Marshall University
Marshall Digital Scholar

Theses, Dissertations and Capstones

2018

Modeling Chlorophyll Concentrations on the Ohio River using Remotely Sensed Data

Thaddaeus Stephen Tuggle
tuggle4@marshall.edu

MODELING CHLOROPHYLL CONCENTRATIONS ON THE OHIO RIVER USING REMOTELY SENSED DATA

A thesis submitted to
the Graduate College of
Marshall University
In partial fulfillment of
the requirements for the degree of
Master of Science
In
Biological Sciences
by
Thaddaeus Stephen Tuggle
Approved by
Dr. Anne Axel, Committee Chairperson
Dr. Gary Schultz
Mr. Steve Foster

Marshall University
May 2018

APPROVAL OF THESIS

We, the faculty supervising the work of Thaddaeus Stephen Tuggle, affirm that the thesis, *Modeling chlorophyll concentrations on the Ohio River using remotely sensed data*, meets the high academic standards for original scholarship and creative work established by the Organismal, Evolutionary and Ecological Biology and the College of Science. This work also conforms to the editorial standards of our discipline and the Graduate College of Marshall University. With our signatures, we approve the manuscript for publication.

Anne C. Axel Dr. Anne Axel, Department of Biology	Committee Chairperson	30 April 2018 Date
Gary E Schultz Jr Dr. Gary Schultz, Department of Biology	Committee Member	4/30/18 Date
Mr. Steve Foster, U.S. Army Corps of Engineers	Committee Member	30 Apr 2018 Date

ACKNOWLEDGMENTS

I want to truly thank my Lord and Savior for blessing me with the opportunity to study more of His Creation. I would also like to thank my wife, Gabrielle for being so supportive throughout this entire journey. Lastly, I would like to thank my parents, Chip and Sharon, for instilling a drive within me at a very early age to learn as much as I can and to strive for excellence in all that I do. Thank you and I love you all!

TABLE OF CONTENTS

List of Tables vii

List of Figures viii

Abstract ix

Chapter 1 1

Introduction 1

Algal Blooms 1

Study Purpose 3

Chapter 2 7

Ohio River Chlorophyll Study 7

Introduction 7

Methods 9

Study Area 9

In situ data 10

Satellite Data 12

Statistical analysis 14

Results 15

Chlorophyll Measurements 15

Models derived from non-cloudy data 15

Cloudy data 19

Discussion 20

Value of Algorithm 22

New Directions 23

References…………………………………………………………………………………..24

APPENDIX A: Office of Research Integrity Approval Letter……………………….……….………27

LIST OF TABLES

Table 1. Chlorophyll concentrations (µg/mL) of the three sites over three sampling days……..15

Table 2. Stepwise regression model results from non-cloudy data for chlorophyll *a*, chlorophyll *ab*, and total chlorophyll.……………………………………..…………………………….16

Table 3. R^2 for models using various satellites estimating chlorophyll concentrations………...21

LIST OF FIGURES

Figure 1. Many variables will contribute to the formation of an algal bloom, but there are still so many unknown variables that cause a particular aquatic environment to produce an algal or cyanobacterial bloom…………………………………………………………………………..…3

Figure 2. Landsat 8 images (USGS) cover a large spatial region, 170 *km* x 183 *km*, resulting in ~325 *km* of Ohio River captured in one satellite image. Polygons in the image on the right reveal the area where samples were taken on the Ohio River, near Huntington, West Virginia…….…...8

Figure 3. One scene of Landsat 8 OLI, as shown in the green image insert, covers Boat Launch, Lock 27, and Harris Park are located within the Greenup Pool of the Ohio River……………...10

Figure 4. A 500 mL Nalgene bottle was attached to the end of a fishing pole and cast into the Ohio River to obtain samples away from shore not using a boat………………..……………....11

Figure 5. Landsat 8 band coverage showing reflectance area. The Aerosal band (1) and the Blue band (2) are next to each other in the spectrum…………………………………………………12

Figure 6. Buffer polygons (200 m x 800 m) were created to extract pixels on the Ohio River from the Landsat 8 images. Polygons were sized to capture the center 50% of the river near the sampling point……….…………………………………………………………………………...14

Figure 7. Actual chlorophyll *a* compared to predicted chlorophyll *a*. Chlorophyll *a* was predicted using band 2 (Blue) and band 5 (NIR) of Landsat 8……………………………………………..17

Figure 8. Actual chlorophyll *a* and *b* compared to predicted chlorophyll *a* and *b*. Chlorophyll *a* and *b* were predicted using band 2 (Blue), band 5 (NIR), and band 6 (SWIR 1) of Landsat 8………………………………………………………………………………………………..18
18

Figure 9. Total chlorophyll compared to predicted total chlorophyll. Total chlorophyll concentrations are predicted by band 2 (Blue), band 5 (NIR), and band 6 (SWIR 1) of Landsat 8………………………………………………………………………………………………..19

Figure 10. Chlorophyll *a* in the cloudy images is not able to be accurately predicted using the non-cloudy derived chlorophyll *a* algorithm. The actual chlorophyll *a* concentrations at the two cloudy sites were 5 and 4.5 µg/L and the algorithm predicted 142.2 and 19.8 µg/L for the sites, respectively……………………………………………………………………………………20

ABSTRACT

Traditional direct water quality methodologies limit the ability to spatially and temporally predict algal blooms in lotic systems due to the size and characteristics of large river systems. Algal blooms potentially can be predicted by knowing the spatial and temporal patterns of change in cyanobacteria concentrations at large scales. Remote sensing studies investigating freshwater algal blooms, some known to secrete harmful toxins, are primarily conducted on lentic systems while large lotic systems are greatly ignored. In this study I developed a chlorophyll concentration estimation model for the Ohio River using a satellite remote sensing approach. Ground-truth water quality measures, including temperature, dissolved oxygen, turbidity, as well as chlorophyll concentrations, were obtained through hand-samples on days the satellite flew over the study area. Concentrations of chlorophyll were correlated with spectral signatures from Landsat-8 OLI satellite imagery. Then a predictive model was developed using two bands of Landsat 8 to predict chlorophyll *a* and the generated model has an $R^2 = 0.879$ (Adj. $R^2 = 0.819$) and a p-value = 0.015. Two other models were generated for estimating both chlorophyll *a* & *b* and total chlorophyll; however, the models were not as robust, $R^2 = 0.801$ (Adj. $R^2 = 0.603$), p-value = 0.141 and $R^2 = 0.764$ (Adj. $R^2 = 0.528$), p-value = 0.18, respectively.

CHAPTER 1

INTRODUCTION

Algal Blooms

Large algal blooms that contain a high percentage of cyanobacteria became an environmental concern over recent decades due in part to their ability to secrete toxins that could potentially harm other aquatic organisms or humans (Carmichael et al., 2001, Hunter, Tyler, Gilvear, & Willby, 2009). However, not all algal blooms with high chlorophyll concentrations that occur have cyanobacteria or toxin present. Toxins are not the only environmental concern; large populations of cyanobacteria or algae can cause anoxic, low oxygen conditions (Havens 2008) and they contribute to decreased biodiversity in aquatic systems (Ribeiro, Andrade, Maizonave, & Crossetti, 2012). These blooms have occurred around the world on every continent, except Antarctica (Carmichael 1992). Algal blooms can become very large and occur in marine, lentic (lake) and lotic (river) systems.

A higher prevalence of these algal and cyanobacterial blooms (often incorrectly referred to as blue-green algae) have caused researchers, state and federal agencies, and many concerned citizens to take an interest in these biological occurrences. Great strides are being made to understand the cause of algal and cyanobacterial growth (Paerl & Fulton, 2006), to investigate the ecology of cyanobacteria (Paerl & Fulton, 2006), to track and monitor movement of cyanobacteria (Kutser, 2004; Kutser, Metsamaa, Strömbeck, & Vahtmäe, 2006; Kutser 2009; Ryan, Davis, Tufillaro, Kudela, & Gao, 2014), and to better understand the risks that are associated with large cyanobacteria blooms (Hunter et al. 2009; Carmichael et. al 2001).

The World Health Organization (WHO) determines what constitutes a ‘bloom’ by the concentration of cyanobacteria (cells/mL) found within the aquatic environment rather than the level of toxin concentration (µg/L) (WHO 2003). According to the World Health Organization, contact with water having cyanobacteria concentration less than 20,000 cells/mL poses a minor health risk to the public. Concentrations of cyanobacteria greater than 20,000 cells/mL, but less than 100,000 cells/mL pose a slightly higher risk; however, exposure that typically results in skin irritation, gastrointestinal illness and other allergenic effects, is not actually due to cyanotoxin toxicity, but rather from other cyanobacterial derived compounds. The World Health Organization has determined cyanobacterial blooms with cell densities of greater than 100,000 cells/mL present a great risk of adverse health impacts due to the cyanobacterial toxin produced (WHO 2003).

Several states have adopted protocols when responding to cyanobacterial blooms; most base their protocols on whether a 'bloom' is present or not by the concentration of cyanobacterial toxins (typically µg/mL). This protocol is different from WHO guidelines, because WHO bases their definition on cell concentration and toxin is not always present when there is a large concentration of cells. For there to be toxins present, there has to be prior cyanobacterial cell growth. This approach makes the WHO guidelines more cautious when it comes to public safety, because it raises public awareness even when a toxin is not present.

Many variables (Figure 1) contribute to cyanobacterial growth: community and population dynamics, intensity of the sun, temperature, and amount of available nutrients (Paerl & Fulton, 2006). However, there are still many unknown variables or combinations of variables that help make up an algal bloom (Graham, Jones, Jones, Downing, Clevenger, 2004). The ability to look at chlorophyll concentrations over a large temporal and spatial scale may reveal unknown dynamics at play in algal bloom creation.

Movements and size of algal blooms in marine (Kutser 2009) and lentic systems (e.g., Dash et al. 2011, Hu, Lee, Ma, Yu, Li, & Shang, 2010) are tracked remotely by detecting chlorophyll *a* in remotely sensed satellite data. However, algal blooms in large lotic systems have been largely overlooked (Simis, Peters, & Gons, 2005; Kutser 2004; Kutser et al. 2006). Measurement and detection of chlorophyll concentrations in rivers pose challenges not typically encountered in marine and lentic systems. Various challenges include narrow width of the river, continual flow of water, and non-stratification of the water column.

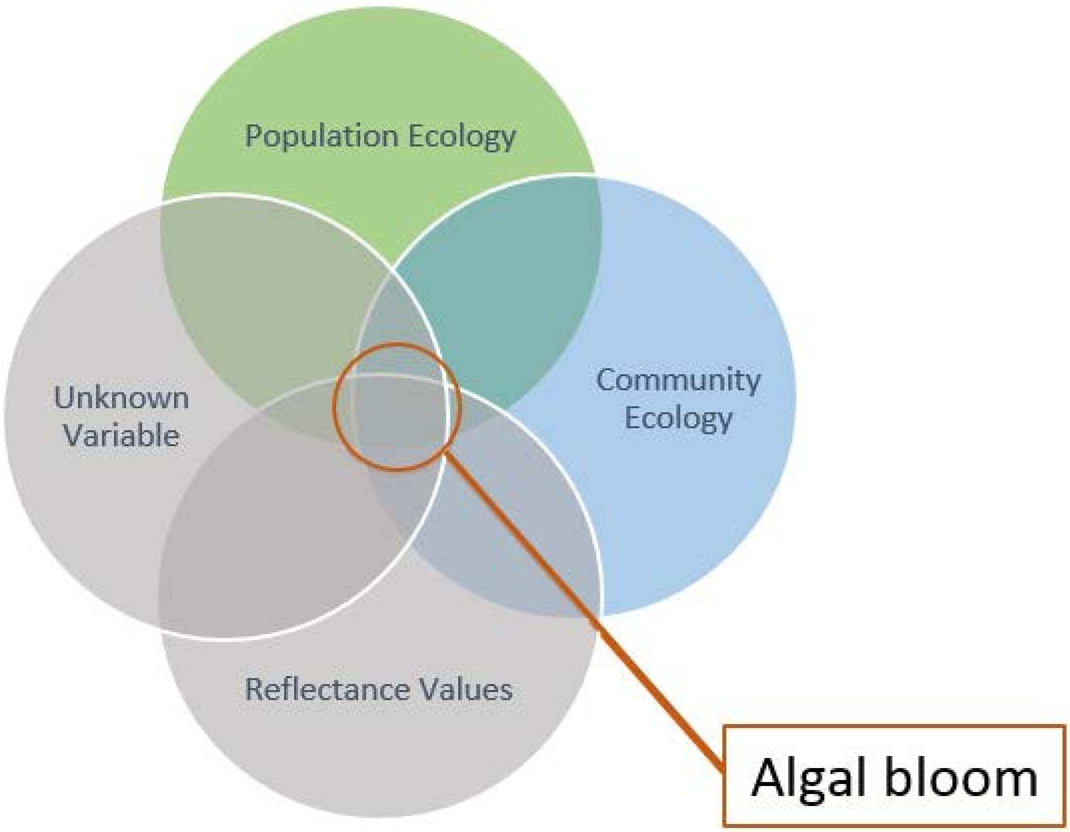

Figure 1. Many variables will contribute to the formation of an algal bloom, but there are still so many unknown variables that cause a particular aquatic environment to produce an algal or cyanobacterial bloom.

Study purpose

Understanding and mapping chlorophyll concentrations within the Ohio River will allow a greater understanding of these cyanobacterial blooms and the general cyanobacterial community. This modeling of chlorophyll concentration is a viable method for estimating cyanobacteria biomass in a large river system. A model was built to remotely predict cyanobacterial concentrations in a larger river system by correlating chlorophyll concentrations calculated from water samples with spectral information collected from Landsat satellite imagery.

Satellite observations of areas within the Ohio River occur on 16-day intervals with Landsat 8 OLI. Short observation intervals are important to not only document how populations are changing within the 16-day time frame, but also because they may also help alert downstream communities of a potential bloom. Most municipals along the Ohio River obtain drinking water from the river, so alerting communities prior to the arrival of high concentrations of chlorophyll (that may potentially carry a cyanobacterial toxin) will be beneficial to the authorities and residents.

Due to the size and characteristics of the river system, the ability to spatially and temporally predict algal blooms in lotic systems is limited using traditional water quality methodologies of collecting chlorophyll readings in unison with dissolved oxygen, pH, turbidity, etc. Manpower, time, and resources needed to cover long stretches of a large river can be financially burdensome for small agencies and researchers on a low budget. Analysis using satellite remote sensing, specifically Landsat 8 OLI imagery, allows researchers to examine river systems at a very large spatial scale (e.g. ~325 river kilometers) within a single scene of the satellite imagery (www.usgs.gov).

Landsat 8 is the most recent space borne satellite in NASA's Landsat program to obtain long-term spatial information of the earth and coverage is expected to continue with the launch of Landsat 9 in 2020. Landsat 8 imagery is freely available and each scene provides coverage of 170 km by 183 km (www.usgs.gov). This large spatial coverage makes it an ideal platform to use for studies of large areas. The satellite has a spatial resolution of 30 m^2 which makes it suitable for use on higher order rivers. The Ohio River, a high order river, is ~0.5 km wide, allowing for approximately sixteen 30 m^2 pixels across the average width of the river. Given the spatial extent of the Ohio River, the higher resolution Landsat data was selected instead of lower resolution sensors typically used for algal detection such as SeaWiFS, MODIS, and AVHRR (250 m^2 to 1 km^2 pixel resolution) (www.usgs.gov).

Using satellite imagery to sample the river using spectral data requires less time and money than water sampling protocols. Remote sensing analysis provides researchers an opportunity to model chlorophyll levels in water on a budget, due to the freely available USGS imagery (www.usgs.gov) and the availability of open source remote sensing software such as GRASS GIS, QGIS, Google Earth Engine, and R.

There are many advantages to using remotely sensed data to understand an aquatic system, but there are also several difficulties that are unique to lotic systems. First, the narrow surface area of rivers, as compared to lakes and oceans, poses a remote sensing challenge. The Ohio River has an average width of ~0.5 km, which is large for most rivers; however, most no-cost satellite imagery has a large spatial resolution (i.e. SeaWiFS, MODIS and AVHRR, 250 m^2 to 1 km^2 pixel resolution) which is not suitable for this narrow a system. Landsat imagery has had a standard pixel size of 30 m^2 since the mid-1980's. This pixel size allows for multiple pixels across the river. Second, canopy cover, bridges, dams, and barge traffic will affect the spectral reflectance of the river. Canopy cover may shade parts of a river.

One of the main characteristics of any river is one-directional flow. In ocean and lentic systems, there is a circulation of water throughout the system due to wind and temperature driven currents, but these currents do not dominate the systems like one-directional flow of a river. Once a river sample is obtained, the particular volume of water that was sampled has moved downstream; consequently the biomass for any particular area in the river is constantly changing. This movement creates a unique challenge for ground-truthing water samples from rivers for satellite remote sensing studies as timing of water sample collection must be made at the very moment (or as close to it) that a satellite flies over the study area. Most ground-truthed samples from ocean and lentic systems are typically collected within 3 days of the satellite flyover (Simis et al., 2007, Yacobi, Giltelson, & Mayo, 1995).

Unlike ocean and lentic systems, it is assumed that this large lotic system was thoroughly mixed (Kovatch, Untitled work on Ohio River Metabolism 2013). This mixing means that the biomass concentration at the surface of the river was representative of the whole column of water and therefore, the biomass of the cyanobacteria was intermixed within the entire water column of the river. In lentic and ocean systems, the photosynthetically active organisms are located at or near the surface of the water, allowing for more accurate estimation of cyanobacteria biomass. Landsat spectral sensors only capture what is occurring on the surface of the river. In low-flow systems, i.e. lentic and ocean systems, cyanobacteria have been shown to regulate their buoyancy using specially regulated gas vesicles within the cell (Kutser, Metsamaa, & Dekker, 2008). Maintaining the proper depth throughout the day allows the cyanobacteria to facilitate photosynthesis throughout the day providing for greater photosynthetic efficiency. This regulation of gas vesicles means cyanobacteria will more likely be near the surface of ocean and lentic systems which allows their spectral reflectance to be detected; however, in lotic systems, the cyanobacteria is most likely distributed throughout the entire water column.

The potential benefits of this project will allow researchers to predict and track cyanobacteria blooms, estimate total biomass of cyanobacteria for the river system, and analyze past algal blooms using historic Landsat imagery. In addition, the model represents a means of monitoring water quality at a large scale. The ability to monitor at large spatial and temporal scales using satellites will help reduce researcher's time and money spent in field sampling.

Tracking cyanobacteria trends over a long temporal scale using historic Landsat data will allow researchers to answer questions about long-term ecological changes and will help to address new questions. Understanding how cyanobacteria populations change over time can benefit ongoing efforts to spatially and temporally predict harmful algal blooms, understand ecological impacts affecting cyanobacteria, and reveal how harmful algal blooms over the last four decades have impacted river ecology.

This research will also help contribute to research on metabolic scaling of the Ohio River. Biomass of an organism coupled with dissolved oxygen readings allows an organism's metabolic rate to be calculated (Brown, Gillooly, Allen, Savage, & West, 2004, Odum 1956). The ability to estimate biomass of cyanobacteria could also determine the carbon sequestration amount in the Ohio River. This particular question has never been addressed on this river and the potential to know how much carbon that is being sequestered in the Ohio River may lead to a better understanding of river ecology

Developing a chlorophyll *a* predictive model for this type of aquatic system will unlock many areas of research within river ecology, but more importantly it will further inform work being done on cyanobacteria and algal growth not just on the Ohio River, but around the world. This model will be useful for universities, state and federal agencies, and any water quality related agency using remote sensing to unlock biological questions within a riverine system.

CHAPTER 2

OHIO RIVER CHLOROPHYLL STUDY

INTRODUCTION

Algal and cyanobacterial blooms (often incorrectly referred to as "blue-green algae") have become more prevalent in recent years. While researchers continue to investigate the cause of these blooms, strides are being made to help detect and model the growth of algal and cyanobacterial blooms. Remotely-sensed satellite data analysis is one promising method of tracking movements and size of algal blooms in marine (Kutser 2009) and lentic systems (e.g., Dash et al. 2011, Hu et al. 2010). Algal blooms in marine systems and large lakes have been tracked remotely by detecting concentrations of chlorophyll *a* (e.g., Bresciani et al. 2012, Hunter et al. 2009, Randolph et al. 2008). However, there are comparatively fewer investigations of blooms in large lotic systems (Simis et al. 2005; Kutser 2004; Kutser et al. 2006).

There are advantages of monitoring algal blooms on the Ohio River using satellite imagery. One scene of Landsat 8 can capture a ~325 km section of the Ohio River (~1,600 km long) per fly over (Figure 2). Researchers can analyze large sections of the river and identify areas with high chlorophyll concentration where the potential for toxicity presence is greater. If an algal bloom occurs on the Ohio River, like the one that occurred in the summer of 2015, researchers will be able to quickly identify areas with high chlorophyll concentration and then target specific field sampling regions.

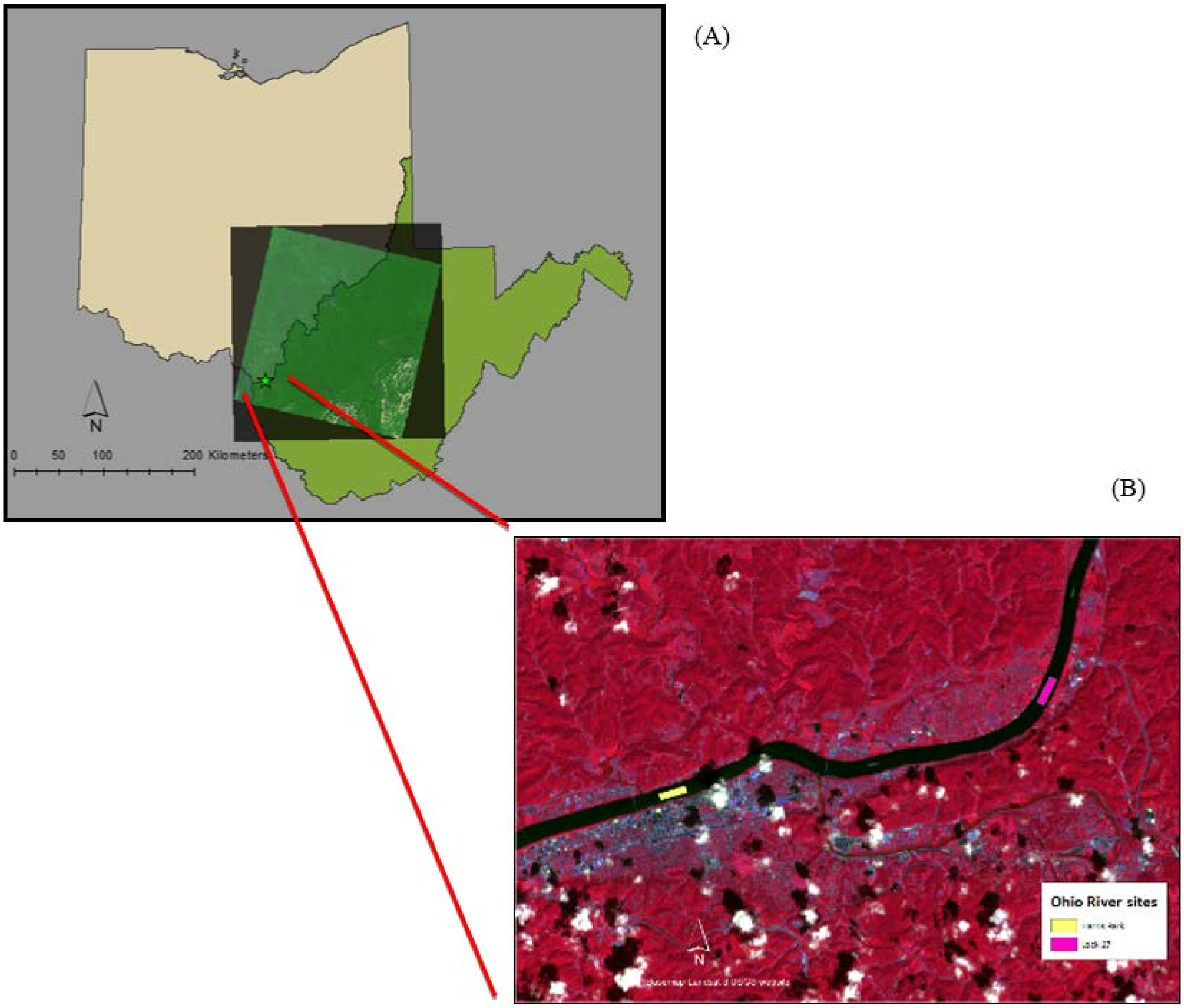

Figure 2. Landsat 8 images (USGS) cover a large spatial region, 170 *km* x 183 *km*, resulting in ~325 *km* of Ohio River captured in one satellite image. Polygons in the image on the right reveal the area where samples were taken on the Ohio River, near Huntington, West Virginia.

While measurement and detection of algae blooms in rivers using remote sensing is promising, the method poses various challenges (Kutser 2009; Kutser et al. 2006) relative to large lake or marine systems. One potential challenge that may occur is the relatively narrow width of the river, which only allows for a few pixels of the satellite across the river. The influence of riparian vegetation may cause shadowing near the river's edge and will give false spectral readings. Also, rapid temporal variation due to the continual flow of water and the non-stratification of the water column could potentially give false spectral readings of the surface of the water.

The main objective of this study was to determine if chlorophyll concentrations in large lotic systems can be predicted using spectral information from Landsat 8 satellite imagery. Single band and multiple band models have been used to detect chlorophyll concentrations, but these models incorporate previous satellite platforms (e.g., Landsat 5 and 7; Han & Jordan 2005), satellites that have a large spatial resolution, ~ 1 km (SeaWiFS) (Dall'Olmo et al. 2005), or hyperspectral platform (Ryan et al. 2014). Landsat 8, launched in February 2013, is freely-available to the public and has a spatial resolution of 30 meters (USGS). The spatial resolution of Landsat 8 makes it ideal for the Ohio River which has an average width of about 0.5 km.

Methods

Study Area

Three sampling locations, Ohio Department of Natural Resources (ODNR) Boat Launch (38°36'41" N, 82°10'25.7" W), Lock 27 (38°27'6.9" N, 82°19'18.3" W), and Harris Park (38°36'41" N, 82°10'25.7" W) (Figure 3) were selected within the Greenup Pool (~110 km long, this pool is one of twenty along the river that is dammed to back up the river to allow for navigation) of the Ohio River (~1600 km long). This large river is located in the Eastern United States and flows into the Mississippi River. The ODNR Boat Launch site was the most upstream site and is about 8 km downstream of the Robert C. Byrd locks and dams in Apple Grove, West Virginia. The most downstream site was Harris Park, which is located along the Ohio River in Huntington, West Virginia.

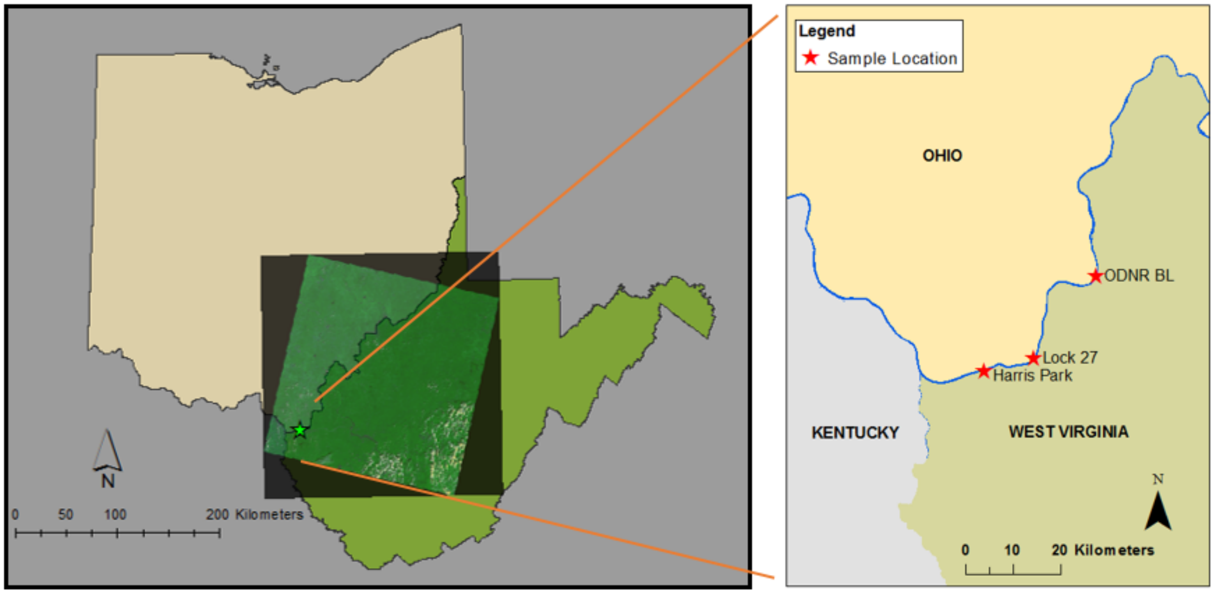

Figure 3. One scene of Landsat 8 OLI, as shown in the green image insert, covers Boat Launch, Lock 27, and Harris Park are located within the Greenup Pool of the Ohio River.

In situ data

The Ohio River is assumed to be homogeneous throughout the main channel of the river (Kovatch, Untitled work on Ohio River Metabolism) resulting in the same amount of chlorophyll concentration on the surface of the river as well as near the substrate. Homogeneity occurs with normal water quality parameters: dissolved oxygen, conductivity, etc. A homogeneous water column is important to note, because the samples collected in this research were collected on the surface, yet the same concentration is expected to be found throughout the entire Ohio River water column. The majority of samples were obtained within 1 minute of the satellite flying over the study area and the remaining samples were collected within an hour. Collecting water samples in a close time frame was an attempt to ensure that the area of the river captured on the satellite image was representative of the water samples.

Figure 4. A 500 mL Nalgene bottle was attached to the end of a fishing pole and cast into the Ohio River to obtain samples away from shore not using a boat.

Water samples were collected using a 500 mL Nalgene bottle attached to the end of a fishing line on a large fishing pole (Figure 4). Sampling with the fishing pole allowed for samples to be taken at or near center channel of the river (>40 m from shore). When the 500 mL bottle was brought back to shore, the water was transferred to a larger container. This process was repeated four times to retrieve two liters of river water for each sample. Once two liters of water were collected, the bottles were wrapped in foil and placed on wet ice which prohibited any outside light source to have any effect on the total chlorophyll concentration. Several water quality parameters were collected *in situ* including dissolved oxygen (DO %), conductivity (μS/cm), temperature (°C), and pH using a YSI 6600 multi-parameter data sonde.

Turbidity (FTU) and chlorophyll (μg L^{-1}) values were calculated in the lab from the two liters of river water collected at the time of satellite flyover. Turbidity (FTU) values were collected using a YSI 9300 photometer, and chlorophyll (μg L^{-1}) concentrations were gathered using chemical extraction according to Standard Method 10200-H (Rice, Baird, Eaton, &

Clesceri, 2012). For chlorophyll concentrations, samples were filtered at the lab within ~2 hours of collection through 0.2 µm PCTE glass fiber filters. The 0.2 µm PCTE filters were used to ensure that the chlorophyll containing synechococcus was captured. A study on the Ohio River found that ~57% of the pelagic bacteria community was comprised of cyanobacteria, with the dominant genus of cyanobacteria being synechococcus (Schultz, Kovatch, & Anneken, 2013) and a cell size 0.5 µm – 0.8 µm (Uysal, 2001). Filters were macerated, steeped for at least 8 hours in 90% alkalized acetone, and then centrifuged. After being centrifuged, the extracted chlorophyll samples were put in a Spectronic Genesys 5 spectrophotometer to obtain spectral readings. These readings were used in a trichromatic equation to determine chlorophyll *a, b,* and *c* (Rice et al. 2012); then these three chlorophyll values were added together to determine total chlorophyll concentration from the sample.

Satellite Data

Satellite imagery is delivered as stacked layers (or bands) collected at different wavelengths of the electromagnetic spectrum (Figure 5).

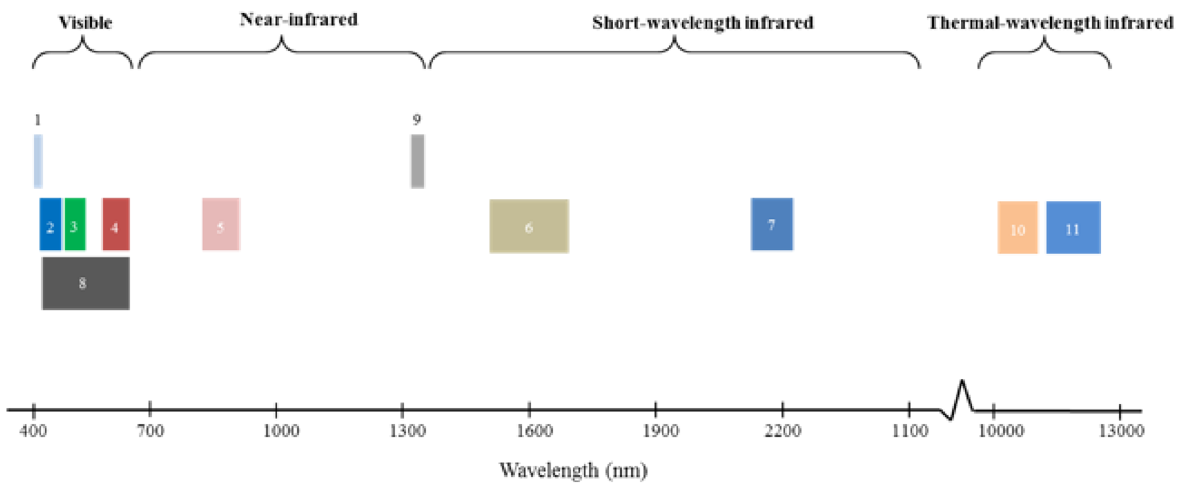

Figure 5. Landsat 8 band coverage showing reflectance area. The Aerosal band (1) and the Blue band (2) are next to each other in the spectrum

Three cloud-free Landsat 8 images were obtained to coincide with *in situ* sampling dates: September 4th (Path 18, Row 33), September 27th (Path 19, Row 33), and October 29th, 2014 (Path 19, Row 33). Landsat 8 flies over every 16 days, but the study area was located in an overlap allowing for more chances to retrieve good quality images. Using ERDAS Imagine

2014, images were pre-processed to Top of Atmospheric (TOA) reflectance and atmospherically corrected using Dark Object Subtraction (DOS).

To ensure the image was representative of the chlorophyll concentration collected from the sample locations, sample points were buffered by polygons designed to match the river's central channel (Figure 6). Based on the average width of the river at sample locations, the center 50% of the river was 200 m across. To account for the flow of the river at the time of the satellite flyover, the length of the polygon was predetermined to be 800 m, making the buffer polygons 200 m wide by 800 m long. Satellite image pixels were extracted from the buffer polygons and then averaged to create a single value.

Figure 6. Buffer polygons (200 m x 800 m) were created to extract pixels on the Ohio River from the Landsat 8 images. Polygons were sized to capture the center 50% of the river near the sampling point.

Statistical analysis

Reflectance values of bands 1-7 and 9 were selected to input into a stepwise regression model as they were deemed to be most biologically relevant. Bands 2-4 were from the visible light spectrum, Bands 1 and 9 were coastal aerosol bands, and Band 5 was near-infrared, and Bands 6 and 7 were short-wave infrared. Reflectance values were extracted from the seven non-cloudy and two cloudy Landsat 8 images. The 179 reflectance values generated from each band were then averaged to a single value for each band and regressed separately against chlorophyll *a*, chlorophyll *a* and *b* and chlorophyll total. Chlorophyll concentrations (µg/L) were predicted using stepwise regression in R Studio as follows: chlorophyll *a*, chlorophyll *a* and *b*, and chlorophyll total. A stepwise regression identifies band combinations that explain the chlorophyll value collected. The best model was selected using Akaike Information Criteria (AIC); the model having the smallest AIC was chosen.

Due to the severe cloudiness during the summer of 2014, there were not many cloud-free satellite images available for the study area. Once the three models were generated from the cloudy-free data, models including the cloudy data were run using the same process. No additional models were successfully generated when the two additional cloudy images were combined with the seven non-cloudy images.

Results

Chlorophyll Measurements

Chlorophyll concentrations varied by site and by date (Table 1). The Ohio River is a large lotic system, so there is a lot of flow, eddies, and upwellings throughout the river. The concentration of chlorophyll decreases as the river flows downstream, the exception being the October 29, 2014 sample. The ODNR boat launch site is only a few kilometers below the R. C. Byrd Locks and Dams, which may have an effect on the overall concentration of chlorophyll.

Table 1. Chlorophyll concentrations (μg/mL) of the three sites over three sampling days.

	9/4/2014	***9/27/2014***	***10/29/2014***
Harris Park	13.2	9.3	7
Lock 27	15	26.4	5.9
ODNR Boat Ramp	24.3	9.9	5.2

Models derived from non-cloudy data

Three models were created from the non-cloudy data. The chlorophyll *a* algorithm was the best fit model (Table 2) (AIC = 33.9) generated, R^2 (adjusted) = 81.9%, is as follows:

$$\boldsymbol{Chla} = (\mathbf{-3.57}) + \mathbf{115.93}(\boldsymbol{Band}\ \mathbf{2}) + \mathbf{143.22}(\boldsymbol{Band}\ \mathbf{5})$$

where Band 2 and Band 5 stand for pixel values generated from a Landsat 8 image that has been pre-processed to Top of Atmospheric (TOA) reflectance haze and a Dark Object Subtraction (DOS).

The second model generated was for chlorophyll *a* and *b* (AIC = 42.9), R^2 (adjusted) = 60.3%, is as follows:

$$\boldsymbol{Chlab} = (\mathbf{-18.95}) + \mathbf{431.57}(\boldsymbol{Band}\ \mathbf{2}) + \mathbf{624.04}(\boldsymbol{Band}\ \mathbf{5}) - \mathbf{838.49}(\boldsymbol{Band}\ \mathbf{6})$$

where Band 2, Band 5, and Band 6 stand for pixel values generated from a Landsat 8 image that has been pre-processed to Top of Atmospheric (TOA) reflectance haze and a Dark Object Subtraction (DOS).

The final model generated with the non-cloudy data is for total chlorophyll (AIC = 48.9), R^2(adjusted) = 52.8%, is as follows:

$$\boldsymbol{Chlt} = (\mathbf{-32.9}) + \mathbf{744.57}(\boldsymbol{Band}\ \mathbf{2}) + \mathbf{1011.53}(\boldsymbol{Band}\ \mathbf{5}) - \mathbf{1491.91}(\boldsymbol{Band}\ \mathbf{6})$$

where Band 2, Band 5, and Band 6 stand for pixel values generated from a Landsat 8 image that has been pre-processed to Top of Atmospheric (TOA) reflectance haze and a Dark Object Subtraction (DOS).

Table 2. Stepwise regression model results from non-cloudy data for chlorophyll *a*, chlorophyll *ab*, and total chlorophyll.

	R^2	Adjusted R^2	F statistic	p-value	AIC
Chlorophyll a	0.879	0.819	14.53 (2,4)	0.015	33.9
Chlorophyll ab	0.801	0.603	4.03 (3,3)	0.141	42.9
Chlorophyll total	0.764	0.528	3.24 (3,3)	0.18	48.9

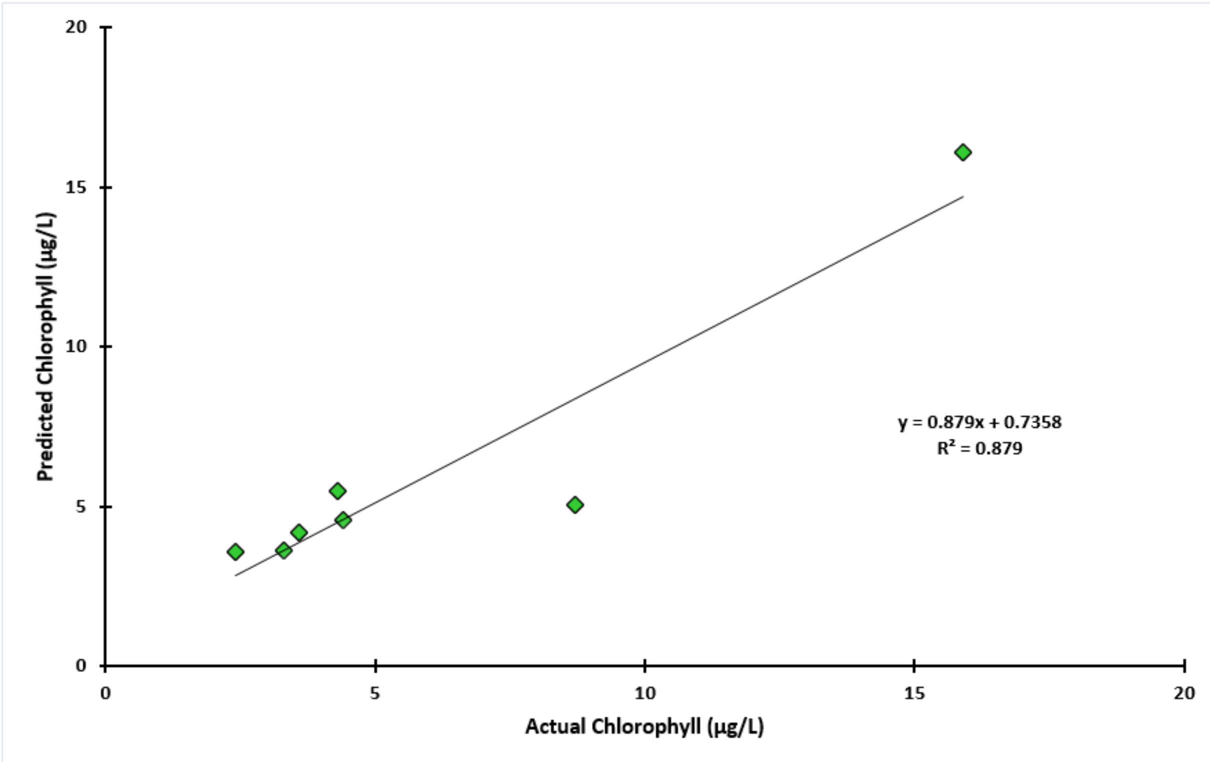

Figure 7. Actual chlorophyll *a* compared to predicted chlorophyll *a*. Chlorophyll *a* was predicted using band 2 (Blue) and band 5 (NIR) of Landsat 8.

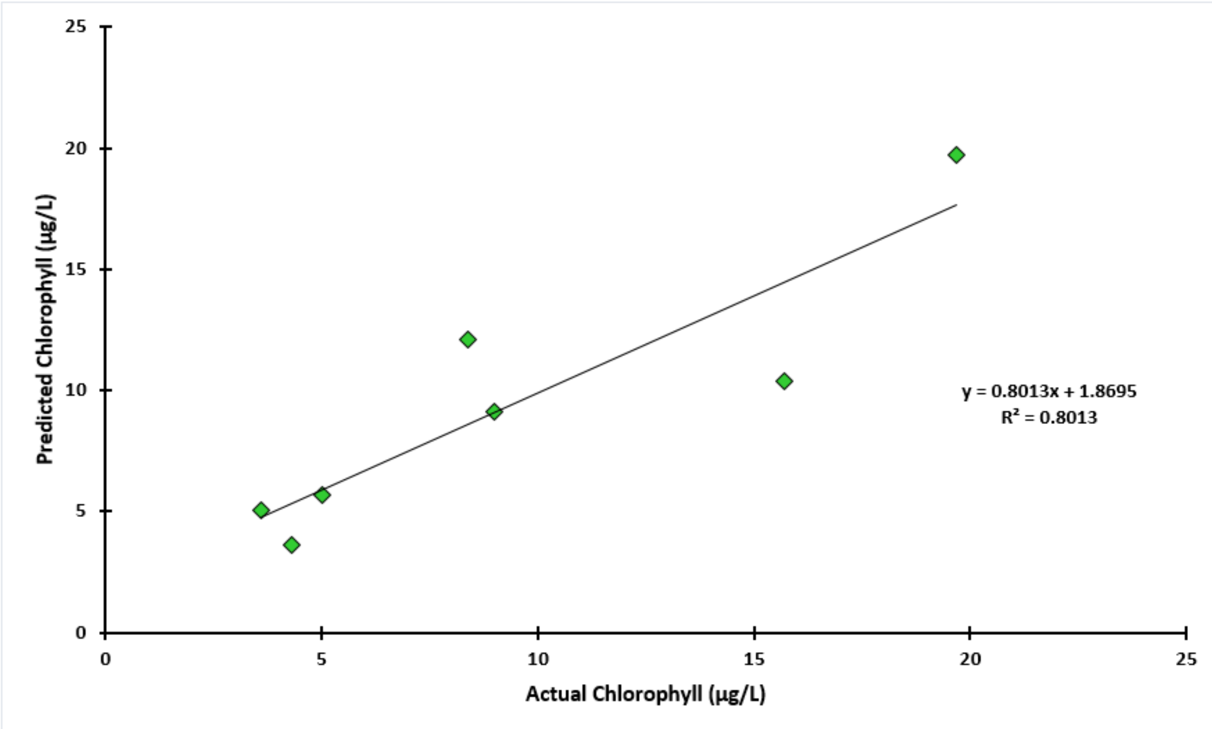

Figure 8. Actual chlorophyll *a* and *b* compared to predicted chlorophyll *a* and *b*. Chlorophyll *a* and *b* were predicted using band 2 (Blue), band 5 (NIR), and band 6 (SWIR 1) of Landsat 8.

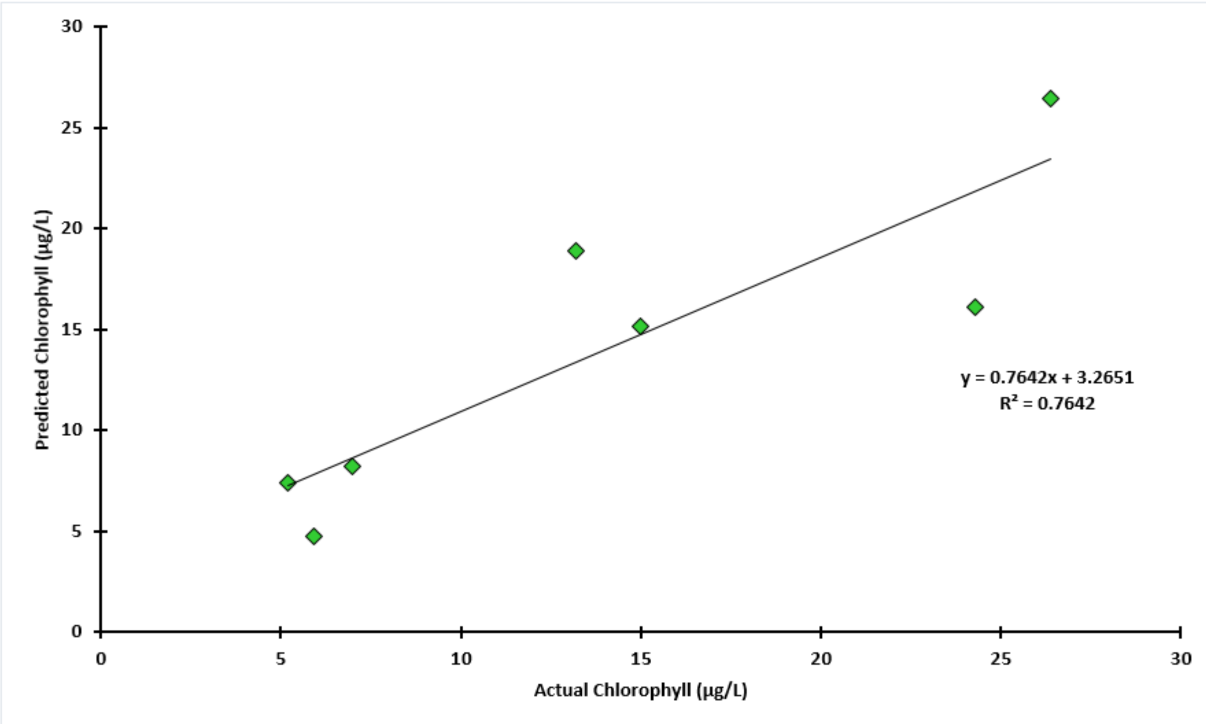

Figure 9. Total chlorophyll compared to predicted total chlorophyll. Total chlorophyll concentrations are predicted by band 2 (Blue), band 5 (NIR), and band 6 (SWIR 1) of Landsat 8.

Cloudy data

In a separate analysis, the data extracted from the cloudy images was added to the data above and the stepwise regression repeated. The model including cloudy satellite data model was unable to accurately predict the chlorophyll concentration (Figure 7). One of the main drawbacks of using this algorithm is the frequently cloudy conditions along the study area.. During the summer of 2014, 45 satellite images were collected over the study area and only 11 of those were somewhat clear of clouds. In the end, only 3 of those 45 were able to be used to help develop the algorithm. The cloudiness of the area may pose a threat to the viability and usability of the algorithm, but on a very clear day, this algorithm will do a great job predicting the concentrations of chlorophyll in the Ohio River.

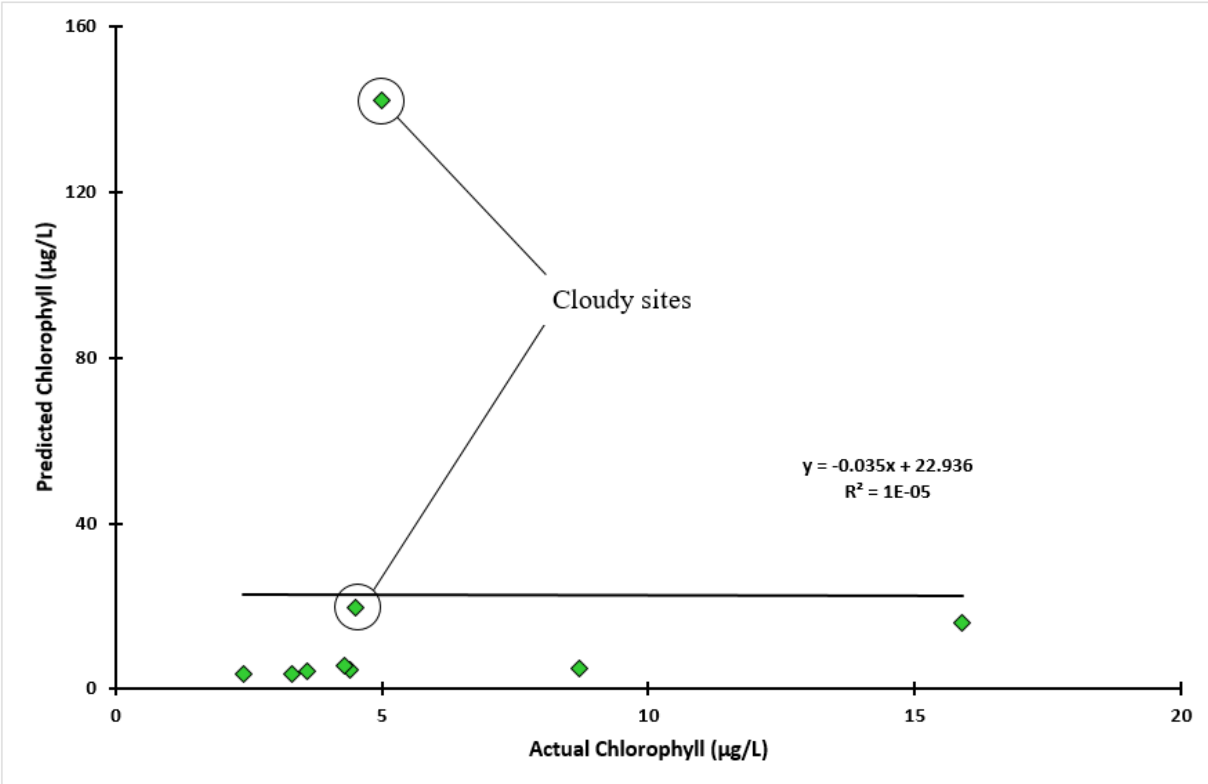

Figure 10. Chlorophyll *a* in the cloudy images is not able to be accurately predicted using the non-cloudy derived chlorophyll *a* algorithm. The actual chlorophyll *a* concentrations at the two cloudy sites were 5 and 4.5 µg/L and the algorithm predicted 142.2 and 19.8 µg/L for the sites, respectively.

DISCUSSION

A fairly robust chlorophyll *a* model was constructed of the Ohio River using Landsat 8 OLI. The most promising finding of this study is that chlorophyll *a* can be estimated on a large lotic system using remotely sensed satellite imagery. Estimating chlorophyll *a* is a huge step forward for monitoring, tracking, and studying algal blooms on the Ohio River and other large riverine systems throughout the world.

There are some limitations to this study, one of them being the cloudy/haziness of some of the images that were collected. Samples were collected from May 2014 through October 2014 and from over the time period there were cloud-free days for this study area on only 3 days of 45 the satellite flew over. A model was attempted to be created from these cloudy data, but no model was created.

The model presented here appears to be as robust as other algorithms created from other earlier satellites (Table 3). It is comparable in goodness-of-fit as a model using MODIS data (Dall'Olmo et al. 2005) and it performs better than those using AVHRR (Svejkovsky & Shandley 2001) or Landsat 7 ETM+ (Han & Jordan 2005),

Table 3: R^2 for models using various satellites estimating chlorophyll concentrations.

System	Satellite	R^2	Author
Ocean	AVHRR	0.64	Svejkovsky & Shandley (2001)
Ocean/Bay	Landsat 7 ETM+	0.67	Han & Jordan (2005)
Lake	MODIS	0.90	Dall'Olmo et al. (2005)
Ohio River	**Landsat 8**	**0.88**	**This study**

Three types of chlorophyll models were generated to test which combination of chlorophyll values would give the most accurate results. Chlorophyll *a*, *ab*, and total all have slightly different spectral signals, which in turn means that different combinations of chlorophyll will yield different model values when predicting the chlorophyll concentration in the satellite image. Chlorophyll *ab* and chlorophyll total models in this study are not as robust as chlorophyll *a*, but when compared to many other remote sensing studies are just as robust (Dall'Olmo et al. 2005, Han & Jordan 2005).

A key limitation to this model is the requirement for a cloudless satellite image. Another limitation is a fairly long satellite repeat cycle (currently 16 days). If this algorithm was to be used by a state agency to help determine if an algal bloom was occurring, they would need more frequent fly-over rates. Typically, most state agencies will need to have an answer within a couple days of the initial report of a potential bloom. There is a potential for a quicker response time if the bloom area is located near an overlap of the satellite scenes which increases the repeat cycle to every 4-8 days depending on the area if it is located in an overlap of two or four images.

A possible limitation with this model is the natural flows, eddies and upwellings of the Ohio River. These types of movements cause chlorophyll to be distributed throughout the entire water column. Chlorophyll is found within the epilimnion of lentic systems where the sunlight is readily available. Most organisms containing chlorophyll can regulate their position within the water column to optimize the amount of energy they receive from the sun; however, in a large river system the constant movement of water impedes this process and possibly distributes the chlorophyll throughout the river and accounts for the difference in chlorophyll concentrations among the sites. Another explanation for the variation in the chlorophyll readings is there are some inputs at different parts of the rivers. These inputs could be from small streams that are carrying nutrients into the river that encourages organismal growth. Light intensity could be another factor affecting the concentration of the chlorophyll in the river. Angle of the sun, shading, and turbidity all contribute to the light intensity of the sun. A higher light intensity means more energy for chlorophyll activity.

Value of algorithm

There are many benefits to this model that predict the amount of chlorophyll *a* within the Ohio River, first being the ability to look at a large spatial scale. The ability to use this model to estimate the amount of chlorophyll *a* in a large section of the Ohio River (~325 km of river per scene) could potentially help researchers ask new questions about the aquatic life and health of a large river.

Two, it is possible to calculate time series of cyanobacterial populations to investigate how concentrations change in the year and in response to specific climatic events. Little is known or studied regarding the cyanobacterial population within the Ohio River throughout the entire year. Monitoring cyanobacterial blooms and cyanobacterial dominate waters is not a new concept (Kutser et al. 2006, Matthews, Bernard, & Winter, 2010, Klemas 2012); however, it has not been successfully tracked on the Ohio River using Landsat 8.

Three, this model potentially unlocks years of historical Landsat data of the Ohio River. Now that chlorophyll can accurately be estimated with Landsat 8 imagery on a large lotic system like the Ohio River, applying this model to historical imagery from previous years could provide the ability to track chlorophyll trends over the past 30 years. Looking back at these trends could help answer questions about the health of the Ohio River related to how phytoplankton populations have varied.

Fourth, this model makes huge strides to help further research on understanding the metabolism of the entire Ohio River. The late Dr. Jeff Kovatch was extremely interested in the metabolism of different organisms, aquatic and terrestrial, but he was especially interested in describing the metabolism of the Ohio River. Estimating the size of the phytoplankton community in the river and pairing that with dissolved oxygen data, the respiration rate of the Ohio River can be determined. From this, the metabolic rate of the River will be able to be determined. Also, by determining the size of the phytoplankton community and the respiration rate of this community, it is possible to estimate the amount of carbon being sequestered within the river.

Lastly, this model provides an accessible and affordable alternative to costly field sampling. Researchers looking to study questions at large spatial and temporal scales can use similar software and freely available Landsat OLI imagery at a relatively low cost. One of the biggest positives about the imagery is that Landsat 8 images are higher resolution (30 m) as compared to other satellites used for chlorophyll estimation (e.g., MODIS).

New Directions

In December 2020, NASA will launch Landsat 9 on a 16-day repeat cycle that will be staggered with the current Landsat 8 cycle. Assuming Landsat 8 remains operational, this will increase temporal resolution of Landsat data collection to an 8-day repeat cycle. Increased temporal resolution will allow for more frequent monitoring of cyanobacterial or algal growth on the Ohio River.

Landsat 8 (and forthcoming 9) is able to reliably detect chlorophyll *a* (720 nm). Future research should focus on sensors that can reliably detect phycocyanins (620 nm). Tracking phycocyanin, a unique photosynthetic pigment within cyanobacteria, will open doors to monitoring cyanobacteria that is moving through the Ohio River. The ability to track such an organism will allow researchers, state and federal agencies a better look into the population dynamics of cyanobacteria in riverine systems.

REFERENCES

Bresciani, Mariano, Giardino, Claudia, Stroppiana, Daniela, Pilkaitytė, Renata, Zilius, Mindaugas, Bartoli, Marco & Razinkovas, Artūras (2012). Retrospective analysis of spatial and temporal variability of chlorophyll-*a* in the Curonian Lagoon. *Journal of Coastal Conservation* 16, 511-519.

Brown, J.H., Gillooly, J.F., Allen, A.P., Savage, V.M., & West, G.B. (2004). Toward a Metabolic Theory of Ecology. *Ecology* 85, 1771-1789.

Carmichael, W.W. (1992). Cyanobacteria secondary metabolites – the cyanotoxins. *Applied Microbiology*, Volume 72, Issue 6, 445-459

Carmichael, W.W., Azevedo, S.M.F.O., An, J.S., Molica, R.J.R., Jochimsen, E.M., Lau, S., Rinehart, K.L., Shaw, G.R., & Eaglesham, G.K. (2001). Human Fatalities from Cyanobacteria: Chemical and Biological Evidence for Cyanotoxins. *Environmental Health Perspectives* 109(7), 663-668.

Dall'Olmo, Giorgio, Gitelson, Anatoly A., Rundquist, Donald C., Leavitt, Bryan, Barrow, Tadd, & Holz, John C. (2005). Assessing the potential of SeaWiFS and MODIS for estimating chlorophyll concentration in turbid productive waters using red and near-infrared bands. *Remote Sensing of Environment* 96, 176-187.

Dash, P.,Walker, N. D., Mishra, D. R., Hu, C., Pinckney, J. L., & D'Sa, E. J. (2011). Estimation of cyanobacterial pigments in a freshwater lake using OCM satellite data. *Remote Sensing of Environment* 115, 3409-3423.

Graham, J., Jones, J., Jones, S., Downing, J., Clevenger, T. (2004). Environmental factors Influencing microcystin distribution and concentration in the Midwestern United States *Water Research* 38, 4395-4404.

Han, Luoheng & Jordan, Karen J. (2005). Estimating and mapping chlorophyll-*a* concentration in Pensacola Bay, Florida using Landsat ETM+ data. *International Journal of Remote Sensing* 26(23), 5245-5254.

Havens, K.E. (2008). Cyanobacteria blooms: effects on aquatic ecosystems. *Advances in experimental medicine and biology*. 619, 733-747.

Hu, C., Lee, Z., Ma, R., Yu, K., Li, D., & Shang, S. (2010) Moderate Resolution Imaging Spectroradiometer (MODIS) Observations of Cyanobacteria Blooms in Taihu Lake, China. *Marine Science Faculty Publications*. Paper 40.

Hunter, P.D., Tyler, A.N., Gilvear, D.J., & Willby, N.J. (2009). Using Remote Sensing to Aid the Assessment of Human Health Risks from Blooms of Potentially Toxic Cyanobacteria. *Environmental Science & Technology* 43(7), 2627-2633.

Klemas, Victor (2012). Remote Sensing of Algal Blooms: An Overview with Case Studies. *Journal of Coastal Research* 28, 1A, 34-43.

Kovatch, J.J. (2013). (Untitled work on Ohio River Metabolism). Marshall University, Department of Biological Sciences.

Kutser, T. (2009). Passive optical remote sensing of cyanobacteria and other intense phytoplankton blooms in coastal and inland waters. *International Journal of Remote Sensing*. 30, 4401-4425.

Kutser, T., Metsamaa, L., & Dekker, A.G. (2008). Influence of the vertical distribution of cyanobacteria in the water column on the remote sensing signal. *Estuarine, Coastal and Shelf Science* 78, 649-654.

Kutser, T., Metsamaa, L., Strömbeck, N., & Vahtmäe, E (2006). Monitoring cyanobacterial blooms by satellite remote sensing. *Estuarine, Coastal and Shelf Science* 67, 303-312.

Kutser, T. (2004). Quantitative detection of chlorophyll in cyanobacterial blooms by satellite remote sensing. *Limnology and Oceanography* 49(6), 2179-2189.

Matthews, Mark W., Bernard, Stewart, & Winter, Kevin (2010). Remote sensing of cyanobacteria-dominant algal blooms and water quality parameters in Zeekoevlei, a small hypertrophic lake, using MERIS. *Remote Sensing of Environment* 114, 9, 2070-2087.

Odum, H.T. (1956). Primary Production of Flowing Waters. *Limnology and Oceanography* 1(2), 102-177.

Paerl, H.W. & Fulton III, R. S. (2006). Ecology of Harmful Cyanobacteria. *Ecology of Harmful Algae* 189, 95-109.

Randoph, Kaylan, Wilson, Jeff, Tedesco, Lenore, Li, Lin, Pascual, D. Lani, & Soyeux, Emmanuel (2008). Hyperspectral remote sensing of cyanobacteria in turbid productive Water using optically active pigments, chlorophyll *a* and phycocyanin. *Remote Sensing of Environment* 112, 4009-4019.

Ribeiro, G.F., Andrade, R. da Rocha, Maizonave, C.R.M., & Crossetti, L.O. (2012). Effects of cyanobacterial summer bloom on the phytoplankton structure in an urban shallow lake, Guaíba Lake, southern Brazil. *Neotropical Biology and Conservation* 7(2), 78-87.

Rice, Eugene W., Baird, Rodger B., Eaton, Andrew D., Clesceri, Lenore S. (2012). *Standard Methods for the Examination of Water and Wastewater 22nd Edition*, 1022-1030.

Ryan, J.P., Davis, C.O., Tufillaro, N.B., Kudela, R.M., & Gao, B.C. (2014). Application of the Hyperspectral Imager for the Coastal Ocean to Phytoplankton Ecology Studies in Monterey Bay, CA, USA. *Remote Sensing* 6(2), 1007-1025.

Schultz, G. E. Jr., Kovatch, J.J., & Anneken, E. M. (2013). Bacterial diversity in a large, temperate, heavily modified river, as determined by pyrosequencing. *Aquatic Microbial Ecology* 70, 169-179.

Simis, S.G.H., Peters, S.W.M., & Gons, H.J. (2005). Remote sensing of the cyanobacterial pigment phycocyanin in turbid inland water. *Limnology and Oceanography*. 50(1), 237-245.

Simis S.G.H., Ruiz-Verdu, A., Domiguez-Gomez, J.A., Pena-Martinez, R., Peters, S.W.M., & Gons, H.J. (2007). Influence of phytoplankton pigment composition on remote sensing of cyanobacterial biomass. *Remote Sensing of Environment*. 106, 414-417.

Svejkovsky, J. & Shandley, J. (2001) Detection of offshore plankton blooms with AVHRR and SAR imagery. *International Journal of Remote Sensing*, Volume 22, Issue 2, 471-485.

Uysal, Zahit (2001). Chroococcoid cyanobacteria Synechococcus spp. in the Black Sea: pigments, size, distribution, growth and diurnal variability. *Journal of Plankton Research* 23 (2): 175-190.

USGS – U.S. Geological Society https://earthexplorer.usgs.gov/

World Health Organization (2003). Guidelines for Safe and Recreational Water Environments. *Coastal and Fresh Water* Chapter 8: Algae and Cyanobacteria in Fresh Water, Volume 1.

Yacobi, Y.Z., Giltelson, A., & Mayo, M. (1995). Remote sensing of chlorophyll in Lake Kinneret using highspectral-resolution radiometer and Landsat TM: spectral features of reflectance and algorithm development. *Journal of Plankton Research* 17(11), 2155-2173.

APPENDIX A: OFFICE OF RESEARCH INTEGRITY APPROVAL LETTER

MARSHALL UNIVERSITY.
www.marshall.edu

Office of Research Integrity

November 27, 2017

Thaddaeus Tuggle
Robert C. Byrd Locks and Dam
Water Quality Operations Center
1744 R. C. Byrd Drive
Gallipolis Ferry, WV 25515

Dear Mr. Tuggle:

This letter is in response to the submitted thesis abstract entitled *"Modeling Chlorophyll Concentrations on the Ohio River using Satellite Imagery"*. After assessing the abstract, it has been deemed not to be human subject research and therefore exempt from oversight of the Marshall University Institutional Review Board (IRB). The Code of Federal Regulations (45CFR46) has set forth the criteria utilized in making this determination. Since the information in this study does not involve human subjects as defined in the above referenced instruction, it is not considered human subject research. If there are any changes to the abstract you provided then you would need to resubmit that information to the Office of Research Integrity for review and a determination.

I appreciate your willingness to submit the abstract for determination. Please feel free to contact the Office of Research Integrity if you have any questions regarding future protocols that may require IRB review.

Sincerely,

Bruce F. Day

Bruce F. Day, ThD, CIP
Director

WE ARE... MARSHALL.

One John Marshall Drive • Huntington, West Virginia 25755 • Tel 304/696-4303
A State University of West Virginia • An Affirmative Action/Equal Opportunity Employer