Taking The Pulse Of The Great Lakes: Using Gaussian Neural Processes To Identify Optimal Observing Sites

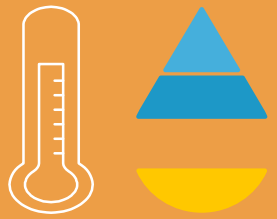
Erin Redding

Mentors: Dani Jones, Russ Miller, Shelby Brunner

UC Berkeley, May 2025

B.A. Data Science

Minor in Conservation & Resource Studies



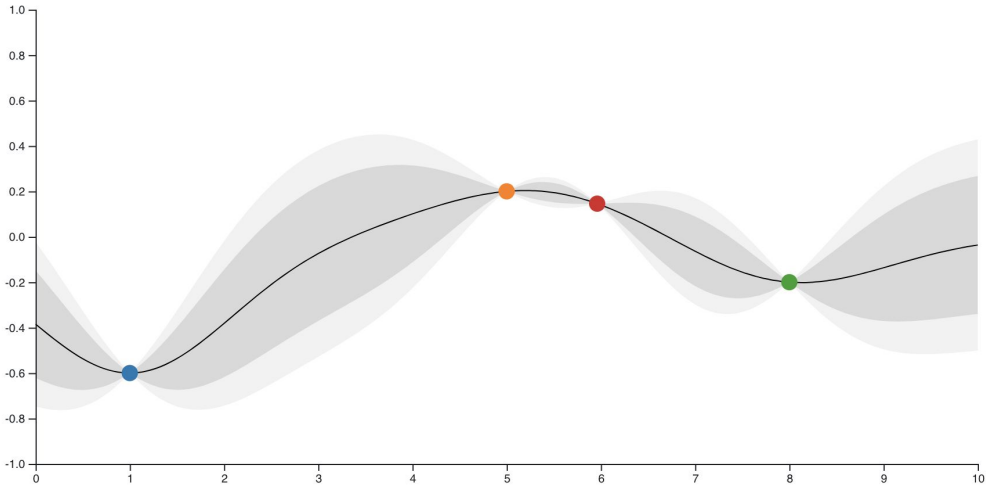
Research Question

Where should the next generation of temperature measurement sensors be placed in order to most efficiently improve our quantitative understanding of Great Lakes surface temperature variability?

What is a Gaussian Neural Process

Gaussian Processes

$$\mathbf{X} \sim \mathcal{N}(\mu, \Sigma)$$



source

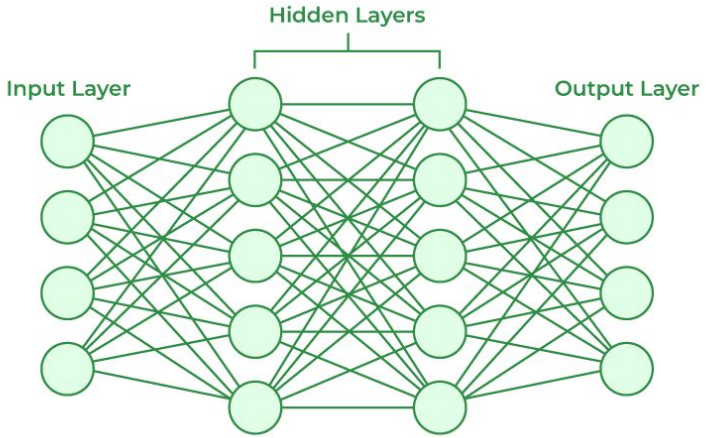
What is a Gaussian Neural Process

Gaussian

Neural

Processes

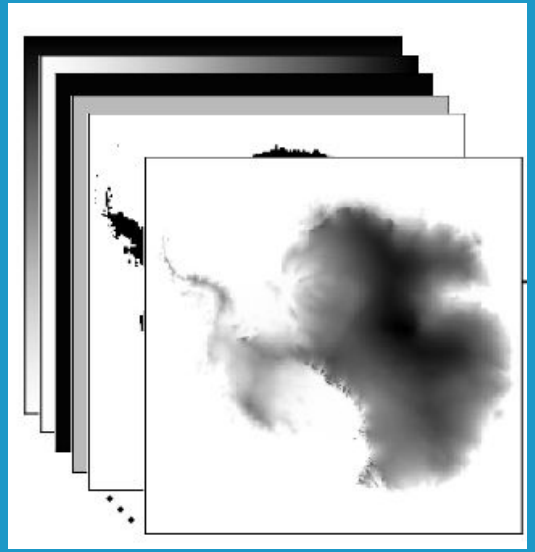
$$\mathbf{X} \sim \mathcal{N}(\mu, \Sigma)$$



Previous Research

Tom Andersson et al. **Environmental Sensor Placement with Convolutional Gaussian Neural Processes**. Environmental Data Science (2023)

- Fuse data sets
  - ERA5, Elevation, DOY cycle, etc.
- Convolution Gaussian Neural Process(ConvGNP) Model

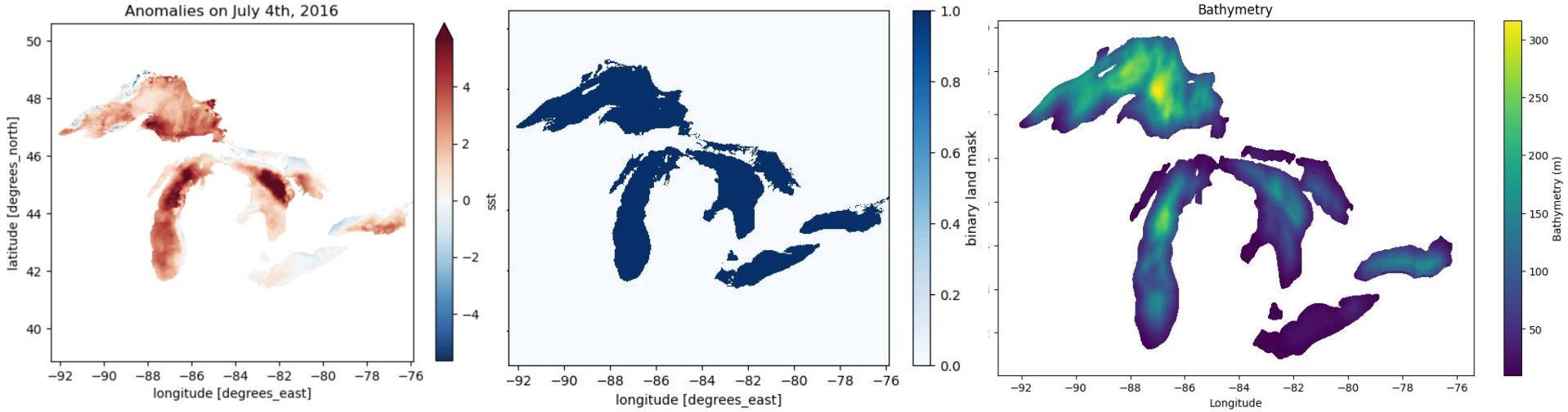


How do we find optimal sensor sites?

The Alan Turing Institute

DeepSensor

Great Lakes DeepSensor Use Case



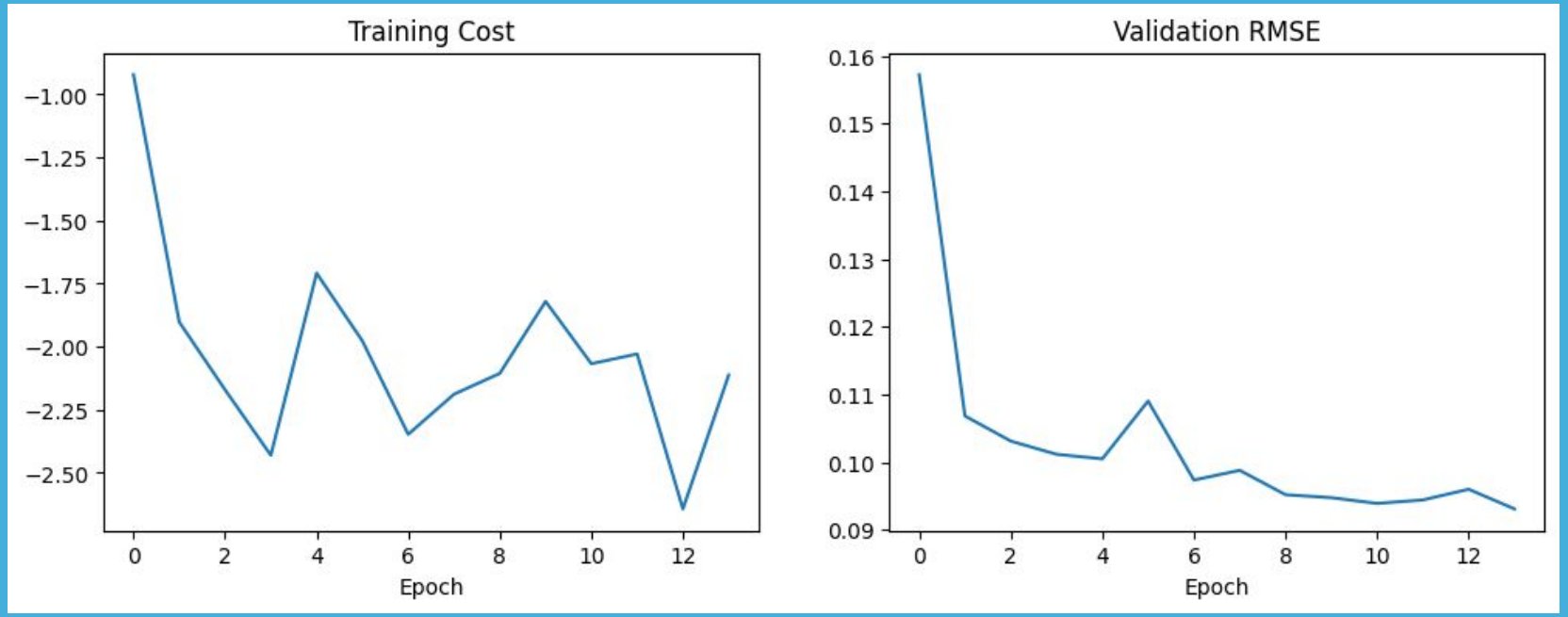
GLSEA3 temperature anomaly

Lake Mask

Bathymetry

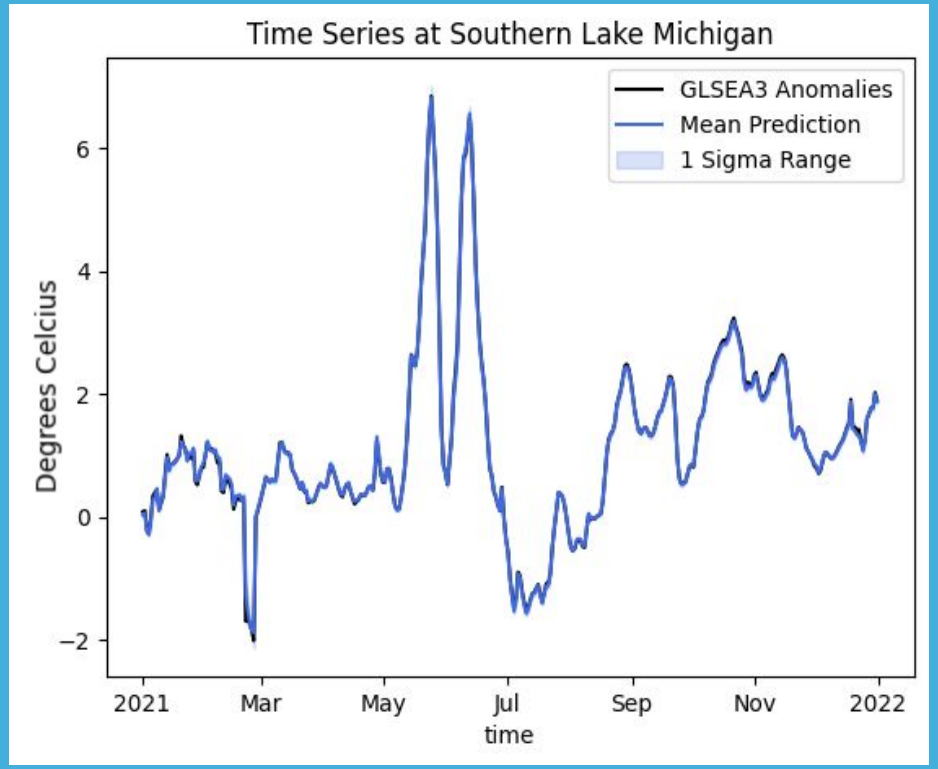
Train: 2007-2014, Validation 2015-2016, Test: 2017-2021

1.) Using All Points, Many Days

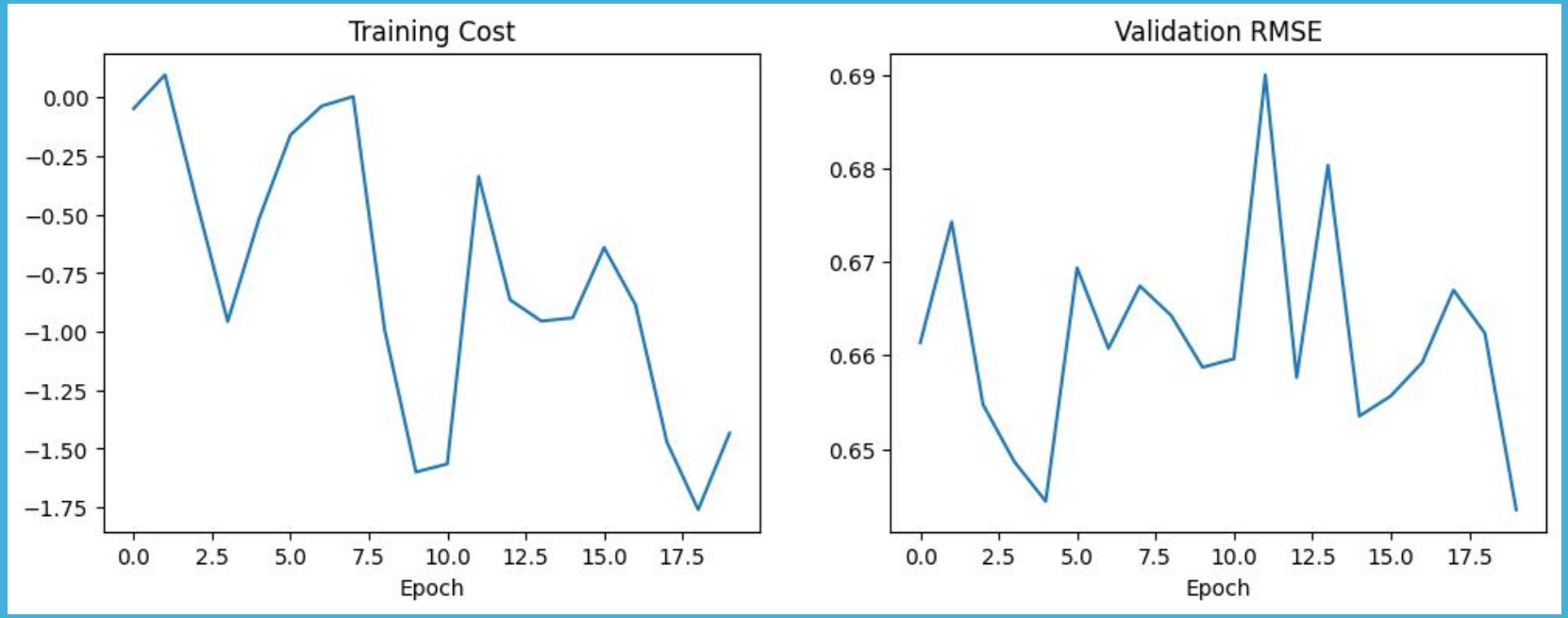


Complete Characterization

- Very few variables in the Great Lakes can be constrained this well

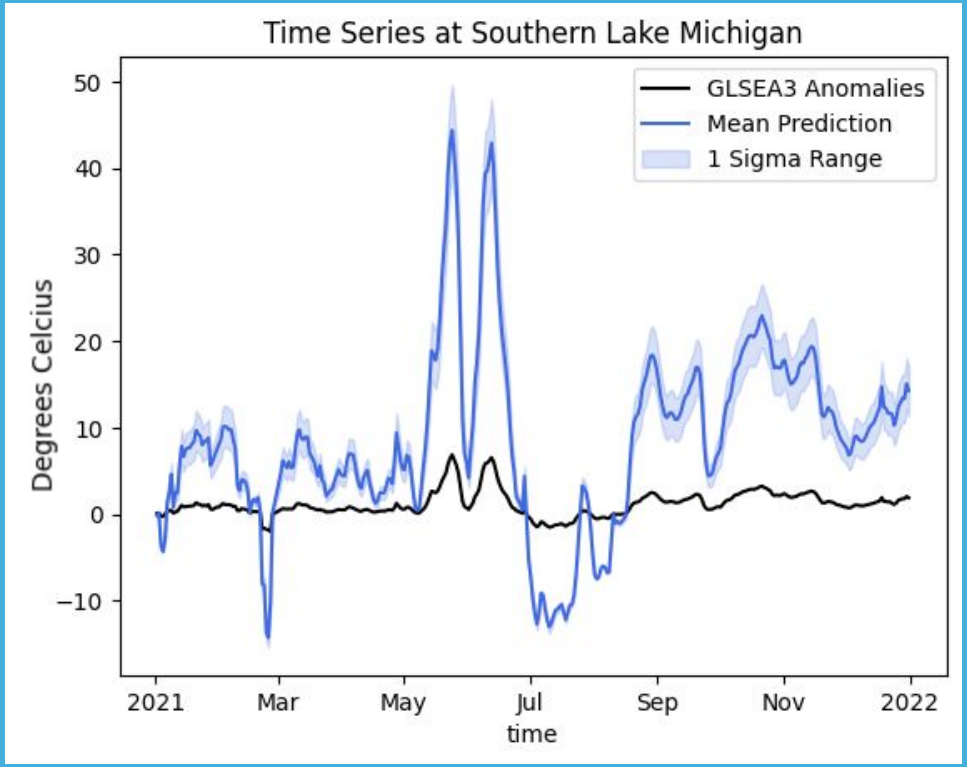


2.) Using Random Sampling, Many Days



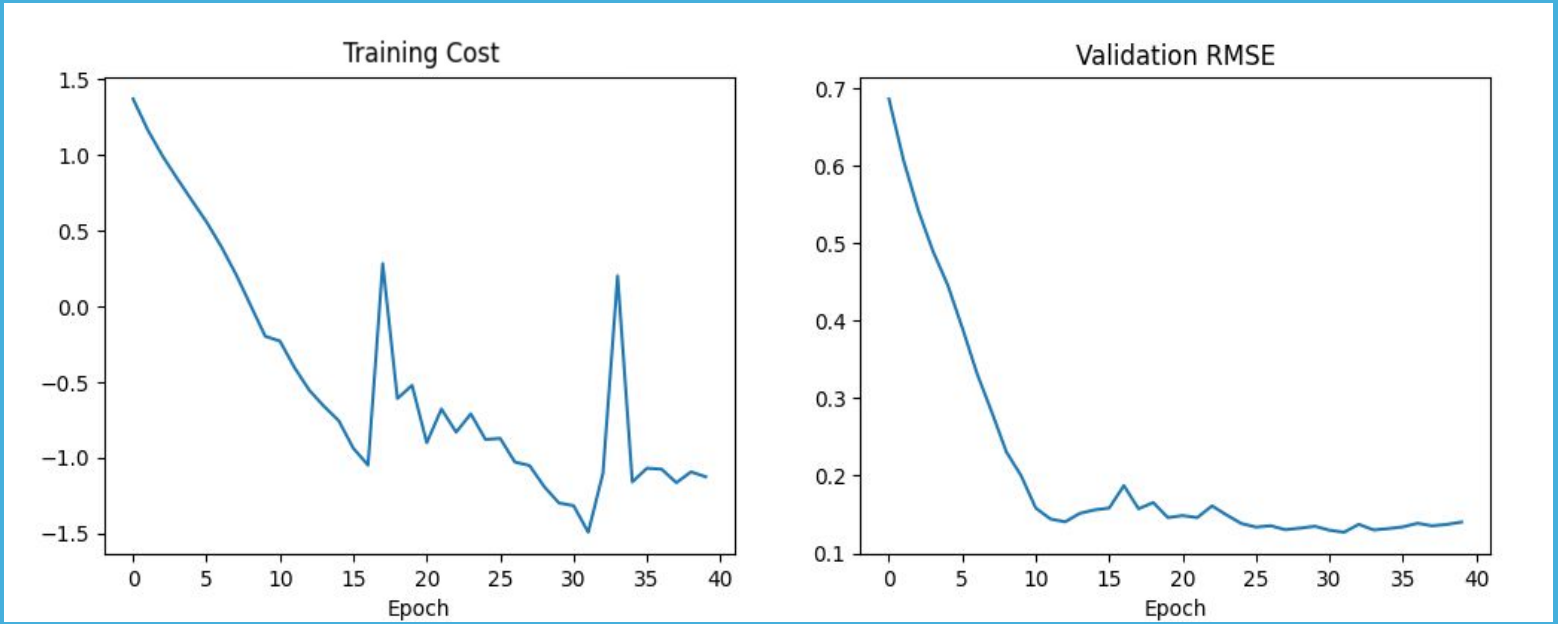
Not Enough Information

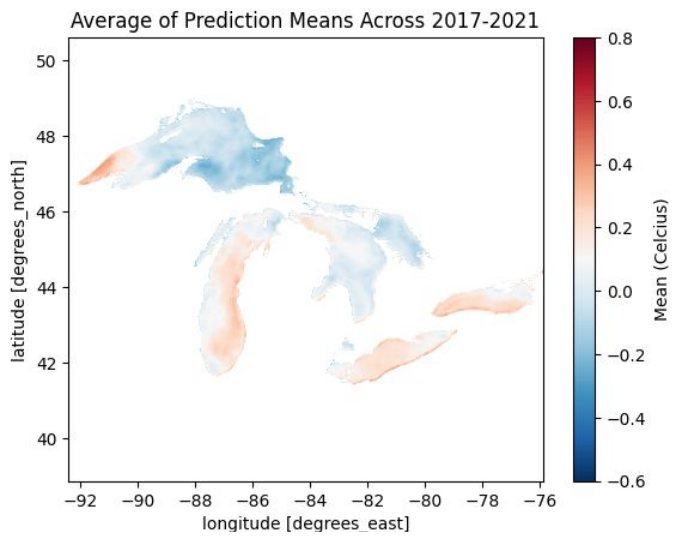
- With insufficient data, the model does not converge

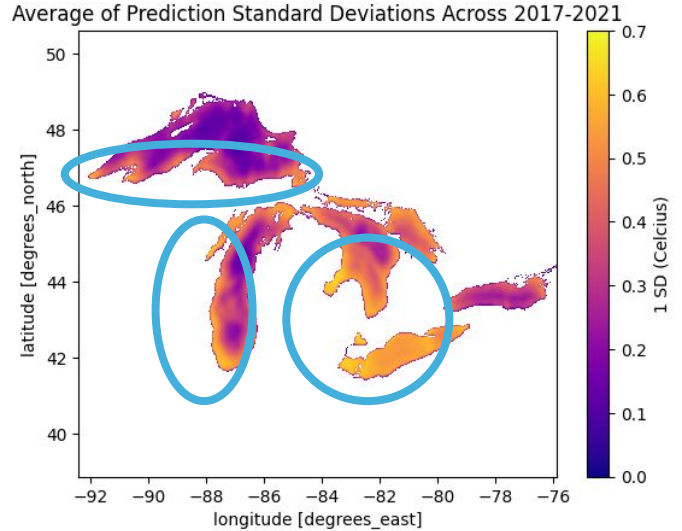


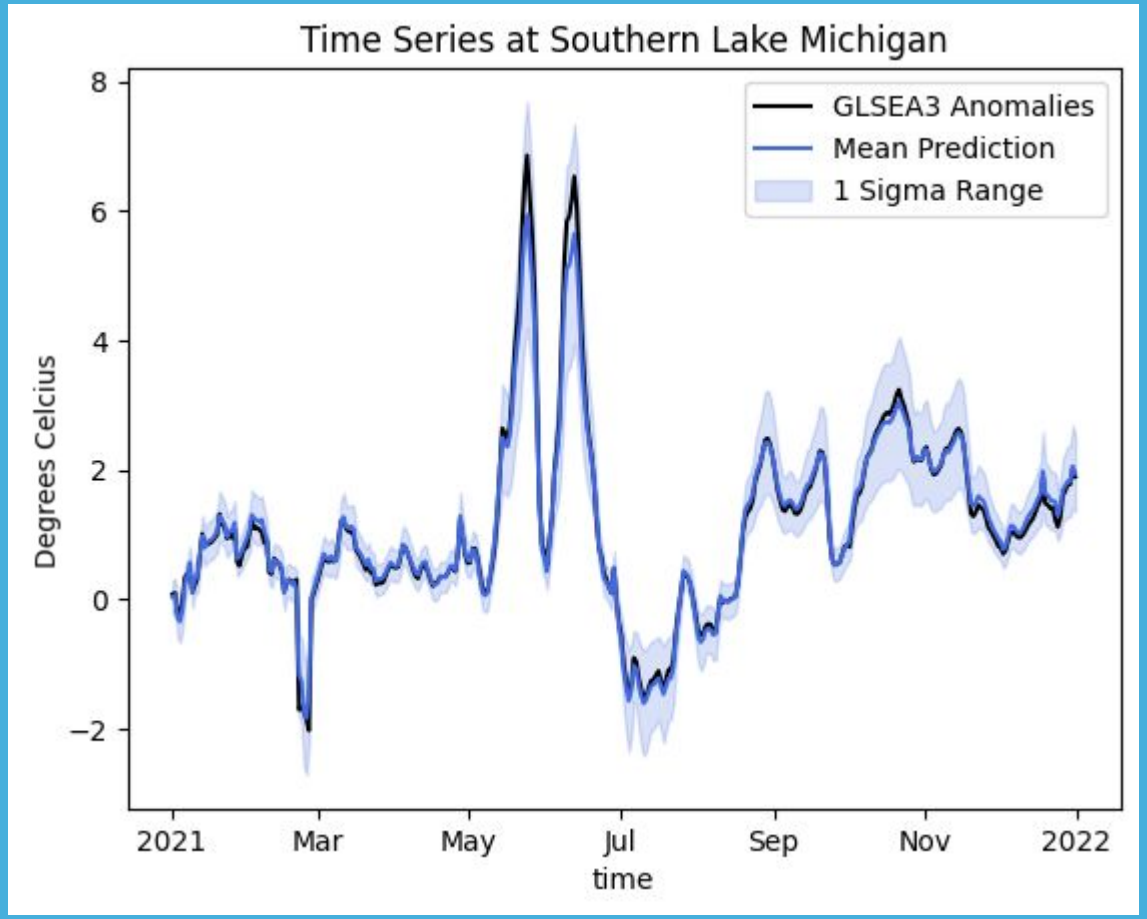
3.) Using All Points, One Single Day

Training Across all the Great Lakes Converges

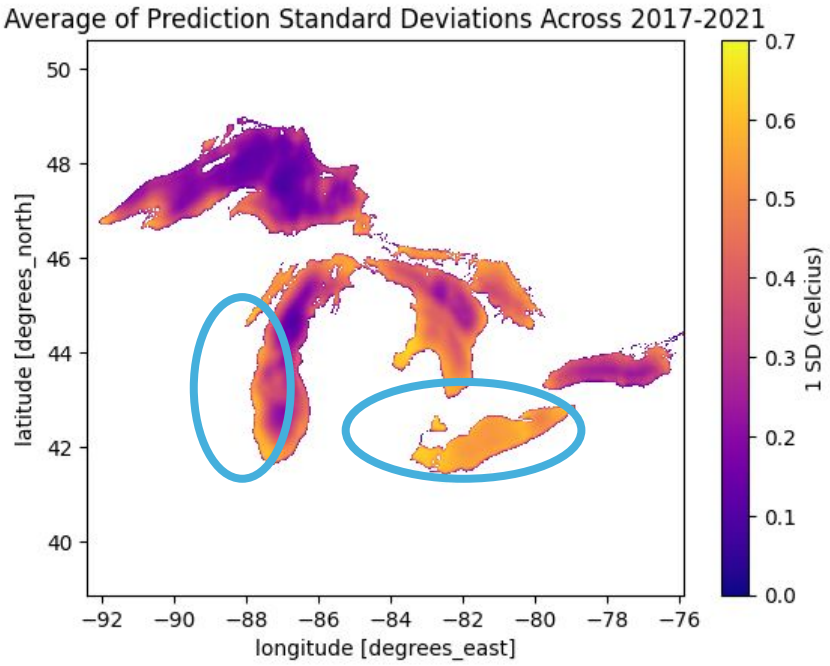


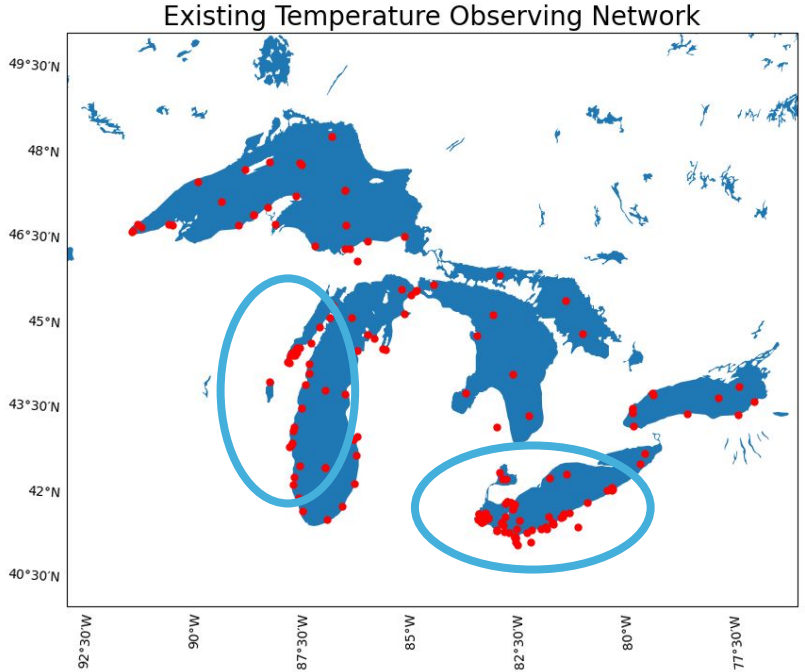

Average of Prediction Means Across 2017-2021
latitude [degrees_north]
longitude [degrees_east]
Mean (Celcius)


Average of Prediction Standard Deviations Across 2017-2021
latitude [degrees_north]
longitude [degrees_east]
1 SD (Celcius)

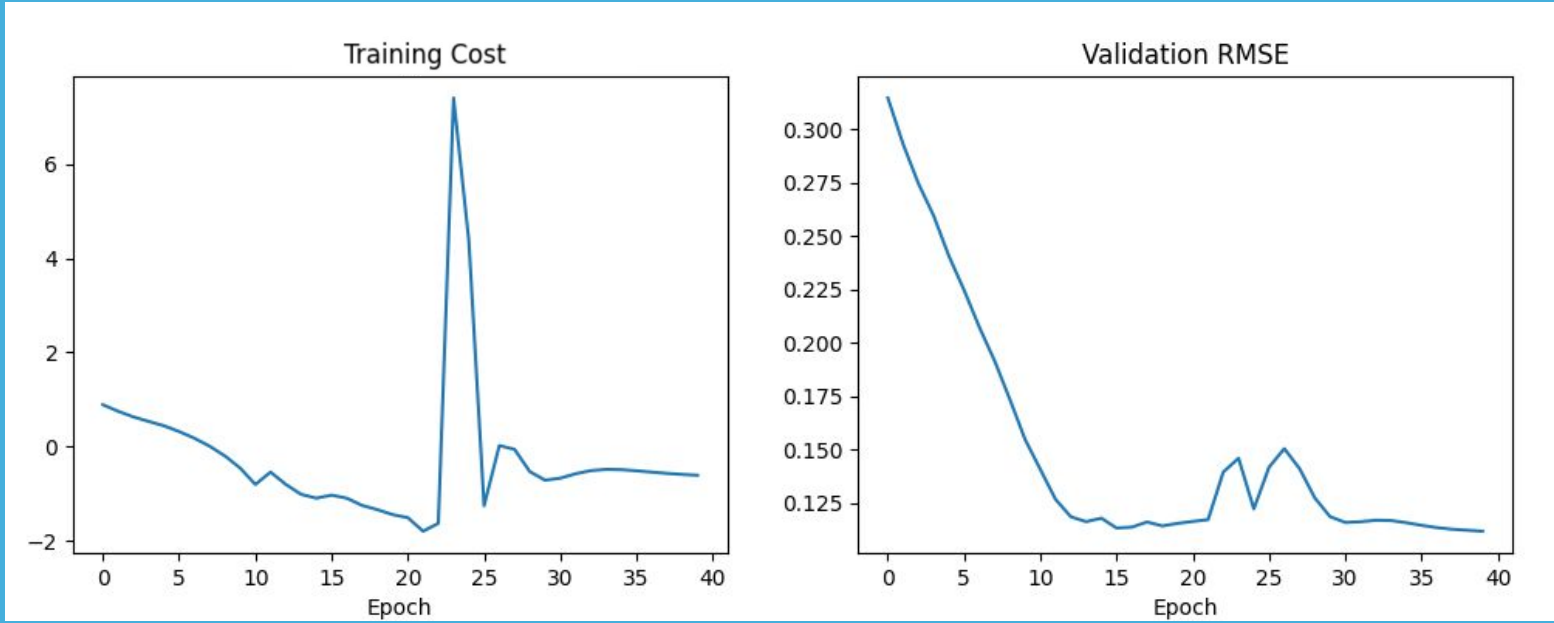

Time Series at Southern Lake Michigan
GLSEA3 Anomalies
Mean Prediction
1 Sigma Range
Degrees Celcius
time

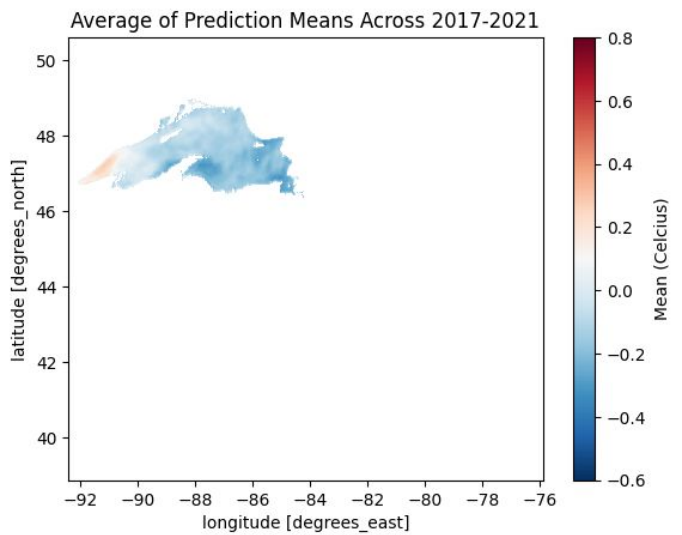
Areas of Uncertainty vs. Deployed Sensors

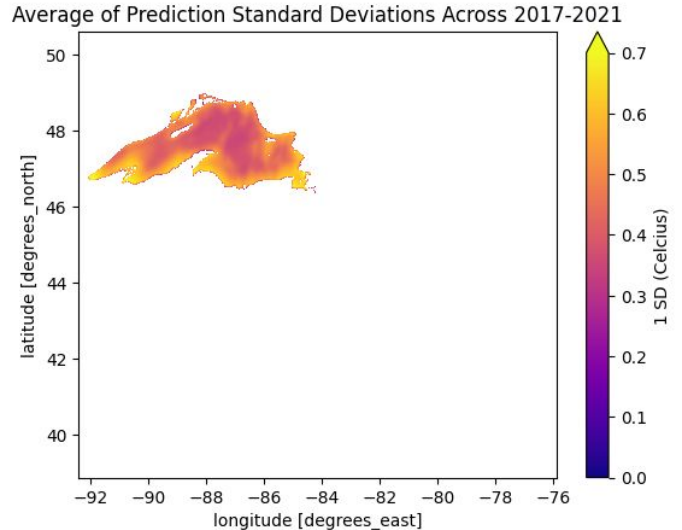


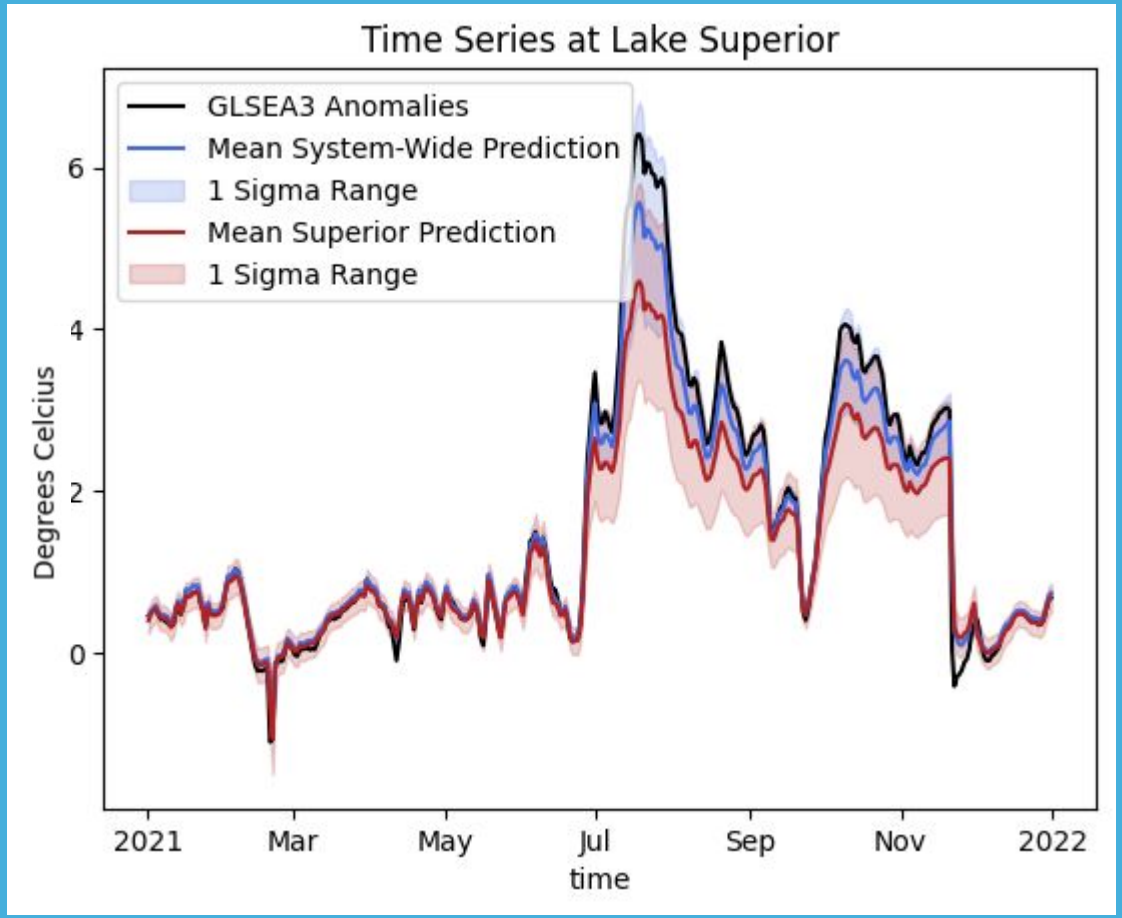


Training on Lake Superior Alone Converges

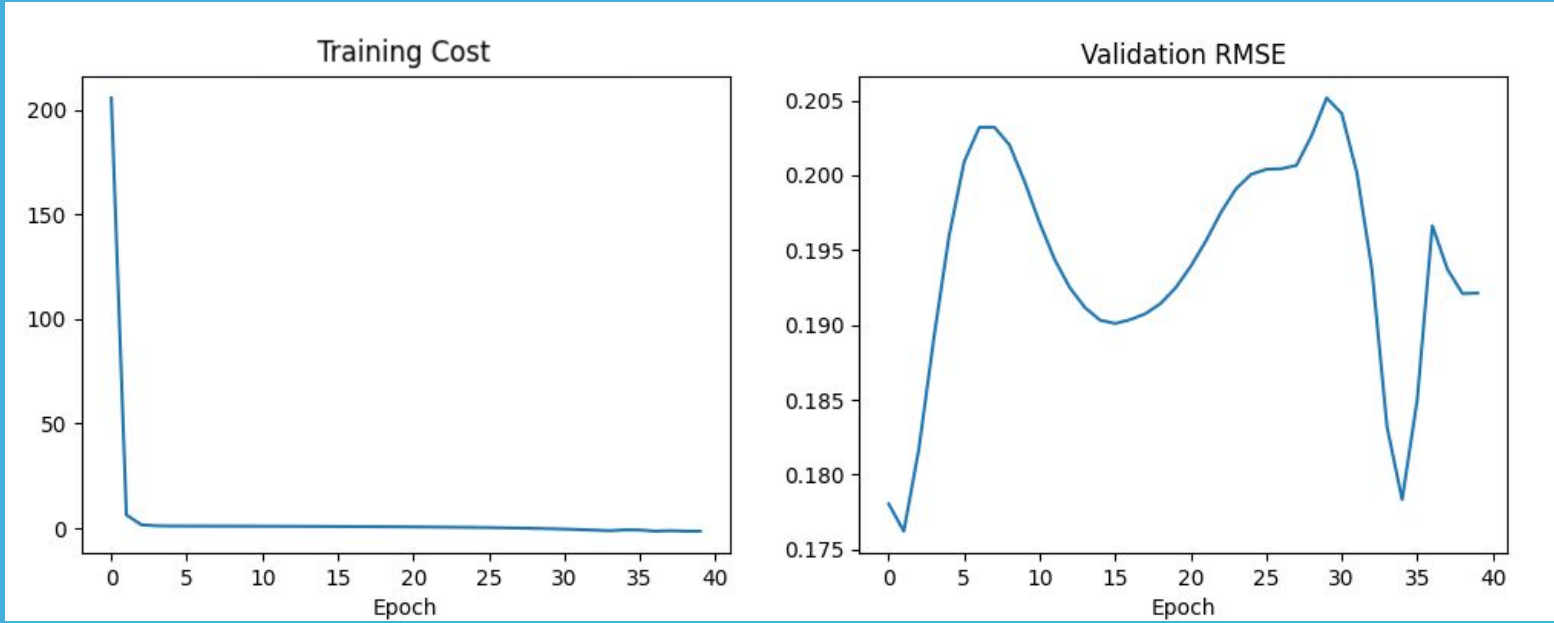


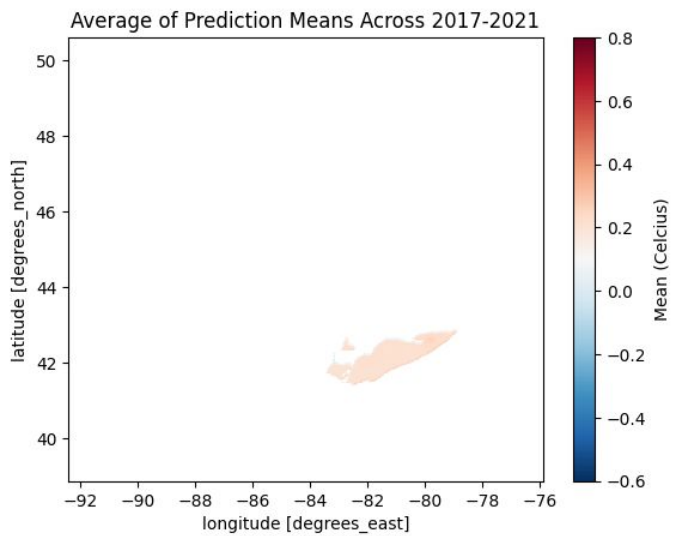

Average of Prediction Means Across 2017-2021
latitude [degrees_north]
longitude [degrees_east]
Mean (Celcius)

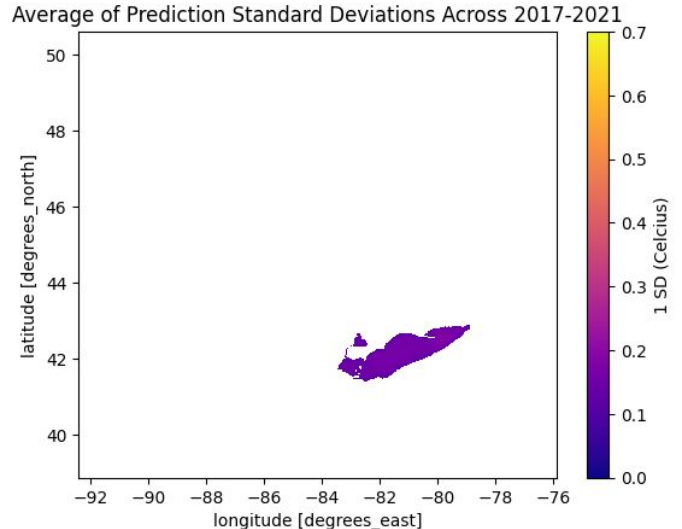

Average of Prediction Standard Deviations Across 2017-2021
latitude [degrees_north]
longitude [degrees_east]
1 SD (Celcius)

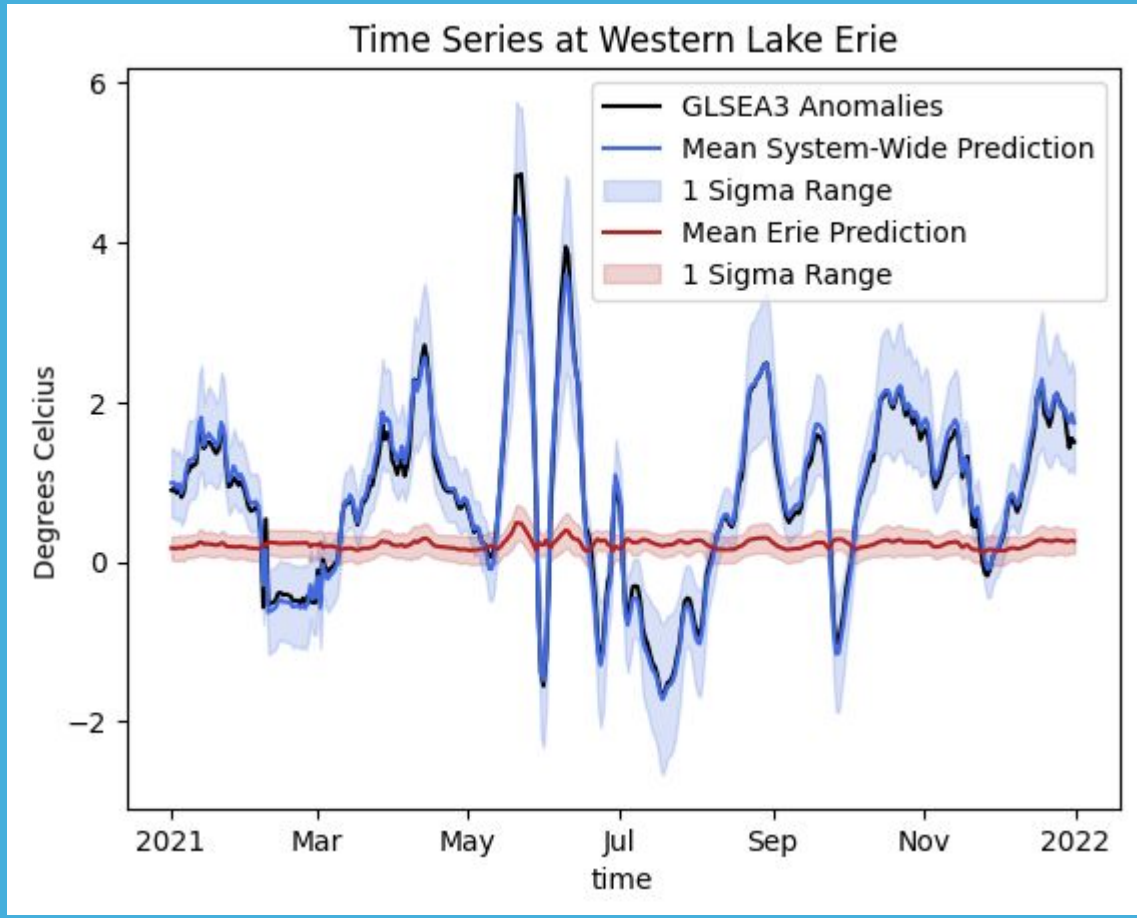

Time Series at Lake Superior
GLSEA3 Anomalies
Mean System-Wide Prediction
1 Sigma Range
Mean Superior Prediction
1 Sigma Range
Degrees Celcius
time

Training on Lake Erie Alone Overfits




Average of Prediction Means Across 2017-2021
latitude [degrees_north]
longitude [degrees_east]
Mean (Celcius)


Average of Prediction Standard Deviations Across 2017-2021
latitude [degrees_north]
longitude [degrees_east]
1 SD (Celcius)


Time Series at Western Lake Erie
GLSEA3 Anomalies
Mean System-Wide Prediction
1 Sigma Range
Mean Erie Prediction
1 Sigma Range
Degrees Celcius
time

Conclusions

- Great Lakes application of DeepSensor
- **Prediction uncertainty highlights shallower, rapidly-changing areas where measurements are needed**
- Explored 3 training regimes
  - 1.) Complete characterisation
  - 2.) More information needed
  - 3.) **Enough** information to constrain the predictions

Further Research

- Refine training
  - Explore different data densities
- More data
  - Add extra information: atmospheric and water quality, GLOS platforms
- Consider other, less well-sampled variables
- Explore other DeepSensor features

Application

- Inflation Reduction Act
- Bipartisan Infrastructure Law

Questions?

Thank you all for listening