Deep learning for Environmental and Ecological Prediction, eValuation and Insight with Ensembles of Water quality

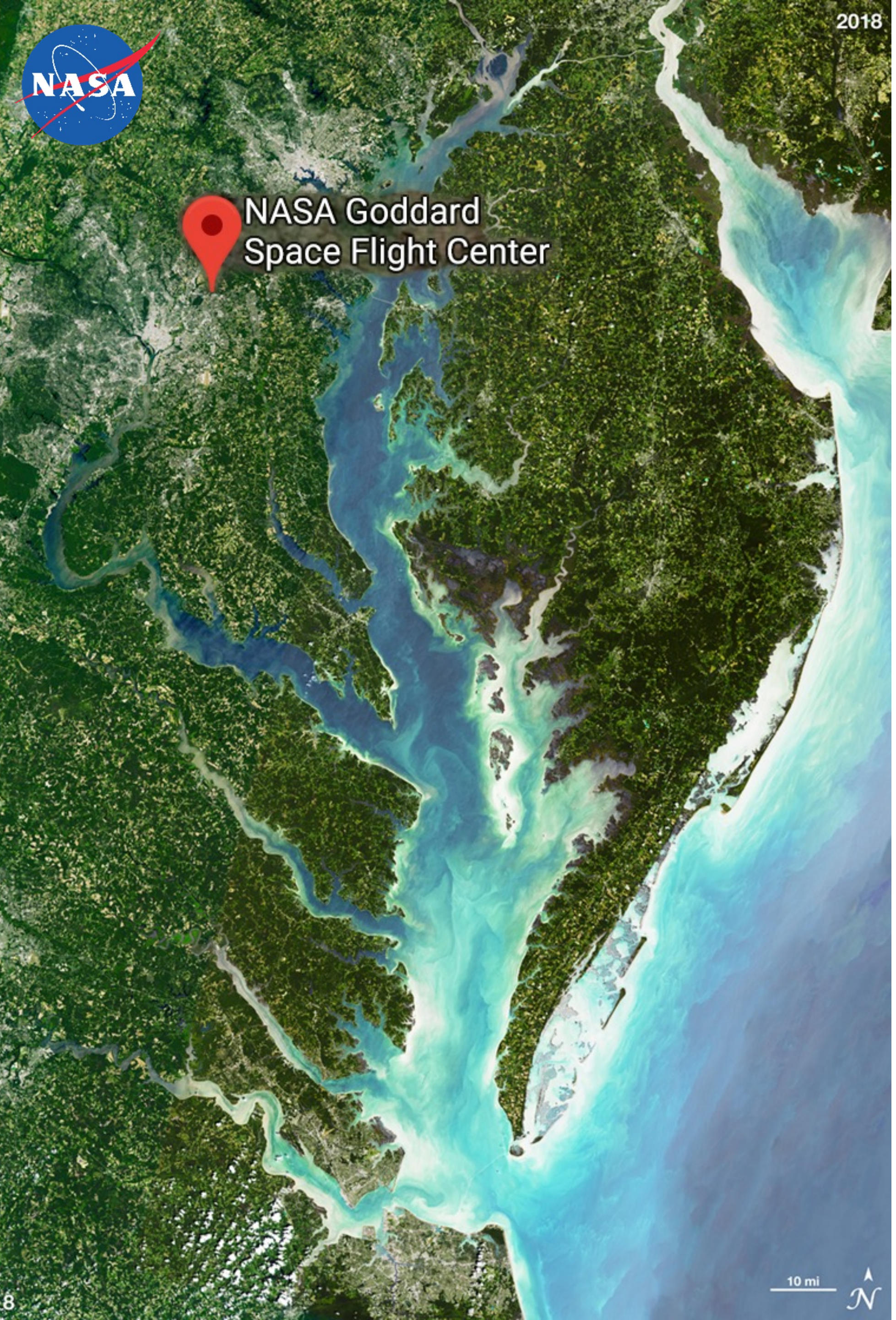
(DEEP-VIEW)

PI: Stephanie Schollaert Uz[1],

Troy Ames[1], J. Blake Clark[1,2], Dirk Aurin[1,3]

[1]NASA Goddard Space Flight Center

[2]University of Maryland Baltimore County, [3]Morgan State University



GODDARD EARTH SCIENCES

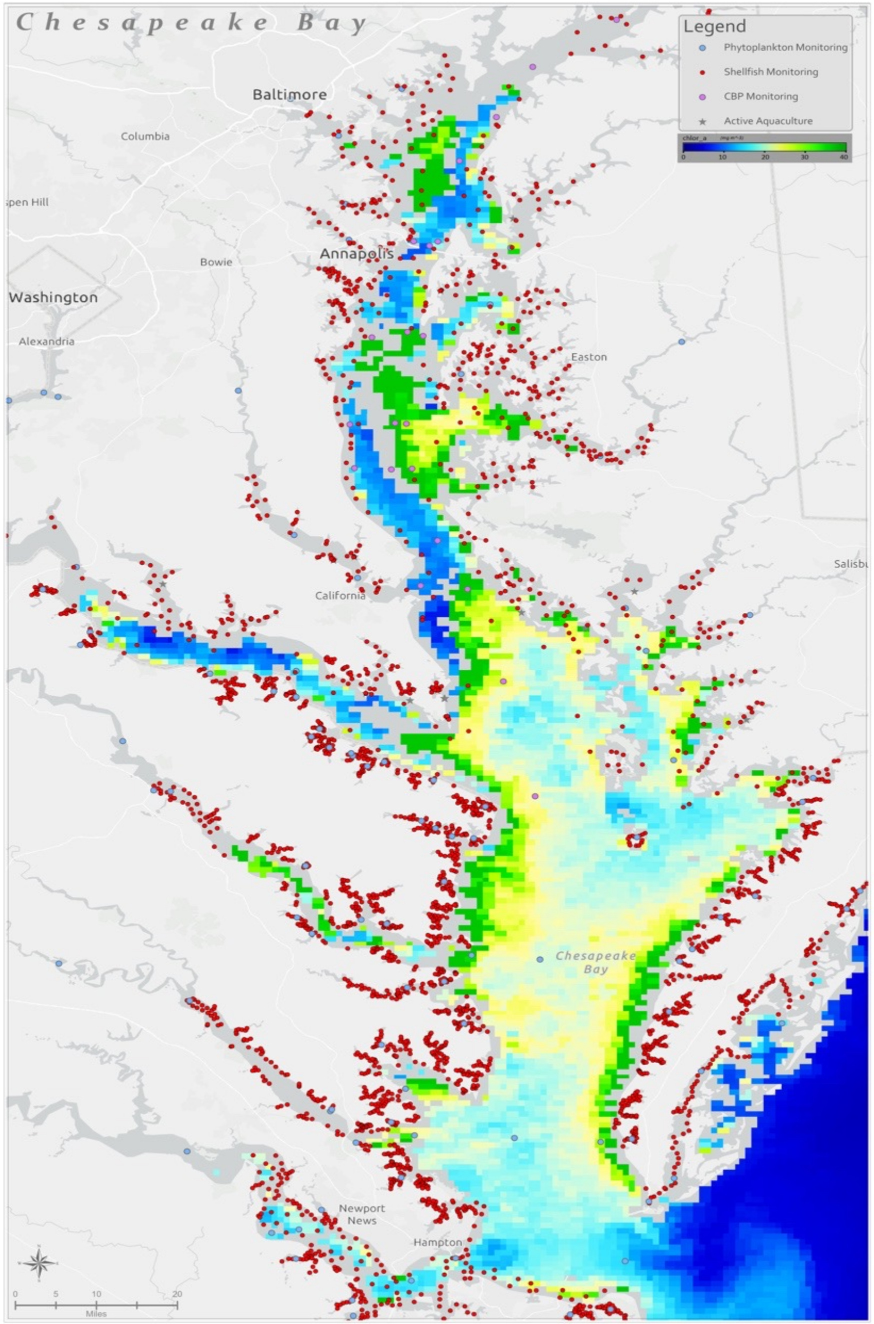
How can we improve coastal water quality data products to address societal challenges locally?

- Challenge – large estuary, increased land use and runoff, increased aquaculture and recreation while water quality management resources shrinking, maximize potential of upcoming NASA hyperspectral missions

- Interagency Chesapeake Bay working group activities increase coordination, collaboration on data collection, calibration activities, new research

- DEEP-VIEW machine learning development to fuse features detected through multiple sources, including hyperspectral satellites, to capture greater variability

Motivation: resource manager need



MODIS chlorophyll-a map from July 2, 2019 with routine sampling sites by Maryland and Virginia superimposed

Parameter name	Water Quality Threshold
Fecal coliform	<14 MPN median per100ml
Bacteriological *Escherichia coli*	< 410 count per 100ml
Dissolved oxygen	> 5 mg/l
Temperature	< 90°F/32°C
pH	6.5 - 8.5
Turbidity	<150 nephelometer turbidity units
Color	< 75 platinum cobalt units
Water clarity	> 13% (tidal fresh)

Maryland shellfish harvesting water quality criteria

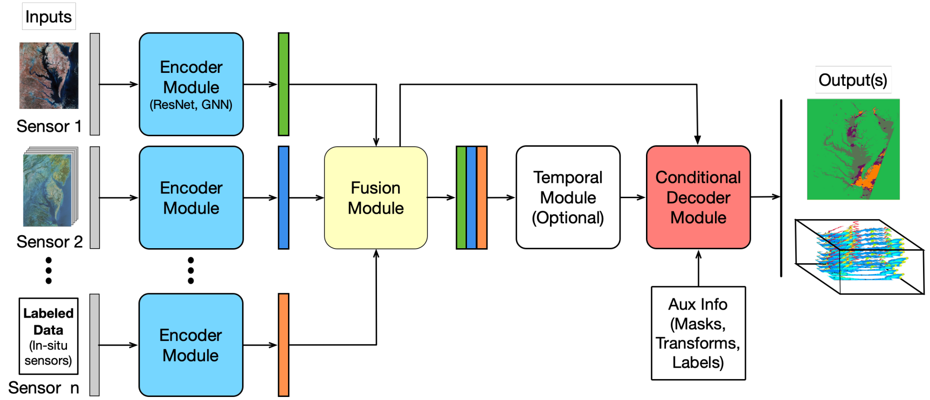
- Can we improve coastal remote sensing to assist resource managers?
- Exploring relationship between satellite data and classification labels: temperature, turbidity, phytoplankton pigments, pollutants

Integration of Observations and Models into Machine Learning for Coastal Water Quality

PI: Stephanie Schollaert Uz, NASA GSFC

Objective

- Continue development and validation of modular framework to integrate data from multiple satellites and models to identify water quality problem areas in the Chesapeake Bay.
- Initial performance goals are >90% accuracy for detection of poor water quality (exceeding thresholds for indicators, e.g. turbidity, harmful algal blooms, pollutants).
- Technology includes feature extraction by machine learning with multiple satellite sensors, physical models, and in situ sampling.
- Improved capability prepares to exploit hyperspectral sensing by future NASA missions, i.e. PACE GLIMR, SBG



Modular framework for detecting water quality features from multi-sensor segmentation using remote sensing and in situ data

Approach

Apply NASA data, science, and technology to support interagency partners (e.g. state agencies, NOAA) in their operations toward the development of a decision support tool for shellfish aquaculture:

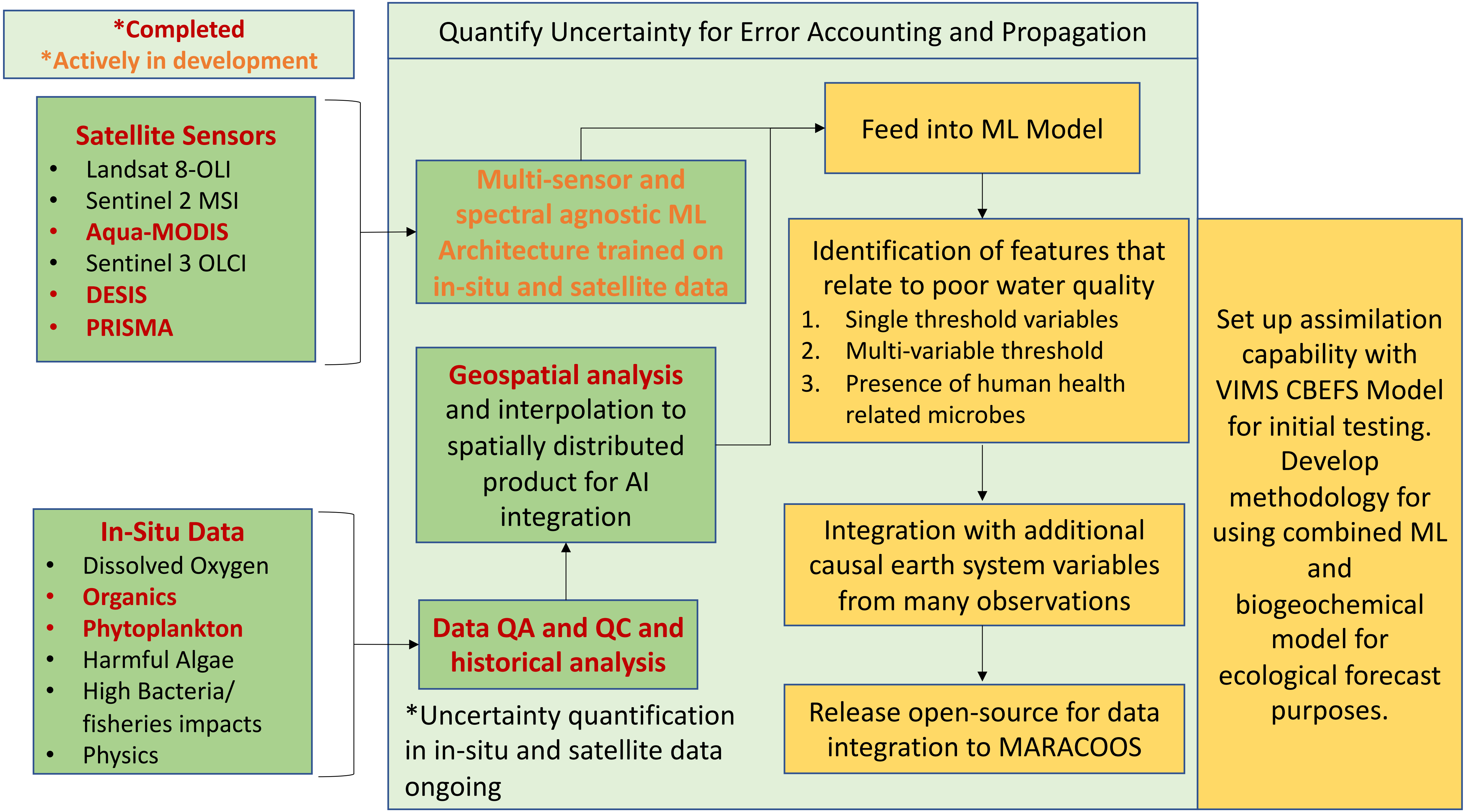
1. Collect and analyze all available in situ and remotely sensed data relevant to Chesapeake Bay water quality.
2. Collect and analyze absorption and fluorescence properties of water constituents at hyperspectral resolution for select sites.
3. Train an ACF ML to identify features that resulted in shellfish bed closures.
4. Refine and validate the ML against current conditions.

Co-Is/Partners: Troy Ames, GSFC; Blake Clark, UMBC/GSFC; Marjorie Friedrichs, VIMS; Chris Brown, NOAA; John McKay, MDE

Key Milestones

- Create interface to data modules (12/22)
- Optimize feature encoders (01/23)
- Develop feature fusion module (03/23)
- Optimize temporal fusion (06/23)
- Machine Learning validation (TRL 5) (06/24)
- Transition ML for Bias Correction (06/25)

TRL_{in} = 3 $TRL_{current}$ = 4


*Completed
*Actively in development
Satellite Sensors
Landsat 8-OLI
Sentinel 2 MSI
Aqua-MODIS
Sentinel 3 OLCI
DESIS
PRISMA
In-Situ Data
Dissolved Oxygen
Organics
Phytoplankton
Harmful Algae
High Bacteria/ fisheries impacts
Physics
Quantify Uncertainty for Error Accounting and Propagation
Multi-sensor and spectral agnostic ML Architecture trained on in-situ and satellite data
Geospatial analysis and interpolation to spatially distributed product for AI integration
Data QA and QC and historical analysis
*Uncertainty quantification in in-situ and satellite data ongoing
Feed into ML Model
Identification of features that relate to poor water quality
1. Single threshold variables
2. Multi-variable threshold
3. Presence of human health related microbes
Integration with additional causal earth system variables from many observations
Release open-source for data integration to MARACOOS
Set up assimilation capability with VIMS CBEFS Model for initial testing. Develop methodology for using combined ML and biogeochemical model for ecological forecast purposes.

AIST18: initially developed ML using satellite data and process model

Using 3-D Virginia Institute of Marine Science (VIMS) model as label data, initial architecture trained on optical satellite data input.

a) Target vs. predicted for every image bin

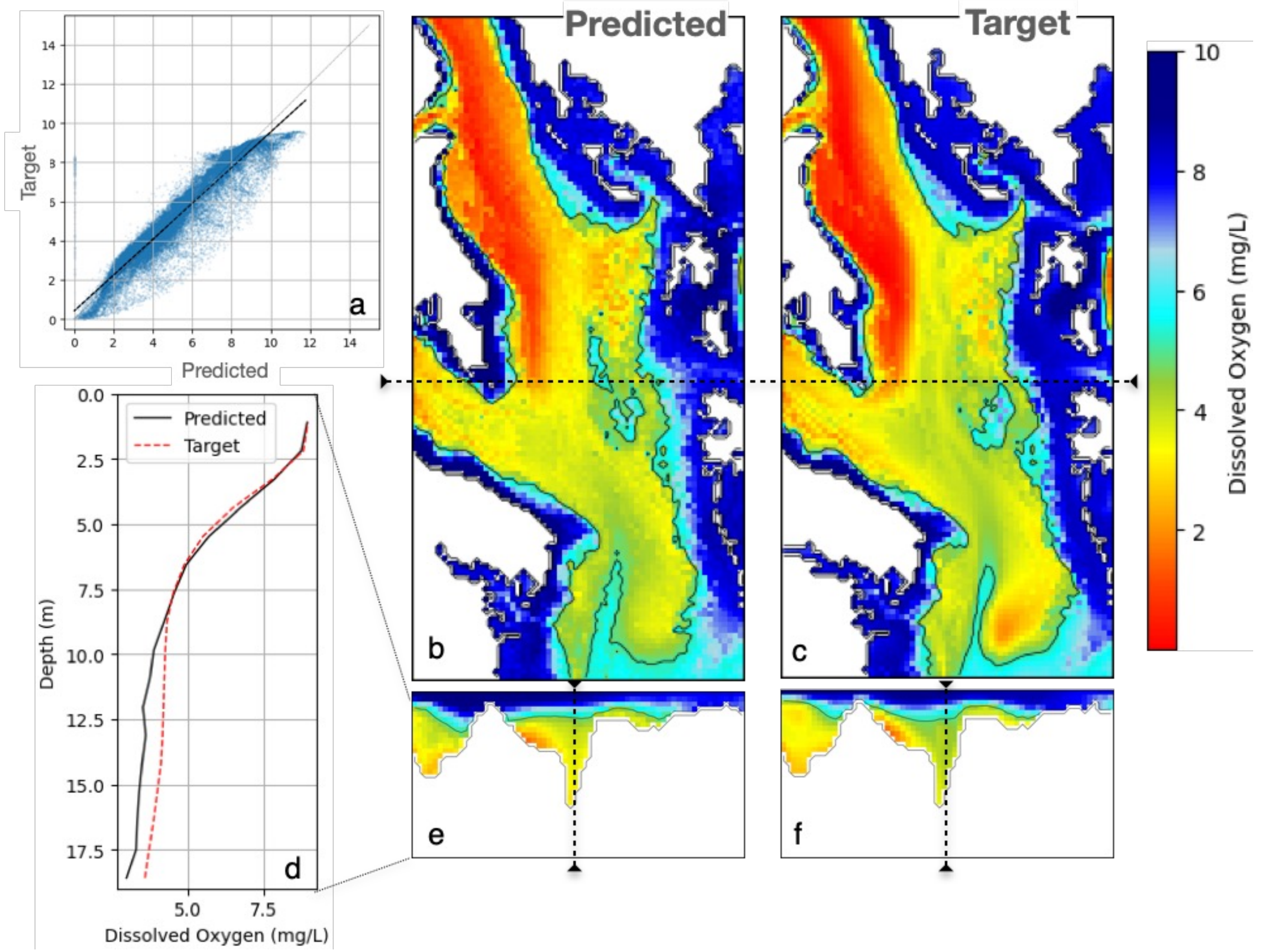
b) Surface predicted dissolved oxygen

c) Surface target dissolved oxygen

d) Predicted vs target vertical contour from the center of map

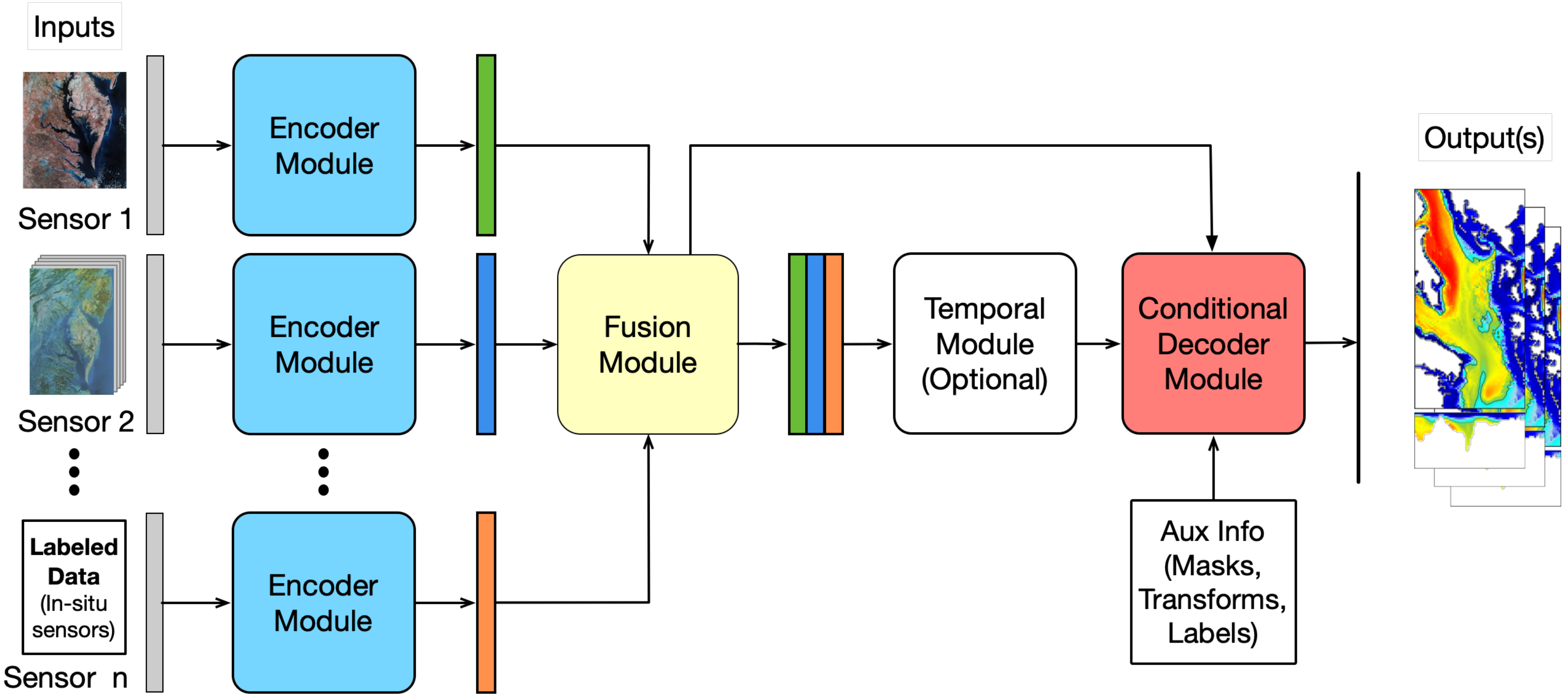
e) Predicted cross-section values at depth

f) Target cross-section at same location



Shellfish harvesting threshold: DO < 5mg/L

Deep learning for Environmental and Ecological Prediction, eValuation and Insight with Ensembles of Water quality (DEEP-VIEW)

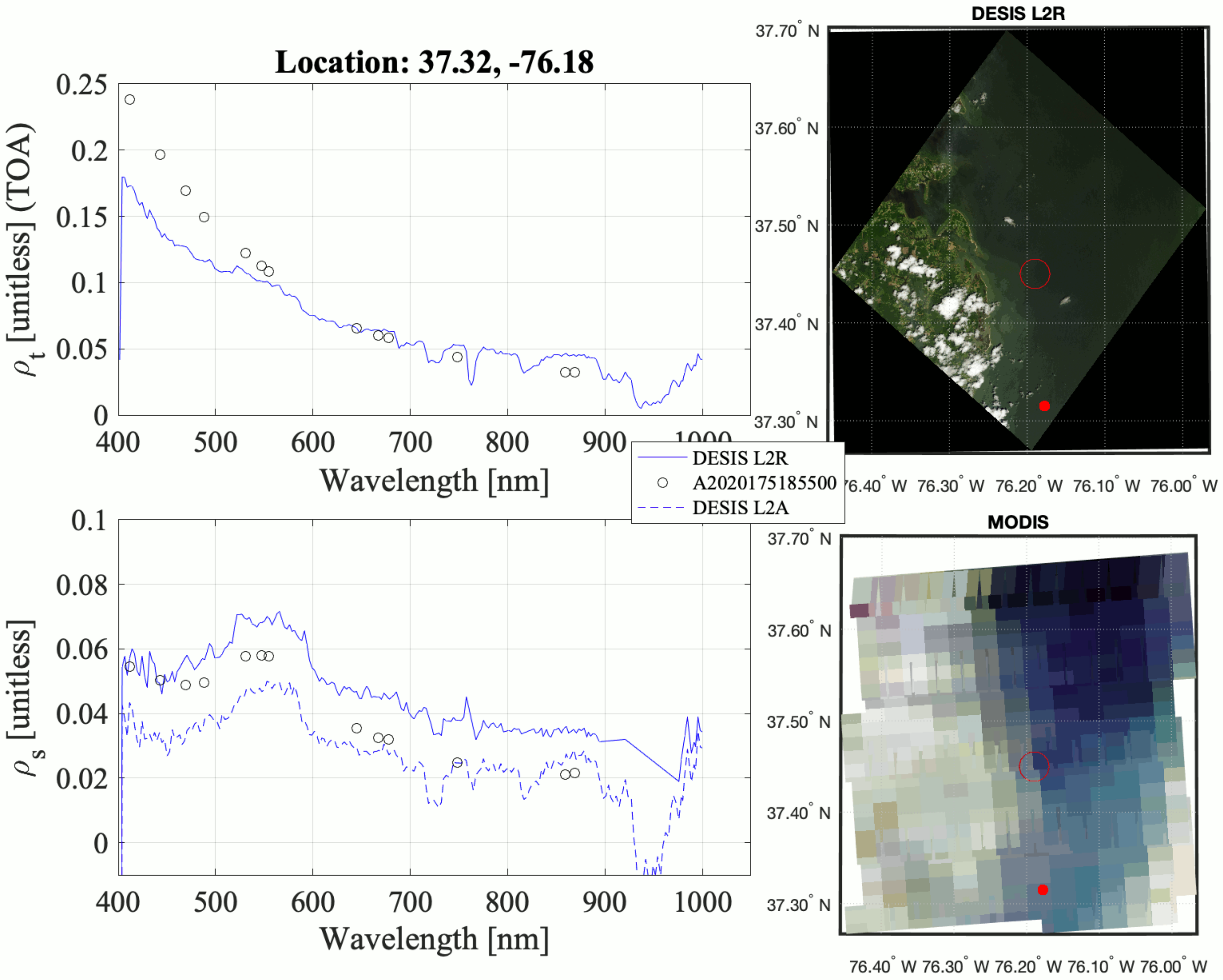


Validating Hyperspectral DESIS Satellite Data

Atmospherically correcting DESIS hyperspectral surface reflectances in Acolite using *https://github.com/acolite/acolite*

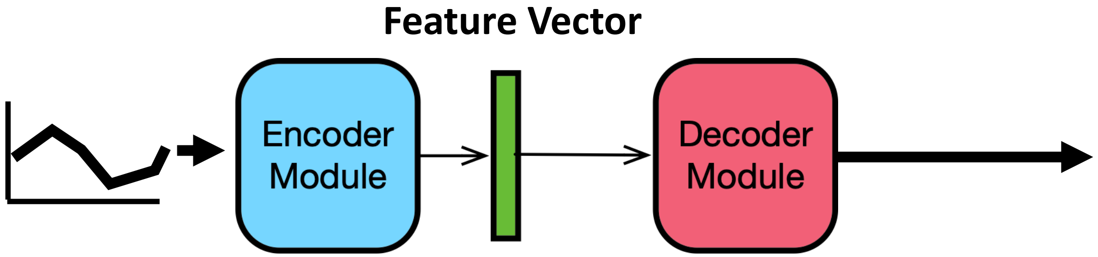
Once these are validated with in situ data (from AERONET-OC and boat), integrate into new machine learning architecture

- 235 bands from 400-1000 nm
- 30 m resolution
- 2020-2022 (156 images)

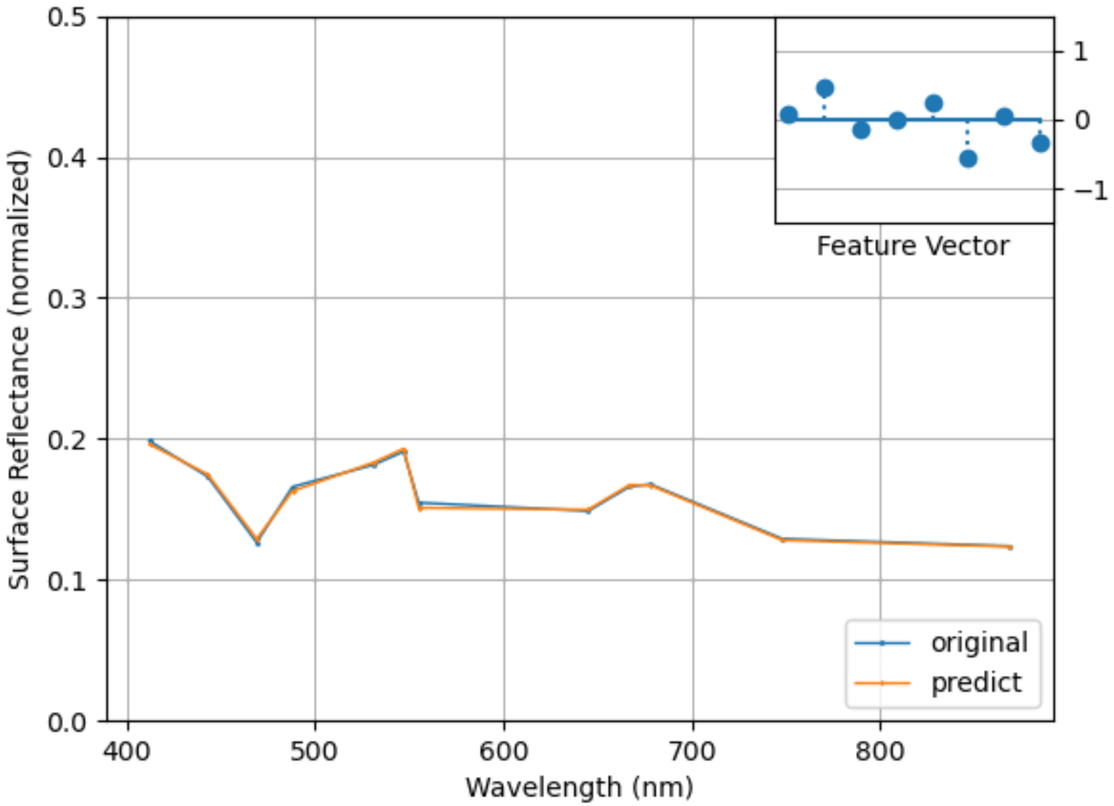


MODIS Satellite Data Spectral Feature Training

Unsupervised Feature Training



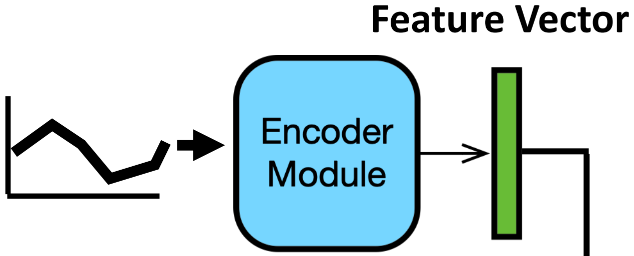
- Unsupervised learning using Autoencoder architecture on spectral data (2003-2021)
- Training sets consisting of 50K, 100K, 250K water pixels
- Trained encoder can then be used for training additional decoder(s) to predict water quality.

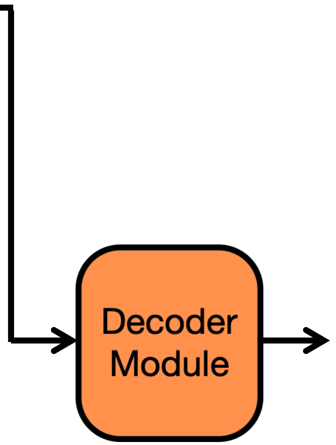


Original spectrum compared to predicted spectrum from 8-feature vector (inset)

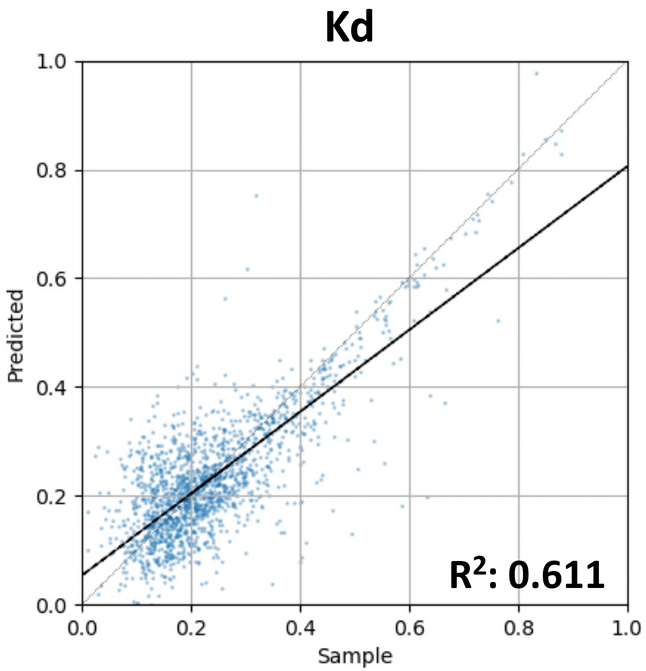
MODIS/In situ Trained ML Output vs In situ Data

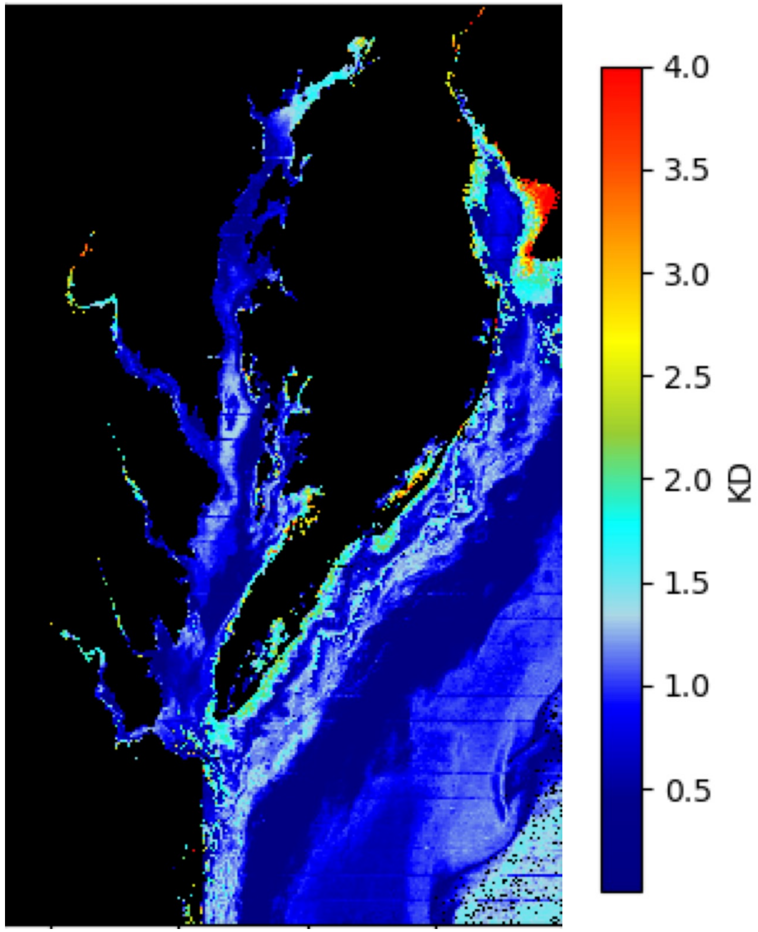
Supervised Feature Training using In situ data



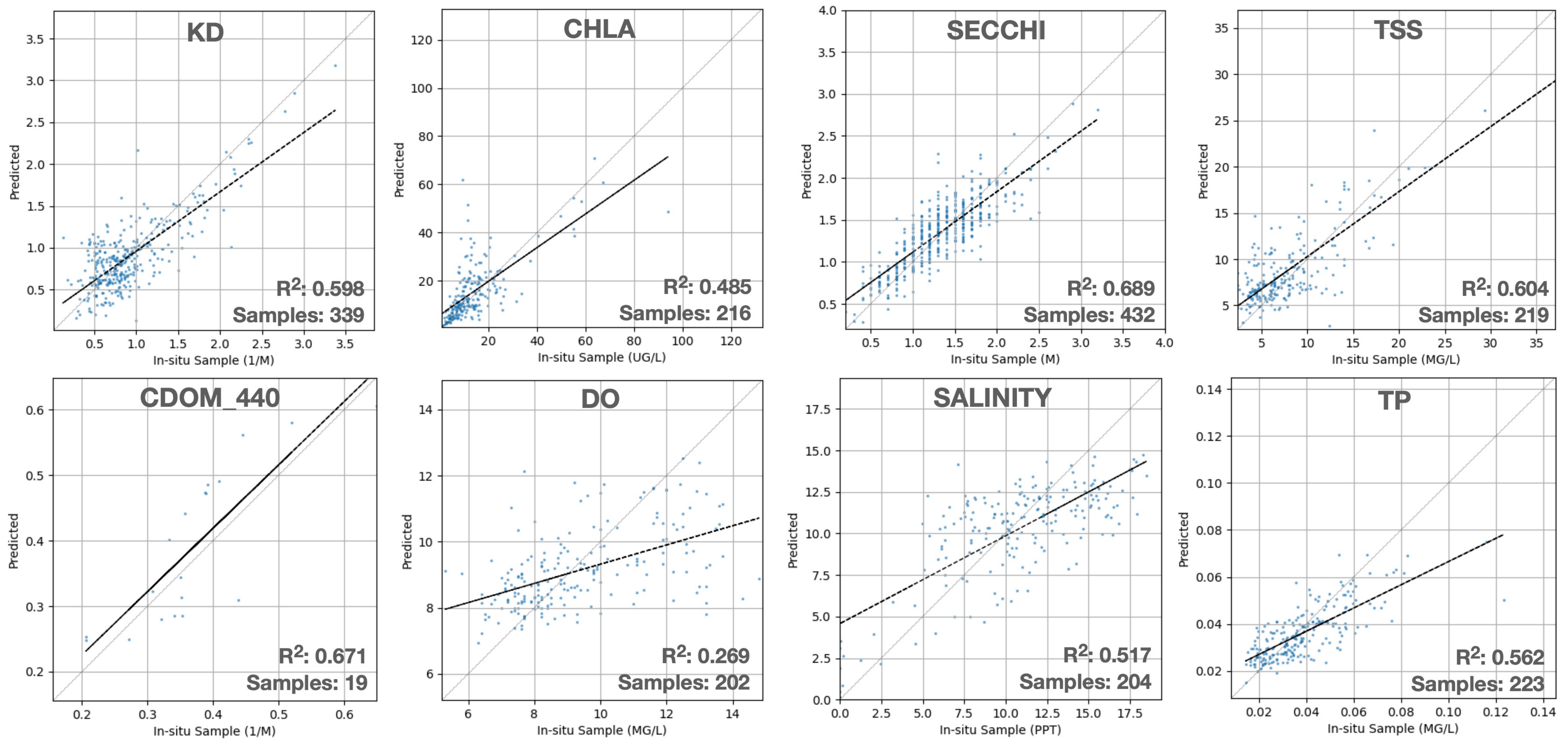


- Supervised learning using decoder module on trained spectral features
  - Correlate learned features with in situ data within 2-6 hours (weighted)
  - Thousands of in-situ matchups
- Utilize longer history of MODIS data and in-situ matchups for transfer learning to other satellite sources

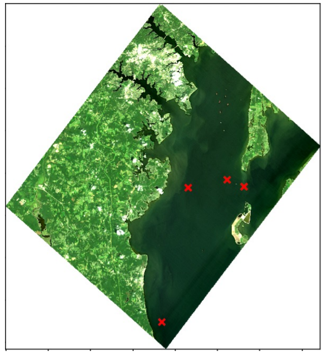


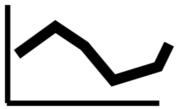


MODIS/In situ Trained ML Output vs In situ Data

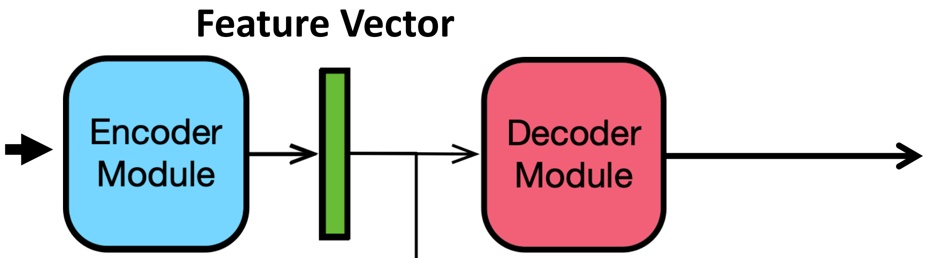


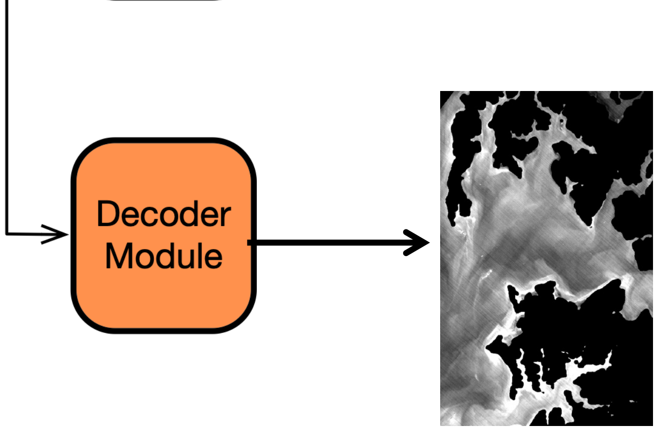
DESIS Satellite Data Unsupervised/Supervised Feature Training



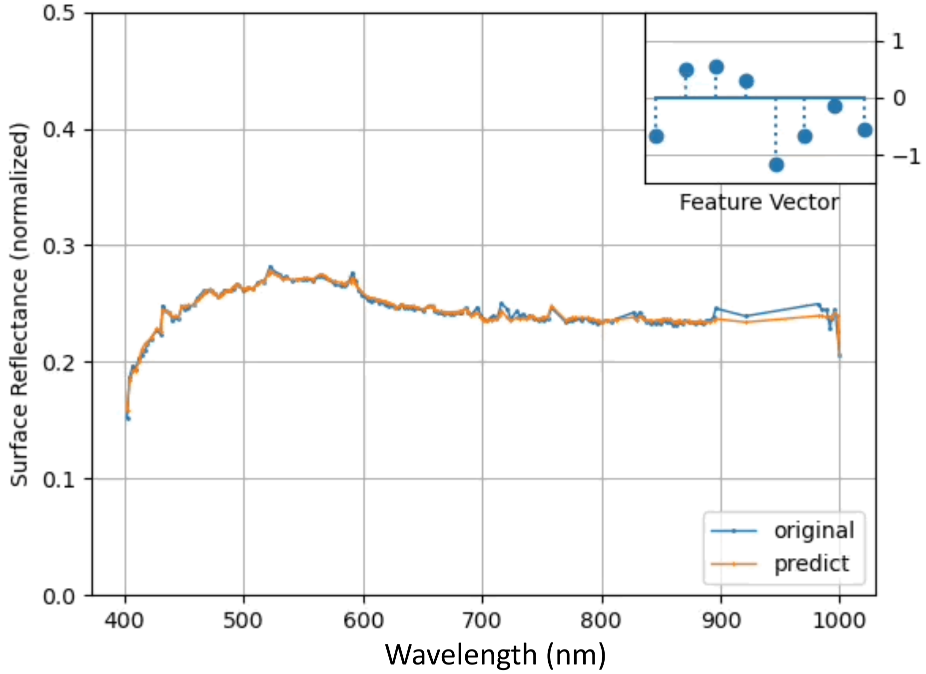


- Currently only 8 scenes have in situ matchups





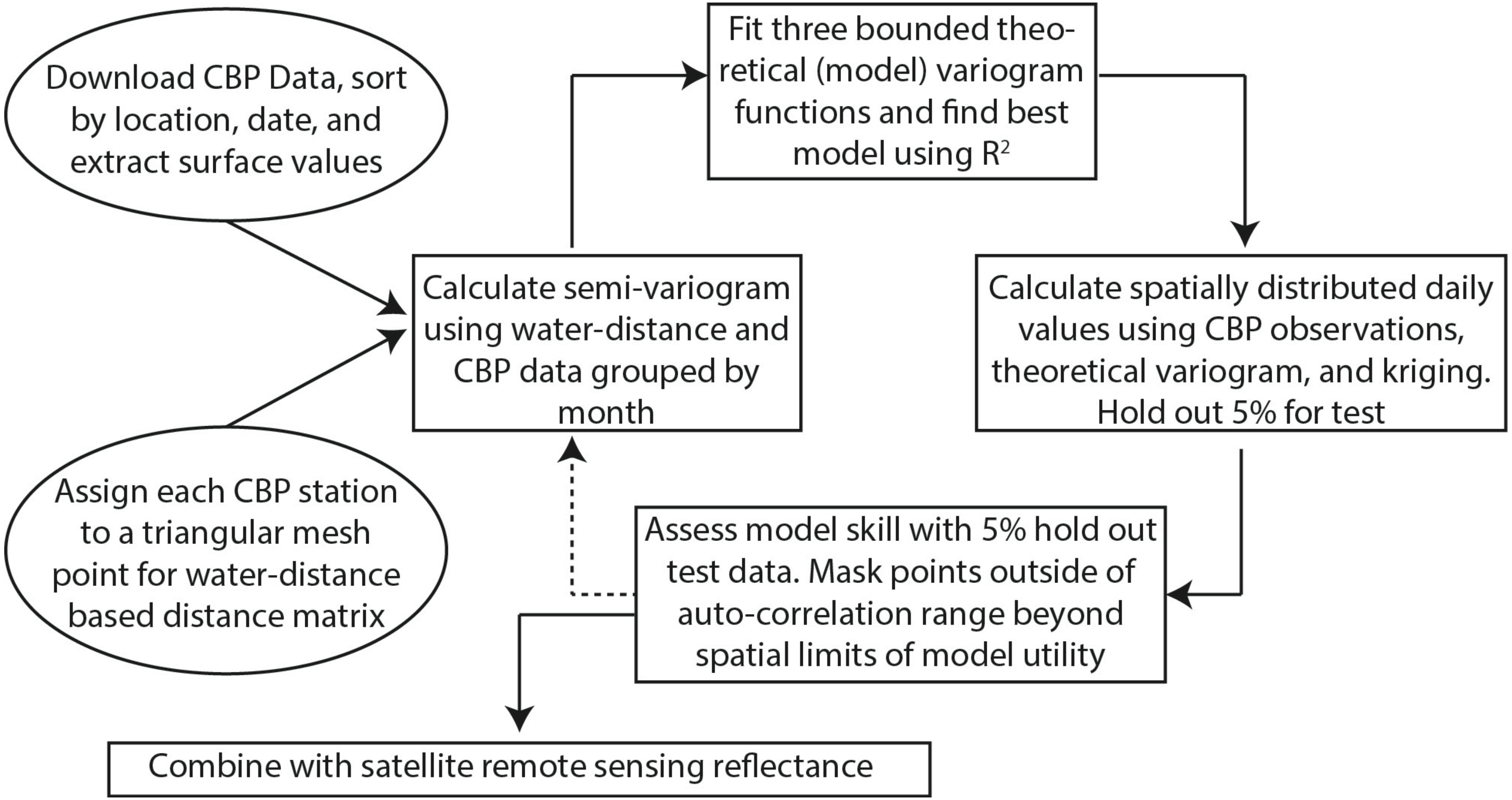
- Unsupervised learning using Autoencoder architecture on hyperspectral pixel data
- Supervised learning using MODIS trained model
- Exploit longer history of MODIS data and in-situ matchups for transfer learning to DESIS model

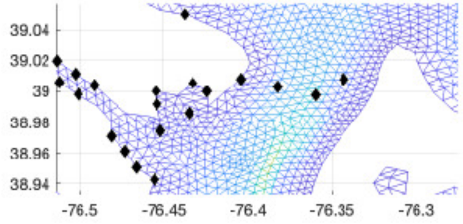


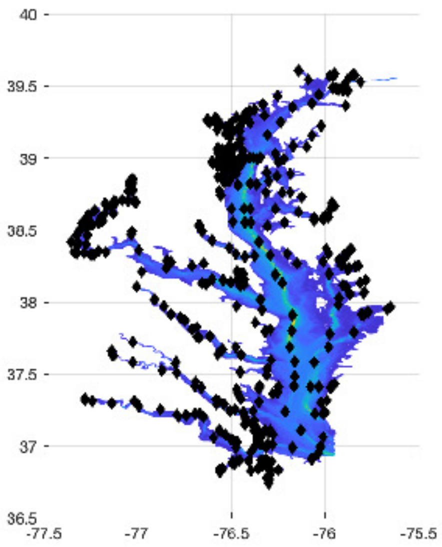
Original spectrum (177 channels) vs reconstructed spectrum from an 8-feature vector (inset)

Kriging In Situ Fields to Increase Labeled Data for ML

Kriging method uses observations weighted by distance in monthly gridded climatology

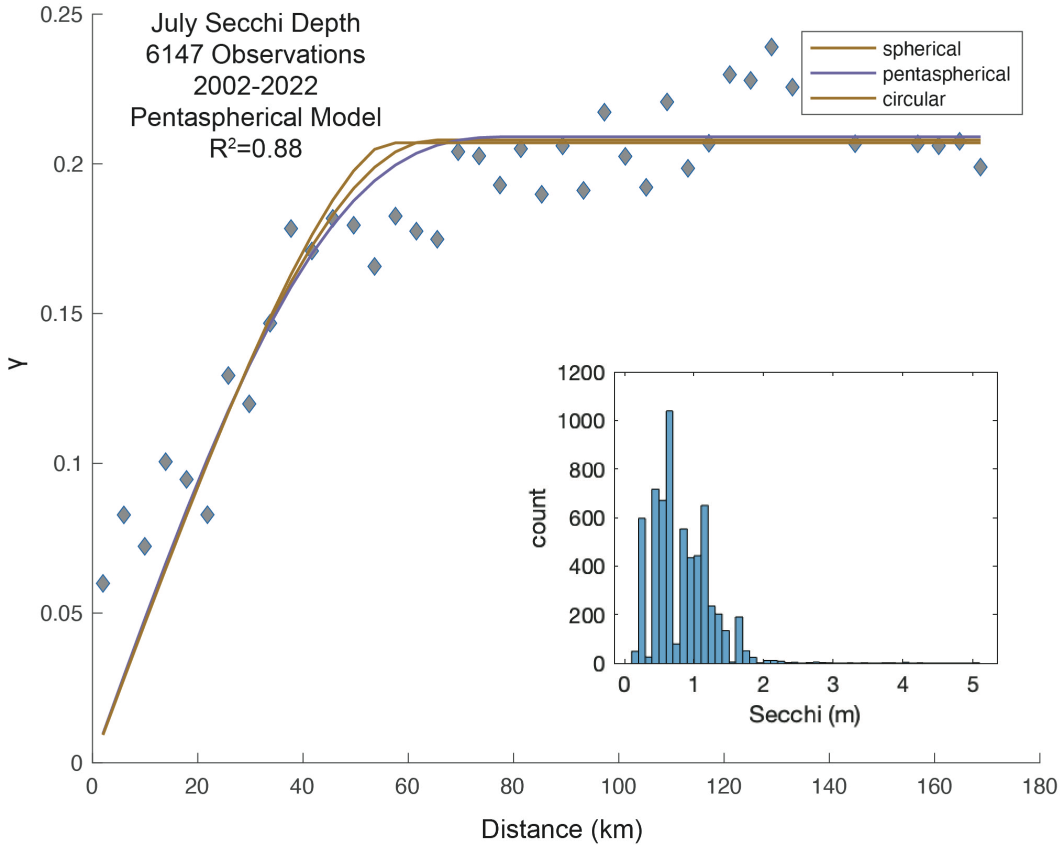


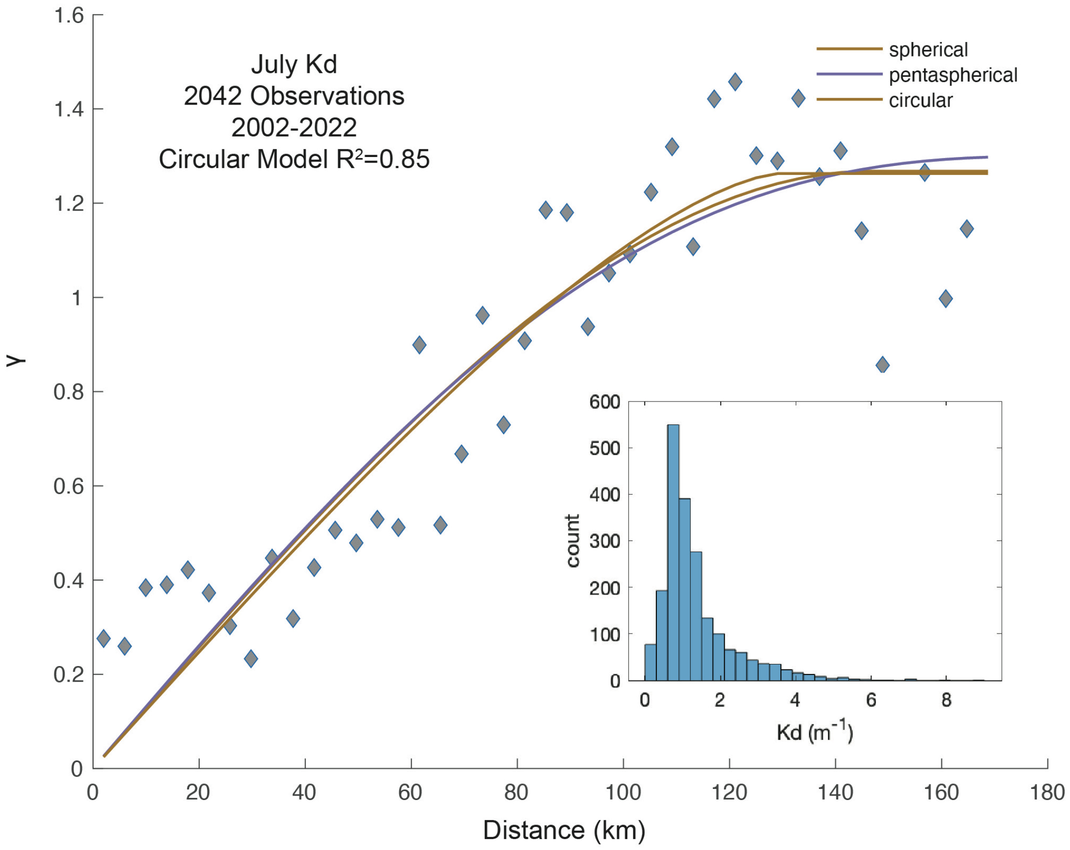




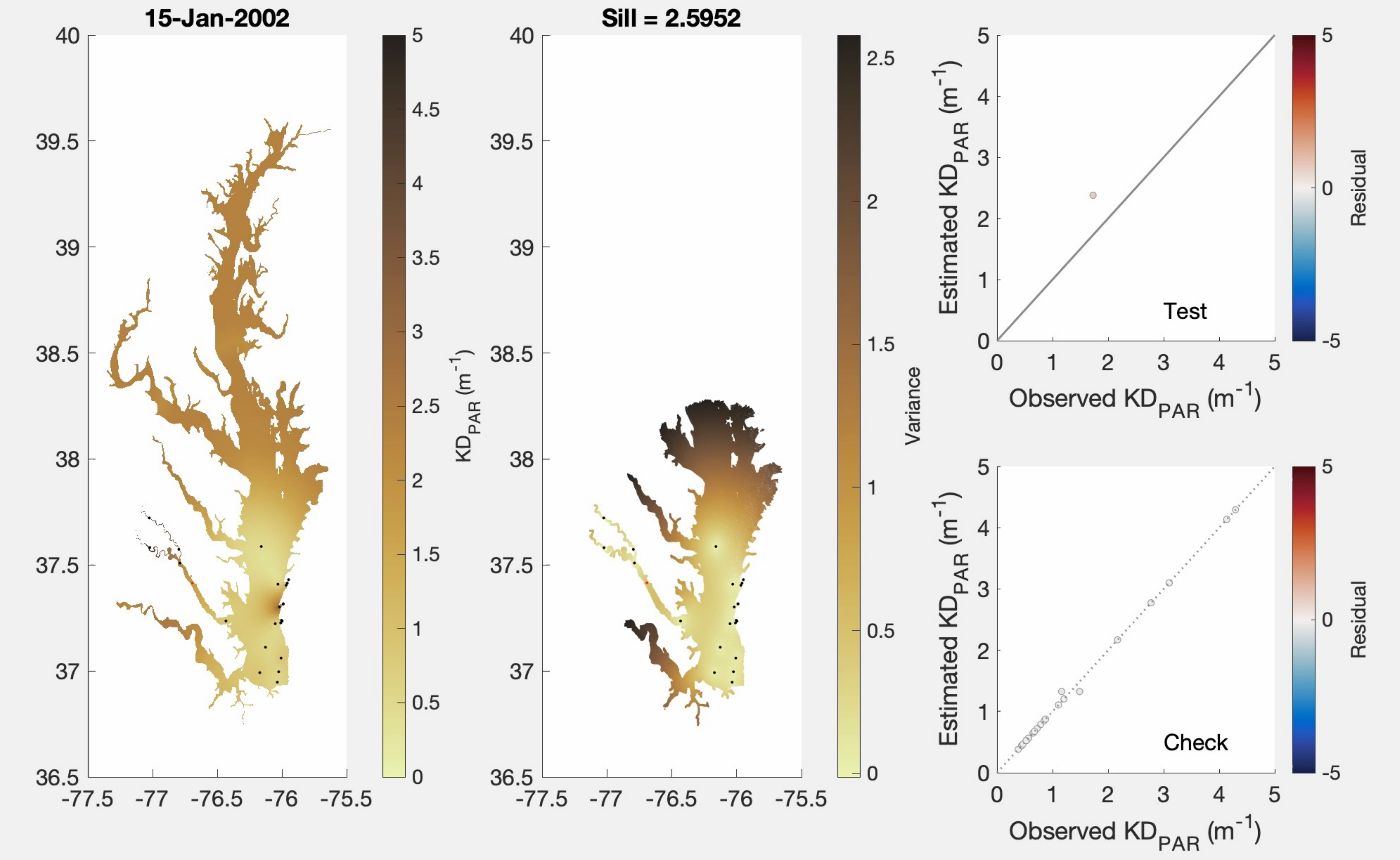
MODIS In-situ Training + Kriging Interpolated Values

Experimental (semi) variograms and theoretical variogram models





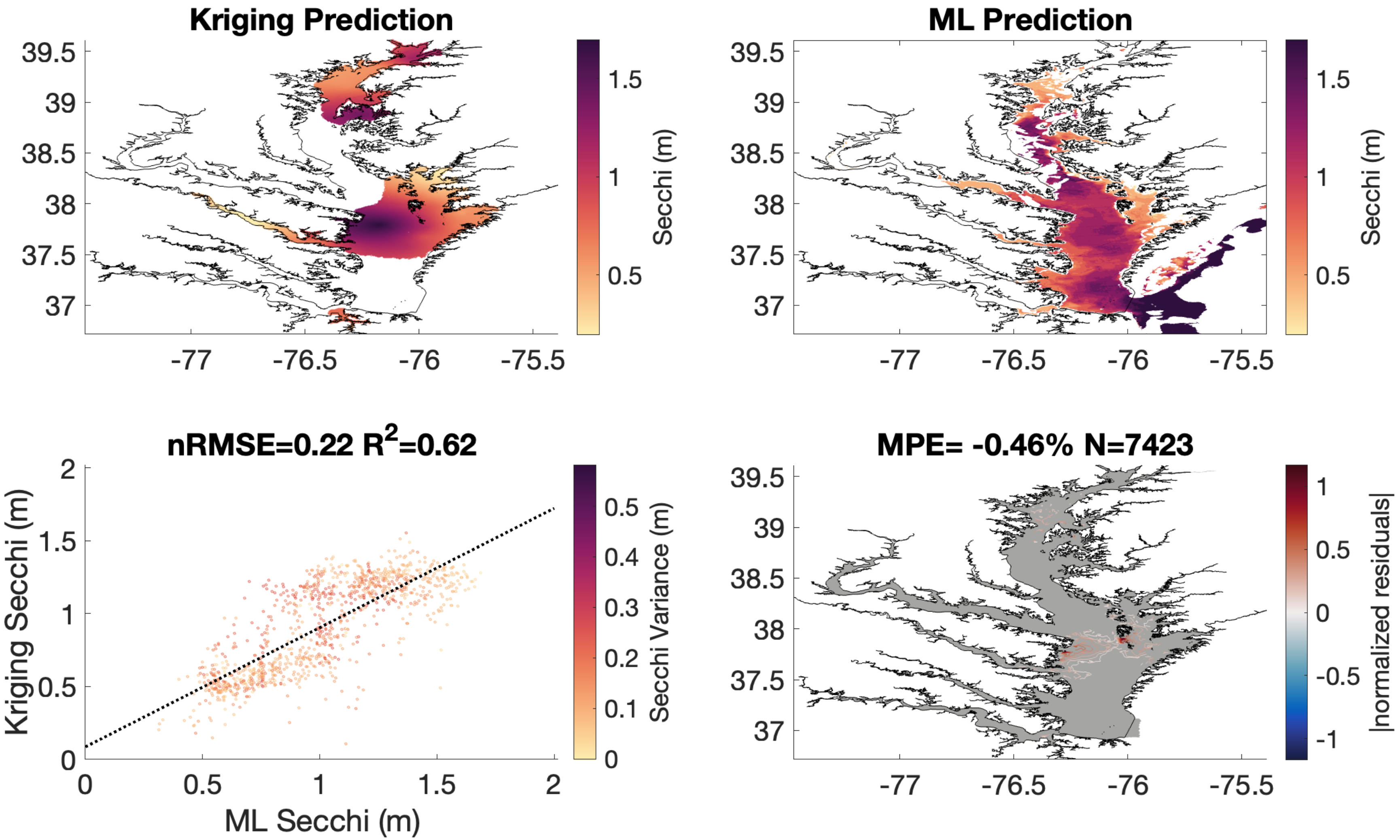
MODIS In-situ Training + Kriging Interpolated Values



Kd 2002-2022

- 547 Individual Days
- Average 15 samples per day
- 8211 input observations
- 554 test observations

MODIS In-situ Training + Kriging Interpolated Values



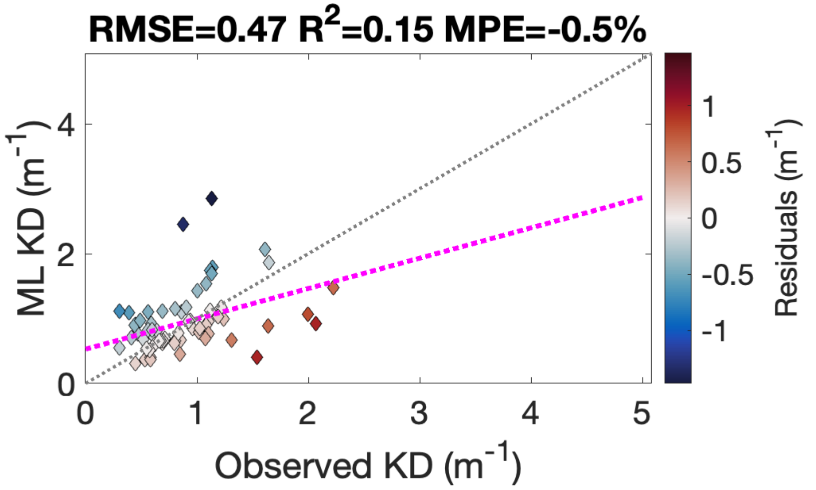
Secchi Depth
March 14, 2019

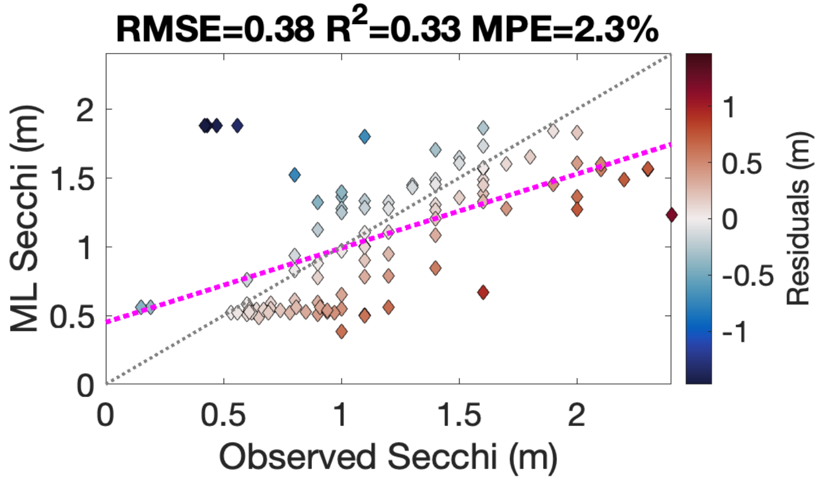
- Kriging product masked by ½ of Sill variance (strict)
- ML prediction using individual image from the same day pixels matched within 50 m

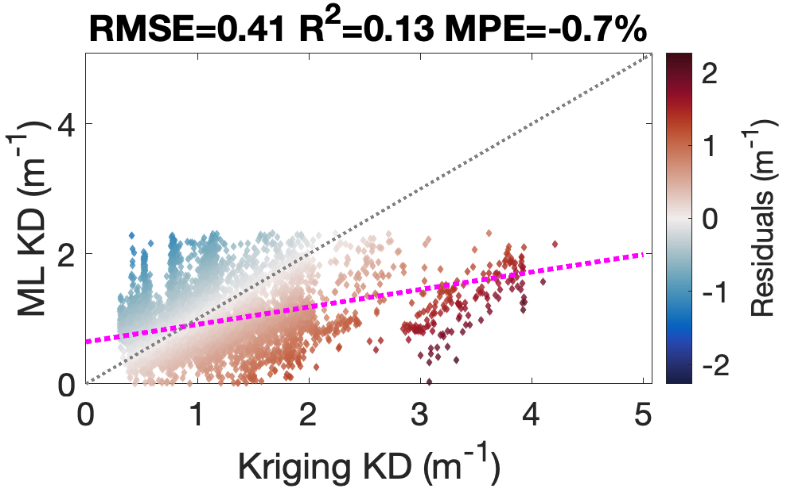
MODIS In-situ Training + Kriging Interpolated Values

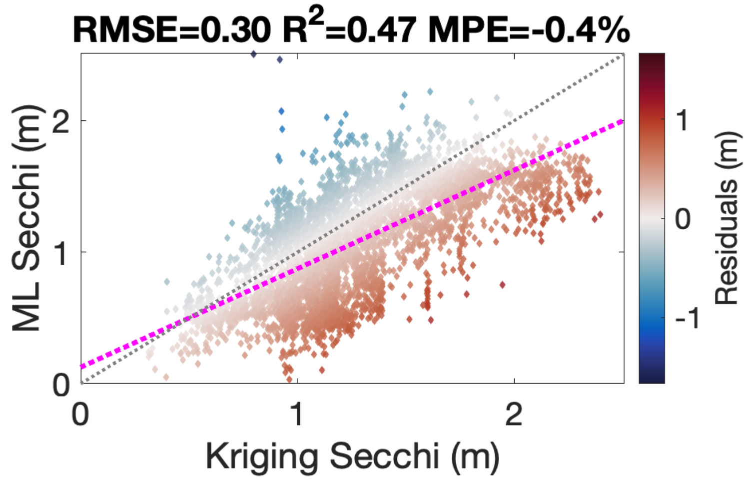
KD

Secchi









Secchi depth and Kd for four initial images

- Upper plots are ML predictions vs. observations each day
- Lower plots are ML predictions vs. Kriging predictions aggregated over the four images

Summary: many preparatory activities are reducing data gaps, quantifying uncertainties, developing DEEP-VIEW to serve resource managers as well as providing a new method for exploiting upcoming hyperspectral satellite data more broadly, e.g. process model assimilation

- Challenges remain – land adjacency for area of greatest interest, clouds and atmospheric correction of satellite data, sparse matchups within 2 hours

- Methodology developed here will be transitioned to open science cloud for interdisciplinary, e.g. land-water research, and scaling to other locations

- Transition from ADAPT to SMCE cloud services

Acronyms

• ADAPT	Advanced Data Analytics PlaTform
• CBEFS	Chesapeake Bay Environmental Forecast System
• CBP	Chesapeake Bay Program
• CDOM	Colored Dissolved Organic Matter
• CNN	Convolutional Neural Network
• CSDAP	Commercial Smallsat Data Acquisition Program
• DEEP-VIEW	Deep learning for Environmental and Ecological Prediction, eValuation and Insight with Ensembles of Water quality
• DESIS	DLR (German Space Agency) Earth Sensing Imaging Spectrometer
• EIS	Earth Information System
• HAB	Harmful Algal Bloom
• HICO	Hyperspectral Imager for the Coastal Ocean
• LSTM	Long Short Term Memory
• MODIS	Moderate-resolution Imaging Spectrometer
• MSI	Multispectral Imager
• NCCS	NASA Center for Climate Simulations
• NWQMC	National Water Quality Monitoring Council
• OLCI	Ocean and Land Color Instrument
• OLI	Operational Land Imager
• PRISMA	(Italian) Hyperspectral Precursor of the Application Mission
• Rhos	Top-of-atmosphere reflectance minus Rayleigh
• Rrs	Remote sensing reflectance
• SAA	Space Act Agreement
• SMCE	Science Managed Cloud Environment
• SST	Sea-Surface Temperature
• S2	Sentinel-2 A&B
• S3	Sentinel-3 A&B