NASA DEVELOP National Program
Arizona – Tempe

Summer 2021

Highland Lakes Water Resources

Using NASA Earth Observations to Improve Detection Systems for Harmful Algal Events in the Highland Lakes in Central Texas

DEVELOP Technical Report

Final – August 12th, 2021

Kaitlynn Hietpas (Project Lead)
Shuyu Chang
Addison Pletcher
Mark Radwin
Emma Waugh

Advisors:
Dr. Erin Urquhart, Science Systems & Applications, Inc., NASA Goddard Space Flight Center (Science Advisor)
Dr. Christine Lee, California Institute of Technology, NASA Jet Propulsion Laboratory (Science Advisor)
Dr. David Hondula, Arizona State University (Science Advisor)

1. Abstract

Beginning in 2019, harmful algal events in Austin, Texas, caused canine deaths in the Lady Bird Lake and Lake Travis reservoirs. These reservoirs are part of the larger Highland Lakes chain, managed by the Lower Colorado River Authority (LCRA) and the City of Austin Department of Watershed Protection (CoA DWP), which fulfill municipal, commercial, and agricultural water demands. Given the recent increase in favorable algal event conditions in central Texas, the LCRA and CoA DWP partnered with NASA DEVELOP to improve algal event early-warning systems through the application of remote sensing and machine learning. An Earth observation-based algal monitoring system will assist the responsible agencies in predicting algal conditions and communicating hazards to the public. The NASA DEVELOP team utilized Landsat 8 Operational Land Imager (OLI) and Sentinel-2 Multispectral Instrument (MSI) data to produce products including chlorophyll-a concentrations, cyanobacteria detections, turbidity, and water surface temperature. Chlorophyll-a concentrations were retrieved with a pre-trained machine learning model (mixture density network) and spectral indices, while the other products were derived from spectral indices. In situ field data were used to validate and quantify uncertainties for each product. The validations show strong correlations for chlorophyll-a and water surface temperature. Time series analyses of chlorophyll-a concentrations show peaks in the severe drought years (2015 and 2016). This project's resulting products enable monitoring of environmental proxies relevant to algal event presence in the Highland Lakes chain and will ultimately support water management, decision making, and risk communication.

Key Terms

machine learning, remote sensing, chlorophyll-a, cyanobacteria, turbidity, water surface temperature, Landsat 8 OLI, Sentinel-2 MSI

2. Introduction

2.1 Background Information

Globally, algal events are increasing in frequency (Lopez et al., 2008). Favorable environmental conditions that can lead to the growth of algae include low hydrologic flow, warm water temperatures, and urban, industrial, and agricultural nutrient sources leading to eutrophic conditions. These conditions are becoming dominant in inland waters due to industrialization and rising global temperatures (Lopez et al., 2008). In addition to depletion of dissolved oxygen in aquatic communities, blue-green algae, or cyanobacteria, can produce substances called cyanotoxins, which can affect the liver, kidney, or nervous system of animals (Lopez et al., 2018). Cyanotoxins pose a threat to the health and wellbeing of water users, including wildlife, humans, and pets.

Our area of interest is the Highland Lakes chain which is comprised of six reservoirs in the lower Colorado River in Texas as shown in Figure 1. In the Highland Lakes, phytoplanktonic algal events, or blooms, have been occurring for years with no apparent cyanotoxin production (Bellinger et al., 2018). Though most cyanotoxin research has focused on phytoplankton, benthic algae proliferations have been recently found to produce neurotoxins such as anatoxin-a and dihydroanatoxin-a (Fetscher et al., 2015 & Bouma-Gregson et al., 2018). The production of these cyanotoxins has been concomitant with the observed increase

of benthic algae proliferations in multiple Highland Lakes since 2019 (Manning et al., 2020). Though algal blooms and proliferations are visibly different, the differences are only detectable through remote sensing technology.

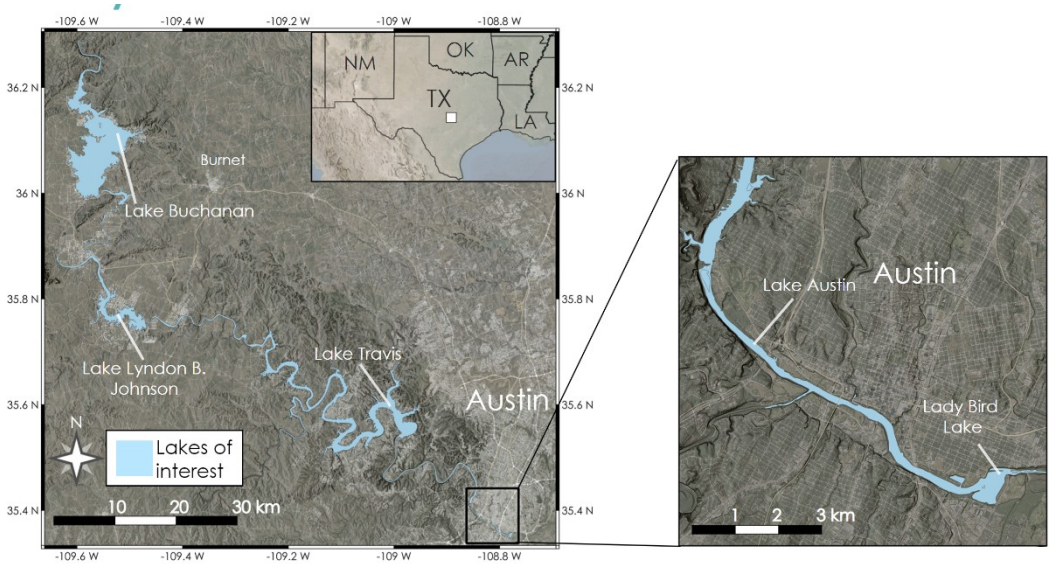

Figure 1. Spatial extent of the study area including all seven reservoirs in the Highland Lakes chain in Central TX, with an inset on the city of Austin and its surrounding reservoirs.

The occurrence of harmful algal events presents risks to ecosystems, public health, water supply, and recreation opportunities. The Highland Lakes, located in and near Austin, Texas, are considered a high-traffic recreation destination for community use. Reports of algae-related canine deaths first began in Lady Bird Lake in Austin, Texas during August 2019 and persisted until November 2020 when conditions were no longer favorable for mat growth (Manning et al., 2020). Cyanotoxins were also found in Lake Travis, upstream of Lady Bird Lake, in February 2021, and have been detected as recently as June 2021. Algal events have prompted recreation and beach closures in these areas, resulting in loss of revenue for recreation facilities and decreased tourism overall. Increased frequency and extent of field sampling for cyanotoxins have incurred additional costs, as well as emergency meetings and enforcement of closed areas (Manning et al., 2020). While these occurrences are relatively new to the Highland Lakes area, *in situ* data and satellite imagery from 2013 to present day were used in this study.

Remotely sensed ocean color data have long been applied to produce effective estimations of near-surface concentration of chlorophyll-a (*chl-a*) and cyanobacteria in coastal and inland waters. Smith et al. (2021) and Pahlevan et al. (2020) adapted and trained a machine-learning algorithm to retrieve *chl-a* concentrations for both inland and coastal water bodies, which outperformed other existing *chl-a* detection algorithms. Zhau et al. (2020) successfully established and utilized a spectral algorithm to achieve cyanobacteria detection in Lake Erie for sensors that lack spectral bands to directly detect cyanobacteria. Because clear and turbid water have different optical signatures in the red and green wavelengths of the spectrum, turbidity can also be detected using remotely sensed data; Lacaux et al. (2007) employed the Normalized Difference Turbidity Index to

classify pond turbidity in Senegal. While these satellite metrics cannot detect actual cyanotoxins, they can provide proxy data that can contribute to management strategies by alerting users to the formation of algal events that may produce toxins.

2.2 Project Partners & Objectives

The City of Austin Department of Watershed Protection (CoA DWP) and the Lower Colorado River Authority (LCRA) are responsible for managing water quality in the Highland Lakes. Both organizations partnered with the NASA DEVELOP Highland Lakes Water Resources team to improve early warning systems for harmful algal events. Historically, these organizations relied on *in situ* sampling to monitor for harmful algal event indicators to warn the public when conditions were hazardous. This project aimed to expand these detection tools through the application of remotely sensed data products. The University of Texas at Austin Department of Molecular Sciences and the Austin Water Utility provided additional information as needed.

The objective of this project was to employ Earth observation data to improve harmful algal event detection systems in the Highland Lakes, Texas, to complement *in situ* sampling which can be time-consuming and expensive. By creating tools to monitor remotely sensed algae indicators, this project aimed to help our partners determine when and where harmful algal events are likely to occur. In turn, this will enable them to take steps to alleviate the impact of these harmful algal events. Additionally, historic satellite imagery was used to identify areas likely to experience algal events to help inform future field sampling efforts.

3. Methodology

3.1 Data Acquisition

To observe relevant changes within the water column of our study area, our team utilized the NASA Landsat 8 satellite with the Operational Land Imager (OLI) and Thermal Infrared (TIRS) sensor as well as the European Space Agency's (ESA) Sentinel-2 satellites with the MultiSpectral Instrument (MSI) sensor. Landsat 8 OLI has a spatial resolution of 30 m for most bands, temporal resolution of 16 days, and 11 spectral bands. Only 7 of Landsat 8 OLI's spectral bands are immediately useful for the visible-to-short-wave-infrared range of the electromagnetic spectrum (Figure 2). The Sentinel-2A and 2B MSI sensors have varying spatial resolutions for the available bands, ranging from 10 to 60 m, a combined temporal resolution of 5-days, and a total of 13 spectral bands geared towards the visible-to-short-wave-infrared range (Table A1). In conjunction, these platforms can provide revisit times as short as ~3 days.

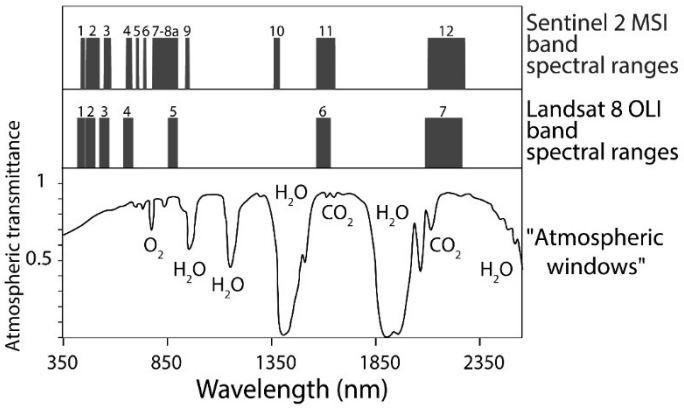

Figure 2. Comparison of bands and respective spectral ranges for the ESA Sentinel-2 MSI and NASA Landsat 8 OLI spaceborne platforms. Bottom panel shows the transmittance of the atmosphere, highlighting the “atmospheric windows” considered when these sensors were designed and the responsible absorbers. Original figure.

Our team retrieved data from the Google Earth Engine Data Catalog, which freely hosts petabytes of Earth observation data ready for processing in the cloud. The chosen image collections in Earth Engine were already atmospherically corrected and pre-processed to surface reflectance to work within the capabilities of the Google Earth Engine API and increase ease of processing for the development of the end product. Thus, the chosen collections were the United States Geological Survey (USGS) Landsat-8 Surface Reflectance Collection 2 Tier 1 and Sentinel-2 MSI: MultiSpectral Instrument Level-2A datasets. The retrieved Landsat data were processed to surface reflectance by the USGS with the NASA Land Surface Reflectance Code (LaSRC; Vermote et al., 2018). The retrieved Sentinel-2 data were processed by the ESA using the Sen2Cor algorithm (Main-Knorn et al., 2017). We filtered each collection to the study area by tile, where scene cloud percentages greater than 90% were filtered out to avoid overly cloudy images. All available dates of imagery were necessary for historical analysis purposes, so no date filter was applied. Our partner organizations provided us a shapefile of the Highland Lakes chain as well as water quality data from approximately monthly *in situ* sampling from 1980 to 2021. Data from the LCRA covered the entire Highland Lakes chain, and parameters include two different metrics of *chl-a* (µg/L), Secchi disk depth (m or ft), water temperature at 0.3 m depth (°C), and turbidity (NTU). Data from the CoA DWP covered Lady Bird Lake and Lake Austin, and parameters included water temperature at 0.1-1.2 m depth (°C), Secchi disk depth (m), and *chl-a* (µg /L).

3.2 Data Processing

Prior to processing imagery for environmental proxies, our team masked out land and cloud pixels to prevent the incorporation of unwanted pixels into analysis and to reduce amount of data processed. Land pixels were segmented out of the image by processing Normalized Difference Water Index (NDWI) images for each image, then masking all pixels below a chosen threshold based on the NDWI image. To accommodate differing lighting, geometry, and environmental conditions between scenes, we implemented a dynamically changing threshold for each NDWI image using the Otsu statistical image segmentation method (Otsu, 1979; Zhou et al.,

2020). Additionally, we masked the imagery based on the shapefile of the Highland Lakes region where we utilized a 30-meter buffer from the shoreline to ensure no contamination between land and water pixels. We performed cloud masking by utilizing the quality control bands present on the Landsat and Sentinel platforms, which indicate whether each pixel in an image is likely to be a cloud or cloud shadow and allowed for masking of cloud pixels. Cloud shadows were not masked out as those areas can still potentially provide insight.

3.3 Data Analysis

3.3.1 Chl-a Concentration and Mixture Density Networks

We retrieved *chl-a* concentrations from atmospherically corrected Sentinel-2 MSI and Landsat 8 OLI imagery through a novel pre-trained machine-learning algorithm named Mixture Density Networks (MDN). MDN, a class of neural networks (Bishop, 1994), was designed to address one-to-many problems (one radiometry value corresponds to more than one potential *chl-a* values) by modeling a conditional probability distribution over the range of the outputs. Standard neural networks directly model the *chl-a* corresponding to the specific radiometric quantity (Equation 1), while MDNs model a conditional probability, mapping as a mixture of various Gaussian distribution functions (Equation 2). The MDN utilizes a subset of bands within the visible (VIS) and near infrared (NID) bands (443 nm, 482 nm, 561 nm, 655 nm for OLI and 443 nm, 490 nm, 560 nm, 665 nm, 705 nm, 740 nm, 783 nm for MSI) Smith et al. (2021) and Pahlevan et al. (2021) have suggested the MDN model can be adopted to predict long-term *chl-a* concentration for inland and nearshore coastal water bodies. They have also suggested MDNs outperform most state-of-the-art benchmarks, though the results are sensitive to noise in remote sensing products. Supplemental to the machine learning values, we implemented the best performing spectral *chl-a* indices for both Sentinel and Landsat, the 2 Band Algorithm (2BDA; Equation 3) and Fluorescence Line Height Violet (FLHV; Equation 4) indices, respectively (Johansen et al., 2020; Dall'Olmo and Gitelson, 2006; Beck et al., 2016). The 2BDA index utilizes the difference between the first red-edge band, centered near 705 nm, to the red band, centered near 665 nm, to quantify the abundance of *chl-a* in the water column. Due to the lack of red-edge bands on the Landsat-8 OLI platform, the utilized FLHV index harnesses the coastal aerosol, green, and red bands to quantify *chl-a*.

$$p\left(Chla \middle| R_{rs}\right) = \sum_{i=1}^{c} \pi_i\left(R_{rs}\right) \Phi_i\left(Chla \middle| R_{rs}\right) \tag{1}$$

$$\Phi_i\left(Chla \middle| R_{rs}\right) = \frac{\exp\left(-(1/2)\left(Chla-\mu_i\right)^T \sum_i^{-1}\left(Chla-\mu_i\right)\right)}{\sqrt{(2\pi)^d \left|\Sigma_i\right|}} \tag{2}$$

$$2BDA = \frac{\rho_{¿\,Edge}(705)}{\rho_{¿}} \tag{3}$$

$$FLHV = \rho_{¿} - \left(\rho_{¿} + \left(\rho_{Aerosol} - \rho_{¿}\right)\right) \tag{4}$$

3.3.2 Turbidity and Normalized Difference Turbidity Index (NDTI)
The widely used Normalized Turbidity Index, combining the green and red bands of the electromagnetic spectrum, is defined as:

$$NDTI = \frac{DC\ s_{red} - DC\ s_{green}}{DC\ s_{red} + DC\ s_{green}} \quad (5)$$

where DC is the digital count, or reflectance values, including radiometric and geometric corrections (Lacaux et al., 2007; Elhag et al., 2019). Suspended sediment in the water column causes higher reflectance in the red wavelengths compared to the green wavelengths, which is used here to detect turbid pixels. Normalization of the spectral index allows for more consistent results even as lighting and environmental conditions vary. In this project, the turbidity levels were estimated based on NDTI and validated with *in situ* measured turbidity (NTU) data.

3.3.3 Broad Wavelength Algae Index (BWAI)
We achieved detection of cyanobacteria mats by implementing a modified cyanobacteria index which can be used for high spatial resolution sensors such as Landsat-8 OLI and Sentinel-2 MSI. Typically, the established and widely used Cyanobacteria Index (CI) quantifies cyanobacteria cell abundance in cells/mL, and relies on multiple bands in the 620-680 nm range absent on the platforms used for this project (Matthews et al., 2012; Lunetta et al., 2015). A recently established algorithm, called the Broad Wavelength Algae Index (BWAI), was designed to emulate cyanobacteria detection like the CI and suppress signals due to turbidity and suspended sediment for more accurate results (Zhao et al., 2020). The BWAI harnesses the algal mat Reflectance Peak Height (RPH, Equation 6) in the near-infrared or green bands and uses a spectral baseline combining the short-wave-infrared and blue bands to normalize the near-infrared reflectance peak. The RPH equation can be implemented as two versions: one version designed to detect cyanobacteria in mesotrophic/eutrophic waters typically plagued by blooms, using the green band, and the other version geared towards hypertrophic waters, typically plagued by vegetative cyanobacteria mats (proliferations), using the NIR band. Our team implemented both versions as the trophic status can vary within the Highland Lakes.

To suppress pixels with high amounts of turbidity and variable amounts of suspended sediment, a depression factor (F_S, Equation 7) and a signal modulation factor (F_C, Equation 8) are calculated, respectively, and weighted to determine how the algorithm will be implemented for each pixel. The depression factor depends on the relationships between the green, red, and near-infrared band reflectances. It is necessary to implement the depression factor as waters with high amounts of suspended sediment can enhance the near-infrared reflectance peak. Likewise, suspended sediment can enhance the blue band reflectance and is accounted for with the signal modulation factor. The final BWAI algorithm (Equation 9) workflow takes the following form:

$$RPH_{(Green,\ NIR)} = \rho_{max} - \rho_{Blue} - \left(\rho_{SWIR} - \rho_{Blue}\right) \times \frac{\lambda_{max} - \lambda_{Blue}}{\lambda_{SWIR} - \lambda_{Blue}} \quad (6)$$

$$F_S = \exp(\rho_{Red} - \rho_{Green} - (\rho_{NIR} - \rho_{Green}) \times \frac{\lambda_{Red} - \lambda_{Green}}{\lambda_{NIR} - \lambda_{Green}} \quad (7)$$

$$F_C = \exp \frac{(\rho_{Green} - \rho_{Blue})}{(\rho_{Green} + \rho_{Blue})} \quad (8)$$

$$BWAI = RPH_{(Green,\ NIR)} \times \begin{cases} F_C, \ln(F_s) \leq T \\ F_S^{-1}, \ln(F_s) > T \end{cases} \quad (9)$$

where ρ represents the reflectance value of a specific band (denoted as subscript), ρ_{max} represents either the green or near-infrared band (depending on use case), and λ represents the numerical wavelength value of a specific band (denoted as subscript). The BWAI algorithm is calculated as the RPH multiplied by one of the modulation factors, depending on the value of the logarithm of the depression factor compared to the turbidity threshold, T. The turbidity threshold is chosen by the user and has a recommended range of –0.015 to 0.003, depending on the noise levels and atmospheric conditions (Zhao et al., 2020). A turbidity threshold of 0.003 is chosen for this project, based on the choice of the BWAI establishing literature. Although the algorithm is modified to emulate the CI, the resulting values are not quantitative cells or mL values, but rather are relative values to indicate low, medium, or high amounts of algal mat presence.

3.3.4 Water Surface Temperature

Our team retrieved water surface temperature using Landsat 8's thermal infrared (TIRS) bands, located between 10600 and 11200 nm. Sentinel-2's MSI does not host a thermal sensor and thus cannot be used for water surface temperature retrieval. The technique of measuring surface temperature via remote sensing harnesses an adaptation of the Planck function, which relates wavelength-dependent spectral radiance of a surface to the temperature of the surface (equation 10). We accomplished retrieval of water surface temperature using the radiative transfer equation method as described by Sekertekin & Bonafoni (2019), using the at-sensor radiance of band 10 alongside the atmospheric transmittance, emissivity, and both upwelling and downwelling radiance values available in Landsat 8's Collection 2 image metadata. We calculated the water surface temperature algorithm as:

$$T_S = \frac{K_2}{\ln ¿¿} \quad (10)$$

Calibration constants K_1 (774.89 Watts/(m^2×srad×μm)) and K_2 (1321.08 Kelvin) are specific to Landsat 8 OLI's band 10 (Sekertekin & Bonafoni 2019). This method was not designed specifically for water surface temperatures, so emissivity estimates will likely be a source of error.

3.3.5 Validation of Products and Statistical Metrics

To provide error estimates and performance metrics for the products, our team compared remotely sensed measurements to *in situ* measurements provided by the partners. To evaluate uncertainty and validate the model performances, we

generated validation matchups through pairing *in situ* measurements with same-day overpassing satellite observations, following strict spatiotemporal filters proposed by Bailey & Werdell (2006). Then, we exported the average pixel value for a 3x3 array centered over the matching sampling locations for each day of sampling. For the MDN *chl-a* product validation, we excluded points for which there were fewer than five valid pixel values. In *chl-a* validation control, outliers around shoreline area with percentage differences more than 200% were removed for quality control. Two outliers were removed from Sentinel-2 *chl*-a products, and three outliers were removed from Landsat 8 *chl-a* products.

Both linear and log-transformed metrics were selected as performance indicators by comparing estimates (E) against measured (M) values. The evaluation metrics include (Equations 11-16):

$$\mathrm{RMSE} = \left[\frac{\sum_{i=1}^{N}\left(E_i - M_i\right)^2}{n}\right]^{1/2} \tag{11}$$

$$\mathrm{RMSLE} = \left[\sum_{i=1}^{N}((\log_{10}(E_i ¿ - \left(\log_{10}(M_i)\right)^2 ¿¿ n\right]^{1/2} \tag{12}$$

$$\mathrm{MAPE} = 100 \times \tilde{r} \text{ where } \tilde{r} \text{ is the median of } \left[\left|E_i - M_i\right| / M_i\right] \tag{13}$$

$$\mathrm{Bias} = 10^{Z} \text{ where } Z = \left[\sum_{i=1}^{n}\left(\log_{10}(E_i) - \log_{10}(M_i)\right) / n\right] \tag{14}$$

$$\mathrm{MAE} = 10^{Y} \text{ where } Y = \left[\sum_{i=1}^{n}\left|\log_{10}(E_i) - \log_{10}(M_i)\right| / n\right] \tag{15}$$

$$r = \frac{\sum\left(M_i - \overline{M}\right)\left(E_i - \overline{E}\right)}{\sqrt{\sum\left(M_i - \overline{M}\right)^2 \sum\left(E_i - \overline{E}\right)^2}} \tag{16}$$

where RMSE is the root mean square error, RMSLE is the root mean squared log-error, MAPE is the median absolute percentage error, Bias represents log-transformed residuals, MAE stands for the mean absolute error computed in log-space, and r is the Pearson Correlation Value.

4. Results & Discussion

4.1 Analysis of Results

4.1.1 Validation

The machine learning model performs the best for both Sentinel and Landsat, compared with other traditional spectral *chl-a* indices (Table A3). From the matchup datasets, Sentinel-1 MDN *chl-a* values strongly correlated with *in situ* measurements (R^2=0.73), while Landsat *chl-a* values are moderately correlated (R^2=0.6) (Figure 3). The improved Sentinel-2 MSI performance can likely be

attributed to its extra spectral bands in the red-edge wavelengths. The spectral *chl-a* indices (2BDA and FLHV) perform poorly, with positive but weak correlations (R^2=0.08 and 0.09, respectively) (Figure 3).

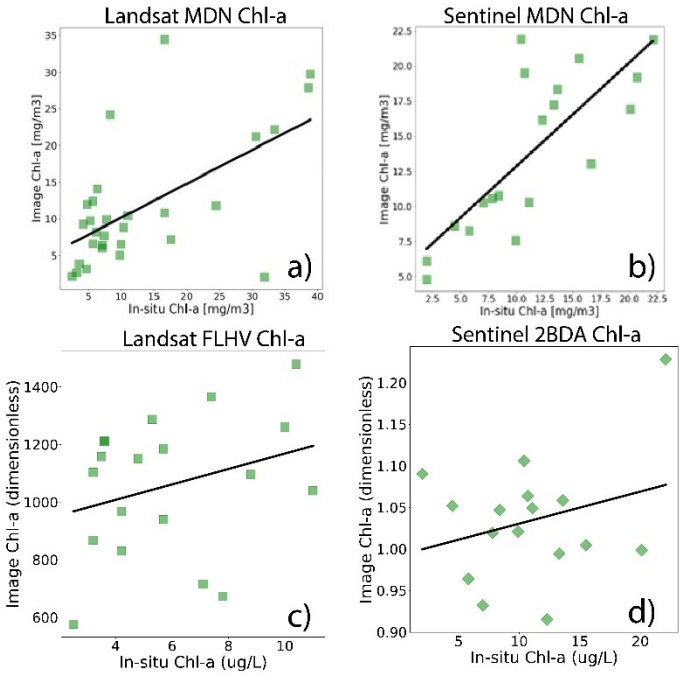

Figure 3. *Chl-a* linear regression comparisons with Landsat derived MDN values (a), Sentinel derived MDN values(b), Landsat derived FLHV values (c), and Sentinel derived 2BDA values (d).

When comparing NDTI turbidity values to *in situ* data, there were weak and negative correlations for both Sentinel-2 MSI and Landsat-8 OLI (R^2 = -0.029 and 0.029, respectively) (Figure 4). The discrepancy in values may be explained by the length of time between image acquisition and water sampling, as suspended sediments can change in short time frames. Additional discrepancies may be attributed to comparing two fundamentally different methods of quantifying turbidity (NDTI vs. NTU). To further investigate the performance of the NDTI, we manually compared color imagery to NDTI imagery and found, in most of the imageries, visually distinguishable suspended sediments plumes are accurately identified. These manual comparisons, along with the widespread use of NDTI, suggest the satellite-derived turbidity product performs much better than the *in-situ* comparisons indicate.

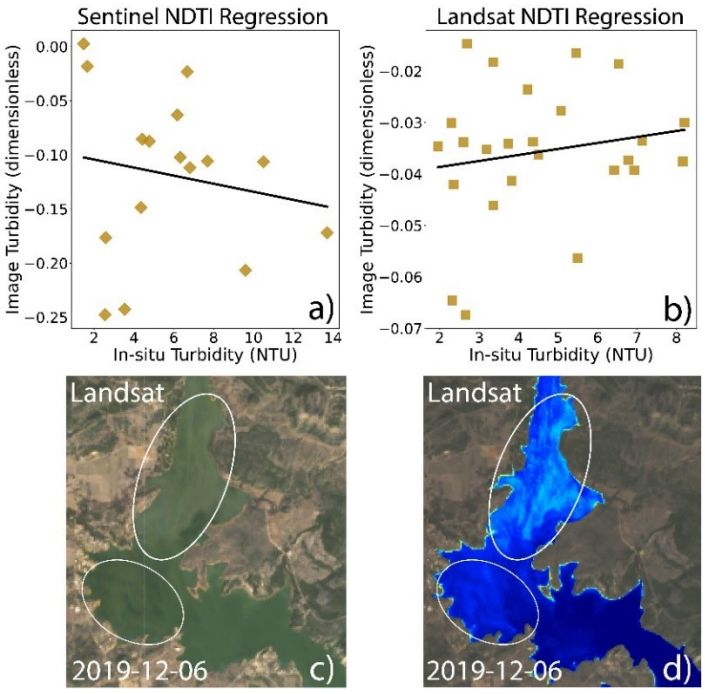

Figure 4. NDTI performance metrics where a-b) are linear regression plots of image vs. *in situ* values for Sentinel and Landsat, respectively. The below photos, c-d), show an example comparison where suspended sediment can be observed in color imagery (c) and is correctly accentuated in the NDTI image (d).

Validation results show that satellite water surface temperature is strongly correlated (R^2=0.77) with *in situ* water surface temperatures (Figure 5). Though remotely sensed water temperature only provides the water's surface uppermost "skin" temperature, its correlation with *in situ* temperatures at 0.3 m depth indicates that it can be a proxy for near-surface temperatures. Due to the unavailability of *in situ* cyanobacteria concentrations, we did not validate the BWAI product using linear regressions, and instead validated the product by checking whether the BWAI indicates positive anomalies during known periods of algal events. Our team computed the average BWAI values for the lake chain for known algal event periods, and then compared with values of times of no algal events. The product successfully indicated larger anomalies during known periods of algal events. However, it is important to note that algal mats tend to be transported to the shoreline and the BWAI might not detect these mats as the imagery was masked to 30 meters inwards from the shoreline.

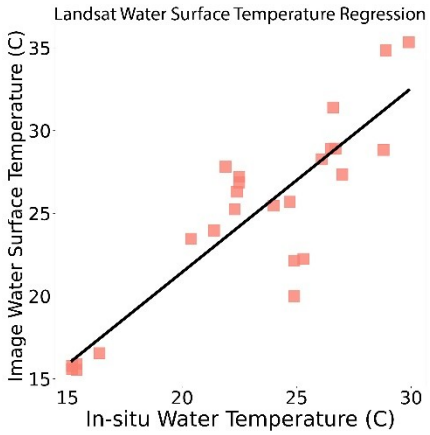

Figure 5. Plot and linear regression of Landsat measured water surface temperature (Celsius, y-axis) against *in situ* water temperature (Celsius).

Overall, satellite – *in situ* validation was limited due to the small number of matchups, which, while sufficient, would be much more robust with more matchup points. Additionally, the time difference between *in situ* sample collection and the satellite overpass likely impacts the validation results. However, each product validation suggests that these products are sufficient for capturing environmental trends and anomalies, which is crucial for data interpretation by the end users.

4.1.2 Historical Environmental Products

To provide the partners with products that will aid in future decision making, our team analyzed all available satellite imagery to date. We reduced image collections to a single image to understand spatial statistics for seasonal variations. Each product was separated into two image collections for each year, then an average value was calculated for each seasonal image collection using every value available for each pixel location. Dates used to split up the years into "seasons" were May 1st and November 1st. We applied such image collection reductions to the *chl-a*, NDTI, and surface temperature products. However, for *chl-a*, we also reduced the MDN *chl-a* products using the median of every value available for each pixel location, as the median can be less sensitive to variability when outliers are present. Additionally, for the BWAI product, we found reducing the image collection by summing all values for every pixel location captured anomaly hotspot locations best and minimized environmental noise. Rather than creating seasonal reduction maps for the BWAI we focused on summer months, June through September, when the Highland Lakes are most prone to algal events. In addition to investigating spatial patterns, we computed time series of each product to track changes for water bodies through time. To create the time series, we computed the average and median for all pixels within each water body for each image. As the temporal resolution for each platform can be sparse during cloudy periods, we chose to use a non-uniform temporal step size for the time series, where the step size is every available image date for each Earth observation platform. Consequently, the time series data provided useful references when deciding what minimum and maximum values to use for stretching the contrast of every product. We found the most effective visualization was using the baseline average low and high time series values.

4.1.3 Historical Product Observations

To detect the temporal pattern of *chl-a* in the Highland Lakes chain, *chl-a* was retrieved from Landsat 8 OLI by the Sentinel-2 MDN from 2013 to 2021. Figure 6 shows the median value for each individual lake with one standard deviation of uncertainty. *Chl-a* has a consistent seasonality in all six lakes, generally peaking in July and August in the summer. Additionally, lakes upstream have relatively higher *chl-a* compared to lakes downstream, which can possibly be explained by different land use and corresponding different nutrients loading to lakes. Trajectories for *chl-a* across the six study lakes also show different temporal patterns. *Chl-a* in Lakes Travis and Austin were relatively stable, compared to other lakes. Notably, our products also captured *chl-a* spikes in the summer of 2015 and 2016 at the Lake Austin, which corresponds to reported severe droughts. Bellinger et al., 2018 suggested that drought-induced discharge reductions promoted earlier onset of warm, stratified waters, lower hypolimnetic oxygen concentrations, and longer hydraulic residence, which altered the peak, amount, and timing of algal bloom. Additionally, *chl-a* spiked in the summer of 2019 at Lake Travis, and Lake Austin also responded to reported algae-related canine deaths and cyanotoxin (Manning et al., 2020).

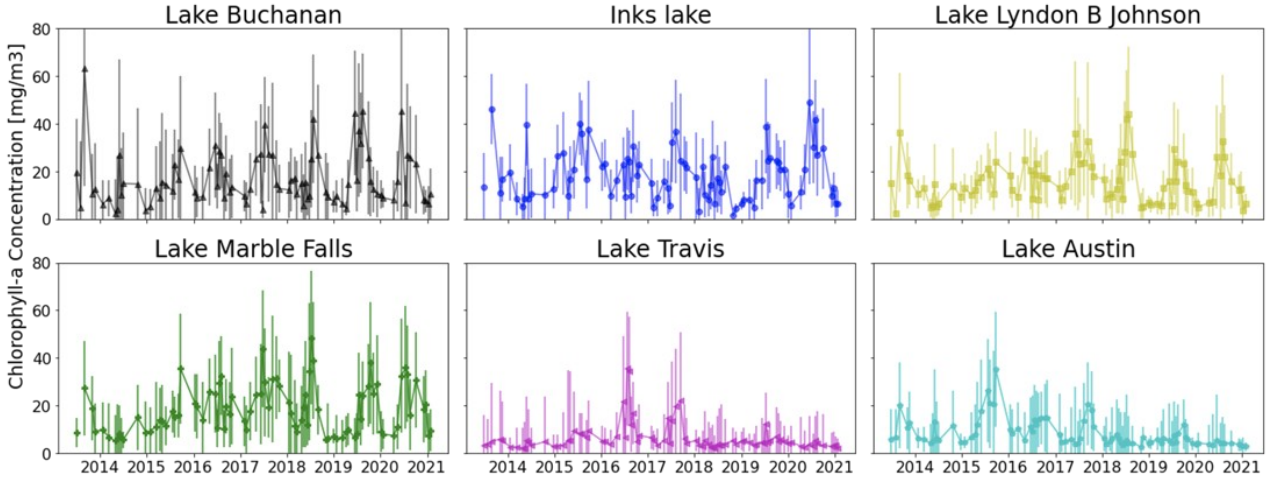

Figure 6. OLI-retrieved median c*hl-a* concentration (mg/m^3) with one standard deviation of uncertainty for the six lakes from upstream to downstream in the highland lakes chain.

We investigated water surface temperature time series for Lake Travis and Lady Bird Lake, two of the recently algae-prone water bodies, as the partners are concerned whether climate change is significantly impacting water body temperatures. To analyze water temperature over time, we calculated an average value for each water body for every image acquired since 2013. The resulting water surface temperature time series illustrate how the product captures annual variability and patterns of temperature differences between the lakes (Figure A1). When inspecting the data to investigate changes over time, there appeared to be no apparent trends to indicate that climate change has significantly affected the temperature of the water bodies yet. However, not enough data were acquired to understand long term trends or changes. In addition, time series analyses were performed for NDTI imagery, for both Landsat and Sentinel. The Landsat-derived time series for turbidity in Lake Travis ranges from spring of 2013 to fall of 2021; however, the annual variability proved difficult to infer due to the sensor's coarse

temporal resolution (Figure A2). However, when performing a time series analysis with Sentinel NDTI imagery, with finer temporal resolution, clear annual variability can be observed (Figure A3). These examples of investigating historical observations help illustrate the usefulness of the tool to the partners, as they will have the ability to implement these statistical reductions to any available time frame or image collection of choice.

4.2 Future Work

Ultimately, accurate detection of algal events is a difficult task with many inherent sources of error and noise. Future work should incorporate data that have gone through atmospheric processing specifically for ocean color studies, which will have less noise and better represent the reflectance spectra directly above the water surface. Utilization of sensors such as Sentinel-3 Ocean and Land Color Instrument (OLCI), with bands specific to cyanobacteria absorption features, would also be beneficial and allow for more quantitative and accurate products. Most notably, incorporation of fluid dynamics, climate, land use, and nutrient flux models would allow for a predictive modelling component important to the end users. Future work will likely be best outside of the Google Earth Engine environment, as the current data acquisition limitations in Google Earth Engine hinder the use of upper-tier ocean color products.

5. Conclusions

The detection of algal events in inland waters involves many factors. This project aimed to create a tool in which many of these factors could be accessed and observed using satellite data. Our team approached this project with a range of tools including MDN machine learning and multiple spectral indices to measure *chl-a*, water surface temperature, cyanobacteria growth, and turbidity. We validated the resulting products with *in situ* measurements and found them to be sufficiently accurate to aid in detecting near-future algae events. They also contribute to a greater understanding of the Highland Lakes system; proxies such as turbidity and surface temperature, although directly related to algal events, also provide helpful context for other environmental conditions impacted by land use. Analysis of imagery dating back to 2013 indicated that *chl-a* concentration is increasing in some lakes and peaked in the 2015 and 2016 drought years. Additionally, seasonal trends in turbidity and temperature became apparent. The MDN successfully estimated *chl-a* values with both Sentinel-2 and Landsat 8 sensors. Estimates of water surface temperatures measured by Landsat 8 TIRS were also accurate. While turbidity and cyanobacteria growth indices seemed to perform poorly, manual comparisons with color imagery and dates of known algal events suggest that time gaps between image acquisition and *in situ* collection may be responsible. The tool that incorporated these products provides users with an improved understanding of their water systems and the algal events that threaten them.

6. Acknowledgments

Our NASA DEVELOP Team would like to thank the following for support and aid in this project:

- Lisa Benton, Lower Colorado River Authority: Senior Aquatic Biologist

- Dr. Brent Bellinger, City of Austin Department of Watershed Protection: Aquatic Ecologist
- Dr. Schonna Manning, University of Texas Department of Molecular Sciences: Research Assistant Professor of Molecular Biosciences
- Teresa Lutes, Austin Water Utility: Managing Engineer
- Dr. Erin Urquhart, NASA Goddard Space Flight Center: Science Advisor
- Dr. Christine Lee, NASA Jet Propulsion Laboratory: Science Advisor
- Dr. David Hondula, Arizona State University: Advisor

This material contains modified Copernicus Sentinel data (2021), processed by ESA.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Aeronautics and Space Administration.

This material is based upon work supported by NASA through contract NNL16AA05C.

7. Glossary

Algae bloom – Growth colony of cyanobacteria distributed throughout the water column
Algae proliferation – Growth colony of cyanobacteria floating as mats on the water surface
Algal event – Event denoting growth of cyanobacteria, blooms, or proliferations
Cyanobacteria – microscopic organisms living throughout the water column in both fresh- and saltwater, also known as blue-green algae
Cyanotoxin – toxins produced by cyanobacteria, especially during blooms and/or proliferations
Depression Factor (F_S) – BWAI variable dependent on relationships between the green, red, and near-infrared band reflectance's; necessary to depress values that interfere with product accuracy
Earth observations – data from satellites, sensors, and ground-based techniques that collect information about physical, chemical, and biological systems on Earth
Google Earth Engine – platform for cloud-based geospatial analysis and Earth observation data repository
Machine learning – type of artificial intelligence in which a computer system utilizes algorithms and statistical models to identify patterns in data and make decisions
Neural network – subset of machine learning, in which underlying relationships in data are recognized via artificial neurons in a process inspired by the human brain
Signal modulation factor (F_C) – BWAI variable which prevents suspended sediment from interfering with cyanobacteria detections

8. References

Bailey, S. W., & Werdell, P. J. (2006). A multi-sensor approach for the on-orbit validation of ocean color satellite data products. *Remote Sensing of Environment, 102*(1), 12–23. https://doi.org/10.1016/j.rse.2006.01.015

Bishop, C. M. (1994). *Mixture Density Networks* (No. 94/004; Neural Computing Research Group Report). Aston University. https://publications.aston.ac.uk/id/eprint/373/1/NCRG_94_004.pdf

Beck, R., Zhan, S., Liu, H., Tong, S., Yang, B., Xu, M., Ye, Z., Huang, Y., Shu, S., Wu, Q., Wang, S., Berling, K., Murray, A., Emery, E., Reif, M., Harwood, J., Young, J., Nietch, C., Macke, D., … Su, H. (2016). Comparison of satellite reflectance algorithms for estimating chlorophyll-a in a temperate reservoir using coincident hyperspectral aircraft imagery and dense coincident surface observations. *Remote Sensing of Environment*, *178*, 15–30. https://doi.org/10.1016/j.rse.2016.03.002

Bellinger, B. J., Richter, A., Porras, A., & Davis, S. L. (2018). Drought and management effects on biophysicochemistry in a rapidly-flushed reservoir. *Lake and Reservoir Management, 34*(2), 182–198. https://doi.org/10.1080/10402381.2017.1384770

Bouma-Gregson, K., Kudela, R. M., & Power, M. E. (2018). Widespread anatoxin-a detection in benthic cyanobacterial mats throughout a river network. *PLOS One, 13*(5), e0197669. https://doi.org/10.1371/journal.pone.0197669

CERES Science Team. (2015). CERES Level 3 SYN1DEGDAYTerra+Aqua netCDF file, Edition 3A.
NASA Atmospheric Science Data Center, accessed 11 February 2016. doi://10.5067/Terra+Aqua/CERES/SYN1degDAY_L3.003A

Chen, J., Zhu, W.-N., Tian, Y. Q., & Yu, Q. (2017). Estimation of Colored Dissolved Organic Matter From Landsat-8 Imagery for Complex Inland Water: Case Study of Lake Huron. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(4), 2201–2212. https://doi.org/10.1109/TGRS.2016.2638828

Dall'Olmo, G., & Gitelson, A. A. (2006). Effect of bio-optical parameter variability and uncertainties in reflectance measurements on the remote estimation of chlorophyll-a concentration in turbid productive waters: Modeling results. *Applied Optics*, *45*(15), 3577. https://doi.org/10.1364/AO.45.003577

Elhag, M., Gitas, I., Othman, A., Bahrawi, J., & Gikas, P. (2019). *Assessment of Water Quality Parameters Using Temporal Remote Sensing Spectral Reflectance in Arid Environments, Saudi Arabia*. 14. https://doi.org/10.3390/w11030556

Fetscher, A. E., Howard, M. D. A., Stancheva, R., Kudela, R. M., Stein, E. D., Sutula, M. A., Busse, L. B., & Sheath, R. G. (2015). Wadeable streams as

widespread sources of benthic cyanotoxins in California, USA. *Harmful Algae, 49*, 105–116. https://doi.org/10.1016/j.hal.2015.09.002

Gordon, H. R., Brown, O. B., Evans, R. H., Brown, J. W., Smith, R. C., Baker, K. S., & Clark, D. K. (1988). A semianalytic radiance model of ocean color. *Journal of Geophysical Research*, *93*(D9), 10909. https://doi.org/10.1029/JD093iD09p10909

Lacaux, J. P., Tourre, Y. M., Vignolles, C., Ndione, J. A., & Lafaye, M. (2007). Classification of ponds from high-spatial resolution remote sensing: Application to Rift Valley Fever epidemics in Senegal. *Remote Sensing of Environment*, 9. https://doi.org/10.1016/j.rse.2006.07.012

Lopez, C. B., Jewett, E. B., Dortch, Q., Walton, B. T., & Hudnell, H. K. (2008). Scientific assessment of freshwater harmful algal blooms. Interagency Working Group on Harmful Algal Blooms, Hypoxia, and Human Health of the Joint Subcommittee on Ocean Science and Technology. www.whoi.edu/fileserver.do?id=41023&pt=10&p=19132

Lunetta, R. S., Schaeffer, B. A., Stumpf, R. P., Keith, D., Jacobs, S. A., & Murphy, M. S. (2015). Evaluation of cyanobacteria cell count detection derived from MERIS imagery across the eastern USA. *Remote Sensing of Environment*, *157*, 24–34. https://doi.org/10.1016/j.rse.2014.06.008

Main-Knorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U., & Gascon, F. (2017). Sen2Cor for Sentinel-2. In L. Bruzzone, F. Bovolo, & J. A. Benediktsson (Eds.), *Image and Signal Processing for Remote Sensing XXIII* (p. 3). SPIE. https://doi.org/10.1117/12.2278218

Manning, S. R., Perri, K. A., & Bellinger, B. J. (2020). Bloom announcement: First reports of dog mortalities associated with neurotoxic filamentous cyanobacterial mats at recreational sites in Lady Bird Lake, Austin, Texas. Data in Brief, 33, 106344. https://doi.org/10.1016/j.dib.2020.106344

Matthews, M. W., Bernard, S., & Robertson, L. (2012). An algorithm for detecting trophic status (chlorophyll-a), cyanobacterial-dominance, surface scums and floating vegetation in inland and coastal waters. *Remote Sensing of Environment*, *124*, 637–652. https://doi.org/10.1016/j.rse.2012.05.032

Mobley, C. D., Werdell, J., Franz, B., Ahmad, Z., & Bailey, S. (2016). *Atmospheric Correction for Satellite Ocean Color Radiometry* (p. 85). NASA Ocean Biology Processing Group.

Ogashawara, I., Kiel, C., Jechow, A., Kohnert, K., Ruhtz, T., Grossart, H.-P., Hölker, F., Nejstgaard, J. C., Berger, S. A., & Wollrab, S. (2021). The Use of Sentinel-2 for Chlorophyll-a Spatial Dynamics Assessment: A Comparative Study on

Different Lakes in Northern Germany. *Remote Sensing*, *13*(8), 1542. https://doi.org/10.3390/rs13081542

Otsu, N. (1979). A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, *9*(1), 62–66. https://doi.org/10.1109/TSMC.1979.4310076

Page, B. P., Olmanson, L. G., & Mishra, D. R. (2019). A harmonized image processing workflow using Sentinel-2/MSI and Landsat-8/OLI for mapping water clarity in optically variable lake systems. *Remote Sensing of Environment*, *231*, 111284. https://doi.org/10.1016/j.rse.2019.111284

Pahlevan, N., Smith, B., Schalles, J., Binding, C., Cao, Z., Ma, R., Alikas, K., Kangro, K., Gurlin, D., Hà, N., Matsushita, B., Moses, W., Greb, S., Lehmann, M. K., Ondrusek, M., Oppelt, N., & Stumpf, R. (2020). Seamless retrievals of chlorophyll-a from Sentinel-2 (MSI) and Sentinel-3 (OLCI) in inland and coastal waters: A machine-learning approach. *Remote Sensing of Environmen*t, 240, 111604. https://doi.org/10.1016/j.rse.2019.111604

Smith, B., Pahlevan, N., Schalles, J., Ruberg, S., Errera, R., Ma, R., Giardino, C., Bresciani, M., Barbosa, C., Moore, T., Fernandez, V., Alikas, K., & Kangro, K. (2021). A Chlorophyll-a Algorithm for Landsat-8 Based on Mixture Density Networks. *Frontiers in Remote Sensing*, 1, 5. https://doi.org/10.3389/frsen.2020.623678

Vermote, E., Roger, J. C., Franch, B., & Skakun, S. (2018). LaSRC (Land Surface Reflectance Code): Overview, application and validation using MODIS, VIIRS, LANDSAT and Sentinel 2 data's. *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 8173–8176. https://doi.org/10.1109/IGARSS.2018.8517622

Zhao, Y., Liu, D., & Wei, X. (2020). Monitoring cyanobacterial harmful algal blooms at high spatiotemporal resolution by fusing Landsat and MODIS imagery. *Environmental Advances*, *2*, 100008. https://doi.org/10.1016/j.envadv.2020.100008

Zhou, S., Kan, P., Silbernagel, J., & Jin, J. (2020). Application of Image Segmentation in Surface Water Extraction of Freshwater Lakes using Radar Data. *ISPRS International Journal of Geo-Information*, *9*(7), 424. https://doi.org/10.3390/ijgi9070424

9. Appendices

Tables A1 & A2. Sensor specifications for Landsat 8 OLI (**Table A1, left**) and Sentinel-2 A and B MSI (**Table A2, right**), the main platforms utilized. Spectral range defines wavelengths within the full width at half maximum (FWHM) range. Ground Sample Distance (GSD) represents spatial resolution in meters.

Landsat 8 OLI			
Band Number	**Band Name**	**Spectral Range (nm)**	**GSD (m)**
1	Coastal aerosol	433–453	30
2	Blue	450–515	30
3	Green	525–600	30
4	Red	630–680	30
5	NIR	845–885	30
6	SWIR 1	1560–1660	30
7	SWIR 2	2100–2300	30
8	Panchromatic	500–680	15
9	Cirrus	1360–1390	30
10	TIRS 1	10600-11200	100
11	TIRS 2	11500-12500	100

Sentinel-2 MSI			
Band Number	**Band Name**	**Spectral Range (nm)**	**GSD (m)**
1	Coastal aerosol	433-453	60
2	Blue	458-523	10
3	Green	543-578	10
4	Red	650-680	10
5	Vegetation red edge	698-713	20
6	Vegetation red edge	733-748	20
7	Vegetation red edge	773-793	20
8	NIR	785-900	10
8a	Narrow NIR	855-875	20
9	Water vapor	935-955	60
10	SWIR – Cirrus	1360-1390	60
11	SWIR	1565-1655	20
12	SWIR	2100-2280	20

Table A3. Validation metrics for image vs. *in situ* comparisons.

	P-value	**Pearson Coefficient**	**R^2**	**RMSE**	**RMSLE**	**MAPE**	**Bias**	**MAE**
Landsat MDN	5.0e-4	0.46	0.6	9.15	0.32	4.56	0.93	1.68

Sentinel MDN	8.0e-5	0.85	0.73	5.33	0.238	5.95	1.42	1.52
Landsat FLHV	0.3	0.22	0.09	N/A	N/A	N/A	N/A	N/A
Sentinel 2BDA	0.3	0.28	0.08	N/A	N/A	N/A	N/A	N/A
Landsat NDTI	0.41	0.17	0.029	N/A	N/A	N/A	N/A	N/A
Sentinel NDTI	0.54	-0.17	-0.029	N/A	N/A	N/A	N/A	N/A
Landsat Temperature	8.0e-9	0.88	0.77	3.3	0.055	1.64	0.935	1.11

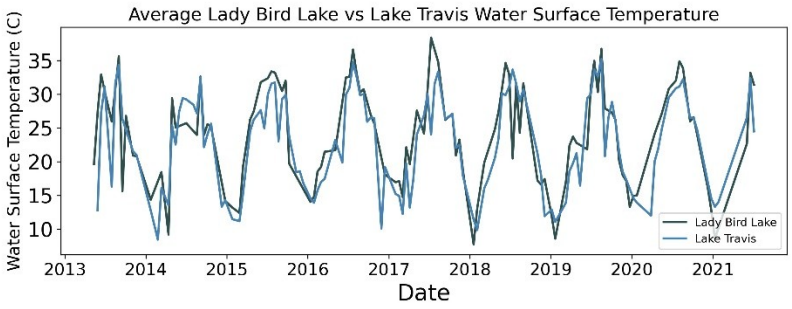

Figure A1. Times series chart of average water body temperature for Lady Bird Lake and Lake Travis from spring of 2013 to fall of 2021. Captured is the annual variability of water temperature between the two lakes as well as common differences in temperatures between the lakes.

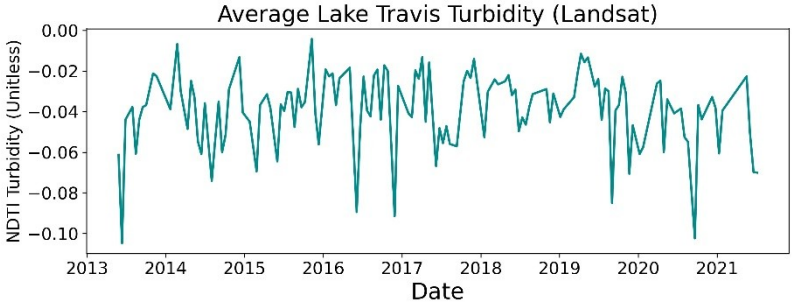

Figure A2. Time series of average turbidity (NDTI) values for Lake Travis from 2013 to 2021.

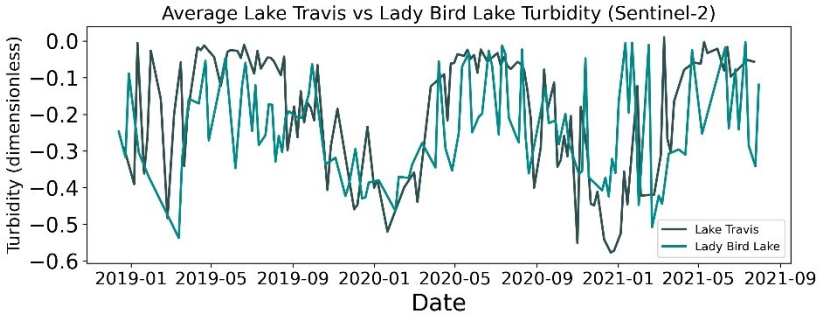

Figure A3. Time series of average turbidity (NDTI) values for Lady Bird Lake and Lake Travis from 2019 through fall 2021.

Dataset tile number: From the Worldwide Reference System (WRS), which Landsat utilizes for image tiles, the path 27 and row 39 tile cover the entire study region. The Sentinel-2 platform requires 3 tiles, Military Grid Reference System (MGRS) 14RPU, 14RNU, and 14RNV tiles, to cover the study area. The 14RNU and 14RNV tiles are part of the same vertical (polar) swath and therefore are acquired on the same day; meanwhile, the 14RPU tile scenes are acquired on different days.