ERDC/EL TR-19-20

US Army Corps
of Engineers®
Engineer Research and
Development Center

Aquatic Nuisance Species Research Program

waterquality: An Open-Source R Package for the Detection and Quantification of Cyanobacterial Harmful Algal Blooms and Water Quality

Richard A. Johansen, Molly Reif, Erich Emery, Jakub Nowosad,
Richard Beck, Min Xu, Hongxing Liu

December 2019

Environmental Laboratory

The US Army Engineer Research and Development Center (ERDC) solves the nation's toughest engineering and environmental challenges. ERDC develops innovative solutions in civil and military engineering, geospatial sciences, water resources, and environmental sciences for the Army, the Department of Defense, civilian agencies, and our nation's public good. Find out more at www.erdc.usace.army.mil.

To search for other technical reports published by ERDC, visit the ERDC online library at http://acwc.sdp.sirsi.net/client/default.

Aquatic Nuisance Species Research Program

ERDC/EL TR-19-20
December 2019

waterquality: An Open-Source R Package for the Detection and Quantification of Cyanobacterial Harmful Algal Blooms and Water Quality

Richard A. Johansen

University of Cincinnati Libraries
240B Braunstein Hall
Cincinnati, Ohio 45221, USA

Molly Reif

ERDC, JALBTCX
US Army Corps of Engineers
Kiln, MS 39556, USA

Erich Emery

Great Lakes and Ohio River Division
US Army Corps of Engineers
Cincinnati, OH 45202, USA

Jakub Nowosad

Institute of Geoecology and Geoinformatics
Adam Mickiewicz University
Krygowskiego 10
61-680 Poznan, Poland

Richard Beck

Department of Geography and GIS
University of Cincinnati
Cincinnati, OH 45221, USA

Min Xu, and Hongxing Liu

Department of Geography
The University of Alabama
204 Farrah Hall
513 University Blvd
Box 870322
Tuscaloosa, AL 35487-0322, USA

Final report

Prepared for Aquatic Nuisance Species Research Program
US Army Engineer Research and Development Center
Environmental Laboratory
Vicksburg, MS 39180-6199

Under Funding Account Code 96 X 3123; AMSCO Code 008284

Abstract

Satellite monitoring of cyanobacterial harmful algal blooms in small freshwater lakes and reservoirs remains challenging. This is partly due to the configurations and resolutions of commonly utilized satellite imagers, which are traditionally designed for large terrestrial applications. The purpose of this report is to provide an efficient methodology for the detection and quantification of harmful algal bloom indicators via remote sensing imagery utilizing the newly developed open-source R package *waterquality*. To accomplish this goal, this report uses Harsha Lake as a case study to demonstrate the use of water quality proxies (chlorophyll-a, phycocyanin, and turbidity) for the evaluation of inland lake and reservoir water quality. This package and associated manuscript were designed to assist researchers and water managers by establishing a flexible and user-friendly workflow to improve water quality monitoring.

Contents

Figures and Tables

Figures

Tables

Preface

This study was conducted for the Aquatic Nuisance Species Research Program (ANSRP) under Funding Account Code 96 X 3123, AMSCO Code 008284. The ANSRP is sponsored by Headquarters, US Army Corps of Engineers (USACE), and is assigned to the US Army Engineer Research and Development Center (ERDC) under the purview of the Environmental Laboratory (EL), Vicksburg, MS. The ANSRP Program Manager is Dr. Linda Nelson.

Appreciation is expressed to Mr. Scott Bourne and Ms. Christina Saltus for their review of this manuscript and also to Dr. Richard Beck, The University of Cincinnati, and Dr. Hongxing Liu, The University of Alabama.

The original Harsha Lake water quality study was funded by the USACE and National Aeronautics and Space Administration Glenn Research Center. The *waterquality* package was developed with funding from the USACE. Additional appreciation is expressed to the University of Cincinnati Library Research and Data Services as well as the University of Cincinnati Space Informatics Lab for their expertise and technical services.

This report was prepared under the general supervision of Mr. Mark Graves, Chief, Environmental Systems Branch; Mr. Mark D. Farr, Chief, Ecosystem Evaluation and Engineering Division; and Dr. Ilker R. Adiguzel, Director, EL.

At the time of the publication of this report, COL Teresa A. Schlosser was the Commander of ERDC, and Dr. David W. Pittman was the Director.

1 Introduction

1.1 Background

There has been a noticeable increase in both the occurrence and severity of cyanobacterial harmful algal blooms (CHABs) in water bodies across the United States (Anderson et al. 2000; Graham 2006; USEPA 2012a). CHABs have the ability to cause adverse effects on human and animal health as well as disrupt local economies (Anderson et al. 2000; Linkov et al. 2009; USEPA 2012a,b). However, consistent in situ monitoring of all at-risk water bodies in the United States would be extremely costly and labor intensive, if even possible. To compound this issue, CHABs are dynamic and respond quickly to environmental and hydrologic changes such as nutrient loads, air and water temperature, and wind speed, which may vary on timescales of days to hours (Dokulil and Teubner 2000; Hunter et al. 2008). This has given rise to the use of remote sensing methods for large-scope water quality studies. In general, satellite-based remote sensing offers a cost-effective and more consistent means of surveying large spatial extents, encompassing most US water bodies. There are multiple, available data sets from satellite sensors, many of which are free to the public or available to the Department of Defense through government contracts, that offer the spatial, spectral, and geographic coverage needed for the study of CHABs. These include such sensors as the Medium Resolution Imaging Spectrometer (MERIS), Moderate Resolution Imaging Spectroradiometer (MODIS), Landsat-8, WorldView-2, and Sentinel-2 (Augusto-Silva et al. 2014; Blondeau-Patissier et al. 2014; Beck et al. 2016; 2017; Johansen et al. 2018a; Klemas 2012; Reif 2011; Stumpf et al. 2012). To avoid confusion, it is important to note that Sentinel-2 is actually a constellation of two identical satellites (Sentinel-2A and Sentinel-2B) with identical configurations. Since these two satellites utilized identical sensors, they are treated as a single satellite throughout the workflow.

However, a major concern with this approach is satellite revisit time, which is frequently worsened by heavy cloud cover during the summer of mid-latitude regions often affected by CHABs. The issue is further compounded by the variation in each satellite imager's spatial and spectral configuration, such as their pixel sizes, band centers, and band widths (Tables 1 and 2). Therefore, it is recommended that water managers and

researchers utilize multiple sensors to increase cloud-free image acquisition as well as capture as much of the water's spectral signature as possible. Although spectral signatures are the foundation of remote sensing, high spectral resolution (many narrow spectral bands) is especially important to water quality remote sensing because water has a much lower signal-to-noise ratio compared to land observations, and algal pigments are often only detectable by narrow peaks (reflectance) and troughs (absorption) in the spectral signature (Beck et al. 2016, 2017; Johansen et al. 2018a). This is vital given that each water body contains unique mixtures of bio-chemical variables and physical characteristics. It is hypothesized that a comprehensive view would allow for precise detection and quantification of water quality for individual water bodies. To achieve this goal of precision-based water quality monitoring, the authors have developed an open-source R software package called *waterquality* (Johansen et al. 2018b; R Core Team 2017). This software package contains a continually growing list of satellite-derived algorithms for the detection and quantification of the following three common water quality proxies: chlorophyll, turbidity, and phycocyanin.

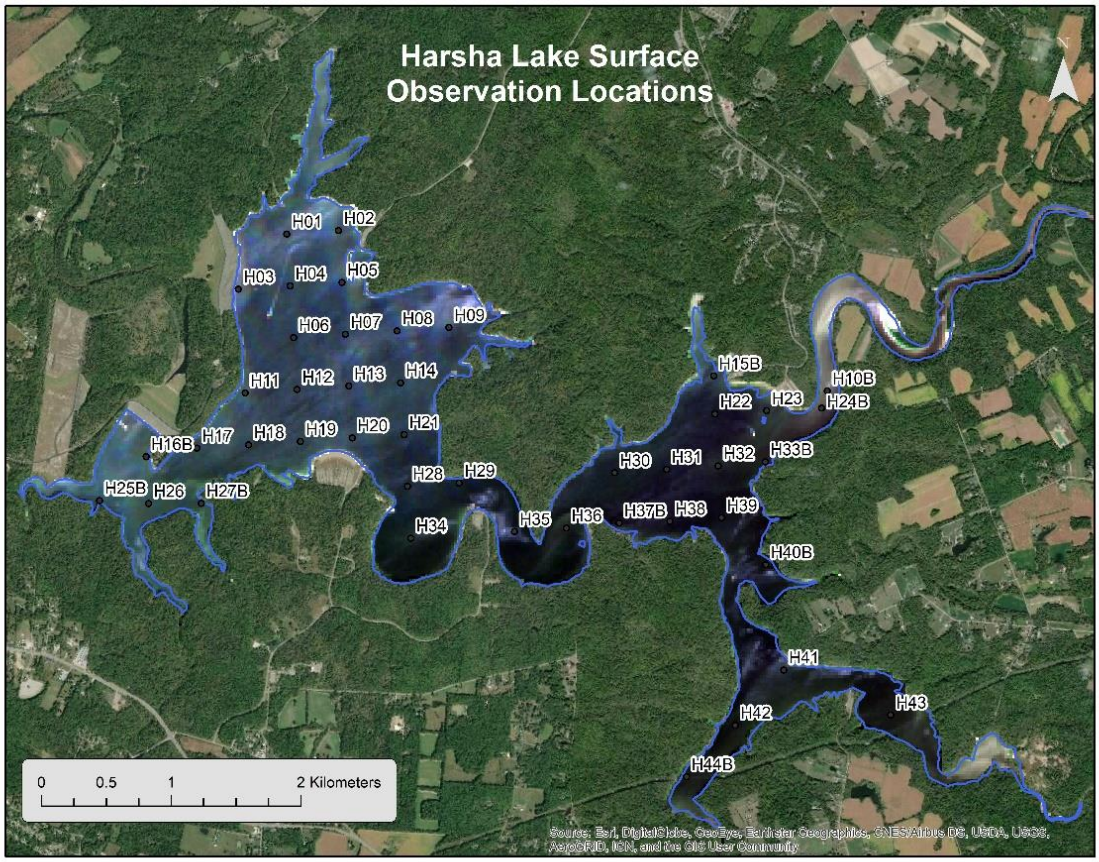
Most commonly, chlorophyll-*a* is used as the measure of chlorophyll because it is the ubiquitous photosynthetic pigment found in both toxic and non-toxic algal species, and it is helpful for the detection of *blooming* conditions as well as blooms containing mixed algae species. However, this report refers to the more generic term *chlorophyll* because researchers have utilized a variety of techniques (cell counts, sonde spectral concentrations, corrected chlorophyll-*a*, etc.) to measure chlorophyll concentrations. Phycocyanin (PC) is an important indicator of CHABs and water quality risk because PC is a cyanobacteria/blue-green algae (BGA)-specific pigment. Fortunately, PC contains a distinct spectral absorption feature centered on 620 nanometers (nm), which makes BGA/PC detectable to sensors with the appropriate spectral configuration (Schalles and Yacobi 2000; Simis et al. 2005; Randolph et al. 2008; Stumpf et al. 2016; Wozniak et al. 2016). Turbidity is the relative clarity of water, and specifically it is the backscattering caused by suspended materials in the water. As such, turbidity is not a measure of CHABs directly but has been shown to be highly correlated with chlorophyll and BGA/PC concentrations in highly productive eutrophic water bodies and can be an early warning sign (Barnes et al. 2015; Dogliotti et al. 2015; Doxaran et al. 2002, 2006; Moore 1980; Olmanson et al. 2013). Unfortunately, none of these parameters are a direct measure of toxicity, and most operational

imaging satellites do not contain the important phycocyanin spectral band associated with cyanobacteria. Some imagers, such as Landsat-8, have been found to have serious limitations for their ability to accurately quantify chlorophyll concentrations (Beck et al. 2016; Hunter et al. 2008; Stumpf et al. 2016). Despite these limitations, this research focused on these five sensors and three parameters because they still provide an effective methodology for estimating CHABs concentration (Beck et al. 2016, 2017; Johansen et al. 2018a; Randolph et al. 2008; Schalles and Yacobi 2000; Simis et al. 2005; Stumpf et al. 2016; Wynne et. al 2008; Wozniak et al. 2016).

1.2 Objective

The main goal of this report is to describe the functionality of the *waterquality* package via a case study of Harsha Lake in Southwest Ohio, United States (Figure 1). The workflow is designed to allow users to efficiently convert atmospherically corrected reflectance imagery from any of the five satellite imagers (MODIS, MERIS, Landsat-8, Sentinel-2, and Worldview-2) into a suite of well-established water quality indices and subsequently evaluate those indices against in situ observations to quantify water quality for a given body of water (Johansen 2018c).

Figure 1. Sentinel-2A image of Harsha Lake Ohio, United States, on August 8, 2016, overlaying high-resolution aerial image of surrounding landscape with the 44 surface sampling locations.



1.3 Approach

The approach of this study is presented in Chapter 2, Data and Methods.

2 Data and Methods

2.1 Imagery acquisition

A suite of imager options to choose from is provided, including MODIS, MERIS, Landsat-8, Sentinel-2, and Worldview-2. These five satellite imagers range widely in their spectral, spatial (1.8 m – 300 m), and temporal resolutions (daily to 16-day revisit times), but all have near-global coverage (Table 1). The data from these imagers are freely available to the public (except WorldView-2, which is available to Department of Defense personnel through a government contract), and require the creation and activation of an account on either the European Space Agency (ESA) Copernicus Hub (https://scihub.copernicus.eu) or the US Geologic Survey (USGS) Earth Explorer (https://earthexplorer.usgs.gov). Once activated, the user can log in to the Open Access Hub or Earth Explorer to search for available imagery based on spatial, temporal, and imagery needs at no charge.

Table 1. List and description of the imaging satellites utilized in this study, including name, spatial resolution, temporal resolution, and source.

Satellite Imager	Spatial Resolution	Temporal Resolution	Source
Worldview-2	1.8m	1.1 days	https://www.digitalglobe.com/resources/satellite-information
Sentinel-2	10-20m	5-10 days	https://sentinel.esa.int/web/sentinel/missions/sentinel-2
Landsat-8	30m	16 days	https://landsat.usgs.gov/landsat-8
MODIS	250m	1-2 days	https://modis.gsfc.nasa.gov/
MERIS (2002-2012)	300m	3 days	https://earth.esa.int/web/guest/missions/esa-operational-eo-missions/envisat/instruments/meris

2.2 Study area and surface measurements

William H. Harsha Lake, also known as East Fork Lake, is a US Army Corps of Engineers (USACE) monitored fresh water reservoir located in Southwest Ohio with a surface area of approximately 8 km^2 (Figure 1). Harsha Lake was chosen for this case study because it is a location of recent and frequently occurring algal blooms, including CHABs, and is a site of previous and ongoing research (Beck et al. 2016, 2017, 2018; Johansen et al. 2018a; Xu et al. 2018). Harsha Lake is also a source of drinking water coupled with heavy recreational use, making re-occurring blooms and their associated toxins potentially dangerous for humans and wildlife, as well as posing potentially significant impacts to the local economy (USEPA 2012a).

Given these conditions, the lake is subject to routine monitoring by the USACE, the US Environmental Protection Agency (USEPA), and the University of Cincinnati, which has provided a wealth of water quality data for this research.

Based on previous findings of Beck et al. (2016) and Johansen et al. (2018a), this study will focus on using the Sentinel-2A sensor for the water quality analysis of Harsha Lake. The imagery and coincident surface water observations were acquired on August 8, 2016. Due to the dynamic nature of CHABs, a host of measurements was collected on the same day as close to the image acquisition time as possible. The Yellow Springs Instrument (YSI) 6600 water quality sonde was used to collect surface fluoresce measurements of chlorophyll-*a* (Chl µg/L), phycocyanin (BGA-PC-reflective units [RFU]), and turbidity (Turbidity+ nephelometric turbidity units [NTU]). YSI sonde measurements were chosen over traditional water sample approaches because they offer a more consistent measurement and require less sophisticated lab equipment and technical expertise. Furthermore, the precision and accuracy of the YSI 6600 are more than sufficient for this research: Range of ~0-400 µg/L, which equates to a range 0-100 relative RFU, a detection limit of ~0.1 µg/L, resolution of 0.1 µg/L, and linearity/R^2 > 0.9999 (YSI 2003). Additional measurements collected that were not directly incorporated in this study include the following: specific conductance, pH, water temperature, dissolved oxygen, air temperature, wind speed, and Analytical Spectral Devices spectroradiometer reflectance measurements.

2.3 Atmospheric correction and image pre-processing

Currently, the ESA and USGS mainly distribute uncorrected top-of-atmosphere (ToA) reflectance or radiance products. It is highly recommended that all users perform an atmospheric correction to convert satellite imagery into bottom-of-atmosphere (BoA) reflectance before conducting imagery analysis on any of the satellite imagery. This is due to the fact that remotely sensed imagery derives the vast majority of the upwelling radiance from atmospheric scattering while only a small fraction is from the water's surface (Bernstein et al. 2012; Gao et al. 2009; Griffin and Burke 2003). Thus, an atmospheric correction should be applied to remove the atmospheric scattering effect, compensate for the weak signal from the water's surface, and subsequently improve pigment detection and concentration accuracy.

Choosing the most appropriate atmospheric correction technique can be challenging and is typically location- and user-dependent. Additionally, image quality, availability of in situ measurements, cost, and industry standards may play a role in this important decision. Currently, atmospheric correction is not automated and may require proprietary software to run specific techniques, such as Fast Line-of-sight Atmospheric Analysis of Hypercubes (FLAASH) and QUick Atmospheric Correction (QUAC), which are available in the remote sensing software, Environment for Visualizing Images (ENVI®). Ultimately, this step is left to the user to decide and implement. However, Xu et al. (2018) provide an evaluation of atmospheric correction techniques including FLAASH, QUAC, Sen2Cor, and Empirical Line Model (ELM), specifically for the detection of CHABs in Harsha Lake using Sentinel-2 imagery. Their findings demonstrate that the ELM approach is the most effective, but this approach is also the most demanding, requiring surface observations and the deployment of massive black and white tarps to be used as ground control targets. Given that this study was designed to be user friendly with minimal technical training, it was decided to use the ESA Sentinel-2-specific atmospheric correction method, Sen2Cor, because it is open-source and can be directly incorporated in the workflow using the R package s*en2r* (Main-Knorn et al. 2017; Ranghetti and Busetto 2018).

Even with the development of the *sen2r* package, the use of raw Sentinel-2 imagery can be challenging because full scenes cover vast geographic areas, making them computationally large and contain spectral bands with varying spatial resolutions (Drusch et al. 2012). Due to this, it is recommended to conduct additional image processing after atmospheric correction, such as spectral band sub-setting and spatial masking. Note that these pre-processing steps are added to the R workflow as supplemental scripts to be implemented before using the *waterquality* package. This increases the flexibility and usability of the workflow to include atmospheric correction, spatial resampling, band sub-setting, and image masking to get an image file that contains the appropriate spatial and spectral information necessary for water quality analysis. Pre-processing an image significantly reduces the size of the image, which dramatically reduces the computation time for the water quality algorithm calculations. Using this approach can reduce a full scene image from approximately 3 gigabytes (GB) to less than 1 megabytes (MB). While this case study focuses on Sentinel-2 imagery, these functions can also be utilized for other satellite imagers as needed.

The final step, before using the *waterquality* package and conducting the water quality index calculations, is to verify that all inputs are in the *standard* format according to the sensor type in Table 2. This step is necessary regardless of which pre-processing techniques and atmospheric correction methods have been applied. In addition, it ensures that all analyses, irrespective of the researcher, are conducted utilizing the same set of band configurations, resulting in comparable and reproducible results.

Table 2. The spectral band configurations of each of the satellite imagers utilized in this study, highlighting the variation in spectral resolutions from one imager to another. The letter "b" denotes which spectral band is being described.

Imager	Center (nm)	Range (nm)	Bandwidth (nm)	GSD (m)
WorldView-2				
b1	425	400–450	50	1.8
b2	480	450–510	60	1.8
b3	545	510–580	70	1.8
b4	605	585–625	40	1.8
b5	660	630–690	60	1.8
b6	725	705–745	40	1.8
b7	832.5	770–895	125	1.8
b8	950	860–1040	180	1.8
Sentinel-2				
b1	443	433-453	20	20
b2	490.5	458–523	65	20
b3	560.5	543–578	35	20
b4	665	650–680	30	20
b5	705.5	698–713	15	20
b6	740.5	733–748	15	20
b7	783	773–793	20	20
b8	842.5	785–900	115	20
b9 (b8a)	865	855–875	20	20
Landsat-8				
b1	440	430–450	20	30
b2	480	450–510	60	30
b3	560	530–590	60	30

Imager	Center (nm)	Range (nm)	Bandwidth (nm)	GSD (m)
b4	655	640–670	30	30
b5	865	850–880	30	30
MODIS				
b1	645	620–670	50	250
b2	858.5	841–876	35	250
MERIS				
b1	407	402–412	10	300
b2	443	438–448	10	300
b3	490	485–495	10	300
b4	510	505–515	10	300
b5	560	555–565	10	300
b6	620	615–625	10	300
b7	665	660–670	10	300
b8	681.5	678–685	7	300
b9	709	704–714	10	300
b10	753.5	750–757	7	300
b11	759.5	757–762	5	300
b12	779.5	772–787	15	300
b13	865	855–875	20	300
b14	885	880–890	10	300
b15	900	895–905	10	300

To assist with this process, it was important to include the output of the pre-processing stage for the case study (Figure 2). For this example, the Sentinel-2 image was downloaded from the ESA Copernicus Open Access Hub and converted from ToA reflectance to BoA reflectance using the *sen2r* package in R. This package requires minimal technical knowledge to run once installed. In fact, converting a Sentinel-2 image from ToA to BoA only requires one function (*sen2cor*) and one argument (input file folder ending in .SAFE).

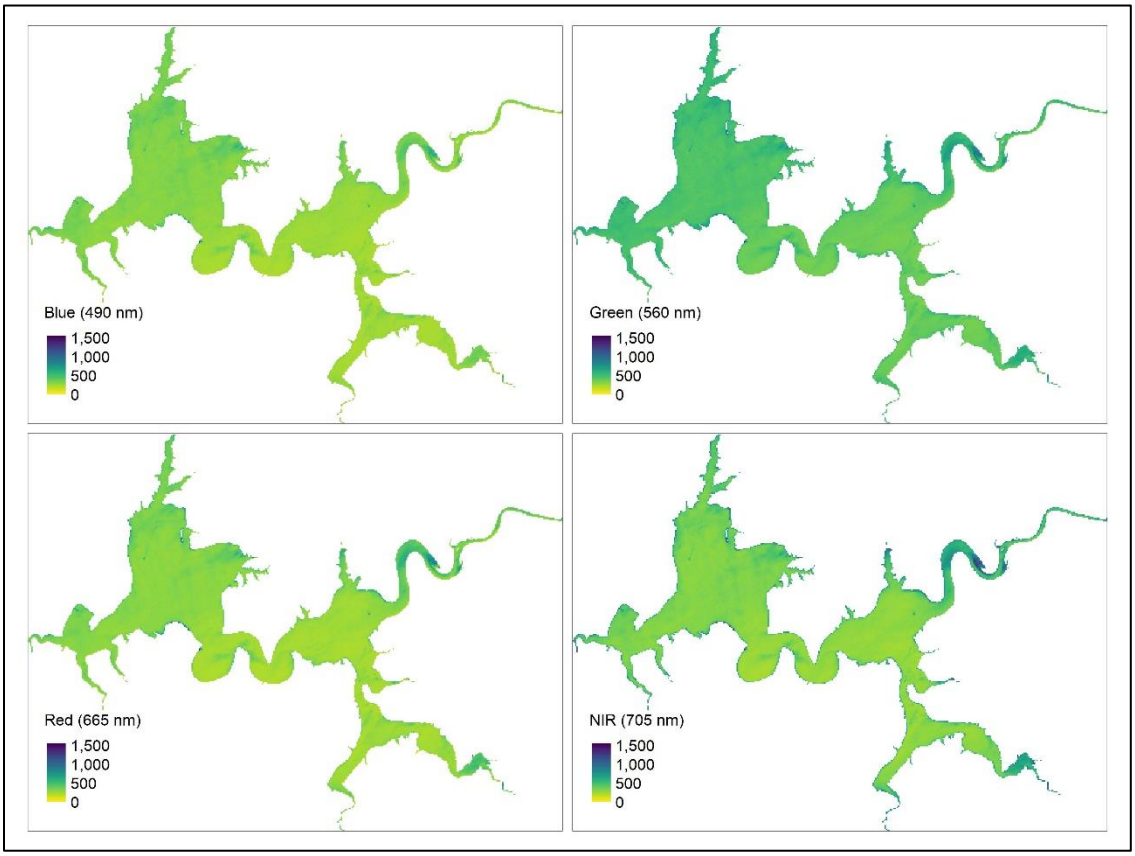
There are other functions contained in the *sen2r* package, including a graphical user interface that allows users to download and pre-process imagery directly, but since this package is still in beta, it was decided to

incorporate only the single *sen2cor* function in this workflow. (For those interested in exploring this package further, please reference https://ranghetti.github.io/sen2r.) The output of the *sen2r* process is a series of folders and JPEG 2000 files organized using an .XML file. A small series of steps as outlined below allows for the conversion of this complicated layered folder structure to be reformatted into a single stacked .TIFF image. A full data flow diagram can be found in the Appendix (A-5).

- Define the image directory to create a list of the JPEG 2000 file names.
- Create raster template using an existing band so all bands have the same pixel dimensions.
- Option 1 (recommended): use an Environmental Systems Research Institute (ESRI) Shapefile (.shp) or GeoPackage (.gpkg) file to clip or mask image to an area of interest (ESRI 1998; OGC, n.d.).
- Option 2: use full scene for water quality analysis. Useful if evaluating multiple water bodies within a single scene.
- Create subset of only the required spectral bands as stated in Table 2.
- Finally, stack the raster and save image as .TIFF.

Following the steps above, it was possible to reduce the Sentinel-2 full scene reflectance imagery and associated data from 1.18 GB of storage to 519 KB for the stacked .TIFF image of Harsha Lake. Thus, the pre-processing steps reduced the imagery size by 99.96% and significantly reduced requirements for storage as well as computation time of running the water quality algorithms. The reduction in image size is the driving force for the recommendation above, and although these steps are not required, users should expect significant increases in computation time with increases in image size. A subset of the final results of the pre-processing steps are displayed in Figure 2, where each plot corresponds to the commonly utilized Sentinel-2 bands: blue, green, red, and near infrared.

Figure 2. Four important Sentinel-2 spectral bands commonly used in the calculation of many water quality algorithms contained in the *waterquality* package suite.



3 Image Analysis Using *waterquality*

Once the data have been pre-processed and atmospherically corrected, the water quality algorithms can be applied using the *waterquality* package. This package makes use of a suite of well-established algorithms designed for the evaluation of water quality using remotely sensed imagery. These algorithms are able to detect the spectral response of only the uppermost portion of the water column (up to approximately 1 m below the surface). Specifically, these algorithms were developed for the detection and quantification of the pigments associated with CHABs, which include the various measurements of chlorophyll/chlorophyll-*a*, the cyanobacteria specific pigment phycocyanin, and turbidity, used as an indication of water clarity (Mishra et al. 2014; Randolph et al. 2008; Simis et al. 2005; Stumpf et al. 2016; Wynne et al. 2008). Studies have demonstrated that these parameters can strongly co-vary, but the degree is influenced by water biochemistry, geography, and atmospheric conditions (Dokulil and Teubner 2000; Stumpf et al. 2016; Wynne et al. 2012, 2015). Localized variations and the subsequent effect on the algorithm's performances are the justification for allowing the user to choose a single algorithm from a comprehensive list of available algorithms. As such, this package aids water managers and researchers in the move towards near real-time monitoring of algal blooms by providing a customizable and user-friendly approach for analyzing remotely sensed imagery for CHABs detection and quantification.

3.1 Water quality algorithms

Currently, the package contains 45 algorithms that can be applied for the detection of three parameters with each one being labeled by the parameter of the original author's intent (Table 3 with references in the Appendix [A-1]). However, given the co-variance of these parameters and overlap between spectral bands, it may be appropriate to use one algorithm that was originally developed for one parameter (i.e., BGA/PC) for the detection of another (i.e, chlorophyll-*a*). Note that not all algorithms are able to be applied to all sensors. This limitation is due to the spectral configurations of each sensor displayed in Table 2. Careful documentation has been made so that each algorithm is searchable within the package, using a "?" in front of the algorithm name (*?Am092Bsub*), which provides the algorithm calculation and reference to the original paper. The output of these algorithms are in relative index values and not the actual estimated concentrations of chlorophyll (µg/L), phycocyanin

(RFU), or turbidity (NTU) values. Relative index values can be easily converted to estimated concentration values using in situ measurements and will be described in more detail in the algorithm evaluation and model validation section.

Table 3. Comprehensive list of water quality algorithms, associated water quality parameter, and band arithmetic. Reference to the original source publication is found in the Appendix (A-1).

Water Quality Algorithm	Water Quality Parameter	Original Band Calculation (w = wavelength in nm)
AI10SABI	chlorophyll	(w857 - w644) / (w458 + w529)
Am092Bsub	chlorophyll	(w681 - w665)
Am09KBBI	BGA/PC	(w686 - w658) / (w686 + w658)
Be162B643sub629	BGA/PC	(w644 - w629)
Be162B700sub601	BGA/PC	(w700 - w601)
Be162BsubPhy	BGA/PC	(w715 - w615)
Be16FLHblue	chlorophyll	(w529) - (w644 + (w458 - w644))
Be16FLHBlueRedNIR	BGA/PC	(w658) - (w857 + (w458 - w857))
Be16FLHGreenRedNIR	BGA/PC	(w658) - (w857 + (w558 - w857))
Be16FLHviolet	chlorophyll	(w529) - (w644 + (w429 - w644))
Be16FLHVioletRedNIR	BGA/PC	(w658) - (w857 + (w444 - w857))
Be16MPI	BGA/PC	((w615) - (w601) - (w644 - w601))
Be16NDPhyl	BGA/PC	(w700 - w622) / (w700 + w622)
Be16NDPhyl644over615	BGA/PC	(w644 - w615) / (w644 + w615)
Be16NDPhyl644over629	BGA/PC	(w644 - w629) / (w644 + w629)
Be16NDTIblue	chlorophyll	(w658 - w458) / (w658 + w458)

Water Quality Algorithm	Water Quality Parameter	Original Band Calculation (w = wavelength in nm)
Be16NDTIviolet	chlorophyll	(w658 - w444) / (w658 + w444)
Be16Phy2BDA644over629	BGA/PC	(w644 / w629)
Da052BDA	BGA/PC	(w714 / w672)
De933BDA	chlorophyll	(w600 - w648 - w625)
Gi033BDA	chlorophyll	((1 / w672) - (1 / w715)) * (w757)
Go04MCI	BGA/PC	(w709 - w681 - (w753 - w681))
HU103BDA	BGA/PC	(((1 / w615) - (1 / w600)) - w725)
Kn07KIVU	chlorophyll	(w458 - w644) / (w529)
Ku15PhyCI	BGA/PC	-1 * (w681 - w665 - (w709 - w665))
Ku15SLH	BGA/PC	((w715) - (w658) + (w715 - w658))
MI092BDA	BGA/PC	(w700 / w600)
MM092BDA	BGA/PC	(w724 / w600)
MM12NDCI	chlorophyll	(w715 - w686) / (w715 + w686)
MM12NDCIalt	BGA/PC	((w700 - w658) / (w700 + w658))
MM143BDAopt	BGA/PC	((1 / w629) - (1 / w659)) * (w724)
SI052BDA	BGA/PC	(w709 / w620)
SM122BDA	BGA/PC	(w709 / w600)
SY002BDA	BGA/PC	(w650 / w625)
TurbBe16GreenPlusRedBothOverViolet	turbidity	((w558 + w658) / w444)
TurbBe16RedOverViolet	turbidity	(w658 / w444)
TurbBow06RedOverGreen	turbidity	(w658 / w558)
TurbChip09NIROverGreen	turbidity	(w857 / w558)
TurbDox02NIRoverRed	turbidity	(w857 / w658)
TurbFrohn09GreenPlusRedBothOverBlue	turbidity	((w558 + w658) / w458)

Water Quality Algorithm	Water Quality Parameter	Original Band Calculation (w = wavelength in nm)
TurbHarr92NIR	turbidity	(w857)
TurbLath91RedOverBlue	turbidity	(w658 / w458)
TurbMoore80Red	turbidity	(w658)
Wy08CI	BGA/PC	-1 * ((w686) - (w672) - (w715 - w672))
Zh10FLH	chlorophyll	(w686) - (w715 + (w672 - w715))

3.2 Water quality index calculation function

The main function of this package is called *wq_calc(),* which calculates water quality indices by using a reflectance raster stack, or multi-band raster layer, as an input, user-defined algorithm(s) selection, and satellite configuration selection corresponding to the following three variables: *raster_stack*, *alg*, and *sat*.

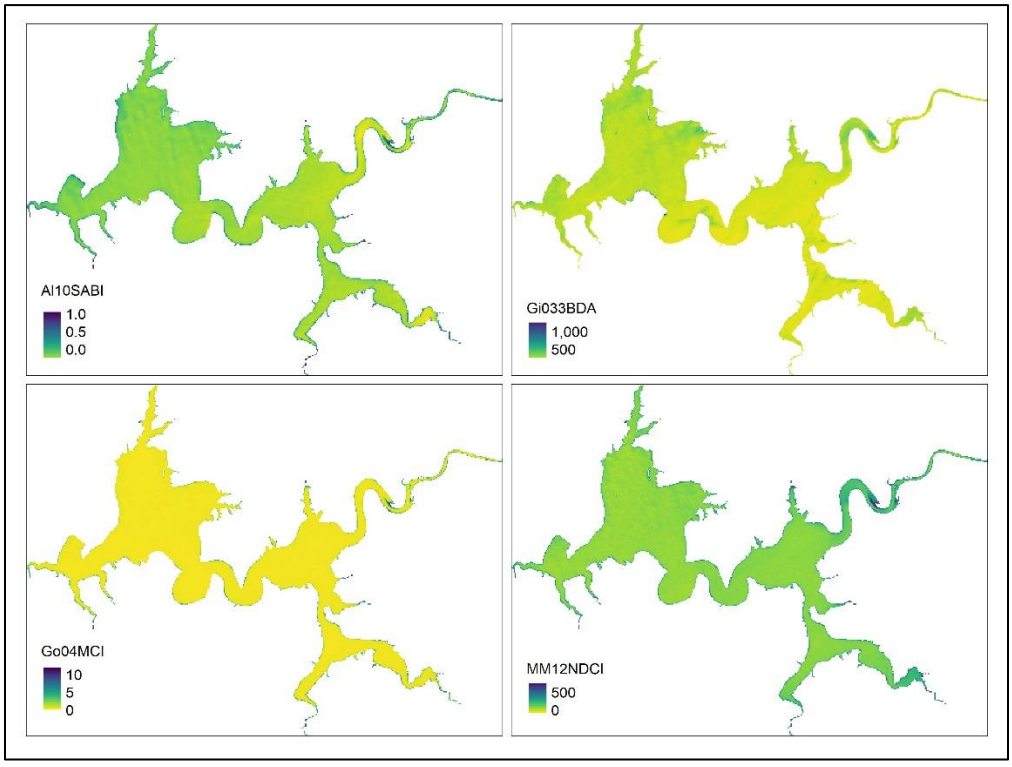
- *raster_stack* - The input reflectance image to be used in band algorithm calculation
- *alg* - Determines the algorithm to be utilized
  - Single Algorithm – "*Zh10FLH"*
  - Multiple Algorithm – c("Zh10FLH, Wy08CI, Am092BSub")
  - Type of Algorithm – "phycocyanin"
  - All Possible Algorithms (default) – "all"
- *sat* - Determines the appropriate spectral configuration associated with the algorithm to be calculated from a predefined list
  - WorldView-2
  - Sentinel-2
  - Landsat-8
  - MODIS
  - MERIS

All algorithms are calculated using the equations listed in Table 3 and the spectral band centers as described in Table 2. Consideration was given so that each band's spectral center was as close as possible to the original algorithm's wavelength. The band arithmetic for each sensor can be examined by opening the ***algorithms.R*** file inside the *waterquality* package. Due to the spectral configurations of Sentinel-2, 24 of the 45 algorithms are able to be calculated for the three water quality parameters.

To demonstrate the comprehensive nature of this tool, all 24 algorithms were calculated, and four are graphed in Figure 3. Each algorithm is calculated using a unique equation, which results in a wide variety of value ranges displayed in Figure 3. At this point in the process, these values are referred to as relative indices because the values do not reflect the values associated with the three water quality parameters. For example, high values for one index may correlate with higher chlorophyll-*a* concentrations, while another might represent an inverse relationship.

Predictive models have been developed using synthetic imagery, which apply these algorithms for multiple lakes over multiple dates, but due to the dynamic nature of CHABs and limited research on the transferability of these algorithms, a standard algorithm-model has yet to be established (Beck et al. 2016, 2017; Johansen et al. 2018a). Therefore, it is recommended that managers apply in situ measurements to develop personalized or precision-based predictive models for the water body of interest. However, this report provides a framework for the production of simple, predictive models using in situ measurements and the results of the *waterquality* package.

Figure 3. Four Sentinel-2 algorithms produced by the *wq_calc* function in the *waterquality* package. Note that measurements of each algorithm are relative index values and do not represent absolute water quality parameter values.



3.3 Algorithm evaluation and model validation

Due to the many potential analytical and statistical options, it was decided to exclude algorithm evaluation or model validation directly from the *waterquality* package. However, it is important to include an example in the workflow given that a standard approach has yet to be established. The initial step of this process requires in situ observations with corresponding spatial coordinates of the observations and one water quality measurement (chlorophyll, BGA/PC, or turbidity). To increase usability, two options are included in the workflow for extracting the pixel values from the water quality parameters. The first option is to import an ESRI Shapefile (.shp) containing the spatial and water quality information. The second option is to take a spreadsheet with these values and convert them into a spatial object called a GeoPackage file (.gpkg). The latter option is preferred because a .gpkg file is a single file that contains all of the spatial information while a shapefile is made up of multiple separate files. Once the spatial object and the raster stack are created and loaded into the R working environment, a single function called *extract* from the *raster* package can be applied to extract the corresponding pixel values of all of the algorithms and combined with the in situ water quality information into a single data table called a data frame in R (Hijmans 2017). This data frame is then utilized as the input for the algorithm evaluation and model validation, or can be easily exported as a .CSV file or Excel spreadsheet for subsequent analysis or archiving.

Before conducting the final statistical analysis, it is recommended to inspect the sample locations on the imagery for any mixed or cloudy pixels that might alter the effectiveness of the predictive model. This can be accomplished by manually viewing the vector points on top of the original reflectance imagery. Any sample point that corresponds to any non-water (clouds, coast, and manmade objects, etc.) pixel should be removed. For this study, only one location (H03) was contaminated by *mixing*, where the land near the lake's beach was combined with water. By removing this location, the study was left with 41 in situ observations to be used in the model. To offer additional context, summary statistics can be produced using the R package *pastecs* and the function *stat.desc*, providing a host of metrics including the minimum, maximum, mean, range, and standard deviation of each variable (Grosjean and Ibanez 2018).

The final step is to evaluate the performance of each algorithm with the in situ measurements collected for each of the three water quality

parameters. This report follows the methodology set forth by Johansen et al. (2018a) by building a function to calculate a repeated k-fold cross-validation using the R package, *caret* (Kuhn et al. 2018; Stone 1974). This process allows for the evaluation of the entire suite of algorithms for the chosen water quality parameter. Due to the size of the data, it was determined that a three-fold with five repeats was most appropriate. There are two types of outputs from this statistical function. First, the k-means cross validation method produces average r^2, root mean square error (RMSE) and mean average error (MAE) values for the 15 models (three folds * five repeats). Second, a linear regression model is calculated using all 41 sample points together to calculate an r^2, p value, slope, and intercept. This function is written so that the user has to define only the data frame, a random seed value to ensure reproducibility (this study used the date of the first model run), the column numbers of the algorithms being evaluated, and the column name of the in situ water quality parameter being evaluated. Since this study acquired in situ measurements for all three water quality parameters, the water quality parameter input name must be manually changed for each to produce three separate statistical output tables.

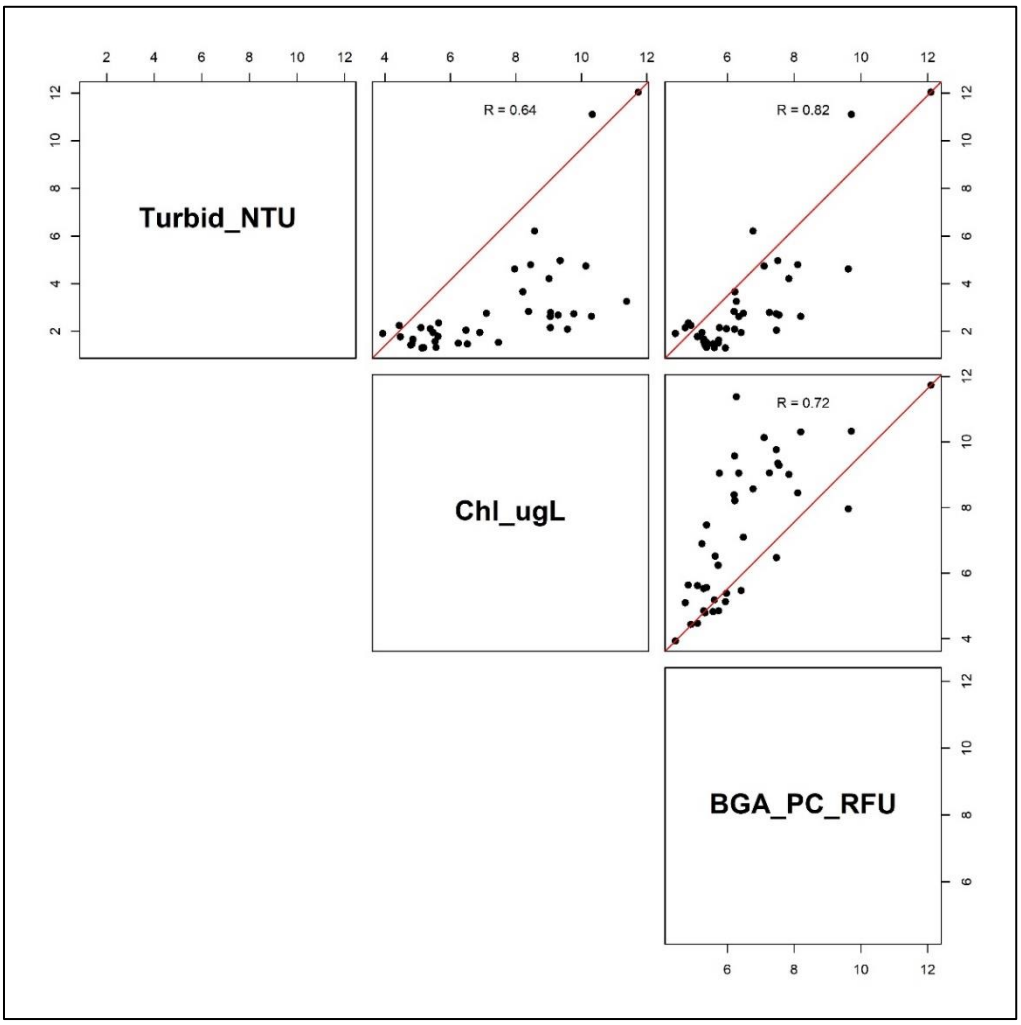
4 Results

The 41 sampled surface locations were used to calculate the descriptive statistics for the three water quality parameters and the following environmental variables at Harsha Lake: secchi depth, wind speed, air temperature, and water temperature (Table 4). YSI sonde measurements indicated that during the sampling campaign, water quality conditions were fair with relatively low concentration levels of all three water quality parameters. Chlorophyll concentrations averaged 7.30 µg/L with a maximum concentration of 11.74 µg/L occurring at sampling location H24B and a minimum concentration of 3.93 µg/L at H16B. These two sites also contain the minimum and maximum concentrations of BGA/PC, indicating a moderate degree of collinearity (Figure 4). This geographic bifurcation of the lake into relatively higher concentrations in the east and relatively lower concentrations in the west as well as the dominance of cyanobacteria as the primary chlorophyll producer has been documented in previous studies of Harsha Lake (Beck et al. 2016, 2017; Johansen et al. 2018a; Xu et al. 2018). A divergence between previous studies and the present study is the relatively low concentration level of all three water quality variables. This deviation allows for additional investigation into potentially detecting the lower limits of these algorithms and determining how algorithm performance is affected by concentration levels. Understanding algorithm thresholds and performance at low concentration levels is pivotal because adverse effects start to emerge that affect both human and animal health at chlorophyll-*a* concentrations as low as 10 µg/ L (Chorus and Bartram 1999; Stumpf et al. 2016).

Table 4. Descriptive statistics of YSI sonde-collected surface measurements for Harsha Lake acquired on August 8, 2016.

	Secchi Depth (cm)	Wind Speed (m/s)	Air Temp (°C)	Water Temp (°C)	Turbidity (NTU)	Chl (µg/L)	BGA/PC (RFU)
Min	50.98	0.13	25.30	28.73	1.30	3.93	4.44
Max	111.94	5.15	34.50	30.07	12.04	11.74	12.10
Range	60.96	5.02	9.20	1.34	10.74	7.81	7.66
Mean	93.06	2.24	30.15	29.31	3.00	7.30	6.44
Std Dev	15.77	1.18	2.47	0.29	2.29	2.17	1.53

Figure 4. Covariance plot of the YSI sonde-measured surface locations for the three water quality parameters, turbidity (Turbid_NTU), chlorophyll (Chl_ugL), and blue-green algae/phycocyanin (BGA_PC_RFU).



4.1 Chlorophyll

Six of the 24 Sentinel-2 algorithms tested produced correlations sufficient for the detection of chlorophyll with r^2 values ranging from 0.505 to 0.665, RMSE values from 1.234 to 1.587, and relative error of 15.8% to 20.3% (Table 5). For the comprehensive results of all chlorophyll algorithms, see Appendix (A-2). The two highest performing algorithms were Al10SABI and G004MCI, which performed well even with low surface chlorophyll concentrations (Figure 5). Of particular interest is the difference between the spectral bands and the calculation of these two algorithms. Al10SABI uses four bands covering much of the visible spectrum to compute a surface algal bloom index (SABI) style algorithm (865 – 665) / (490.5 + 560.5) while G004MCI uses three spectrally concentrated bands to

calculate a Maximum Chlorophyll Index (MCI) style algorithm (705.5 – 665 – (740.5 – 665). Even though both make use of the 665 nm band, these appear quite different and capture very different spectral features associated with chlorophyll and BGA/PC.

4.2 Blue-green algae/Phycocyanin (BGA/PC)

Only two of the 24 algorithms tested produced high enough correlations to be considered for this study. For the comprehensive results of all BGA/PC algorithms, see the Appendix (A-3). The two algorithms were Go04MCI and Al10SABI, which were the same as the top performing chlorophyll algorithms, but in reversed order. The r^2 values were 0.657 and 0.576, RMSE of 0.957 and 1.052, and relative error of 12.5% and 13.7% for Go04MCI and Al10SABI, respectively (Table 5). An unexpected result was that there were only two acceptable BGA/PC algorithms, since previous studies demonstrated that the dominant algal species of Harsha Lake during bloom conditions was cyanobacteria; thus, a stronger overlap between the chlorophyll and BGA/PC algorithms was expected (Johansen et al 2018a). Stumpf et al. (2016) offer a possible solution, who found that when cyanobacteria is the dominant algal species, chlorophyll algorithms are the preferred choice. In addition, this might be further exemplified with relatively low concentration levels of both pigments. An additional contributing factor to note is that the Sentinel-2 imager does not contain the 620 nm spectral band associated with BGA/PC.

4.3 Turbidity

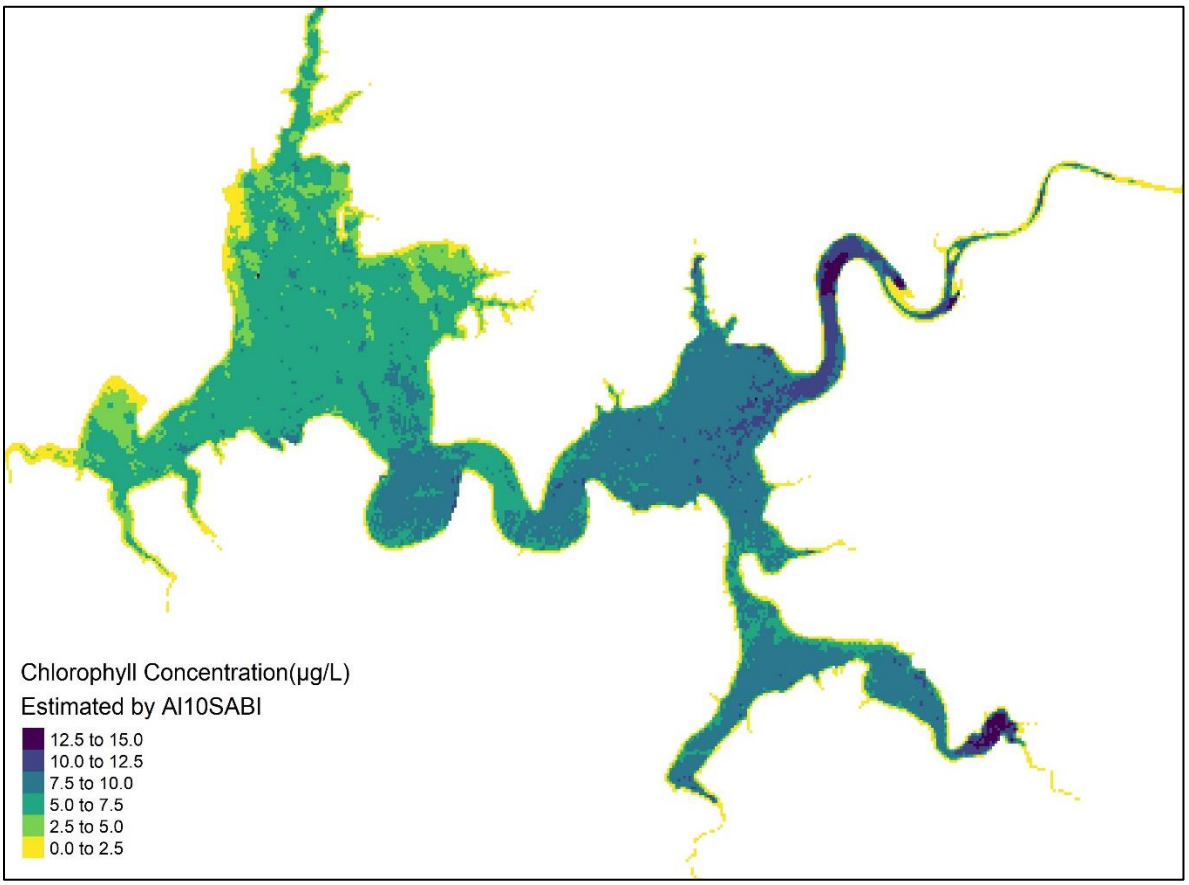
Finally, 5 of the 24 turbidity algorithms evaluated were deemed sufficient for the detection of turbidity (NTU) as measured by the water quality sonde. The turbidity algorithms had r^2 values ranging from 0.506 to 0.693, RMSEs from 1.760 to 1.414, and relative errors of 13.2% to 16.4% (Table 5). For the comprehensive results of all turbidity algorithms, see the Appendix (A-4). The highest performing algorithm was the Go04MCI algorithm followed by a three-way tie of the Am092Bsub, Wy08CI, and Ku15SLH algorithms. This tie is likely the result of the original algorithms being forced into specific Sentinel-2 bands, which results in multiple algorithms creating the same index values. Only the Go04MCI algorithm is shared among all three water quality parameters, and there is no overlap between chlorophyll and BGA/PC for the remaining four acceptable performing turbidity algorithms. Given that turbidity is a general water

clarity metric, this makes sense because these algorithms encompass additional spectral features than the more specific pigment algorithms.

Table 5. Evaluation of top-performing algorithms for each water quality parameter at Harsha Lake according to Pearson's r test (Type-1), linear regressions, and k-folds cross-validation.

		Global Model				Cross-Validated Average		
Algorithms	Water Quality Parameter	r^2	Slope	Intercept	p-value	r^2	RMSE	MAE
Al10SABI	Chl	0.665	-33.305	-1.352	0.000	0.677	1.234	0.958
Go04MCI	Chl	0.645	0.035	4.611	0.000	0.673	1.337	1.04
TurbDox02NIRoverRed	Chl	0.556	-13.001	11.552	0.000	0.568	1.459	1.162
Da052BDA	Chl	0.511	28.531	-24.368	0.000	0.523	1.535	1.25
MM12NDCI	Chl	0.511	64.421	3.982	0.000	0.520	1.526	1.22
Be16FLHGreenRedNIR	Chl	0.505	0.073	17.937	0.000	0.545	1.587	1.265
Go04MCI	BGA/PC	0.657	0.025	4.523	0.000	0.659	0.958	0.763
Al10SABI	BGA/PC	0.576	-21.885	0.753	0.000	0.613	1.052	0.789
Go04MCI	Turbidity	0.693	0.038	0.052	0.000	0.726	1.435	1.047
Am092Bsub	Turbidity	0.66	0.117	-0.296	0.000	0.537	1.454	1.117
Wy08CI	Turbidity	0.66	0.117	-0.296	0.000	0.533	1.415	1.085
Ku15SLH	Turbidity	0.66	0.059	-0.296	0.000	0.585	1.414	1.100
Be16FLHBlueRedNIR	Turbidity	0.506	0.071	3.441	0.000	0.501	1.76	1.257

Figure 5. Estimated Chlorophyll-*a* concentration (µg/L) for Harsha Lake, Ohio for August 8, 2016. Estimated concentrations were calculated by applying the AI10SABI algorithm derived linear model (-1.352x – 33.305) to the AI10SABI index values.



5 Conclusions

The goals of this report are two-fold: (1) To establish a user-friendly workflow for evaluating satellite-based remote sensing of water quality with a particular focus on CHABs (see the Appendix [A-6]), and (2) To investigate the effectiveness of using Sentinel-2 imagery for the detection and quantification of three water quality parameters using the newly developed open-source R package *waterquality*. This package allows for the easy conversion of reflectance imagery from five satellite imagers into 45 water quality algorithms. Furthermore, this report outlines a standardized workflow that guides researchers from data acquisition to statistical analysis of these algorithms with user-provided in situ measurements. Comprehensive R scripts and materials are available by referencing Johansen, 2018c. Study goals were accomplished by conducting a case study using Sentinel-2 imagery for Harsha Lake in Southwest Ohio and dense surface observations collected the same day as the satellite overpass.

The results of this case study are promising because 13 algorithm-parameter pairs had r^2 values of 0.5 or greater. Of these 13 pairs, 6 were highly correlated with chlorophyll, 2 with BGA/PC, and 5 with turbidity (Table 5). There was significant overlap between two of the highest performing algorithms, which confirms the principle of transferability of these algorithms between pigments or what Johansen et al. (2018a) termed "portability." For example, the G004MCI algorithm had r^2 values of greater than 0.5 for all three water quality parameters and was the top-performing algorithm for both BGA/PC and turbidity (second best for chlorophyll). Also of note is the Al10SABI algorithm, which was the highest-performing algorithm for chlorophyll and second best for BGA/PC but did not meet the r^2 threshold for turbidity. Given the complexity and dynamic nature of CHABs, it is encouraging that many simple reflectance algorithms demonstrate moderate to high levels of correlation between in situ observations and the three parameters. Each water quality parameter had at least one algorithm with an r^2 value greater than 0.65, indicating a strong correlation between the remotely sensed algorithms and in situ measurements. More importantly, the relative errors associated with these algorithms are very low (Chlorophyll RMSE ranged from 1.234 to 1.587 µg/L), which is accurate enough to be implemented into an early warning system. For example, if Harsha Lake implemented the Chorus and Bartram (1999) chlorophyll-*a* threshold of 10 µg/L, any location that

exceeded this concentration should be subject to additional testing. Given this threshold, there are two areas (northeast and southeast) that are visually identified as exceeding 10 µg/L in purple in Figure 5. These two areas should be subjected to further testing to evaluate algal species as well as test for any potential harmful toxins. Since algorithm performance will vary to some degree for each water body, concentration thresholds and error acceptance levels should be established locally by water management professionals.

Overall, algorithm-water quality parameter pairs correlate strongly enough with the surface observations of chlorophyll, BGA/PC, and turbidity to assist in the monitoring of water quality in freshwater bodies. These results continue to build upon a series of papers that consistently demonstrate moderate to high levels of accuracy in detection and quantification of water quality that implement simple and transferable methods. Additional investigation is still needed to explore how concentrations levels of all three water quality parameters influence algorithm performance. This is especially important around threshold standards, such as 25 µg/L of chlorophyll, where harmful effects of algal blooms may start to manifest (USEPA 2012a; Miltner 2018). The *waterquality* R package and subsequent workflow developed for this study was specifically designed to increase the usability and coverage of CHABs detection by using a multi-sensor, multi-algorithm approach. Ultimately, this study adds value to the water quality and CHABs community by assisting researchers and managers to conduct near real-time water quality monitoring required to understand and protect against future CHABs events.

References

Anderson, D. M., P. Hoagland, Y, Kaoru, and A. W. White. 2000. *Estimated Annual Economic Impact from Harmful Algal Blooms (HABs) in the United States*. Woods Hole, MA: Woods Hole Oceanographic Institution.

Augusto-Silva, P. B., I. Ogashawara, C. C. F. Barbosa, L. A. S. de Carvalho, D. S. F. Jorge, C. I. Fornari, and J. L. Stech. 2014. "Analysis of MERIS Reflectance Algorithms for Estimating Chlorophyll-*a* Concentration in a Brazilian Reservoir." *Remote Sens*. 6(12): 11689–117077.

Barnes, B. B., C. Hu, C. Kovach, and R. N. Silverstein. 2015. "Sediment Plumes Induced by the Port of Miami Dredging: Analysis and Interpretation Using Landsat and MODIS Data." *Remote Sens. Environ* 170: 328-339.

Beck, R. A., S. Zhan, H. Liu, S. Tong, B. Yang, M. Xu, Z. Ye, and Y. Huang. 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 Sens. Environ*. 178: 15–30.

Beck, R., M. Xu, S. Zhan, R. Johansen, H. Liu, S. Tong, B. Yang, and S. Shu. 2018. "Comparison of Satellite Reflectance Algorithms for Estimating Turbidity and Cyanobacterial Concentrations in Productive Freshwaters Using Hyperspectral Aircraft Imagery and Dense Coincident Surface Observations." *Journal of Great Lakes Research* 45(3): 413-433.

Beck, R. A., M. Xu, S. Zhan, H. Liu, R. A. Johansen, S. Tong, B. Yang, and S. Shu. 2017. "Comparison of Satellite Reflectance Algorithms for Estimating Phycocyanin Values and Cyanobacterial Total Biovolume in a Temperate Reservoir Using Coincident Hyperspectral Aircraft Imagery and Dense Coincident Surface Observations." *Remote Sensing* 9: 53.

Bernstein, L. S., X. Jin, B. Gregor, and S. M. Adler-Golden. 2012. "Quick Atmospheric Correction Code: Algorithm Description and Recent Upgrades." *Opt. Eng*. 51. https://doi.org/10.1117/1.OE.51.11.111719

Blondeau-Patissier, D., J. F. R. Gower, A. G. Dekker, S. R. Phinn, and V. E. Brando. 2014. "A Review of Ocean Color Remote Sensing Methods and Statistical Techniques for the Detection, Mapping and Analysis of Phytoplankton Blooms in Coastal and Open Oceans." *Prog. Oceanogr*. 123: 123–144.

Chorus, E. I., and J. Baartram. 1999. *Toxic Cyanobacteria in Water: A Guide to Their Public Health Consequences, Monitoring, and Management*. London: E & FN Spon. https://apps.who.int/iris/handle/10665/42827

Dogliotti, A. I., K. G. Ruddick, B. Nechad, D. Doxaran, and E. Knaeps. 2015. "A Single Algorithm to Retrieve Turbidity from Remotely Sensed Data in All Coastal and Estuarine Waters." *Remote Sens. Environ*. 156: 157-168.

Dokulil, M. T., and K. Teubner. 2000. "Cyanobacterial Dominance in Lakes." *Hydrobiologia* 438(1-3): 1–12.

Doxaran, D., J. M. Froidefond, and P. Castaing. 2002. "A Reflectance Band Ratio Used to Estimate Suspended Matter Concentrations in Sediment-Dominated Coastal Waters." *Int. J. Remote Sens.* 23(23): 5079-5085.

Doxaran, D., N. Cherukuru, and S. J. Lavender. 2006. "Apparent and Inherent Optical Properties of Turbid Estuarine Waters: Measurements, Empirical Quantification Relationships, and Modeling." *Appl. Opt.* 45(10): 2310-2324.

Drusch, M., U. Del Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, B. Hoersch, C. Isola, P. Laberinti, P. Martimort, A. Meygret, F. Spoto, O. Sy, F. Marchese, and P. Bargellini. 2012. "Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services." *Remote Sens. Environ.* 120: 25-36.

ESRI (Environmental Systems Research Institute). 1998. *ESRI Shapefile Technical Description*. ESRI White Paper, July 1998. https://www.esri.com/library/whitepapers/pdfs/shapefile.pdf

Gao, B. C., M. Montes, C. O. Davis, and A. Goetz. 2009. "Atmospheric Correction Algorithms for Hyperspectral Remote Sensing Data of Land and Ocean." *Remote Sens. Environ.* 113(Supplement 1): S17-S24.

Graham, J. L. 2006. *Harmful Algal Blooms*. USGS Fact Sheet, 2006-3147. Reston, VA: USGS.

Griffin, M. K., and H. K. Burke. 2003. "Compensation of Hyperspectral Data for Atmospheric Effects." *MIT Lincoln Lab. J.* 14: 29-54.

Grosjean, P., and F. Ibanez. 2018. *Pastecs: Package for Analsis of Space-Time Ecological Series*. R package version 1.3.21. https://CRAN.R-project.org/package=pastecs

Hijmans, R. J. 2016. *Raster: Geographic Data Analysis and Modeling*. R package version 2. https://CRAN.R-project.org/package=raster

Hunter, P. D., A. N. Tyler, N. J. Willby, and D. J. Gilvear. 2008. "The Spatial Dynamics of Vertical Migration by *Microcystis aeruginosa* in a Eutrophic Shallow Lake: A Case Study Using High Spatial Resolution Time-Series Airborne Remote Sensing." *Limnology and Oceanography* 53(6): 2391–2406.

Johansen, R. A., R. Beck, J. Nowosad, C. Nietch, M. Xu, S. Shu, B. Yang, and H. Liu. 2018a. "Evaluating the Portability of Satellite Derived Chlorophyll-*a* Algorithms for Temperate Inland Lakes Using Airborne Hyperspectral Imagery and Dense Surface Observations." *Harmful Algae* 76: 35-46.

Johansen, R., J. Nowosad, M. Reif, and E. Emery. 2018b. *waterquality: Satellite Derived Water Quality Detection Algorithms*. R package version 0.2.2. https://doi.org/10.5281/zenodo.1493487

Johansen, R. 2018c. *waterquality_workflow: A Case Study and Workflow for Detecting and Quantifying Cyanobacterial Harmful Algal Blooms (CHABs) from Sentinel-2 Imagery*. Version (v.0.2.) Zenodo. http://doi.org/10.5281/zenodo.2003619

Klemas, V. 2012. "Remote Sensing of Algal Blooms: An Overview with Case Studies." *J. Coast. Res.* 28(1A): 34–43.

Kuhn, M., J. Wing, S. Weston, A. Williams, C. Keefer, A. Engelhardt, T. Cooper, Z. Mayer, B. Kenkel, the R Core Team, M. Benesty, R. Lescarbeau, A. Ziem, L. Scrucca, Y. Tang, C. Candan, and T. Hunt. 2018. *caret: Classification and Regression Training*. R package version 6.0-80. https://CRAN.R-project.org/package=caret

Linkov, I., F. K. Satterstrom, D. Loney, and J. A. Steevans. 2009. *The Impact of Harmful Algal Blooms on USACE Operations*. ANSRP Technical Notes Collection. ERDC/TNANSRP-09-1. Vicksburg, MS: US Army Engineer Research and Development Center.

Main-Knorn, M., B. Pflug, J. Louis, V. Debaecker, U. Müller-Wilm, and F. Gascon. 2017. *Sen2Cor for Sentinel-2*. In *SPIE Proceedings, Vol. 10427: Image and Signal Processing for Remote Sensing XXIII*.

Miltner, R. J. 2018. "Eutrophication Endpoints for Large Rivers in Ohio, USA." *Environ. Monit. Assess*. 190: 55. http://dx.doi.org/10.1007/s10661-017-6422-4

Mishra, S., D. R. Mishra, and Z. P. Lee. 2014. "Bio-Optical Inversion in Highly Turbid and Cyanobacteria Dominated Waters." *IEEE Trans. Geosci. Remote Sens*. 52(1): 375-388.

Moore, G. K. 1980. "Satellite Remote Sensing of Water Turbidity." *Hydrol. Sci*. 25(4): 407-422.

Olmanson, L. G., P. L. Brezonik, and M. E. Bauer. 2013. "Airborne Hyperspectral Remote Sensing to Assess Spatial Distribution of Water Quality Characteristics in Large Rivers: The Mississippi River and Its Tributaries in Minnesota." *Remote Sens. Environ*. 130: 254-265.

OGC (Open Geospatial Consortium). n.d. *Getting Started With GeoPackage*. http://www.geopackage.org/guidance/getting-started.html

Randolph, K., J. Wilson, L. Tedesco, L. Li, D. L. Pascual, and E. Soyeux. 2008. "Hyperspectral Remote Sensing of Cyanobacteria in Turbid Productive Water Using Optically Active Pigments, Chlorophyll a and Phycocyanin." *Remote Sens. Environ*. 112: 4009–4019.

Ranghetti, L., and L. Busetto. 2018. *sen2r: an R toolbox to find, download and preprocess Sentinel-2 data*. R package version 0.3.2. doi: 10.5281/zenodo.1240385

R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. *R Foundation for Statistical Computing*. Vienna, Austria. https://www.R-project. org/

Reif, M. 2011. *Remote Sensing for Inland Water Quality Monitoring: A US Army Corps of Engineers Perspective*. ERDC/EL TR-11–13. Vicksburg, MS: US Army Engineer Research and Development Center.

Schalles, J., and Y. Yacobi. 2000. "Remote Detection and Seasonal Patterns of Phycocyanin, Carotenoid and Chlorophyll-*a* Pigments in Eutrophic Waters." *Archiv fur Hydrobiologie* (Special Issues Advances in Limnology) 55: 153–168.

Simis, S. G. H., S. W. M. Peters, and H. J. Gons. 2005. "Remote Sensing of the Cyanobacteria Pigment Phycocyanin in Turbid Inland Water." *Limnol. Oceanogr.* 50(1): 237–245.

Stone, M. 1974. "Cross-Validatory Choice and Assessment of Statistical Predictions." *Journal of the Royal Statistical Society* 36(2): 111-133.

Stumpf, R. P., T. T. Wynne, D. B. Baker, and G. L. Fahnenstiel. 2012. "Interannual Variability of Cyanobacterial Blooms in Lake Erie." *PLoS One* 7: 1–11.

Stumpf, R. P., T. W. Davis, T. T. Wynne, J. L. Graham, K. A. Loftin, T. H. Johengen, D. Gossiaux, D. Palladino, and A. Burtner. 2016. "Challenges for Mapping Cyanotoxin Patterns from Remote Sensing of Cyanobacteria." *Harmful Algae* 54: 160–173.

USEPA (US Environmental Protection Agency). 2012a. *Cyanobacteria and Cyanotoxins: Information for Drinking Water Systems*. Washington, DC: EPA-810F11001.

USEPA. 2012b. "The National Lakes Assessment Fact Sheet." *National Aquatic Resource Surveys*. Washington, DC.

Wynne, T. T., R. P. Stumpf, M. C. Tomlinson, R. A. Warner, P. A. Tester, and J. Dyble. 2008. "Relating Spectral Shape to Cyanobacterial Blooms in the Laurentian Great Lakes." *Int. J. Remote Sens.* 29(12): 3665–3672.

Wynne, T. T., R. P. Stumpf, M. C. Tomlinson, and J. Dyble. 2015. "Characterizing a Cyanobacterial Bloom in Western Lake Erie Using Satellite Imagery and Meteorological Data." *Limnol. Oceanogr.* 55(5): 2025–2036.

Wynne, T. T., and R. P. Stumpf. 2015. "Spatial and Temporal Patterns in the Seasonal Distribution of Toxic Cyanobacteria in Western Lake Erie from 2002–2014." *Toxins* 7(5): 1649-1663.

Wozniak, M., K. M. Bradtke, M. Darecki, and A. Krezel. 2016. "Empirical Model for Phycocyanin Concentration Estimation as an Indicator of Cyanobacterial Bloom in the Optically Complex CoastalWaters of the Baltic Sea." *Remote Sens.* 8(3): 1–23.

Xu, Min, Hongxing Liu, Richard Beck, John Lekki, Bo Yang, Song Shu, Emily L. Kang, Robert Anderson, Richard Johansen, Erich Emery, Molly Reif, and Teresa Benko. 2018. "A Spectral Space Partition Guided Ensemble Method for Retrieving Chlorophyll-*a* Concentration in Inland Waters from Sentinel-2A Satellite Imagery." *Journal of Great Lakes Research* 45(3): 454-465. https://doi.org/10.1016/j.jglr.2018.09.002

YSI. 2003. *ADV6600 Environmental Monitoring System: Operations Manual.* https://www.ysi.com/File%20Library/Documents/Manuals%20for%20Discontinued%20Products/ADV6600.pdf

Appendix: Water Quality Algorithms, Workflow, and Computation

A-1. Comprehensive list of water quality algorithms and the reference to original works

Water Quality Index	Original Reference
Al10SABI	Alawadi, F. Detection of surface algal blooms using the newly developed algorithm surface algal bloom index (SABI). Proc. SPIE 2010, 7825.
Am092Bsub	Amin, R.; Zhou, J.; Gilerson, A.; Gross, B.; Moshary, F.; Ahmed, S. Novel optical techniques for detecting and classifying toxic dinoflagellate Karenia brevis blooms using satellite imagery. Opt. Express 2009, 17, 9126–9144.
Am09KBBI	Amin, R.; Zhou, J.; Gilerson, A.; Gross, B.; Moshary, F.; Ahmed, S.; Novel optical techniques for detecting and classifying toxic dinoflagellate Karenia brevis blooms using satellite imagery, Optics Express, 2009, 17, 11, 1-13.
Be162B643sub629	Beck, R.; Xu, M.; Zhan, S.; Liu, H.; Johansen, R.A.; Tong, S.; Yang, B.; Shu, S.; Wu, Q.; Wang, S.; Berling, K.; Murray, A.; Emery, E.; Reif, M.; Harwood, J.; Young, J.; Martin, M.; Stillings, G.; Stumpf, R.; Su, H.; Ye, Z.; Huang, Y. Comparison of Satellite Reflectance Algorithms for Estimating Phycocyanin Values and Cyanobacterial Total Biovolume in a Temperate Reservoir Using Coincident Hyperspectral Aircraft Imagery and Dense Coincident Surface Observations. Remote Sens. 2017, 9, 538.
Be162B700sub601	Beck, R.; Xu, M.; Zhan, S.; Liu, H.; Johansen, R.A.; Tong, S.; Yang, B.; Shu, S.; Wu, Q.; Wang, S.; Berling, K.; Murray, A.; Emery, E.; Reif, M.; Harwood, J.; Young, J.; Martin, M.; Stillings, G.; Stumpf, R.; Su, H.; Ye, Z.; Huang, Y. Comparison of Satellite Reflectance Algorithms for Estimating Phycocyanin Values and Cyanobacterial Total Biovolume in a Temperate Reservoir Using Coincident Hyperspectral Aircraft Imagery and Dense Coincident Surface Observations. Remote Sens. 2017, 9, 538.
Be162BsubPhy	Beck, R.; Xu, M.; Zhan, S.; Liu, H.; Johansen, R.A.; Tong, S.; Yang, B.; Shu, S.; Wu, Q.; Wang, S.; Berling, K.; Murray, A.; Emery, E.; Reif, M.; Harwood, J.; Young, J.; Martin, M.; Stillings, G.; Stumpf, R.; Su, H.; Ye, Z.; Huang, Y. Comparison of Satellite Reflectance Algorithms for Estimating Phycocyanin Values and Cyanobacterial Total Biovolume in a Temperate Reservoir Using Coincident Hyperspectral Aircraft Imagery and Dense Coincident Surface Observations. Remote Sens. 2017, 9, 538.
Be16FLHblue	Beck, R.A. and 22 others; Comparison of satellite reflectance algorithms for estimating chlorophyll-*a* in a temperate reservoir using coincident hyperspectral aircraft imagery and dense coincident surface observations, Remote Sens. Environ., 2016, 178, 15-30.

Be16FLHBlueRedNIR	Beck, R.; Xu, M.; Zhan, S.; Liu, H.; Johansen, R.A.; Tong, S.; Yang, B.; Shu, S.; Wu, Q.; Wang, S.; Berling, K.; Murray, A.; Emery, E.; Reif, M.; Harwood, J.; Young, J.; Martin, M.; Stillings, G.; Stumpf, R.; Su, H.; Ye, Z.; Huang, Y. Comparison of Satellite Reflectance Algorithms for Estimating Phycocyanin Values and Cyanobacterial Total Biovolume in a Temperate Reservoir Using Coincident Hyperspectral Aircraft Imagery and Dense Coincident Surface Observations. Remote Sens. 2017, 9, 538.
Be16FLHGreenRedNIR	Beck, R.; Xu, M.; Zhan, S.; Liu, H.; Johansen, R.A.; Tong, S.; Yang, B.; Shu, S.; Wu, Q.; Wang, S.; Berling, K.; Murray, A.; Emery, E.; Reif, M.; Harwood, J.; Young, J.; Martin, M.; Stillings, G.; Stumpf, R.; Su, H.; Ye, Z.; Huang, Y. Comparison of Satellite Reflectance Algorithms for Estimating Phycocyanin Values and Cyanobacterial Total Biovolume in a Temperate Reservoir Using Coincident Hyperspectral Aircraft Imagery and Dense Coincident Surface Observations. Remote Sens. 2017, 9, 538.
Be16FLHviolet	Beck, R.A. and 22 others; Comparison of satellite reflectance algorithms for estimating chlorophyll-*a* in a temperate reservoir using coincident hyperspectral aircraft imagery and dense coincident surface observations, Remote Sens. Environ., 2016, 178, 15-30.
Be16FLHVioletRedNIR	Beck, R.; Xu, M.; Zhan, S.; Liu, H.; Johansen, R.A.; Tong, S.; Yang, B.; Shu, S.; Wu, Q.; Wang, S.; Berling, K.; Murray, A.; Emery, E.; Reif, M.; Harwood, J.; Young, J.; Martin, M.; Stillings, G.; Stumpf, R.; Su, H.; Ye, Z.; Huang, Y. Comparison of Satellite Reflectance Algorithms for Estimating Phycocyanin Values and Cyanobacterial Total Biovolume in a Temperate Reservoir Using Coincident Hyperspectral Aircraft Imagery and Dense Coincident Surface Observations. Remote Sens. 2017, 9, 538.
Be16MPI	Beck, R.; Xu, M.; Zhan, S.; Liu, H.; Johansen, R.A.; Tong, S.; Yang, B.; Shu, S.; Wu, Q.; Wang, S.; Berling, K.; Murray, A.; Emery, E.; Reif, M.; Harwood, J.; Young, J.; Martin, M.; Stillings, G.; Stumpf, R.; Su, H.; Ye, Z.; Huang, Y. Comparison of Satellite Reflectance Algorithms for Estimating Phycocyanin Values and Cyanobacterial Total Biovolume in a Temperate Reservoir Using Coincident Hyperspectral Aircraft Imagery and Dense Coincident Surface Observations. Remote Sens. 2017, 9, 538.
Be16NDPhyI	Beck, R.; Xu, M.; Zhan, S.; Liu, H.; Johansen, R.A.; Tong, S.; Yang, B.; Shu, S.; Wu, Q.; Wang, S.; Berling, K.; Murray, A.; Emery, E.; Reif, M.; Harwood, J.; Young, J.; Martin, M.; Stillings, G.; Stumpf, R.; Su, H.; Ye, Z.; Huang, Y. Comparison of Satellite Reflectance Algorithms for Estimating Phycocyanin Values and Cyanobacterial Total Biovolume in a Temperate Reservoir Using Coincident Hyperspectral Aircraft Imagery and Dense Coincident Surface Observations. Remote Sens. 2017, 9, 538.

Be16NDPhyI644over615	Beck, R.; Xu, M.; Zhan, S.; Liu, H.; Johansen, R.A.; Tong, S.; Yang, B.; Shu, S.; Wu, Q.; Wang, S.; Berling, K.; Murray, A.; Emery, E.; Reif, M.; Harwood, J.; Young, J.; Martin, M.; Stillings, G.; Stumpf, R.; Su, H.; Ye, Z.; Huang, Y. Comparison of Satellite Reflectance Algorithms for Estimating Phycocyanin Values and Cyanobacterial Total Biovolume in a Temperate Reservoir Using Coincident Hyperspectral Aircraft Imagery and Dense Coincident Surface Observations. Remote Sens. 2017, 9, 538.
Be16NDPhyI644over629	Beck, R.; Xu, M.; Zhan, S.; Liu, H.; Johansen, R.A.; Tong, S.; Yang, B.; Shu, S.; Wu, Q.; Wang, S.; Berling, K.; Murray, A.; Emery, E.; Reif, M.; Harwood, J.; Young, J.; Martin, M.; Stillings, G.; Stumpf, R.; Su, H.; Ye, Z.; Huang, Y. Comparison of Satellite Reflectance Algorithms for Estimating Phycocyanin Values and Cyanobacterial Total Biovolume in a Temperate Reservoir Using Coincident Hyperspectral Aircraft Imagery and Dense Coincident Surface Observations. Remote Sens. 2017, 9, 538.
Be16NDTIblue	Beck, R.; Xu, M.; Zhan, S.; Liu, H.; Johansen, R.A.; Tong, S.; Yang, B.; Shu, S.; Wu, Q.; Wang, S.; Berling, K.; Murray, A.; Emery, E.; Reif, M.; Harwood, J.; Young, J.; Martin, M.; Stillings, G.; Stumpf, R.; Su, H.; Ye, Z.; Huang, Y. Comparison of Satellite Reflectance Algorithms for Estimating Phycocyanin Values and Cyanobacterial Total Biovolume in a Temperate Reservoir Using Coincident Hyperspectral Aircraft Imagery and Dense Coincident Surface Observations. Remote Sens. 2017, 9, 538.
Be16NDTIviolet	Beck, R.; Xu, M.; Zhan, S.; Liu, H.; Johansen, R.A.; Tong, S.; Yang, B.; Shu, S.; Wu, Q.; Wang, S.; Berling, K.; Murray, A.; Emery, E.; Reif, M.; Harwood, J.; Young, J.; Martin, M.; Stillings, G.; Stumpf, R.; Su, H.; Ye, Z.; Huang, Y. Comparison of Satellite Reflectance Algorithms for Estimating Phycocyanin Values and Cyanobacterial Total Biovolume in a Temperate Reservoir Using Coincident Hyperspectral Aircraft Imagery and Dense Coincident Surface Observations. Remote Sens. 2017, 9, 538.
Be16Phy2BDA644over629	Beck, R.; Xu, M.; Zhan, S.; Liu, H.; Johansen, R.A.; Tong, S.; Yang, B.; Shu, S.; Wu, Q.; Wang, S.; Berling, K.; Murray, A.; Emery, E.; Reif, M.; Harwood, J.; Young, J.; Martin, M.; Stillings, G.; Stumpf, R.; Su, H.; Ye, Z.; Huang, Y. Comparison of Satellite Reflectance Algorithms for Estimating Phycocyanin Values and Cyanobacterial Total Biovolume in a Temperate Reservoir Using Coincident Hyperspectral Aircraft Imagery and Dense Coincident Surface Observations. Remote Sens. 2017, 9, 538.
Da052BDA	Wynne, T. T., Stumpf, R. P., Tomlinson, M. C., Warner, R. A., Tester, P. A., Dyble, J.; Relating spectral shape to cyanobacterial blooms in the Laurentian Great Lakes. Int. J. Remote Sens., 2008, 29, 3665–3672.
De933BDA	Dekker, A.; Detection of the optical water quality parameters for eutrophic waters by high resolution remote sensing, Ph.D. thesis, 1993, Free University, Amsterdam.
Gi033BDA	Gitelson, A.A.; U. Gritz, and M. N. Merzlyak.; Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. J. Plant Phys. 2003, 160, 271-282.

Go04MCI	Gower, J.F.R.; Brown,L.; Borstad, G.A.; Observation of chlorophyll fluorescence in west coast waters of Canada using the MODIS satellite sensor. Can. J. Remote Sens., 2004, 30 (1), 17–25.
HU103BDA	Hunter, P.D.; Tyler, A.N.; Willby, N.J.; Gilvear, D.J.; The spatial dynamics of vertical migration by Microcystis aeruginosa in a eutrophic shallow lake: A case study using high spatial resolution time-series airborne remote sensing. Limn. Oceanogr. 2008, 53, 2391-2406
Kn07KIVU	Kneubuhler, M.; Frank T.; Kellenberger, T.W; Pasche N.; Schmid M.; Mapping chlorophyll-*a* in Lake Kivu with remote sensing methods. 2007, Proceedings of the Envisat Symposium 2007, Montreux, Switzerland 23–27 April 2007 (ESA SP-636, July 2007).
Ku15PhyCI	Kudela, R.M., Palacios, S.L., Austerberry, D.C., Accorsi, E.K., Guild, L.S.; Application of hyperspectral remote sensing to cyanobacterial blooms in inland waters, Torres-Perez, J., 2015, Remote Sens. Environ., 2015, 167, 1-10.
Ku15SLH	Kudela, R.M., Palacios, S.L., Austerberry, D.C., Accorsi, E.K., Guild, L.S.; Application of hyperspectral remote sensing to cyanobacterial blooms in inland waters, Torres-Perez, J., 2015, Remote Sens. Environ., 2015, 167, 1-10
MI092BDA	Mishra, S.; Mishra, D.R.; Schluchter, W. M., A novel algorithm for predicting PC concentrations in cyanobacteria: A proximal hyperspectral remote sensing approach. Remote Sens., 2009, 1, 758–775.
MM092BDA	Mishra, S.; Mishra, D.R.; Schluchter, W. M., A novel algorithm for predicting PC concentrations in cyanobacteria: A proximal hyperspectral remote sensing approach. Remote Sens., 2009, 1, 758–775.
MM12NDCI	Mishra, S.; and Mishra, D.R. Normalized difference chlorophyll index: A novel model for remote estimation of chlorophyll-*a* concentration in turbid productive waters, Remote Sens. Environ., 2012, 117, 394-406
MM12NDCIaIt	Mishra, S.; Mishra, D.R.; A novel remote sensing algorithm to quantify phycocyanin in cyanobacterial algal blooms, Env. Res. Lett., 2014, 9 (11), DOI:10.1088/1748-9326/9/11/114003
MM143BDAopt	Mishra, S.; Mishra, D.R.; A novel remote sensing algorithm to quantify phycocyanin in cyanobacterial algal blooms, Env. Res. Lett., 2014, 9 (11), DOI:10.1088/1748-9326/9/11/114003
SI052BDA	Simis, S. G. H.; Peters, S.W. M.; Gons, H. J.; Remote sensing of the cyanobacteria pigment phycocyanin in turbid inland water. Limn. Oceanogr., 2005, 50, 237–245
SM122BDA	Mishra, S. Remote sensing of cyanobacteria in turbid productive waters, PhD Dissertation. Mississippi State University, USA. 2012.
SY002BDA	Schalles, J.; Yacobi, Y. Remote detection and seasonal patterns of phycocyanin, carotenoid and chlorophyll-*a* pigments in eutrophic waters. Archiv fur Hydrobiologie, Special Issues Advances in Limnology, 2000, 55,153–168

TurbBe16GreenPlusRedBothOverViolet	Beck, R.; Xu, M.; Zhan, S.; Liu, H.; Johansen, R.A.; Tong, S.; Yang, B.; Shu, S.; Wu, Q.; Wang, S.; Berling, K.; Murray, A.; Emery, E.; Reif, M.; Harwood, J.; Young, J.; Martin, M.; Stillings, G.; Stumpf, R.; Su, H.; Ye, Z.; Huang, Y. Comparison of Satellite Reflectance Algorithms for Estimating Phycocyanin Values and Cyanobacterial Total Biovolume in a Temperate Reservoir Using Coincident Hyperspectral Aircraft Imagery and Dense Coincident Surface Observations. Remote Sens. 2017, 9, 538
TurbBe16RedOverViolet	Beck, R.; Xu, M.; Zhan, S.; Liu, H.; Johansen, R.A.; Tong, S.; Yang, B.; Shu, S.; Wu, Q.; Wang, S.; Berling, K.; Murray, A.; Emery, E.; Reif, M.; Harwood, J.; Young, J.; Martin, M.; Stillings, G.; Stumpf, R.; Su, H.; Ye, Z.; Huang, Y. Comparison of Satellite Reflectance Algorithms for Estimating Phycocyanin Values and Cyanobacterial Total Biovolume in a Temperate Reservoir Using Coincident Hyperspectral Aircraft Imagery and Dense Coincident Surface Observations. Remote Sens. 2017, 9, 538
TurbBow06RedOverGreen	Bowers, D. G., and C. E. Binding. 2006. “The Optical Properties of Mineral Suspended Particles: A Review and Synthesis.” Estuarine Coastal and Shelf Science 67 (1–2): 219–230. doi:10.1016/j.ecss.2005.11.010
TurbChip09NIROverGreen	Chipman, J. W.; Olmanson, L.G.; Gitelson, A.A.; Remote sensing methods for lake management: A guide for resource managers and decision-makers. 2009.
TurbDox02NIRoverRed	Doxaran, D., Froidefond, J.-M.; Castaing, P. ; A reflectance band ratio used to estimate suspended matter concentrations in sediment-dominated coastal waters, Remote Sens., 2002, 23, 5079-5085
TurbFrohn09GreenPlusRedBothOverBlue	Frohn, R. C., & Autrey, B. C. (2009). Water quality assessment in the Ohio River using new indices for turbidity and chlorophyll-*a* with Landsat-7 Imagery. Draft Internal Report, US Environmental Protection Agency.
TurbHarr92NIR	Schiebe F.R., Harrington J.A., Ritchie J.C. Remote-Sensing of Suspended Sediments—the Lake Chicot, Arkansas Project. Int. J. Remote Sens. 1992;13:1487–1509
TurbLath91RedOverBlue	Lathrop, R. G., Jr., T. M. Lillesand, and B. S. Yandell, 1991. Testing the utility of simple multi-date Thematic Mapper calibration algorithms for monitoring turbid inland waters. International Journal of Remote Sensing
TurbMoore80Red	Moore, G.K., Satellite remote sensing of water turbidity, Hydrological Sciences, 1980, 25, 4, 407-422
Wy08CI	Wynne, T. T., Stumpf, R. P., Tomlinson, M. C., Warner, R. A., Tester, P. A., Dyble, J.; Relating spectral shape to cyanobacterial blooms in the Laurentian Great Lakes. Int. J. Remote Sens., 2008, 29, 3665–3672.
Zh10FLH	Zhao, D.Z.; Xing, X.G.; Liu, Y.G.; Yang, J.H.; Wang, L. The relation of chlorophyll-*a* concentration with the reflectance peak near 700 nm in algae-dominated waters and sensitivity of fluorescence algorithms for detecting algal bloom. Int. J. Remote Sens. 2010, 31, 39-48

A-2. Evaluation of all chlorophyll algorithms for Harsha Lake according to Pearson's r test (Type-1), linear regressions, and k-folds cross-validation

Chl (µg/L)	Global Model				Cross-Validated Average		
Algorithms	r^2	Slope	Intercept	p-value	r^2	RMSE	MAE
Al10SABI	0.665	-33.305	-1.352	0.000	0.677	1.234	0.958
Go04MCI	0.645	0.035	4.611	0.000	0.673	1.337	1.040
TurbDox02NIRoverRed	0.556	-13.001	11.552	0.000	0.568	1.459	1.162
Da052BDA	0.511	28.531	-24.368	0.000	0.523	1.535	1.250
MM12NDCI	0.511	64.421	3.982	0.000	0.520	1.526	1.220
MM12NDCIalt	0.511	64.421	3.982	0.000	0.522	1.521	1.216
Be16FLHGreenRedNIR	0.505	0.073	17.937	0.000	0.545	1.587	1.265
TurbChip09NIROverGreen	0.473	-18.557	11.165	0.000	0.507	1.603	1.291
TurbHarr92NIR	0.457	-0.036	10.439	0.000	0.482	1.622	1.269
Am092Bsub	0.385	0.085	4.921	0.000	0.387	1.711	1.488
Ku15PhyCI	0.385	0.085	4.921	0.000	0.425	1.744	1.513
Wy08CI	0.385	0.085	4.921	0.000	0.396	1.710	1.488
Ku15SLH	0.385	0.042	4.921	0.000	0.433	1.732	1.516
Be16FLHBlueRedNIR	0.244	0.047	7.589	0.001	0.303	1.911	1.539
Be16NDTIblue	0.224	26.466	7.627	0.002	0.271	1.962	1.570
TurbLath91RedOverBlue	0.223	13.369	-5.781	0.002	0.250	1.970	1.574
Be16FLHviolet	0.211	-0.026	12.185	0.002	0.309	2.045	1.543
Gi033BDA	0.167	30.513	5.038	0.008	0.207	2.007	1.782
TurbFrohn09GreenPlusRedBothOverBlue	0.156	5.492	-6.516	0.010	0.222	2.027	1.634
TurbBow06RedOverGreen	0.096	15.948	-2.861	0.049	0.151	2.118	1.832
TurbMoore80Red	0.041	-0.010	9.908	0.206	0.211	2.397	1.928
Am09KBBI	0.041	-7.405	5.876	0.206	0.121	2.210	1.935
Be16FLHVioletRedNIR	0.024	-0.008	8.749	0.337	0.126	2.280	1.912
TurbBe16GreenPlusRedBothOverViolet	0.021	-0.271	9.743	0.366	0.146	2.352	1.949
Be16NDTIviolet	0.016	-5.867	10.533	0.428	0.178	2.347	1.984
TurbBe16RedOverViolet	0.000	-0.070	7.545	0.909	0.130	2.344	2.001

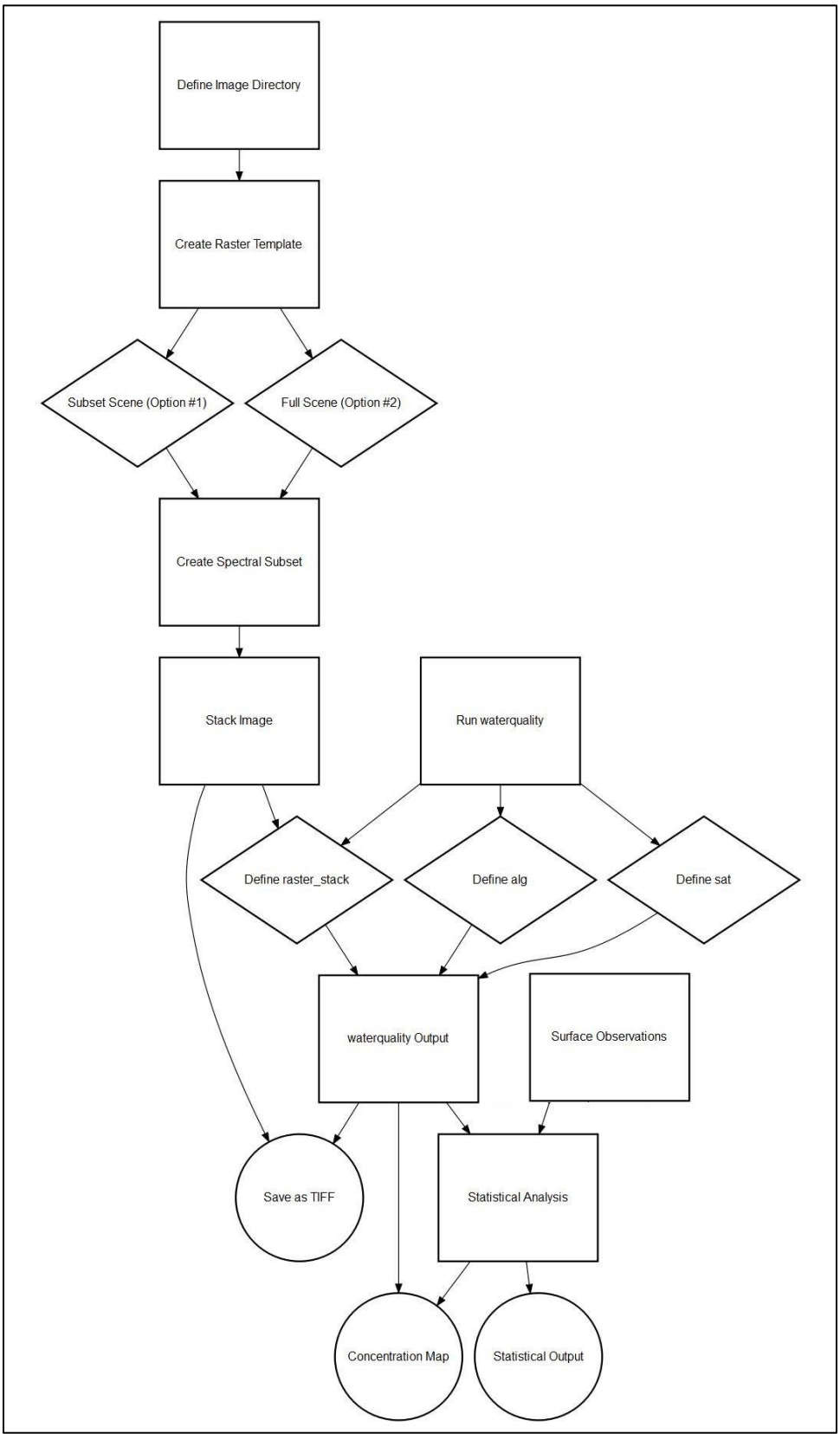
A-3. Evaluation of all BGA/PC algorithms for Harsha Lake according to Pearson's r test (Type-1), linear regressions, and k-folds cross-validation

BGA/PC (RFU)	Global Model				Cross-Validated Average		
Algorithms	r^2	Slope	Intercept	p-value	r^2	RMSE	MAE
Go04MCI	0.657	0.025	4.523	0.000	0.659	0.958	0.763
Al10SABI	0.576	-21.885	0.753	0.000	0.613	1.052	0.789
Be16FLHGreenRedNIR	0.467	0.049	13.655	0.000	0.518	1.137	0.868
Am092Bsub	0.428	0.063	4.668	0.000	0.343	1.255	0.999
Ku15PhyCI	0.428	0.063	4.668	0.000	0.386	1.213	0.966
Wy08CI	0.428	0.063	4.668	0.000	0.376	1.195	0.954
Ku15SLH	0.428	0.031	4.668	0.000	0.352	1.244	0.977
TurbDox02NIRoverRed	0.365	-7.438	8.871	0.000	0.413	1.227	0.905
Da052BDA	0.350	16.673	-12.067	0.000	0.335	1.287	1.055
MM12NDCI	0.336	36.913	4.537	0.000	0.318	1.274	1.037
MM12NDCIalt	0.336	36.913	4.537	0.000	0.328	1.325	1.074
TurbBow06RedOverGreen	0.310	20.274	-6.478	0.000	0.334	1.317	1.039
Be16FLHBlueRedNIR	0.289	0.036	6.660	0.000	0.311	1.343	0.971
TurbLath91RedOverBlue	0.284	10.672	-4.004	0.000	0.316	1.327	0.984
TurbChip09NIROverGreen	0.270	-9.907	8.502	0.000	0.354	1.280	0.966
Be16NDTIblue	0.268	20.433	6.691	0.001	0.325	1.344	1.002
TurbHarr92NIR	0.226	-0.018	7.996	0.002	0.304	1.330	0.953
TurbFrohn09GreenPlusRedBothOverBlue	0.096	3.035	-1.197	0.049	0.188	1.528	1.104
TurbBe16RedOverViolet	0.065	0.676	4.064	0.109	0.197	1.621	1.243
Gi033BDA	0.061	12.977	5.477	0.121	0.136	1.563	1.194
Be16NDTIviolet	0.029	5.541	3.386	0.287	0.186	1.599	1.216
Be16FLHviolet	0.024	-0.006	7.611	0.329	0.150	1.550	1.126
Be16FLHVioletRedNIR	0.022	0.005	5.455	0.356	0.164	1.661	1.237
TurbBe16GreenPlusRedBothOverViolet	0.013	0.153	5.058	0.470	0.138	1.565	1.180
TurbMoore80Red	0.012	0.004	5.442	0.497	0.219	1.654	1.244
Am09KBBI	0.008	-2.297	5.997	0.581	0.045	1.538	1.177

A-4. Evaluation of all turbidity algorithms for Harsha Lake according to Pearson's r test (Type-1), linear regressions, and k-folds cross-validation

Turbidity (NTU)	Global Model				Cross-Validated Average		
Algorithms	r^2	Slope	Intercept	p-value	r^2	RMSE	MAE
Go04MCI	0.693	0.038	0.052	0.000	0.726	1.435	1.047
Am092Bsub	0.660	0.117	-0.296	0.000	0.537	1.454	1.117
Ku15PhyCI	0.660	0.117	-0.296	0.000	0.534	1.479	1.125
Wy08CI	0.660	0.117	-0.296	0.000	0.533	1.415	1.085
Ku15SLH	0.660	0.059	-0.296	0.000	0.585	1.414	1.100
Be16FLHBlueRedNIR	0.506	0.071	3.441	0.000	0.501	1.760	1.257
TurbLath91RedOverBlue	0.488	20.949	-17.498	0.000	0.435	1.758	1.226
Al10SABI	0.485	-30.103	-4.819	0.000	0.547	1.708	1.168
Be16NDTIblue	0.455	39.901	3.493	0.000	0.454	1.776	1.292
Da052BDA	0.413	27.143	-27.127	0.000	0.409	1.859	1.291
MM12NDCI	0.395	59.960	-0.088	0.000	0.398	1.853	1.305
MM12NDCIalt	0.395	59.960	-0.088	0.000	0.385	2.020	1.393
TurbBow06RedOverGreen	0.385	33.855	-18.568	0.000	0.365	1.974	1.447
TurbBe16RedOverViolet	0.300	2.184	-4.669	0.000	0.292	2.087	1.615
Be16FLHGreenRedNIR	0.248	0.054	10.890	0.001	0.276	2.099	1.424
TurbDox02NIRoverRed	0.242	-9.084	5.972	0.001	0.320	1.941	1.192
TurbFrohn09GreenPlusRedBothOverBlue	0.234	7.103	-14.869	0.001	0.309	2.078	1.328
TurbBe16GreenPlusRedBothOverViolet	0.206	0.899	-5.095	0.003	0.260	2.258	1.672
Be16NDTIviolet	0.202	21.935	-9.084	0.003	0.289	2.265	1.721
Be16FLHVioletRedNIR	0.187	0.023	-1.314	0.005	0.328	2.357	1.760
TurbChip09NIROverGreen	0.163	-11.540	5.404	0.009	0.281	2.008	1.209
TurbMoore80Red	0.151	0.021	-2.315	0.012	0.238	2.313	1.712
Gi033BDA	0.123	27.754	0.944	0.024	0.161	2.198	1.514
TurbHarr92NIR	0.091	-0.017	4.482	0.056	0.212	2.130	1.310
Am09KBBI	0.033	-7.034	1.648	0.257	0.127	2.241	1.476
Be16FLHviolet	0.003	0.003	2.417	0.747	0.212	2.441	1.602

A-5. Data flow diagram, waterquality_workflow



A-6. Comprehensive R code for waterquality_workflow (Johansen 2018c)

```
#Title: waterquality_workflow: A case study and workflow for
detecting and quantifying cyanobacterial harmful algal blooms
(CHABs) from Sentinel-2 Imagery
# Author: Richard A. Johansen
# Date: December 6th 2018
# Source: https://github.com/RAJohansen/waterquality_workflow
# Citation: Richard A. Johansen. (2018).
RAJohansen/waterquality_workflow: waterquality_workflow: A case
study and workflow for detecting and quantifying cyanobacterial
harmful algal blooms (CHABs) from Sentinel-2 Imagery (Version
v.0.2). Zenodo. http://doi.org/10.5281/zenodo.2003619

### Initial Requirements and R Packages ------------------------
---------------
#Packages must be installed using install.packages("Package
Name") or
# devtools::install_github("Package Name"), if this is the first
time you are using these packages.
library(tidyverse)
library(raster)
library(waterquality)
library(sen2r)
library(sf)
library(gdalUtils)
library(magrittr)
library(rgdal)
library(caret)

#### Sen2Cor Atmospheric Correction-----------------------------
---------------
#Make Sure Dependencies are installed
check_sen2r_deps()
#Once all are installed close dependencies GUI

#Run sen2cor
#default output will be in a new folder at the same level as the
input level 1C
#Calculate time elapsed using system.time Process takes ~30
minutes
sen2cor(".../S2A_MSIL1C_20160808T162342_N0204_R040_T16SGJ_2016080
8T162611.SAFE")

### Preprocessing imagery --------------------------------------
---------------
#Recommended*** Option 1: Preprocess Clipped & Masked Reflectance
Imagery

#Image Directory Folder
Image_Directory =
"C:/temp/S2A_MSIL2A_20160808T162342_N0204_R040_T16SGJ_20160808T16
2611.SAFE/GRANULE/L2A_T16SGJ_A005900_20160808T162611/IMG_DATA/R20
m"
```

```
# Extracts all raster files from Image Directory with extention
.jp2
Rasters = dir(Image_Directory, pattern = "*.jp2$", full.names =
TRUE)

#Import Shapefile of Area of Interest
AOI = st_read("C:/temp/waterquality Vignette/Harsha_Lake.gpkg")

#Reproject AOI if needed (Example - UTM Zone 16)
Projection = "+proj=utm +zone=16 +datum=WGS84 +units=m +no_defs
+ellps=WGS84 +towgs84=0,0,0"
AOI = st_transform(AOI, Projection)
AOI = as(AOI, "Spatial")

# Create Template for Final Raster Stack
#For Sentinel-2 we want our final output to be in 20m so use Band
5
raster_B5 = raster(Rasters[[5]])
raster_template = raster(raster_B5)

# Resample, Crop, & Stack All Images
raster_stack = Rasters %>%
 lapply(raster) %>%
 lapply(resample, raster_template) %>%
 lapply(crop, AOI) %>%
 stack()

# Mask cropped image for further reduction of the stacked image
S2_Ref_Image <- mask(raster_stack,AOI)

#Band Subset
S2_Harsha <- S2_Ref_Image[[1:9]] #Sentinel-2 Algorithms Only Use
Bands 1-8A

#Save Final Stacked Image as Tiff
writeRaster(x = S2_Harsha,
 filename= "C:/temp/S2_Harsha_08082016.tif", # save as a tif
 datatype ="FLT4S", # save as a float 4 significant digits
 overwrite = FALSE) #Overwrites same named file


#Option 2: Preprocess Full Image Reflectance Imagery
#Image Directory Folder
Image_Directory =
"C:/temp/S2A_MSIL2A_20180918T154911_N0206_R054_T18STD_20180918T20
5506.SAFE/GRANULE/L2A_T18STD_A016925_20180918T160118/IMG_DATA/R20
m"
# Extacts all raster files from Image Directory with extention
.jp2
Rasters = dir(Image_Directory, pattern = "*.jp2$", full.names =
TRUE)
S2_Ref_Image = stack(Rasters)

#Band Subset
S2_Harsha <- S2_Ref_Image[[1:9]] #Sentinel-2 Algorithms Only Use
Bands 1-8A
```

```
#Save Final Stacked Image as Tiff
writeRaster(x = S2_Harsha,
 filename= "C:/temp/S2_Harsha.tif", # save as a tif
 datatype ="FLT4S", # save as a float 4 significant digits
 overwrite = FALSE) #Overwrites same named file


###Calculate Water Quality indices using waterquality------------
---------------
S2_raster <-
stack("C:/R_Packages/USACE_WQ/Data/S2_Harsha_08082016.tif")
S2_wq <- wq_calc(raster_stack = S2_raster, alg = "all", sat =
"sentinel2")
writeRaster(x = S2_wq,
 filename= "C:/temp/S2_Harsha_WQ_08082016.tif", # save as a tif
 datatype ="FLT4S", # save as a float 4 significant digits
 overwrite = FALSE) #Overwrites same named file
### Extract Values from Raster imagery from Shapefile------------
---------------
#Option 1: Import Shapefile
wq_points <- shapefile('C:/temp/samples.shp')

#Option 2: Create spatial file from spreadsheet
#Read water quality data from csv
library(sp)
wq_df <-
read.csv("C:/R_Packages/USACE_WQ/Data/Harsha_WQ_Measurements_0808
2016.csv")

# Get long and lat from your data.frame. Make sure that the order
is in long/lat.
xy <- wq_df[,c(4,3)]
# Create spatial objects (points) using xy and rest of water
quality data
#define projection! Harsha Lake is in utm zone 16
wq_points <- SpatialPointsDataFrame(coords = xy, data = wq_df,
proj4string = CRS("+proj=longlat +datum=WGS84 +ellps=WGS84
+towgs84=0,0,0"))
##View points on map
#require(mapview)
#mapview(wq_points)
writeOGR(obj=wq_points,
dsn="C:/R_Packages/USACE_WQ/Data/Harsha_wq_points.gpkg",
layer="wq_points", driver="GPKG") # this is in geographical
projection

#Extract values from tiff using points
#Input raster image
raster <-
stack("C:/R_Packages/USACE_WQ/Data/S2_Harsha_WQ_08082016.tif")
##View raster on map
#require(mapview)
#mapview(raster[[1]])
waterquality_data <- data.frame(wq_points, extract(S2_wq,
wq_points))
write.csv(waterquality_data, file =
"C:/R_Packages/USACE_WQ/Data/Harsha_WQ_Algs_08082016.csv")
```

```
### Inspect and Remove mix pixels--------------------------------
---------------
#Manually inspected image for mixed pixels and clouds.
#One sample site (H03) removed because pixel appears mixed land
covers due to beach
#Read Data frame
df <-
read.csv("C:/R_Packages/USACE_WQ/Data/Harsha_WQ_Algs_08082016.csv
")

#Removed H03
df <- df[c(1:2,4:42),]

### Calculate Descriptive Stats and Create Scatterplot matrix----
---------------
#Calculate Summary statistics
df_stats <- pastecs::stat.desc(df[,c(5:7,9,15:18)])

#Scatter Plot of highest performing algorithms
#Optional: Add Text to graph
# Run fit model
fit <- lm(Chl_ugL~BGA_PC_RFU, data = df)
rmse <- round(sqrt(mean(resid(fit)^2)), 2)
coefs <- coef(fit)
b0 <- round(coefs[1], 2)
b1 <- round(coefs[2],2)
r2 <- round(summary(fit)$r.squared, 2)
eqn <- bquote(italic(y) == .(b0) + .(b1)*italic(x) * "," ~~
 r^2 == .(r2) * "," ~~ RMSE == .(rmse))

#Chl & PC & Turbidity Scatterplot Matrix
panel.lm <- function (x, y, pch = par("pch"), col.lm = "red",
...) {
 ymin <- min(y)
 ymax <- max(y)
 xmin <- min(x)
 xmax <- max(x)
 ylim <- c(min(ymin,xmin),max(ymax,xmax))
 xlim <- ylim
 points(x, y, pch = pch,ylim = ylim, xlim= xlim,...)
 ok <- is.finite(x) & is.finite(y)
 if (any(ok))
 abline(lm(y[ok]~ x[ok]),
 col = col.lm, ...)
}

pairs(df[c(15,16,18)], panel=panel.lm)


#Chl & Al10SABI
ggplot(df, aes(Chl_ugL, Al10SABI)) +
 geom_point() +
 geom_smooth(method = "lm",se = FALSE) +
 theme_classic() +
 labs(x = "Chlorophyll (µg/L)", y = "Al10SABI", subtitle = eqn) +
 theme(axis.title.y = element_text(angle = 0, vjust = 0.5))
```

```
#BGA_PC_RFU & Go04MCI
ggplot(df, aes(BGA_PC_RFU, Go04MCI)) +
 geom_point() +
 geom_smooth(method = "lm",se = FALSE) +
 theme_classic() +
 labs(x = "BGA/PC (RFU)", y = "Go04MCI", subtitle = eqn) +
 theme(axis.title.y = element_text(angle = 0, vjust = 0.5))

#Turbidity & Go04MCI
ggplot(df, aes(Turbid_NTU, Go04MCI)) +
 geom_point() +
 geom_smooth(method = "lm",se = FALSE) +
 theme_classic() +
 labs(x = "Turbidity (NTU)", y = "Go04MCI", subtitle = eqn) +
 theme(axis.title.y = element_text(angle = 0, vjust = 0.5))

###Conduct Cross-Validated linear regression Analysis------------
--------------
# Set seed using date created for reproducibility
set.seed(2018-11-09)

# Run LM and Cross-Validation Function
# DONT NOT ALTER
extract_lm_caret = function(y, x, df){
 my_formula = as.formula(paste(y, "~", x))
 caret_model = train(form = my_formula,
 data = df,
 method = "lm",
 na.action = na.exclude,
 #repeated k-fold validation
 trControl = trainControl(method = "repeatedcv",
 number = 3, repeats = 5))
 my_lm = caret_model$finalModel
 CV_R_Squared = getTrainPerf(caret_model)[, "TrainRsquared"]
 RMSE = getTrainPerf(caret_model)[, "TrainRMSE"]
 MAE = getTrainPerf(caret_model)[, "TrainMAE"]
 R_Squared = summary(my_lm)$r.squared
 P_Value = summary(my_lm)$coefficients[8]
 Slope = summary(my_lm)$coefficients[2]
 Intercept = summary(my_lm)$coefficients[1]
 data_frame(R_Squared = R_Squared, Slope = Slope, Intercept =
Intercept, P_Value = P_Value,
 CV_R_Squared = CV_R_Squared, RMSE = RMSE, MAE = MAE)
}

#Define Parameters
#Column numbers of water quality indices from df
indices <- 24:49
#Water quality parameter to evaluate against indices
WQ_parameter <- "Turbid_NTU" #In situ water quality parameter
(Must match name from column in df)

#Run LM and Cross-Validation
Algorithms = names(df)[indices]
names(Algorithms) = Algorithms
Harsha_Turbidity_08082016 = Algorithms %>%
```

```
 map_dfr(~extract_lm_caret(y = WQ_parameter, x = ., df = df),
.id="Algorithms")

### Export Results ------------------------------------------------
---------------
write_csv(Harsha_Turbidity_08082016,
path=".../Harsha_08082016_Turbidity_Results.csv")
```

Acronyms and Abbreviations

BGA	cyanobacteria/blue-green algae
BoA	bottom-of-atmosphere
CHABs	cyanobacterial harmful algal blooms
ELM	Empirical Line Model
ENVI	Environment for Visualizing Images
ESA	European Space Agency
ESRI	Environmental Systems Research Institute
FLAASH	Fast Line-of-sight Atmospheric Analysis of Hypercubes
MAE	mean average error
MERIS	Medium Resolution Imaging Spectrometer
MODIS	Moderate Resolution Imaging Spectroradiometer
NTU	nephelometric turbidity units
PC	Phycocyanin
QUAC	QUick Atmospheric Correction
RFU	reflective units
RMSE	root mean square error
SABI	surface algal bloom index
ToA	top-of-atmosphere
USACE	US Army Corps of Engineers
USEPA	US Environmental Protection Agency
USGS	US Geologic Survey
YSI	Yellow Springs Instrument

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED *(From - To)*
December 2019	Final Report	

4. TITLE AND SUBTITLE
waterquality: An Open-Source R Package for the Detection and Quantification of Cyanobacterial Harmful Algal Blooms and Water Quality

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Richard A. Johansen, Molly Reif, Erich Emery, Jakub Nowosad, Richard Beck, Min Xu, Hongxing Liu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) (see reverse)

8. PERFORMING ORGANIZATION REPORT NUMBER
ERDC/EL TR-19-20

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Aquatic Nuisance Species Research Program
US Army Engineer Research and Development Center
Environmental Laboratory
Vicksburg, MS 39180-6199

10. SPONSOR/MONITOR'S ACRONYM(S)
ANSRP

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
Funding Account Code 96 X 3123; AMSCO Code 008284

14. ABSTRACT
Satellite monitoring of cyanobacterial harmful algal blooms in small freshwater lakes and reservoirs remains challenging. This is partly due to the configurations and resolutions of commonly utilized satellite imagers, which are traditionally designed for large terrestrial applications. The purpose of this report is to provide an efficient methodology for the detection and quantification of harmful algal bloom indicators via remote sensing imagery utilizing the newly developed open-source R package *waterquality*. To accomplish this goal, this report uses Harsha Lake as a case study to demonstrate the use of water quality proxies (chlorophyll-a, phycocyanin, and turbidity) for the evaluation of inland lake and reservoir water quality. This package and associated manuscript were designed to assist researchers and water managers by establishing a flexible and user-friendly workflow to improve water quality monitoring.

15. SUBJECT TERMS
Algal blooms—Monitoring, Environmental management, Remote-sensing images, Reservoirs—Water quality, Water quality management

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Molly Reif
Unclassified	Unclassified	Unclassified	SAR	54	19b. TELEPHONE NUMBER *(Include area code)* 228-252-1134

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) (continued)

University of Cincinnati Libraries
240B Braunstein Hall
Cincinnati, Ohio 45221, USA

ERDC, JALBTCX
US Army Corps of Engineers
Kiln, MS 39556, USA

ERDC, JALBTCX
US Army Corps of Engineers
Kiln, MS 39556, USA

Institute of Geoecology and Geoinformatics
Adam Mickiewicz University
Krygowskiego 10
61-680 Poznan, Poland

Department of Geography and GIS
University of Cincinnati
Cincinnati, OH 45221, USA

Department of Geography
The University of Alabama
204 Farrah Hall
513 University Blvd
Box 870322
Tuscaloosa, AL 35487-0322, USA