# Enabling construction of Multimission water quality products through mixture density networks

Nima Pahlevan, Ryan O'Shea, Brandon Smith, Arun Saranathan,
AND 30+ Data Contributors

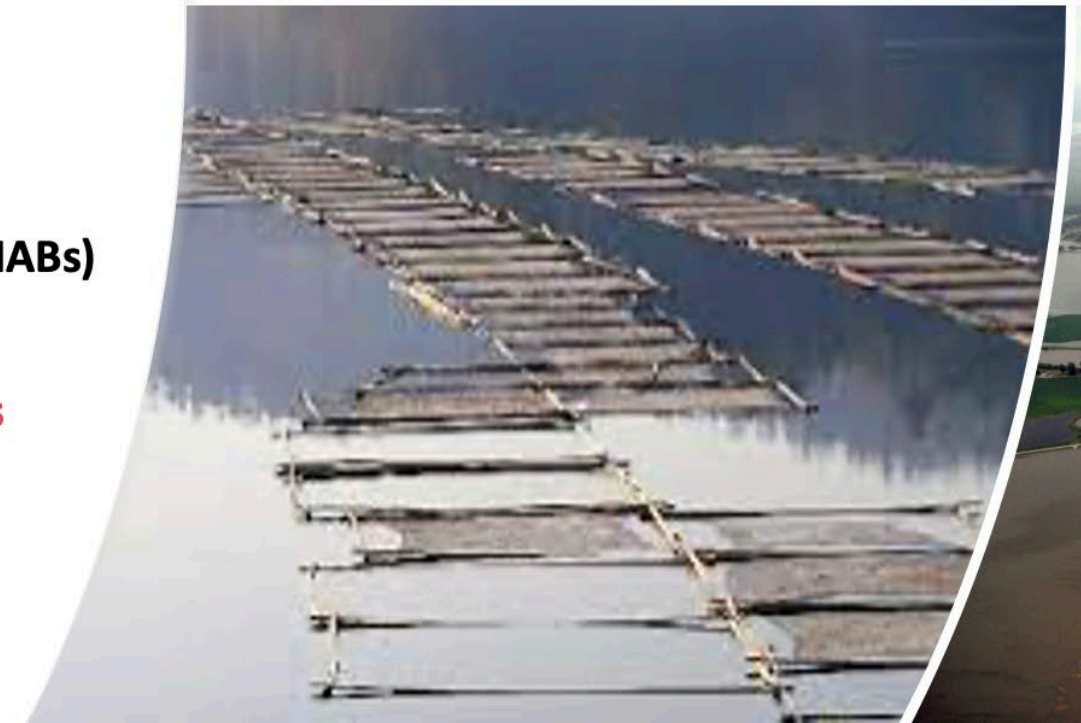
SSAI, Freshwater Sensing Group
NASA Goddard Space Flight Center

Living Planet Symposium
Bonn, May 2022

science for a changing world

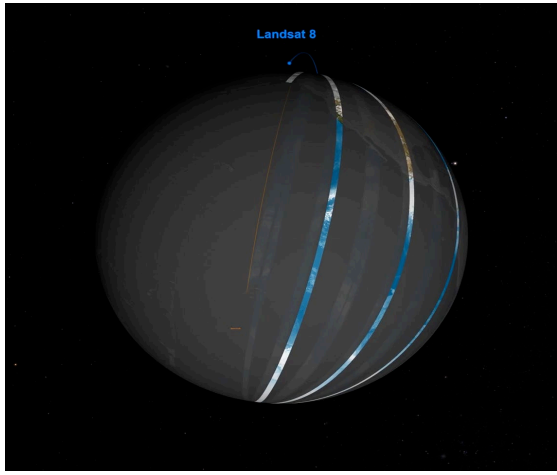
# Aquatic Ecosystems Are Under Threat

- **Stressors**
  - Wildfires, Flooding,…
  - Food Production, Mining, Recreation,…
- **Impacts**
  - Water Pollution
  - Harmful Algal Blooms (HABs)
- **Consequences & Implications**
  - Public Health
  - Economy

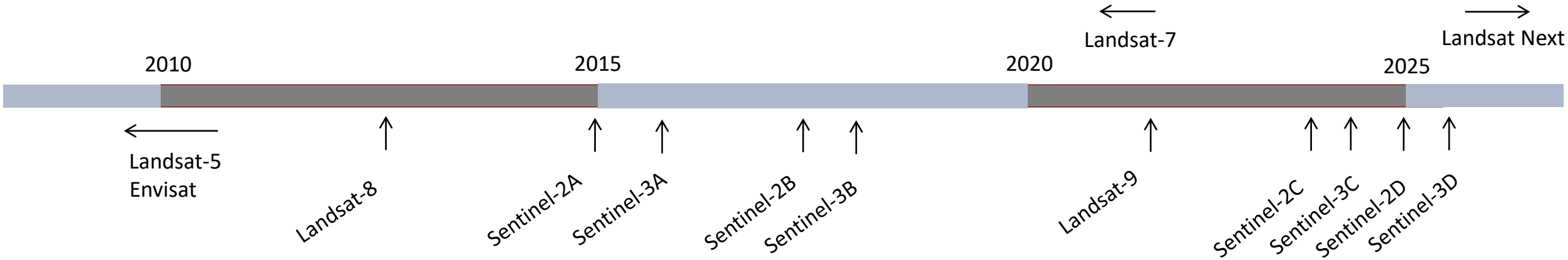


NASA Scientific Visualization Studio

# Global Multimission Data Processing/Production




- Open-ocean products have been produced
- Lake products have been lately released (demonstration-level)

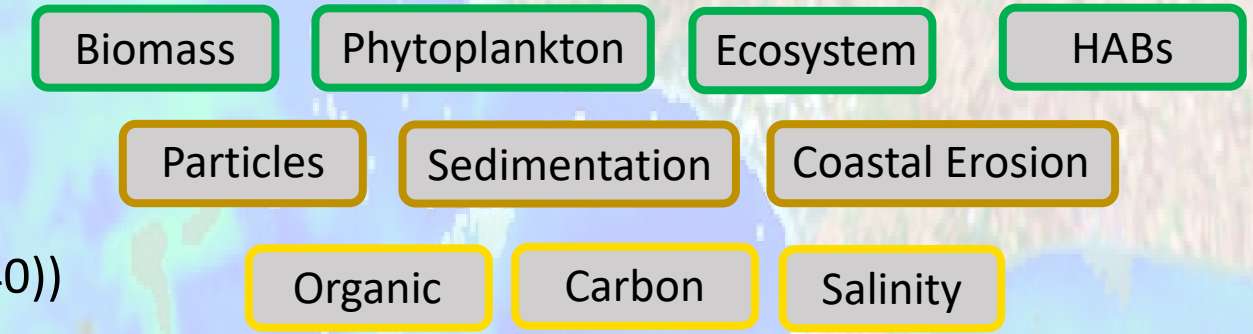



**Algorithmic challenges still remain over freshwaters and coastal ecosystems**

# Synopsis

- Objective
  - Generate **globally** consistent and advanced products from **a universal algorithm**

- Products
  - Chlorophyll-a (Chl*a*)
  - Total Suspended Solids (TSS or SPM)
  - CDOM absorption @ 440 nm ($a_{cdom}(440)$)



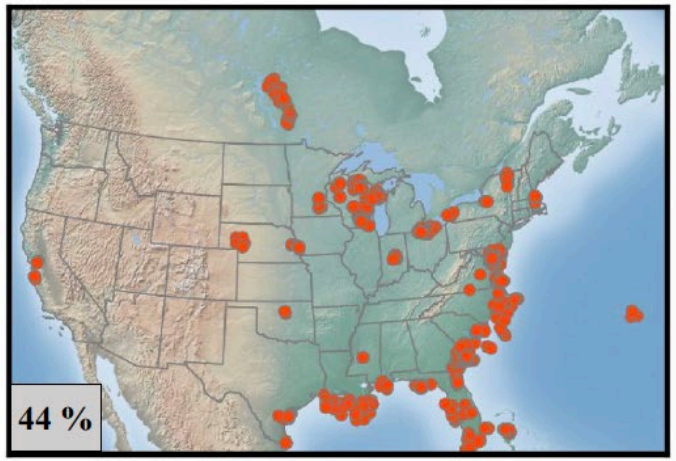

- Satellite sensors
  - Landsat-8/OLI
  - Sentinel-2/MSI
  - Sentinel-3/OLCI

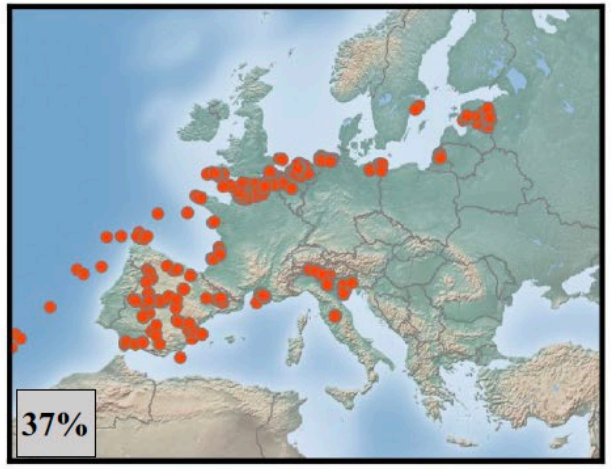
- Algorithm
  - A machine-learning model termed **Mixture Density Networks (MDNs)**

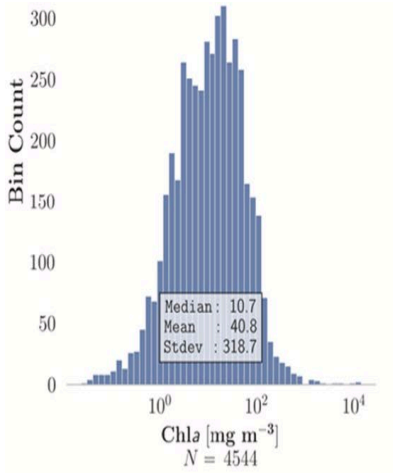
Lake Erie
HICO; Oct 30th 2013

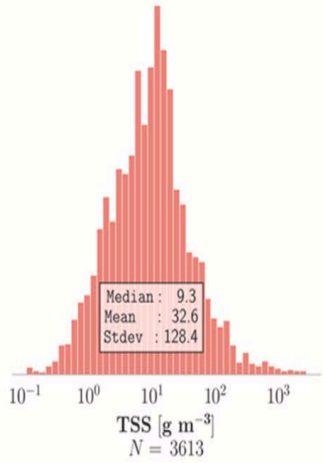
# Development Data

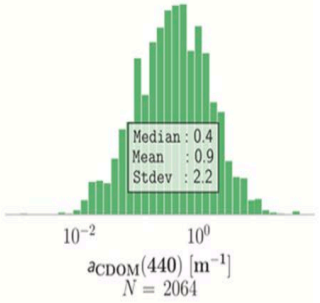
- Source: Field campaign data with co-located
  - $R_{rs}$ (@ 1-3 nm)
  - Chla/TSS/$a_{cdom}$(440)



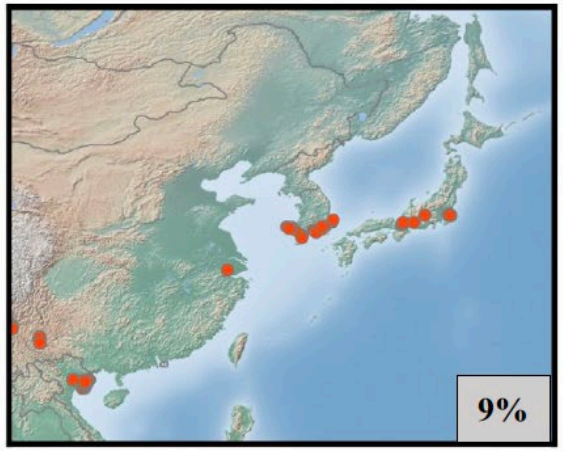



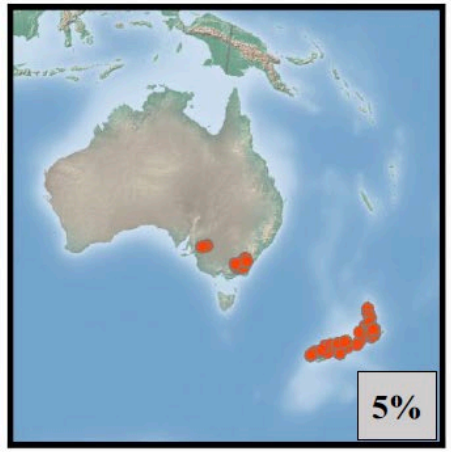



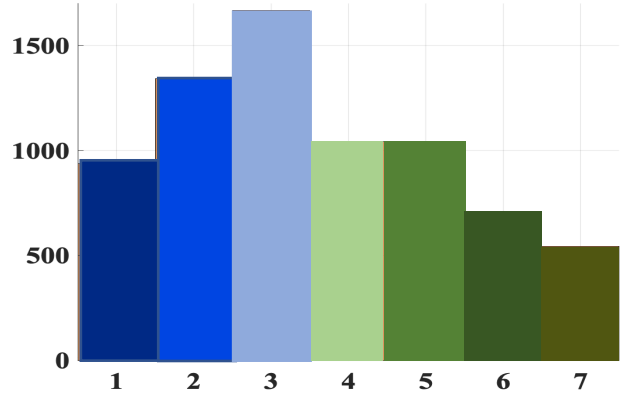








Optical Water Types

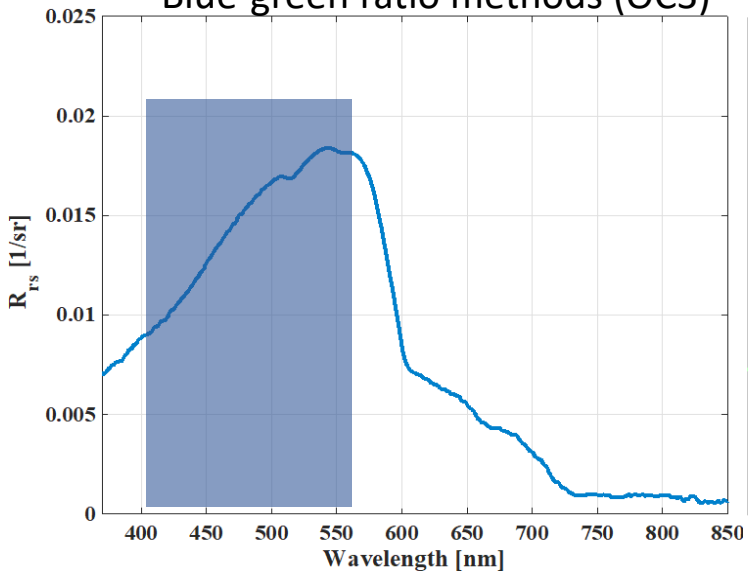


# State-of-the-art Models: Chlorophyll-*a*

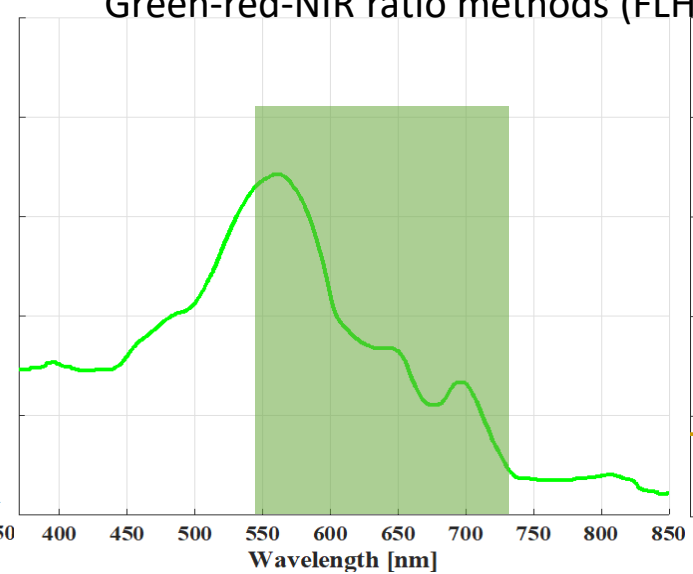
- $R_{rs}$- based approaches
  - Band combinations in different spectral regions
    - O'Reilly et al. 1998, Gilerson et al. 2010, Moses et al. 2012 & …
  - Machine-learning models
    - Doerffer and Schiller 2007 (e.g., NNs trained on simulated data built into C2RCC)



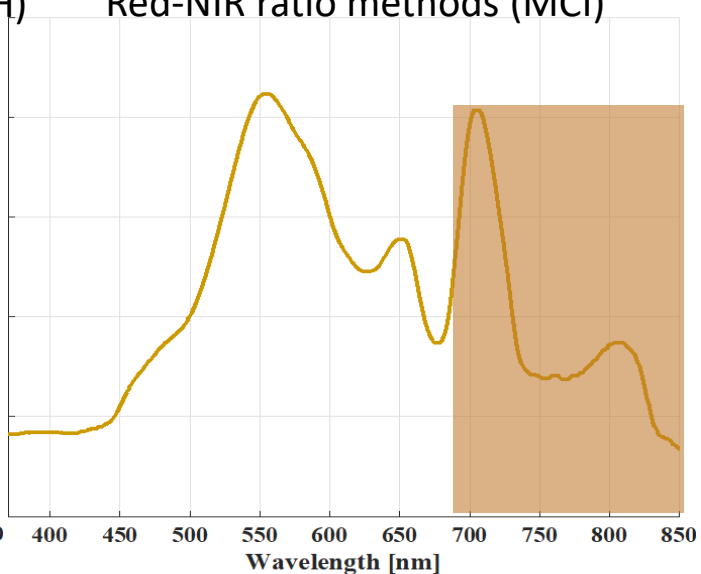
Blue-green ratio methods (OC3)



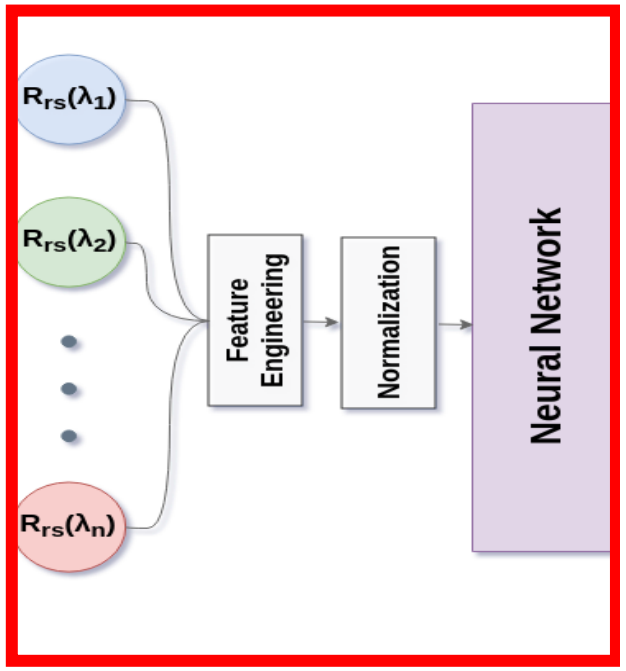

Green-red-NIR ratio methods (FLH)

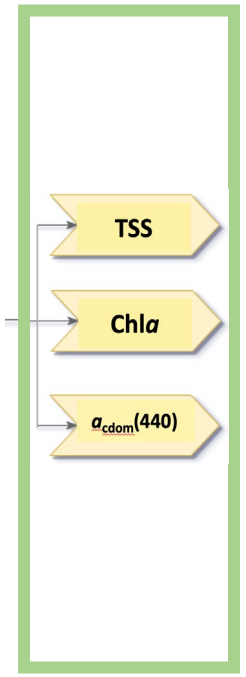


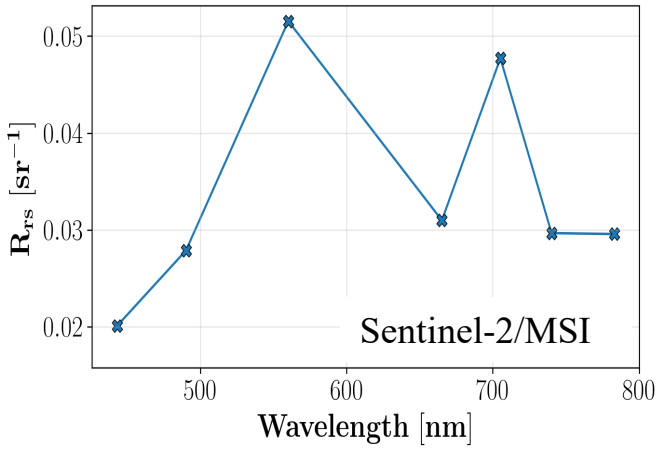

Red-NIR ratio methods (MCI)

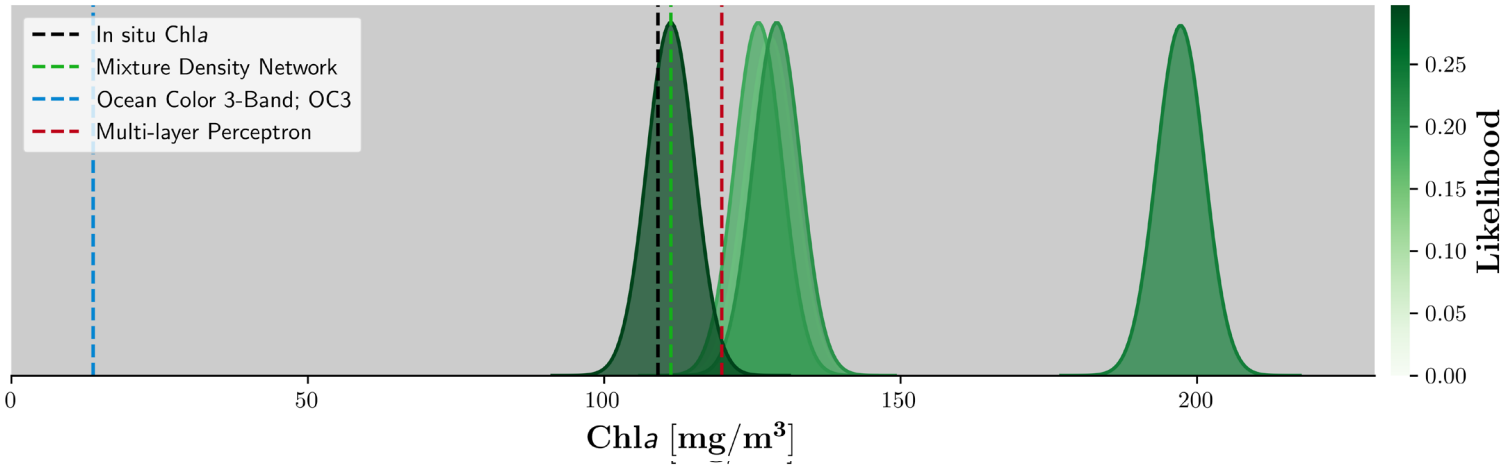


# Inverse Modeling Approach (MDNs)



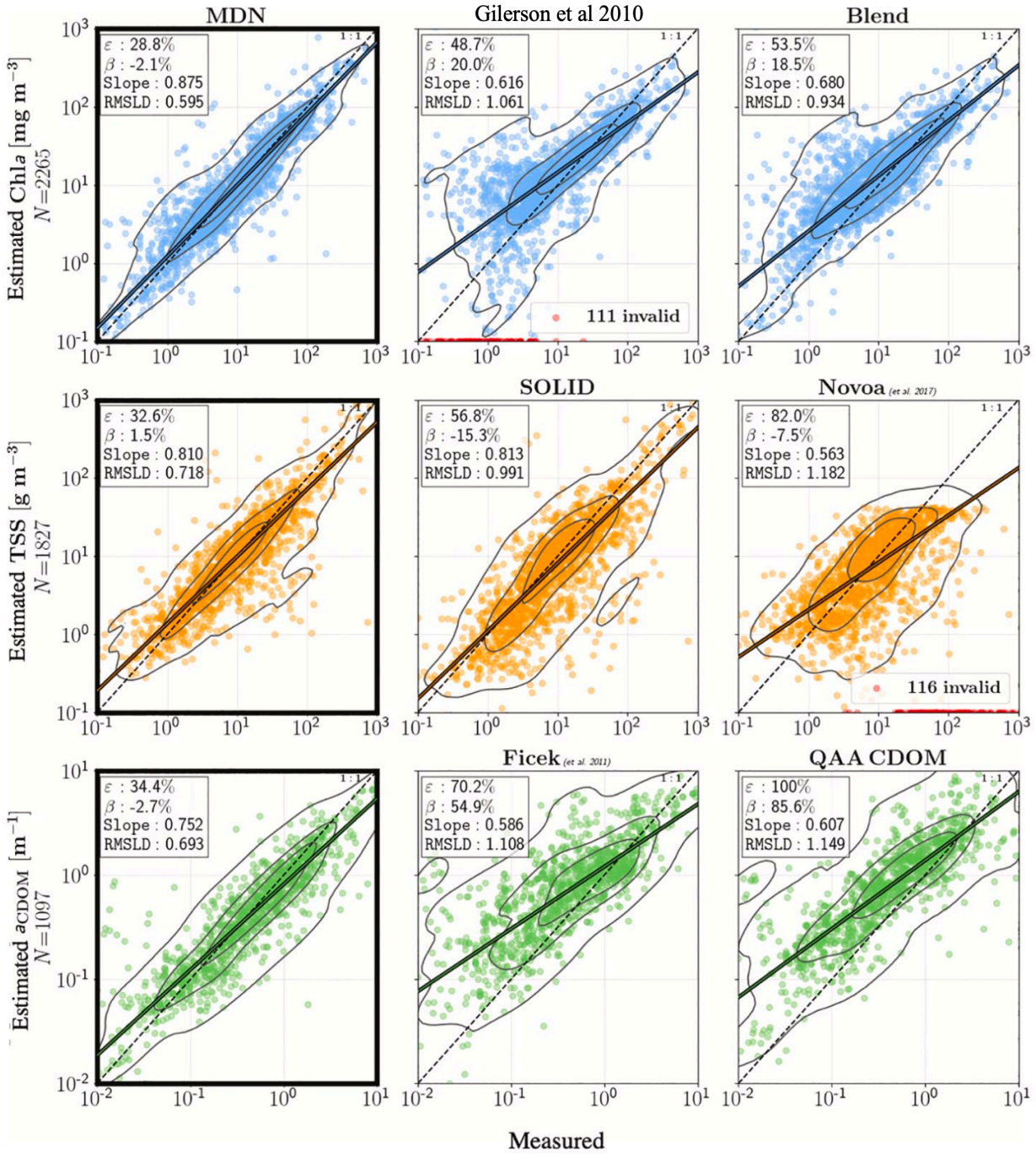



# Performance Assessment (Sentinel-2/MSI)

1. Hold-out (50-50 split)
   - Chl*a*: 10+ algorithms
   - TSS: 8 algorithms
   - $a_{cdom}(440)$: 5 algorithms

| Model | Reference | Chl-a Range (mg/m³) |
|---|---|---|
| Drozd | Drozd et al., 2019 | 4 - 4700 |
| Gurlin (2 Bands) | Gurlin et al., 2011 | 0-100 |
| Gurlin (3 Bands) | Gurlin et al., 2011 | 0-100 |
| MDN | Pahlevan et al., 2020 | < 150 |
| Mishra NDCI | Mishra & Mishra, 2012 | < 30 |
| Moses (2 Bands) | Moses et al., 2009a | 10- 40 |
| Moses (3 Bands) | Moses et al., 2009a | 10- 40 |
| OC2 | O'Reilly & Werdell, 2019 | < 78 |
| OC3 | O'Reilly & Werdell, 2019 | < 78 |
| OC4 | O'Reilly & Werdell, 2019 | < 78 |
| OC5 | O'Reilly & Werdell, 2019 | < 78 |
| OC6 | O'Reilly & Werdell, 2019 | < 78 |
| Smith Blend | Smith et al., 2018 | < 221 |
| Yang Bandindex | Yang et al., 2011 | < 140 |

Chl*a* benchmarks



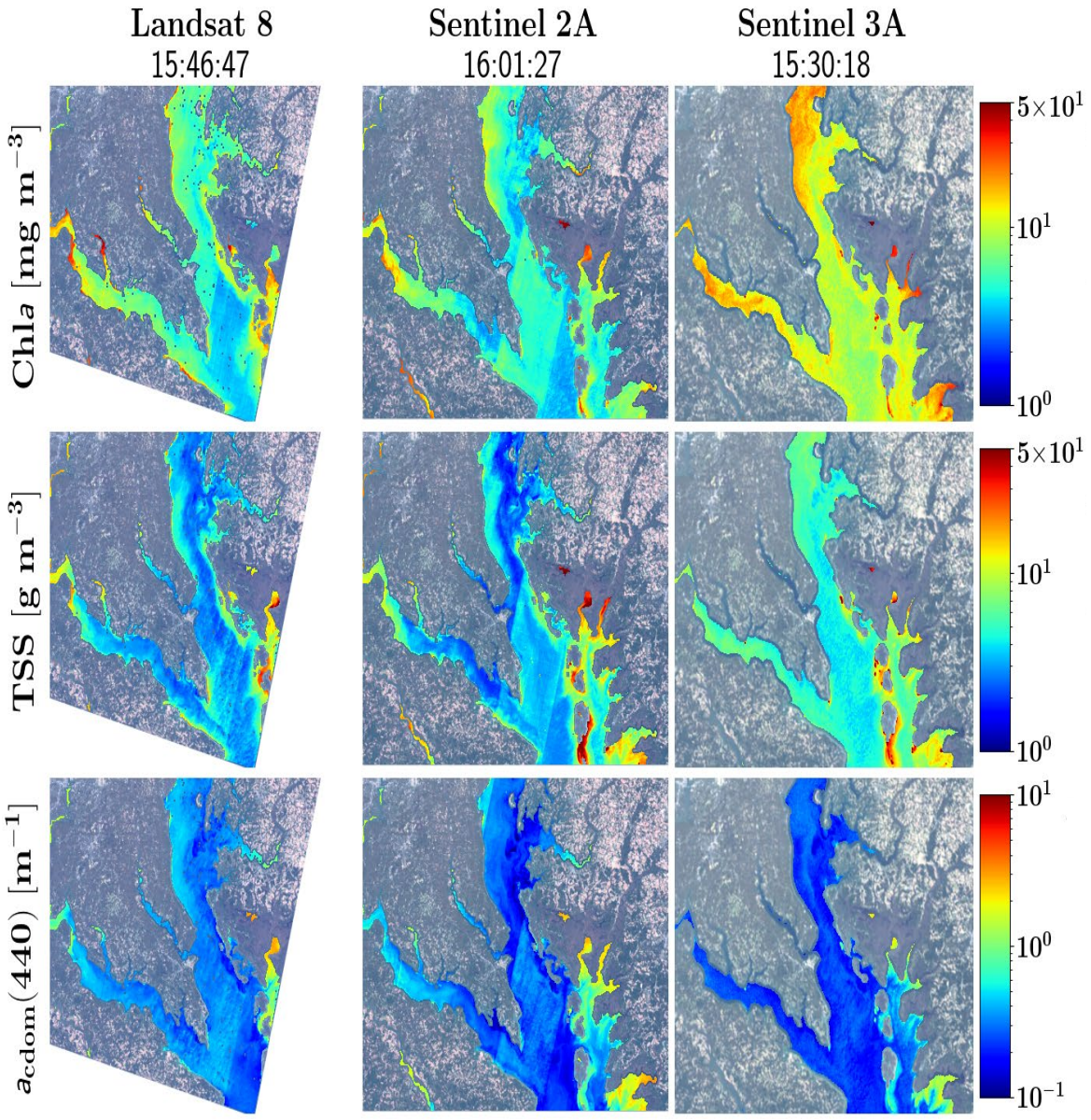
# Performance Assessment (Sentinel-2/MSI)

2. Leave-one-out (LOO)
   - Independently assess the performance using one data source left out in the training
   - Take median across all sources


- With the two assessment methods
  - 32 – 48% (TSS from Sentinel-2/MSI)
  - 26 – 46% (Chla from Sentinel-3/OLCI)
  - 40 – 45% (TSS from Landsat-8/OLI)

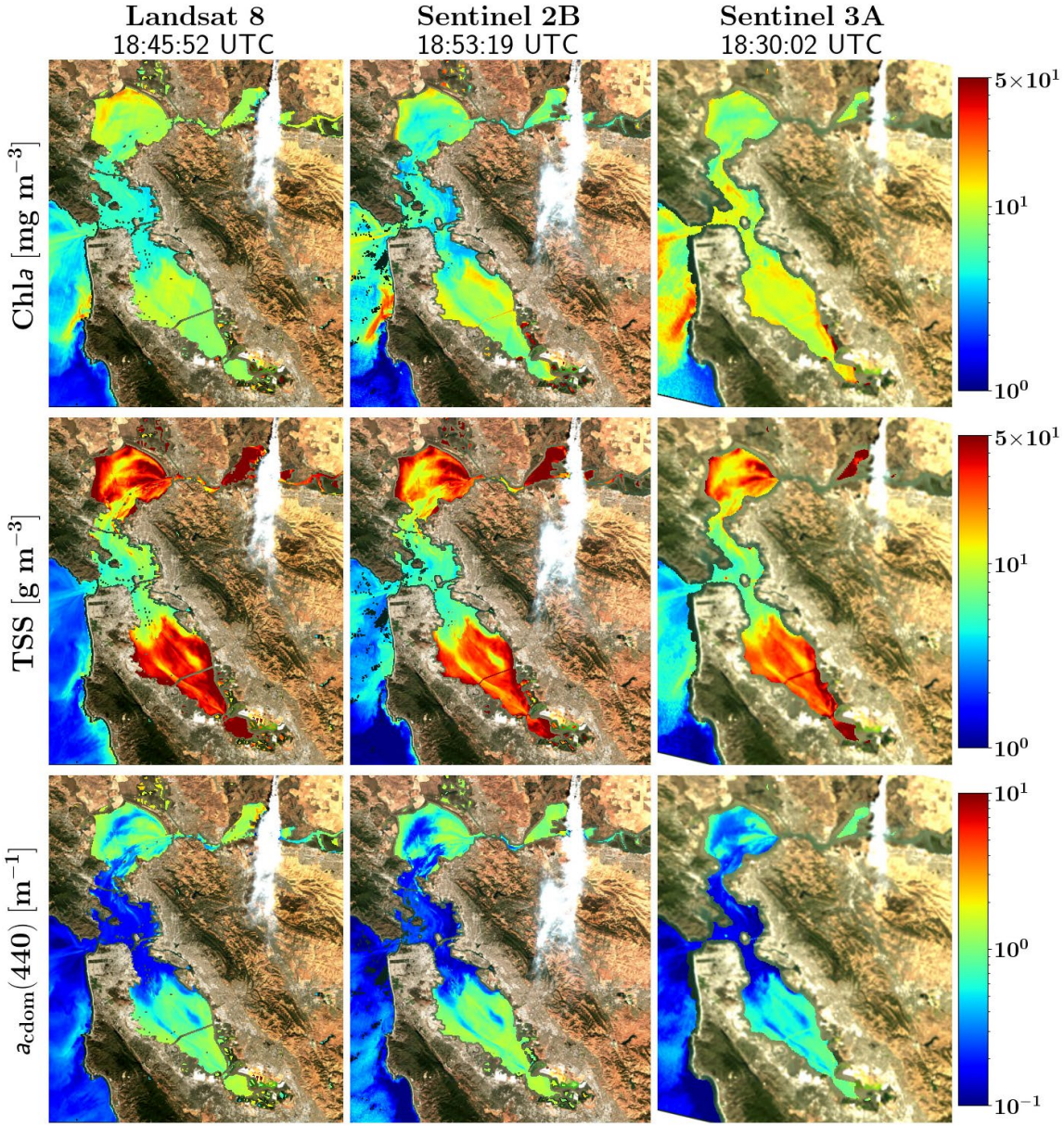
Hold-out LOO

**Nov 7th 2016**




Chesapeake Bay (U.S.)

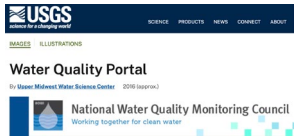
**Oct 7th 2018**



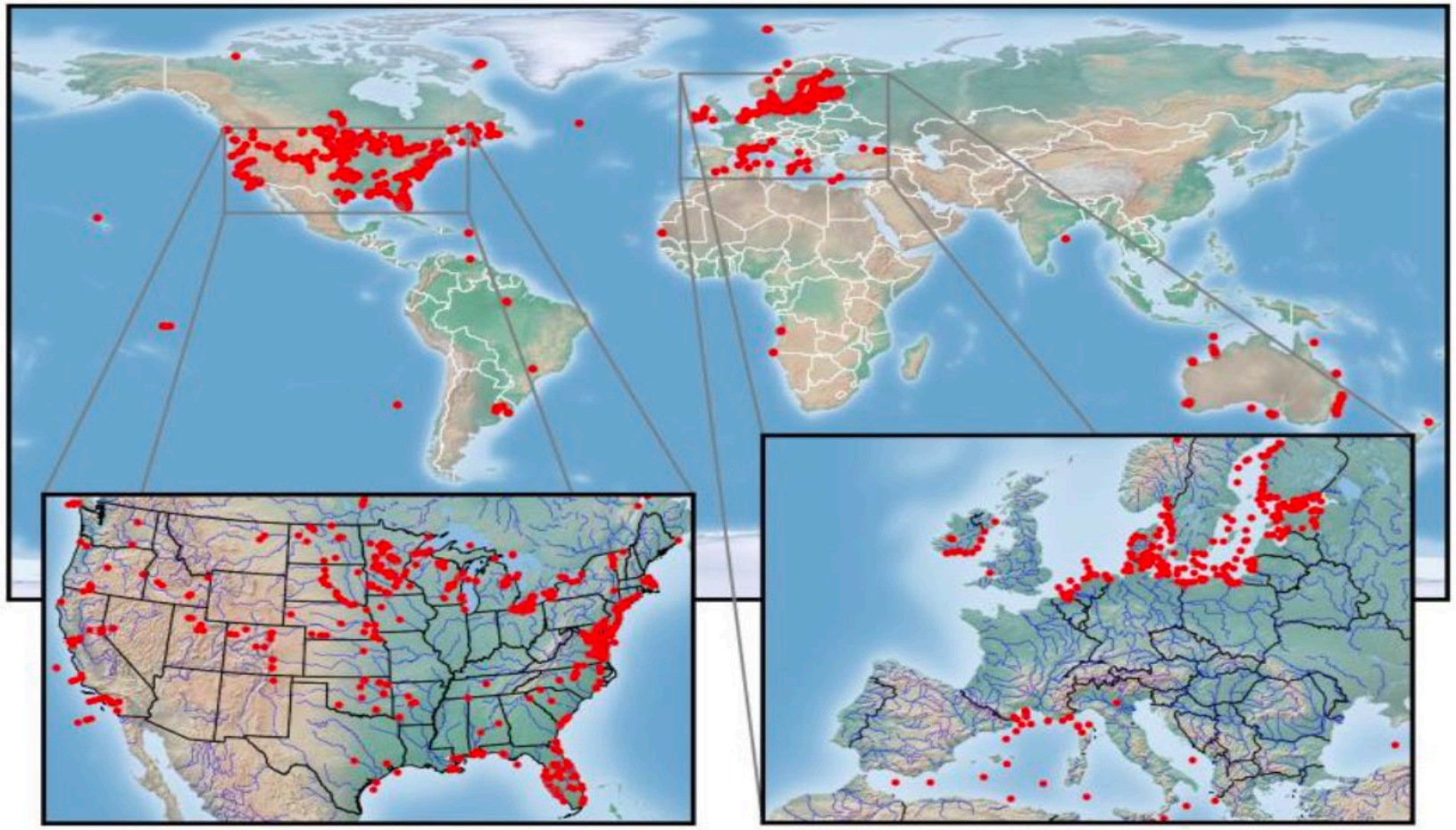

San Francisco Bay (U.S.)

# Validation Data

- **Chl*a*:** USGS Water Quality Portal, USGS National Water Information System, NOAA World Ocean Data (N ~ 8000)
- **TSS:** USGS Water Quality Portal, USGS National Water Information System (N ~ 1000)
- **$a_{cdom}(440)$:** Environment and Climate Change Canada, Lake Pulse Network, University of Minnesota (N ~ 250)



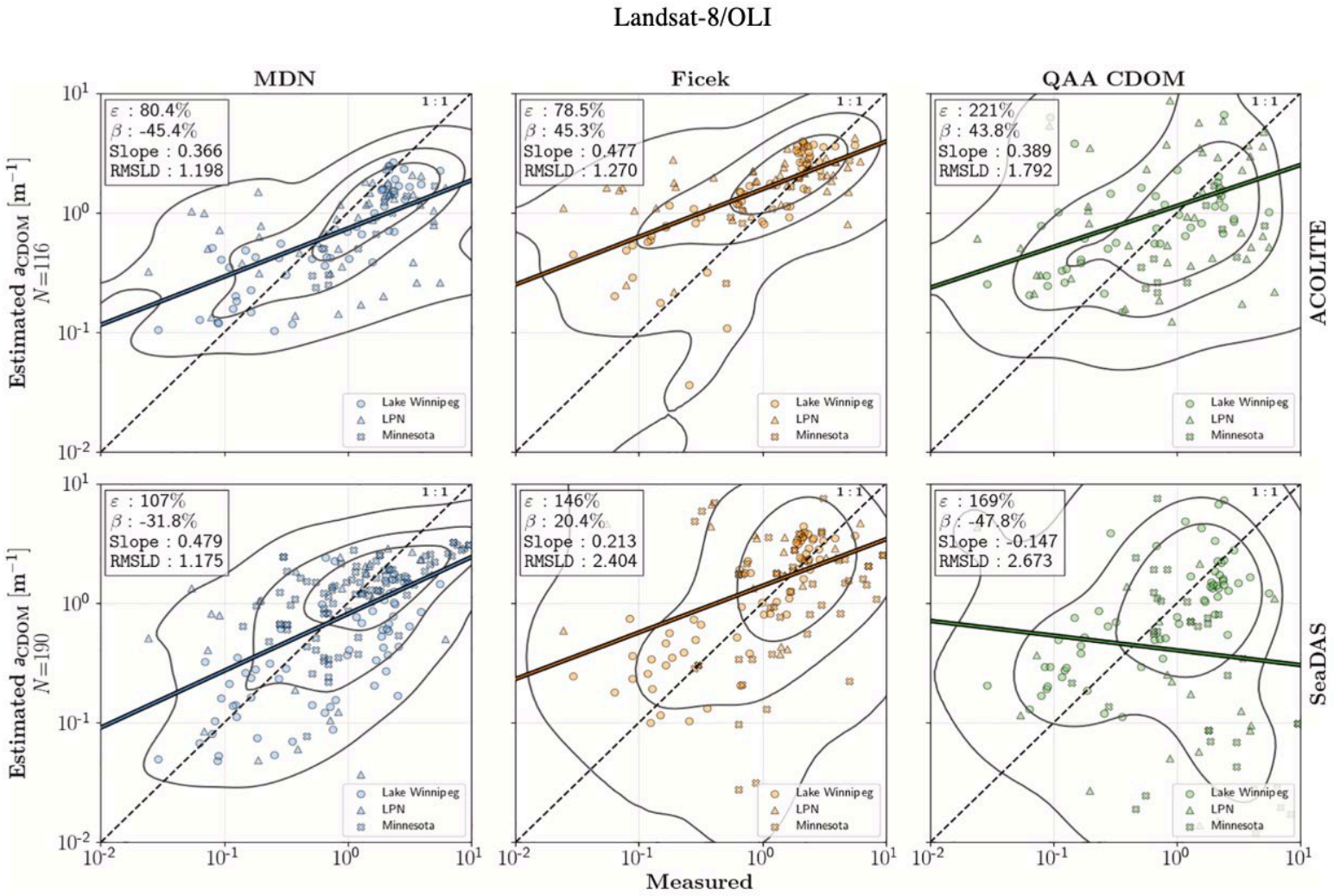


Environment and Climate Change Canada

NSERC Canadian LakePulse.ca Network

# Implementation and Matchup Analysis

- Atmospheric correction processors
  - SeaDAS, ACOLITE, and POLYMER
- Benchmark algorithms
  - Chl*a*: 3 algorithms
  - TSS: SOLID and Novoa
  - $a_{cdom}(440)$: Ficek and QAA CDOM
- Missions
  - Landsat-8 & Sentinel-2



# Summary

- Global validity
  - Model performance is uniform across a broad range of aquatic conditions
    - $0.5 < Chla < 150$ mg $m^{-3}$
    - $0.5 < TSS < 200$ g $m^{-3}$
    - $0.03 < a_{cdom}(440) < 10$ $m^{-1}$
- Product accuracy

| **Sentinel-2 & -3** | |
|---|---|
| Chl*a*, TSS, $a_{cdom}(440)$ | 25 – 55% |
| **Landsat-8 & -9** | |
| TSS & $a_{cdom}(440)$ | 30 – 60% |
| Chl*a* | 50 – 90% |

- Model performance degrades in the presence of uncertainties from atmospheric correction
- Under development
  - Per-pixel uncertainty estimates
  - Extension to hyperspectral domain
  - IOP retrievals
  - Atmospheric correction
- Code access: https://github.com/STREAM-RS
- **GLORIA data access: Early 2023 (See Daniela Gurlin's poster #300)**

Pahlevan, N., Smith, B., *et al*. 2022. Simultaneous retrieval of optical water quality indicators from the Landsat-8, Sentinel-2, and Sentinel-3 missions. *Remote Sensing of Environment, 270, 1128060.*

Lake Erie; West Sister Island
Landsat/OLI; Sep 27th 2014



## ACKNOWLEDGEMENT

**United States Geological Survey**

*Landsat Science Team*

**NASA Earth Science Programs**

*Ocean Biology and Biogeochemistry*

*Remote Sensing of Water Quality*



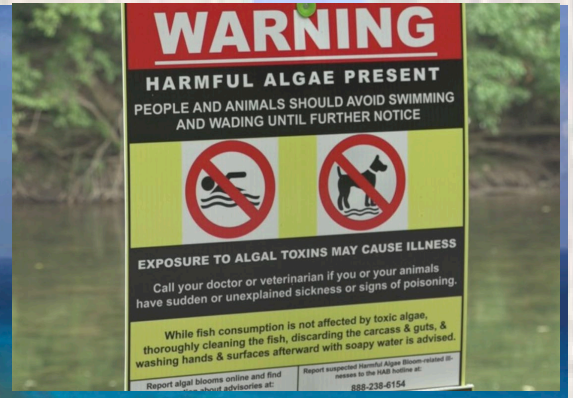
Contact Info:
Nima Pahlevan
nima.pahlevan@nasa.gov

# Backup

# Challenges Surrounding Coastal and Freshwater Ecosystems

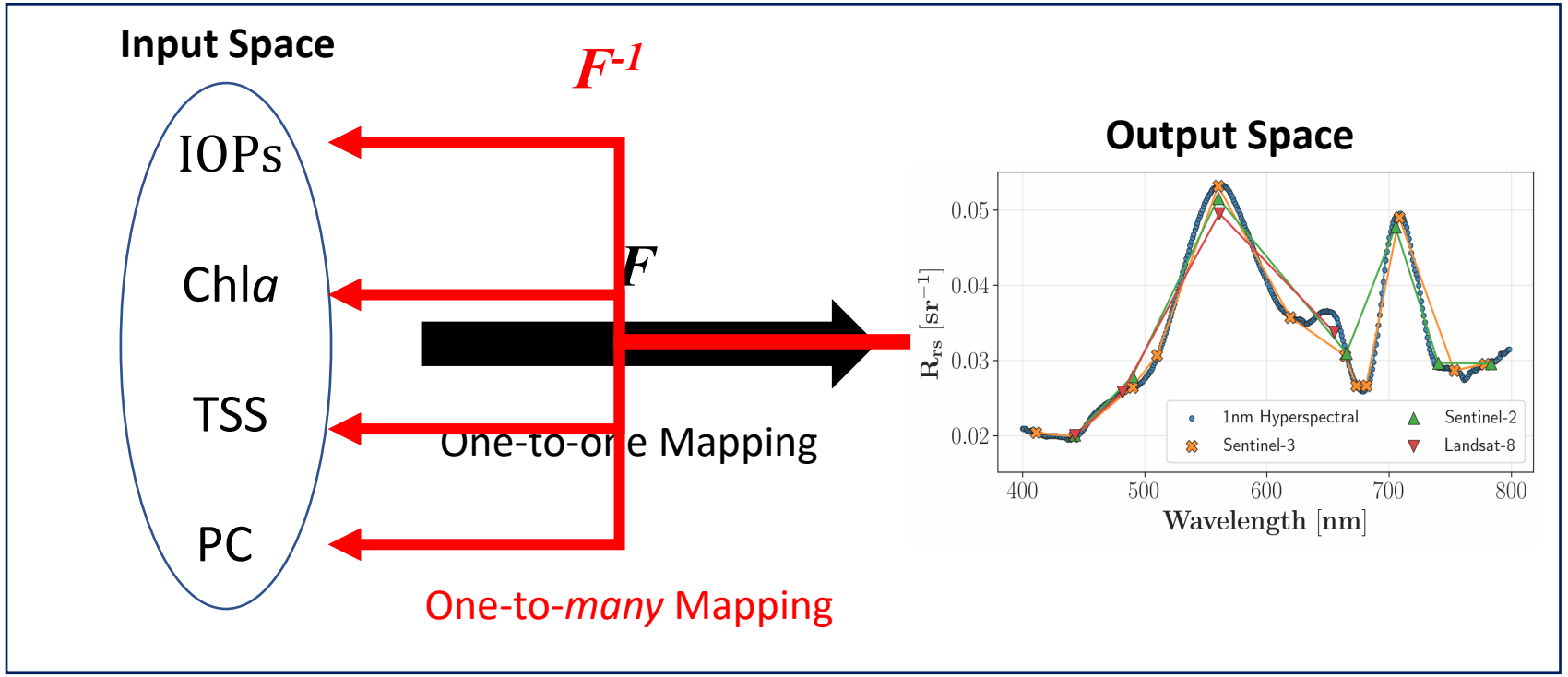
- Aquatic plants multiply under favorable conditions
  - Temperature
  - Nutrients
  - Physical forcing
- Sedimentation
  - Pollution transport
  - Trigger algal blooms
- Dissolved organic matter
  - Warming waters

Freshwater HABs result in economic losses of more than US$4 billion annually in the United States


$$R_{rs} = f(\text{IOP}, \text{Chl}a, \text{TSS}, \dots)$$

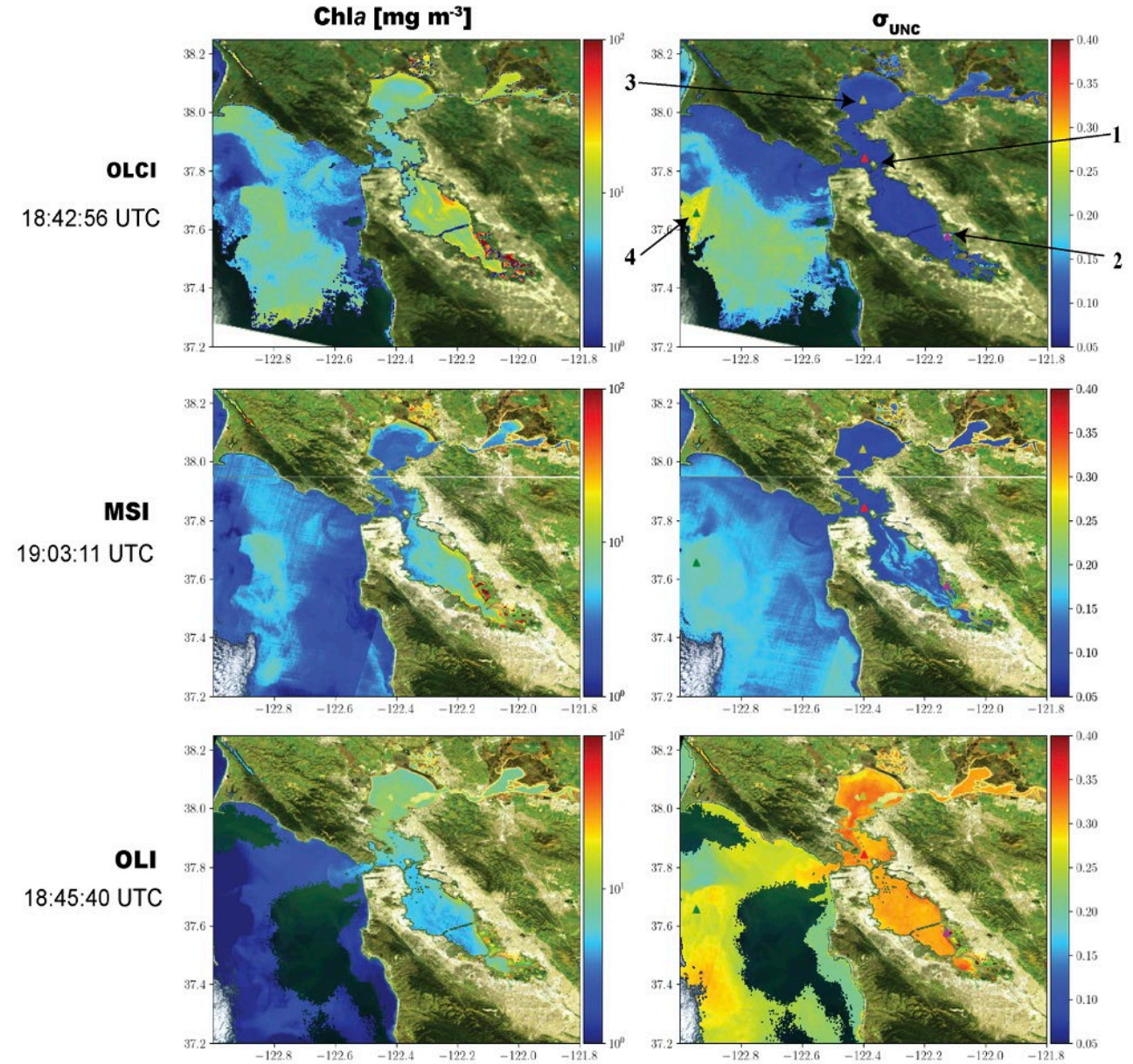
# C. Inverse Problem




- **I**nherent **O**ptical **P**roperties (**IOPs**)
  - $a_{ph}$, $b_{bp}$, $a_{nap}$, $a_{cdom}$, VSF, …
- *Pigment concentrations*
  - Chl*a,* phycocyanin *(PC),* phycoerythrin *(PE),* …

- Other dependencies
  - Total Suspended Solids (TSS)
  - Acquisition geometry

# Per-pixel Uncertainty Estimation



# Landsat-8 Matchups