DOE SBIR Phase I

Final Report

Performing Organization
Flying Gybe Inc
Portland, Oregon
30 March 2022

Authors
Nicholas Tufillaro, Adam Belmonte

Report Period
06/29/2020-12/30/2021
No cost extension due to limited site access because of Covid restrictions

SBIR Phase I Project Title
Correlation of Remote Sensing Water Quality Maps
with an *in situ* Sensor Network

Funding Office
Office of Science, U.S. Department of Energy

Work Performed under DOE SBIR Grant
Number DE-SC0020843

Abstract

The application of high spatial (~10 m) remote sensing satellites, such as Sentinel-2, to detect water quality issues faces two key technical challenges: (1) the accurate retrieval of water leaving radiances from satellite imagery (*i.e.* the atmospheric correction, surface reflectance and related problems), and (2) extending remote sensing water quality product maps beyond current quantities such as turbidity to new quantities such as concentrations of nutrients (e.g. nitrates, phosphorus) and indicators of toxins (phycocyanin, an indicator of harmful algal blooms) in waterways. Gybe is developing solutions to both of these technical challenges by using an *in situ* spectral sensor network that operates on the ground in tandem with satellite remote sensing imagery.

The Phase I proposal focused on problem (2) - the generation of remote sensing water quality maps for nutrients (specifically, nitrates), and indicators of harmful algal blooms (specifically, phycocyanin). Sentinel-2 is unable to measure either of these quantities directly because of fundamental physical limitations. In both cases, the Sentinel-2 does not have spectral bands targeted at the absorption features of either of these constituents in water - for nitrates it is spectral features in the UV that the satellite water imagers cannot detect, and for phycocyanin the absorption band (~630 nm) is not in the band set of the satellite's multispectral imager. In this SBIR Phase I we provided a demonstration, based on field trials, of how to measure both these quantities by fusing data from Gybe's *in situ* hyper-spectral sensor with Sentinel-2 imagery. The demonstration sites were on the Kansas River at Desoto, KS, and the San Luis Reservoir near Los Banos, CA. Both multivariate adaptive regression splines and convolutional neural nets were used to create surrogate data models allowing for extended water quality product retrievals for both phycocyanin and nitrates. The methodology can be applied to similar nutrients (*e.g.* phosphorous) and pigments.

These new remote sensing data products provide spatial awareness of water quality throughout a reservoir, river, or watershed, enabling water managers to make better data driven decisions for flow plans and operations that balance metrics for power generation, water needs, and ecological sustainability, *i.e.*, Sustainable or Green Hydropower.

Table of Contents

Abstract	2
Table of Contents	3
Summary	4
Introduction	7
Phase I Technical Objectives	7
Phase I Accomplishments	8
Details of results from work plan	12
Task 1: Site planning and sensor installation	12
(a) Nutrients - Desoto, Kansas Site	12
(b) Harmful Algal Blooms - San Luis Reservoir, California Site	13
Task 2: Data flow, processing, and modeling framework	15
Task 3: Model development and validation	19
(a) Nitrate retrievals	19
(b) Phycocyanin retrievals	22
Conclusion	27
References	28
Acknowledgments	31

Summary

DOE's 2016 Hydropower VISION report highlights the need to operate hydropower in a ecologically sustainable manner:

"*Ensure that hydropower's contributions toward meeting the nation's energy needs are consistent with the objectives of environmental stewardship and water use management.*"

To develop business opportunities in support of this goal, section 13(b) "Innovative Sensing and Data Platforms for Water and Hydropower" of the FY2020 DOE SBIR request from the Water Power Technologies Office (WPTO) calls for "Innovate Sensing and Data Platforms for Water and Hydropower" which can include both remote sensing data sets and new *in situ* sensor networks that "enhance the performance and value of hydropower systems" with respect to multiple metrics including "ecological, operational, safety, and hydrological," that is, flow and water quality.

This goal is well aligned with Gybe's goal to develop hydropower decision support tools based on real-time, actionable, data streams providing a comprehensive view of the flow and water quality state of a reservoir, river system, or watershed,

Gybe's goal is to assist with the sustainable use of water resources by closing the loop between on the ground operations and measurable outcomes, visible at the watershed scale using satellite imagery.

The application of remote sensing satellites, such as Landsat-8 and Sentinel-2, to water quality issues faces two key technical challenges: (1) the accurate retrieval of water leaving radiances from satellite imagery (i.e. the atmospheric correction, surface reflectance and related problems), and (2) extending remote sensing water quality product maps beyond current quantities such as turbidity to new quantities such as concentrations of nutrients (e.g. nitrates, phosphorus) and indicators of toxins (phycocyanin, and indicator of harmful algal blooms) in waterways. Gybe is developing solutions to both of these technical challenges by the use of an *in situ* sensor network that Gybe has created which provides a complete spectral signature of a specific location within a satellite image, and then propagates this information to the whole remote sensing scene.

The Phase I proposal was focused on problem (2) - the generation of remote sensing water quality maps for nutrients (specifically, nitrates), and indicators of harmful algal blooms (specifically, phycocyanin).

Currently, Landsat-8 or Sentinel-2 are unable to measure either of these quantities directly because of fundamental physical limitations. In both cases, the current sensors do not have spectral bands targeted at the absorption features of either of these products - for nitrates it is spectral features in the UV that the satellite imagers cannot detect, and for phycocyanin the absorption band (~630 nm) is not in the band set of the satellite multispectral imagers. In this SBIR Phase I we provided a demonstration, based on field trials, of how to measure both of these quantities by fusing data from Gybe's *in situ* spectral sensor with Sentinel-2 imagery.

In collaboration with USGS, Gybe installed its sensor network at the De Soto, Kansas USGS Super Gage on the Kansas River, downstream from power operations in Lawrence, KS, by Bowersock Hydropower and several USACE reservoir operations. The Kansas river is a conduit for nitrates from the midwest traveling eventually to the Mississippi River, and the Gulf of Mexico. It experiences a high range of nitrate levels from 0.1 to 10 mg/L, usually peaking during spring when fertilizer is applied to fields after spring plowing. The Gybe sensor was installed during 2021 and provided the data record for algorithm development and validation.

As a first step to creating a model for nitrates concentrations we build a convolutional neural network that correlates the target product (nitrate concentration) to satellite measurable quantities (turbidity, chlorophyll-a concentration) using only *in situ.* data. As mentioned, nitrates are not directly measurable by satellites, but turbidity and chlorophyll-a concentrations are measurable from Sentinel-2 imagery. Therefore, we first built models using machine learning (specifically splines, and neural networks) between these *in situ* data sets to discover the parameters that are measurable and best correlate with nitrates. The result of using neural networks trained on data from 2016-2020 at the De Soto showed excellent correlation for in sample testing. We next validated the model with an out of sample test data set during 2021. The excellent out of sample agreement demonstrates the feasibility of the method given a sufficient training set. To next test the method with Sentinel-2 remote sensing data we found all corresponding satellite matches in the 2021 time frame and implemented a similar model training run. The model for the satellite data shows good correlation for the main nitrate events - spikes in nitrate values. The results are expected to improve as the data set grows. The USGS only data set contains ~ 100,000 instances, and the remote sensing data set contains ~ 70 instances. The product is improved when an algorithm using data fusion from the Gybe sensor is incorporated - here the product is based not just on turbidity and chlorophyll-a, but an optimization over the primary spectral bands - that is all the spectral data available. We thus met the first technical objective, showing how to estimate and map nitrate concentrations from Sentinel-2 imagery with a field trial along the nutrient rich Kansas river.

For phycocyanin, the site selected was the San Luis Reservoir, the main storage reservoir along the canal of the Central Valley Project. San Luis reservoir experiences seasonal harmful algal blooms (HABs) due to its warm weather and abundant sunshine in the California central valley. Phycocyanin concentration varies from 0 to greater than 50 mg/L throughout the year. The Gybe sensor was installed on the pier of the intake pipe, in collaboration with the California Department of Water Resources (CA DWR), and has been collecting data since early 2021.

For phycocyanin our target product does have a signature in a visible band (~ 630 nm), but this band is not covered by the Sentinel-2 or Landsat-8 multispectral imaging sensor band sets. Therefore, we used the Gybe sensors' complete UV-VIS-NIR band set to detect the phycocyanin, and then used a linear correlation between bands to create an empirical model using only the bands available on Sentinel-2. Training data validation data were both collected during 2021, and band correlation between the Gybe sensors ~ 630 nm band and Sentinel-2's visible bands products showed a good correspondence to the variation in the 630nm phycocyanin absorption band, though matches in the data set were a very limited sample size.

We emphasize that both methods (for nitrates and phycocyanin) are empirical and site specific. They require the availability of contemporaneous *in situ* data to train the models, and propagate those calibrations across the entire image region as the satellite passes overhead. The spatial extent of the range of the validation of the models is still a topic of research and is presumably limited by changes with distance of atmospheric state and variations in water constituents due to extraneous inputs.

The Phase 1 work demonstrates the technical feasibility of estimating previously invisible water quality parameters from operational multispectral sensing satellites, namely phycocyanin and nitrates from Sentinel-2 imagery when fused with *in situ* sensor measurements. Methods were validated against field data from the USGS and CA DWR. The Phase 1 project opens the way to a basin-wide retrial of water quality maps which is spatially dense and unlocks new information from the growing archive of earth observation satellite images acquired daily. Applications include: planning of grab samples, understanding of water/land interactions and processes, identification of point and nonpoint sources of water pollution, the implementation of flow plans to balance multiple power and ecological metrics, and the evaluation of water ecology restoration projects.

Introduction

Water quality parameters such as turbidity and chlorophyll-a are currently being produced from satellite imagery such as Landsat-8 and Sentinel-2 which have a 30 meter and 10 meter spatial resolution respectively. Given their wide spatial coverage, these new water quality map services complement ground based point sensor measurements by providing a watershed wide view of water quality and its connection to land processes. There are many water quality parameters such as nutrients and other pigments associated with harmful algal blooms (HAB's) which are not accessible with current remote sensing satellites. For nutrients this is because the nutrients have a spectral signature in the UV, and not the visible part of the spectrum, and for HABs, the pigments spectral signature is not part of the limited band coverage of the satellite imagers.

In the Phase 1 SBIR we proposed and demonstrated that it is possible to extend the products available from Sentinel-2 by augmenting its imagery with on the ground, contemporaneous, spectral data from a spectral radiometer developed by Gybe that covers the full visible spectrum. Specifically we chose target products of nitrates and phycocyanin since these are water quality products of high public concern. Nitrates because it is a principal driver of eutrophication and subsequent dead zones in water bodies, and phycocyanin because it is often associated with HAB's.

Phase I Technical Objectives

The overall technical objective is to *demonstrate the retrieval of water quality maps for quantities that are not directly measurable by operational high spatial resolution multispectral imagers such as Sentinel-2.* We accomplished this by fusing the multispectral imagery with a ground based sensor network collecting contemporaneous data for a specific region of interest (ROI). Specifically, the target products are nitrates and phycocyanin concentrations, both of which have absorption features that are outside of the band set for Sentinel-2 and Landsat-8.

We accomplished this goal in three steps:

(1) identification of appropriate sites to run trials for data collection;

(2) installation of the Gybe sensor system and data collection and processing over seasons where the target products vary greatly, and

(3) the development of new algorithms for data fusions between the Gybe sensor and Sentinel-2 which are validated against field data.

Phase I Accomplishments

For nitrates the site selected was the USGS Super Gage at on the Kansas river in De Soto, KS, east of Lawrence, KS (Fig 1). This site is centrally located in a midwest agriculture region and has hydropower generation at Lawrence, KS. It is a conduit for nitrates from the midwest traveling eventually to the Mississippi River, and the Gulf of Mexico. It experiences a high range of nitrate levels from 0.1 to 10 mg/L, usually peaking during spring when fertilizer is first applied to plowed agricultural fields. The Gybe sensor, in collaboration with the USGS, was installed during 2021 and provided the data record for algorithm development and validation.

Figure 1. (Left) Gybe sensor installed above Kansas River at Desoto, KS colocated with USGS super-gage site measuring nitrates. (Right) Sentinel 2 image from 10 July 2021 of Kansas River above Gybe sensor.

For phycocyanin, the site selected was the San Luis Reservoir, the main storage reservoir along the canal of the Central Valley Project near Los Banos, CA (Fig 2). San Luis reservoir experiences seasonal HAB's due to its warm weather and abundant sunshine. Phycocyanin concentration varies from 0 to greater than 30 mg/L throughout the year. The Gybe sensor was installed on the pier of the intake pipe, in collaboration with CA DWR, and has been collecting data since early 2021.

New algorithms were developed and tested for both nitrates and phycocyanin. The basis for both algorithms is what the USGS calls the method of 'surrogate data' [1]. The method is empirical and uses numerical models to correlate the target product (e.g. nitrate concentration) to measurable quantities (eg. turbidity, chlorophyll-a concentration). As mentioned, nitrates are not directly measurable by satellites, but turbidity and chlorophyll-a concentrations are measurable.

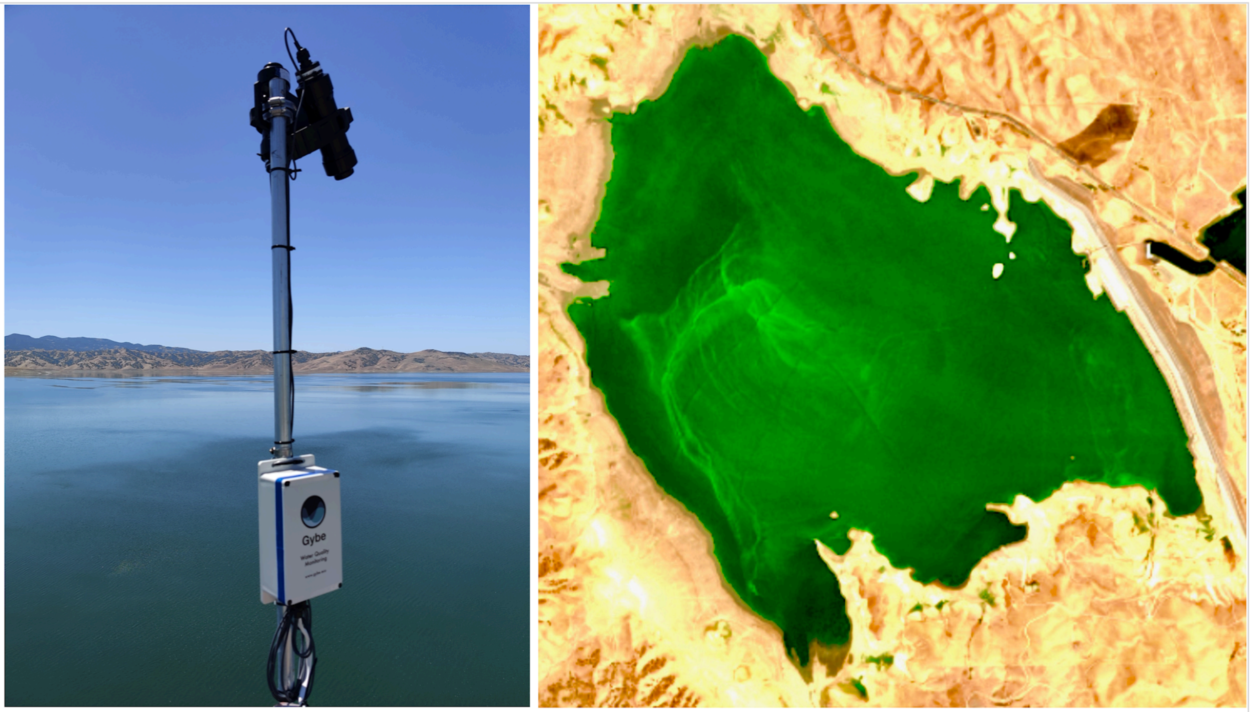

Figure 2. (Left) Gybe sensor installed at San Luis reservoir near Los Banos, California. (Right) Sentinel-2 image showing a Harmful Algal Bloom on 27 August 2021.

Therefore, in our model development we first built models using machine learning methods (specifically splines, and neural nets) between *in situ* data sets to demonstrate the feasibility of the surrogate method for the target products. The result of using neural nets trained on data from 2016-2020 at the Desoto site is shown in Fig 3.

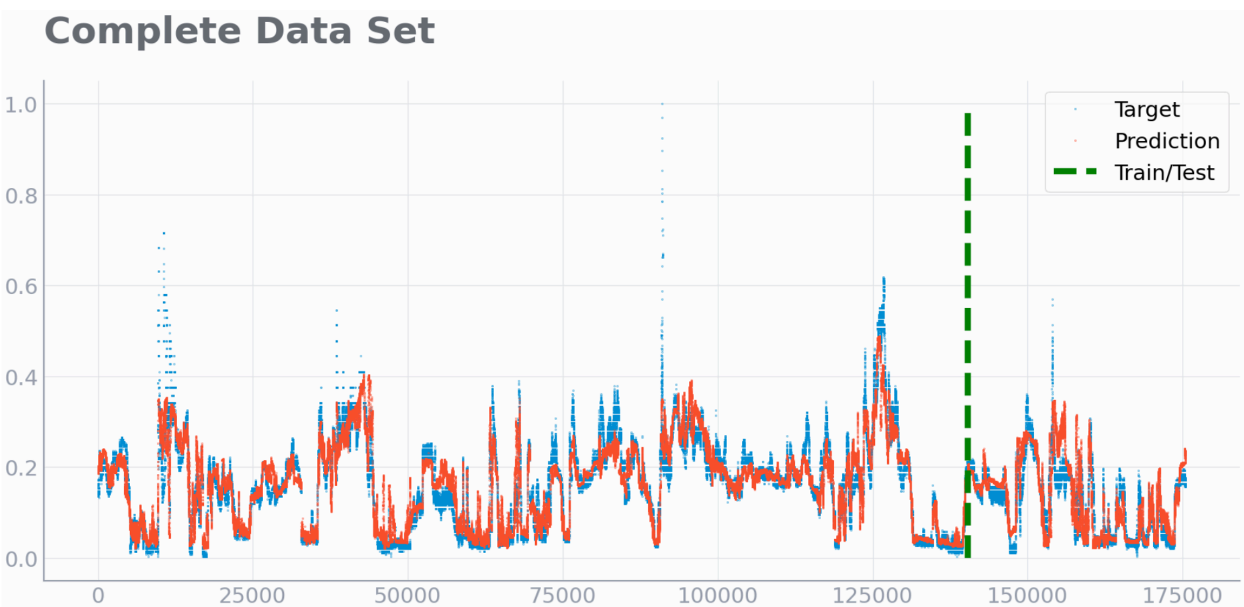

Figure 3. Nitrate values (normalized to 1) at USGS Desoto Gage from 2016-2020 used for training (left of green line) a neural net surrogate data model for nitrates based only on inputs from chlorophyll and turbidity. Out of sample test of validation (right of the green line) in of surrogate data model using data from 2021.

The test set was the out of sample data set shown during 2021. The good out of sample agreement demonstrates the feasibility of the method with an ample training set. To extend the method to Sentinel-2 remote sensing data we found all corresponding satellite matches in the same time frame and implemented a similar model training run. The model for the satellite data shows good correlation for the main nitrate events - spikes in nitrate values (Fig 4.).

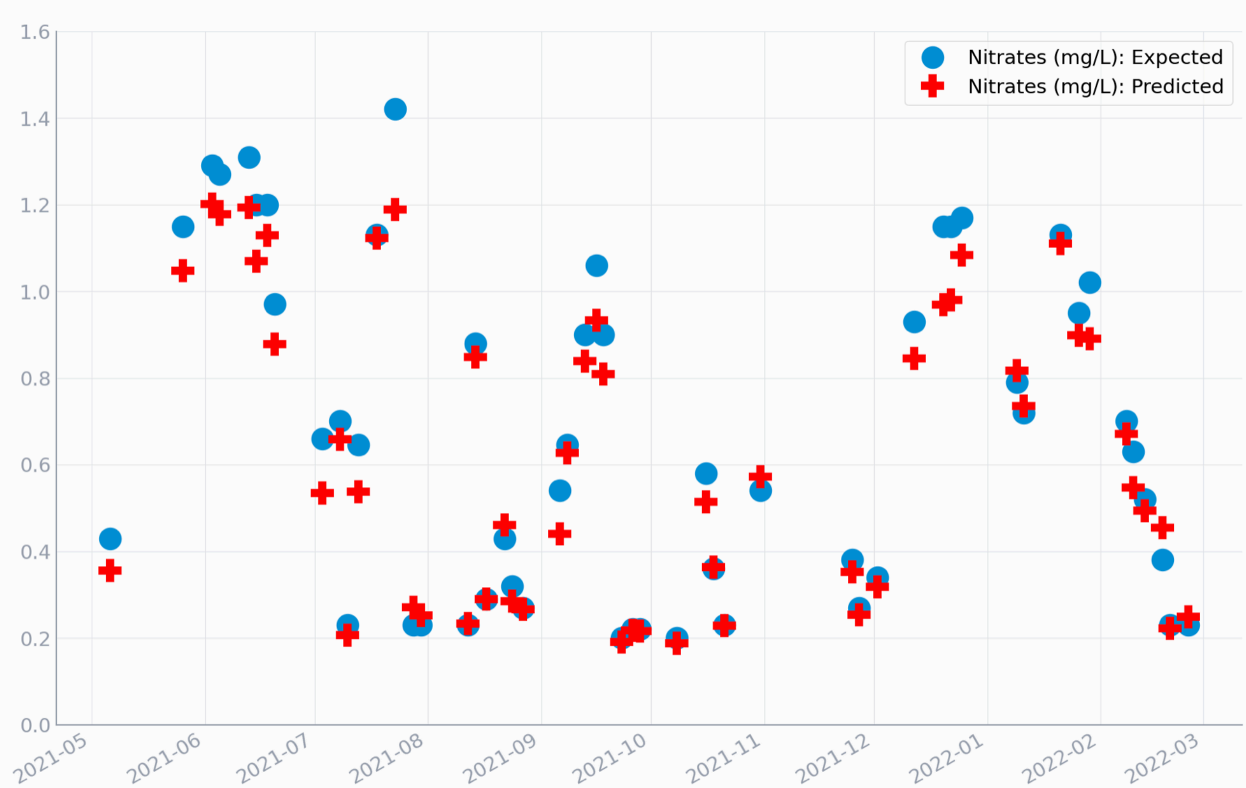

Figure 4. Surrogate model of nitrates using only Sentinel-2 Turbidity and Chlorophyll remote sensing products. Data from Sentinel-2 clear overpasses during 2021.

The results are expected to improve as the data set grows. The USGS only data set is ~ 100,000 instances, and the remote sensing data set is ~ 70 instances. The product is improved when algorithms using data fusion from the Gybe sensor are incorporated - here the product is based not just on turbidity and chlorophyll-a, but an optimization over the primary spectral bands - that is all the spectral data available [3].

For phycocyanin our target product does have a signature in a visible band (~630 nm), but this band is not covered by the Sentinel-2 or Landsat-8 sensors. Therefore, we used the Gybe sensor to detect the phycocyanin and then used empirical correlation between bands (Fig. 5) to create an empirical model using the bands on Sentinel-2. Sample images showing both green and blue-green algae blooms are shown in Figure 6.

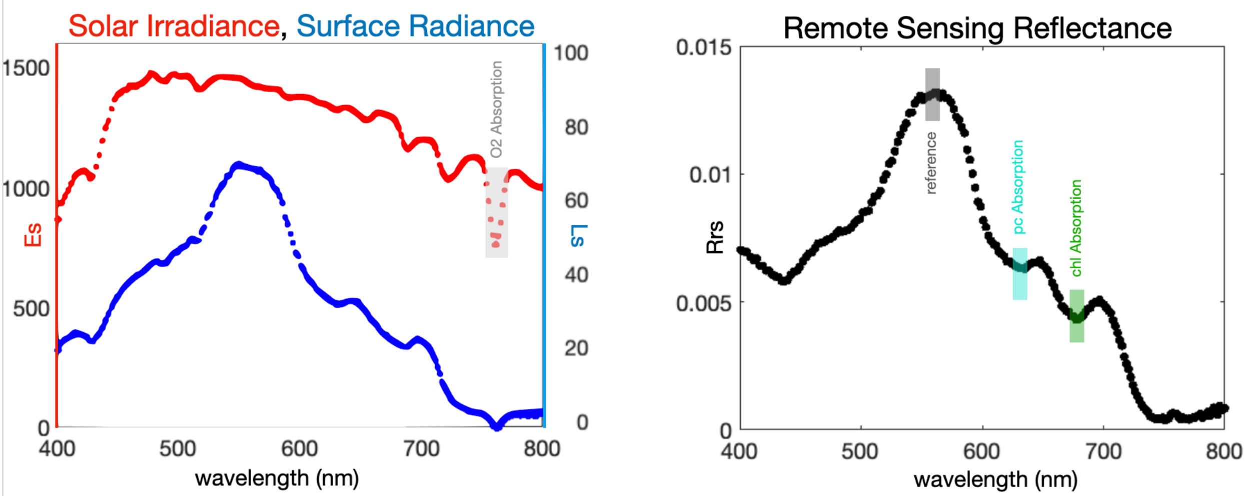

Figure 5. Spectra from Gybe Sensor mounted at San Luis Resevoir at noon on 20 September 2021. Left is the primary spectra, Solar irradiance (Es, downwelling mW/(m^2 x nm)), and Surface radiance (Ls, upwelling). Right is the ratio (after corrections for sky reflection) termed **Remote Sensing Reflectance (Rrs)**. An absorption band for Chlorophyll and Phycocyanin are indicated.

The Phase 1 work demonstrates the technical feasibility for estimating new methods and products for water quality parameters from remote sensing satellites, namely phycocyanin and nitrates. The methods are validated against field data from the USGS and CA DWR. The Phase 1 project opens the way to a basin wide retrial of water quality maps which is spatially dense. Data of this type has many applications including: planning of grab samples, understanding of water/land interactions and processes, identification of point and nonpoint sources of water pollution, and the evaluation of water ecology restoration projects.

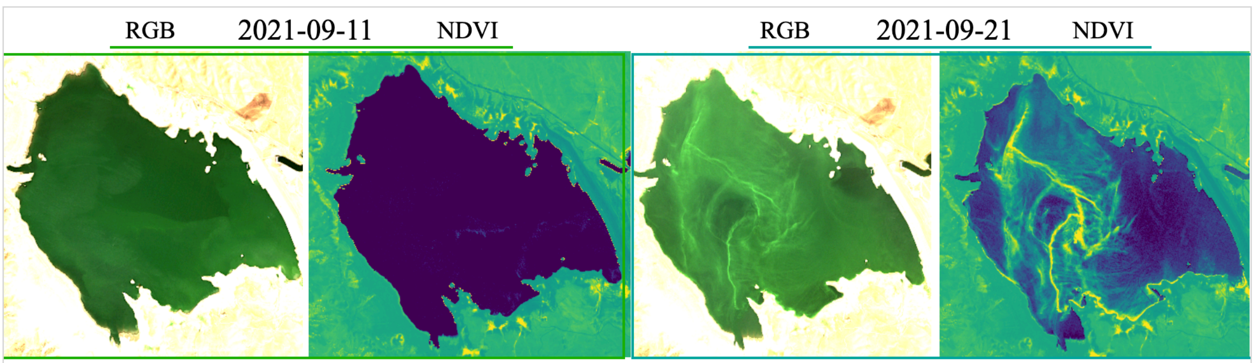

Figure 6. Sentinel-2 images from 2021-09-11. Left a bloom that looks like it is dominated by 'green' algae earlier in September switches over to a bloom dominated on the surface by 'blue-green' algae later in September 2021. RGB images are brighten, and the NDVI highlights the surface algae floating on the surface.

Details of results from work plan

Task 1: Site planning and sensor installation

High nutrient levels in source waters can lead to water quality issues ranging from algal blooms locally, to hypoxic conditions globally. A good example of this is the Kansas river system which locally has high nitrate levels from leaching of fertilizers into the Kansas river causing, in some instances, HAB's. More globally the Kansas River feeds into the Missouri River and finally into the mouth of the Mississippi River contributing nutrients that have led to an oceanic 'dead zone.' The Kansas River experiences a wide range of nitrate levels, typically peaking in the spring after plowing and the first application of fertilizers to agricultural fields. The USGS also maintains a Super Gage which makes continuous measurements of nitrate levels [4]. This site was chosen for our nitrate study since it met our planning criterium for continuous measurements of nitrates, colocation with a USGS Super Gage, good visibly by Sentinel-2, and an engaged collection of stakeholders that includes a private hydropower operator upstream in Lawrence, KS. Our original proposal described the use of a site on the Willamette River in Oregon, however, further discussions with the Portland, OR office of the USGS made this site less attractive because the nitrate variation was not as wide as the Kansas site, and the Portland USGS sensor location did not provide good visibility to the river (downwelling light was shadowed by a bridge structure).

For our HAB pigment study the San Luis Reservoir near Los Banos, California was selected [5]. The site is the largest reservoir on the Central Valley Project and experiences a regular pattern of HAB's due to the warm weather and abundant sunshine. Dam operations are managed by California Department of Water Resources (CA DWR).

We had limited access to both sites at the start of the project due to the COVID-19 outbreak which led to a delay of sensor installation plans. Sensor installations were originally planned for the Fall of 2020, but we were only able to get permitting and access to both sites in the Spring of 2021. This led to a no cost project extension from July 2021 to December 2021 so that we could collect and analyze data from the Fall season.

(a) Nutrients - Desoto, Kansas Site

The USGS De Soto, Kansas Gage (site number: 06892350) is located at 38.9822 N by -94.0648 W at the De Soto bridge crossing of the Kansas River on Kansas Hwy 2. The site is a USGS 'Super Gage' and includes sensors for water quality pigments including chlorophyll-a (chl-a), phycocyanin (pc), turbidity, and nitrates (NO2 + NO3) [6]. A continuous record (15 minute sampling) for these parameters dates back at least to 2015, though short and long term gaps exist in individual sensor records because of sensor impairments, or the failure of data to pass USGS's strict quality control protocols.

With the assistance of USGS Lawrence, Kansas office we installed the Gybe sensor (Fig. 1) on the bridge above the USGS's gage. The Gybe sensor is a radiometer system which measures downwelling solar irradiance and upwelling water radiance to facilitate more reliable match ups with remote sensing satellites, in particular Sentinel-2 image records. The Gybe sensor measures light from ~400-800 nm with ~10 spectral resolution every 5 seconds.

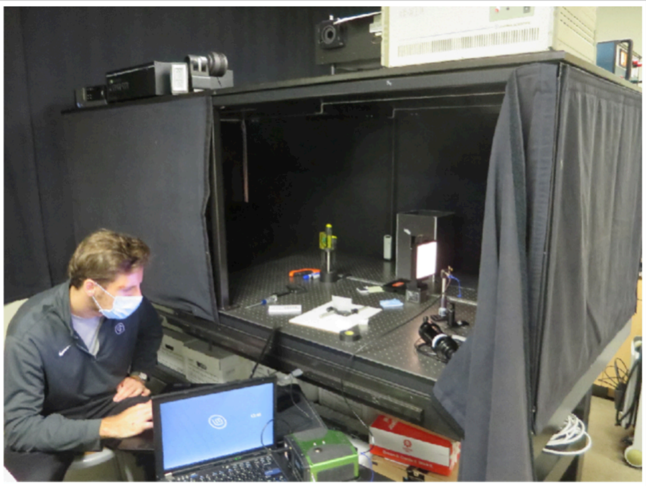

Figure 7. Undergraduate intern, Clemente Borgogni from Georgetown University calibrating Gybe sensors for DOE project during the Fall 2020.

Actual sampling integration times range from 100 microseconds to 2 seconds, and data is averaged to get an aggregated spectrum.

The integration time varies automatically with light conditions. Achieving reproducible results with the Gybe sensor requires radiometric calibration. As part of this project we hired a student intern to assist with both the sensor construction, sensor calibration, and field installation (Fig. 7). The Desoto site location is on a bridge that faces north-south.

This allowed us to point the sensor on the east side of the bridge, toward the middle of the Kansas river at an angle of approximately 45 deg NE from North. The best data retrievals are at 135 deg from the solar azimuth to minimize sun glint from the water surface [7,8]. The sensor is also pointing about 45 deg up from the water surface. Both view angles are chosen so that the sensor is approximately in alignment with the NASA above water ocean color radiometric field measurement protocols for viewing when the solar angle is at high noon (Fig. 1).

(b) Harmful Algal Blooms - San Luis Reservoir, California Site

The San Luis Reservoir sensor is located on a large concrete structure above the intake pipe for water operations (. 8). The sensor is located at 37.0674 N, -121.0854 W on the South West corner of the structure and the sensor faces approximately due West (solar noon azimuth 90 deg), looking at the water surface at 45 deg from vertical. The structure is typically 10 meters (depending on water level) above the water surface and 150 meters from the shore, still there can be some shadowing on the water surface before noon.

The Gybe sensor was installed in the spring 2021 after access and permitting was granted by CA DWR and CA Parks. There is a regular water sampling program including HAB toxins at

Figure 8. Installation of Gybe sensor at San Luis Reservoir, California by project scientist Adam Belmonte (2021).

several locations in the reservoir. The most commonly reported toxin is microcystin. Also tested for are cylindrospermopsin, saxitoxin, and anatoxin-a.

The most extensive testing is at Basalt Road boat ramp in the southern part of the reservoir. Previous work has identified microcystis as the the typical source species for the toxins, which forms long colonial strands near the surface in the day which provides an easily imaged feature in remote sensing imagery as seen in Fig 2. [9]. Drought conditions have been common in California in recent years and 2021 was no exception. Danger levels for HABs were issued for much of August and September of 2021 (Fig 9).

Figure 9. Television (KFSN) news broadcast from from 7 September 2021 announcing dangerous level of toxic blue-green algae in San Luis Reservoir, California.

Task 2: Data flow, processing, and modeling framework

Gybe sensor data is streamed via cellular connection to our data servers. The Gybe raw sensor data is organized into a Level 0 (L0) data file to which lab calibrations are applied which convert digital numbers (DN's) to Level 1 (L1) radiances (upwelling) or irradiance (downwelling). The Rrs values in this report use lab calibrations for radiances and irradiances, our operational codes for Remote sensing reflectances (Rrs), Level 2a (L2a) processing use the method described by Grötsch et. al. [2]. These Rrs values are then used to compute Level 2b (L2b) water products. Standard satellite band methods for chlorophyll-a, turbidity and other quantities are used. A range of standard algorithms were tested from OC5 to Mishra; similarly, for turbidity the Dogliotii band algorithm was used [10,11]. For both sites, these algorithms are 'tuned' by a linear regression to the model coefficients with any ancillary field data or available gauge data. The algorithms used were chosen to be identical to the algorithms for Level 2 product generation used by Acolite [12,13]. An alternative method we are developing uses a nonlinear optimization method recently described by Grötsch which uses all the bands [2]. This spectral optimization method is seeded by a range of water parameters which we estimate for each site, and proceeds to find the best fitting spectrum built from a fixed bio-optical model of the water [14].

On the remote sensing side, Sentinel-2 imagery was downloaded from European Space Agencies' (ESA) Copernicus Hub Server. Both a L1 and L2 product are available from the ESA. Sentinel-2 has a nominal 10-20 meters spatial resolution in the visible spectrum, and bands B1-B8 are centered at 443, 490, 560, 665, 705, 740, 783, 842, 865 nm. We worked with the L1 ESA data (top of atmosphere radiances) and produced L2 products (Rrs, and water quality parameters) using the open source program Acolite development by Vanhellemont [15]. References for all the band water quality products and their associated band algorithms are described in the Acolite Manual, and they were the same band algorithms used by the Gybe sensor (up to parameter tuning) [10,11,12].

For model development, we first tested different types of models using the full data sets from USGS, and then selected the best models for further development for product generation from remote sensing data. For readability, we will describe the general framework for our model generation before describing more details about how we processed the ancillary USGS or DWR data for the target products of nitrate and phycocyanin.

Surrogate modeling is a term used by the USGS to describe the estimation of a water quality parameter, say phosphorus concentration, by modeling its correlation to other water parameters, such as turbidity and discharge [1,16]. The utility of the model is that the target parameter is usually more difficult or expensive to measure than source parameters. For instance, turbidity gauges are both less expensive and more reliable than nitrate gauges.

Specific surrogate models are empirical and site specific. Surrogate models are common in many fields of engineering and go under numerous names. In machine learning they are called tabular models [17], in electrical engineering behavioral models [18], and in time series analysis Multiple Input, Single Output (MISO) models. In all cases, the mathematical problem is similar, the estimation of a target variable,

$$y = f(x_1, x_2, \ldots, x_n)$$

in terms of multiple source variables $x_1, x_2, x_3, \ldots$, given a discrete sample set for both input and output variables. The above problem is referred to as 'static' since it has no explicit temporal dependence. In the examples described in this report we also assume that the models are continuous. The target product model is a surface (or more generally manifold [19]) which can be estimated by explicitly formulating the underlying assumptions in the model, and the uncertainty in the data - that is with a statistical framework.

In contrast, A 'dynamic' model includes explicit temporal dependence. The system is described by a state vector that traces out a continuous trajectory in a 'state space.' These are commonly said to be systems 'with memory.' In one formulation they are modeled by fitting a vector field,

$$\frac{dy}{dt} = f(x_1(t), x_2(t), x_3(t), \ldots)$$

and model predictions are obtained by the integration of the model, and not just by functional evaluation, as with 'static' models. Examples of dynamic models are time series models with feedback (NARMAX), or recurrent neural networks. The discussion above is very informal: under suitable mathematical assumptions, dynamic models can be translated to static models, and continuous models parametrized by t can be reformulated in terms of a discrete delay parameter, τ. The above discussion is simply meant to help orient the reader about the different approaches available for data driven modeling.

In this report we describe the development of surrogate models for water quality from remote sensing data such as that provided by Sentinel-2 [20]. Most water quality applications of surrogate models correlate water samples collected in the field (and measured in the lab) to *in situ* real time gauges. For instance, correlating data from turbidity gauges to weekly nitrate estimations from water grab samples [1]. Since the source variables are continuous measurements (e.g. 15 minute turbidity measurements from USGS gauges) it is possible to consider either static or dynamic models. However, if the underlying data for the source variables are not sampled well enough to be rendered as a continuous sequence - as is the case with many satellite observations which have revisit times of order of days and not minutes - then we are constrained to the use of static models.

We discuss extensions of surrogate models for water quality data by moving from linear modeling functions to nonlinear modeling functions, specifically splines and neural networks, and discussing the functional estimation problem of static modeling from a set theoretic perspective, with what is called 'manifold learning.' The perspective aids understanding of underlying physical processes as well as providing guidance on creating nonlinear surrogate models that complement the statistical guidelines described in several USGS reports [21]. The described modeling considerations are helpful for extending surrogate modeling to inputs that include remote sensing data which is capable of providing data of greater spatial coverage than gauge data alone.

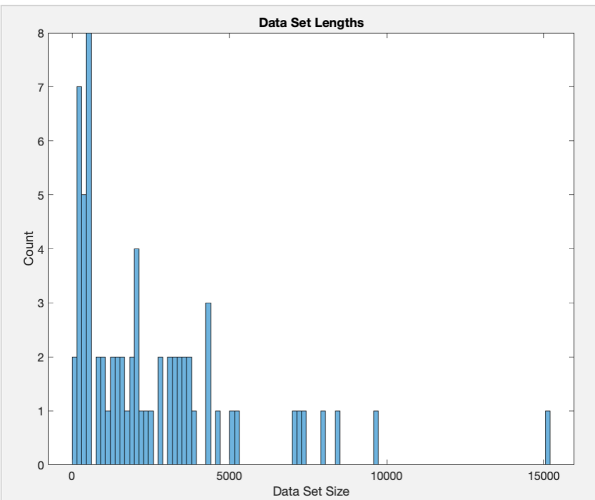

Figure 10. Histogram showing size of continuous (15 minute sampling interval) data sets of turbidity, chlorophyll, and nitrates at Desoto, KS from 2016-2021 breaking data sets with gaps greater than 3 hours and minimum of 1 day (96 samples).

To estimate models of nitrate from possible remote sensing quantities, we first start with the USGS and curate and organize the time series data into contiguous records with no gaps. We start with all qualified data from the USGS portal from 2016 (the start of the Sentinel-2 record) till the end of 2021 [6]. Next we specify a collection of input and output variables and break the initial USGS data (with gaps) into smaller records that are contiguous and have no time gaps. All data gaps less than 3 hours or less (12 samples at a 15 minute sampling interval) are interpolated to provide continuous record. If a gap is longer than 3 hours, then the data set is split, and a new contiguous data set is begun.

Figure 10 shows a histogram of the sequence of contiguous data sets that are found when two input variables are selected - turbidity and chlorophyll-a - and the output is nitrates. Sequences less than a minimum length are also excluded from modeling, typically 7 days (or sequences less than 672 points). Continuous data sets are desirable when forming state space models and for creating ancillary variables (such as derivatives) which can be used for filtering. These subsets are then examined for any anomalies or outliners.

The first artifact noticed in the data records is digitization noise. Data reported by USGS typically contains 3 significant digits. When plotted there are small visible jumps in the data which are artifacts of the effective digitization of the supplied data. Therefore, the next step is data smoothing.

We used a moving mean filter, typically with a window length of 12 hours (48 points). Though not utilized here, we also experiment with polynomial smoothing filters, such as a Savitzky-Golay (SG) filter, which permit the estimating of smooth derivatives. We note that a moving average filter creates a linear correlation of points within the averaging window which may, or may not, be in the original data set. Effectively, the moving average filter acts as a low pass filter, in this case minimizing signal oscillations below 6 hours (see Fig 11).

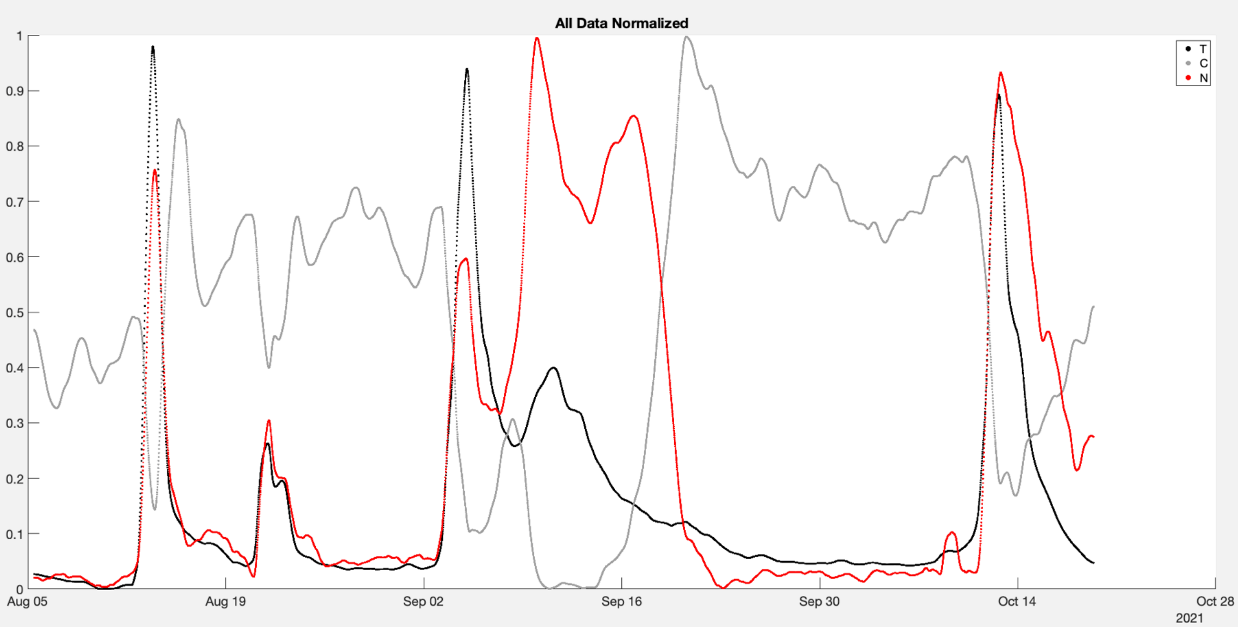

Figure 11. Smoothed (moving mean) and normalized data from Desoto, Kansas during fall 2021. Black- Turbidity, Gray - Chlorophyll, Red - Nitrates. Spikes in data are run off after storms.

The subsequences are then collected into a single training set. The training set is curated to provide representative samples of each nitrate event -- a spike in the nitrate values. In machine learning this is called 'instance selection,' and can help with both training speed and model accuracy [22]. As an example, long sequences of near zero values could be removed to shorten the training set. In effect this step reweighs data values and can also be accomplished in the functional estimation step by adjusting the weighting functions. It is not an essential step for this data set and we did little or no instance selection in building our training data sets.

The next standard step before function fitting is data normalization to aid with numerical estimations. The quantities in the USGS time series are all inherently positive. So instead of centering on the mean, we normalize the data records to a magnitude of one with

$$Nx = (x - \min(x))/(\max(x) - \min(x))$$

This transformation is reversible, so no information is lost in the final prediction of products in physical units. The normalizations are chosen over the minimums and maximums over the full training set.

Task 3: Model development and validation

(a) Nitrate retrievals

Sentinel-2 imagery is available every 2nd and 3rd day at the Desoto, Kansas gauge. A catalog of imagery is shown in Figure 12. Many images are not useful because of cloud cover, ice cover, or other variable atmospheric conditions that make above water retrieval of radiances difficult. In total, 49 images were usable during 2021, and of those only 39 had complete sets of matching *in situ* data. Still, the available imagery is very informative and covers the full range of water states. For instance, Fig. 13 shows how the water changes from a highly turbid state (brownish water) to a high chlorophyll-a state (greenish water) in just 3 days during the Fall of 2021. The brownish water and higher water levels is due to a storm between 8 - 14 August 2021. The jump in turbidity can also be seen in Figure 11.

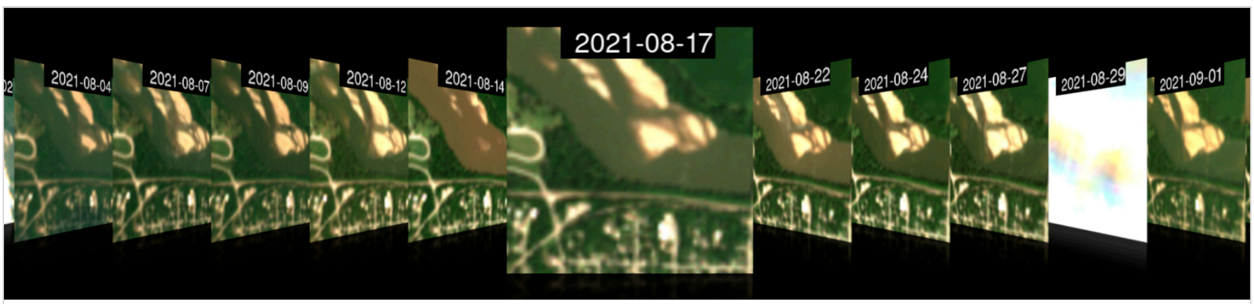

Figure 12. Kansas river at Desoto, KS Sentinel-2 imagery during Fall 2021.

Figure 13. Kansas river at Desoto, KS Sentinel-2 imagery during Fall 2021. Note dramatic change of water color from 14 August (left) to 17 August (right).

Figure 14 shows the dates of *in situ* data matches to good Sentinel-2 imagery. The data in the Figure is not vicariously calibrated. However, it still shows a few sensible trends. First, there is little data recovery in the winter (February) because of either clouds, snow or ice. Second, in general, turbidity values are inversely correlated to chlorophyll-a concentrations. This is reasonable, high turbidity values limit light which is essential for growing phytoplankton communities. It is also consistent with the type of band ratio algorithms used. Similarly, as seen

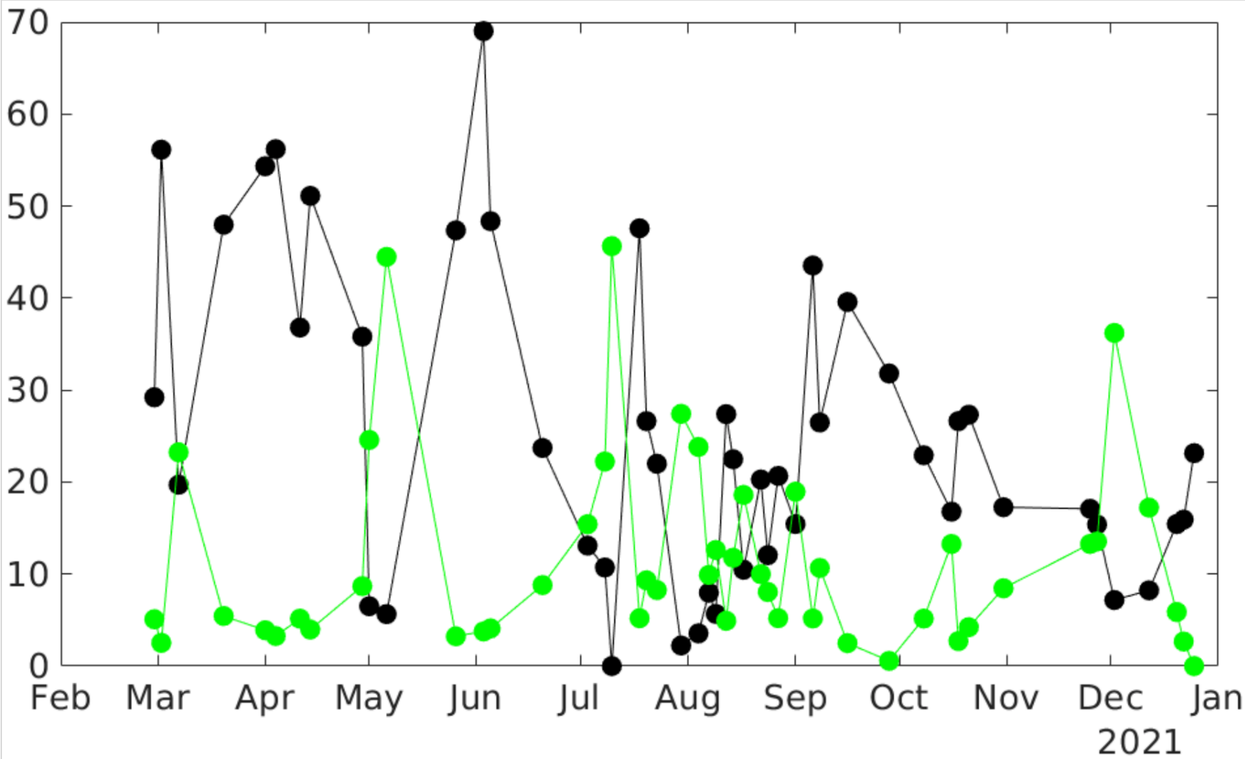

Figure 14. Nominal (i.e. no vicarious calibration) Sentinel-2 retrieval's for turbidity (black) [12] and chlorophyll (green) [11] for 2021. Note approximate inverse correlation between chlorophyll concentrations (mg/L) and turbidity (FNU).

in Figure 12, elevated chlorophyll-a values tend to pull down future nitrate values as phytoplankton consume nutrients.

To vicariously calibrate the satellite data products we can regress them against the USGS gauge data records. Figure 15 shows an estimate of the linear correlation (gain correction) needed to adjust the satellite retrievals of turbidity and chlorophyll-a to the gauge values. The gain adjustment for both products is ~ 2.

The scatter shown is typical for satellite retrievals, which typically need long data sets (< 200 pts) to show greater convergence.

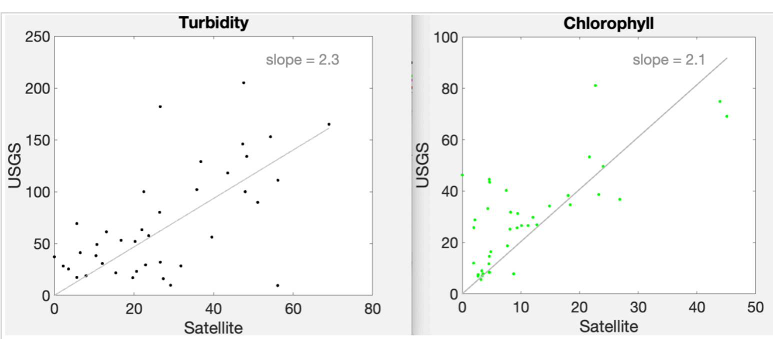

Figure 15. Linear regressions using 2021 matches for satellite calibration of turbidity (FNU) and (mg/L).

The next step is to create a static model to quantitatively correlate turbidity and chlorophyll-a values to nitrates. We tested two types of nonlinear functional methods, Multivariate Adaptive Regression Splines (MARS) [23], and convolutional neural networks (NN) [18]. For the MARS functions we used the Python py-earth library, or a

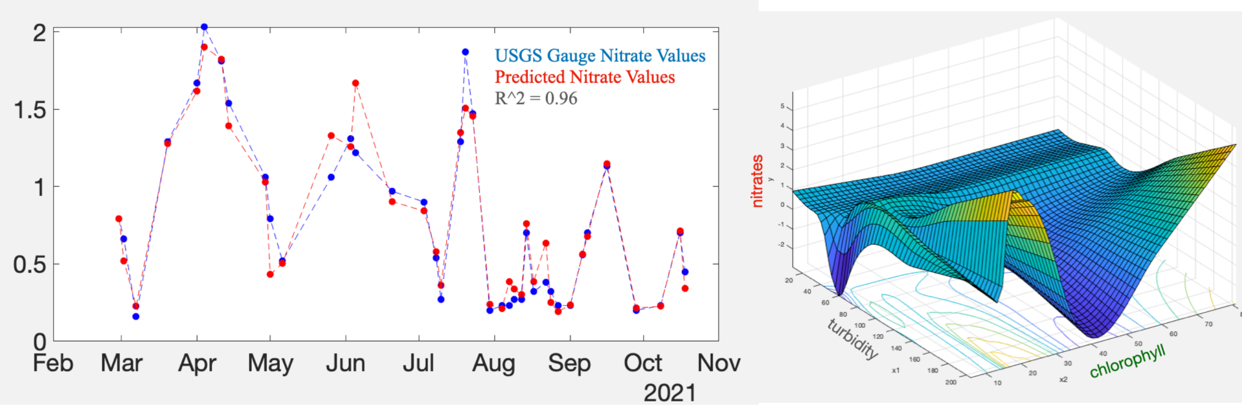

Figure 16. (left) Spline (MARS model) fit of input (turbidity, chlorophyll) to output (nitrates) for Desoto Kansas during 2021. (Right) Response surface for the model.

similar implementation in R or Matlab [23].

For MARS, the main parameters to tune are maximum number of basis functions (maxFuncs), maximum degree of interactions of input variables (maxInteractions), and the degree of the spline (linear, cubic, ...). For smaller data sets (<100 pts), we found that the most critical parameter was maxInteractions. Values chosen for this modeling exercise was maxFuncs = 40, maxInteractions = 5, spline type = cubic. A typical fit is shown in Figure 16 with a R^2 value of 0.96. The more critical test of the model is the 'out of sample' performance.
We trained on the first 80% of the data (Mar-Aug) and tested on the last 20% (Sept-Oct). The R^2 drops to 0.81 for the out of sample fit, but the result is still accurate enough for many applications (Fig. 17).

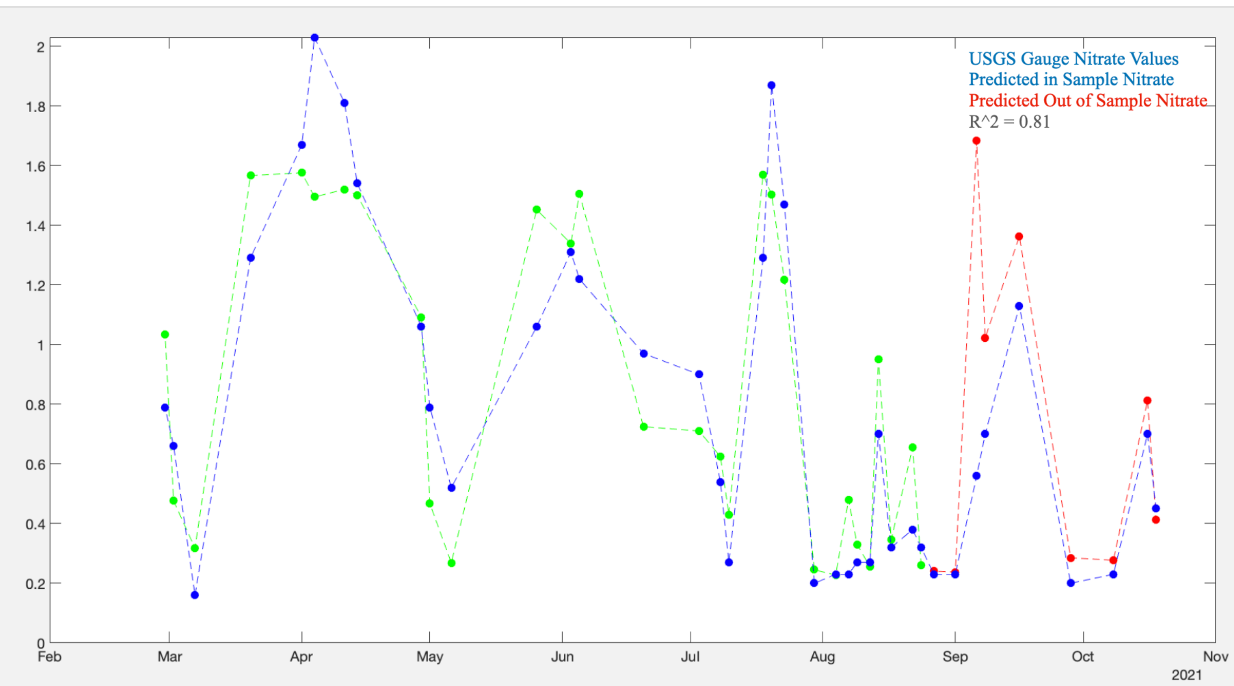

Figure 17. Out of sample test of spline (MARS) nitrate model. Last 20% is not in the training set (Sept-Oct).

As we increased the size of the training set we found that the spline models did not improve in accuracy. We conjecture that this is due to a lack of injectivity in the data set as the number of data training points increases --- that is there is a wider range of responses for similar inputs. We are preparing a paper that explains this theory in more detail [24].

In an attempt to improve the out of sample prediction, particularly for large data sets, we next turned to a black box a convolutional neural network model we formulated in Python PyTorch. Figure 3 shows a trading set based on 2016-2020 USGS data as input and the output again is nitrates. Though we are still in the early stages of these exercises, the conclusion is clear -- *accurate predictions of nitrates for specific sites is possible from nonlinear models using only chlorophyll-a and turbidity*. either from other gauge data (as has been noted previously), or from the same product fields derived from remote sensing.

Again we emphasize that the models are empirical and site specific. More work is needed to see how general these results are, and under what degree of concentration variations they might hold. Still, the first results do provide solid evidence that where nitrate value variations are of order 10X, and there are associated patterns of color variation in water (due to runoff or phytoplankton growth), quantitative correlations are feasible and useful to create.

(b) Phycocyanin retrievals

Sentinel-2 L1b data was downloaded from ESA's Copernicus Hub and processed to a L2 product including band Remote sensing reflectance (443, 490, 560, 665, 705, 740, 783, 842, 865 nm) [25]. In addition basic water quality products were produced for turbidity, chlorophyll-a, and a Normalized Difference Chlorophyll Index (NDCI) which is useful for detecting surface blooms [11-14]. All product data were interpolated to 10 meters. A browser for the processed Satellite Data of the San Luis Reservoir for 2021 is shown in Figure 18.

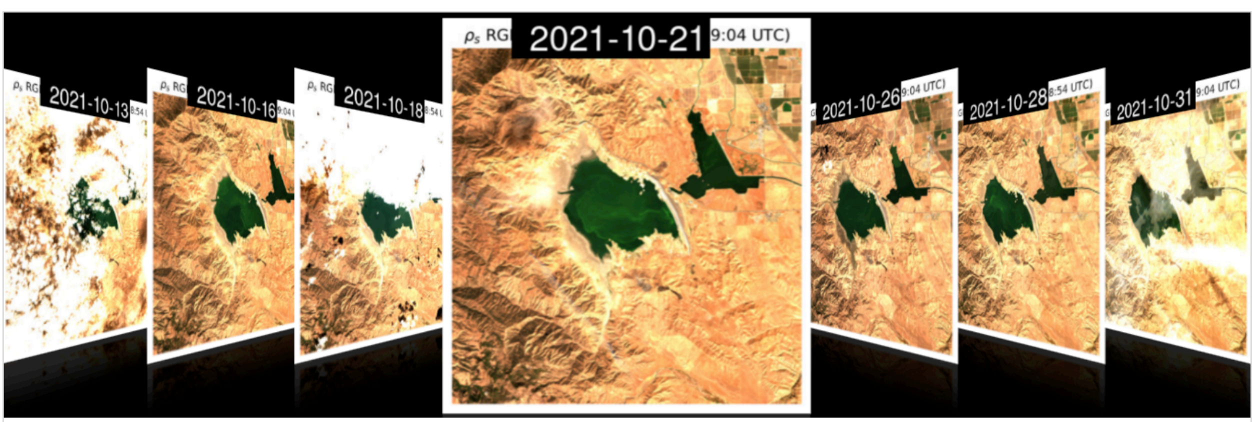

Figure 18. Sentinel-2 image browser for San Luis Resevoir, California.

During the second half of 2021, CA DWR collected water samples at a few locations in the San Luis Reservoir. The main sites are the Basalt Boat Ramp and the Pacheco Pumping Plant. The samples are analyzed for various toxins including microcystins, cylindrospermopsin, saxitoxin, and anatoxin-a. Microcystins are the most prevalent. Some species identification is also done. The most common species later in the year is microcystis, and earlier in the year dolichospermum is observed (Fig. 19).

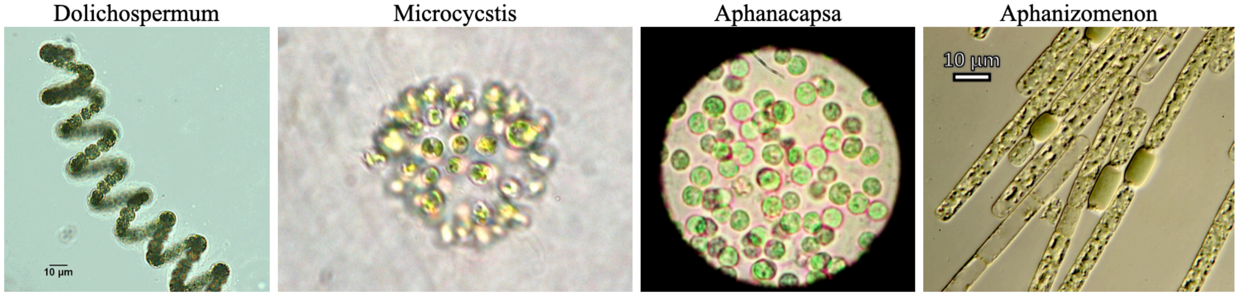

Figure 19. Four most common blue-green algae in San Luis Resevoir. Microcycstis is the dominant species later in the year.

In order to estimate the toxic (blue-green) algae in the reservoir we look at an associated pigment, phycocyanin. A recent study by Nardelli and Twardowski discusses the utility of using absorption bands to estimate pigment concentrations in field studies [26]. Specifically, they examine Chl-a absorption at 676 nm. Where possible to utilize, absorption provides a direct physical mechanism for detection of a pigment. The Gybe sensor can detect these absorption bands as illustrated in Figure 5. On the left side of Figure 5, Gybe sensor data for downwelling solar irradiance (including O2 absorption at ~760 nm) is plotted, and upwelling surface radiance is also shown. On the right side of Figure 5 the key absorption and reference bands are indicated -- a maximum back scatter band that is often used as a reference at 560 nm, and the phycocyanin band at ~630 nm, and a chl-a band at ~676 nm.

We will not go into great detail here, but a rough estimate is made of the so-called 'sky-glint' in the computation of Remote Sensing Reflectence (Rrs) [27]. The sensor placement had two non-ideal factors which affect the estimation of Rrs. First, the sensor was on the West side of a concrete structure above the intake pipe for the reservoir. The structure causes some blockage of sunlight preventing a full exposure of Es on the water (shadowing) before the solar zenith for the day (~1 PM PDT). We used a single (spectrally constant) scale factor to account for this shadowing. Of more concern is the sky-glint. The sensor is facing due West, so in the afternoon the relative view angle has increasing contamination by sky-glint which is easy to see in the Lt spectral data. We did a crude fix for this by subtracting off a spectral constant from Lt to ensure that the deep red signal is close to zero. Additionally, we did not use data past solar noon to keep sky glint contamination to a minimum.

For the reasons just discussed and a number more, the absorption bands are a proxy for estimating absorption directly, and its response may only be approximately linear (as opposed, say, to an instrument designed to measure absorption, such as the SeaBird's ac-s [28]). Still, as

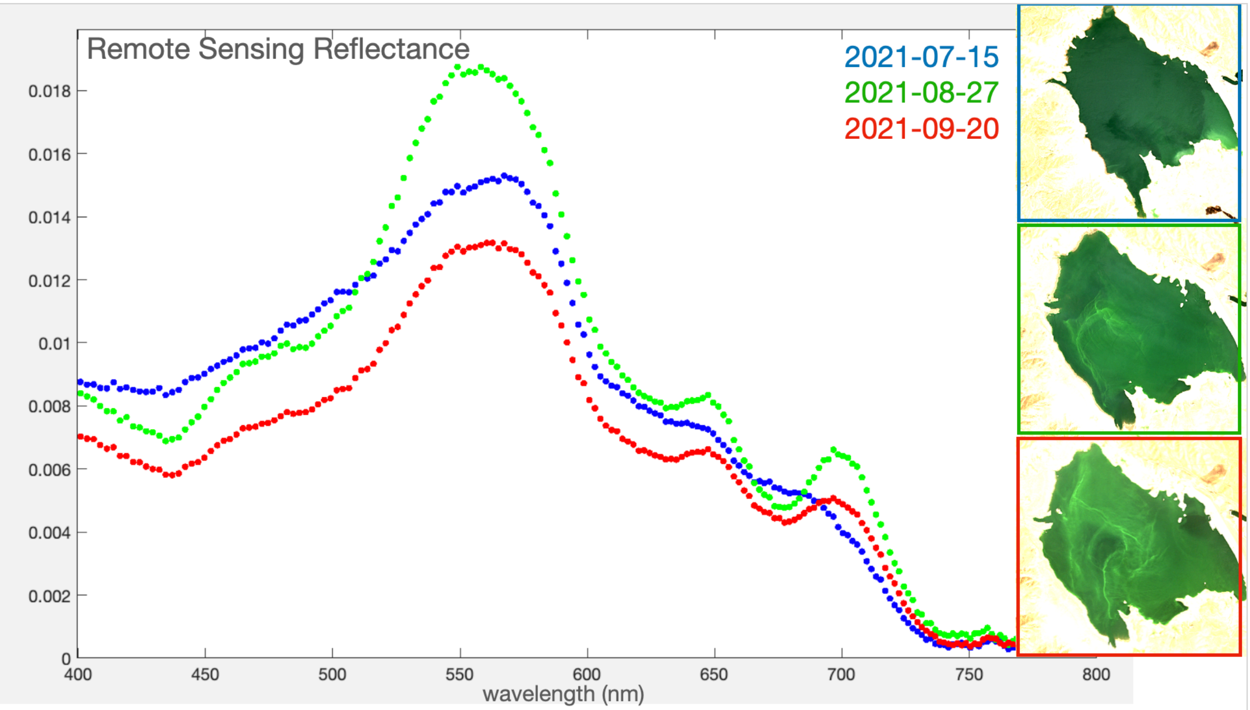

Figure 20. Remote sensing reflectance spectra from the Gybe sensor from three dates spanning summer to fall 2021. The increasing depth of the absorption bands (relative to the reference bands) provides a indicator of the concentrations of the pigments chlorophyll (green algae) and phycocyanin (blue-green algae). Sentinel-2 images are on the right provide for each day and also provide a relative indication of algae concentrations.

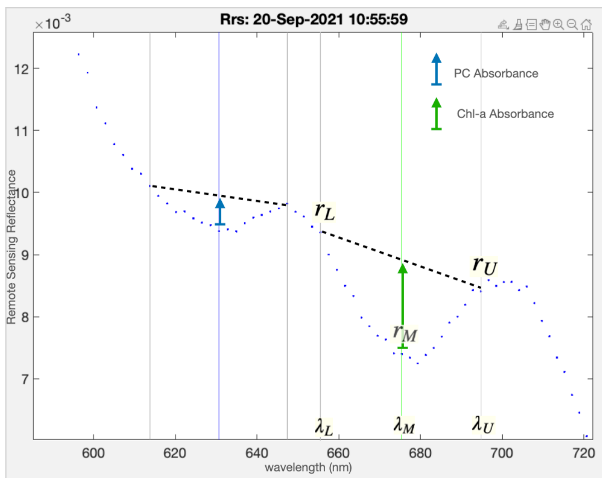

Figure 21. Illustration of computation of the Line Height Absorption method for both phycocyanin (~630 nm) and Chlorophyll-a (~676 nm) for data from the Gybe Sensor at San Luis Reservoir.

is commonly done, we used a Line Height algorithm to correlate the absorption (chl, pc concentrations) to the remote sensing reflectance spectra [29]. Figure 20, for instance, shows three scenes with increasing bloom intensity and (presumably) different compositions (green vs. blue-green algae population percentages). The Line Height method (here a line 'dip') uses only the local min or max in the spectrum, and can be less sensitive to other confounding factors. The specific method is illustrated in Figure 21. For each pigment three bands are chosen -- L (Lower), M (Middle), U (Upper) -- and the local dip is estimated by computing the distance from location of the value at the middle band, to a line formed by the values of the Lower and Upper band directly above the critical point. The specific formula is (see Fig. 21):

$$l_h = \delta \cdot r_U - r_M + (1 - \delta) \cdot r_L, \quad \delta = \frac{(\lambda_M - \lambda_L)}{(\lambda_U - \lambda_L)}$$

For this application the bands were all chosen based on the best bands for the spectra collected at San Luis reservoir with the Gybe Sensor. Specifically, for Chl-a the band set was (655, 676, 695) nm, and for PC the band set was (614, 631, 647) nm. The line height for phycocyanin is then regressed against the water samples. As is all too common in ocean color field studies, the locations do not match closely. The Gybe sensor views water on the Northeast side of the reservoir, and the water samples are from the Basalt Boat Launch on the south side. As some of the imagery indicates, at least for surface concentrations, the phytoplankton are not well mixed.

YEAR	MO	DA	MC	S2	GY
2021	06	28	0.40	1	1
2021	07	06	2.53	0	1
2021	07	12	2.10	0	1
2021	07	21	4.90	0	1
2021	07	26	3.45	0	1
2021	08	02	2.75	0	1
2021	08	09	13.80	1	1
2021	08	16	3.54	0	1
2021	08	23	8.60	0	1
2021	08	30	32.00	0	1
2021	09	07	43.00	0	1
2021	09	13	6.60	1	1
2021	09	20	17.70	0	1
2021	09	27	1.46	0	0
2021	10	04	12.70	0	0
2021	10	11	1.86	1	0
2021	10	18	2.50	1	0
2021	10	25	2.80	0	0
2021	11	01	0.53	0	0
2021	11	08	1.69	0	0
2021	11	15	0.53	0	0
2021	11	22	0.70	1	0
2021	11	29	0.00	0	0
2021	12	06	1.22	0	0
2021	12	13	0.00	0	0
2021	12	20	0.00	0	0

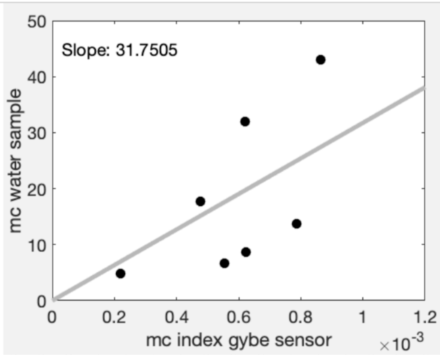

Figure 22. (Left) Water sampling data provided by CA DWR. MC is the microsystin concentration (ng/mL), and S2 and GY indicated availability of matches to Sentinel-2 or Gybe spectral data respectively. (Right) A correlation between the phycocyanin Line Height and the microsystin field samples is estimated by a linear regression.

At the surface, the microcystis can bunch together in long rows. The green algae at depth does appear more well mixed. Additionally, we observed a large scatter in the matchups for values below 4 ng/mL, so we exclude these samples. This is similar to the observation that Nardelli and Twardowski made for chl-a concentrations below 2 ng/mL [26]. After excluding these low values, we did observe a good correlation, though the data set is limited to only 7 samples (Fig. 22).

The last step is to relate the linear regression in Figure 22 to a subset of Sentinel-2 products (the spectral bands and/or water quality products). Given the small data set, we did not perform a more detailed study in surrogate modeling, but we were able to get a usable correlation with only the Normalized Difference Chlorophyll Index (NDCI), which we then used to estimate the microcystin concentration for clear Sentinel-2 images for 2021 (Fig 23).

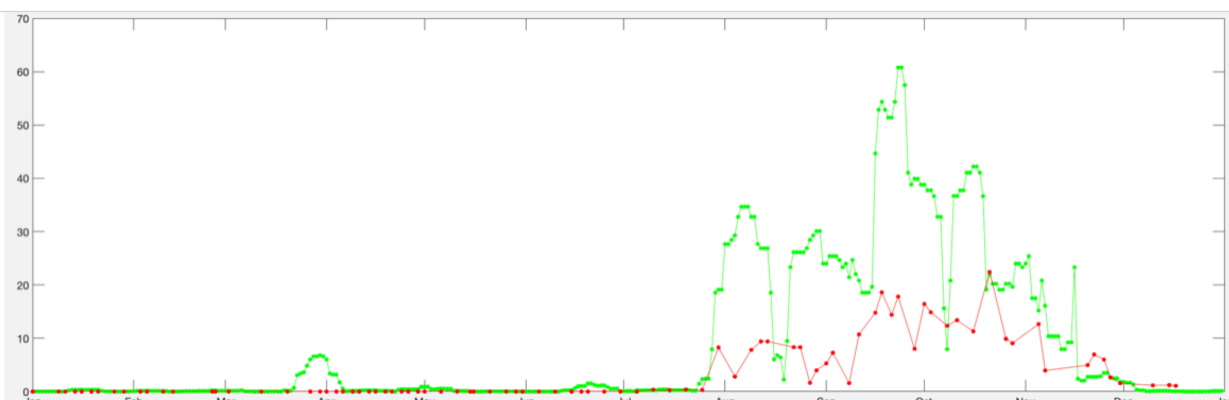

Figure 23. Estimates of blooms intensities at San Luis Reservoir, CA as tracked by chlorophyll and phycocyanin (microsystin) from satellite data for 2021. Green is Sentinel-3 NOAA product. Red is Sentinel-2 Gybe Product calibrated with CA DWR water samples. Blooms in late Fall and early winter are easily detected showing (presumably) and mixture of green and blue-green algae.

Because of the limited data these results should only be taken as indication of a correlation, and not for any quantitative predictions at this stage.

For completeness we compared the Sentinel-2 phycocyanin retrievals to chlorophyll-a retrievals for Sentinel-3 from the NOAA product (Fig 23) [30]. Both retrievals are based on averaging pixels (for Sentient-2 a 100x100 box) from the center of the reservoir.
A more detailed analysis is called for based on these preliminary results. For instance, there is some evidence that the blue-green algae crowds out the green algae in the Fall, that is, there is a sharp change in the community composition throughout the year.

Additional field data coordinated with historical remote sensing products should be able to test this hypothesis given the preliminary results and methods described in this study. Though these results are preliminary and limited, we think they provide compelling evidence that fusing contemporaneous *in situ* spectral data with operational remote sensing imagery, it is now possible to track independently both green algae and blue-green (HAB producing) algae independently.

This greatly increases the confidence of a remote sensing product in the monitoring of HABs, in particular, in tracking their movements (*e.g*. concentrations across a reservoir, proximity to intake pipes) with high spatial resolution. *Therefore, we have demonstrated how to fuse in situ spectral data with Sentinel-2 imagery to produce a phycocyanin retrieval, albeit for a very limited data set.* We caution though, that additional validation data sets are required to further refine the method, and gauge better the uncertainties, limitations, and utility in water management operations.

Conclusion

This study set out to test the hypothesis that *the combination of contemporaneous remote sensing imagery from operational sensors, in particular Sentinel-2, and an easily deployed sensor network of spectral radiometers -- the Gybe sensor net -- enables the retrieval of new water quality parameters, in particular, nitrates and phycocyanin*. Each quantity -- a nutrient and a harmful algal bloom indicator -- are considered to be among the most pressing needs of water quality managers. In addition to the deployment of the network and the collection of the required validation data, this study also greatly *extended the method of 'surrogate data modeling' currently in operational use by USGS, to nonlinear functional fitting with splines and neural networks. These new deep learning algorithms showed compelling results in their ability to track nitrates in the Kansas River using only inputs of turbidity and chlorophyll-a either from other gauge data, or from Sentinel-2 remote sensing data.* To the best of our knowledge these results provide a new level of accuracy for surrogate data estimations, which are of great concern for water quality managers.

Similarly, our study in the San Luis Reservoir, though based on limited validation data, also shows that *HAB pigments detected by an in situ spectral sensor network can be correlated to existing multispectral imager products.* With additional data, we also expect that the application of deep learning algorithms will produce results similar to our initial results with nutrients along the Kansas River.

The wide spatial coverage of the remote sensing of these new water quality products is useful for the water quality operations on several fronts including the identification of point and nonpoint runoff sources of nutrients and pollutants, the planning of effective sampling operations, the timing of reservoir releases for ecological impacts (*i.e*. river flushing), and the balancing of flow operations to meet metrics both for power operations and ecological sustainability.

References

[1]. T. J. Williams, *Linear regression model documentation and updates for computing water-quality constituent concentrations or densities using continuous real-time water-quality data for the Kansas River, Kansas*, July 2012 through September 2019, USGS Open-File Report 2021-1018 (2021).

[2]. P. M. M. Groetsch, R.Foster, and A.Gilerson, *Exploring the limits for sky and sun glint correction of hyperspectral above-surface reflectance observations,* Appl. Opt. 59, 2942-2954 (2020).

[3]. J. Pitarch, M/ Talone, G. Zibordi, and P. Groetsch, *Determination of the remote-sensing reflectance from above-water measurements with the "3C model": a further assessment*, Opt. Express 28, 15885-15906 (2020).

[4]. B. Pellerin, B. Bergamaschi, B. Downing, J. Saraceno, J. Garrett, L. Olsen, *Optical techniques for the determination of nitrate in environmental waters,* Chapter 5 of Section D Water Quality Book 1, Collection of Water Data by Direct Measurements, USGS (2013).

[5]. T.E.C. Kraus, B.A. Bergamaschi, P.J. Hernes, D. Doctor, C. Kendall, B.D. Downing & R.F. Loseel, *How reservoirs alter drinking water quality: Organic matter sources, sinks, and transformations*, Lake and Reservoir Management, 27:3, 205-219 (2011).

[6]. Web Link (verified 2022-03-14): https://waterdata.usgs.gov/monitoring-location/06892350/

[7]. C. D. Mobley, *Estimation of the remote-sensing reflectance from above-surface measurements*. *Appl. Optics* 38(36), 7442-7455 (1999).

[8]. K. G. Ruddick et. al., *A Review of Protocols for Fiducial Reference Measurements of Water-Leaving Radiance for Validation of Satellite Remote-Sensing Data over Water, Remote Sens. 11*, 2198 (2019).

[9]. C. Lee., N. Tufillaro,. B. Downing, E Hestir, and L. Wakamatsu, *Evaluation of satellite-derived water quality variables using satellite remote sensing and in situ data in San Luis Reservoir*, American Geophysical Union, Fall 2018 abstract #H43G-2520 (2018).

[10]. S. Mishra, and D. Mishra, *Normalized difference chlorophyll index: A novel model for remote estimation of chlorophyll-a concentration in turbid productive waters*. Remote Sensing of Environment 117, 394–406 (2012).

[11]. A. Dogliotti et. al., *A single algorithm to retrieve turbidity from remotely-sensed data in all coastal and estuarine waters*. Remote Sensing of Environment 156, 157–168 (2015).

[12]. Q. Vanhellemont, and K. Ruddick, *ACOLITE For Sentinel-2: Aquatic Applications of MSI imagery,* ESA Special Publication SP-740. Presented at the ESA Living Planet Symposium held in Prague, Czech Republic (2016).

[13]. Q. Vanhellemont, *Adaptation of the dark spectrum fitting atmospheric correction for aquatic applications of the Landsat and Sentinel-2 archives*. Remote Sensing of Environment 225, 175–192 (2019).

[14]. S. Ackleson, J. Rhea, S. Blaser, F. Wilkerson, R. Dugdale, C. Davis, and N. Tufillaro, SFE-O: *An optical model of the San Francisco Estuary*, Proceedings of Ocean Optics XXII (2016).

[15]. Q. Vanhellemont, Acolite Github Link (verified 2022-03-14): https://github.com/acolite/

[16]. T. J. Rasmussen, A. C. Ziegler, and P. P. Rasmussen, *Estimation of constituent concentrations, densities, loads and yields in lower Kansas River, northeast Kansas, using regression models and continuous water-quality monitoring*, January 2000 through December 2003: U.S. Geological Survey Scientific Investigation report 2005-5165, 117p (2005).

[17]. J. Howard and S. Gugger, *Deep Learning for Coders with Gastai and PyTourch: AI Applications Without a PhD*, O'Reilly Publications (2020).

[18]. J. Wood, D. E. Root and N. B. Tufillaro, *A behavioral modeling approach to nonlinear model-order reduction for RF/microwave ICs and systems*, in IEEE Transactions on Microwave Theory and Techniques, vol. 52, no. 9, pp. 2274-2284 (2004).

[19]. A. K. Ziemann, *A manifold learning approach to target detection in high-resolution hyperspectral imagery*, Ph.D. Thesis, Rochester Institute of Technology (2015).

[20]. Web Link (retrieved on 2022-03-14): https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-2-msi/msi-instrument

[21]. P. Rasmussen, HJ. Gray, G. Glysson, and A. Ziegler, *Guidelines and procedures for computing time-series suspended sediment concentrations and loads from in-stream turbidity sensor and streamflow data: US Geological Survey techniques and Methods*, book3, chap. C4, 53p. (2009).

[22]. J. Olivera-López, et. al, *A review of instance selection methods.* Artif. Intell. Rev. 34. 133-143 (2010).

[23]. G. Jekabsons, ARESLab: *Adaptive Regression Splines toolbox for Matlab/Octave*, 2016, available at http://www.cs.rtu.lv/jekabsons/

[24]. N. Tufillaro, *Measures of injectivity for surrogate models, in preparation.*

[25]. Web Link (verified 2022-03-23): https://scihub.copernicus.eu/dhus/#/home

[26]. S. Nardelli and M. Twardowski, *Assessing the link between chlorophyll concentration and absorption line height at 676 nm over a broad range of water types*. Opt Express, 24(22):A1374-A1389 (2016).

[27]. P. Gege and P. Groetsch, *A spectral model for correcting sun glint and sky glint*, Ocean Optics XXIII Conference paper, Victoria Canada (2016).

[28]. Web Link (verified 2022-03-23): https://www.seabird.com/ac-s-spectral-absorption-and-attenuation-sensor/product?id=60762467715

[29]. Chuanmin Hu and Lian Feng, M*odified MODIS fluorescence line height data product to improve image interpretation for red tide monitoring in the eastern Gulf of Mexico,* J. Appl. Rem. Sens. 11(1) 012003 (2016).

[30]. Web Link (verified 2022-03-23): https://coastwatch.noaa.gov/cw/satellite-data-products/ocean-color/near-real-time/olci-sentinel3.html

Acknowledgments

It is a pleasure to thank the many individuals and organizations that assisted with these studies. We thank the Lawrence, Kansas office of the USGS, in particular Dr. Thomas Williams and Ryan Walters, for their assistance in the installation and operation of the De Soto sensors. We thank Brianne Sakata of CA DWR for the collection and processing of the water samples at San Luis Reservoir. The assistance of the staff at the San Luis Reservoir site for help with the installation and maintenance of the Gybe Sensor. We thank Dr. Dan Sousa at San Diego State University for his recommendations and illustrations of how to use MARS models for surrogate data modeling. We thank Dr. Philipp Grötsch, Omar Zurita, Sara De Moitié, Alexis Nelson, Henry Kobin, Ivan Lalović, and the staff at Gybe for their assistance in preparing this report. We thank Clemente Borgogni of Georgetown University for his assistance in the manufacture, calibration, and data analysis of the Gybe sensors used in this project, and Dr. Scott Prahl of the Oregon Institute of Technology for assistance with the calibration facilitates and set up.

This work in no way serves as an endorsement of Gybe, or any of its products, by the persons or organizations acknowledged. Any errors in this report are solely the fault and responsibility of the authors.