Water Level Predictions through AI Models

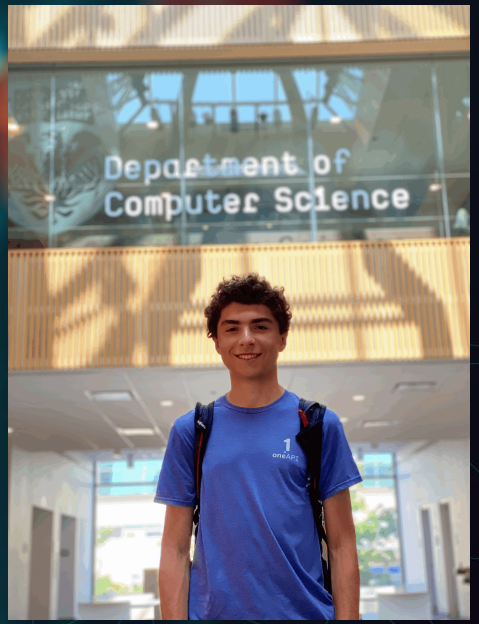
Lucas Raicu
Advisors: Daniel Grizenda, Prof. Kyle Chard
July 1st, 2025

globus labs

About me

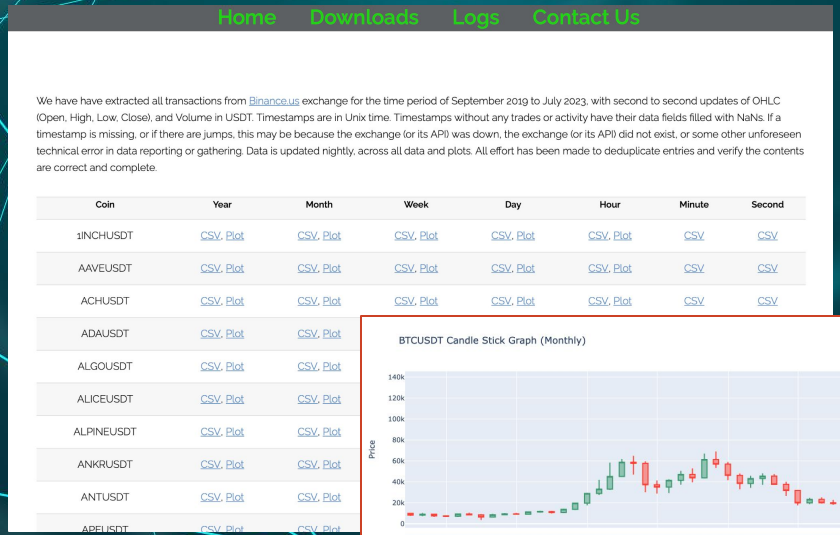
- Rising Senior at Glenbrook South High School
- Prior Knowledge

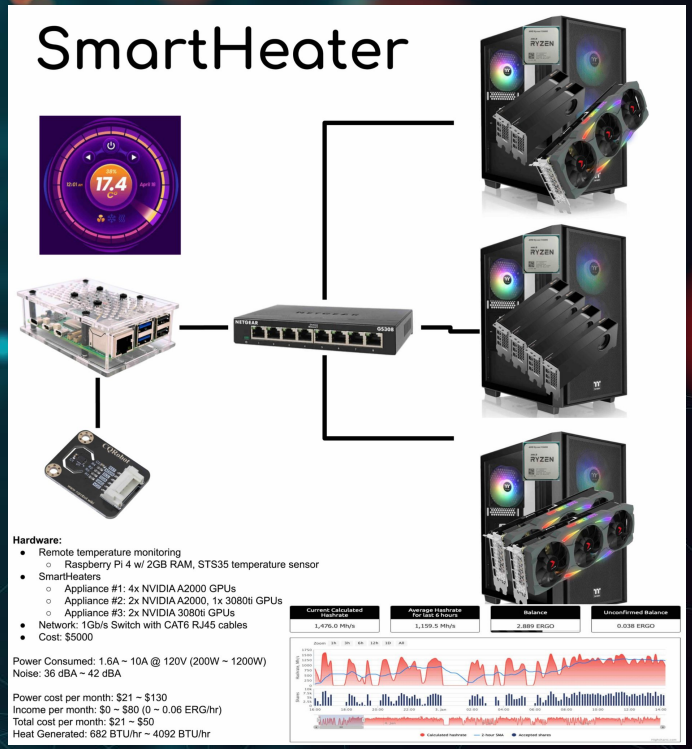
- Coursework
  - Honors CS and AP CS A
  - Fundamentals of AI through Oakton College



Personal Website:
https://lucasraicu.github.io

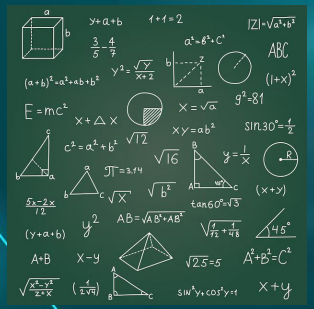
Past Research Projects





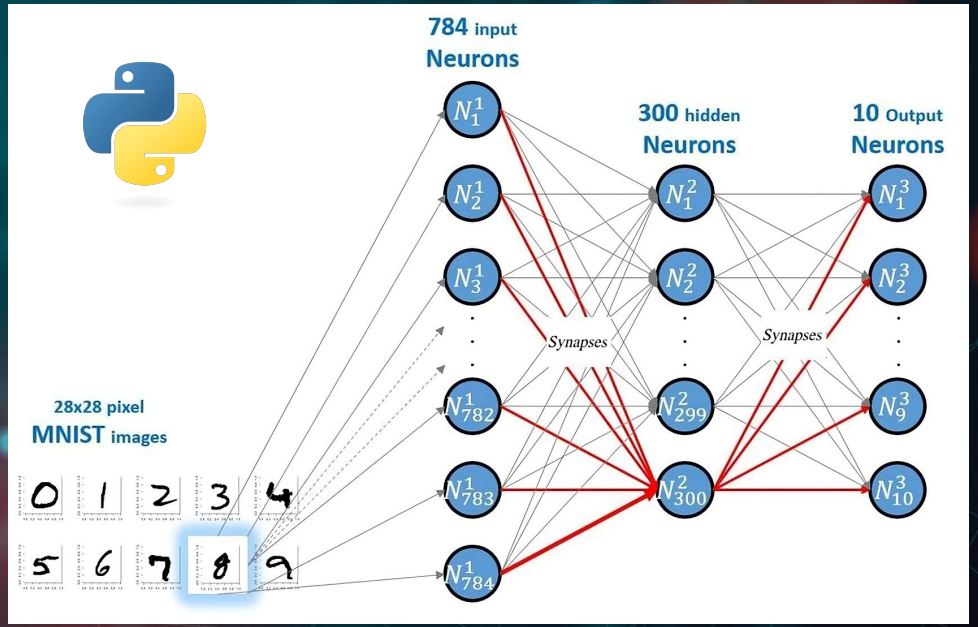
Research Link:
http://crypto.cs.iit.edu/datasets

Preliminary Work (MNIST)

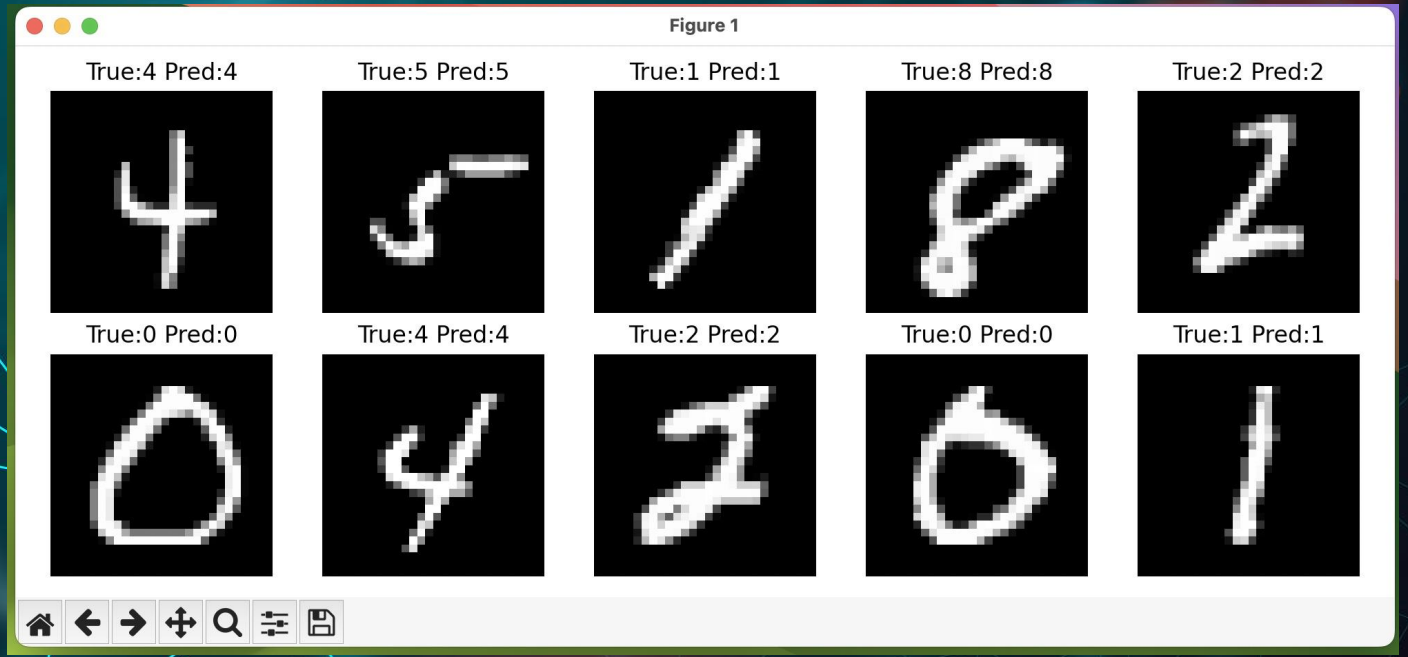


Methods

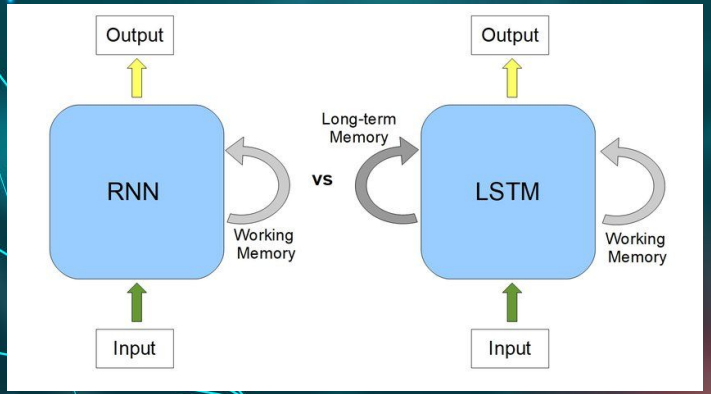
PyTorch

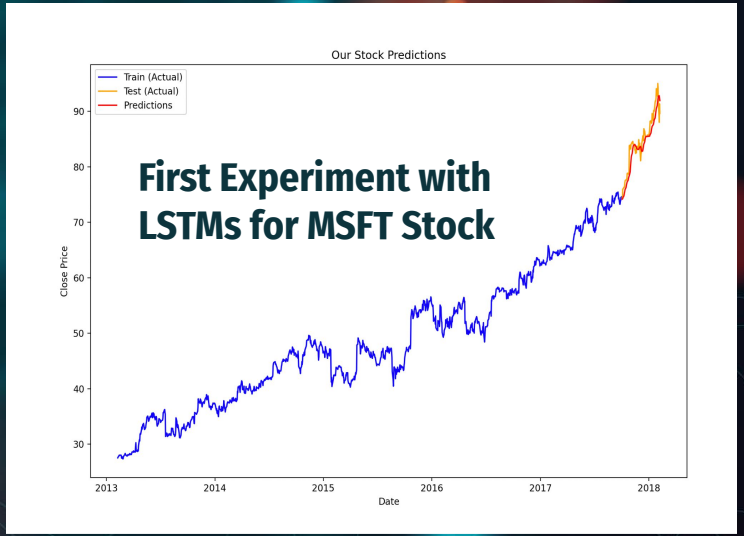


Preliminary Work (MNIST)



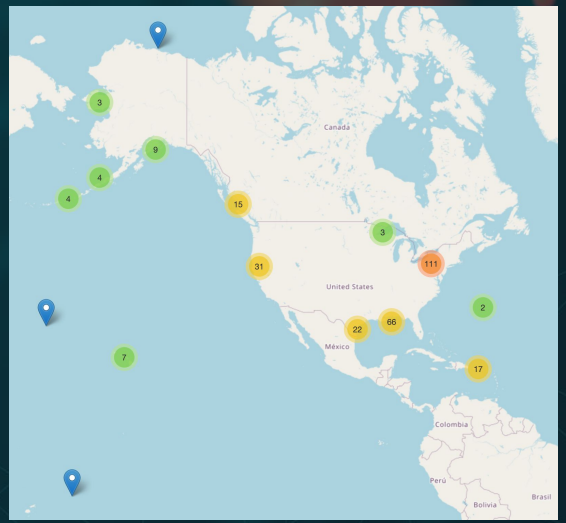
LSTMs as an Option



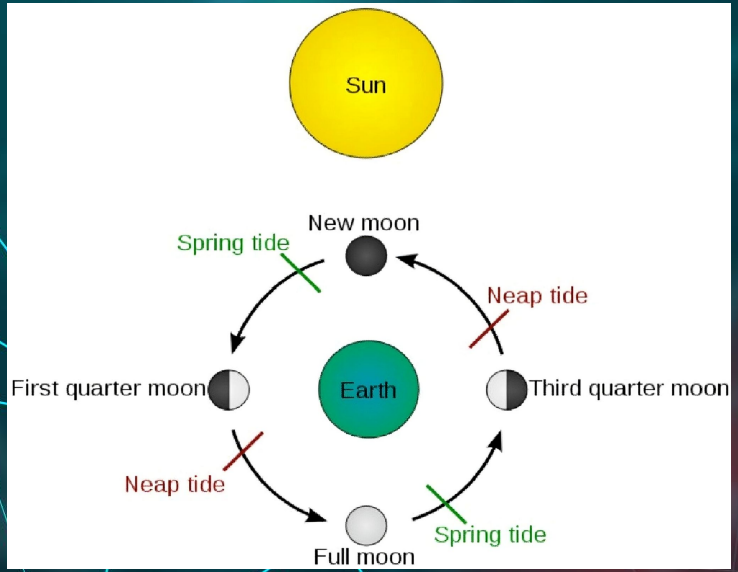


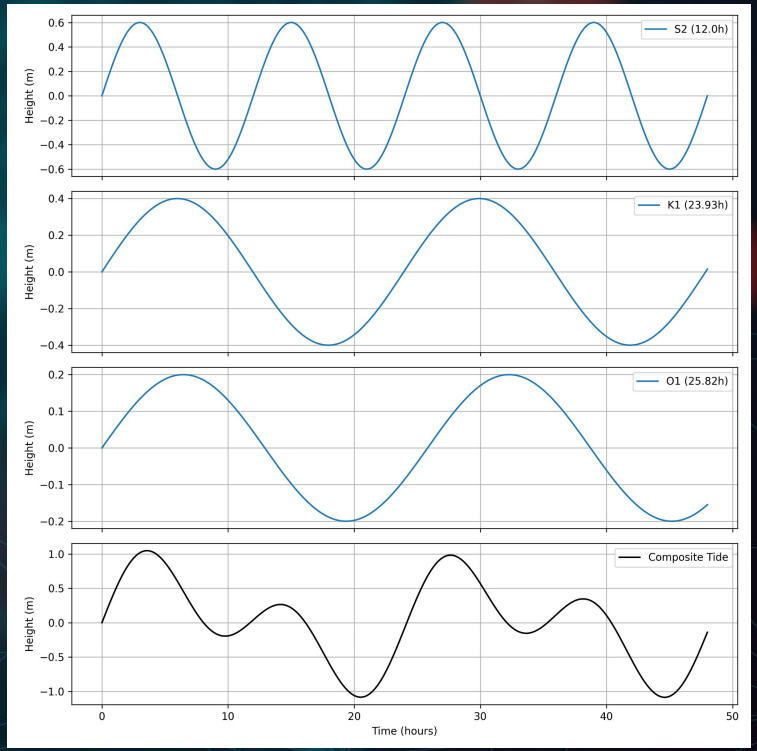
Project Overview

- Goal: temporal-spatial water-level data -> machine learning -> accurate future trend predictions



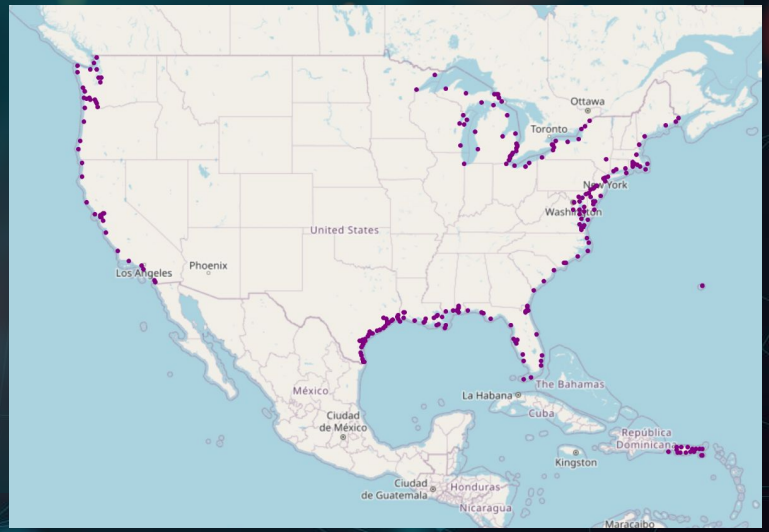
Harmonic Analysis





Fountain Framework

- Github Repository
  - Data (82gb)
    - 217 coastal water level stations
    - Over 127 million data points (6-minute granularity)
  - Scripts
    - Training and results
    - Visualization



Data Format (FOUNTAIN)

Stations Info (tsv)

~/Documents/summer-2025/fountain/data/processed/noaa/stations/stations.tsv lucas

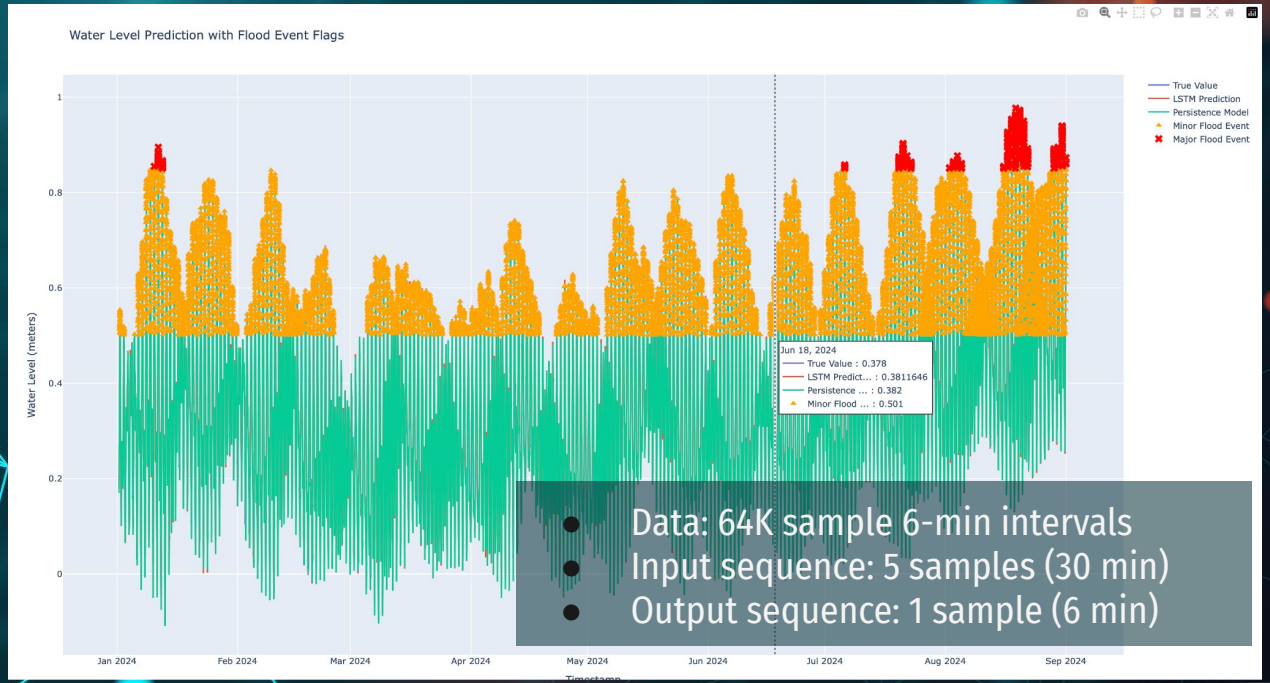
	tidal	greatlakes	shefcode	state	timezone	timezonecorr	id	name	lat	lng	affiliations	tideType	established	removed	origyear	missing_percent
1	True	False	NWWH1	HI	HAST	-10	1611400	Nawiliwili	21.9544	-159.3561	NWLON	Mixed	1954-11-24 00:00:00.0	1991-02-16 00:00:00.0		1.6903284477206767e-06
2	True	False	OOUH1	HI	HAST	-10	1612340	Honolulu	21.303333	-157.864528	NWLON	Mixed	1905-01-01 00:00:00.0	1989-01-20 00:00:00.0		0.0
3	True	False	PRHH1	HI	HAST	-10	1612401	Pearl Harbor	21.3675	-157.963898	PORTS	Mixed	2023-05-31 00:00:00.0	2023-05-31 00:00:00.0		
4	True	False	MOKH1	HI	HAST	-10	1612480	Mokuoloe	21.433056	-157.79	NWLON	Mixed	1957-05-03 00:00:00.0	1989-01-21 00:00:00.0		0.0
5	True	False	KLIH1	HI	HAST	-10	1615680	Kahului, Kahului Harbor	20.895	-156.469167	NWLON	Mixed	1946-12-19 00:00:00.0	1989-01-29 00:00:00.0		0.0

Sorted Station Data (.tsv) - Nawiliwili

~/Documents/summer-2025/fountain/data/lucas/model_data_less_than_20_missing_no_igld_no_dups_1611400_sorted.tsv lucas

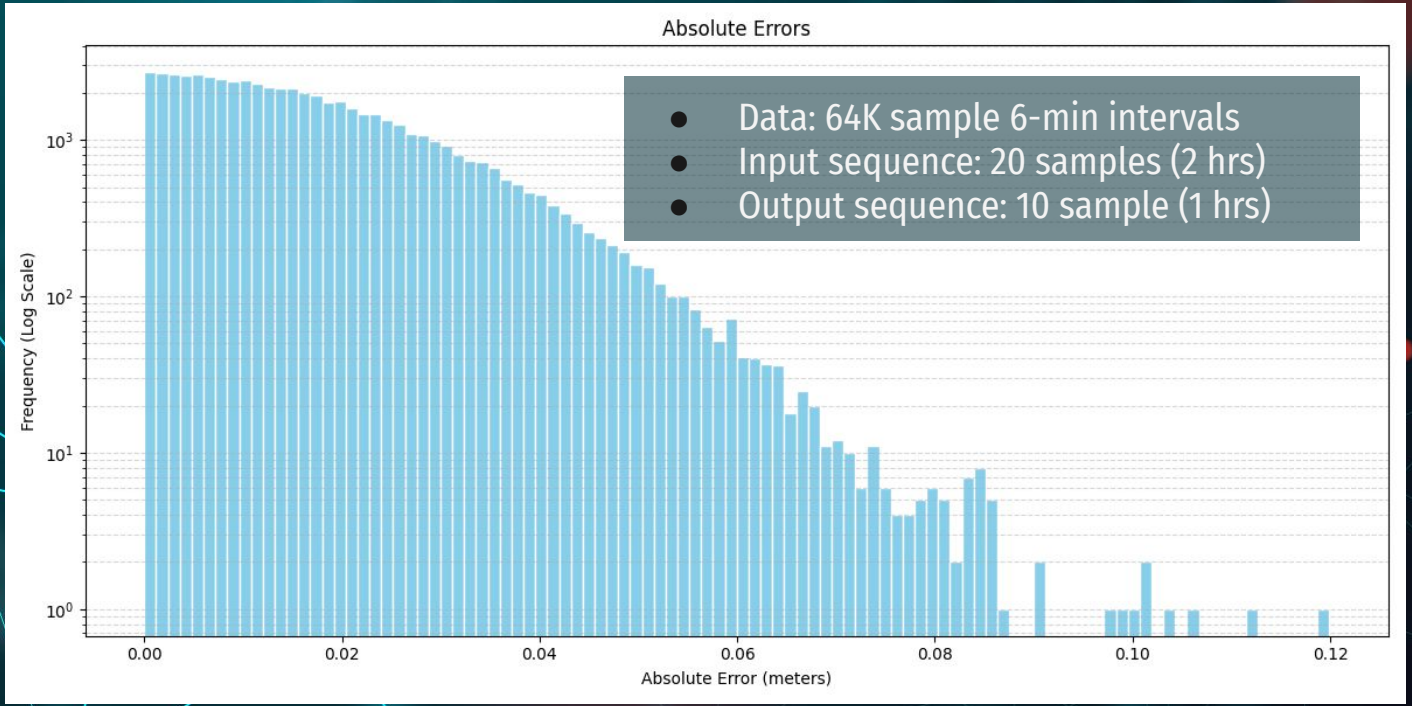
	time	value	sigma	quality	inferred		flat		roc	threshold	station	datum
1	time	value	sigma	quality	inferred		flat		roc	threshold	station	datum
2	2018-01-01 00:00:00	0.273	0.002	v	0	0	0	0	1611400	mllw		
3	2018-01-01 00:06:00	0.278	0.003	v	0	0	0	0	1611400	mllw		
4	2018-01-01 00:12:00	0.277	0.003	v	0	0	0	0	1611400	mllw		
5	2018-01-01 00:18:00	0.276	0.006	v	0	0	0	0	1611400	mllw		

Visualization (Nawalili)



- Data: 64K sample 6-min intervals
- Input sequence: 5 samples (30 min)
- Output sequence: 1 sample (6 min)

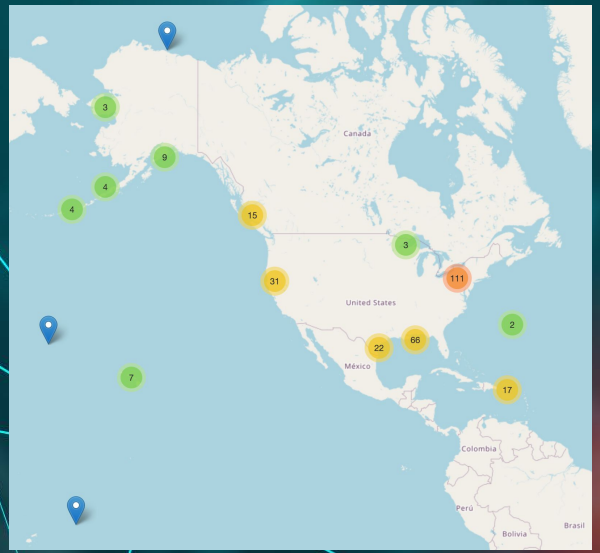
Visualization (cont.)

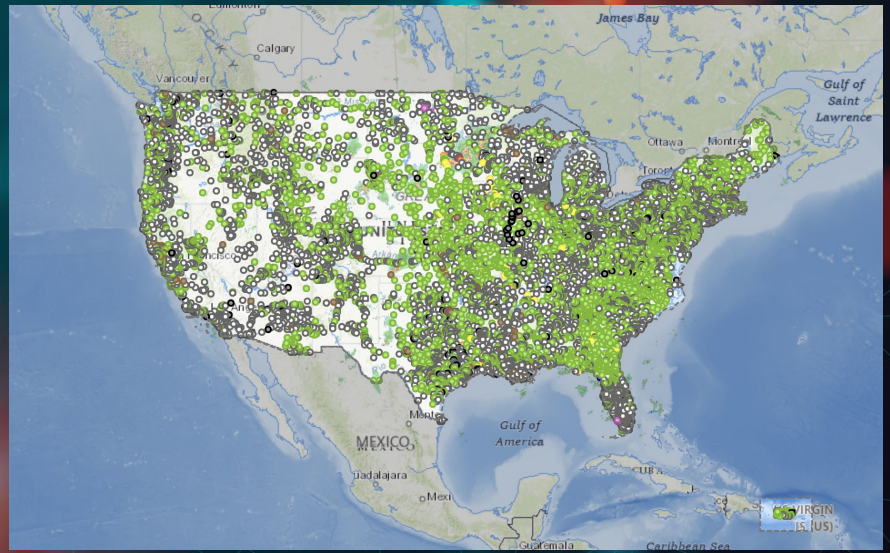


Next Steps

- LSTM Parameters
  - Hidden Layers (2), inputs (input sequence), outputs (1)
    - Input sequence (5, 240, 720, 1200, 1680): 30-min, 1-day, 3-day, 5-day, 7-day (in 6 min samples)
    - Output sequence (1): 6-min, 1-day, 3-day, 5-day, 7-day
  - Loss functions
    - MSE: Mean-Squared Error
    - RMSE: Root Mean Squared Error
- Testbed
  - Hardware: MacOS, 6-core Intel CPU 2.6GHz, 16GB RAM
  - Time to train 500K 6-min samples for two epochs: 2hrs
  - Will need more compute resources: Chameleon, Mystic, etc
- Other Models
  - Dynamic Graph Neural Networks (DGNN)

Long-term Goals:
Coastal vs River Predictions





Questions