Exploring Sea Surface Temperature Data Using Machine Learning

23rd GHRSST International Science Meeting

Madolyn Kelm

TABLE OF CONTENTS

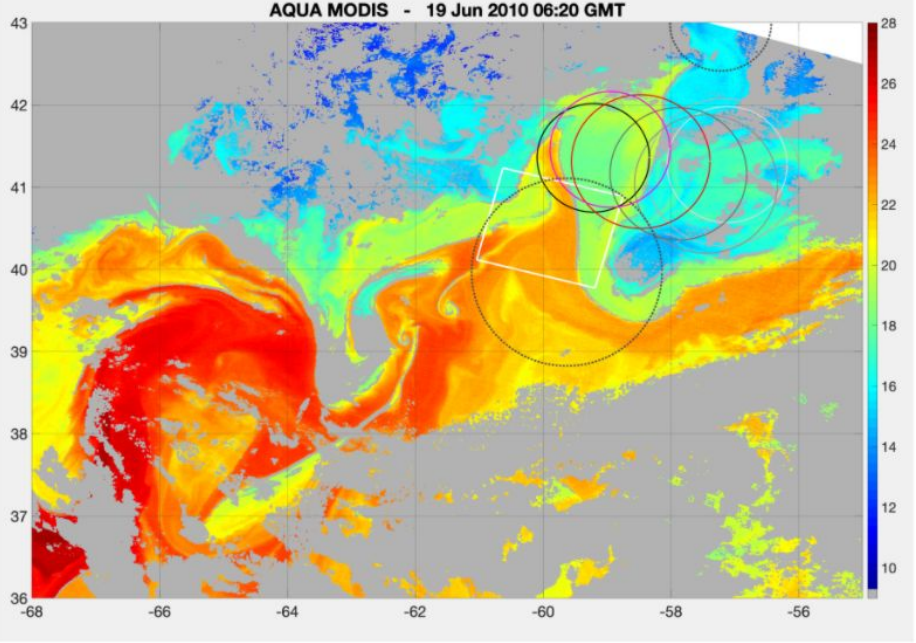
01

BACKGROUND

Problem

- Sea surface temperature (SST) datasets are massive and show complex patterns
- Current standard oceanographic measures (e.g. power spectrum) fail to describe the full complexity

Possible Solution



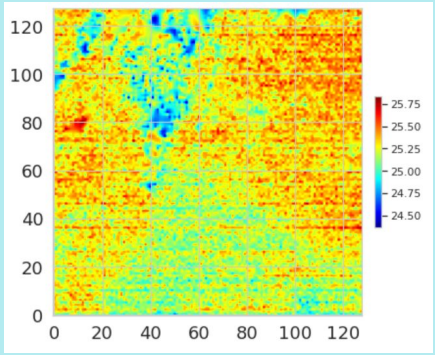
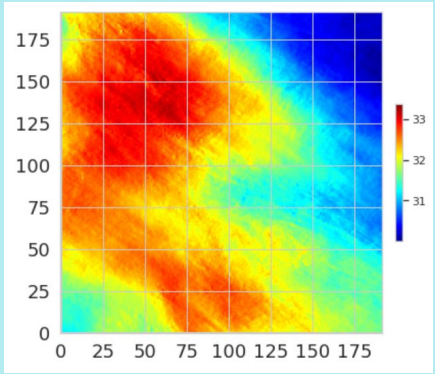
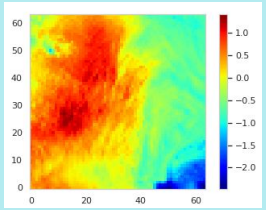
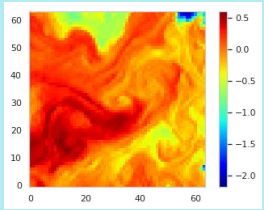
THE DATA COLLECTION

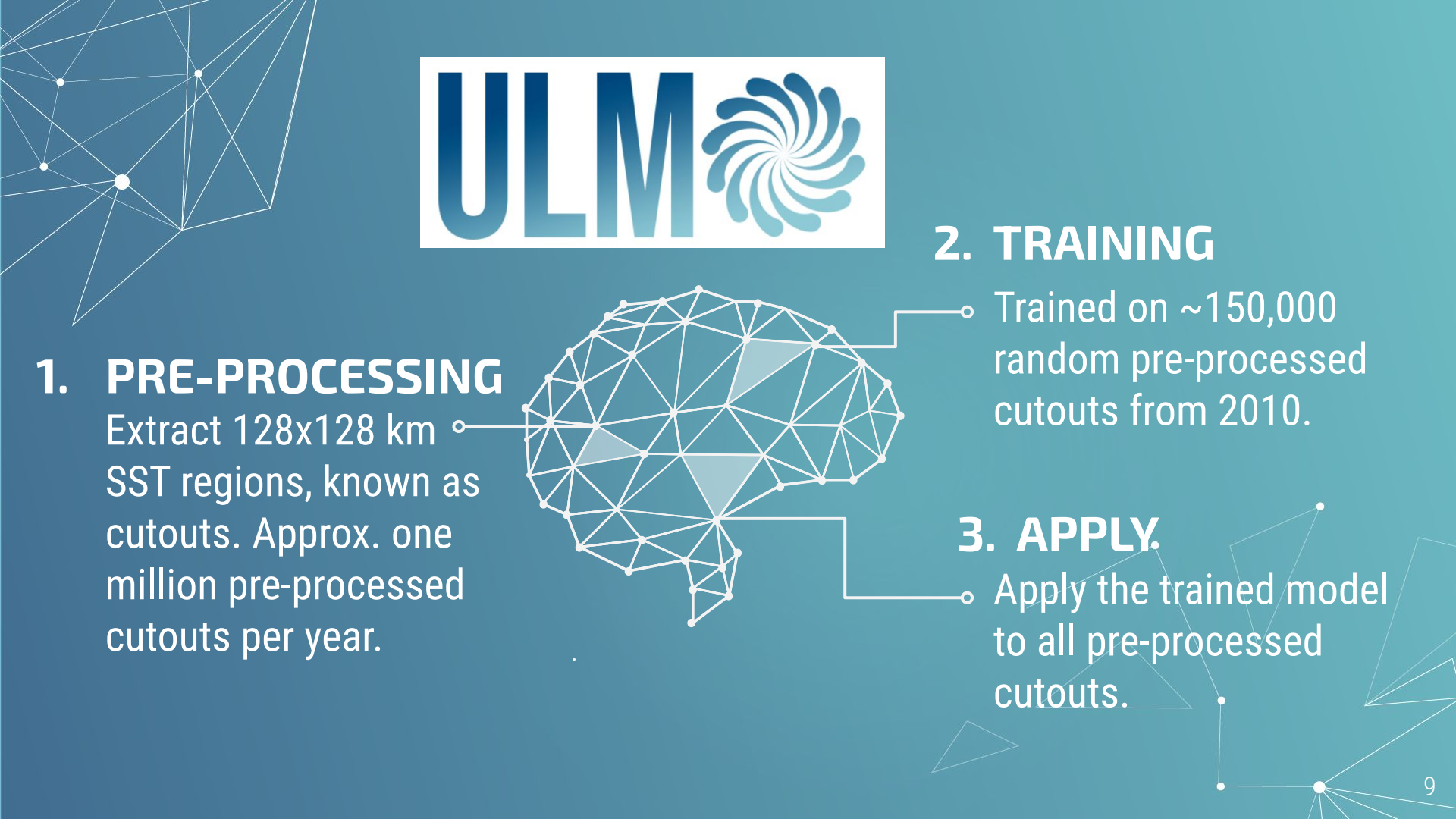
	MODIS	VIIRS
Sampling Period		
Spatial Resolution		

1. PRE-PROCESSING

Extract 128x128 km SST regions, known as cutouts. Approx. one million pre-processed cutouts per year.

	MODIS	VIIRS
Sample Size		
Cloud Free Percentage		
Final Avg. Region		

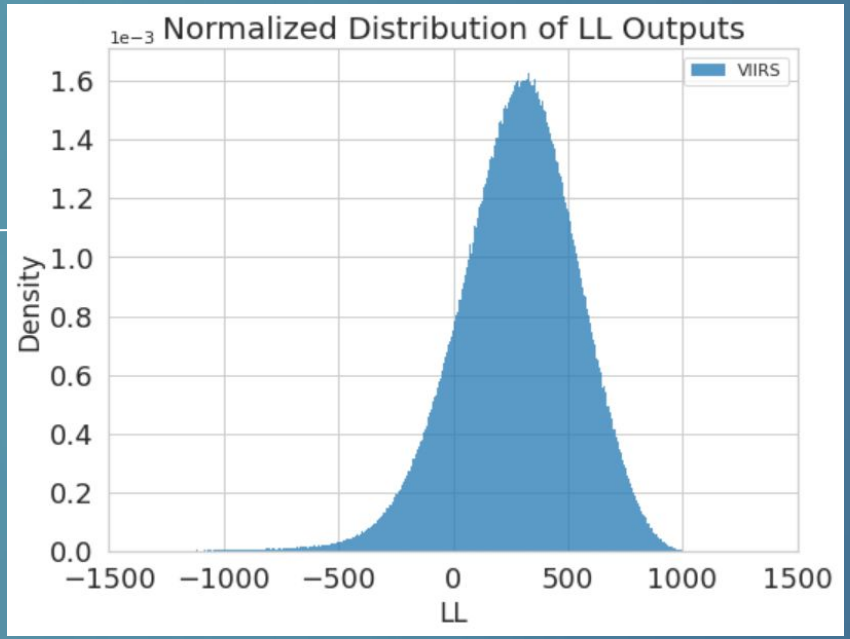
	MODIS	VIIRS
Sample Size	128 x 128 km 128 x 128 pixel 	192 x 192 km 128 x 128 pixel 
Cloud Free Percentage		
Final Avg. Region	64 x 64 pixel 	64 x 64 pixel 


ULM
1. PRE-PROCESSING
Extract 128x128 km SST regions, known as cutouts. Approx. one million pre-processed cutouts per year.
2. TRAINING
Trained on ~150,000 random pre-processed cutouts from 2010.
3. APPLY.
Apply the trained model to all pre-processed cutouts.

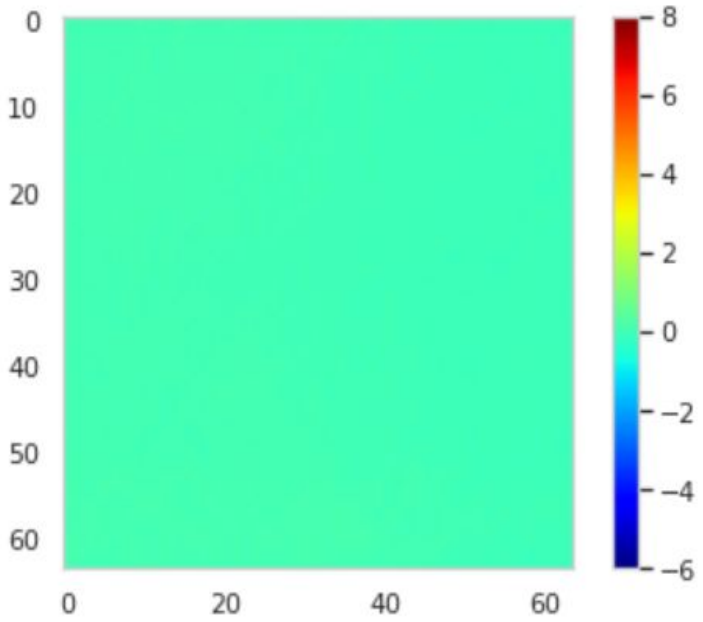
Machine Learning Metric: LL

THE OUTPUT

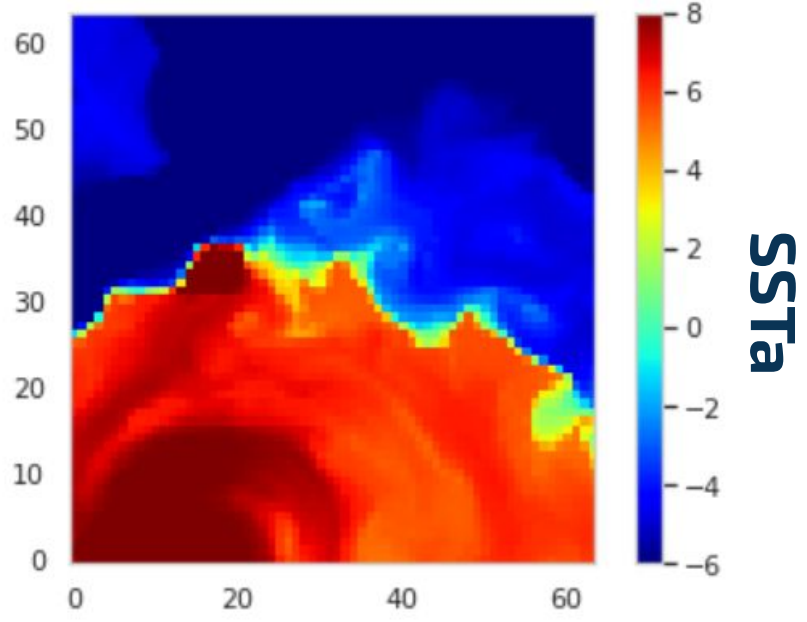
Each cutout is assigned a single number that represents the frequency of images similar to it within the full distribution.



EXAMPLES OF CUTOUTS WITH HIGH and LOW LL VALUES



HIGH
LL Value: 1060.52



LOW
LL Value: -21033.65

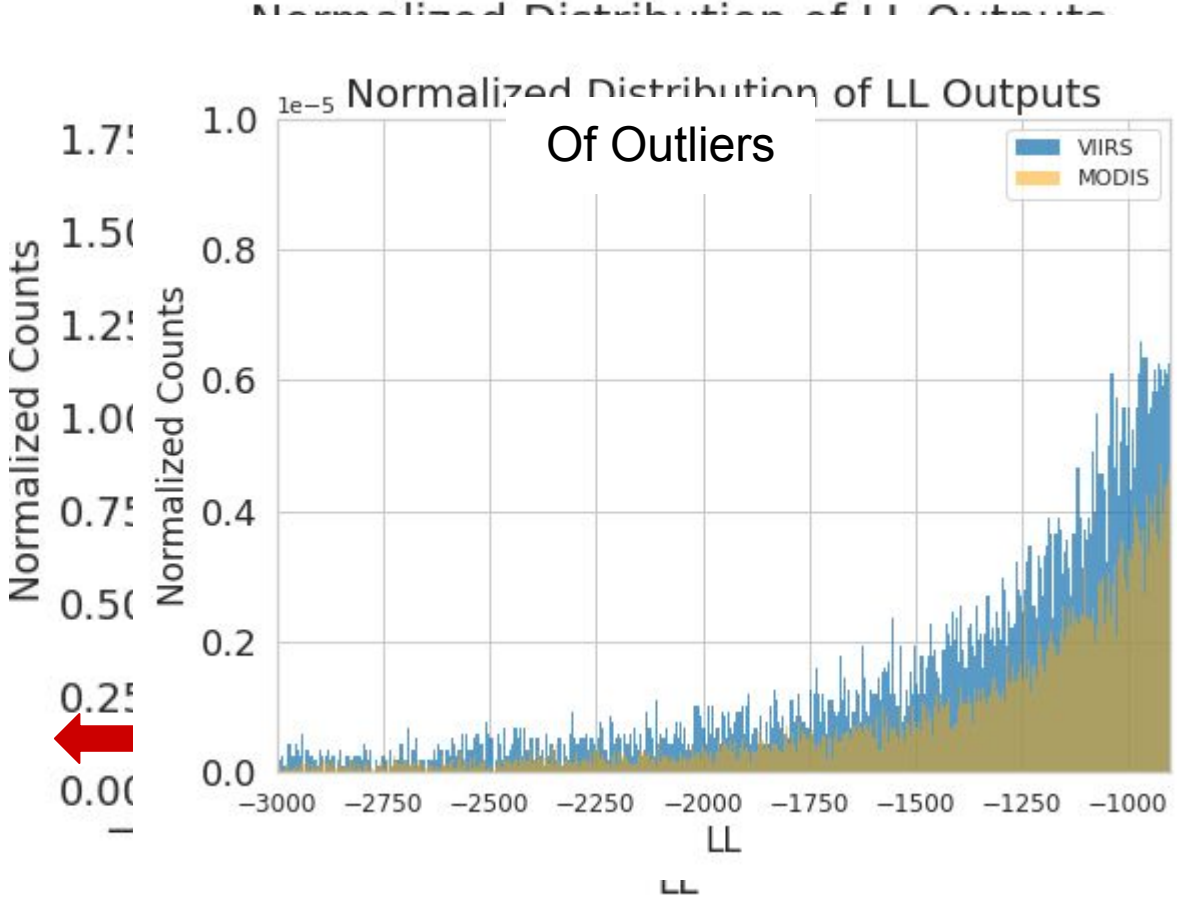
02

GOALS AND RESULTS

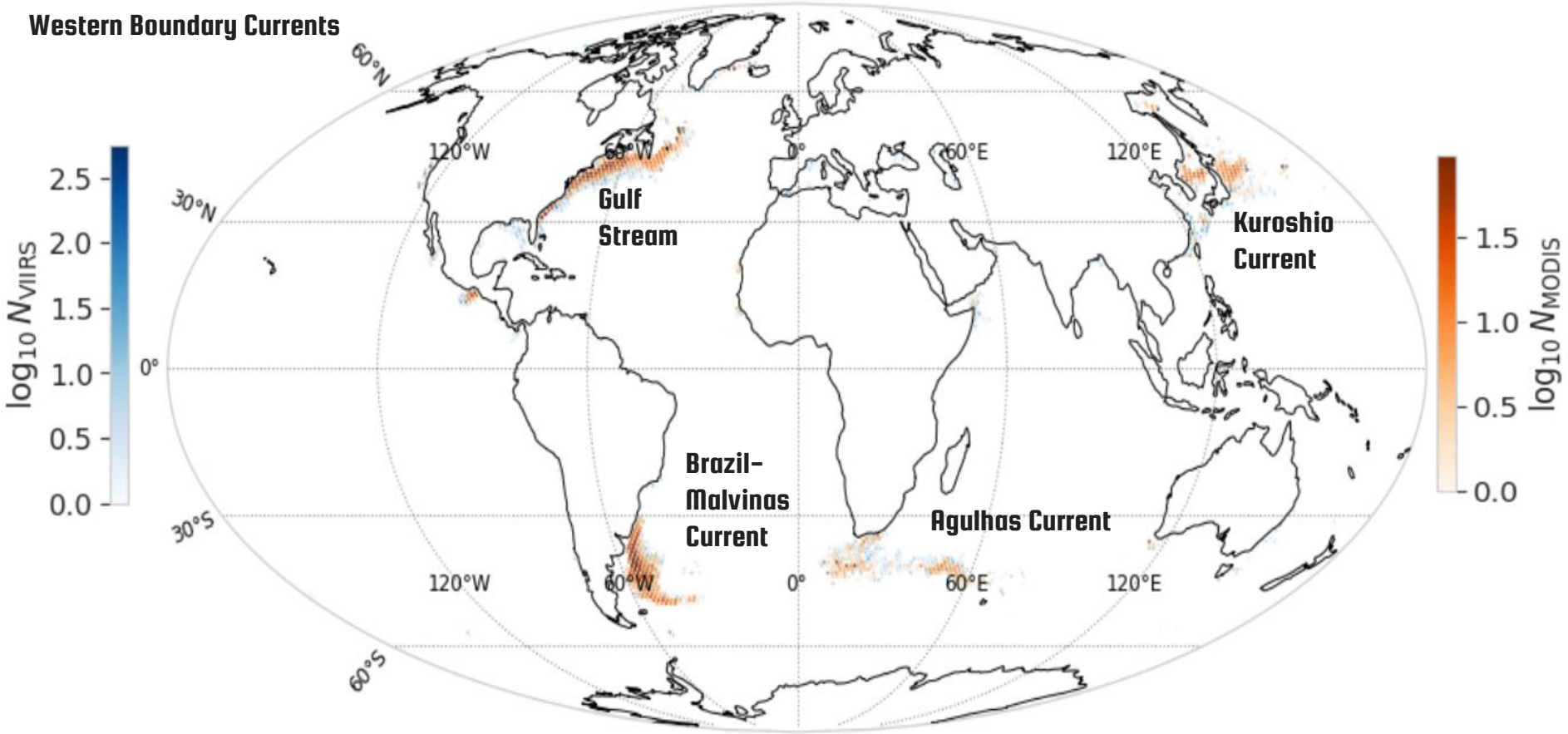
INITIAL GOALS

VERIFY ULMO

Use the MODIS trained ULMO on VIIRS data


Normalized Distribution of LL Outputs
Of Outliers
1e−5
1.0
0.8
0.6
0.4
0.2
0.0
Normalized Counts
VIIRS
MODIS
−3000
−2750
−2500
−2250
−2000
−1750
−1500
−1250
−1000
LL

Cutouts with Log Likelihood Values <-1000 for MODIS and VIIRS



VERIFY ULMO

HIGHLIGHTS

- The MODIS trained ULMO was able to recognize similar, but not identical, patterns in the VIIRS dataset
- Moving forward with understanding the pattern recognition within VIIRS

INITIAL GOALS

VERIFY ULMO

Use the MODIS trained ULMO on VIIRS data

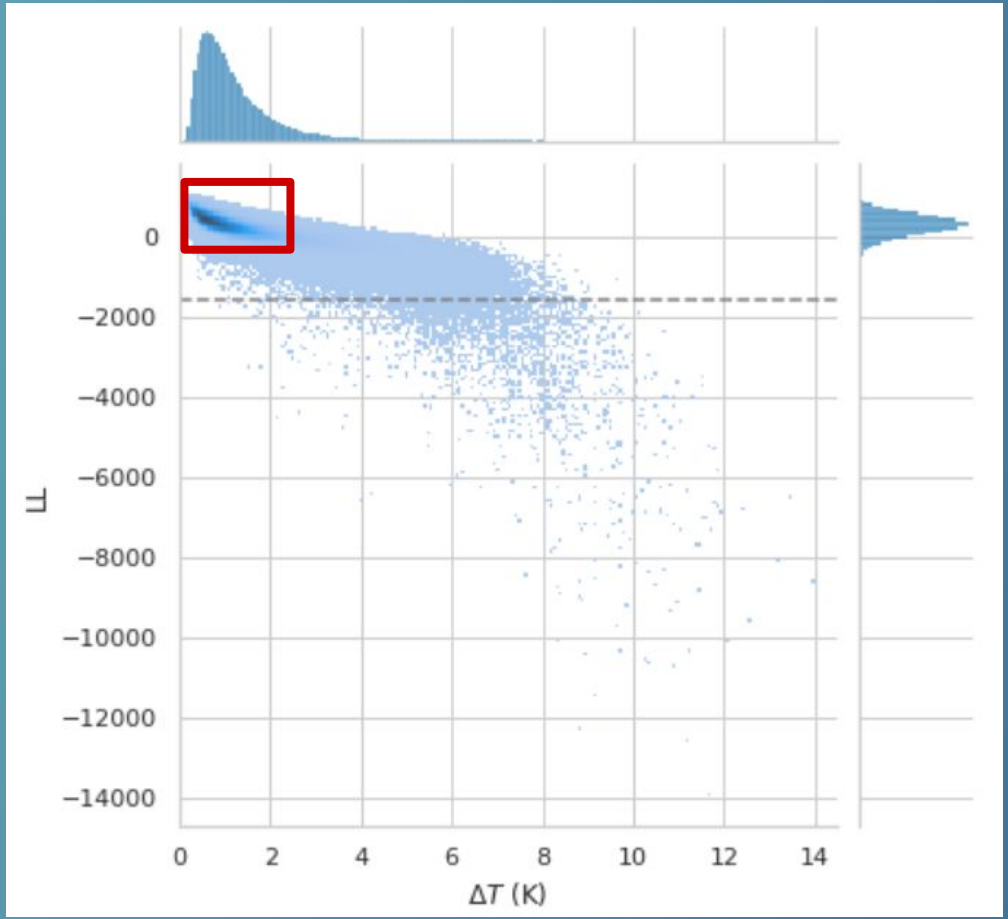
EXPLORE

Explore the patterns ULMO has recognized

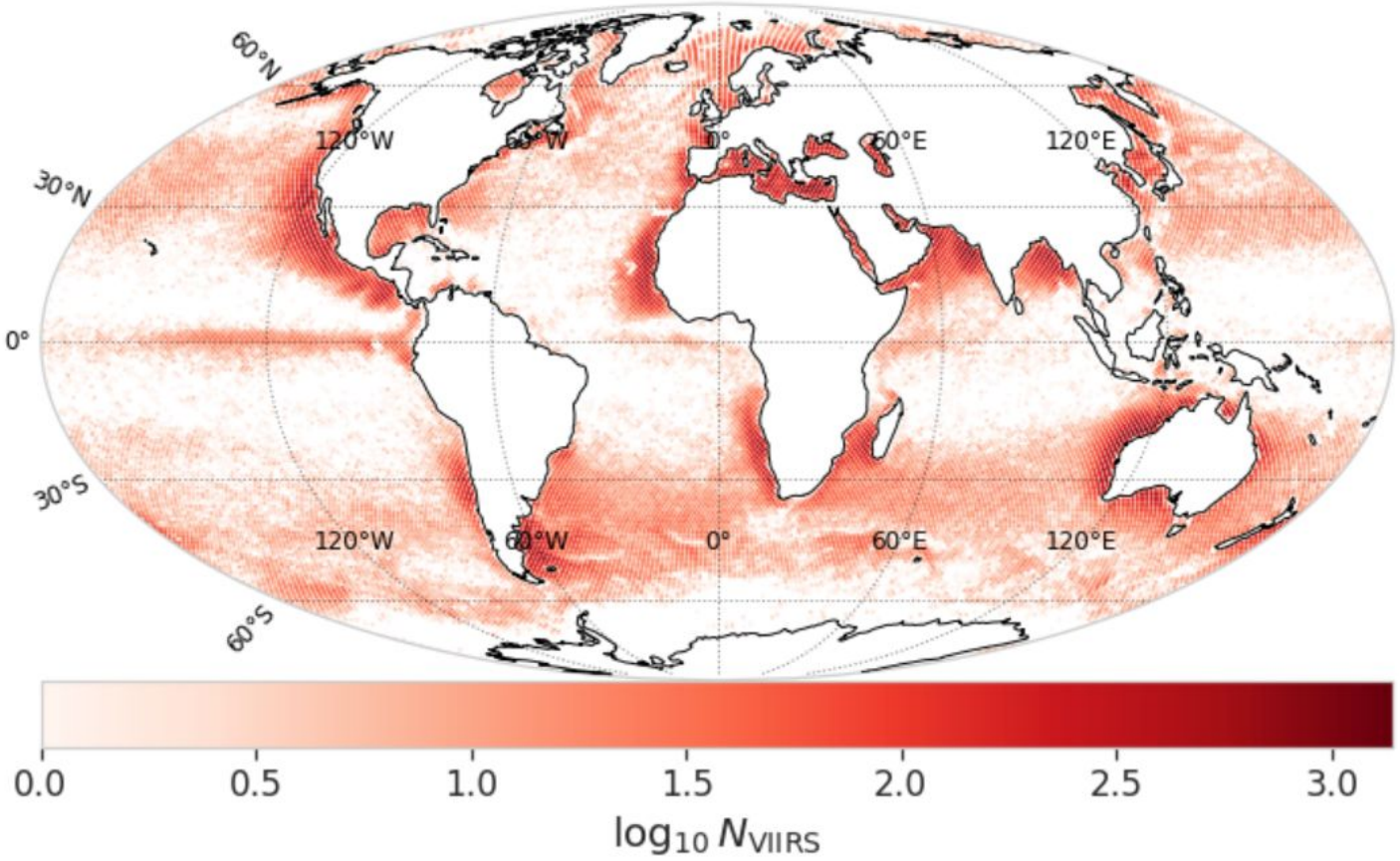
INVESTIGATION OF THE LL VALUES

LL Values

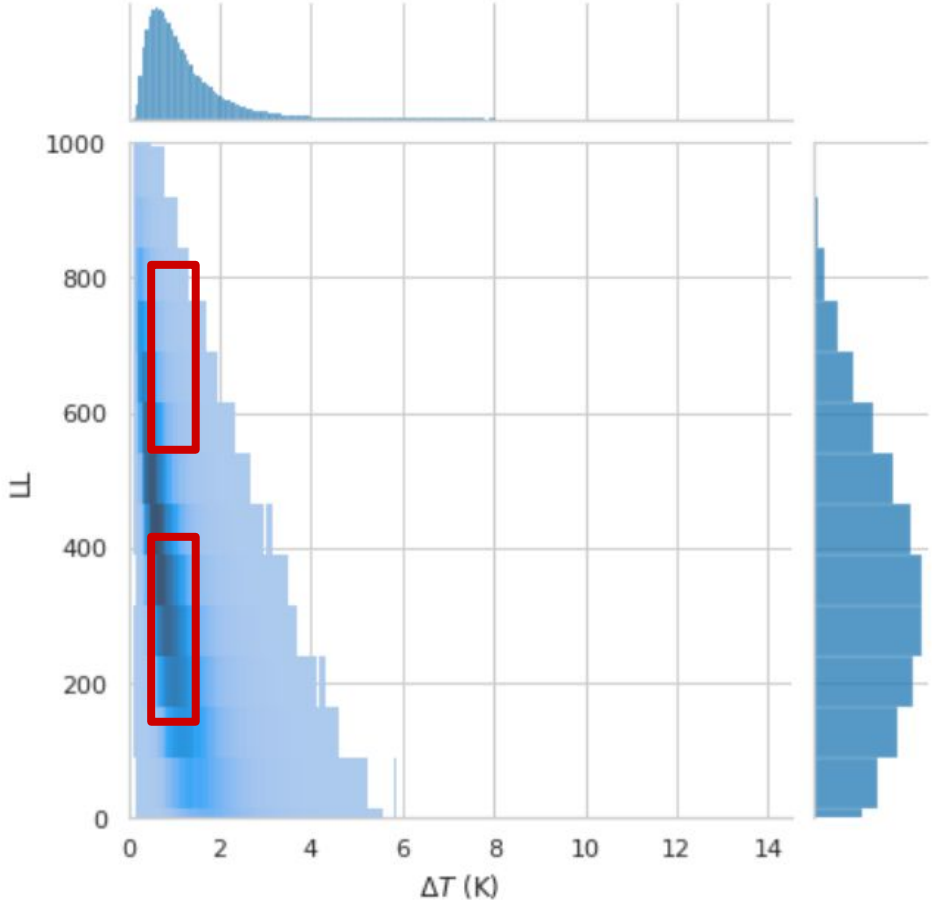
- We have started to see that the outliers (LL< -1000) show more structure with some dependence on temperature range.
- Now I will explore the larger portion of the distribution



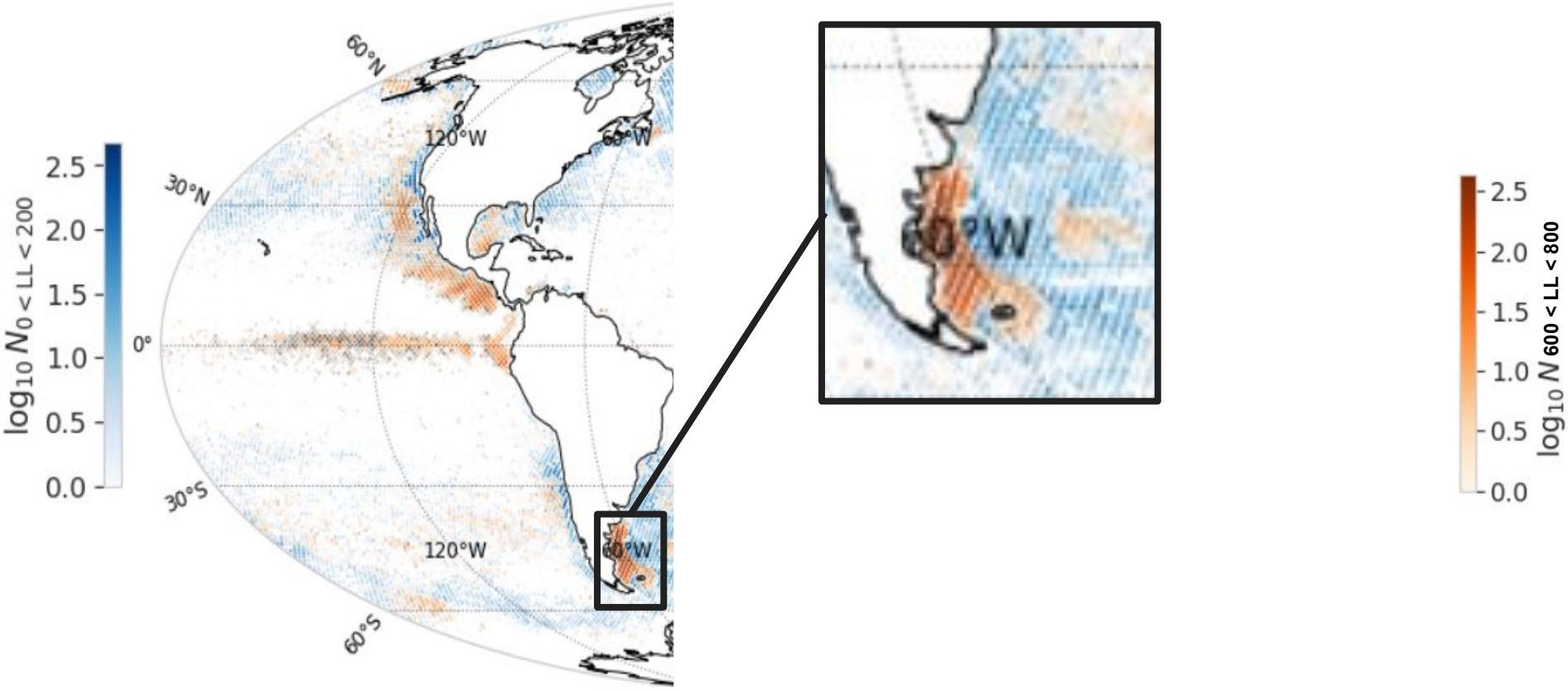
Cutouts with Temperature Range 1<ΔT<1.5 VIIRS



Cutouts with Temperature Range 1<ΔT<1.5 VIIRS



Cutouts with 1<ΔT<1.5 and Specified LL Value Range



03

CONCLUSION NEXT STEPS

CONCLUSIONS

HIGHLIGHTS

- ULMO highlighted differences in SST datasets
- LL is a good discriminator over the entire range of the LL rather than just the tail

FUTURE

- Use this to now explore the LL space
  - What patterns exist in the LL distribution
  - How LL values relate to the basic geophysical properties

TAKE HOME MESSAGE

The ULMO algorithm is successful at identifying complex patterns within extremely large sea surface temperature datasets, even upwards of 100 TB.

Its machine learning metric (LL) highlights geophysical processes in distinct locations.

ACKNOWLEDGEMENTS

National Science Foundation

Mentors

Peter Cornillon and
J. Xavier Prochaska:
5257-LEARNING THE FUNDAMENTAL PATTERNS OF SST IMAGERY ON SUB-MESOSCALES Slides at:
https://tinyurl.com/x-at-osm2022

University of Rhode island

AND THANKS TO YOU!
Any Questions?