A WEST COAST ESTUARINE CASE STUDY: A PREDICTIVE APPROACH TO MONITOR ESTUARINE EUTROPHICATION

Vedant Janapaty[1]
[1]Silver Creek High School, San Jose, CA, United States

Abstract – **Estuaries are wetlands where freshwater from streams mixes with salt water from sea. Also known as "kidneys of our planet"- they are extremely productive environments that filter pollutants, absorb floods from sea level rise, and shelter unique ecosystems. However, eutrophication and loss of native species are ailing California wetlands. There is a lack of uniform data collection and sparse research on correlations between satellite data and in situ measurements. Remote sensing (RS) has shown great promise in environmental monitoring. This study proposes using satellite data by correlating metrics with in situ observations, collected at five West Coast estuaries. Images for satellite data were processed to calculate 7 bands (SIs) using Python. Average SI values were calculated per month for 23 years. Publicly available data, from sites at each estuary, was used to obtain 10 parameters (OPs). Average OP values were calculated per month for 23 years. Linear correlations between the 7 SIs and 10 OPs were made and found to be inadequate (correlation = 1 to 64%). Fourier transform analysis on 7 SIs was performed. Dominant frequencies and amplitudes were extracted for 7 SIs and a machine learning (ML) model was trained, validated, and tested for 10 OPs. Better correlations were observed between SIs and OPs, with certain time delays (0, 3, 4, 6 month delay) and ML was again performed. The OPs saw improved R^2 values in the range of 0.2 to 0.93. This approach can be used to obtain the periodic analysis of an overall wetland health with satellite indices. The research proposes that remote sensing can be used to develop correlations with critical parameters that measure eutrophication in situ data and can be used by practitioners to easily monitor wetland health.**

NOMENCLATURE

SI	Satellite Index
OP	In Situ parameter Parameter
DFT	Discrete Fourier Transform
WH	Wetland Health
ML	Machine Learning
CRAM	California Rapid Assessment Method
i	Image
p	Pixel
d	Datapoint
s	Station

I. INTRODUCTION

Estuaries are wetlands where freshwater from streams mixes with salt water from sea. They are extremely productive environments that create the most organic matter annually, and shelter a unique ecosystem [1].

Coastal watersheds provide clean, clear water by filtering nutrients, and pollutants. Estuaries also absorb flood water and dissipate storms, protecting uplands [1].

Estuaries face 3 main threats: habitat loss from human expansion, eutrophication from farms and rising sea levels. Since the 1700s, 91% of our wetlands have been lost [2].

While efforts are being made to minimize human expansion, and thereby manage wetland extent, according to the EPA, there remains a gap in measuring wetland health (WH) and little is known about their actual ecological health.

In the United States, there is well documented, publicly available data for monitoring and reporting water quality and nutrient levels. All in situ data is publicly available through the Centralized Data Management Office (CDMO) [3]. The conservation and restoration of estuaries are a major priority in California, because salt marshes are among the most threatened in the state, having suffered an 80% loss since industrialization [4].

Remote sensing (RS) is a great tool for monitoring wetland health and progress from restoration efforts. RS offers continuous, low cost, large spatial, spectral and temporal coverage over point sampling methods. As a result, they can be used reliably and efficiently for environmental monitoring, protection, and sustainability.

A. Current Methods

The National Wetland Condition Assessment (NWCA) conducts a field sampling every five years [5]. The U.S Fish and Wildlife Service (FWS) measures changes in wetland acreage [6] [7]. The California Rapid Assessment Method (CRAM) uses data from flora and fauna to assess WH, but is cost and labor intensive, and thus conducted infrequently [8] [9]. In the west coast, stations are sampled monthly and in situ data is collected from stations using sondes daily.

B. Research on using Remote Sensing to monitor Wetland Health

Research by scientists at NCCOS combine observations from different satellites and found some correlations between NDVI and 2 sample months. Other research by Guo and Li compares sensors for wetland classification [10]. A correlative analysis between satellite indices and in situ data using Fourier transform is not yet available.

II. LITERATURE PRECEDENT

A. Water Quality and Nutrients

Phosphorous (P) and Nitrogen (N) are essential nutrients for the aquatic food web [11]. However, even modest changes in P and N can set of a domino effect of algal blooms, low dissolved oxygen, and loss of flora and fauna. The chemistry of eutrophication can be described by [12]:

$$106\,CO_2 + 16\,NO_3^- + HPO_4^{2-} + 122H_2O + 18\,H^+ \xrightarrow{\text{Energy}} C_{106}H_{263}O_{11}N_{16}P \text{ (bioplasm of algae)} + 138\,O_2$$

Fig. 1. Chemistry of Eutrophication

Wetlands help solve eutrophication, by transforming nitrates to free nitrogen and phosphates to phosphorus for plant adsorption. However, excessive nutrient levels contribute to wetland destruction and marsh dieback. The health of a wetland can be measured by monitoring nutrient levels, chlorophyll-A, dissolved oxygen, turbidity, pH, temperature, specific conductivity and salinity.

B. Discrete Fourier Transform

DFT decomposes a waveform, which is a function of time, into frequencies, enabling us to find the spectrum of the signal. DFT is often used in wireless technologies to disentangle transmissions from different users [13].

DFT is also used in data compression. A few samples of the entire data can be used to represent the weighted sum. Frequencies with low weights are discarded, allowing the data to be represented with fewer bits, and with minimal loss of fidelity [13].

III. HYPOTHESIS

This research proposes that if satellite indices such as vegetation, temperature, and dissolved oxygen are suitably correlated with in situ data, then the health of a wetland can be predicted [14].

IV. MATERIALS AND METHODOLOGY

In Situ Data	*Units*
Temperature	C°
Specific Conductivity	mS/cm
Salinity	psu
Dissolved Oxygen	mg/L
pH	
Turbidity	FNU/NTU
NO_{23}	mg/L
PO_4	mg/L
Chlorophyll-A	mg/L
NH4	mg/L

Bands	*Pixel Size*	*Wavelength*	*Description*
Band 1	30	0.45 - 0.52 $\mu\,m$	Blue
Band 2	30	0.52 - 0.60 $\mu\,m$	Green
Band 3	30	0.63 - 0.69 $\mu\,m$	Red
Band 4	30	0.77 - 0.90 $\mu\,m$	Near Infrared
Band 5	30	1.55 - 1.75 $\mu\,m$	SW Infrared 1
Band 6	60	10.40 - 12.50 $\mu\,m$	Th Infrared 1
Band 7	30	2.08 - 2.35 $\mu\,m$	SW Infrared 2

Bands 1-7 (SIs) were obtained from Landsat 7 using Google Earth Engine(GEE). The sample band data for all 7 bands was obtained. The five, largest west coast estuaries were analyzed. Pixelwise SI data were obtained using shapefiles. In situ data (OPs) for 10 critical parameters were obtained from publicly available datasets provided by CDMO. About 10 million in situ records and 2 million images were processed using Python code.

Over a million images were obtained for the 5 west coast estuaries using Google Earth Engine and Colab. A shapefile of each estuary was applied to the images to obtain pixel-wise band data for each image. Cloud masks used to remove cloudy data from images. Pixel-wise averages were computed for every month. Similarly, in situ data was obtained from publicly available datasets from CDMO for the 5 estuaries. Monthly averages were for 271 months for the last 23 years, since 1999. Figure 4 shows the image preprocessing done. Equations 1-4 show the methods used to obtain monthly data. In situ data are hereafter referred to as OPs, while the satellite bands are referred to as SIs.

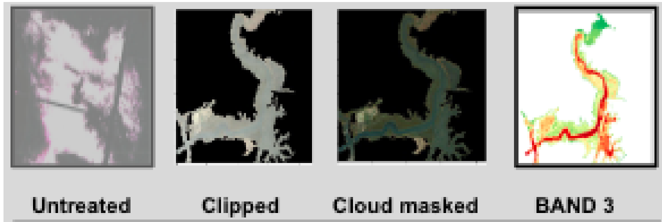

Fig. 2. Image preprocessing for each Satellite Index (SI)- Elkhorn Slough, as an example

A. Methods Used For Image, Pixel, and Datapoint Processing

The following equations were used for data processing of the SIs and OPs.

$$[SI_{\mathrm{m,y}}]_{\mathrm{image}} = \frac{1}{N_{\mathrm{i}}} \times \sum_{i=1}^{N_{\mathrm{i}}} SI_{\mathrm{i,m,y}}$$
$$[SI_{\mathrm{m,y}}]_{\mathrm{overall}} = \frac{1}{N_{\mathrm{p}}} \times \sum_{p=1}^{N_{\mathrm{p}}} SI_{\mathrm{p,m,y}}$$
$$[OP_{\mathrm{m,y}}]_{\mathrm{stations}} = \frac{1}{N_{\mathrm{d}}} \times \sum_{d=1}^{N_{\mathrm{d}}} OP_{\mathrm{d,m,y}}$$
$$[OP_{\mathrm{m,y}}]_{\mathrm{overall}} = \frac{1}{N_{\mathrm{s}}} \times \sum_{s=1}^{N_{\mathrm{d}}} OP_{\mathrm{m,y}}$$

B. Padilla Bay

Padilla Bay is right in the heart of the Salish Sea, with a massive eelgrass meadow. It totally measures 8,000 acres, making it the second largest estuary on North America's Pacific Coast. Padilla Bay is one of the many estuaries, identified by the National Estuarine Research Reserve, established to protect coastal land for long-term research.

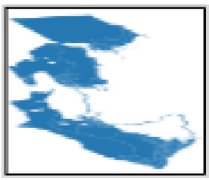

Fig. 3. Padilla Bay, Washington

C. South Slough

The South Slough National Estuarine Research Reserve measures nearly 7,000 acres of natural areas along the coast of Oregon. The NERRS's main environmental goals are to protect and enhance water quality in the estuary, improve habitat through restoration, improve research and understanding of coastal issues, and conserve land for flood reduction and improve surrounding ecological resilience for landowners.

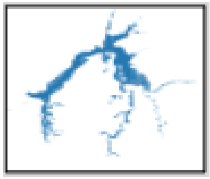

Fig. 4. South Slough, Oregon

D. San Francisco Estuary

Currently the largest west coast estuary, the San Francisco Estuary is an extremely diverse system **?** [15]. Measuring 400,000 acres, the San Francisco Bay accounts for 77 percent of California's remaining estuaries. However, in recent years, several portions have been diked off to increase agricultural territory and urban development. Since the estuary has a low elevation, there is an increased risk of flooding. Sea level rise and increased rainfall exacerbate this problem.

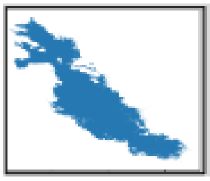

Fig. 5. San Francisco, CA

E. Elkhorn Slough

Elkhorn Slough provides habitats for a variety of species, ranging from raptors to milkweed. Elkhorn Slough Foundation researchers stress the importance of water quality monitoring. Poor water quality has reduced wildlife diversity in Elkhorn Slough wetlands. Research shows that increased tidal flow improves water quality grades.

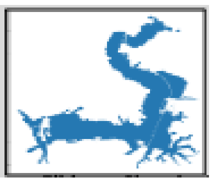

Fig. 6. Watsonville, CA

F. Tijuana Bay

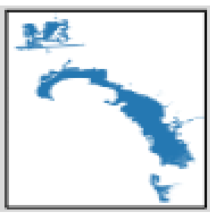

Fig. 7. San Diego, CA

The Tijuana estuary, a shallow water habitat, is termed as an "intermittent estuary". Intermittent estuaries are subjected to vast changes in stream flow [16]. For example, due to the California drought crisis, several parts of the estuary are left dry, but flooding continues to innundate other parts. Sewage from Mexico continue to pollute Tijuana Bay water.

V. RESULTS - LINEAR CORRELATION

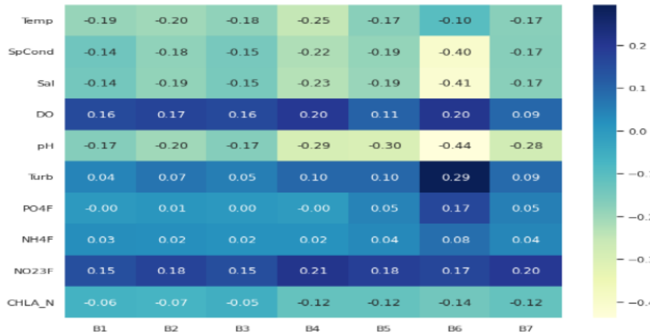

Fig. 8. Poor Linear Correlation Between SIs and OPs in the Range of 0 to 44 %

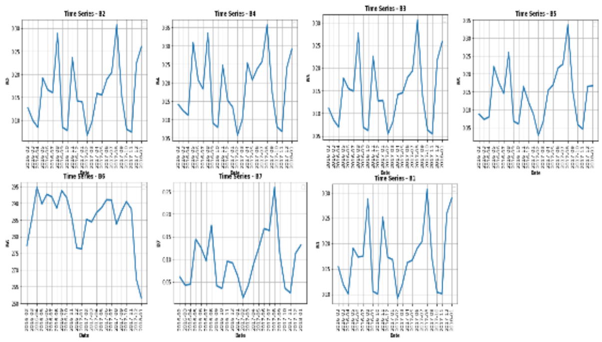

Fig. 9. Observed wave-like periodic trends for all SIs

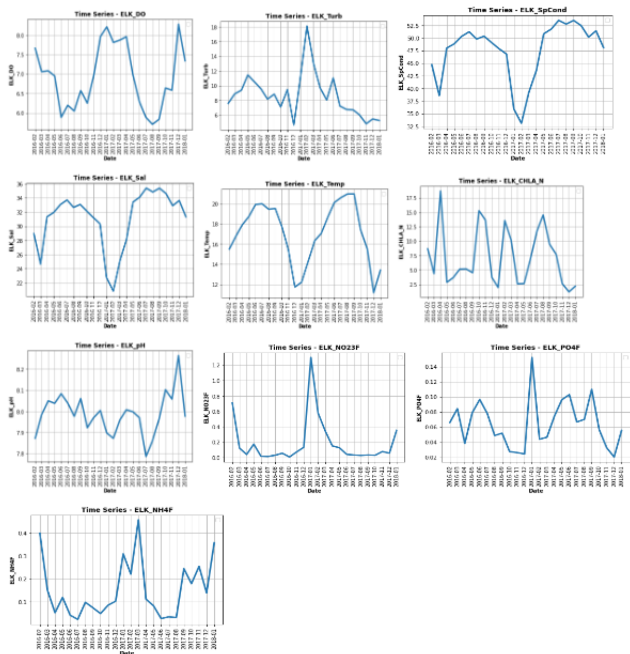

Fig. 10. Observed wave-like periodic trends for all OPs

SI parameters were modelled with equations with dominant frequencies and corresponding amplitudes. These equations are used for further analysis and machine learning (ML).

Wave-like periodic trends were observed for all seven bands (SIs) and 10 in situ parameters (OPs).

VI. RESULTS - FOURIER TRANSFORMATION EXTRACTED EQUATIONS

The equation for the Fourier series in terms of sines and cosines is:

$$x(t) = \sum_{n=0}^{N/2} a_n cos(2\pi nt / N\Delta t) + b_n sin(2\pi nt / N\Delta t)$$

SI parameters are modeled with equations with dominant frequencies and corresponding amplitudes. These equations are used for further analysis and ML.

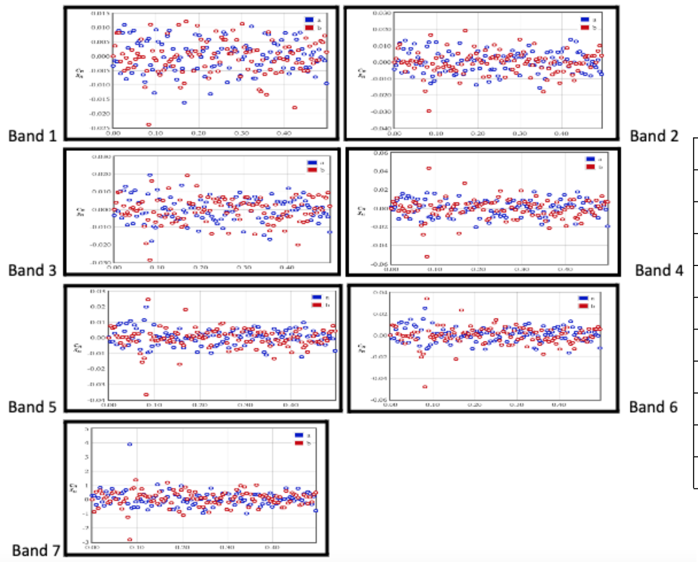

Fig. 11. 7 SIs set to Fourier Transform

SI parameters were modeled with equations with dominant frequencies and corresponding amplitudes. These equations were used for further analysis and machine learning (ML).

The sample equations for each of the SIs after Fourier Transformation were

Band 1: $0.00517605 \times cos(\frac{2\times\pi\times19\times t}{233}) - 0.015231553 \times sin(\frac{2\times\pi\times19\times t}{233}) + 0.0048606412 \times cos(\frac{2\times\pi\times39\times t}{233}) + 0.023020227 \times sin(\frac{2\times\pi\times39\times t}{233})$

Band 2: $0.019250147 \times cos(\frac{2\times\pi\times19\times t}{233}) - 0.026618878 \times sin(\frac{2\times\pi\times19\times t}{233}) + 0.010267703 \times cos(\frac{2\times\pi\times39\times t}{233}) + 0.018949636 \times sin(\frac{2\times\pi\times39\times t}{233})$

Band 3: $0.019566245 \times cos(\frac{2\times\pi\times19\times t}{233}) - 0.028545021 \times sin(\frac{2\times\pi\times19\times t}{233}) - 0.010427037 \times cos(\frac{2\times\pi\times39\times t}{233}) - 0.01930217 \times sin(\frac{2\times\pi\times39\times t}{233})$

Band 4: $0.016815123 \times cos(\frac{2\times\pi\times19\times t}{233}) - 0.052004476 \times sin(\frac{2\times\pi\times19\times t}{233}) - 0.0043411076 \times cos(\frac{2\times\pi\times39\times t}{233}) + 0.027178362 \times sin(\frac{2\times\pi\times39\times t}{233})$

Band 5: $0.025196353 \times cos(\frac{2\times\pi\times19\times t}{233}) - 0.048089676 \times sin(\frac{2\times\pi\times19\times t}{233}) - 0.0014710099 \times cos(\frac{2\times\pi\times39\times t}{233}) + 0.023377723 \times sin(\frac{2\times\pi\times39\times t}{233})$

Band 6: $3.8920651 \times cos(\frac{2\times\pi\times19\times t}{233}) - 2.8288747 \times sin(\frac{2\times\pi\times19\times t}{233})$

Band 7: $0.011137073 \times cos(\frac{2\times\pi\times19\times t}{233}) - 0.038350374 \times sin(\frac{2\times\pi\times19\times t}{233}) + 0.0036044988 \times cos(\frac{2\times\pi\times39\times t}{233}) - 0.015166007 \times sin(\frac{2\times\pi\times39\times t}{233})$

SI parameters are modelled with equations with dominant frequencies and corresponding amplitudes. These equations are used for further analysis and machine learning (ML).

VII. RESULTS - MACHINE LEARNING WITH EXTRACTED SIs

OPs	*KNN*	*Rand Forest*	*D. Tree*	*Auto ML*
Temp	91.16%	96.66%	97.68%	99.0%
SP Cond	49.55%	89.43%	97.03%	98.2%
SAL	63.58%	89.98%	96.89%	98.1%
DO	79.23%	91.22%	90.95%	99.1%
pH	20.83%	77.36%	85.35%	96.6%
TURB	48.07%	80.99%	92.51%	96.1%
PHOS	63.65%	87.54%	91.05%	97.7%
NH4	44.08%	58.37%	60.42%	98.1%
NO_{23}	72.25%	87.58%	92.2%	96.4%
CHLA	20.74%	82.58%	92.67%	98.9%

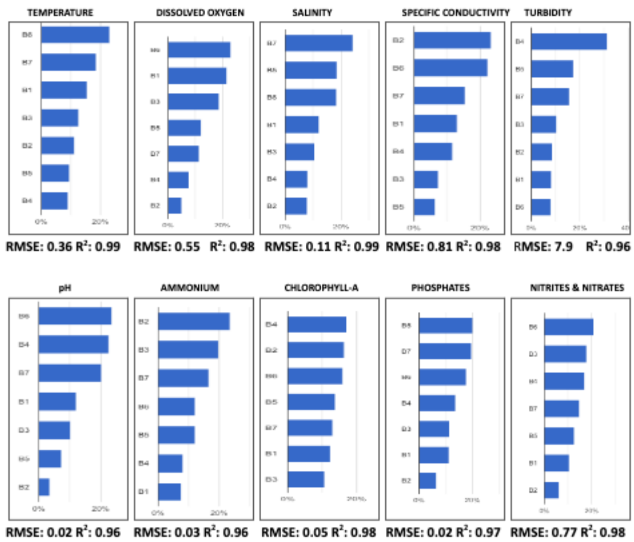

Fig. 12. Feature Importance for 10 OPs for 7 SI bands

The extracted equations for each SI and the raw data for each of the ten OPs were trained, validated, and tested using ML **?** [17]. The results above show the feature importance for each OP, and the Root Mean Square Error (RMSE) [18] and R^2 value for each OP parameter. Four regression based machine learning algorithms were used to analyze the relationships between the 7 SIs and each of the 10 OPs (Table 4). While all algorithms yielded good results, the results for Auto-ML were the best. They showed a drastic improvement with R^2 values in the range of 96.1% to 99% for all 10 critical parameters necessary to track eutrophication.

Especially strong R^2 values were found for temperature (99%), dissolved oxygen (99%), salinity (98.1%), and specific conductivity (98.2%). Crucial parameters for eutrophication showed very strong results - Dissolved Oxygen (99.1%), chlorophyll-A (98.9%), Ammonium (98.1%), Phosphates (97.7%), pH (96.6%), and Nitrites Nitrates (96.4%). Strong correlations were also found for optically insensitive OPs such as pH, PO_4, and NO_{23}.

The extracted equations for each SI and the raw data for each of the ten OPs were trained, validated, and tested using ML. The results above show the feature importance for each OP, and the Root Mean Square Error (RMSE) and R^2 value for each OP parameter. Four regression based machine learning algorithms were used to analyze the relationships between the 7 SIs and each of the 10 OPs (Table 4). While all algorithms yielded good results, the results for Auto-ML were the best. They showed a drastic improvement with R^2 values in the range of 96.1% to 99% for all 10 critical parameters necessary to track eutrophication. Especially strong R^2 values were found for temperature (99%), dissolved oxygen (99%), salinity (98.1%), and specific conductivity (98.2%). Crucial parameters for eutrophication showed very strong results - Dissolved Oxygen (99.1%), chlorophyll-A (98.9%), Ammonium (98.1%), Phosphates (97.7%), pH (96.6 %), and Nitrites Nitrates (96.4%). Strong correlations were also found for optically insensitive OPs such as pH, PO_4, and NO_{23}.

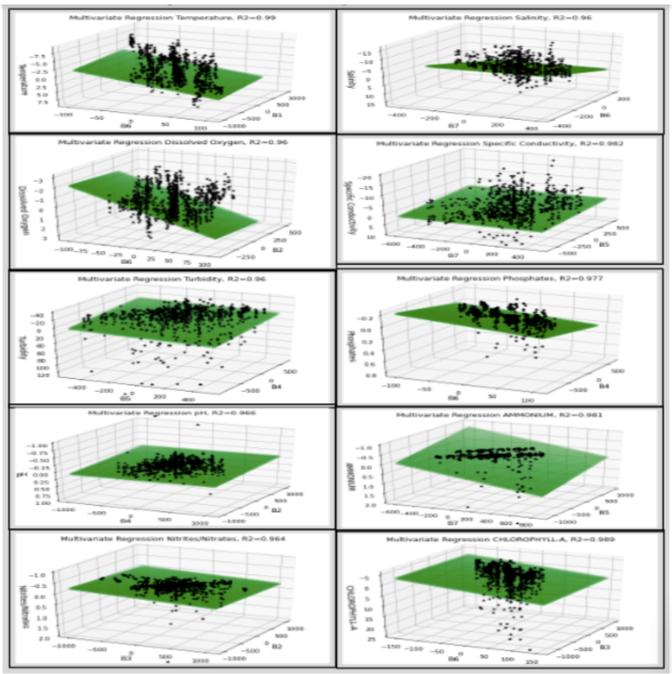

Fig. 13. Multivariate Regression for 10 OPs

VIII. DISCUSSION

GEE combined with Landsat 7 images allowed pixel wise sampling for the region of interest for all West coast estuaries

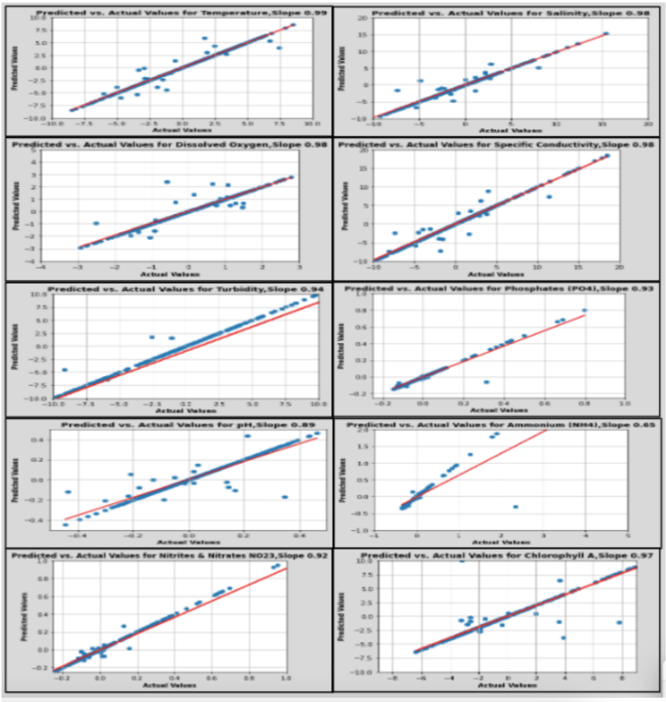

Fig. 14. Predicted vs. Actual values for 10 OPs (using Decision Tree Rg ML) with slopes close to 1

for 23 years. After experimenting with 21 SIs, raw band data 1-7 yielded the best results. Overall, the project supports the hypothesis that correlations can be established between SIs and OPs.

Applying Discrete Fourier Transform (DFT) to the seven SIs and extracting frequencies yields drastic improvements to correlations. Periodic 6-month and annual cycles observed in time series charts (Figs 7, 8) led to DFT analysis being performed on the 7 SIs (Fig 10). Dominant frequencies and amplitudes were extracted for SIs and used for the machine learning model (Eq 5) (Table 3).

Four regression based machine learning algorithms were used to analyze the relationships between the 7 SIs and each of the 10 OPs (Table 4). While all algorithms yielded good results, the results for Auto-ML were the best. They showed a drastic improvement with R^2 values in the range of 96.1% to 99% for all 10 critical parameters necessary to track eutrophication. Especially strong R^2 values were found for temperature (99%), dissolved oxygen (99%), salinity (98.1%), and specific conductivity (98.2%). Crucial parameters for eutrophication showed very strong results - Dissolved Oxygen (99.1%), chlorophyll-A (98.9%), Ammonium (98.1%), Phosphates (97.7%), pH (96.6%), and Nitrites Nitrates (96.4%). Strong correlations were also found for optically insensitive OPs such as pH, PO_4, and NO_{23}.

This research proposes that using raw band data, with extracted DFT frequencies coupled with time delay and machine learning was developed. Additionally, the feature importance for OPs can be used to develop models for other estuaries.

This research proposes that combining in situ data with satellite data, offers a powerful tool compared to using in situ data alone, because it allows for monitoring recovery efforts, historic perspectives and causes and effects.

IX. CONCLUSION

This project was successful in several areas:

- It supports the hypothesis that remote sensing can be used to develop correlations with in situ data for all west coast estuaries.
- Employed 7 unique satellite bands, without overlap.
- Data from 5 West coast estuaries over 23 years and 271 months were analyzed.
- SI data considers the entire region of interest, instead of individual sampling stations.
- In the absence of strong Linear correlation, a novel method of Fourier transform analysis was performed.
- A low cost, powerful tool that offers ability to analyze actual raster images.
- The ML Model with delay improves to R^2 between 0.961 to 0.99 for all critical parameters.
- It is fast and delivers predictive analysis within 5 minutes once trained.

X. FUTURE WORK

This present study combines in situ data with satellite data, offering a powerful tool that allows for monitoring, remediation efforts, historic perspectives, and a better understanding of wetland processes.

Future work aims to build an app that correlates data between Satellite Indices and in situ data of any wetland using anomalies and climatologies.

A further step is to analyze correlations between band data and in situ data to provide insight into how climate change is affecting wetlands.

REFERENCES

[1] EPA "basic information about estuaries". https://www.epa.gov/nep/basic-information-about-estuaries. Accessed: 2010-09-30.

[2] J. B. Zedler. Ecological issues in wetland mitigation: an introduction to the forum. *Ecological Applications*, 6(1):33–37, 1996.

[3] D. E. Porter, G. Osburn-Matthews, W. Jefferson, and F. Taylor. Development and implementation of the nerrs centralized data management office. 1995.

[4] WEF "why wetlands are our best climate hack". https://www.weforum.org/agenda/2022/03/wetlands-climate-change/. Accessed: 2010-09-30.

[5] M. E. Kentula and S. G. Paulsen. The 2011 national wetland condition assessment: overview and an invitation. *Environmental Monitoring and Assessment*, 191(1):1–18, 2019.

[6] N. P. Reed and D. Drabelle. *The US Fish And Wildlife Service*. Routledge, 2019.

[7] L. M. Cowardin and F. C. Golet. Us fish and wildlife service 1979 wetland classification: A review. *Classification and inventory of the world's wetlands*, pages 139–152, 1995.

[8] E. D. Stein, A. E. Fetscher, R. P. Clark, A. Wiskind, J. L. Grenier, M. Sutula, J. N. Collins, and C. Grosso. Validation of a wetland rapid assessment method: use of epa's level 1-2-3 framework for method testing and refinement. *Wetlands*, 29(2):648–665, 2009.

[9] W. N. Heady, R. P. Clark, K. O'Connor, C. Clark, C. Endris, S. Ryan, and S. Stoner-Duncan. Assessing california's bar-built estuaries using the california rapid assessment method. *Ecological Indicators*, 58:300–310, 2015.

[10] M. Guo, J. Li, C. Sheng, J. Xu, and L. Wu. A review of wetland remote sensing. *Sensors*, 17(4):777, 2017.

[11] C. Humborg, D. J. Conley, L. Rahm, F. Wulff, A. Cociasu, and V. Ittekkot. Silicon retention in river basins: far-reaching effects on biogeochemistry and aquatic food webs in coastal marine environments. *AMBIO: A Journal of the Human Environment*, 29(1):45–50, 2000.

[12] X.-e. Yang, X. Wu, H.-l. Hao, and Z.-l. He. Mechanisms and assessment of water eutrophication. *Journal of zhejiang university Science B*, 9(3):197–209, 2008.

[13] J. A. Richards. Spatial domain image transforms. In *Remote Sensing Digital Image Analysis*, pages 203–246. Springer, 2013.

[14] R. Albarakat and V. Lakshmi. Comparison of normalized difference vegetation index derived from landsat, modis, and avhrr for the mesopotamian marshes between 2002 and 2018. *Remote sensing*, 11(10):1245, 2019.

[15] P. B. Moyle, J. R. Lund, W. A. Bennett, and W. E. Fleenor. Habitat variability and complexity in the upper san francisco estuary. *San Francisco Estuary and Watershed Science*, 8(3), 2010.

[16] N. Saintilan, K. Rogers, C. Toms, E. D. Stein, and D. Jacobs. Intermittent estuaries: linking hydro-geomorphic context to climate change resilience. *Journal of Coastal Research*, (75 (10075)):133–137, 2016.

[17] B. Mahesh. Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet]*, 9:381–386, 2020.

[18] T. Chai and R. R. Draxler. Root mean square error (rmse) or mean absolute error (mae). *Geoscientific Model Development Discussions*, 7(1):1525–1534, 2014.