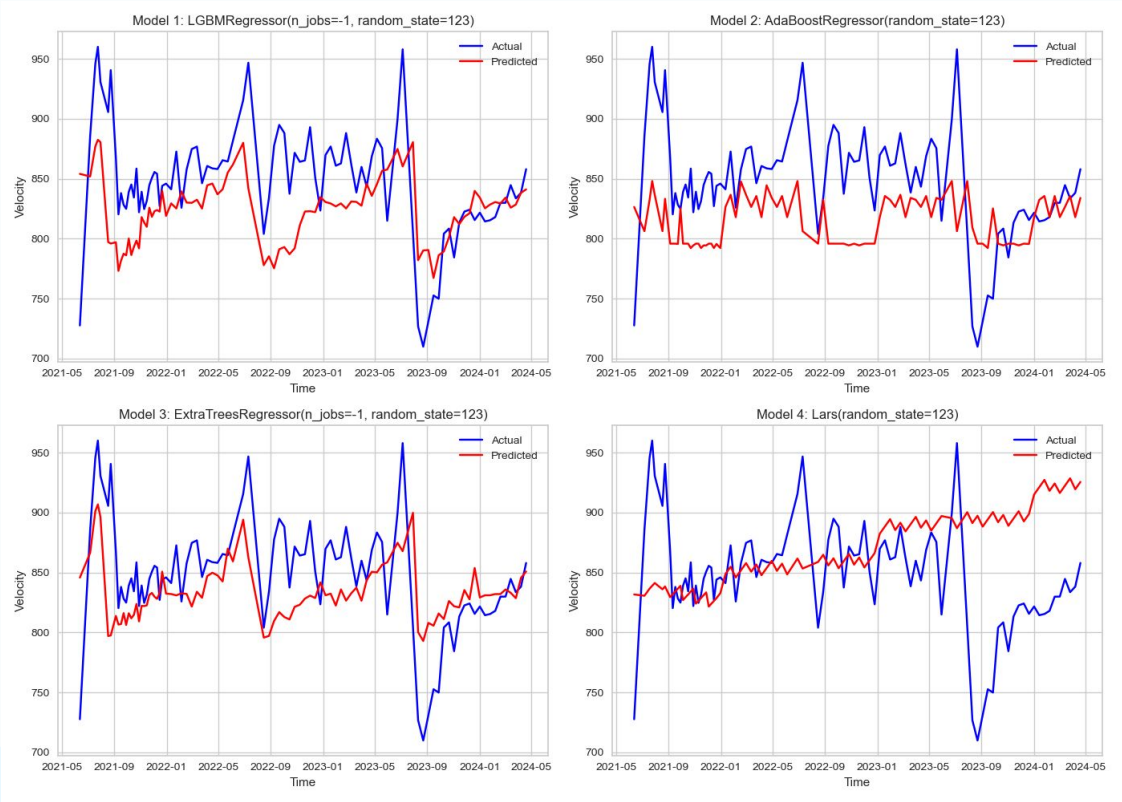






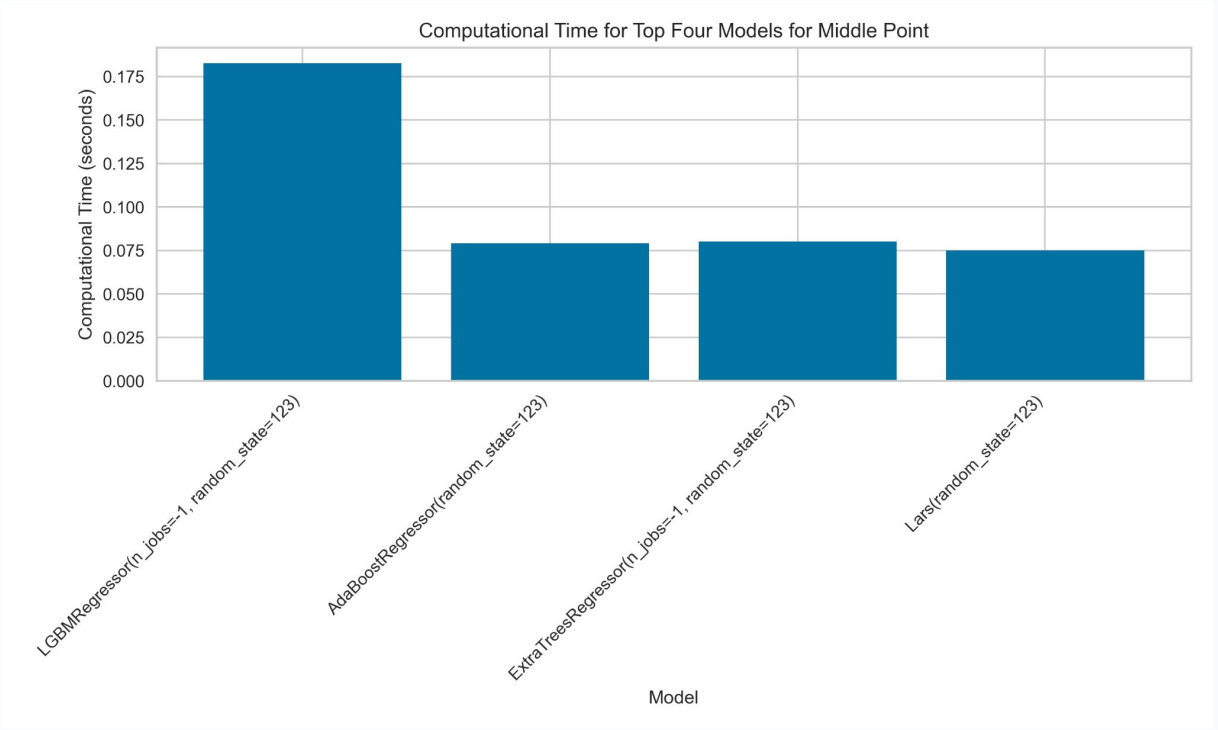
CML - Hyperparameter Tuning

- Used PyCaret library to select optimal model for each point
- LightGBM and ExtraTrees **capture seasonal trends**
- Tuned and chose ExtraTrees (splits randomly rather than optimally)



CML - Computational Time

- LightGBM was slowest
- Others are relatively similar
  - Not much of a concern right now...

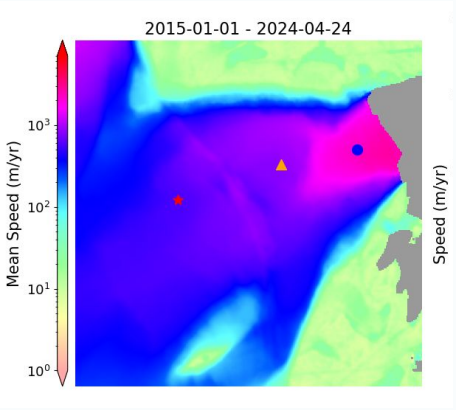


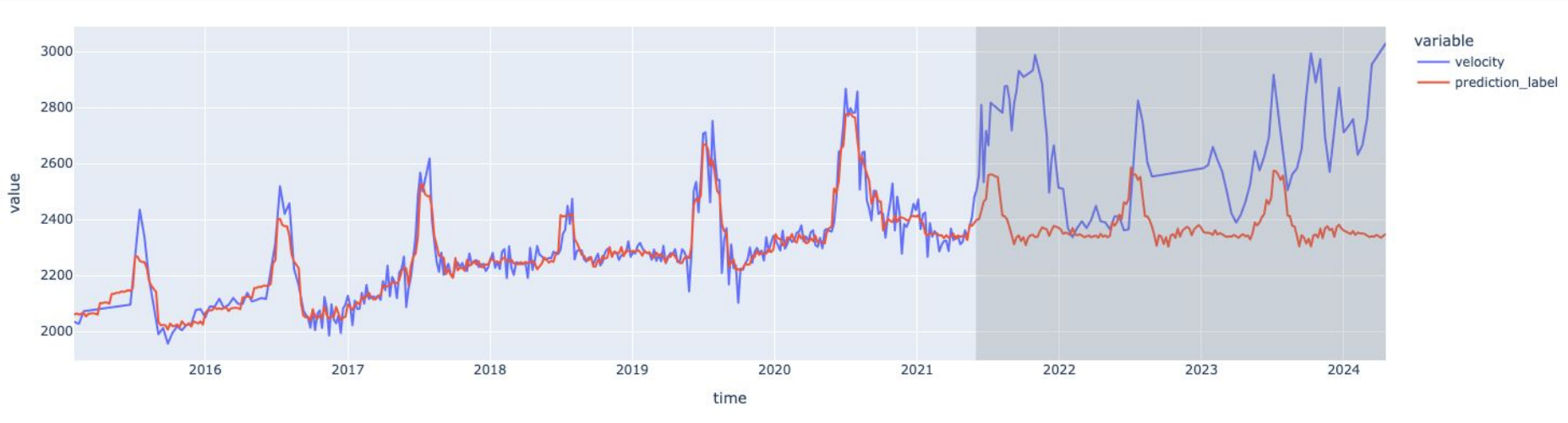
CML - Train/Test

- Train (red) and test (red, highlighted gray) vs. ground truth (blue)
- Paying attention to seasonality

CML - Train/Test (terminus)

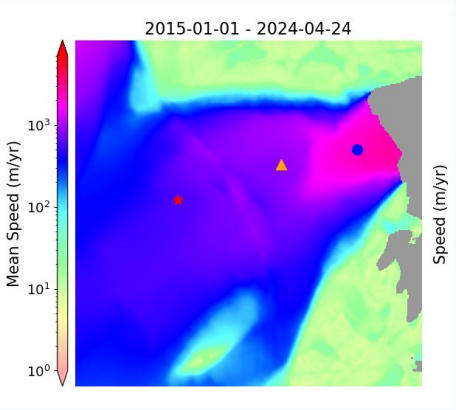
- Train (red) and test (red, highlighted gray) vs. ground truth (blue)





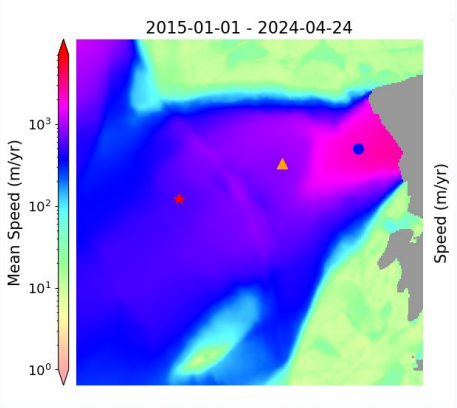
CML - Predictions (middle)

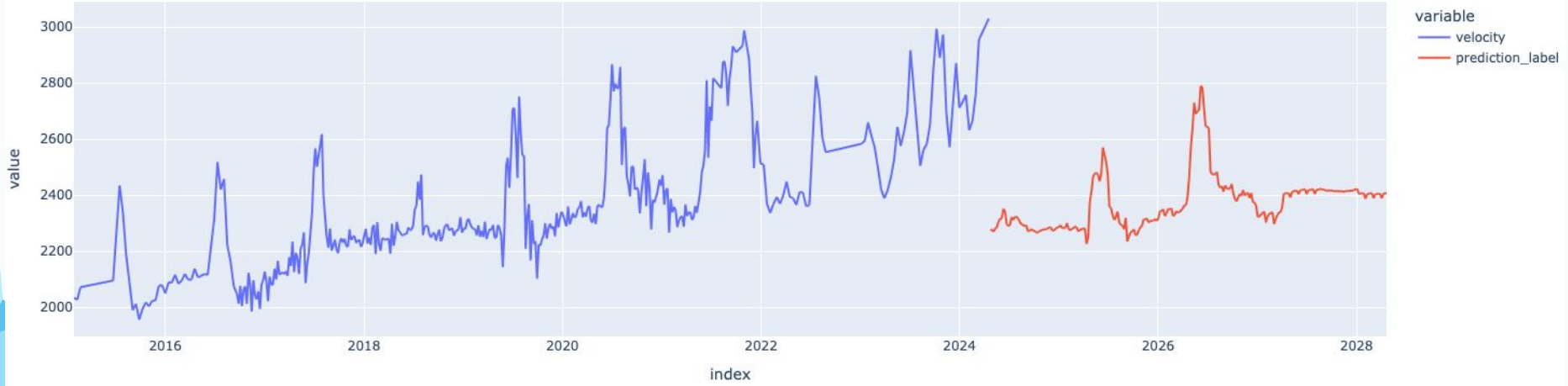
- Used **ExtraTrees** to predict values 2 years in the future
- Prediction (red) vs. ground truth (blue)



CML - Predictions (terminus)

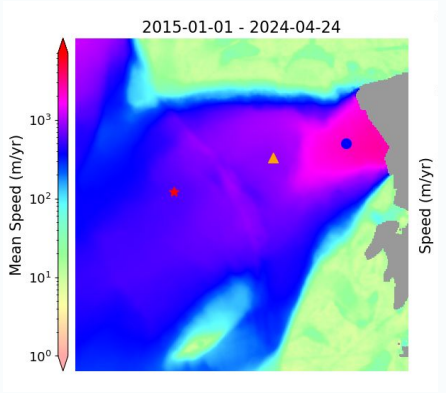
- Used **LightGBM** to predict values 2 years in the future
- Prediction (red) vs. ground truth (blue)

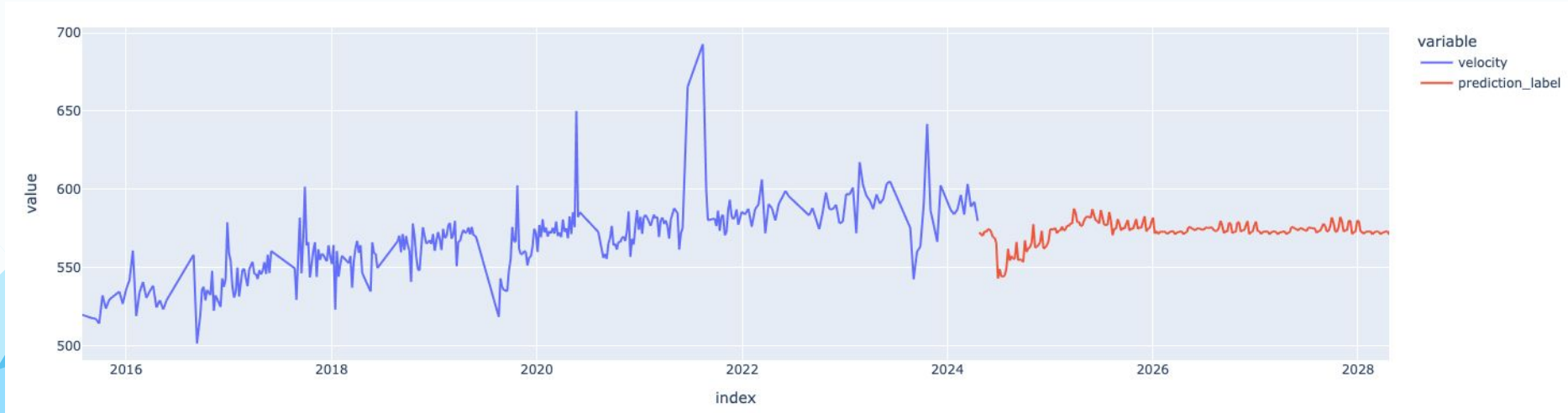




CML - Predictions (upstream)

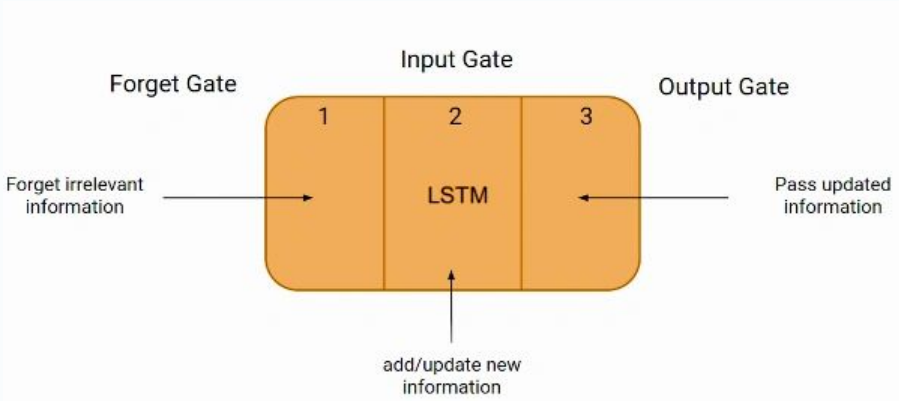
- Used **RandomForest** to predict values 2 years in the future
- Prediction (red) vs. ground truth (blue)





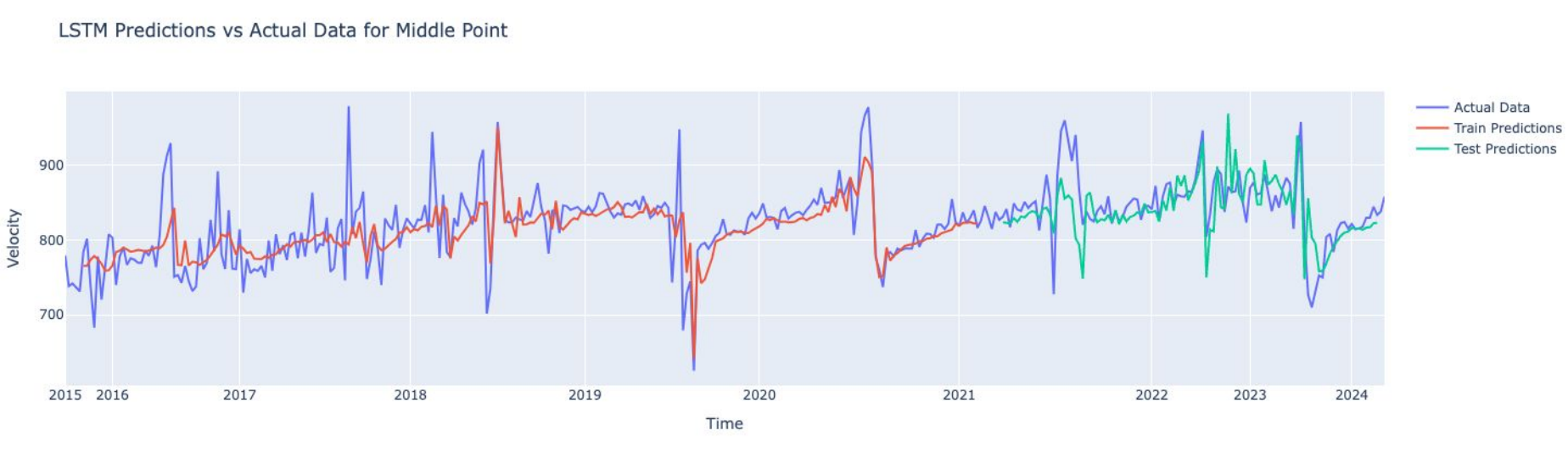
DL - LSTM

- LSTM uses input, forget, output gates to store memory
- Must choose optimal lookback (memory length) for predictions
- Trained 500 epochs with 0.2 dropout, lookback of 5



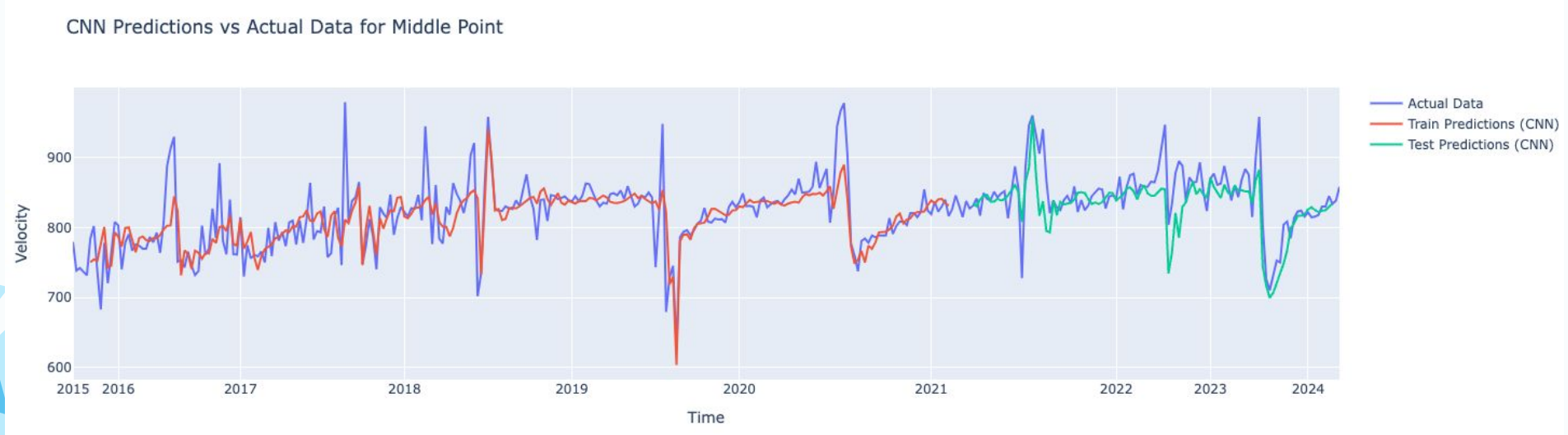
DL - LSTM

- Trained (red) and tested (green) on ground truth (blue)



DL - CNN

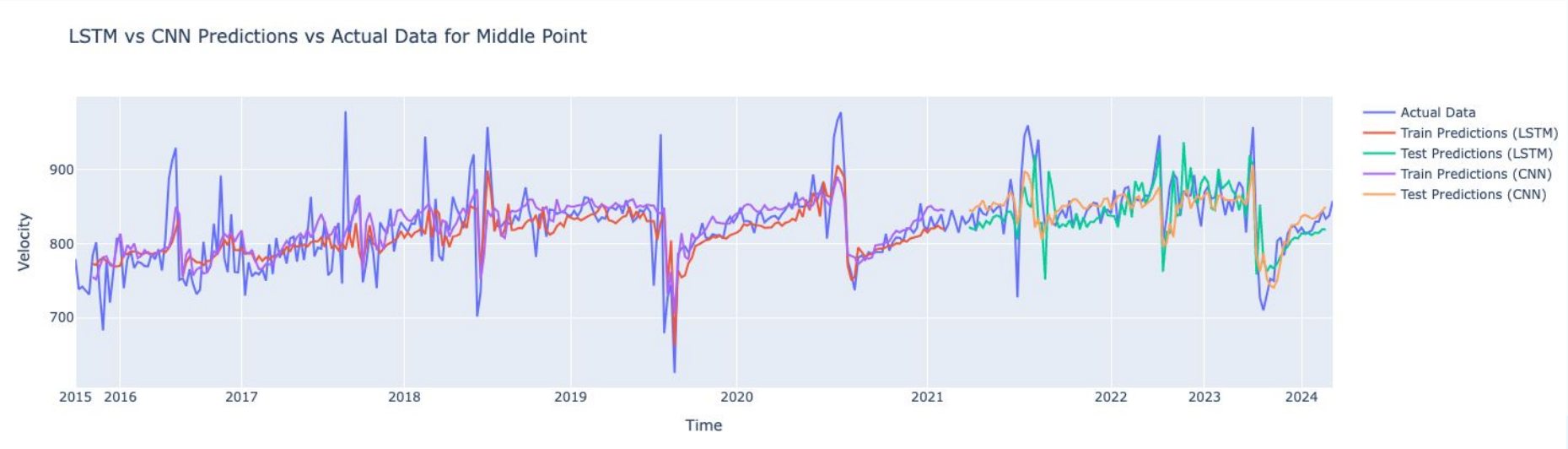
- Trained (red) and tested (green) on ground truth (blue)



DL - Comparison

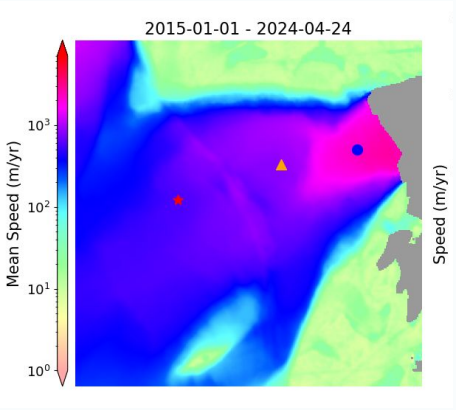
LSTM Train Score: 36.88 RMSE
LSTM Test Score: 41.10 RMSE
CNN Train Score: 821.66 RMSE
CNN Test Score: 843.57 RMSE

Horrible RMSE scores but do predictions capture seasonality?



DL - Comparison

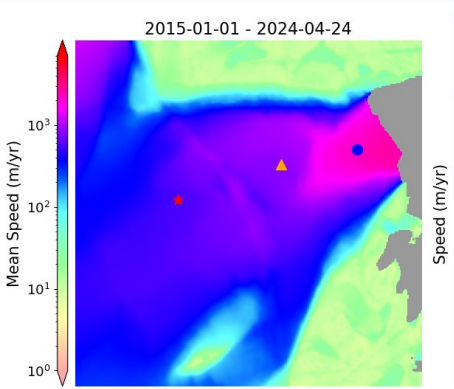
LSTM Train Score: 36.88 RMSE
LSTM Test Score: 41.10 RMSE
CNN Train Score: 821.66 RMSE
CNN Test Score: 843.57 RMSE

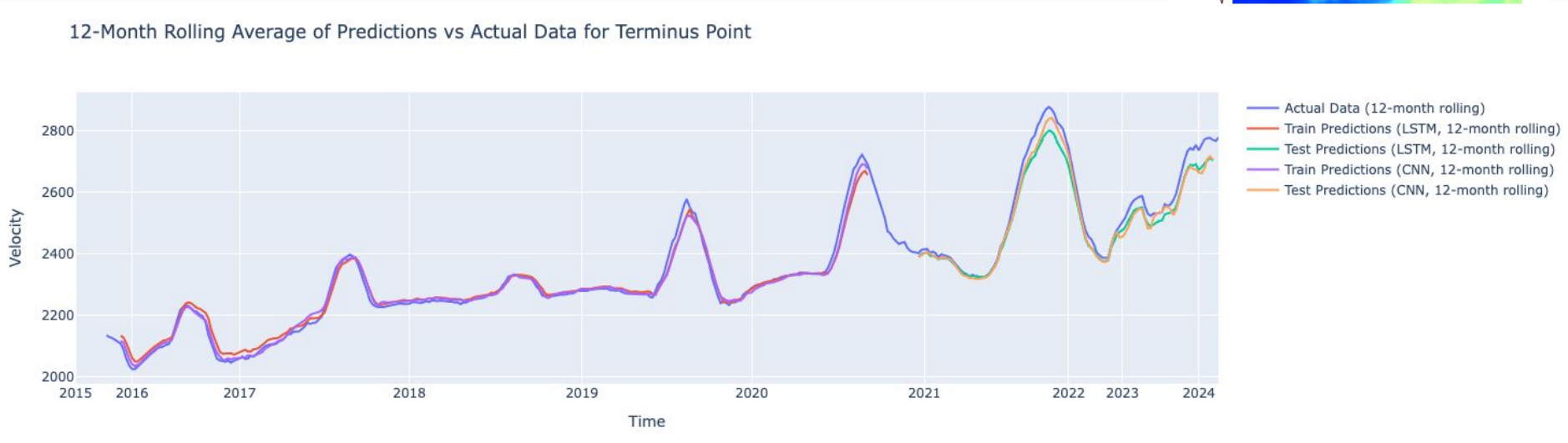


DL - Comparison

For the terminus point, yes!

... but is it overfitting??

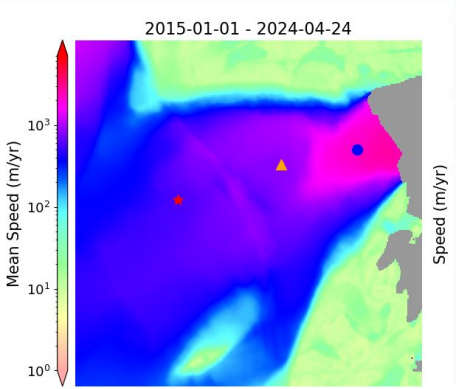


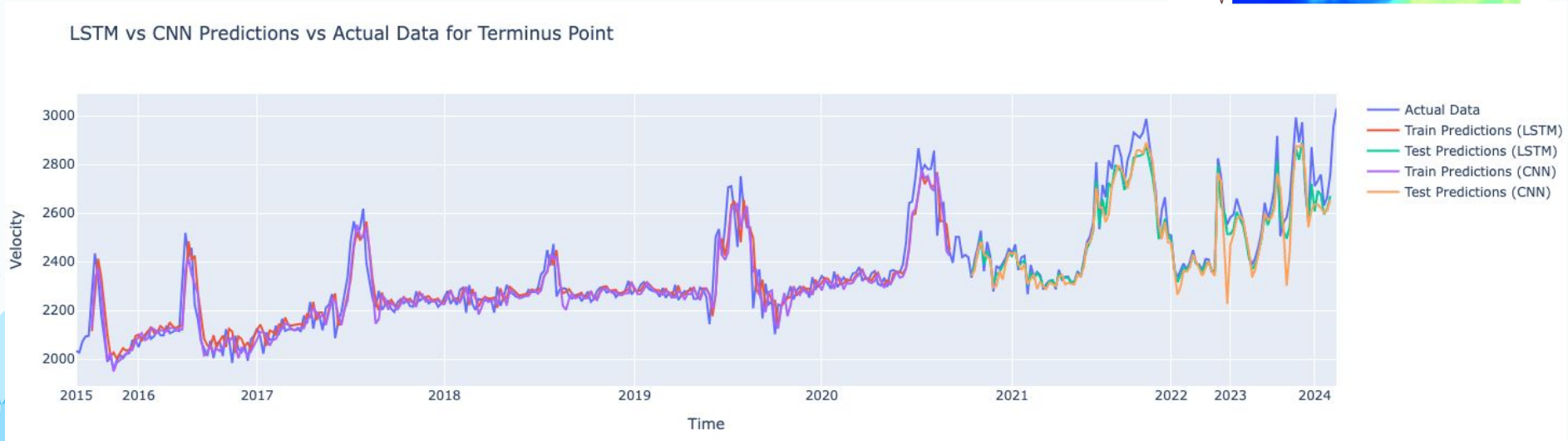


DL - Comparison

For the terminus point, yes!

... but is it overfitting??





Scaling Up

Exogenous Variables

Incorporate temperature, meltwater runoff, etc.

Frequency

Test monthly predictions

Spatial Coverage

Incorporate many glaciers or many points along flowline

If different models perform better on different parts of the glacier, can one accurately describe the entire glacier??

Python library to automate point extraction, training, model dev?

Scaling Up

Exogenous Variables

Incorporate temperature, melt water runoff, etc

Frequency

Test monthly predictions

Spatial Coverage

Incorporate many glaciers or many points along flowline

If different models perform better on different parts of the glacier, can one accurately describe the entire glacier??

Python library to automate point extraction, training, model dev?

FIX OVERFITTING