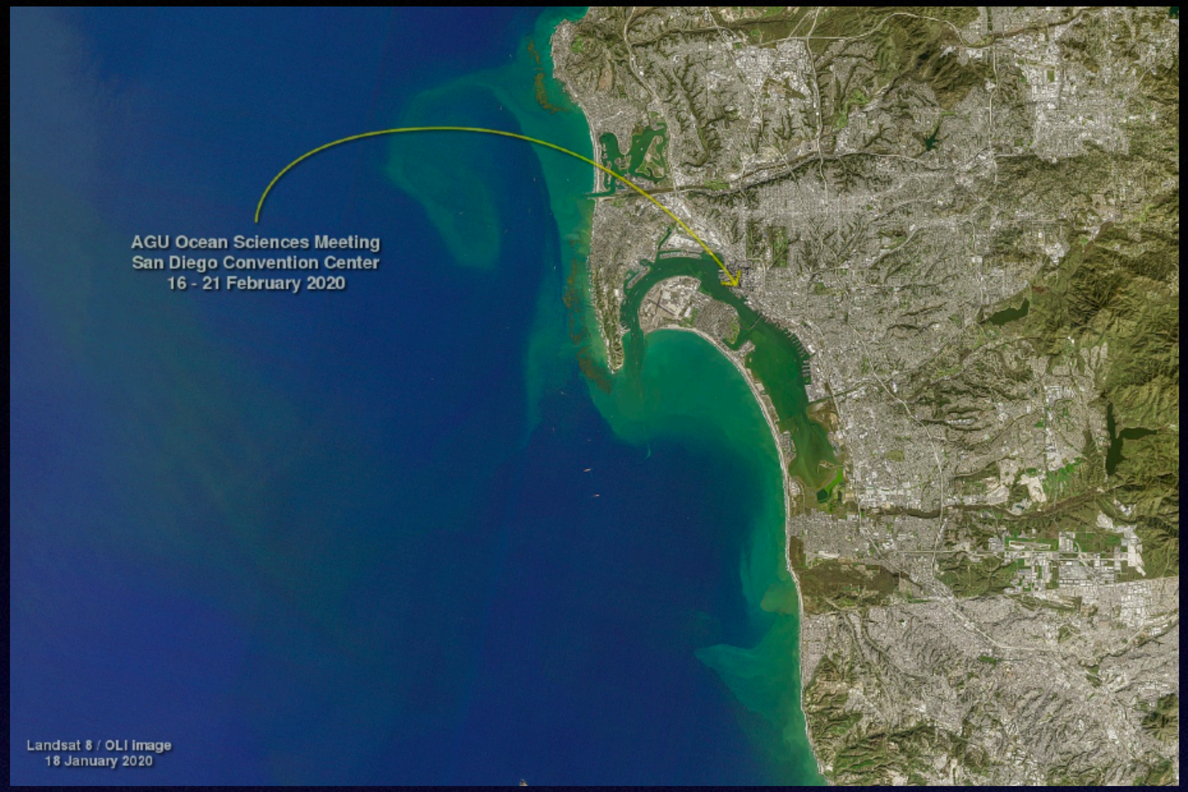
Remote Sensing of Ocean Color in the Coastal Zone: Challenges and Solutions

Susanne E. Craig
NASA Goddard Space Flight Center, Ocean Ecology Laboratory/UMBC

Ocean Color in the Coastal Ocean: The Problem

- Chlorophyll (Chl) and inherent optical properties (IOPs) are proxies of biogeochemical processes in the ocean



- Can be accurately estimated from satellite ocean color in many regions of the world ocean ➙ band ratio & semi-analytical algorithms

- However, frequently not able to <u>quantitatively</u> utilize satellite ocean color in a variety of scenarios, e.g.:
  - model assumptions violated
  - difficulty in removing the atmospheric contribution to the signal detected by the satellite

    ➙ *typically over coastal & inland waters*

- Alternative approaches required if we are to fully exploit satellite ocean color data collected over these systems

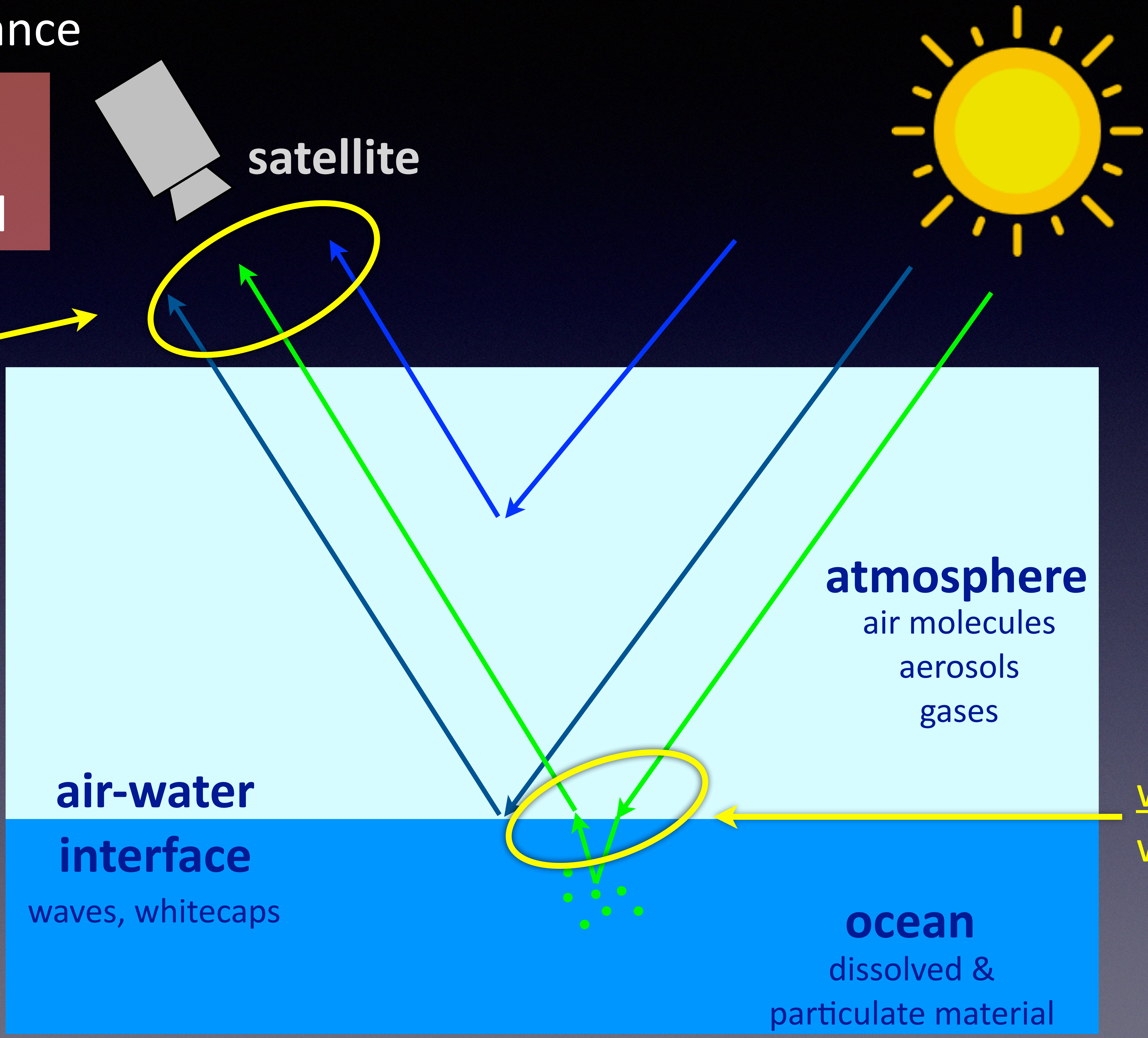
Imagery: <u>https://oceancolor.gsfc.nasa.gov</u>

Ocean Color

L_t - top of atmosphere radiance

water-leaving signal contributes a max. of 10% of the total signal

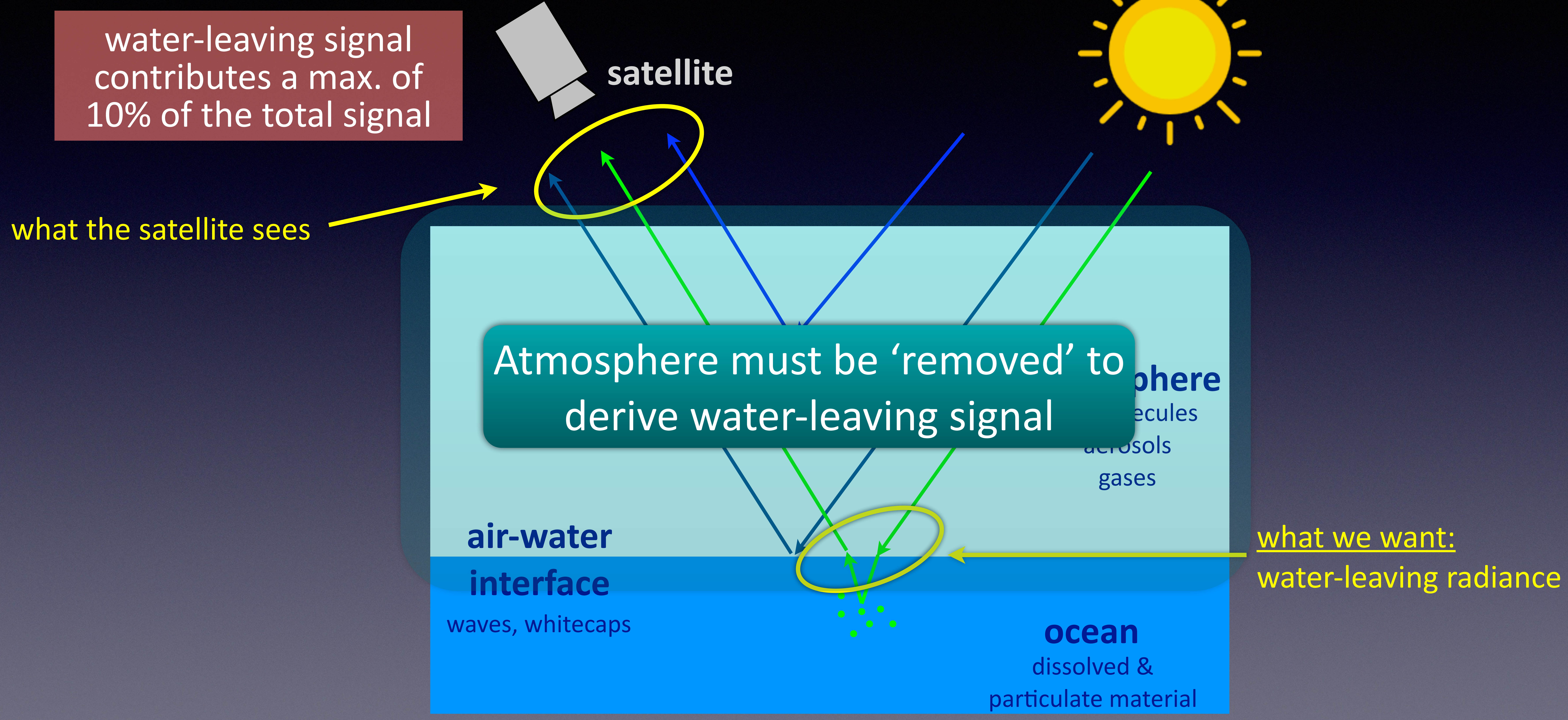
what the satellite sees



what we want:
water-leaving radiance

Ocean Color

L_t - top of atmosphere radiance



Atmospheric Correction of Ocean Color

top of atmos.

white caps

water-leaving

$$L_t(\lambda) = \underbrace{L_r(\lambda) + L_a(\lambda) + L_{ra}(\lambda)}_{\text{Rayleigh \& aerosol}} + \overbrace{t(\lambda)L_{wc}(\lambda)}^{\text{white caps}} + \underbrace{T(\lambda)L_g(\lambda)}_{\text{sun glint}} + t(\lambda)t_0(\lambda)\cos\theta_0\left[L_w(\lambda)\right]_N$$

Rayleigh & aerosol

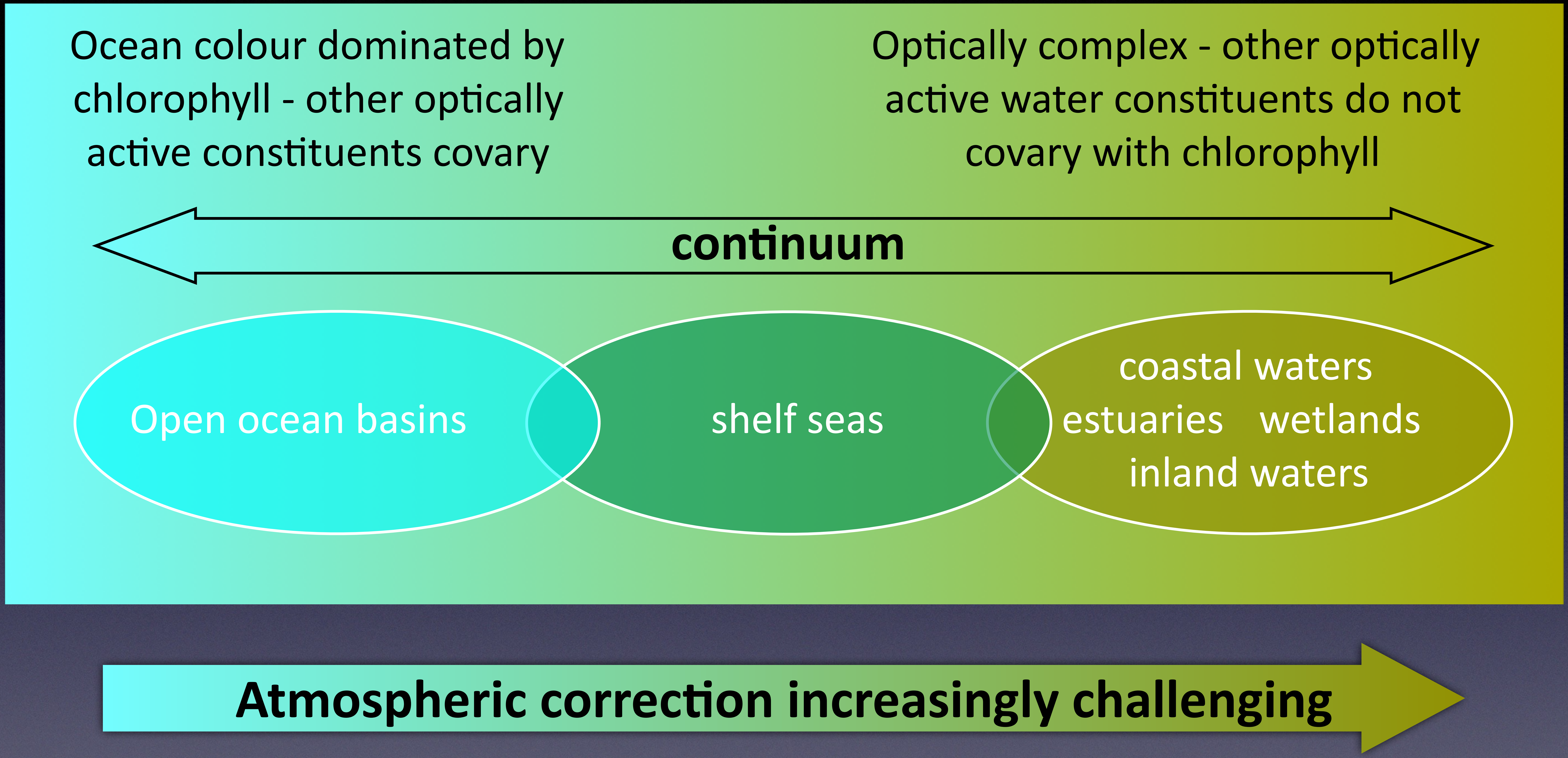
sun glint

What the satellite sees

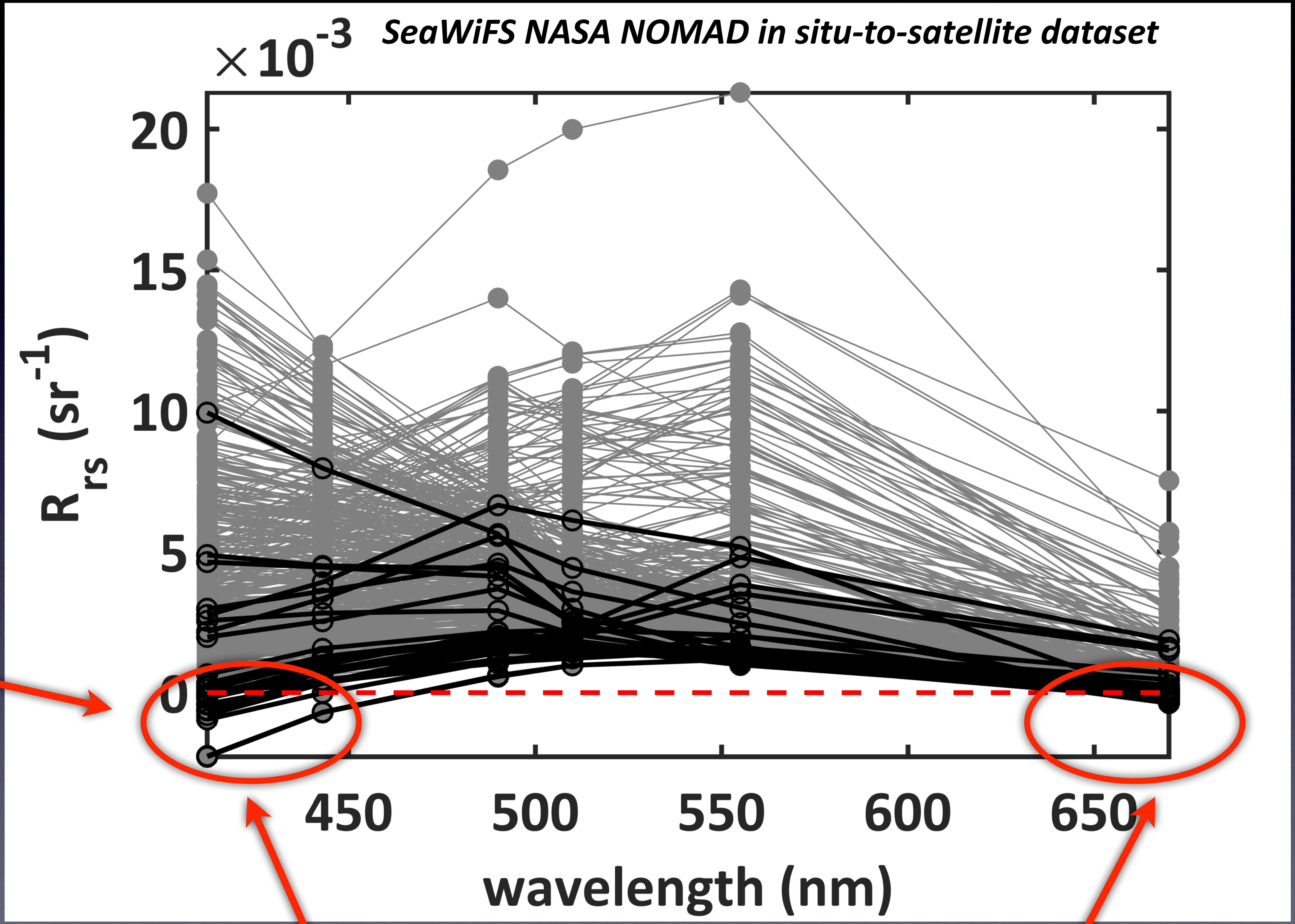
Atmosphere ‘subtracted’ using measurements, models & assumptions

What we want

Atmospheric Correction of Ocean Color



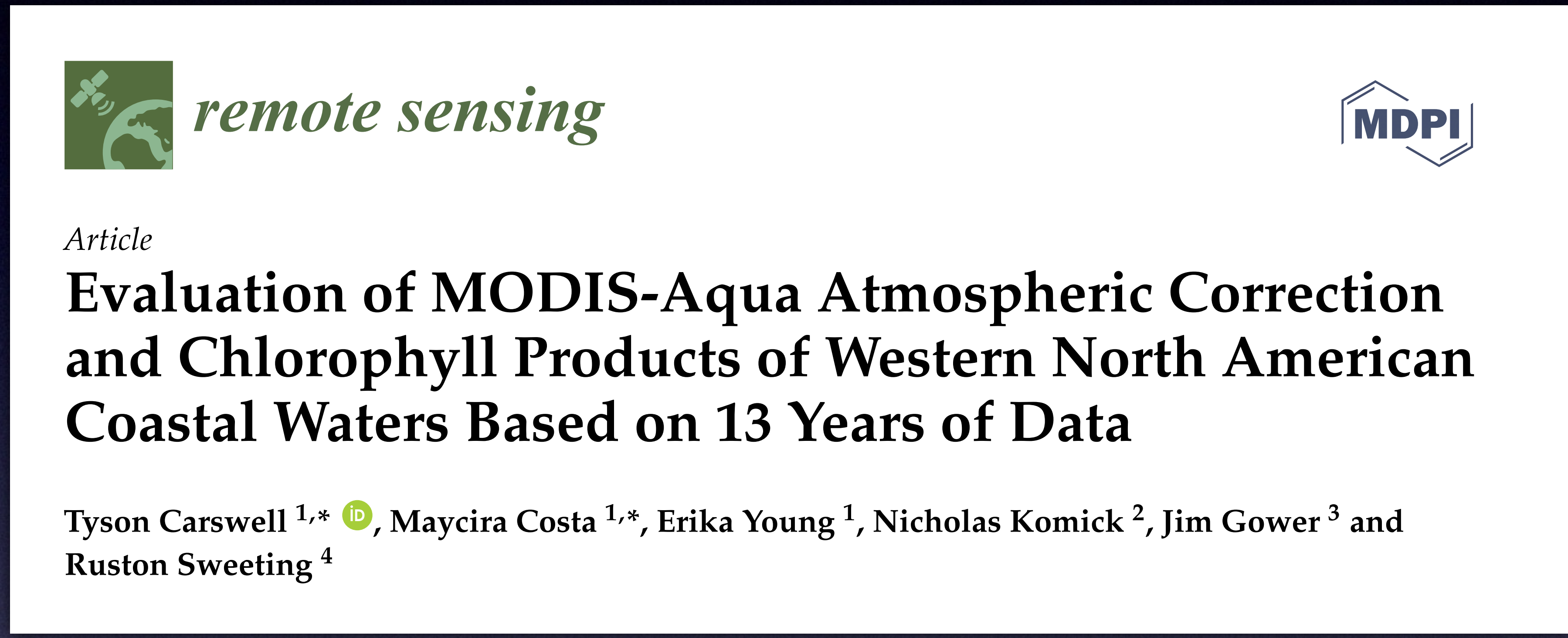
Atmospheric Correction Produces Negative Reflectance



AC failure typically affects blue wavelengths most strongly

‘Too much’ atmosphere subtracted → negative reflectance

Pixels near coastlines are frequently flagged for failing atmospheric correction


remote sensing

MDPI

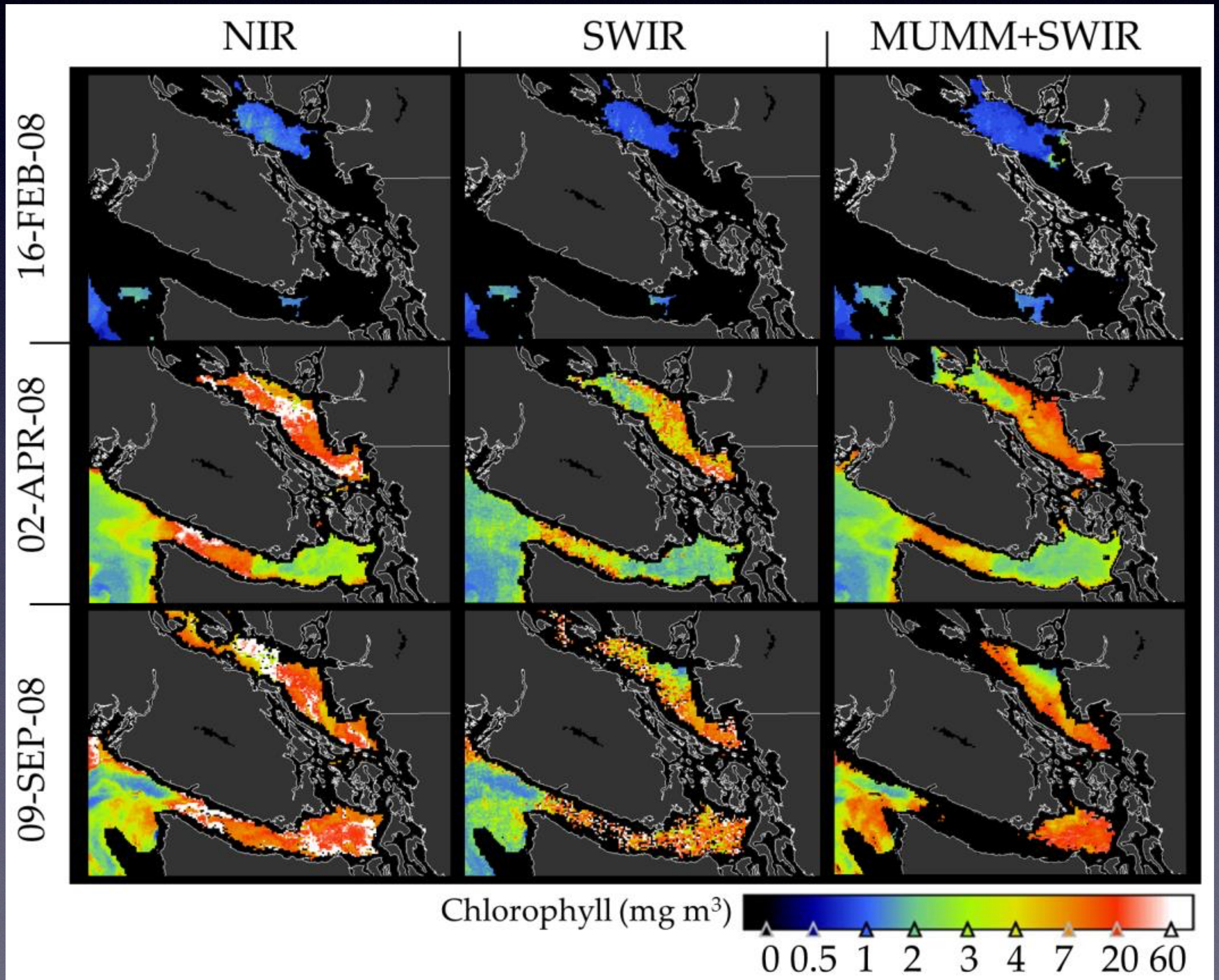
Article

Evaluation of MODIS-Aqua Atmospheric Correction and Chlorophyll Products of Western North American Coastal Waters Based on 13 Years of Data

Tyson Carswell [1,*], Maycira Costa [1,*], Erika Young [1], Nicholas Komick [2], Jim Gower [3] and Ruston Sweeting [4]

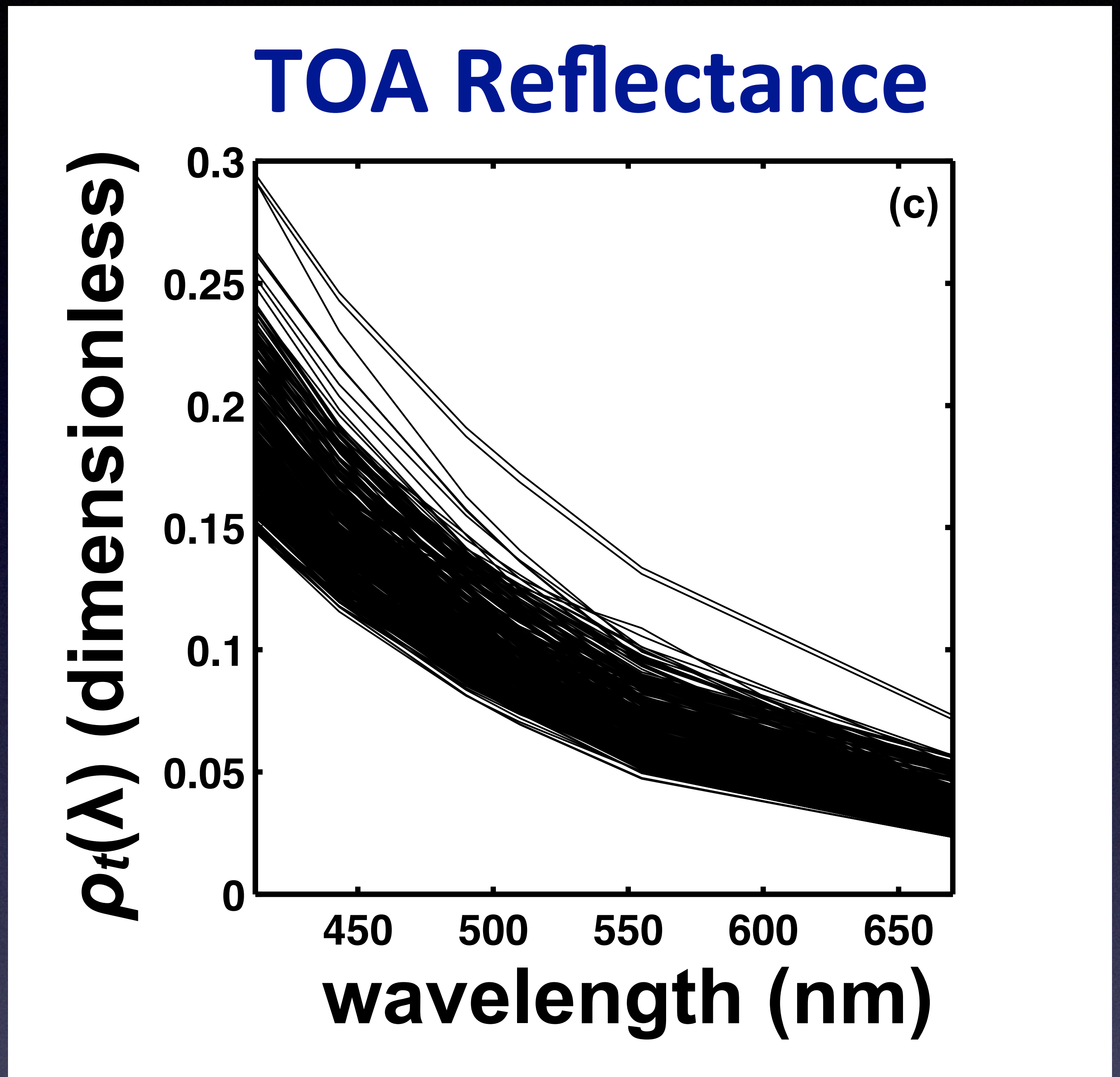
Carswell et al. (2017). Remote Sensing, 9, 1063, https://doi.org/10.3390/rs9101063

Example of MODIS-derived chlorophyll distribution



An Alternative Approach

1.



2. EOF analysis

stepwise selection of EOF scores

3. Model to estimate IOPs/Chl

Multiple linear regression

IOPs, *Chl*

Response

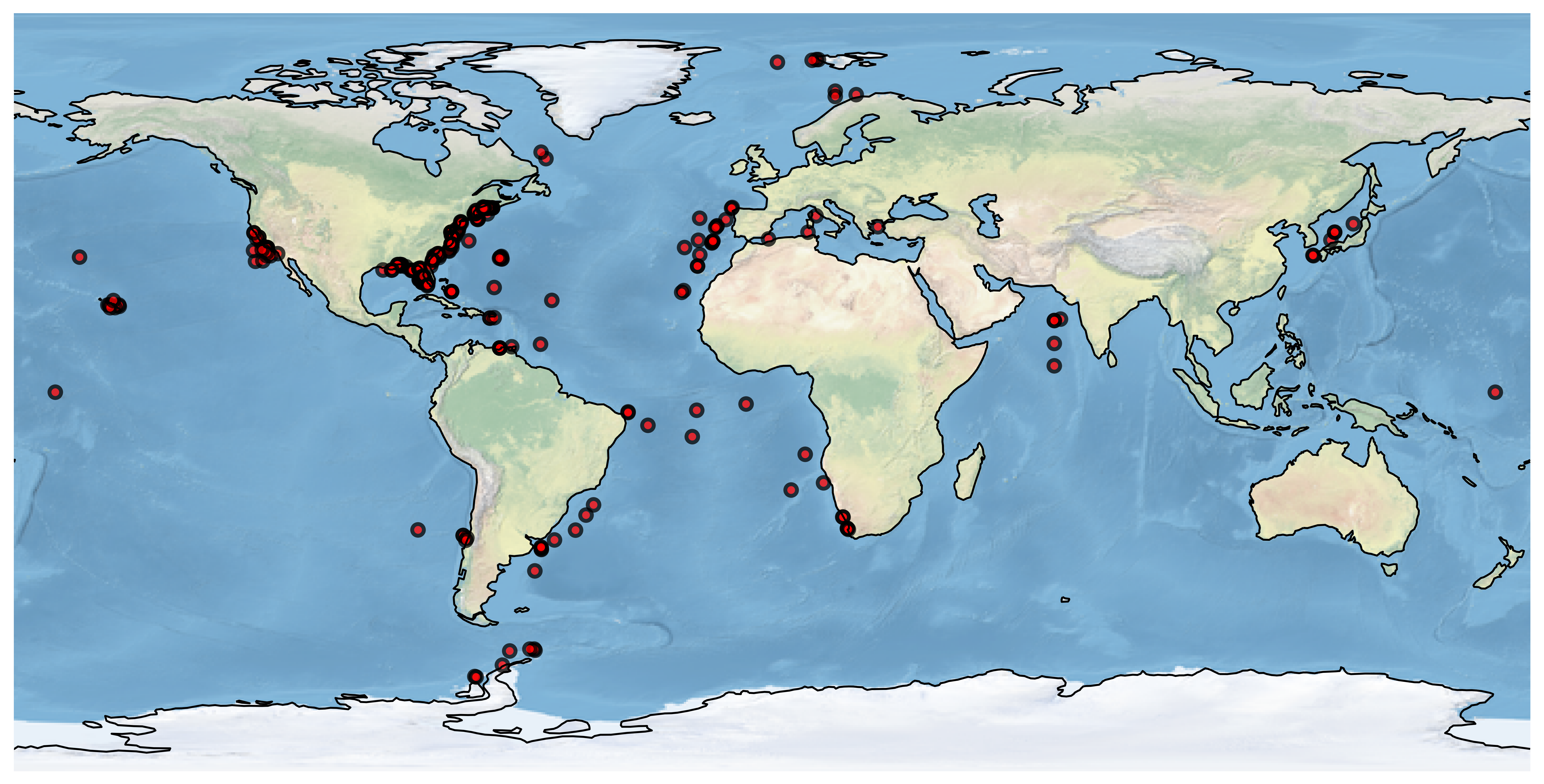
EOF scores

Predictor variables

c.f. Craig et al. (2012), Remote Sensing of Envitonment, 119, 72-83

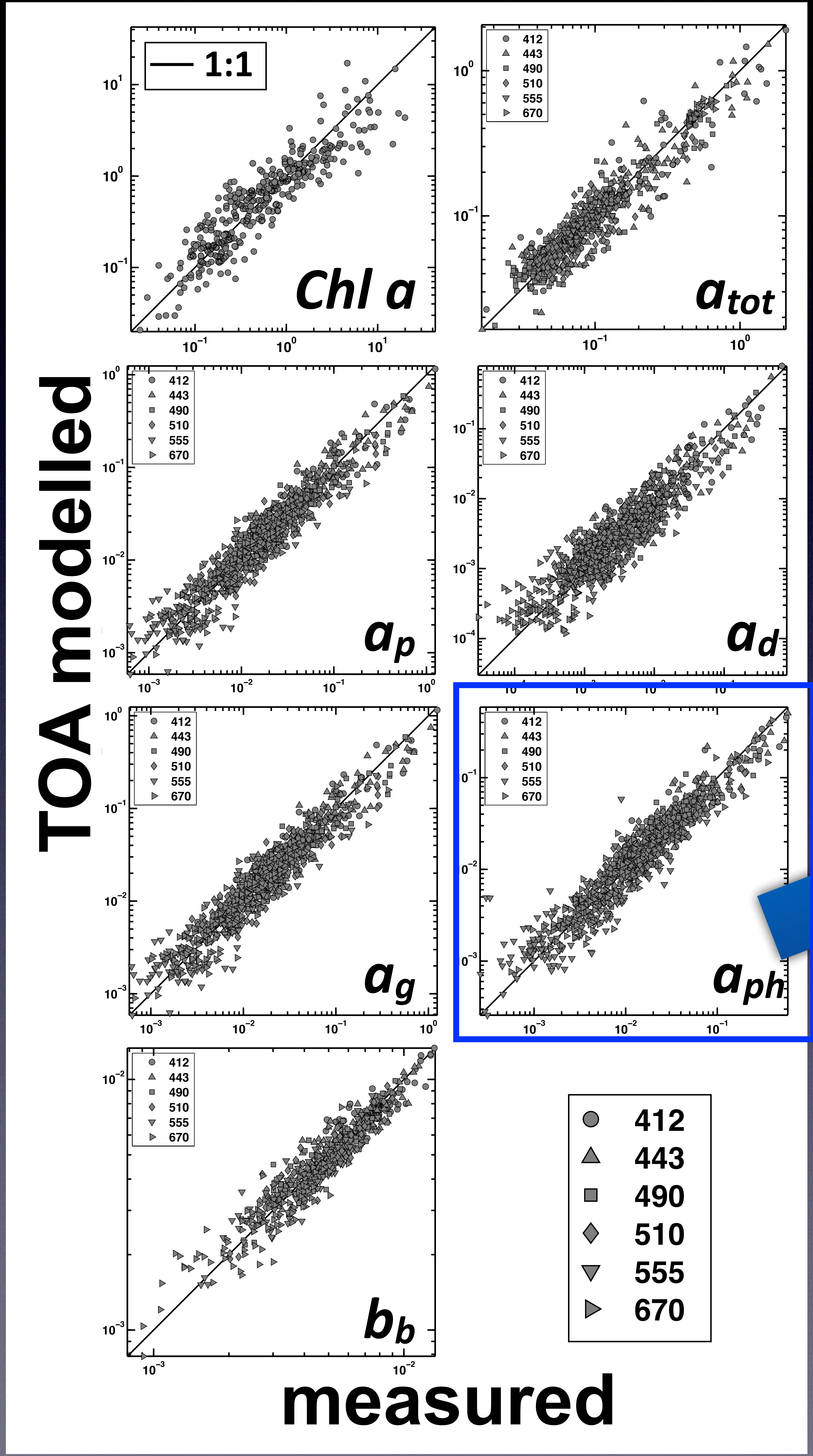
- Don’t try to remove the atmosphere’s contribution
- Use EOF analysis to detect the underlying variance due to water signals

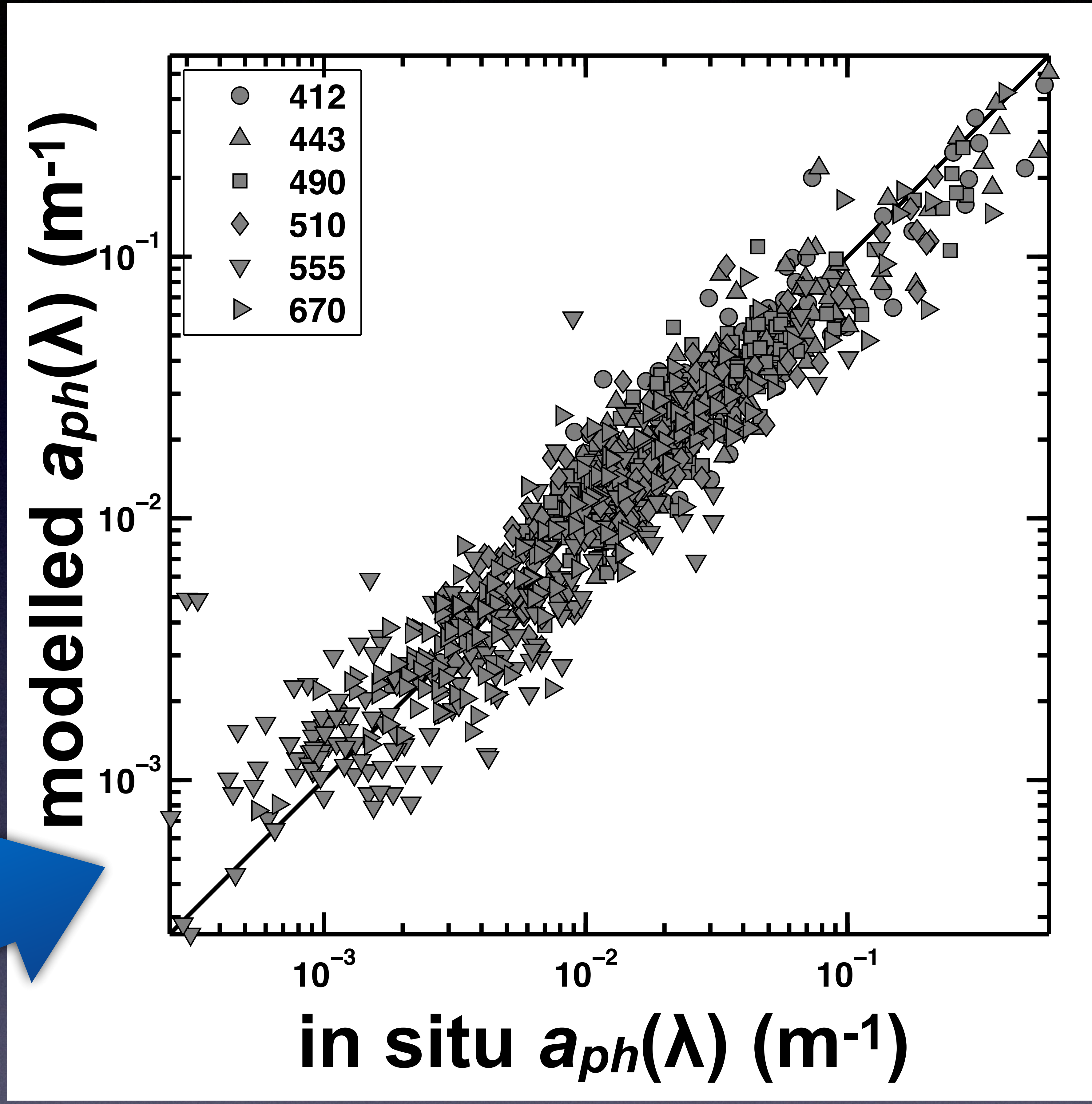
Geographical Distribution of Data Points Used in Model Development



NASA NOMAD SeaWiFS-to-in situ matchup dataset
https://seabass.gsfc.nasa.gov/

Results - EOF Analysis of Top of Atmosphere (TOA) Spectra





a_{ph}: optical metric of phytoplankton biomass

Cross-validation statistics

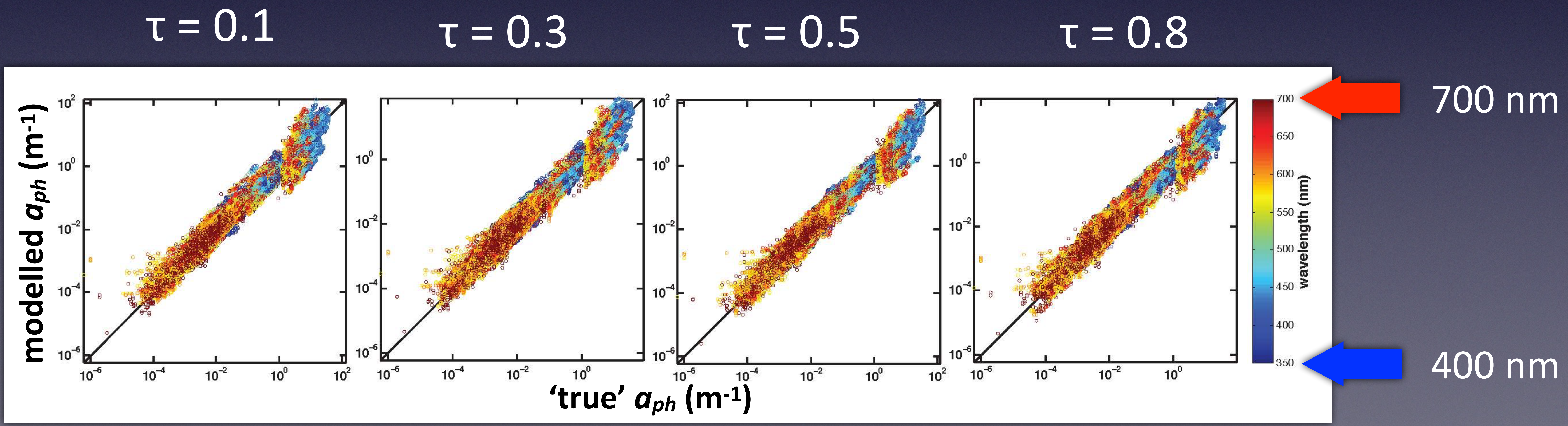
λ(nm)	r^2 (N = 163)	RMSE
412	0.801	0.178
443	0.808	0.169
490	0.825	0.164
510	0.811	0.187
555	0.706	0.301
670	0.867	0.204

N = 163, training:test 80:20, 5000 trials

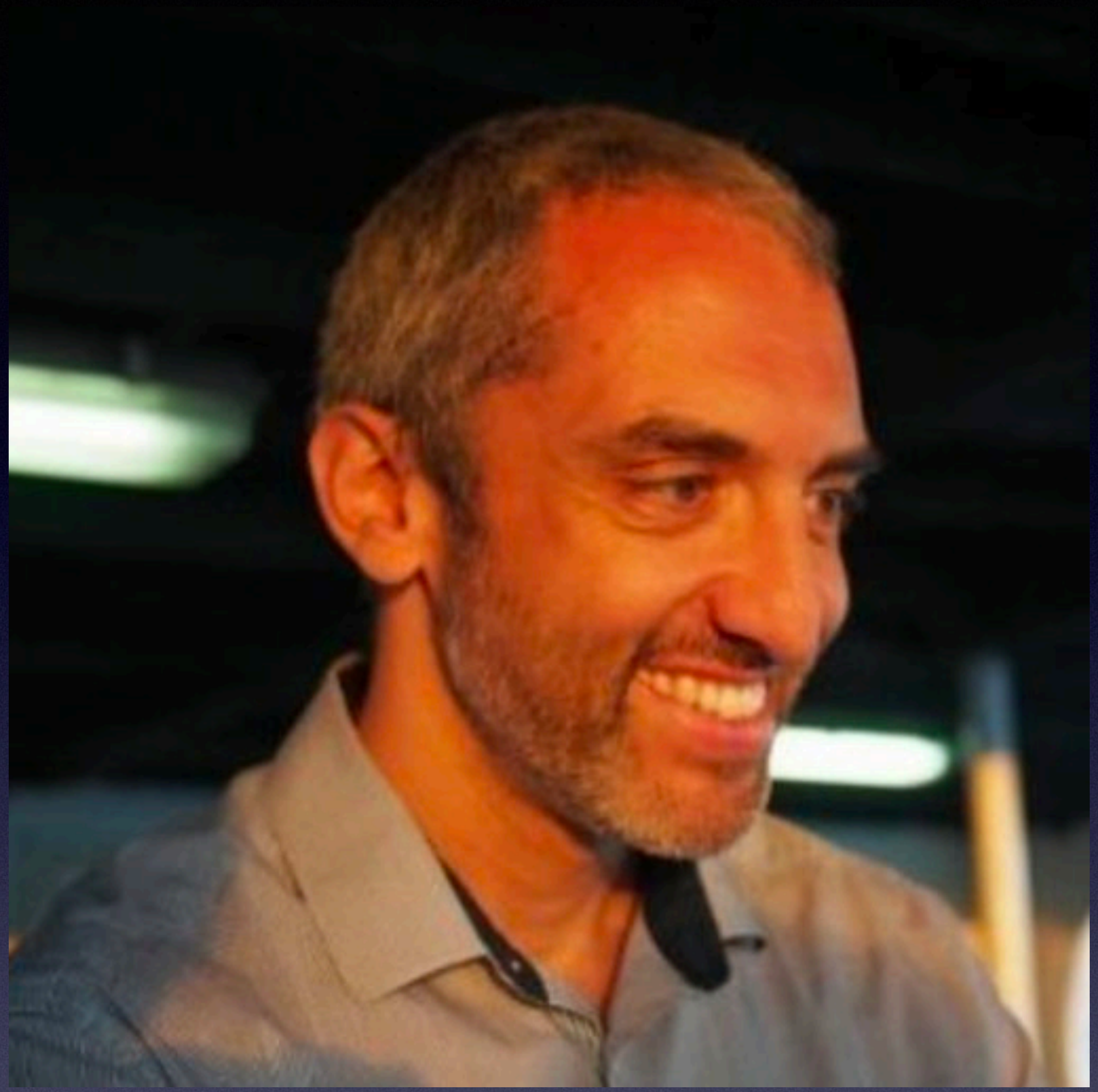
- Very similar statistics compared with model fitted to entire dataset
- Robust, likely not overtrained

Sensitivity Analysis - variable atmosphere & water constituents

- Synthetic dataset created using a coupled ocean-atmosphere radiative transfer model
- Variable IOPs (c.f IOCCG Report No.5), variable AOD (τ), & absorbing aerosols
- TOA-EOF models derived for IOPs & Chl
- Found to perform well & cross-validation suggested robust, generalizable models



Approach evolved into Bayesian machine learning model



Erdem Karaköylü - Data Science Consulting

Bayesian Models for Deriving Biogeochemical Information from Satellite Ocean Color

Susanne E. Craig[1,2] **and Erdem M. Karaköylü**[1,3]

[1]Ocean Ecology Laboratory, NASA Goddard Space Flight Center; [2]Universities Space Research Association; [3]Science Applications International Corporation

Craig and Karaköylü (2019), Earth ArXiv, https://eartharxiv.org/repository/dashboard/557/10.31223/osf.io/shp6y

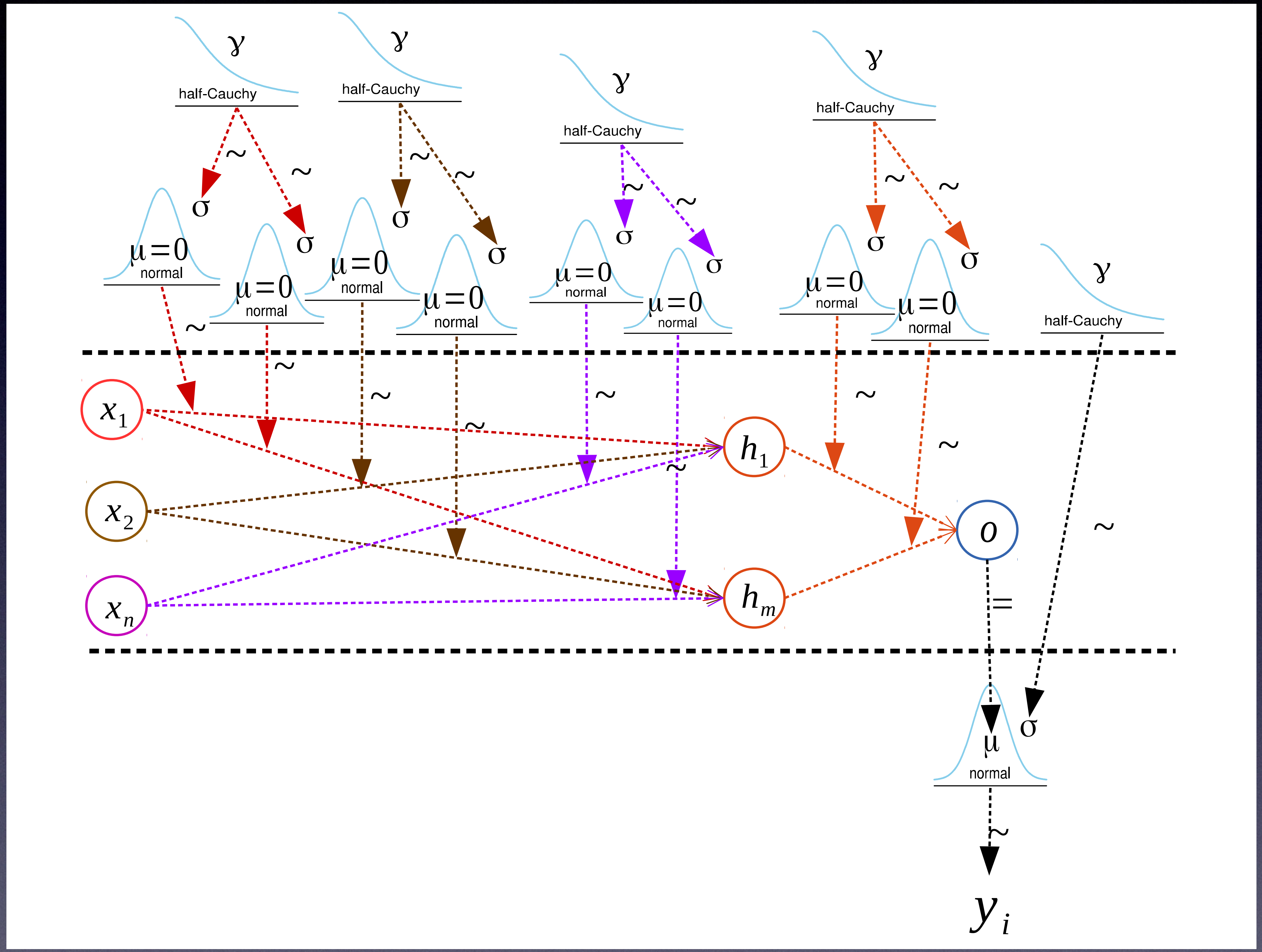
Bayesian Machine Learning Approaches

- The ocean is chronically under sampled ➝ sparse data ➝ overfitted, non-generalizable models
- Bayesian machine learning approaches:
  - Intrinsically mitigate overfitting
  - Principled modeling approach - allows the inclusion of prior knowledge:
    - ✳ ***Requires a close collaboration between the domain specialist and the data scientist***
    - ✳ ***Embrace what the human practitioner knows about the system***
  - Can integrate measurement uncertainty to give more accurate overall uncertainty estimation - probability distributions come for free
  - Readily adaptable to the ‘clustering’ that often occurs in nature
  - Model can be updated as more labeled data is collected - satellites, *in situ* campaigns, autonomous assets

Bayesian Neural Network for IOP Prediction from TOA Spectra

- Same NOMAD SeaWiFS-to *in situ* dataset used
- Models derived for spectral phytoplankton absorption, $a_{ph}(\lambda)$
- Built for automatic relevance determination
- Model features:
    - ✦ EOFs of TOA spectra, as in original models
    - ✦ Metrics of latitude & longitude
    - ✦ Day of year
    - ✦ Sea surface temperature
    - ✦ Bathymetry

Inference diagram of the Bayesian NN



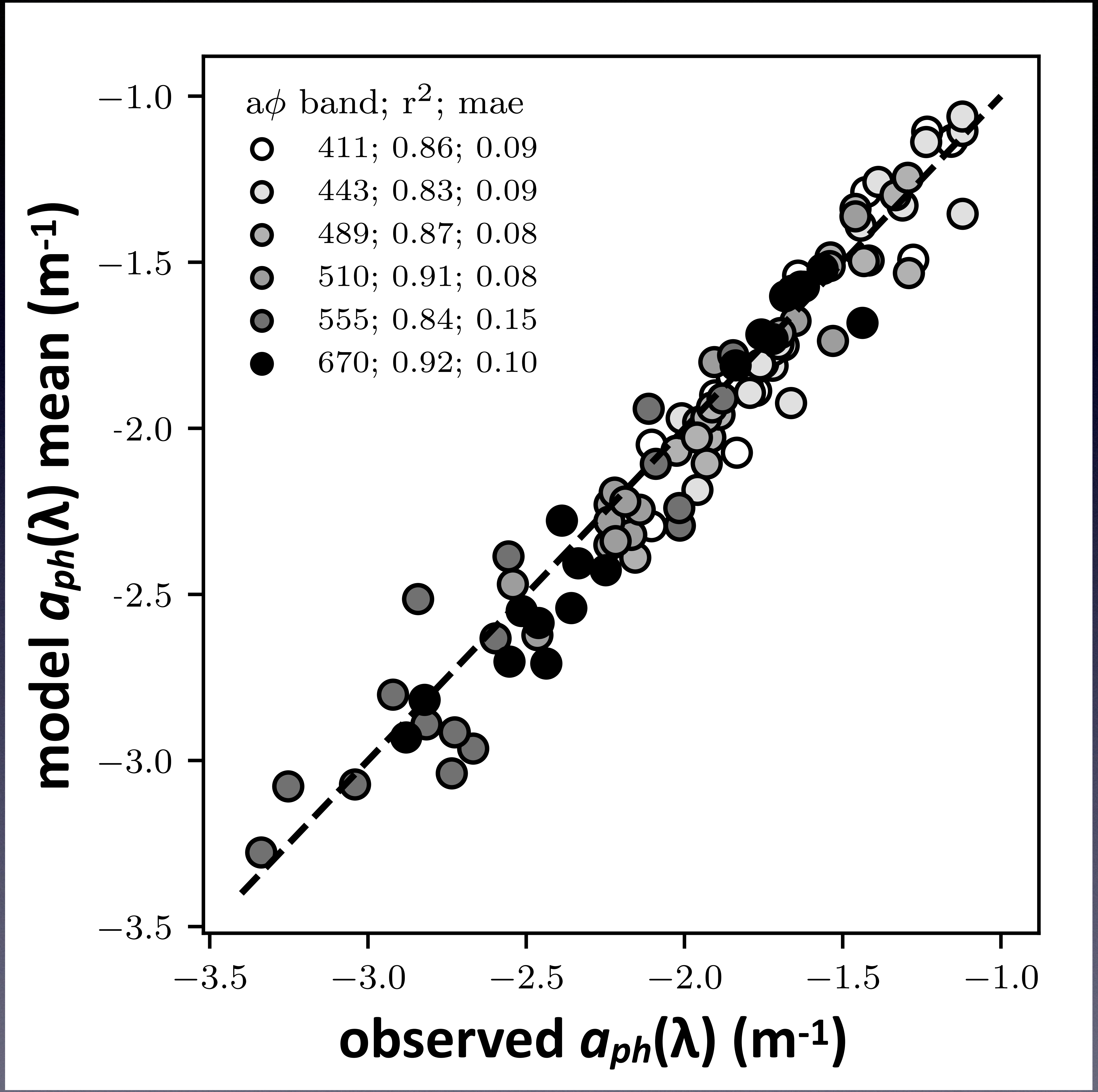
Bayesian Models for Deriving Biogeochemical Information from Satellite Ocean Color

Susanne E. Craig[1,2] and Erdem M. Karaköylü[1,3]

[1]Ocean Ecology Laboratory, NASA Goddard Space Flight Center; [2]Universities Space Research Association; [3]Science Applications International Corporation

Earth ArXiv (2019), https://eartharxiv.org/repository/dashboard/557/10.31223/osf.io/shp6y

Results - Bayesian Neural Network for $a_{ph}(\lambda)$



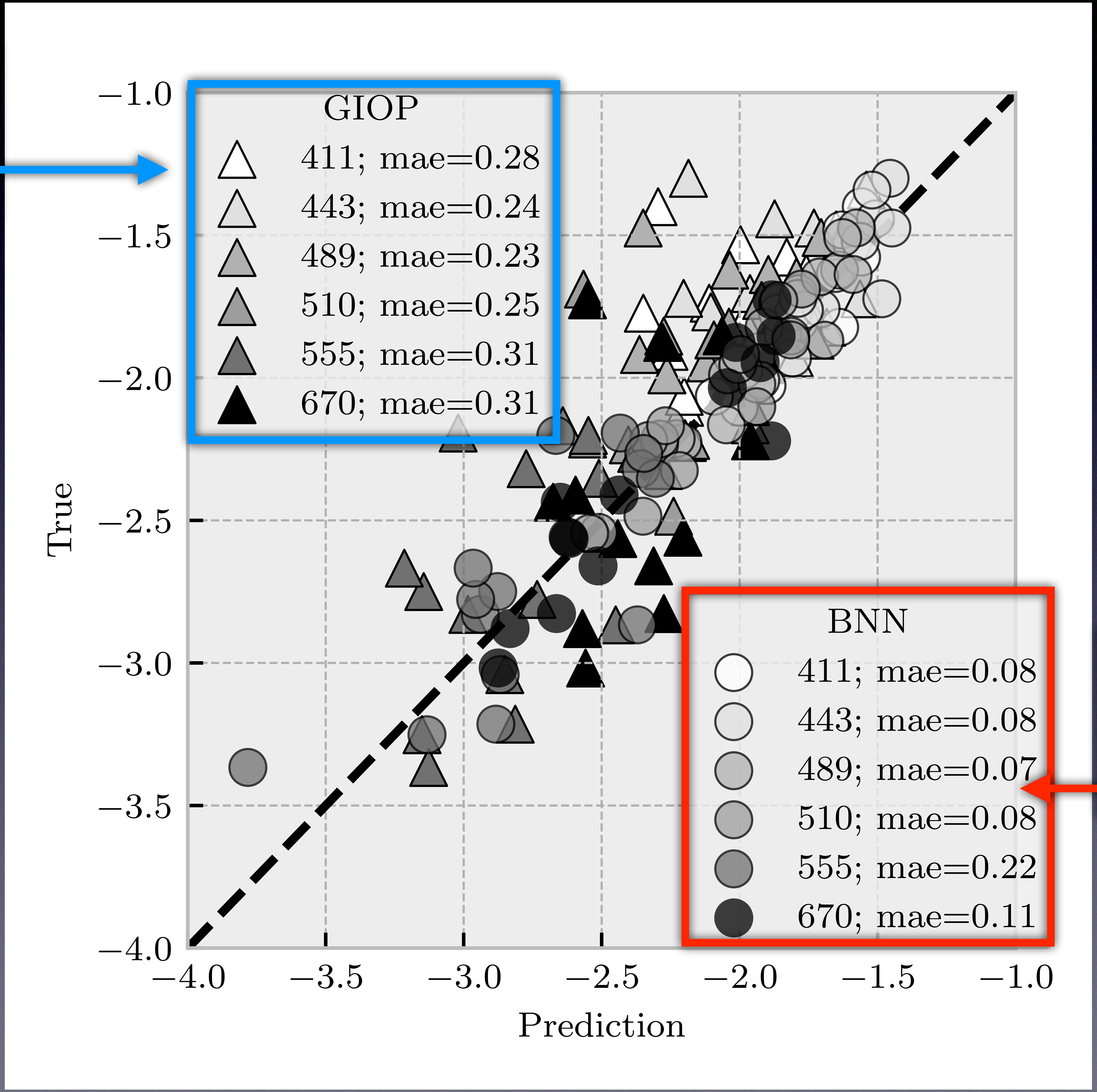
- Out-of-sample observed vs. prediction mean
- $R^2(\lambda) \geq 0.8$
- Mean absolute error(λ) = 0.08-0.15
- Accurate estimate at blue wavelengths
  - ✳ In conventional approaches, atmospheric correction is the most challenging in the blue

N.B. These are means of a predicted distribution

Comparison with a Conventional Semi-Analytical Model

▲ Semi-analytical model

- BNN trained with a test dataset
- Same test data used in GIOP model
- BNN performs significantly better - especially at blue wavelengths



● Bayesian neural net

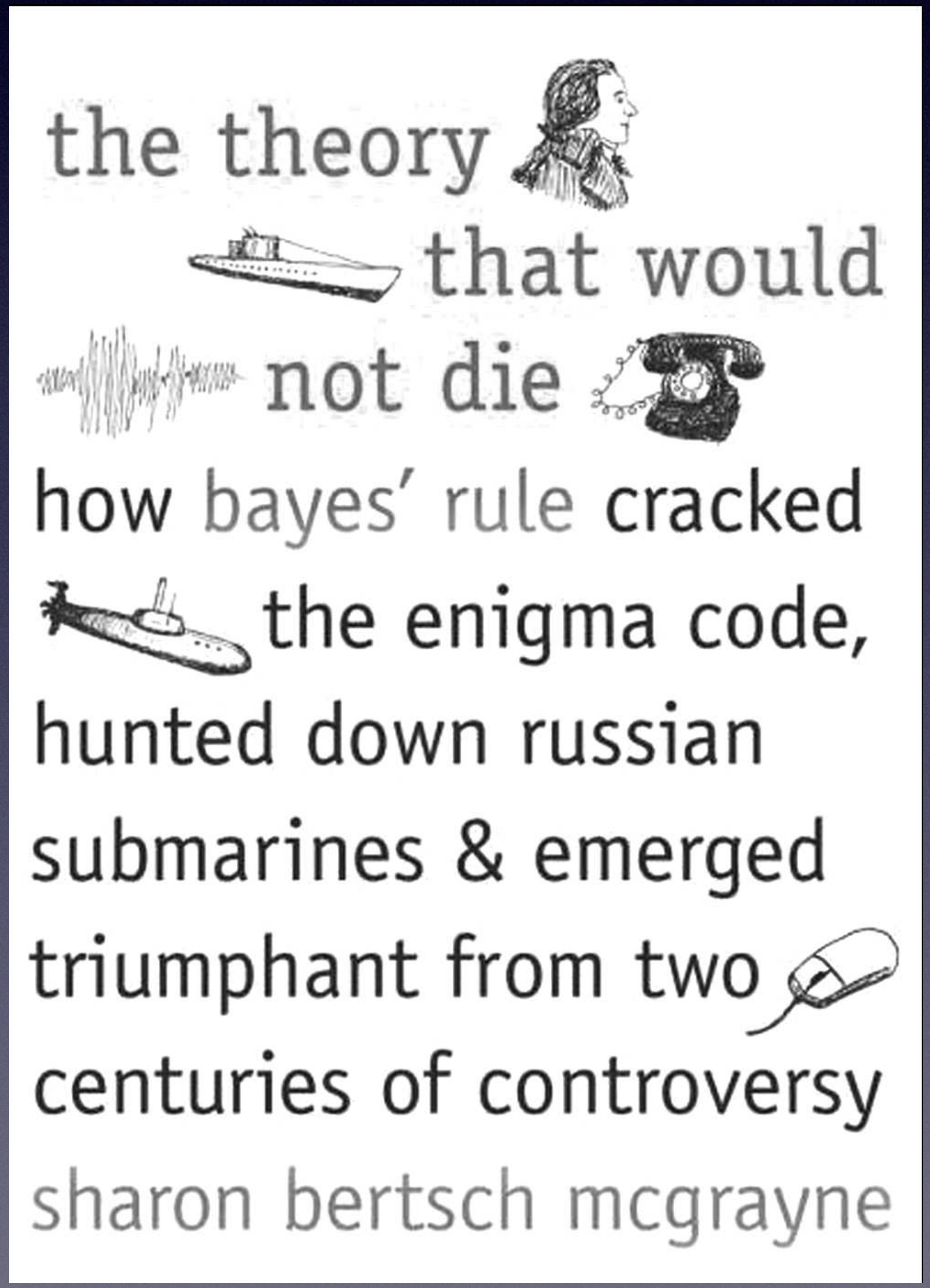
GIOP model - default configuration

(Werdell et al. (2013). Applied Optics, 52, 2019)

Take Home Messages

- Bayesian neural network models were able to accurately estimate optical biomass proxy (and Chl - not shown) from top of atmosphere spectra
  - ***<u>Standard atmospheric correction was not required</u>***
  - ***<u>Particularly relevant for coastal, inland and optically complex waters</u>***
- Bayesian neural networks:
  - Provide robust estimates of uncertainty
  - Resistant to overfitting
  - Improve as more labeled data is acquired

More Bayes…
(The Theory that Would Not Die)



Machine Learning Approaches for Predicting Phytoplankton Community Composition (PCC) from Ocean Color

PI: Susanne E. Craig[1]

Team: Ian Carroll[1], Jason Graff[2], Susanne Menden-Deuer[3], Erdem Karaköylü[4], Cecile Rousseaux[5]

[1]Ocean Ecology Laboratory, Goddard Space Flight Center/UMBC, [2]Oregon State University, [3]University of Rhode Island GSO, [4]Machine Learning Consultant, [5]Global Modeling & Assimilation Office, Goddard Space Flight Center

Problem Statement

<u>Context</u>

Phytoplankton community composition (PCC) models have historically relied on deriving empirical relationships between HPLC pigments & ocean color

Pigments are an imperfect proxy of PCC

Co-occur across very different taxonomic groupings/sizes

Environmental + physiological factors
→ Relative abundances are highly dynamic

➡ Climate forcing → restructuring of water column
➡ perturb physiology → pigments
➡ ∴ relationships no longer valid

Problem Statement

Context

Phytoplankton community composition (PCC) models have historically relied on deriving empirical relationships between HPLC pigments & ocean color

Pigments are an imperfect proxy of PCC

But it was all we had for several decades!

Co-occur across very different taxonomic groupings/sizes

Environmental + physiological factors
➝ Relative abundances are highly dynamic

- ➡ Climate forcing ➝ restructuring of water column
- ➡ perturb physiology ➝ pigments
- ➡ ∴ relationships no longer valid

Proposed Solution

Machine learning + less ambiguous metrics of PCC

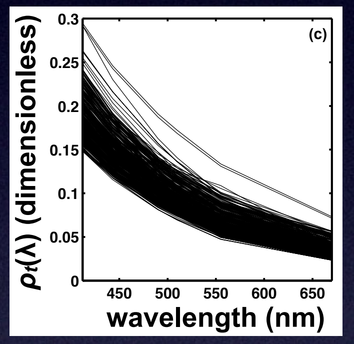
Ingredients

PCC metrics

- Imaging flow cytotbot
- Flow cytometry
- Microscopic enumeration & ID
- Flocam
- Molecular information

\+

Top of atmosphere reflectance

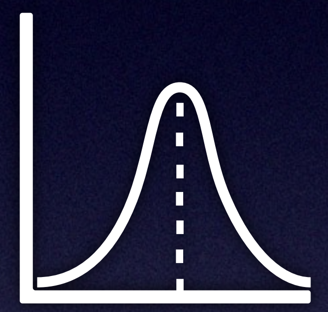


\+

Contextual environmental observations

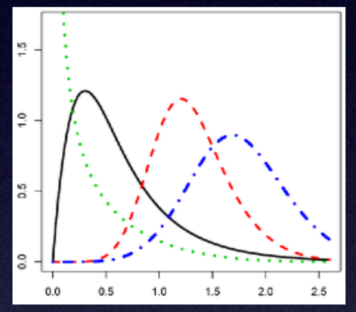
\+

Associated measurement uncertainties



\+

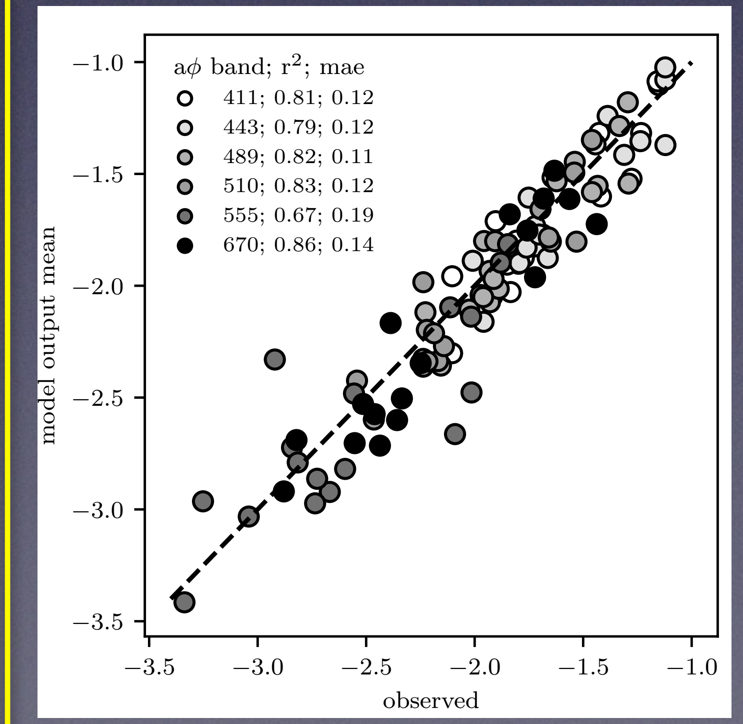
Known/unknown unknowns



Bayesian Models for Deriving Biogeochemical Information from Satellite Ocean Color

Susanne E. Craig[1,2] and Erdem M. Karaköylü[1,3]

[1]Ocean Ecology Laboratory, NASA Goddard Space Flight Center; [2]Universities Space Research Association; [3]Science Applications International Corporation



Circumvents the need for perfect atmospheric correction
➞ Coastal, optically complex, inland

Craig & Karaköylü, 2019, Earth ArXiv, https://doi.org/10.31223/osf.io/shp6y

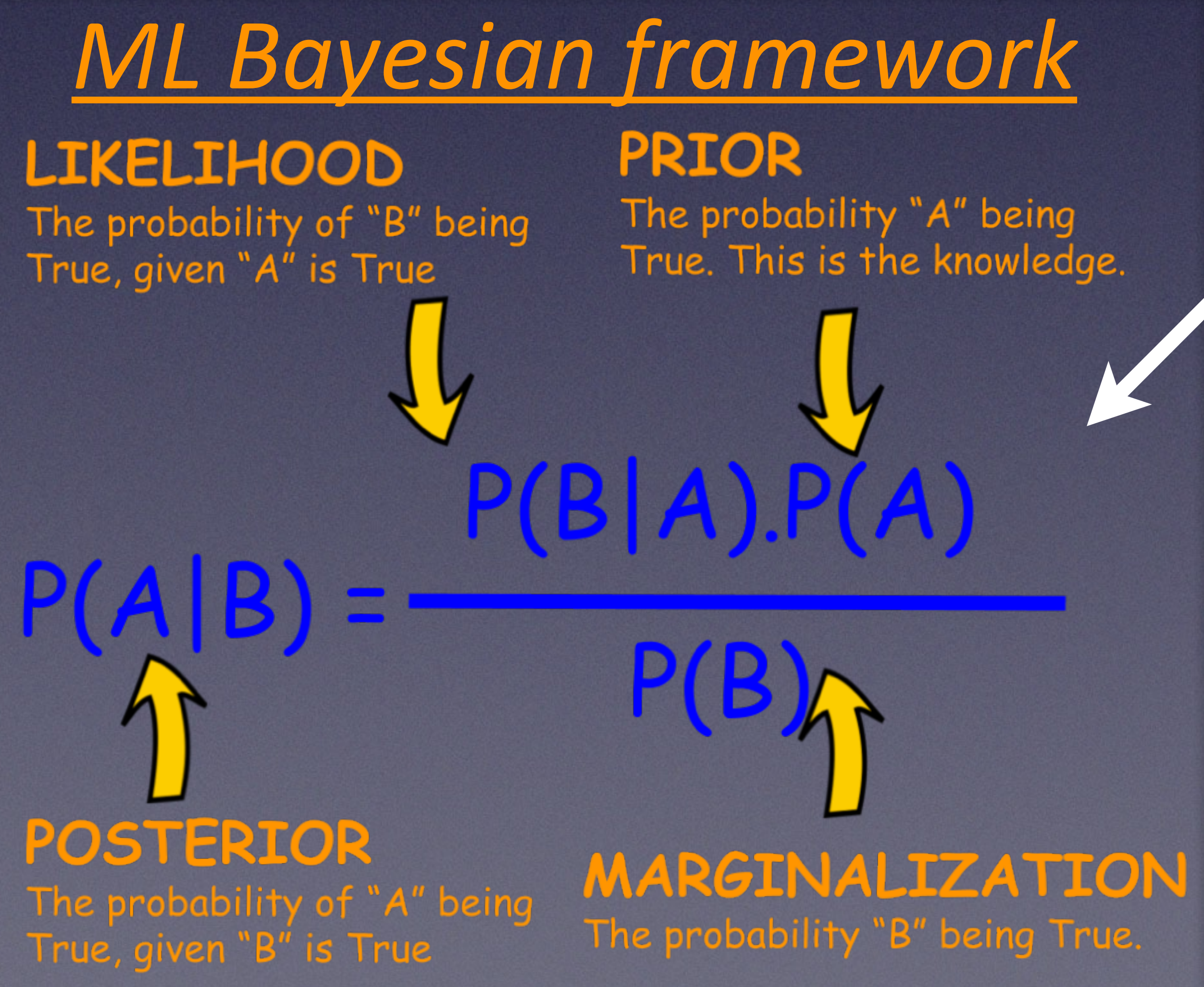


Figure: https://towardsdatascience.com/

- Explore data using ML approaches in a Bayesian framework
- Mitigates:
  - ➡ Overfitting
  - ➡ Data sparsity
  - ➡ Provides robust estimates of uncertainty

Where we are now…

Big Question????
What metric of PCC should be inferred?

Initial ML approach:
Exploratory data analysis of synthetic dataset
- MERRA-NOBM (NASA Ocean Biogeochemical Model)*
- Coupled ocean-atmosphere radiative transfer model
- 6 phytoplankton groups

Greg & Rousseaux, 2017, Front. Mar. Sci., http://dx.doi.org/10.3389/fmars.2017.00060

Can be guided by what labelled data will realistically be available for model training

Approach:
- Detect optical signatures ✓
- Codify ecological principles in Bayesian model(s)
- Conduct sensitivity analysis
  ➡ Reduce predictor variables to those that are available for training


PACE
Plankton, Aerosol, Cloud, ocean Ecosystem

PACE will support studies of:

- ocean biology, ecology, & biogeochemistry
- atmospheric aerosols
- clouds
- land

Primary hyperspectral radiometer:

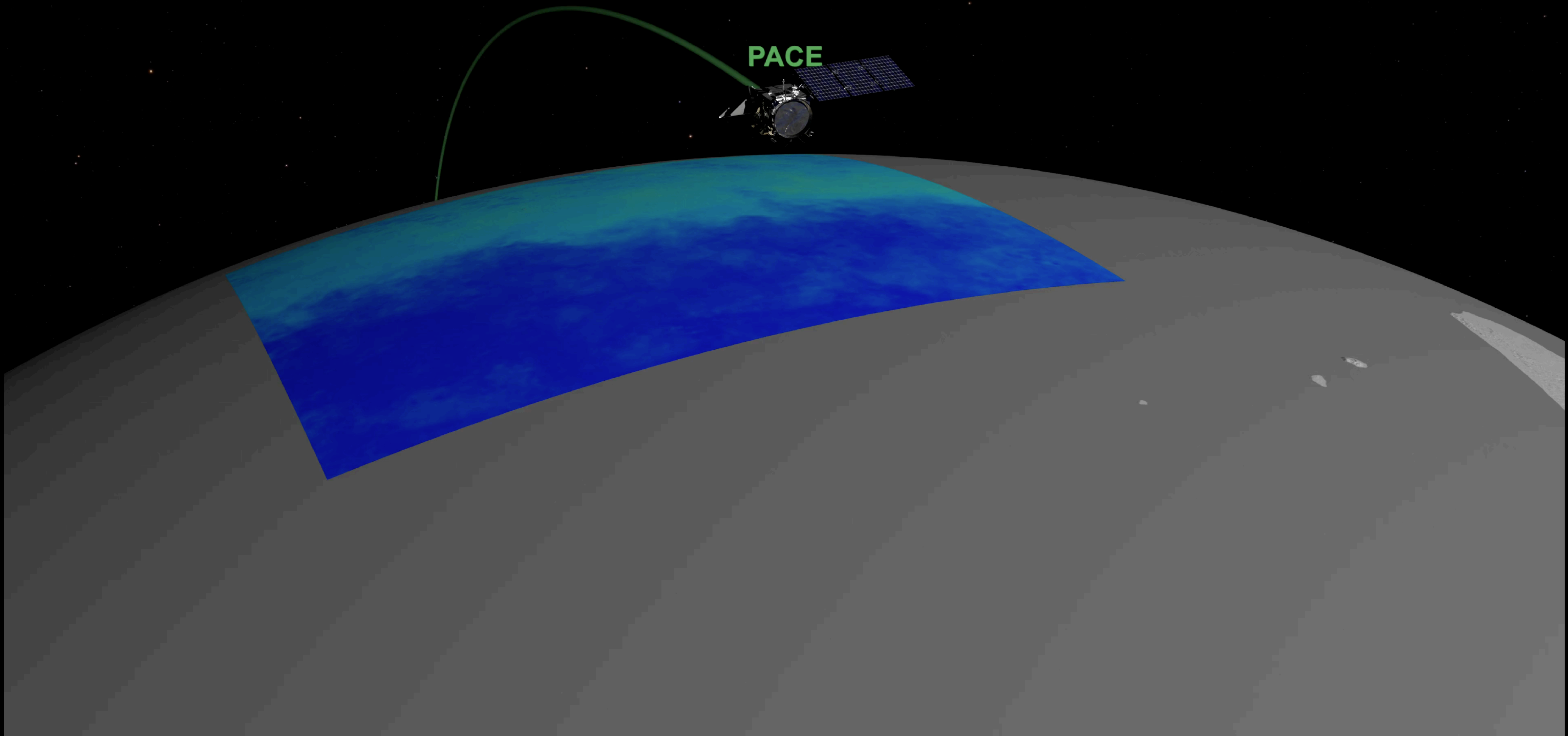
- Ocean Color Instrument (OCI) (GSFC)

2 contributed multi-angle polarimeters:

- HARP2 (UMBC)
- SPEXone (SRON/Airbus)

Mission elements:

- Competed science teams (ESD)
- Competed SVC teams (ESD)
- Science analysis & processing (GSFC)
- Spacecraft (GSFC)
- Mission operations (GSFC)



Legacies:

- SeaWiFS, MODIS, VIIRS
- POLDER, MISR

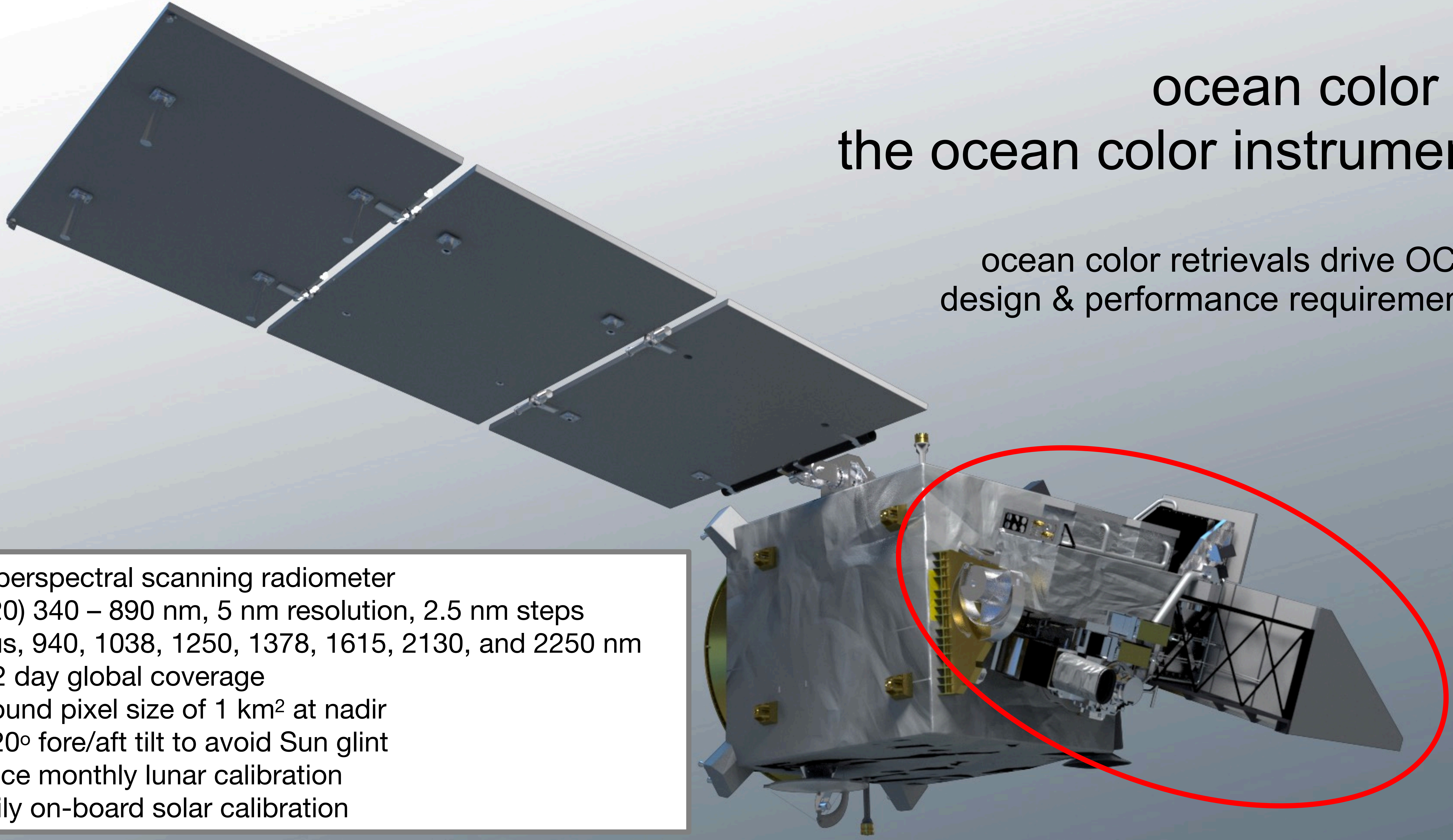
Key characteristics:

- Jan. 2024 launch
- Falcon 9 from KSC/Cape Canaveral
- 676.5 km altitude
- polar, ascending, Sun synchronous orbit; 98° inclination
- 13:00 local Equatorial crossing
- 3-yr design life; 10-yr propellant

Slides courtesy of Jeremy Werdell, PACE Project Scientist NASA GSFC

ocean color & the ocean color instrument

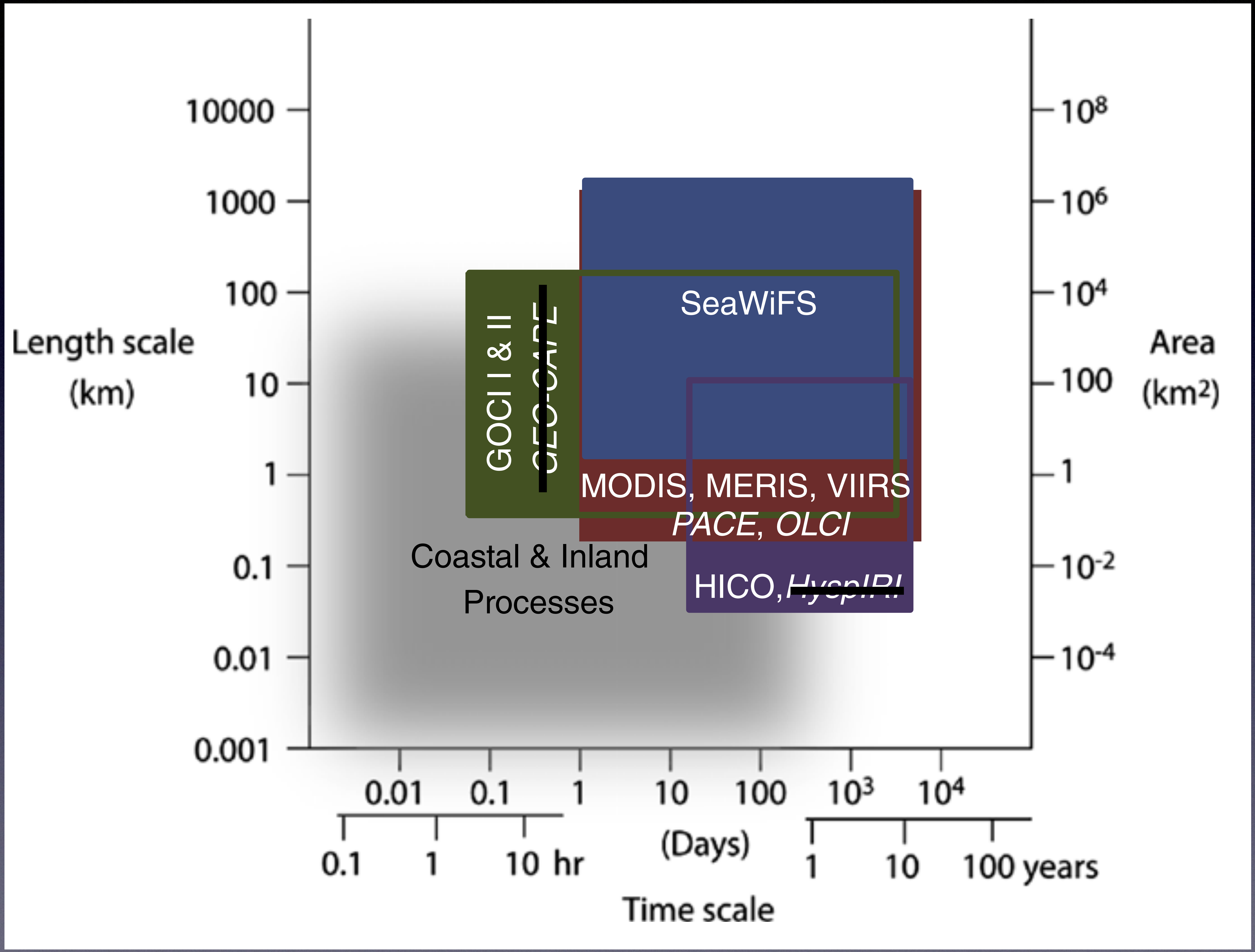
ocean color retrievals drive OCI's design & performance requirements



- hyperspectral scanning radiometer
- (320) 340 – 890 nm, 5 nm resolution, 2.5 nm steps
- plus, 940, 1038, 1250, 1378, 1615, 2130, and 2250 nm
- 1-2 day global coverage
- ground pixel size of 1 km^2 at nadir
- ± 20^o fore/aft tilt to avoid Sun glint
- twice monthly lunar calibration
- daily on-board solar calibration

Slides courtesy of Jeremy Werdell, PACE Project Scientist NASA GSFC

PACE Measurement Scales



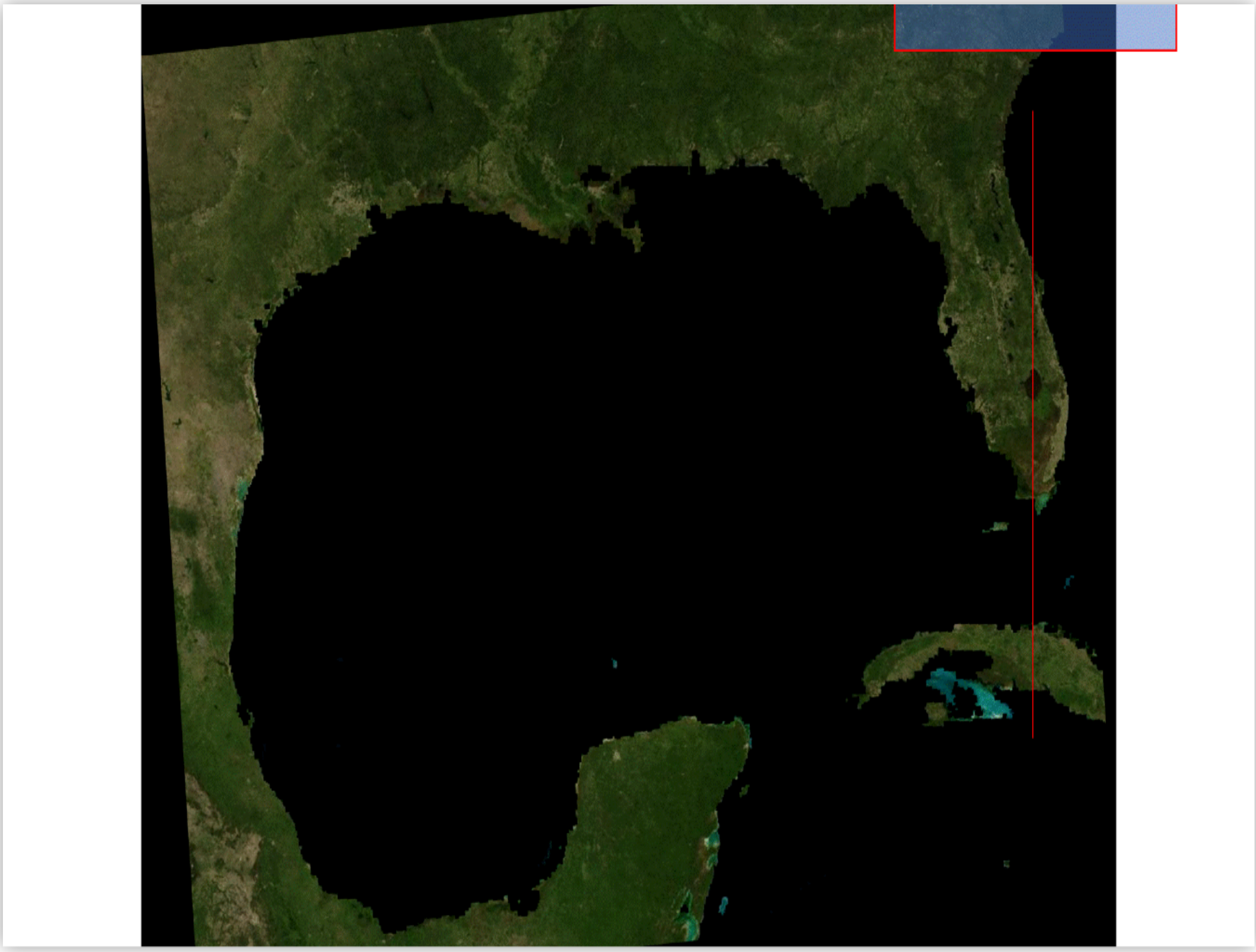
Mouw et al. (2015). Remote Sensing of Environment 160, 15-30, https://doi.org/10.1016/j.rse.2015.02.001

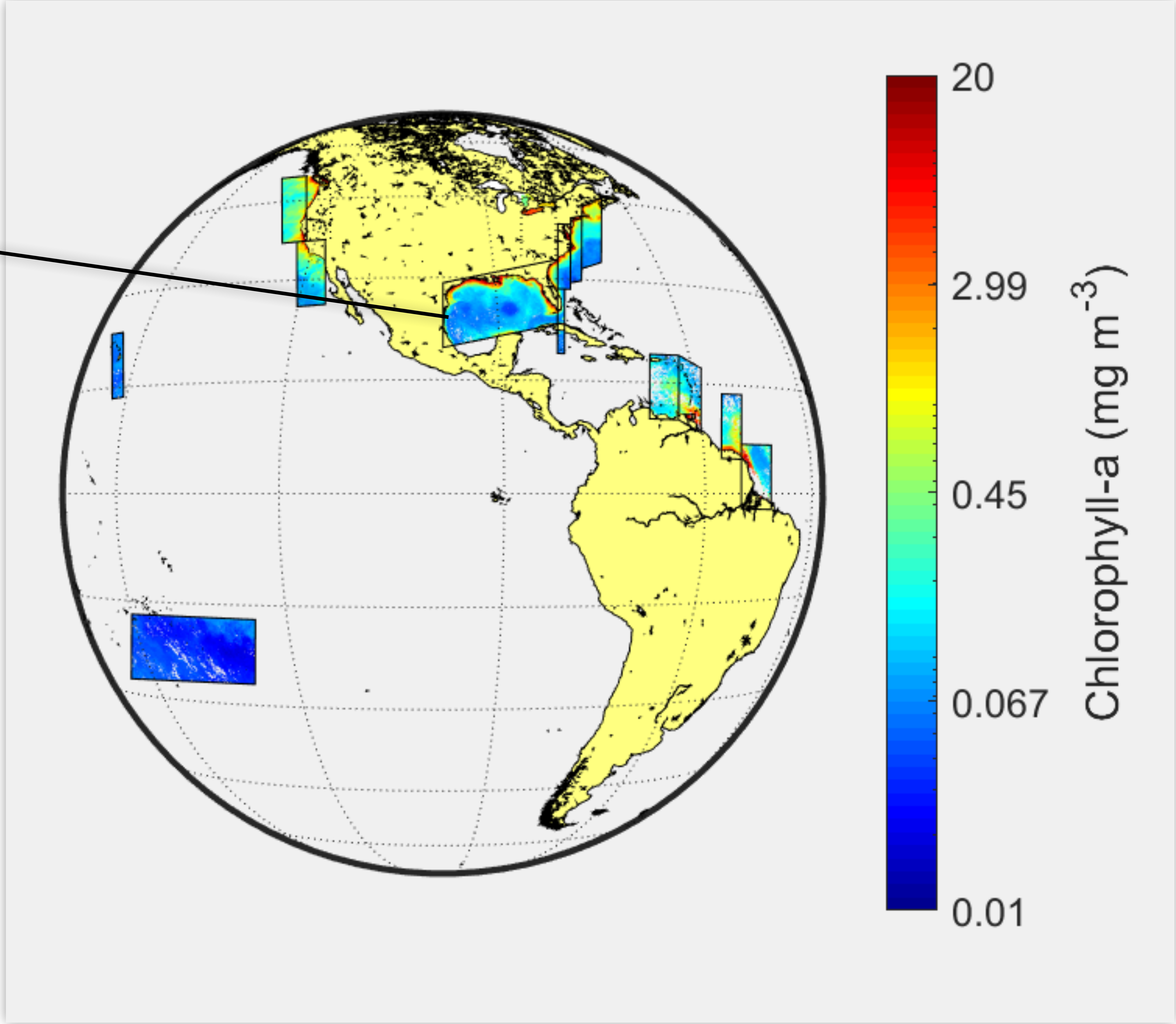
Geosynchronous Littoral Imaging and Monitoring Radiometer (GLIMR)

- Funded under NASA's Earth Venture Instrument (EVI) portfolio
- PI: Joe Salisbury, UNH
- Deputy PI: Antonio Mannino, NASA GSFC
- Anticipated launch date 2026/27
- Geostationary orbit

NASA | University of New Hampshire | NASA Goddard Space Flight Center | Raytheon Technologies | SwRI Southwest Research Institute | Kamel Engineering

GLIMR





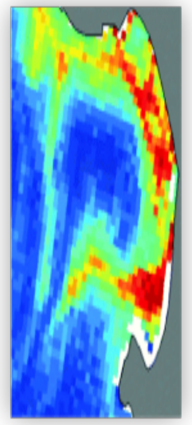
Primary Science Scans

- 6x/day Gulf of Mexico (GoMex)
- 2x/day US East Coast
- 2x/day US West Coast
- 2x/day Amazon River Plume ROI
- 2x/day Caribbean Sea ROI
- 3x/day other HAB target sites
- Calibration Sites (MOBY/S. Pacific/PACE)

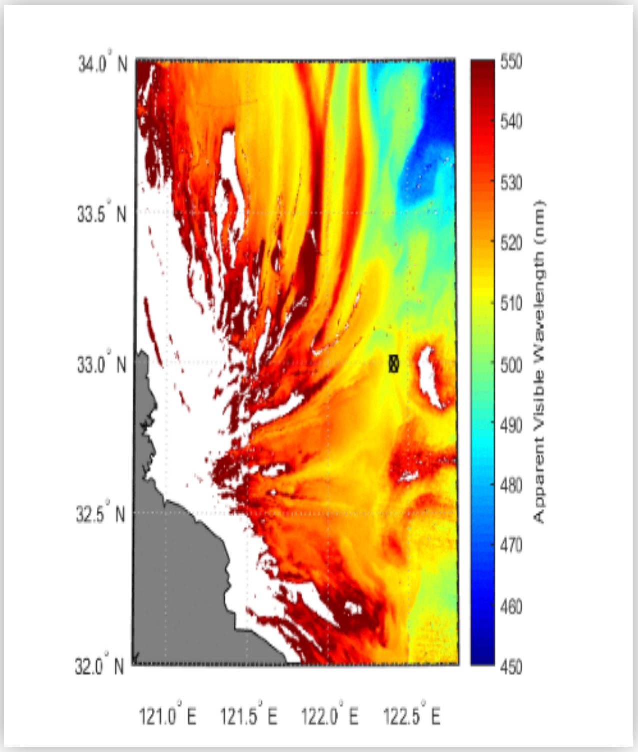
Slides courtesy of Ryan Vandermuelen, GLIMR Investigator, NASA GSFC

A glimpse into GLIMR

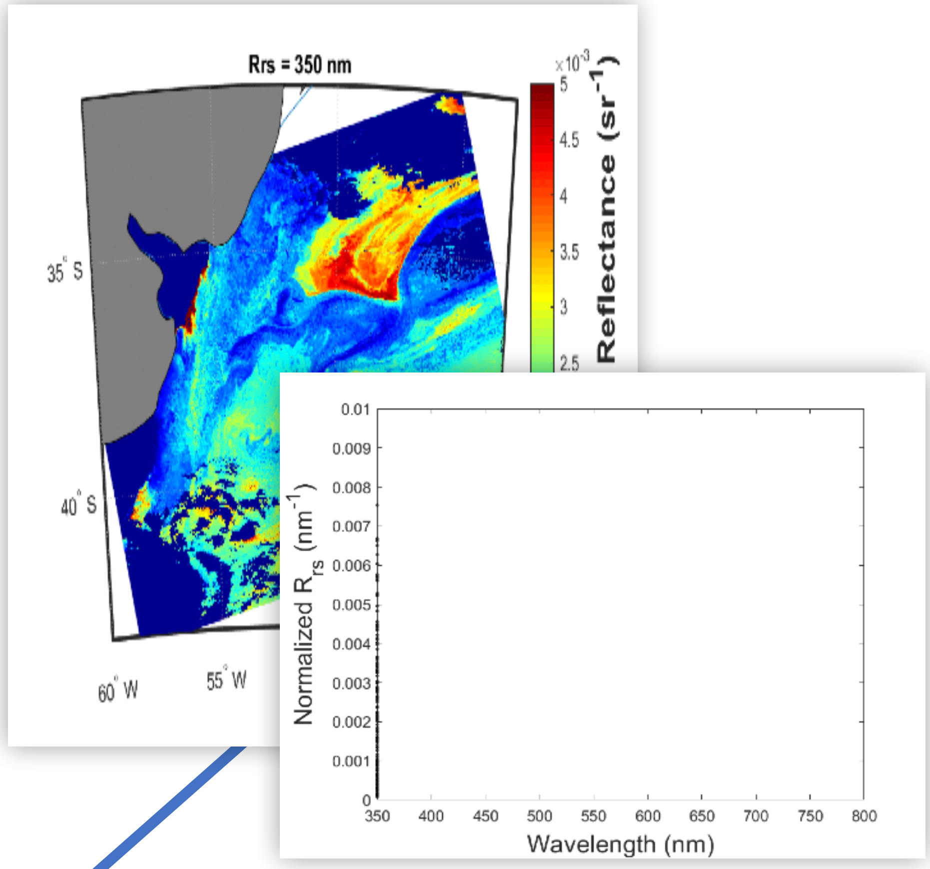
Spatial



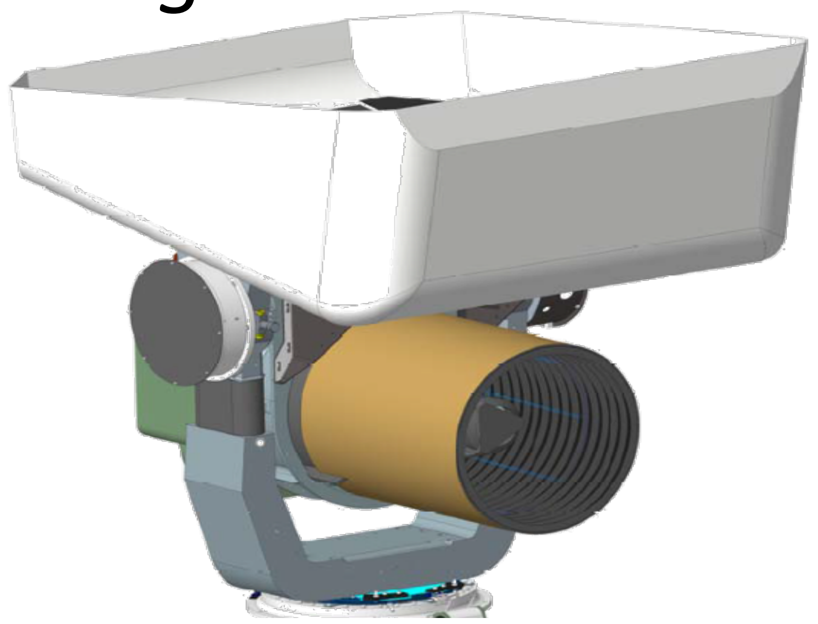
Temporal

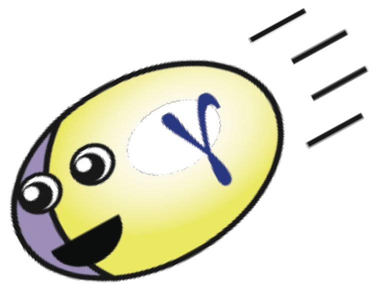


Spectral



Signal to Noise





Telescope mounted on a 2-axis gimbal that actively scans an imaging spectrometer across the Gulf of Mexico.

Geostationary **L**ittoral **I**maging and **M**onitoring **R**adiometer

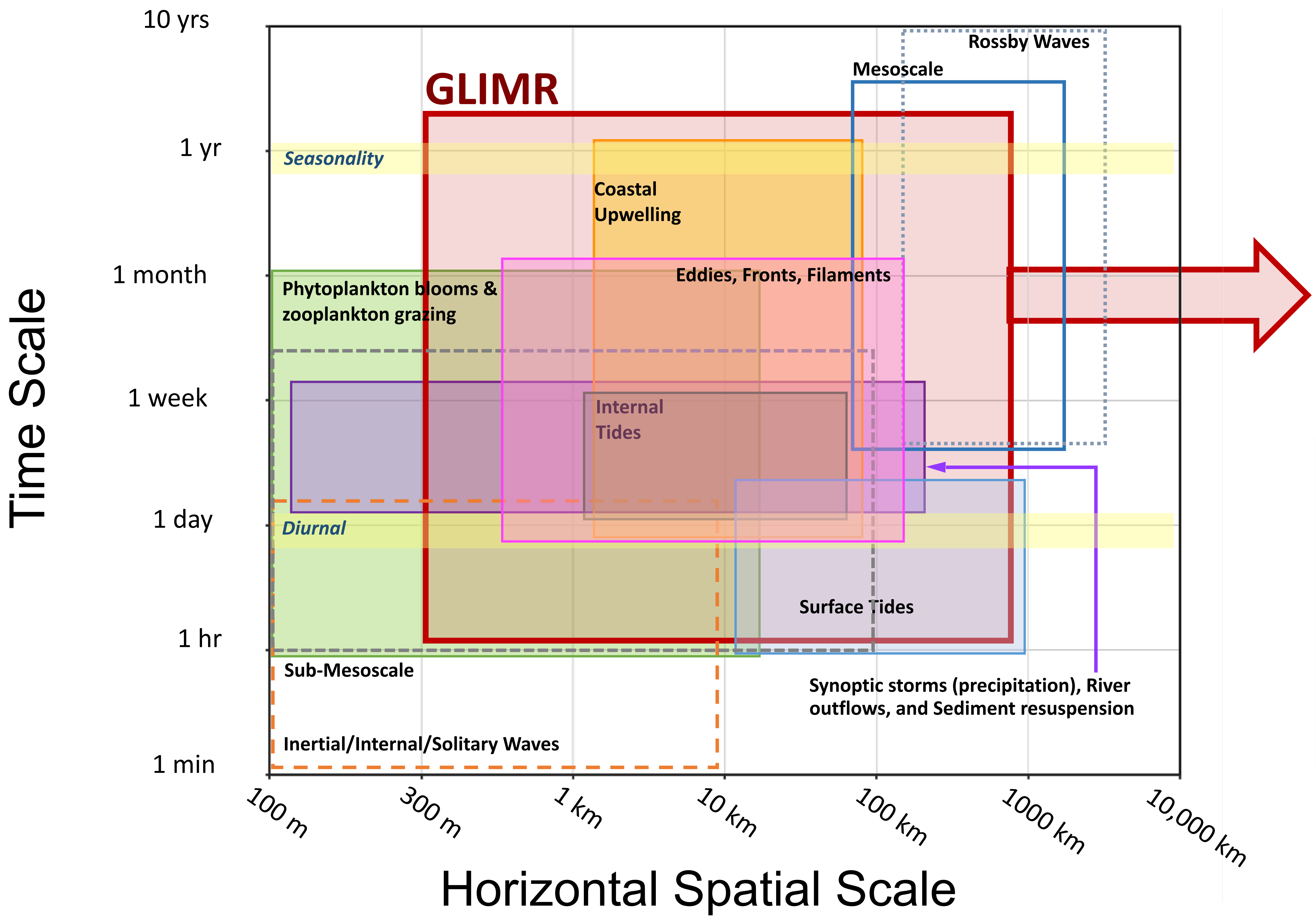
Hyperspectral
- 340-1040 nm
- <10 nm UV-Vis resolution
- <5 nm UV-Vis sampling

High Temporal
- ~hourly scans of Gulf of Mexico (6x/day)
- 2x/day other regions
- 3x/day HAB target sites

High Spatial
- 300 m GSD nadir
- ~328 m Gulf of Mexico
- <500 m over coastal CONUS

High SNR
- > 420, UV
- > 1000, 400-580 nm
- > 750, 580-650 nm
- > 580, 650-890 nm

Slides courtesy of Ryan Vandermuelen, GLIMR Investigator, NASA GSFC

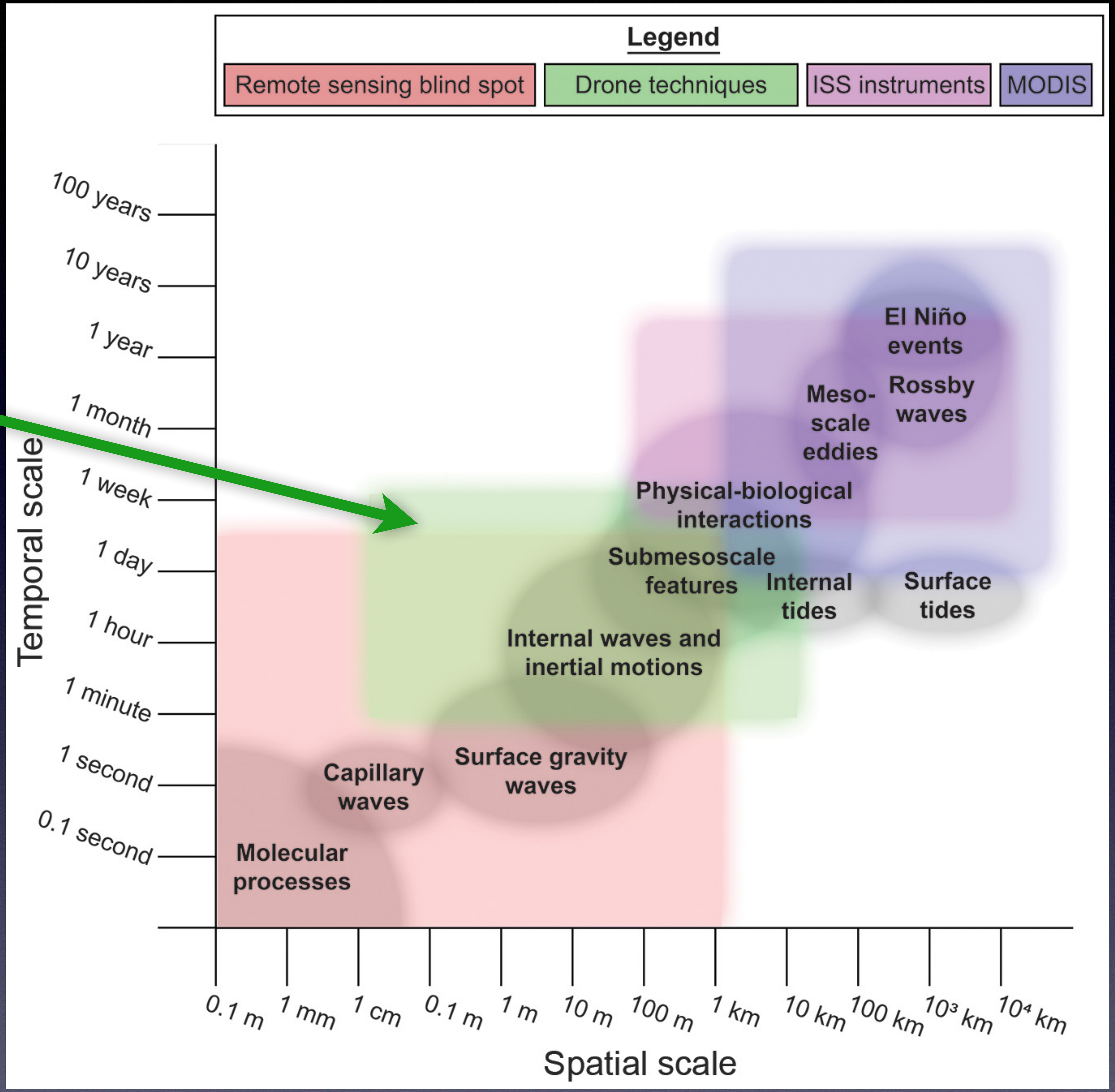


The temporal cadence of GLIMR will enable the observation of physical processes that regulate the spatial-temporal dynamics of biological and biogeochemical processes and constituent distributions.

Slides courtesy of Ryan Vandermuelen, GLIMR Investigator, NASA GSFC

Suborbital Ocean Color Platforms

- UAVs
- Moorings
- Ships of opportunity
- Gliders
- Wavegliders
- SailDrones…



Gray et al. (2022). Frontiers in Ecology and the Environment, https://doi.org/10.1002/fee.2472

- These all address sub-pixel variability and fill satellite spatial/temporal gaps
- Autonomous measurements will become increasingly important for ocean observations & model validation

Q: How to promote data sharing and the creation of integrated archives with consistent data quality and format requirements?

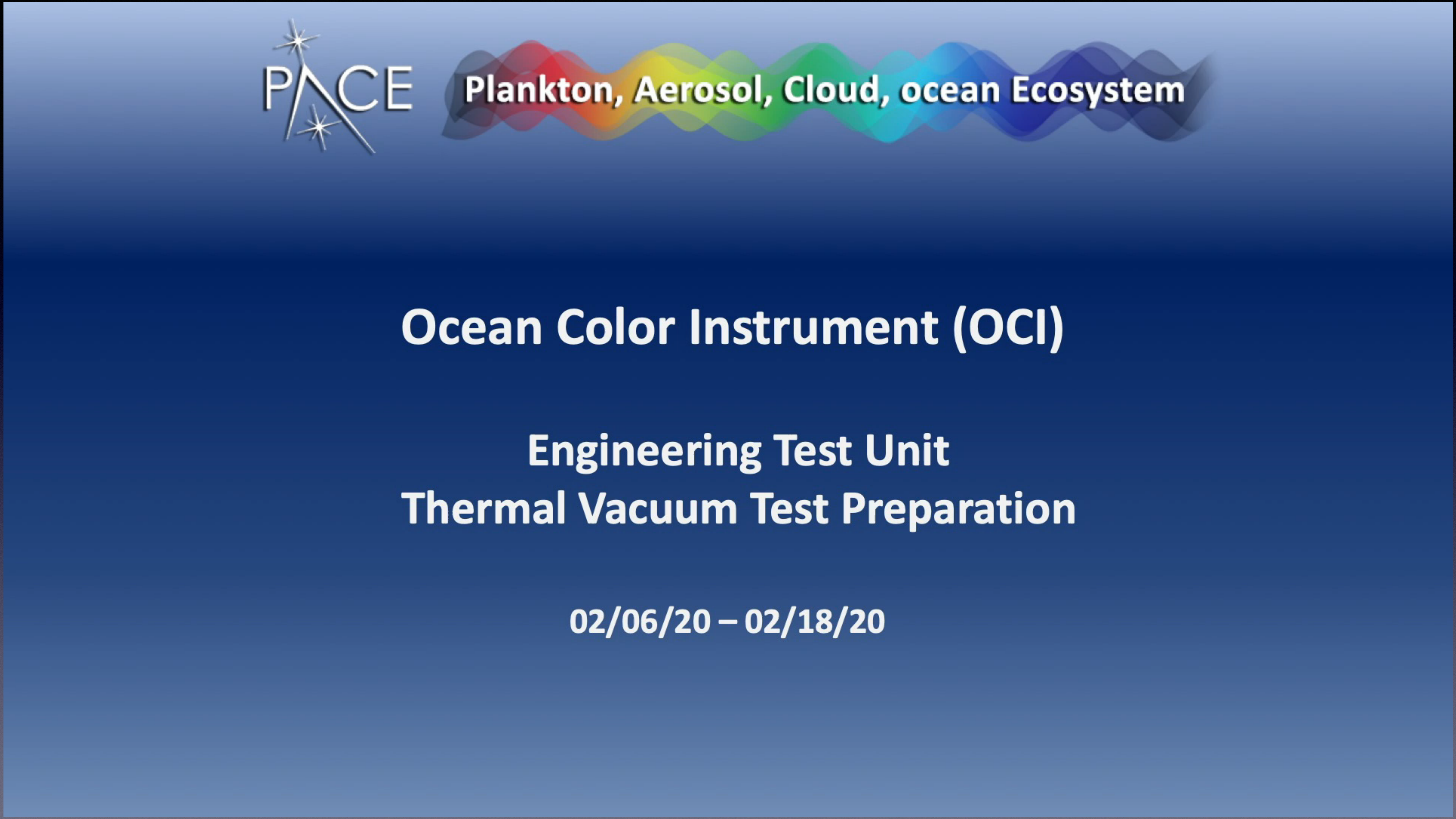
- All NASA-funded investigators are obliged to submit their data (ocean color and oceanographic) to the SeaBASS repository (https://seabass.gsfc.nasa.gov/)
- The community should strive to adhere to FAIR data principles:
  - Findable, Accesible, Interoperable, Reusable (https://www.go-fair.org/)
  - However, this requires funded support to achieve!
- NASA is strongly encouraging a push to open science:
  - Transform to Open Science (TOPS https://science.nasa.gov/open-science/transform-to-open-science)

Q: How to promote data sharing and the creation of integrated archives with consistent data quality and format requirements?

- The PACE mission has a dedicated Applications Team
  - Erin Urquhart, Natasha Sadoff
  - Early adopters program to prime future community for using PACE data products
  - Community of Practice
  - https://pace.oceansciences.org/applications.htm

Erin Urquhart, Project Applications Coordinator
erin.u.jephson@nasa.gov

Natasha Sadoff: Project Applications Deputy Coordinator
natasha.sadoff@nasa.gov



Movie courtesy of Jeremy Werdell, PACE Project Scientist NASA GSFC