Validation of 2015 Lake Erie MODIS image spectral decomposition using visible derivative spectroscopy and field campaign data

Dulcinea M. Avouris*, **Joseph D. Ortiz**

Department of Geology, Kent State University, Kent OH

Kent State Department of Geology, 221 McGilvrey Hall, 325 S. Lincoln St., Kent, OH 44242 USA

*Corresponding author: davouris@kent.edu; voice 1- 330 672-2680

Abstract

Timely identification of color-producing agents (CPAs) in Lake Erie is a challenging, but vital aspect of monitoring Harmful Algal Blooms (HABs). In particular, HABs that include large amounts of cyanobacteria (CyanoHABs) can be toxic to humans, posing a threat to drinking water, in addition to recreational and economic use of Lake Erie. The optical signal of Lake Erie is complex (Becker et al., 2009; Moore et al., 2017), typically comprised of phytoplankton, cyanobacteria, colored dissolved organic matter (CDOM), detritus, and terrigenous inorganic particles, varying in composition both spatially and temporally. The Kent State University (KSU) spectral decomposition method effectively partitions CPAs using a varimax-rotated, principal component analysis (VPCA) of visible reflectance spectra measured using lab, field or satellite instruments (Ali et al., 2013; Ortiz et al., 2017, 2013). We analyze 2015 imagery acquired by the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor and field samples collected during the early 2015 cyanoHAB season. We identified four primary CPA spectral signatures, and the spatial distribution of each identified CPA, in the reflectance spectra datasets of both the MODIS and lab-measured water samples. The KSU spectral decomposition method results in mixtures of specific pigments, pigment degradation products, and minerals that describe the optically complex water. We found very good agreement between the KSU VPCA spectral decomposition results and *in situ* measurements, indicating that this method may be a powerful tool for rapid CyanoHAB monitoring and assessment in large lakes using instruments that provide moderate resolution imagery (0.3 to 1 km^2).

Keywords: remote sensing; MODIS; cyanobacteria; Harmful Algal Blooms; Lake Erie; VPCA

Introduction

Lake Erie is an economic and social resource, providing much of the regional drinking water for surrounding communities, particularly in Northern Ohio. Lake Erie is divided into three basins (Figure 1A) formed during the last glacial advance, which differ in average depth, due to bedrock geology (Sly, 1976). The Western Basin of Lake Erie (WBLE) extends from the western coast of the lake to a line drawn from Point Pelee, ON to Lorain, OH, and is the shallowest sub-basin of Lake Erie. Sandusky Bay, the largest marginal basin in Lake Erie, is located along the southern shore of the western basin. The Detroit and Maumee Rivers are the major rivers that discharge into the WBLE, while the Sandusky River, the third largest source of fluvial input to Lake Erie, feeds in via Sandusky Bay. The Maumee watershed, located primarily in Ohio and Indiana, but which also includes a small portion of southern Michigan, is dominated by agricultural activity as is the Sandusky River watershed. Perennial toxic cyanobacterial and Harmful algal blooms (CyanoHABS) develop during the summer as lake conditions become increasingly eutrophic due to nutrient input from the rivers that feed into Lake Erie (Baker et al., 2014). The shallow depth, warm water, and input of agriculturally-derived nutrients from the Maumee and Sandusky Rivers combine to create ideal conditions for algal and cyanophyte growth in the Western Basin and Sandusky Bay (Paerl and Otten, 2013).

Of specific concern are blooms containing toxic cyanobacteria. The toxin, microcystin, produced by *Microcystis* and *Planktothrix*, which flourish in Lake Erie and Sandusky Bay respectively, can cause liver damage, while *Lyngbya wollei* produces a freshwater analogue of saxitoxin, which can damage the nervous system (Carmichael and Boyer, 2016). The Maumee River is the source of the heavy springtime nutrient loading to the Western Basin that drives the predominantly *Microcystis* CyanoHAB; the Sandusky River drives the *Planktothrix* CyanoHAB that forms in Sandusky Bay (Davis et al., 2015; Ho and Michalak, 2017; Kane et al., 2014; Stow et al., 2015; Watson et al., 2016; Wetzel et al., 2003). Significant CyanoHABS have become an almost yearly occurrence over the last decade, presenting serious risks to human and animal health, and can impart an unpleasant taste and odor to drinking water (Becker et al., 2009; Stumpf et al., 2012). The impact of CyanoHABs on drinking water supplies was clearly demonstrated during the water use ban issued by the city of Toledo, OH on 2 August 2014 (Steffen et al., 2017), when microcystin exceeded 1μg L^{-1}, the safe level as established by the World Health Organization (WHO). The Toledo municipal water supply was shut down, leaving nearly 500,000 people without running water for three days, until the ban ended on 4 August 2014. The shutdown was enforced until lines could be flushed and appropriate treatment measures implemented to prevent the CyanoHAB bloom that continued to surround the water intake at the Toledo Crib from further contamination of the water supply. This situation highlights that rapid identification of color producing agents (CPAs) that include CyanoHABs in the lake's waters is vitally needed to provide timely data for water management purposes, to establish recreational use warnings such as beach closures, and for the detection of CyanoHABs. In addition to cyanobacteria, other CPAs present may include other phytoplankton pigments, colored dissolved organic matter (CDOM), phytodetritus, and terrigenous inorganic particles, all contributing to the complexity of the optical signal. The need for effective analysis methods to distinguish between CPAs is emphasized by the similarities in optical characteristics of some of these CPAs.

Remote Sensing of CyanoHABs

Historically, the optical complexity of Lake Erie has made remote sensing applications challenging (Ali et al., 2013; Ali and Ortiz, 2016; Becker et al., 2009; Moore et al., 2017; Ortiz et al., 2013; Witter et al., 2009). One aspect of this challenge is that both cyanobacteria and other phytoplankton contain chlorophyll *a* (hereafter, chl a), making that primary pigment less than ideal for distinguishing between potentially toxic CyanoHABs and nuisance blooms. To address that challenge, we focus on identification of chl *a* as well as accessory pigments that distinguish algal and cyanobacterial groups. In addition, various types of sediment as well as CDOM have been shown to have unique spectral characteristics and can be identified by spectroscopic analysis (Balsam and Deaton, 1996; Clark, 1995; Clark et al., 2003; Deaton and Balsam, 1991; Menken et al., 2006). To identify the mixtures of CPAs present, we apply the Kent State University (KSU) spectral decomposition method (Ortiz, 2011; Ortiz et al., 2013), which uses varimax-rotated, principal component analysis (VPCA) to differentiate constituents based on their spectral characteristics (Davis, 1986; Kaiser, 1958).

VPCA spectral decomposition

Principal component analysis is a dimensional reduction statistical tool that reduces a multivariate dataset to a few independent principal components. This method is ideal for analyzing highly correlated data as well as separating the environmental signal from the noise present in multivariate datasets, such as reflectance spectra obtained from optically complex waters (Ortiz et al., this issue, 2017, 2013). In this application, the VPCA functions as a data-adaptive filtering and unsupervised, soft, classification system (Romero et al., 2016; Singh, 1989; Xia et al., 2014).

A VPCA of the visible spectrum (400-700 nm) from the Medium Resolution Imaging Spectrometer (MERIS) sensor has produced excellent results in detection of chl a concentrations (Ali et al., 2014), leading to an effective method for discriminating between various CPAs in optically complex waters. All reflectance bands in the visible are highly correlated, which requires use of some method to deconvolute spectral signals (Ortiz et al., this issue, 2017, 2013). While MERIS ceased operations in 2012, the spectral bands of the MODIS instruments are also suitable for detection of HABs in Lake Erie, particularly by spectral shape decomposition methods like VPCA that work based on the correlation structure of the data set (Ortiz et al., this issue, 2017, 2013). Furthermore, MODIS imagery has been successfully used in various algorithms focused on detection of CyanoHABs and retrieval of chl a (Becker et al., 2009; Binding et al., 2012; Shuchman et al., 2006, 2013; Stumpf et al., 2016; Wynne et al., 2013, 2010; Witter et al., 2009). This previous work provides background and motivation for our study.

MODIS sensors

Complementary MODIS sensors are currently orbiting onboard the NASA Aqua and Terra satellites. Our dataset is comprised of images acquired by MODIS Aqua. The Aqua satellite is on the ascending node, acquiring data at 1:30 pm local time. This sensor provides daily imagery sampled around the entire Earth and data is freely available within hours of data acquisition. MODIS bands 8-14 have a high signal to noise ratio (SNR > 750), and are used in our analysis (NASA MODIS specifications: https://modis.gsfc.nasa.gov/about/specifications.php). Bands 1-4, designed for Land/Cloud/Aerosol Boundary detection, while also within the VNIR region of interest, have lower SNR (<250), which makes them less suitable for our purpose in comparison to the higher SNR bands. The MODIS instruments have a ground resolution of 1 km^2, which is adequate for study of large water bodies like Lake Erie.

2015 spring meteorological conditions

According to the National Oceanic and Atmospheric Administration (NOAA) National Center for Environmental Information, 2015 spring precipitation in Indiana and Ohio was average to above average in March, April, and May, 2015. June 2015 was a record wettest month for both states, based on average precipitation records for the period of 1895-2017 (https://www.ncdc.noaa.gov/temp-and-precip/us-maps/). This intense precipitation caused a peak discharge of 2806 m^3 s^{-1} from the Maumee River on 30 June 2015, recorded at USGS gauge #04193500 located in Waterville, OH just upstream of Toledo, OH (National Water Information System https://waterdata.usgs.gov/nwis). As a result of the spring discharge, the WBLE had high levels of suspended sediment through June and early July, frequently re-suspended by strong winds associated with storms during the spring precipitation season. The discharge contributed to high nutrient loading, but delayed bloom conditions due to sediment turbidity and thus, low light conditions (Ho and Michalak, 2017; Watson et al., 2016). The suspended sediment is visible in the MODIS imagery from this time, contributing to the complexity of the optical signal.

Project Objectives

The multi-faceted problem of CyanoHAB occurrence in Lake Erie, and the challenge of identifying in-water constituents in a timely fashion, defines the purpose of this study. We are interested in providing a methodology to aid in rapidly determining the composition of CPAs in Lake Erie, and their distribution at times in the year before conventional remote sensing methods can be applied (due to the presence of mixed spectral signatures). This will allow earlier incipient bloom identification at lower detection limits. To that end, this study brings together remote sensing and *in situ* field datasets for analysis and comparison. Here we present *in situ* field and lab-filtered data for comparison with near-concurrent MODIS imagery. The lab-filtered water samples were collected by KSU, while the *in situ* field measurements were collected by the Cooperative Institute for Great Lakes Research (CIGLR), a multi-institution collaborative affiliated with NOAA Great Lakes Environmental Research Laboratory (NOAA-GLERL). Four reasonably cloud-free images were acquired by MODIS Aqua on 21, 23, 27, & 28 July 2015. These images span a 1 – week time period, and comprise our remote sensing dataset which documents the inception of the bloom in mid-July as reported by the NOAA HAB Bulletin. The HAB Bulletin, using the NOAA-CI algorithm (Wynne et al., 2013), is produced by NOAA CO-OPS, and provides information on current and predicted CyanoHAB locations (Lake Erie HAB-OFS Bulletin Guide https://tidesandcurrents.noaa.gov/hab/hab_publication/Lake_Erie_HAB_Bulletin_Guide.pdf). We decompose the reflectance spectra of both remote sensing data and lab-measured field samples collected by KSU to extract the primary components of the optical reflectance signal. We then compare our results with the field and lab measurements to validate our analysis. We find that spectral decomposition of remote sensing imagery is a highly effective method for distinguishing the primary CPA signals in optically complex waters, and can be used in conjunction with well-planned, field sampling campaigns to assess CyanoHAB formation and development.

Methods

Field Data Collection

We collected water samples at 11 locations in Sandusky Bay and along the coast between Lorain and Sandusky, OH (Figure 1B, Table 2). These sites were re-visited weekly from 8 June – 2 September 2015. Surface water was collected at each site using a bucket to skim the surface (~10-30 cm) and poured into 1 L Nalgene sample bottles. The samples were transported to the lab on ice, and filtered immediately through a 47 mm GF/F filter. The lab processing protocol followed that of Ortiz et al. (2013). These filters retain particulate matter larger than 0.7 μm from the surface water sample, but excludes all CDOM from the sample. The GF/F filter was then oven dried, to remove all water from the particulate matter. The samples were then measured using a benchtop ASD (now Malvern Panaltyical) LabSpec Pro FR UV/VIS/NIR spectrometer, equipped with a high-intensity contact probe, which has an internal light source of known properties. This instrument measures reflectance from 350 nm - 2500 nm, at ~4 nm spectral resolution in the visible and 10 nm in the NIR. To exclude any ambient light during measurement, each GF/F filter was placed on top of a circular Spectalon^{TM} plate (9 cm diameter) sitting on a scissors jack, and the instrument probe was brought into contact with the GF/F filter. Results are oversampled by the ASD software and reported at 1 nm resolution to fully characterize peak position and amplitude. A total of 900 measurements (collected in groups of 30) were averaged to produce a single, high signal-to-noise (SNR) reflectance spectra for each GF/F sample. A blank GF/F filter was also measured in the same manner during each processing run, to determine the spectral characteristics of the filter paper. We removed the filter signal from the measured signal by dividing the sample measurement by the blank measurement, isolating the spectral response of the particulate matter (Ortiz et al., 2013). The visible part (400-700 nm) of the measured reflectance spectra were extracted from the full spectrum and collected into a single database. We then interpolated this dataset from 1 nm spectral resolution to 10 nm over the spectral range from 400 nm to 700 nm, to further increase the SNR, and to remove the effect of spectral oversampling by the ASD software, while maintaining the hyperspectral nature of this dataset. We then calculated the center-weighted, first derivative for each sample. In addition, the 1 nm resolution spectra were band-averaged to replicate MODIS bands 8-14, using the MODIS bandwidth definitions, and band naming conventions. The center-weighted, first derivative of each simulated MODIS-resolution GF/F reflectance spectra was calculated for each sample in the dataset for comparison. The lab measurements from these field-collected samples are referred to as the "GF/F samples" throughout this paper, and the MODIS resolution GF/F sample spectra are referred to as GF/F_{MODIS}.

In situ data was collected by CIGLR from 8 June to 5 October 2015 on a nearly weekly basis. Samples are collected at fixed sampling sites (Figure 1B, Table 3). In addition to Secchi depth, other parameters measured in the field with a SeaBird CTD include: temperature, specific conductivity, beam attenuation, transmissivity, chl *a*, phycocyanin, and dissolved oxygen. In addition, water samples were collected at each site, and measured in the CIGLR lab for extracted phycocyanin, particulate and dissolved microcystin, turbidity (using a HACH benchtop turbidimeter), and extracted chl *a* using a Turner 10AU fluorometer.

MODIS Imagery Data – acquisition and pre-analysis processing

We obtain all our MODIS imagery from the NASA Ocean Biology Processing Group (OBPG) at NASA Goddard, which provides MODIS data products, and the open-source processing software SeaDAS, through the Ocean Color Web (https://oceancolor.gsfc.nasa.gov). The remote sensing images are obtained as a Level-1 product, and processed to Level-2 remote sensing reflectance using SeaDAS, which facilitates further satellite data processing and image analysis. Level 1A MODIS imagery of Lake Erie and Lake St. Clair acquired within a ±3 day window centered on field campaign dates were downloaded and qualitatively assessed for cloud cover. While conditions can change rapidly in Lake Erie, previous work has shown that images acquired within ±1 day of field sampling campaigns best capture the in-water signal, but that a time lag of ±3 day between samples and imagery is acceptable (Bailey and Werdell, 2006; Moses et al., 2009). For comparisons with field data, valid pixels were extracted from each image that was closest in time to the sample collection within this temporal window. Four reasonably cloud free (determined by visual inspection, focusing on the Western Basin) images from the Aqua sensor were available for 21, 23, 27 & 28 July 2015. These images are within ±1 day of field sampling dates. We use only the images acquired by MODIS Aqua for this study as the images acquired by Terra on 21 & 23 July 2015 exhibit significant striping due to a known mirror-side banding issue (https://mcst.gsfc.nasa.gov/calibration/mirror-side-striping). Removal of the mirror-side banding as an image pre-processing step is beyond the scope of this paper.

The standard SeaDAS processing stream applies a series of corrections to the L1 radiance values of bands 1, 3, 4, 8 – 14 (visible light range) to L2 surface reflectance values, with an approximate range of values from -0.015 to 0.115 sr^{-1} (https://oceancolor.gsfc.nasa.gov/docs/format/Ocean_Level-2_Data_Products.pdf); this process also provides geophysical parameter flags. The flags provide information on pixel quality, and create land and cloud masks for each image (help manual: https://seadas.gsfc.nasa.gov/help/). The atmospheric correction step, accounts for the effect of the atmosphere on the signal received by the sensor. MODIS, as with all satellite based sensors, measure the total radiance at the top of the atmosphere (L_t^{TOA}). The total measured radiance is a combination of the water-leaving radiance of interest with reflected radiance from the surface of the water, atmospheric scattering due to aerosols and atmospheric gasses such as ozone, (Mobley et al, 2016). The contributions to the TOA radiance measurements by factors other than water-leaving radiance must be estimated and subtracted from the measurement (Franz et al., 2007; Gordon, 1997; Wang and Bailey, 2001). The various atmospheric contributions are influenced by atmospheric conditions such as the amount and type of aerosols, relative humidity, and particulate matter present, as well as by the angle of the sun and the viewing angle of the instrument. The atmospheric correction process is a step-wise algorithm that assesses the conditions in which each image was acquired.

The calculations for these steps that the OBPG (SeaDAS) implements as standard processing protocol are described in Mobley et al (2016). These steps include correction for gas absorption, correction for polarization, a removal of foam reflectance, correction for Rayleigh scattering, sun glint removal, and aerosol correction. The corrected TOA radiance measurement is then transformed to a normalized water-leaving radiance. An assessment is then made to determine if the near-infrared (NIR) values have changed significantly from the measured TOA radiance. If this change is large, the algorithm iterates through the sun glint removal, aerosol removal, and normalization steps until the change in NIR radiance is small. Then a correction is applied to remove any measurement of radiance outside each individual bandwidth, called an ‘out-of-band’ correction (Mobley et al, 2016).

The last step in the OBPG atmospheric correction is the bidirectional reflectance distribution function (BRDF) effect correction (Mobley et al, 2016). The BRDF effect is the effect of sky radiance distribution, viewing geometry, and water optical properties on upwelling radiance distribution. The key aspect of this correction is the water optical properties. These properties are largely dependent on the amount of chl a in the water, and the standard BRDF correction is based on Case 1 water body models with a known chl a concentration (Mobley et al, 2016). Case 1 water bodies, typically oceans, are those in which chl a concentration is the dominant factor affecting in-water scattering (Morel and Prieur, 1977). This step, along with the use of the NIR bands as an assessment of sun glint and aerosol removal, potentially introduces error into the OBPG level 2 products for optically complex, turbid coastal environments such as Lake Erie because the scattering is due to multiple factors. The reflectivity of turbid, high chl a waters is non-negligible in the NIR, making those bands less effective for removing the effects of the atmosphere (Shi and Wang, 2009; Siegel et al., 2000). The challenge of effective atmospheric correction for turbid, coastal waters is an ongoing focus of research (e.g. Hu et al., 2000; Ruddick et al., 2000), although recently the development of algorithms using the SWIR bands have been successful (Wang and Gordon, 2018; Wang and Shi, 2007).

The SeaDAS level 2 processing tool allows for different options in the processing stream. One of the options is a cloud mask threshold that is HAB specific for coastal, optically complex waters, and is suitable for use in Lake Erie (Urquhart et al., 2017; Sean Bailey, NASA OBPG, personal communication, 2017)
https://oceancolor.gsfc.nasa.gov/forum/oceancolor/topic_show.pl?pid=28087). This same cloud mask threshold is employed with the NOAA CI (Wynne and Stumpf, 2015; Sean Bailey, NASA

OBPG, personal communication, 2017) https://oceancolor.gsfc.nasa.gov/forum/oceancolor/topic_show.pl?pid=28087). The shallow nearshore, sediment laden, turbid water of the WBLE causes the standard Ocean Color cloud mask algorithm to flag pixels with high reflectance values, typically associated with sediment or HABs, as clouds. Use of a standard OceanColor mask would have resulted in an underestimation of valid data pixels in the image (Banks and Mélin, 2015; Wang and Shi, 2006). Using the HAB specific cloud mask SeaDAS processor reduces the number of valid pixels that would have been inaccurately flagged as clouds. A first-look test was run on the 28 July 2015 MODIS image, using the standard processing protocol, the standard processing protocol with the HAB-specific cloud mask, and with the atmospheric option using SWIR instead of NIR bands, and the HAB-specific cloud mask. The results indicate that the HAB-specific cloud mask improves the retrieval of water pixels in the image, but that there is no difference between the standard atmospheric correction and the SWIR-based option. Therefore, the standard OBPG parameters were used for the atmospheric correction routine (Mobley et al, 2016), including the standard NIR band value check, with the HAB-specific cloud mask option enabled. This also allows a more direct comparison with the NOAA CI product which used the same cloud masking option. After processing to Level 2 products, the images were geo-rectified and cropped to isolate Lake Erie, using SeaDAS tools.

VPCA Spectral Decomposition

Varimax-rotated, principal component analysis (VPCA) is an eigenvalue-eigenvector based statistical matrix rotation procedure that maximizes the variance within a dataset along orthogonal axes, and reduces dimensionality of a multivariate dataset. The resultant eigenvectors describe the direction of the new axes, and the eigenvalues describe the length of the axes, or how much signal variance is associated with that eigenvector. The eigenvector-eigenvalue pairs are the VPCA components. The longest eigenvectors, corresponding to the largest eigenvalues, are retained. Noise is partitioned into the trailing components, referred to as the noise floor, and discarded. When we apply this method to remote sensing imagery, the input variables are the wavelength bands, while the input samples are the image pixels. The components are orthogonal, and therefore independent of each other, which addresses any correlation between sensor wavelength bands (Davis, 1986). The varimax rotation maximizes the differences between the small and large component loadings, which vary as a function of wavelength, simplifying the spectral shape functions, while maintaining orthogonality, which aids in the interpretation of the spectral shapes (Kaiser, 1958). The resultant VPCA component loadings describe the spectral signature of the in-water color producing agents represented by that component while the component scores describe the spatial distribution of each component. The VPCA method was applied to each image separately. The spectral signatures from each separate image were grouped into patterns based on their extracted spectral shapes (component loadings). The spectral signatures from each of the images in a pattern group were then averaged to produce an average spectral signature pattern. The identification of individual image spectral signatures and the average pattern signature were very similar although notable differences in identification are discussed below.

Each VPCA component loading (spectral signature) and average of spectral signature, is identified using a library of known reflectance derivative spectra of water quality constituents (Table 1), which includes 44 algal & cyanobacterial pigments, accessory pigments and chlorophyll degradation products, compiled from the literature (Ortiz et al., this issue, 2013 and references therein), and using mineral spectra from the United States Geological Survey (USGS) Spectral Library (Clark, 1995; Kokaly et al., 2017) or measured in the lab (Ortiz et al., this issue, 2013). This identification process – employed with both the MODIS and the filtered GF/F sample spectra uses a stepwise, forward multiple linear regression of the component loading against known standardized spectra, a form of principal component regression. The Variance Inflation Factor (VIF) is used to ensure that the regressions do not exhibit multicollinearity (Ortiz, et al, and references therein, this issue). This forward, stepwise, multiple linear regression creates a null model and adds and removes variables until the match is optimized (Ortiz et al., this issue, 2017, 2013). In this usage, each variable is a standard spectra from the library. This can result in a VPCA spectral signature identified as a mixture of pigments, minerals or degradation products; however, this is reasonable within this complex environment and is determined by the correlation structure of the data set. The spectral library is hyperspectral at 10 nm resolution from 400-700 nm. For this study, we resampled the library to MODIS resolution, based on the bandwidth of each MODIS band, before the identification analysis was performed.

Relevant examples of the differences between the original, hyperspectral library spectra and the multispectral MODIS resolution spectra are shown in figure 2 for three categories of in-water constituents: the dinoflagellate accessory pigment peridinin and the cyanobacterial accessory pigment myxoxanthophyll (figure 2A); geothite, hematite & smectite minerals (figure 2B); and chl *a* and its degradation products (figure 2C). The standard spectra for peridinin and for myxoxanthophyll are shown in figure 2A. When these spectra are resampled to MODIS band resolution, the distinguishing peaks at 500 nm, 510 nm, and 520 nm are muted, and the spectra for the pigments become similar, highlighting one of the limitations of multispectral instruments relative to hyperspectral ones. In particular, this loss of reflectance peak amplitude and structure is particularly accentuated in data from coarse resolution, multispectral instruments, where gaps exist in the visible spectrum due to band placement decisions implemented during instrument design.

The standard spectra for minerals can be distinguished because the derivative spectra exhibit different peaks (Figure 2B). In particular, the smectite derivative spectrum is higher in the blue range (400 nm – 450 nm) and then decreases towards the red wavelengths (600 nm) before leveling off from 600 nm – 650 nm. In contrast, the derivative spectra of goethite and hematite are low in the blue end, and then increase to a peak at 550 nm – 570 nm for goethite and at 570 nm – 590 nm for hematite. Goethite and hematite are similar at MODIS resolution, but goethite has a lower trough at 670 nm than hematite (figure 2B) and a secondary peak at 440 nm, which is absent from the hematite spectrum.

The derivative spectra of chl *a* and the three chl *a* degradation products are shown in Figure 2C. These derivative spectra are similar, with a level response between 580 nm & 670 nm, increasing at 680 nm, although the depth and width of the trough and peak between 650 nm and 700 nm differ for the three constituents. Phaeophytin *a* and phaeophorbide *a* have a higher response in the blue wavelengths (410 nm), while chlorophyllide *a* has a lower response at 410 nm. The peaks in the chl *a* derivative spectra reside at 450 nm & 680 nm in the 10 nm resolution spectra, offset from the peaks in the phaeophorbide *a* and phaeophytin *a* degradation products, but are similar in placement to peaks in the chlorophyllide *a* derivative spectra. At 10 nm resolution, the relative amplitude of the peaks on the blue and red ends of the spectra differ as do their relative placement, helping to differentiate these compounds from each other when conducting a whole waveform decomposition. However, at multispectral MODIS resolution, the chl *a* spectrum is similar to the chlorophyllide *a* spectrum, except for a higher response at 665 nm, yielding some ambiguity in differentiating these two constituents. The similarity of these derivative spectra indicate some ambiguity in differentiating chl a from chlorophyllide *a* and in differentiating phaeophytin *a* and phaeophorbide *a* from each other. Notice that algorithms based on bands centered only on the red edge would have even more difficulty differentiating between chl a, chlorophyllide *a*, phaeophytin *a* and phaeophorbide *a*.

Field Data VPCA

As with the remote sensing observations, the visible spectra measured from the filtered water samples represent mixed signals that depend on the weighted average of the collected particulate matter, and must also be spectrally unmixed. The database of measured GF/F sample reflectance derivative spectra, band averaged to match MODIS bandwidths, was analyzed with varimax-rotated, principal component analysis (VPCA) using SPSS statistics software by IBM. The field sample data includes 93 samples. Each row in the data set represents a daily sample composed of a center-weighted, visible reflectance derivative spectrum with the variables (columns) defined at the center-weighted wavelength of the MODIS bands, derived from the hyperspectral observations.

MODIS Imagery VPCA

The reflectance spectra of all pixels in each of the MODIS images is scaled by multiplying all Rrs values by a factor of 10^5, to avoid very small derivative values for plotting, then analyzed using the ENVI function Forward PCA Rotation. Forward PCA Rotation calculates eigenvectors and eigenvalues for the dataset of derivative spectra, and basic statistics, i.e. the standard deviation, minimum, maximum and mean of the dataset. The varimax rotation of the principal components is carried out in IDL, using in-house and open source algorithms.

Pixel Extraction in-water data matching

The location of each 2015 field sampling site (Table 2) used by CIGLR in the WBLE was identified in the MODIS Aqua image VPCA component score maps from 21, 23, 27, & 28 July 2015, and the pixel value of each of the component scores closest to the field sampling location was extracted. Eight CIGLR station locations were sampled on 27 July 2015, and each MODIS image pixel at the sampling locations had valid data, providing eight data pairs for comparison. The VPCA score values were regressed against the field data provided by CIGLR. The critical Pearson's R-value for 6 degrees of freedom (df = 6) at the p=0.05 confidence level is R-crit = 0.707. We also extracted the 21 July 2015 MODIS image VPCA score at each location where spectra from filtered KSU GF/F sample measurements were most closely collected in time and space (Table 3). There were eleven sampling locations visited on 20 July 2015, and the closest matching, 21 July 2015, MODIS image had valid data in pixels containing seven of those sampling locations. We regress the GF/F sample VPCA scores against the VPCA scores of MODIS image pixels at the KSU sampling locations. The critical Pearson’s R-value for this dataset is R-crit = 0.754 ($df = 5$, p=0.05).

Comparison to NOAA CI

We constructed an image for comparison to the 28 July 2015 NOAA CI image by adding each of the 28 July 2015 MODIS Aqua VPCA spectral decomposition components, weighted by the percent variance explained by each component. This resulted in a single band image that can be visually compared to the published NOAA CI image. The components were added sequentially until the visual match was optimized. In this case, on 28 July 2015, all extracted VPCA components that included a red edge response were included in the constructed CI.

Results and discussion

The VPCA spectral decomposition analysis of all datasets resulted in four principal components that fell into distinct loading patterns representing the spectral signature of each component, with associated component spatial distributions. These components are the result of the VPCA, and often represent a mixture of individual in-water constituents.

VPCA Spectral Signature Identification

The forward, stepwise multiple regression of spectral signatures against the standard spectral library resulted in mixtures of pigments as the best fit for each component. Figure 3 shows the spectral signatures of each component extracted from the GF/F dataset at 10 nm as well as at MODIS resolution (Figure 3A – 3D) along with the components from the four MODIS Aqua images acquired on 21, 23, 27 & 28 July 2015 (Figure 3E - 3H). The spectral signature of each MODIS Aqua image is averaged by pattern (spectral shape) and shown with the individual image components. The average pattern is shown with the identification spectra ‘fit’, which is the weighted combination of the matched standard spectra from the forward stepwise multiple regression (Figure 3I – 3L). VPCA decomposition of the in-lab measured reflectance spectra (GF/F dataset) of the water samples collected by the KSU field team yielded four primary components similar to the components extracted from the MODIS images. The GF/F spectral signature averaged to MODIS band resolution (GF/F_{MODIS}; orange) exhibit some differences in spectral features, but when we regress the GF/F_{MODIS} spectral signature against the average MODIS image spectral signature for each pattern, the R-values for patterns A, B, & C are statistically correlated at $p < 0.05$ (Table 4). Therefore, the GF/F sample component identification for GF/F sample patterns A, B & C are identical to the MODIS pattern A, B, & C identification, within error. The GF/F sample spectral pattern D did not correlate significantly with MODIS spectral pattern D. These spectral patterns were identified independently as is discussed further below.

The identification of each VPCA component is often, though not always, a mixture of pigments, minerals and degradation products representing various in-water constituents. The correlations may be negative or positive, indicting a positive or negative correlation with that constituent. The association of each pixel reflectance spectra to the component spectral signature is described by the spatial distribution of each component. In these figures, the red/warm colored pixels increase with positively correlated constituents, and (if applicable) the blue/cool colored pixels increase with negatively correlated constituents.

The spectral signature results of the VPCA spectral decomposition of the 10 nm resolution hyperspectral GF/F sample dataset (blue) is plotted for comparison purposes (figures 3A – 3D). The hyperspectral nature of the 10 nm GF/F dataset can be used to provide additional information to validate the identifications determined from the multispectral data. Increased spectral information provides greater insight for determining the CPAs captured by this field data. However, some distinguishing spectral peaks, such as the 620-630 nm absorption of the cyanobacterial pigment phycocyanin (Bryant, 1981; Sathyendranath et al., 1987), are lost when the GF/F 10 nm dataset is resampled to multi-spectral MODIS bands, in particular between 550-670 nm, were MODIS does not have a high SNR band.

The MODIS spectral signature pattern A (Figure 3E) represents a negative correlation with hematite and a positive correlation with peridinin and/or myxoxanthophyll. Peridinin is a dinoflagellate pigment (Song et al., 1976), while myxoxanthophyll is a cyanobacterial pigment (Paerl et al., 1983). The ANOVA results for the forward stepwise regression of the average pattern A spectral signature are shown in Table 5. The average image identified as a combination of hematite and peridinin, however, the pattern A spectral signature for the first day in the series, the 21 July 2015 MODIS Aqua image, is identified as a combination of hematite and myxoxanthophyll (Table 5). This component likely represents a combination of dinoflagellates, cyanobacteria, and iron-oxide rich, suspended sediment comprising the CyanoHAB signal. It is worth noting that dinoflagellates and cyanobacteria, such as *Planktothrix*, are known to be reside in the western basin (Watson et al., 2016), and to be adapted to living in turbid waters (Oberhaus et al., 2007; Scheffer et al., 1997).

The MODIS image pattern B spectral signature (Figure 3F) represents a mixture of goethite, smectite, hematite, and phaeophorbide *a* (the end member chl a degradation product). The ANOVA results (Table 6) for most of the images and the average pattern spectral signature indicate that a sediment mixture of goethite & smectite are present, while the results for the earliest day in the series, 21 July 2015, indicate the presence of hematite and phaeophorbide *a*. This component, therefore is a sediment and phaeophorbide *a* laden constituent, with a composition that may vary with time, and suggests the presence of chl a that has largely degraded.

Pattern C (Figure 3G) represents a positive correlation with hematite and a negative correlation with the chl a degradation product, chlorophyllide *a* (Table 7). This component is thus comprised of different sediment and chl a degradation products than pattern B. The potential ambiguity in differentiating chl a from chlorophyllide *a* (Figure 2C) suggests that this component may indicate the presence of chl a at varying stages of degradation. Furthermore, the identification of the pattern C spectral signature was consistent across all four images and the average spectral signature, suggesting that the optical signal from hematite and chl a degradation products was not temporally varying throughout the study.

While MODIS pattern D and GF/F_{MODIS} pattern D are similar in their spectral shape, the two are not statistically correlated. These components represent a small fraction of the variance in the MODIS image (1.8%), although a larger percent of variance in the GF/F samples (14.6%). In addition to the partitioning of variance of these two datasets, the physical filtering process, which removes all material smaller than 0.7 μm from the GF/F samples, may explain the difference in variance between these last components and their lack of statistical correlation. The small fraction of variance places these components close to the noise floor, and the sparse multispectral data likely makes their identification more difficult relative to hyperspectral data. MODIS pattern D (Figure 3H) represents a positive correlation with the Cryptophyta algae, a member of the kingdom Chromista (Algaebase.org; Table 8). The GF/F_{MODIS} pattern D (Figure 3D) represents a positive correlation with both hematite and phaeophytin *a*, a degradation product of chl a (Table 8). Given the uncertainty in the data this pattern could represent either of these two potential identifications. In the section on *in situ validation* we compare the spatial patterns of these components to see if they are consistent or unique in spatial distribution. Given the VPCA orthogonality constraint, a common spatial pattern would suggest these two components represent the same in-water constituents while a unique one would indicate that they are likely distinct components.

The identification results from patterns A – D suggest that the first image in the series, acquired on 21 July 2015, may represent slightly different dominant compositional make-up than the latter three days, which span the time period from 23 – 28 July, 2015. This is consistent with observed riverine discharge from the Maumee, Sandusky & Portage Rivers to the WBLE, which were decreasing during this week - long time period (https://waterdata.usgs.gov/nwis). Some differences between the results from the GF/F_{MODIS} and MODIS image VPCA spectral decompositions are to be expected due to several issues: the water samples were collected exclusively in Sandusky Bay and the nearshore waters just east of Sandusky Bay in Lake Erie while the MODIS image analysis includes the reflectance signal from the entire Lake. However, analysis of a subsampled image that matched the geographic distribution of the samples produced similar results to the full image analysis presented here. The field collection dates are also not identical to the MODIS image acquisition dates, but differ by +1 day. Furthermore, the differences in scale between the MODIS image pixel samples and the GF/F samples are significant in that the GF/F samples represent a 1 L, filtered point sample, with only particulate matter ≥ 0.7 μm captured, and all water removed by the physical filtering and drying process, while the MODIS pixels represent an integrated signal from 1 km^2 of Lake Erie, acquired at a height of 705 km, with an atmospheric correction applied to produce a surface reflectance value. Due to physical filtration, the GF/F samples do not include CDOM as a constituent because the dissolved CDOM is lost during the filtering process. CDOM is an important constituent in the Great Lakes (Becker et al., 2009; Binding et al., 2012, 2008; Moore et al., 2017), and is thus incorporated in the MODIS reflectance spectra. However, despite all these spatial and temporal differences, the spectral signatures of the MODIS image and the GF/F field dataset are remarkably similar and statistically significant, indicating a consistent CPA presence. The similarities between the two VPCA spectral decomposition results based on independent observations from different instruments are remarkable, because the VPCA method removes extraneous stochastic noise, and partitions unrelated signals into different components based on their correlation structure. The method is relatively insensitive to atmospheric errors and addresses the mixed pixel problem (Ortiz et al., this issue, 2017, 2013).

In situ validation

The component scores are the projection of the derivative-transformed data onto the principal component loading axis, and as such provide a measure of the amount of signal explained by each component loading. The component score indicates the relationship between each pixel and the component spectral signature, creating a spatial distribution for each component. A value of zero reflects a mean contribution to the derivative reflectance spectra from that component, while positive or negative values indicate standard deviations greater than or less than the mean of the original derivative reflectance spectra for each component spectral signature. We used a simple linear regression to quantify the relationship between the 21 July 2015 MODIS image VPCA score values and the 20 July 2015 GF/F VPCA score values. These two near-concurrent dates provided the 7 clear pixels in the MODIS image corresponding to the KSU sampling locations. There were no cloud-free pixels available for the 27 July 2015 image at the KSU sample site locations, although the image is sufficiently clear at other locations to process. We regressed the 21 July 2015 MODIS score values against the 20 July 2015 GF/F score values for each spectral signature. For these four regressions, R-crit = 0.754 (*df* =5, p=0.05). Despite the slight temporal offset, the R-value for pattern A is 0.73; for pattern B the R-value is -0.80; for pattern C the R-value is 0.88; and for pattern D the R-value is 0.77 (Figure 4). These results indicate very good statistically significant agreement between the *in situ* measurements and remote sensing image analysis for patterns B, C, & D, and provide confidence that the remote sensing analysis partitions the in-water reflectance signal as effectively as the analysis of spectra measured in the lab from *in situ* water samples. As noted above, there is some ambiguity in the identification of pattern D, although the correlation of the spatial patterns for the GF/F_{MODIS} and MODIS VPCA scores suggest they represent the same mixture of constituents. Carefully planned field sampling campaigns, designed to coordinate with satellite overpasses can increase the number of location matchups used for validation. Newer satellites, with a smaller pixel resolution, such as the European Space Agency's (ESA) Sentinel-3A Ocean and Land Color Instrument (OLCI) (300 m ground resolution) also have the potential to increase sample size, particularly if field sites are located close together.

Four of the lab-measured parameters collected by CIGLR on 27 July 2015 for Secchi depth, beam attenuation, in-water chl *a*, and in-water phycocyanin were regressed against the VPCA spectral decomposition results of the MODIS Aqua image acquired on 27 July 2015 and found to yield significant results. We extracted the component score values at each CIGLR sampling location from each component pattern map. The spectral signature results from this individual image analysis matched the patterns described previously. The regression of the eight VPCA Pattern C score values from 27 July 2015 against the CIGLR measured chl *a* values (Figure 5B) yields an R-value of 0.88 (*df*=6, $p<0.05$), indicating statistically significant agreement between the in-water data and remote sensing analysis results, particularly when the data were collected on the same day and has closer temporal coincidence than the KSU GF/F sample comparisons presented earlier. However, some of the slightly weaker fit observed to the lab measurements could also result from the filtration process. Figure 5A maps a close-up of the sampling locations on the VPCA spatial distribution of pattern C in the WBLE, which represents a mixture of hematite and chl *a* degradation products calculated based on data from the entire lake. These results show that the two independent sets of regional validation data, the 20 July 2015 GF/F samples from Sandusky Bay (Figure 4) and the 27 July 2015 CIGLR samples (Figure 5) from the WBLE correlate well with the closest temporal MODIS image match, despite the fact that the MODIS image VPCA spectral decomposition was conducted on the entire image, which includes both Lake Erie and Lake St. Clair. These results document the effectiveness of the VPCA spectral decomposition method to extract signal, such as chl *a*, despite the range of optical conditions across the lake.

The red colors in Figure 5b indicate that the pixels have a higher proportion of sediment than chlorophyllide *a*, while the blue colors in Figure 5b indicate pixels where the mixture is higher in chlorophyllide *a*. In all pixels, chlorophyllide *a* contributes to the signature of the VPCA component, suggesting that chl a is present, or present in some stage of degradation, given the potential ambiguity the spectral signature identification as chl a or chlorophyllide *a* (Figure 2C). The distribution of sampling sites spans a range of scores nearly equal to the full observed range, capturing the variability of this component, although intermediate values are somewhat under-sampled given the static sampling points and the geographic extent of the component scores on this date. Despite the limitation of sample size, these results indicate that VPCA spectral decomposition results from remote sensing data are significantly correlated with the in water measurements.

GF/F component spatial distribution

Our filtered field sample measurements (GF/F sample spectra) indicate that the spatial distribution of each component changes from day to day, and provides quantitative insights into the temporal evolution of the CPAs that each component represents, as well as changes in the composition of the CyanoHAB over time. The GF/F pattern A spectral signature (Figure 6) represents sediment and the cyanobacterial pigment myxothanthophyll. The hematite signal has a negative correlation, so negative values in the spatial distribution plots (Figure 6) correspond to higher than average concentration of hematite in this component, while the correlation with myxothanthophyll & peridinin pigments are positive. During the month of June the spatial distribution of CPA at all sites is varied, highlighting the rapidly changing conditions within a relatively small range of temporal and spatial samples. This month was also exceptionally rainy, and the influx of water, sediment, and debris could have delayed the initiation of the 2015 CyanoHAB. On 6 July 2015 there was a large amount of sediment in the water, particularly in the western end of the Bay near Muddy Creek, creating highly turbid conditions [9 July 2015 NOAA HAB Bulletin] Additionally, there was an unusual amount of woody debris in the water noted during the weekly sample collection. Beginning on 6 July 2015, the spatial distribution of GF/F component A followed a pattern of higher predicted concentration of *Planktothrix* cyanobacteria in Sandusky Bay and a lower concentration outside the Bay, inferred from the correlation with myxothanthophyll. These results point out the utility of VPCA spectral decomposition of optically complex images early in the bloom season before NOAA CI predictions are initiated. The NOAA HAB Monitoring work typically shifts from prediction to bulletin in mid-July, when streamflow and sediment content in the Maumee River plume decreases, creating conditions in which the *Microcystis* bloom can expand in Western Basin of Lake Erie.

GF/F Pattern B spectral signature (Figure 6) indicates the presence of sediment and the chl *a* degradation product, phaeophorbide *a* when these component scores are positive. This component generally increases from Sandusky Bay into Lake Erie (figure 6) on each day sampled. This trend likely reflects the movement of the early season, precipitation driven, sediment fluxes from the Sandusky River towards Lake Erie. On 27 July 2015 the signal from this component is more evenly distributed across the sampling transect, indicating that the heavy sediment input had been dispersed by this time.

GF/F pattern C (Figure 6) represents the chl *a* degradation product chlorophyllide *a*, and the mineral hematite. Positive scores have a higher sediment to chlorophyllide *a* proportion, while negative scores indicate the opposite – a higher chlorophyllide *a* to sediment proportion. Two distinct trends are visible in the scores associated with this component (figure 6). On 8 and 15 June 2015 as well as 6 July 2015 the scores generally increase from Sandusky Bay to Lake Erie. On 22 and 28 June 2015 as well as 13 & 20 July 2015 the signal from this component is much higher at the furthest west sites: SND 4 & SND 6, drops in the outer Bay, and increases along the transect into Lake Erie. This result indicates that chlorophyllide *a* is present in Sandusky Bay and coastal Lake Erie throughout the sampling season although suspended sediment sometimes becomes the dominant optical signal.

The spectral signature for pattern D (Figure 6) represents a combination of hematite and the chl a degradation product, phaeophytin *a*, when the component scores are positive. This signal fluctuates over the course of the sampling season, with a neutral contribution on 8 and 15 June 2015, with the strongest signal from this component within inner Sandusky Bay. On the other sampling days, the signal from this component has a weaker signal in the samples taken from sites in the outer Bay than from the inner Bay or Lake Erie. While this mixture of CPAs is similar to that of Pattern C, the chl *a* degradation products are the result of different pathways. Chlorophyllide *a* forms due to phytol chain removal through hydrolosis, while phaeophytin *a* arises from Mg^{2+} ion removal through demetalation (Hendry, 1982). These two chl a degradation products are spectrally different, but both represent the decay of chl a, albeit through different processes. For our purposes, these pigments indicate the presence of older chl a in the water column, and therefore the two degradation products can exist in the same water sample, but likely in different proportions. Further investigation of the spatial distribution of these processes would require additional study, but future work could shed light on the conditions favorable for each degradation pathway, providing useful information regarding the spatial distribution of various chl a degradation rates.

MODIS component spatial distribution

Figure 7 shows the spatial distribution of component patterns for the VPCA spectral decomposition of the individual day images acquired on 21, 23, 27, & 28 July 2015. In all images, the continuum from positive – to – negative values for individual pixels is shown as a red – to – purple color scheme. Color bars are included for each component’s spatial distribution.

The red areas in Figure 7A, 7E, 7I, and 7M indicate where the in-water reflectance signal is increasing with pattern A, a mixture of cyanobacteria and dinoflagellates. This is within the Maumee Plume, in the southern part of Maumee Bay, and eastward to the Lake Erie Islands, into Sandusky Bay, which is the area where, according to the Experimental Lake Erie Harmful Algal Bloom Bulletin released on 15 July 2015, the 2015 bloom initiated around 11 July 2015 (https://www.glerl.noaa.gov), and is present on the four days included in this analysis. The CPAs described by pattern A are also present in Sandusky Bay, and could represent *Planktothrix* and dinoflagellates in the turbid water present in the persistent Sandusky Bay bloom. This component captures a range of variability in the individual images from 38% to 47% of the image variance. The daily images indicate the temporal change of this component over the course of a week. On 21 July 2015, this component is located in a large area toward the western end of the Maumee Plume, and a long streamer extending from the shore to the Lake Erie Islands. The two areas begin to combine on 23 July 2015 and by 28 July 2015 this component has expanded and increased northward and eastward.

The pattern B spectral signature representing clay sediments and the chl *a* degradation phaeophorbide *a*, found in the discharge plumes of the Detroit River, the Portage River, and streams entering Lake Erie from the Ottawa National Wildlife Refuge, in Lake St. Clair, and along the northern shore near Long Point, ON (Figures 7B, 7F, 7J, & 7N). This component explains 25% to 33% in the individual image variance on a day by day basis. This component is largely absent from the Maumee River Plume in all images, but is a clear indicator of the input from Lake St. Clair and the Detroit River on all days. On 28 July 2015 this component highlights all the input river plumes, and is present even in the Maumee River plume (Figure 7N).

Component pattern C is a sediment and chlorophyllide *a* constituent, and is located in the Maumee River Plume, along the western shore of Maumee Bay, as well as in Sandusky Bay. This could represent the signal from the degradation of the very early CyanoHAB bloom (15 July 2015 HAB Bulletin) that began on 11 July 2015 in the western-most part of Maumee Bay, but which drifted offshore to the east by 21 July 2015 (HAB Bulletin). This component explains 21% to 28% of the variability in the individual image analyses. The red areas where suspended sediment is more prominent than degradation products in the overall mixture, are located further west in the mouth of Maumee Bay on 21 & 23 July 2015 (Figures 7C & 7G) and are distinct from the area of cyanobacteria and dinoflagellates indicated by pattern A (Figures 7A & 7E). These two signals (pattern A & pattern C) begin to overlap on 27 July 2015 (Figures 7I & 7K), and overlap even more in area on 28 July 2015 (Figures 7M & 7O). While not all this area is addressed by the NOAA CI (29 July 2015 HAB Bulletin), the area in the Maumee plume along the southern shore of Maumee Bay where the CPAs associated with pattern C are located is identified as cyanobacteria. Our results suggest that this area is older CyanoHAB, where the chl a signal has begun to degrade as the CyanoHAB shifts to the east. The blue areas in all the images, which indicate an in-water signal where chlorophyllide *a* is prominent over the hematite signal in the overall mixture, are similar to the areas indicated by pattern B (Figures 7B, 7F, 7J, & 7N).

The pattern D spectral signature is found along the southern coast of Lake St. Clair and Lake Erie. This cryptophyta component captures about 1.6% to 4% of the variance in the individual images (Figures 7D, 7H, 7L, and 7P). The spatial distribution of this component is complementary to the distribution of pattern A, which is consistent with observations because pattern D represents in-water constituents typically associated with algal blooms in Lake Erie. Notably, on 28 July 2015, Pattern D is present as a border for the area of the bloom identified as Pattern A (Figure 7P). Additionally, the filamentous swirl of pattern D extending up towards Point Pelee on both 27 & 28 July 2015 (Figures 7L and 7P) is quite visible in the NOAA CI image published in the 29 July 2015 HAB Bulletin (Figure 8).

We can compare our results to the published NOAA CI (https://www.glerl.noaa.gov//res/HABs_and_Hypoxia/) as a weighted average of the components that include a signature within the red edge. This includes all four extracted components presented in this study (Figure 3E – 3H). Qualitative comparison of the NOAA CI calculated from the same 28 July 2015 MODIS image (29 July HAB Bulletin) with the constructed CI, based on the four orthogonal spectral signatures, indicates that the KSU VPCA spectral decomposition extracts a more detailed distribution of CPAs than the NOAA CI. The spatial distribution of pattern A (Figure 3I) and the CI index identify the Maumee Plume as cyanobacteria, likely *Microcystis*, in the WBLE (Figure 8). However, when all 4 components are included in the constructed CI (Figure 8), the heterogeneity of the signal shown in the NOAA CI in the WBLE is more closely matched. This result suggests that the KSU spectral decomposition method identifies CyanoHAB related constituents with greater specificity than the NOAA CI, and warrants further investigation.

Conclusion:

VPCA Spectral decomposition of derivative reflectance spectra acquired from *in situ* samples and by satellite sensors provide a powerful tool for determining the composition and distribution of CPAs in Lake Erie, and other optically complex water bodies. Hyperspectral ground data can be used to help identify the signals extracted from multispectral satellite data sets as was successfully demonstrated in this project. We identified four distinct patterns of CPAs and noted the temporal change in composition and distribution of each identified CPA. The spatial distribution of VPCA patterns, in conjunction with the identification of loadings indicate the distribution of algae, cyanobacteria, sediment, and degradation products of chl a during the early 2015 CyanoHAB season. We also highlighted that the Detroit River plume has a different set of constituents than that of the CyanoHAB-associated signals from Maumee River and Sandusky River to the south.

This analysis was conducted using images acquired after a particularly stormy spring, with high levels of discharge from the Maumee, Detroit, and Sandusky Rivers, as well as frequent high-wind storm events that re-suspend sediment in the shallow WBLE, enhancing the contribution of sediment to the visible reflectance signature. The spectral decomposition method effectively partitioned the sediment signal in all images we analyzed. These images capture the beginning of the 2015 CyanoHAB, including a green algae signature, and our analysis supports this assessment. The correlation of all spectral signature patterns with lab-measured water samples from 21 July 2015, and the strong correlation of pattern C with *in situ* chl a measurements on 27 July 2015 supports the conclusion that this method effectively partitions the optical signal from remote sensing imagery.

The KSU spectral decomposition method has the potential to contribute to future CyanoHAB monitoring and assessment efforts, as it can guide field season planning for sampling locations and yields information about the distribution of the mixtures of constituents present and how they change through time. Moreover, enhanced coordination of *in situ* measurements and remote sensing datasets has the potential to improve imagery analysis as well as provide powerful monitoring tools for researchers and water management authorities alike.

Acknowledgements

Field data was provided by the NOAA Great Lakes Environmental Research Laboratory and the Cooperative Institute for Great Lakes Research with partial funding support from the EPA Great Lakes Restoration Initiative. George Bullerjahn and Taylor Tuttle from Bowling Green State University, Kristin Slodysko from University of Washington, along with the BGSU & KSU field sampling teams were instrumental in data collection. We also thank George Bullerjahn for providing a constructive manuscript review prior to submission. We thank the Ohio Department of Natural Resources – Watercraft for access to sampling vessels.

References

Ali, K., Witter, D., Ortiz, J., 2014. Application of empirical and semi-analytical algorithms to MERIS data for estimating chlorophyll a in Case 2 waters of Lake Erie. Environ. Earth Sci. 71, 4209–4220. doi:10.1007/s12665-013-2814-0

Ali, K.A., Ortiz, J.D., 2016. Multivariate approach for chlorophyll-a and suspended matter retrievals in Case II type waters using hyperspectral data. Hydrol. Sci. J. 61, 200–213. doi:10.1080/02626667.2014.964242

Ali, K.A., Witter, D.L., Ortiz, J.D., 2013. Multivariate approach to estimate colour producing agents in Case 2 waters using first-derivative spectrophotometer data. Geocarto Int. 1–26.

Bailey, S.W., Werdell, P.J., 2006. A multi-sensor approach for the on-orbit validation of ocean color satellite data products. Remote Sens. Environ. 102, 12–23. doi:10.1016/J.RSE.2006.01.015

Baker, D.B., Confesor, R., Ewing, D.E., Johnson, L.T., Kramer, J.W., Merryfield, B.J., 2014. Phosphorus loading to Lake Erie from the Maumee, Sandusky and Cuyahoga rivers: The importance of bioavailability. J. Great Lakes Res. 40, 502–517. doi:10.1016/j.jglr.2014.05.001

Balsam, W.L., Deaton, B.C., 1996. Determining the composition of late Quaternary marine sediments from NUV, VIS, and NIR diffuse reflectance spectra. Mar. Geol. 134, 31–55. doi:http://dx.doi.org/10.1016/0025-3227(96)00037-0

Banks, A.C., Mélin, F., 2015. An assessment of cloud masking schemes for satellite ocean colour data of marine optical extremes. Int. J. Remote Sens. 36, 797–821.

Becker, R.H., Sultan, M.I., Boyer, G.L., Twiss, M.R., Konopko, E., 2009. Mapping cyanobacterial blooms in the Great Lakes using MODIS. J. Great Lakes Res. 35, 447–453.

Binding, C.E., Greenberg, T.A., Bukata, R.P., 2012. An analysis of MODIS-derived algal and mineral turbidity in Lake Erie. J. Great Lakes Res. 38, 107–116.

Binding, C.E., Jerome, J.H., Bukata, R.P., Booty, W.G., 2008. Spectral absorption properties of dissolved and particulate matter in Lake Erie. Remote Sens. Environ. 112, 1702–1711. doi:10.1016/J.RSE.2007.08.017

Bryant, D.A., 1981. The Photoregulated Expression of Multiple Phycocyanin Species: A General Mechanism for the Control of Phycocyanin Synthesis is Chromatically Adapting Cyanobacteria. Eur. J. Biochem. 119, 425–429.

Carmichael, W.W., Boyer, G.L., 2016. Health impacts from cyanobacteria harmful algae blooms: Implications for the North American Great Lakes. Harmful Algae 54, 194–212. doi:10.1016/J.HAL.2016.02.002

Clark, R.N., 1995. Reflectance spectra. AGU Ref. Shelf 3, 178–188.

Clark, R.N., Hoefen, T.M., Swayze, G.A., Livo, K.E., Meeker, G.P., Sutley, S.J., Wilson, S.A., Brownfield, I.K., Vance, J.S., 2003. Reflectance spectroscopy as a rapid assessment tool for the detection of amphiboles from the Libby, Montana region, Open-File Report - U.S.Geological Survey. U. S. Geological Survey : Reston, VA, United States, U. S. Geological Survey, Denver, CO, United States.

Davis, J.C., 1986. Statistics and data analysis in geology. 1986.

Davis, T.W., Bullerjahn, G.S., Tuttle, T., McKay, R.M., Watson, S.B., 2015. Effects of Increasing Nitrogen and Phosphorus Concentrations on Phytoplankton Community Growth and Toxicity During Planktothrix Blooms in Sandusky Bay, Lake Erie. Environ. Sci. Technol. 49, 7197–7207. doi:10.1021/acs.est.5b00799

Deaton, B.C., Balsam, W.L., 1991. Visible Spectroscopy--A Rapid Method for Determining Hematite and Goethite Concentration in Geological Materials: RESEARCH METHOD PAPER. J. Sediment. Res. 61.

Franz, B.A., Bailey, S.W., Werdell, P.J., McClain, C.R., 2007. Sensor-independent approach to the vicarious calibration of satellite ocean color radiometry. Appl. Opt. 46, 5068–5082. doi:10.1364/AO.46.005068

Gordon, H.R., 1997. Atmospheric correction of ocean color imagery in the Earth Observing System era. J. Geophys. Res. Atmos. 102, 17081–17106.

Hendry, M.J., 1982. Hydraulic Conductivity of a Glacial Till in Alberta a. Groundwater 20, 162–169.

Ho, J.C., Michalak, A.M., 2017. Phytoplankton blooms in Lake Erie impacted by both long-term and springtime phosphorus loading. J. Great Lakes Res. 43, 221–228. doi:10.1016/J.JGLR.2017.04.001

Hu, C., Carder, K.L., Muller-Karger, F.E., 2000. Atmospheric correction of SeaWiFS imagery over turbid coastal waters: a practical method. Remote Sens. Environ. 74, 195–206.

Kaiser, H.F., 1958. The varimax criterion for analytic rotation in factor analysis. Psychometrika 23, 187–200.

Kane, D.D., Conroy, J.D., Peter Richards, R., Baker, D.B., Culver, D.A., 2014. Re-eutrophication of Lake Erie: Correlations between tributary nutrient loads and phytoplankton biomass. J. Great Lakes Res. 40, 496–501.

Kokaly, R.F., Clark, R.N., Swayze, G.A., Livo, K.E., Hoefen, T.M., Pearson, N.C., Wise, R.A., Benzel, W.M., Lowers, H.A., Driscoll, R.L., others, 2017. USGS spectral library version 7.

Menken, K.D., Brezonik, P.L., Bauer, M.E., 2006. Influence of Chlorophyll and Colored Dissolved Organic Matter (CDOM) on Lake Reflectance Spectra: Implications for Measuring Lake Properties by Remote Sensing. Lake Reserv. Manag. 22, 179–190. doi:10.1080/07438140609353895

Mobley, C.D., Werdell, P.J., Franz, B., Ahmad, Z., Bailey, S., 2016. Atmospheric Correction for Satellite Ocean Color Radiometery: A Tutorial and Documentation of the Algorithms Used by the NASA Ocean Biology Processing Group.

Moore, T.S., Mouw, C.B., Sullivan, J.M., Twardowski, M.S., Burtner, A.M., Ciochetto, A.B., McFarland, M.N., Nayak, A.R., Paladino, D., Stockley, N.D., others, 2017. Bio-optical Properties of Cyanobacteria Blooms in Western Lake Erie. Front. Mar. Sci. 4, 300.

Morel, A., Prieur, L., 1977. Analysis of variations in ocean color 1. Limnol. Oceanogr. 22, 709–722.

Moses, W.J., Gitelson, A.A., Berdnikov, S., Povazhnyy, V., 2009. Estimation of chlorophyll-a concentration in case II waters using MODIS and MERIS data—successes and challenges. Environ. Res. Lett. 4, 45005.

Oberhaus, L., Briand, J.F., Leboulanger, C., Jacquet, S., Humbert, J.F., 2007. Comparative effects of the quality and quantity of light and temperature on the growth of Planktothrix agardhii and P. rubescens1. J. Phycol. 43, 1191–1199. doi:10.1111/j.1529-8817.2007.00414.x

Ortiz, J.D., 2011. Application of Visible/near Infrared derivative spectroscopy to Arctic paleoceanography. IOP Conf. Ser. Earth Environ. Sci. 14, 12011. doi:10.1088/1755-1315/14/1/012011

Ortiz, J.D., Avouris, D., Schiller, S., Luvall, J.C., Lekki, J.D., Tokars, R.P., Anderson, R.C., Shuchman, R., Sayers, M., Becker, R., 2017. Intercomparison of Approaches to the Empirical Line Method for Vicarious Hyperspectral Reflectance Calibration. Front. Mar. Sci. 4, 296. doi:10.3389/fmars.2017.00296

Ortiz, J.D., Witter, D.L., Ali, K.A., Fela, N., Duff, M., Mills, L., 2013. Evaluating multiple colour-producing agents in Case II waters from Lake Erie. Int. J. Remote Sens. 34, 8854–8880. doi:10.1080/01431161.2013.853892

Paerl, H.W., Otten, T.G., 2013. Harmful Cyanobacterial Blooms: Causes, Consequences, and Controls. Microb. Ecol. 65, 995–1010. doi:10.1007/s00248-012-0159-y

Paerl, H.W., Tucker, J., Bland, P.T., 1983. Carotenoid enhancement and its role in maintaining blue-green algal (Microcystis aeruginosa) surface blooms. Limnol. Oceanogr. 28, 847–857.

Romero, A., Gatta, C., Camps-Valls, G., 2016. Unsupervised Deep Feature Extraction for Remote Sensing Image Classification. IEEE Trans. Geosci. Remote Sens. 54, 1349–1362. doi:10.1109/TGRS.2015.2478379

Ruddick, K.G., Ovidio, F., Rijkeboer, M., 2000. Atmospheric correction of SeaWiFS imagery for turbid coastal and inland waters. Appl. Opt. 39, 897–912.

Sathyendranath, S., Lazzara, L., Prieur, L., 1987. Variations in the spectral values of specific absorption of phytoplankton. Limnol. Oceanogr. 32, 403–415.

Scheffer, M., Rinaldi, S., Gragnani, A., Mur, L.R., van Nes, E.H., 1997. On the dominance of filamentous cyanobacteria in shallow, turbid lakes. Ecology 78, 272–282. doi:10.1890/0012-9658(1997)078[0272:OTDOFC]2.0.CO;2

Shi, W., Wang, M., 2009. An assessment of the black ocean pixel assumption for MODIS SWIR bands. Remote Sens. Environ. 113, 1587–1597. doi:10.1016/J.RSE.2009.03.011

Shuchman, R., Korosov, A., Hatt, C., Pozdnyakov, D., Means, J., Meadows, G., 2006. Verification and application of a bio-optical algorithm for Lake Michigan using SeaWiFS: a 7-year inter-annual analysis. J. Great Lakes Res. 32, 258–279.

Shuchman, R.A., Leshkevich, G., Sayers, M.J., Johengen, T.H., Brooks, C.N., Pozdnyakov, D., 2013. An algorithm to retrieve chlorophyll, dissolved organic carbon, and suspended minerals from Great Lakes satellite data. J. Great Lakes Res. 39, 14–33.

Siegel, D.A., Wang, M., Maritorena, S., Robinson, W., 2000. Atmospheric correction of satellite ocean color imagery: the black pixel assumption. Appl. Opt. 39, 3582–3591. doi:10.1364/AO.39.003582

Singh, A., 1989. Review Article Digital change detection techniques using remotely-sensed data. Int. J. Remote Sens. 10, 989–1003. doi:10.1080/01431168908903939

Sly, P.G., 1976. Lake Erie and its basin. J. Fish. Board Canada 33, 355–370.

Song, P.-S., Koka, P., Prezelin, B.B., Haxo, F.T., 1976. Molecular topology of the photosynthetic light-harvesting pigment complex, peridinin-chlorophyll a-protein, from marine dinoflagellates. Biochemistry 15, 4422–4427.

Steffen, M.M., Davis, T.W., McKay, R.M.L., Bullerjahn, G.S., Krausfeldt, L.E., Stough, J.M.A., Neitzey, M.L., Gilbert, N.E., Boyer, G.L., Johengen, T.H., Gossiaux, D.C., Burtner, A.M., Palladino, D., Rowe, M.D., Dick, G.J., Meyer, K.A., Levy, S., Boone, B.E., Stumpf, R.P., Wynne, T.T., Zimba, P. V, Gutierrez, D., Wilhelm, S.W., 2017. Ecophysiological Examination of the Lake Erie Microcystis Bloom in 2014: Linkages between Biology and the Water Supply Shutdown of Toledo, OH. Environ. Sci. Technol. 51, 6745–6755. doi:10.1021/acs.est.7b00856

Stow, C.A., Cha, Y., Johnson, L.T., Confesor, R., Richards, R.P., 2015. Long-Term and Seasonal

Trend Decomposition of Maumee River Nutrient Inputs to Western Lake Erie. Environ. Sci. Technol. 49, 3392–3400. doi:10.1021/es5062648

Stumpf, R.P., Johnson, L.T., Wynne, T.T., Baker, D.B., 2016. Forecasting annual cyanobacterial bloom biomass to inform management decisions in Lake Erie. J. Great Lakes Res. 42, 1174–1183. doi:10.1016/J.JGLR.2016.08.006

Stumpf, R.P., Wynne, T.T., Baker, D.B., Fahnenstiel, G.L., 2012. Interannual variability of cyanobacterial blooms in Lake Erie. PLoS One 7, e42444.

Urquhart, E.A., Schaeffer, B.A., Stumpf, R.P., Loftin, K.A., Werdell, P.J., 2017. A method for examining temporal changes in cyanobacterial harmful algal bloom spatial extent using satellite remote sensing. Harmful Algae 67, 144–152. doi:10.1016/J.HAL.2017.06.001

Wang, M., Bailey, S.W., 2001. Correction of sun glint contamination on the SeaWiFS ocean and atmosphere products. Appl. Opt. 40, 4790–4798. doi:10.1364/AO.40.004790

Wang, M., Gordon, H.R., 2018. Sensor performance requirements for atmospheric correction of satellite ocean color remote sensing. Opt. Express 26, 7390–7403. doi:10.1364/OE.26.007390

Wang, M., Shi, W., 2007. The NIR-SWIR combined atmospheric correction approach for MODIS ocean color data processing. Opt. Express 15, 15722–15733. doi:10.1364/OE.15.015722

Wang, M., Shi, W., 2006. Cloud Masking for Ocean Color Data Processing in the Coastal Regions. IEEE Trans. Geosci. Remote Sens. 44, 3105–3196. doi:10.1109/TGRS.2006.876293

Watson, S.B., Miller, C., Arhonditsis, G., Boyer, G.L., Carmichael, W., Charlton, M.N., Confesor, R., Depew, D.C., Höök, T.O., Ludsin, S.A., Matisoff, G., McElmurry, S.P., Murray, M.W., Peter Richards, R., Rao, Y.R., Steffen, M.M., Wilhelm, S.W., 2016. The re-eutrophication of Lake Erie: Harmful algal blooms and hypoxia. Harmful Algae 56, 44–66. doi:10.1016/J.HAL.2016.04.010

Witter, D.L., Ortiz, J.D., Palm, S., Heath, R.T., Budd, J.W., 2009. Assessing the application of SeaWiFS ocean color algorithms to Lake Erie. J. Great Lakes Res. 35, 361–370.

Wynne, T.T., Stumpf, R.P., 2015. Spatial and temporal patterns in the seasonal distribution of toxic cyanobacteria in western lake erie from 2002–2014. Toxins (Basel). 7, 1649–1663.

Wynne, T.T., Stumpf, R.P., Briggs, T.O., 2013. Comparing MODIS and MERIS spectral shapes for cyanobacterial bloom detection. Int. J. Remote Sens. 34, 6668–6678.

Wynne, T.T., Stumpf, R.P., Tomlinson, M.C., Dyble, J., 2010. Characterizing a cyanobacterial bloom in Western Lake Erie using satellite imagery and meteorological data. Limnol. Oceanogr. 55, 2025–2036. doi:10.4319/lo.2010.55.5.2025

Xia, J., Du, P., He, X., Chanussot, J., 2014. Hyperspectral Remote Sensing Image Classification Based on Rotation Forest. IEEE Geosci. Remote Sens. Lett. 11, 239–243. doi:10.1109/LGRS.2013.2254108

Funding: This work was supported by the National Aeronautics and Space Administration [NNC15VB05P]; Ohio Sea Grant [R/ES-021-PD, R/ER-105-PD]; the Ohio Space Grant Consortium [Doctoral Fellowship; Campus Allotment]; the Ohio Department of Natural Resources [in-kind]; and the Kent State University Department of Geology. No funding agency was involved in planning, execution, or analysis of this research.

Table 1. Spectral Library contents analyzed.

Algal and Cyanobacterial groups (9)	*Primary and accessory pigments (27)*	*Pigment degradation products (6)*	*Minerals and mineral mixtures (41)*
Bacillariophyceae (2 spectra)	α-Carotene	Chlorophyllide-a	Actinolite (2 spectra)
Chlorophyceae (2 spectra)	allophycocyanin	Chlorophyllide-b	Anhydrite (2 spectra)
Cryptophyta	Alloxanthin	Phaeophorbide-a	Calcite (2 spectra)
Cyanobacteria (2 spectra)	Antheraxanthin	Phaeophorbide-b	Calcite+Dolomite
Dinophyta	β-carotene	Phaeophytin-a	Chlorite
Haptophyta	Carotenoids	Phaeophytin-b	Chlorite+Smectite (2 spectra)
	Chlorophyll a + carotenoids		Diatomite
	Chlorophyll a (3 spectra)		Dolomite
	Chlorophyll b (2 spectra)		Dolomite (Ferroan)
	Chlorophyll c		Epidote
	Diadinoxanthin		Glauconite
	Dinoxanthin		Goethite
	Echinenone		gypsum (2 spectra)
	Fucoxanthin		Hematite
	Lutein		Illite (5 spectra)
	Myxoxanthophyll		Kaolinite (2 spectra)
	Neoxanthin		Kaolinite+Hematite
	Nostoxanthin		Muscovite (2 spectra)
	Peridinin		Natrolite
	Phycocyanin (2 spectra)		Opal
	Phycoerythrin		Opal Hyalite
	Trans-neoxanthin		Pyrolusite
	Violaxanthin		Quartz (3 spectra)
			Sanidine Feldspar
			Smectite (2 spectra)
			Smectite+Chlorite+Illite
			Sphalerite
			Tremolite

Table 2. Locations of field sites where CIGLR collected water samples for analysis. Dates indicate the datasets used in this study.

Site Name	Latitude (°N)	Longitude (°W)	Date Sampled (2015)
WE2	41.7648	-83.3314	6/8; 7/20; 7/27
WE4	41.8259	-83.1944	6/8; 7/20; 7/27
WE6	41.7123	-83.3781	6/8; 7/20; 7/27
WE8	41.8336	-83.3637	6/8; 7/20; 7/27
WE12	41.7036	-83.2581	6/8; 7/20; 7/27
WE13	41.7441	-83.1381	7/20; 7/27
WE14	41.7177	-83.0148	7/20; 7/27
WE15	41.6177	-83.0124	7/27

Table 3. Field site locations and dates where KSU/BGSU collected in situ data.

Samples collected in 2015 on: 6/8, 6/15, 6/2, 6/29, 7/6, 7/13, 7/20, and 7/27.

Site Name	**Latitude (°N)**	**Longitude (°W)**
	Sandusky Bay	
SND 4	41.453333	-82.960767
SND 6	41.457300	-82.898655
SND 2	41.479817	-82.782867
SND 1	41.477367	-82.739783
EC 1163	41.469000	-82.715000
	Central Basin Costal Zone	
Bells	41.511667	-82.657967
CBCZ 2	41.442802	-82.630832
CBCZ 3	41.404250	-82.561904
CBCZ 4	41.386335	-82.513187
CBCZ 5	41.434531	-82.373920
CBCZ 6	41.456671	-82.217642

Table 4. Pearson's Correlation for GF/F_{MODIS} VPCA spectral

signature against average MODIS VPCA spectral signature. * Denotes significant correlation with R-crit

= 0.754 (df = 5, p = 0.05) ,

VPCA Spectral Signature Pattern	**R-value**
Pattern A	0.90*
Pattern B	0.90*
Pattern C	0.85*
Pattern D	0.57*

Table 5. ANOVA results for the forward, stepwise multiple linear regression of the pattern A for the average MODIS Aqua loadings and the 21 July 2015 MODIS Aqua image spectral signatures against standard spectra.

Pattern A – average MODIS spectral signature

R	*R-Squared*	*Adjusted R-Squared*	*S*	*F*	*p-value*
0.96	0.92	0.88	0.33	24.00	0.00591

Constituent	*Coeff.*	*Standard Error*	*Beta*	*t*	*p-value > t*	*VIF*
hematite	-0.91	0.14	-0.91	-6.378	0.0031	1.07
peridinin	0.61	0.14	0.61	4.25	0.01306	1.07

Pattern A – 21 July 2015 spectral signature

R	*R-Squared*	*Adjusted R-Squared*	*S*	*F*	*p-value*
0.97	0.94	0.92	0.27	36.72	0.00267

Constituent	*Coeff.*	*Standard Error*	*Beta*	*t*	*p-value > t*	*VIF*
hematite	-0.86	0.11	-0.86	-7.50	0.00169	1.02
myxo-xanthophyll	0.61	0.11	0.61	5.31	0.00602	1.02

Table 6. ANOVA results for the forward, stepwise multiple linear regression of the pattern B for the average MODIS Aqua loadings and the 21 July 2015 MODIS Aqua image spectral signatures against standard spectra.

Pattern B – average MODIS spectral signature						
R	*R-Squared*	*Adjusted R-Squared*	*S*	*F*	*p-value*	
0.97	0.94	0.91	0.28	34.11	0.00307	

Constituent	*Coeff.*	*Standard Error*	*Beta*	*t*	*p-value > t*	*VIF*
smectite	0.72	0.12	0.72	5.80	0.00438	1.13
geothite	-0.44	0.12	-0.44	-3.51	0.0245	1.13

Pattern B – 21 July 2015 spectral signature						
R	*R-Squared*	*Adjusted R-Squared*	*S*	*F*	*p-value*	
0.99	0.98	0.97	0.15	129.72	0.00023	

Constituent	*Coeff.*	*Standard Error*	*Beta*	*t*	*p-value > t*	*VIF*
hematite	-0.79	0.06	-0.79	-12.49	0.00024	1.05
phaeophorbide-a	0.44	0.06	0.44	6.97	0.00222	1.05

Table 7. ANOVA results for the forward, stepwise multiple linear regression of the pattern C for the average MODIS Aqua loading spectral signatures against standard spectra.

Pattern C – average MODIS spectral signature					
R	*R-Squared*	*Adjusted R-Squared*	*S*	*F*	*p-value*
0.96	0.92	0.88	0.33	24.75	0.00559

Constituent	*Coeff.*	*Standard Error*	*Beta*	*t*	*p-value > t*	*VIF*
hematite	0.95	0.13	0.95	6.85	0.00237	1.03
chlorophyllide-a	-0.38	0.13	-0.38	-2.79	0.04896	1.03

Table 8: ANOVA results for the forward, stepwise multiple linear regression of the pattern D for the average MODIS Aqua loading spectral signatures and the GF/F (MODIS resolution) spectral signature against standard spectra.

Pattern D – average MODIS spectral signature

R	*R-Squared*	*Adjusted R-Squared*	*S*	*F*	*p-value*
0.75	0.57	0.48	0.71	6.75	0.04834

Constituent	*Coeff.*	*Standard Error*	*Beta*	*t*	*p-value > t*	*VIF*
cryptophyta	0.75	0.29	0.75	2.59	0.04834	1.

Pattern D – GF/F spectral signature

R	*R-Squared*	*Adjusted R-Squared*	*S*	*F*	*p-value*
0.96	0.93	0.90	0.30	31.17	0.00363

Constituent	*Coeff*	*Standard Error*	*Beta*	*t*	*p-value > t*	*VIF*
hematite	0.88	0.12	0.88	7.02	0.00216	1.05
phaeophytin-a	0.63	0.12	0.63	5.05	0.00719	1.05

Figure Captions

Figure 1. A) Bathymetry map of Lake Erie, delineating the Western, Central, and Eastern Basins. Available from NOAA (Lake Erie and Lake Saint Clair: ngdc.noaa.gov) B) Field sampling locations. The 11 KSU sites in Sandusky Bay and along the southern coast of the Central Basin are shown in inset, as well as the sites sampled by CIGLR. These field sampling sites were sampled weekly through the summer and fall 2015.

Figure 2. A) Standardized derivative reflectance spectra of peridinin and myxoxanthophyll at 10 nm resolution and at MODIS band resolution. B) Standardized derivative reflectance spectra of the minerals goethite, hematite, and smectite at 10 nm resolution and at MODIS band resolution. C) Standardized derivative reflectance spectra of chl *a* and chl *a* degradation products phaeophytin *a*, chlorophyllide *a*, and phaeophorbide *a,* at 10 nm resolution and at MODIS band resolution. Note that distinguishing peaks in the spectra are lost when resolution is re-sampled, but that there are differences between pigments and minerals that allow for identification of specific constituents.

Figure 3. VPCA spectral decomposition results for the GF/F lab-measured reflectance spectra and MODIS Aqua images acquired on 21, 23, 27 and 28 July 2015. Spectral signature patterns extracted by VPCA spectral decomposition of the GF/F dataset (blue) at 10 nm resolution; the GF/F dataset (orange) at MODIS band resolution (A-D). Figures E-H show the MODIS image spectral signature patterns, with the average signature shown in black. Identified CPA represented by each component is listed below.

Figure 4. Regression of GF/F score values on 20 July 2015 against the 21 July 2015 MODIS Aqua image score values at KSU sampling locations. R-crit = 0.754 (df=5, p=0,05). A) spectral signature pattern A; R-value = 0.73. B) spectral signature pattern B; R-value = -0.80. C) spectral signature pattern C; R-value = 0.88. D) spectral signature pattern D; R-value = 0.77.

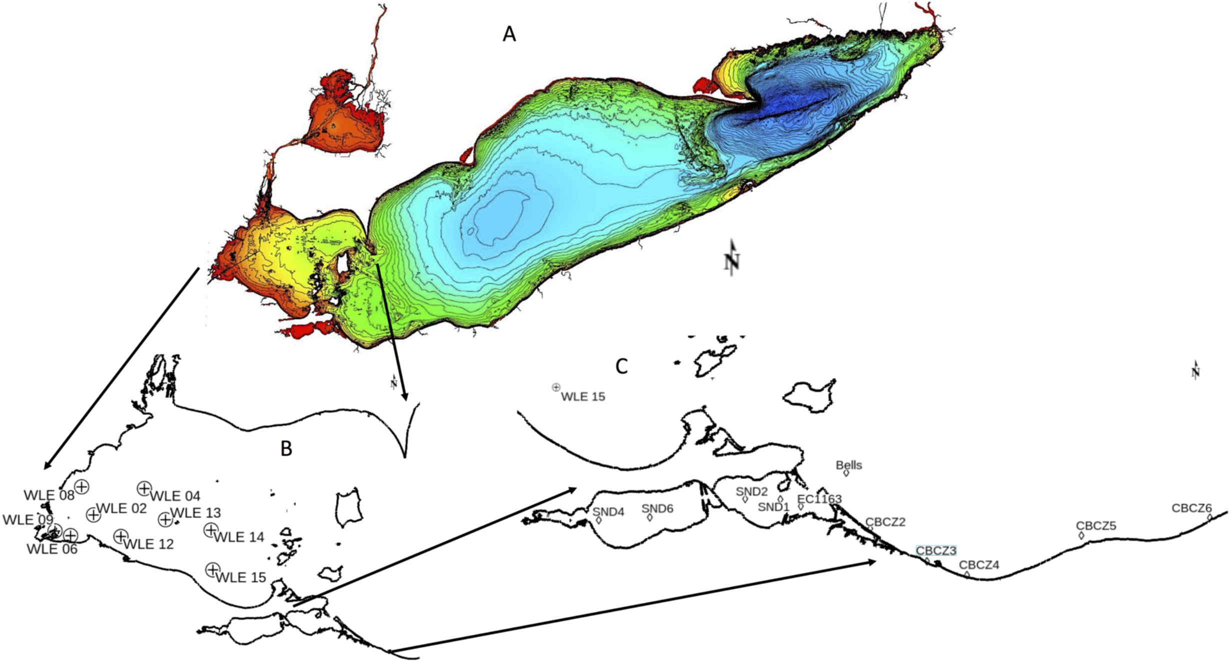
Figure 5. A) 27 July 2015 MODIS image score value for pattern C at the NOAA CIGLR sampling locations plotted against *in situ* chl *a* measured by NOAA CIGLR on 27 July 2015. R-value = 0.88 (df=6, p=0.05). B) Location of NOAA CIGLR sampling locations plotted on the spatial distribution of the 27 July 2015 MODIS image pattern C.

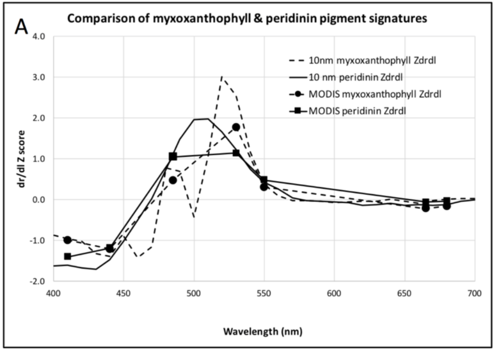
Figure 6. VPCA spectral decomposition results for the GF/F dataset. Points marked with a closed circle are field sampling sites in Sandusky Bay (SND), open circles denote field sampling sites along the southern central basin coast (CBCZ). Pattern A spectral signature and spatial distribution. Pattern A represents sediment, dinoflagellates, and cyanobacteria. Pattern B spectral signature and spatial distribution. Pattern B represents smectite & goethite and chl *a* degradation products. Pattern C spectral signature and spatial distribution. Pattern C represents hematite and chl *a* degradation products. Pattern D spectral signature and spatial distribution. Pattern D represents hematite and phaeophytin a.

Figure 7. Spatial distribution of each spectral signature pattern for the four MODIS Aqua images acquired on 21, 23, 27 and 28 July 2015. Cyanobacteria and dinoflagellates are present in the red areas of pattern A (panels A, E, I, and M). The red areas of pattern B indicate sediment laden

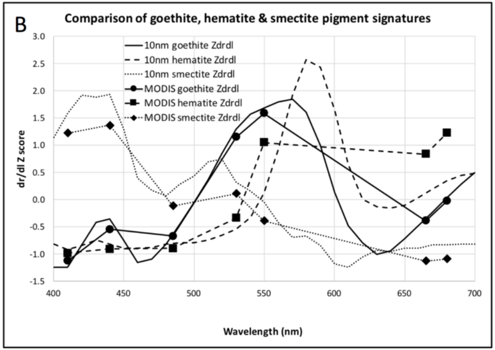
riverine discharge is present (panels B, F, J, and N). Pattern C (panels C, G, K, and O) highlight the presence of hematite and chlorophyll degradation products. Red areas in pattern D (panels D, H, L, and P) indicate the presence of cryptophytes.

Figure 8: The ‘constructed CI’ based on the weighted addition of the 4 VPCA patterns extracted from the 28 July 2015 MODIS Aqua image. Shown here in comparison to the NOAA CI calculated from the same image. The NOAA CI figure was taken from the publicly available HAB bulletin (https://www.glerl.noaa.gov//res/HABs_and_Hypoxia/) hosted by NOAA-GLERL.

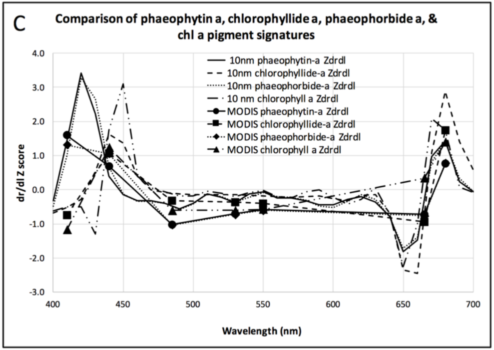

A
N
B
N
WLE 08
WLE 04
WLE 02
WLE 13
WLE 09
WLE 14
WLE 06
WLE 12
WLE 15
C
N
WLE 15
Bells
SND2
SND1
EC1163
SND4
SND6
CBCZ6
CBCZ2
CBCZ5
CBCZ3
CBCZ4


A

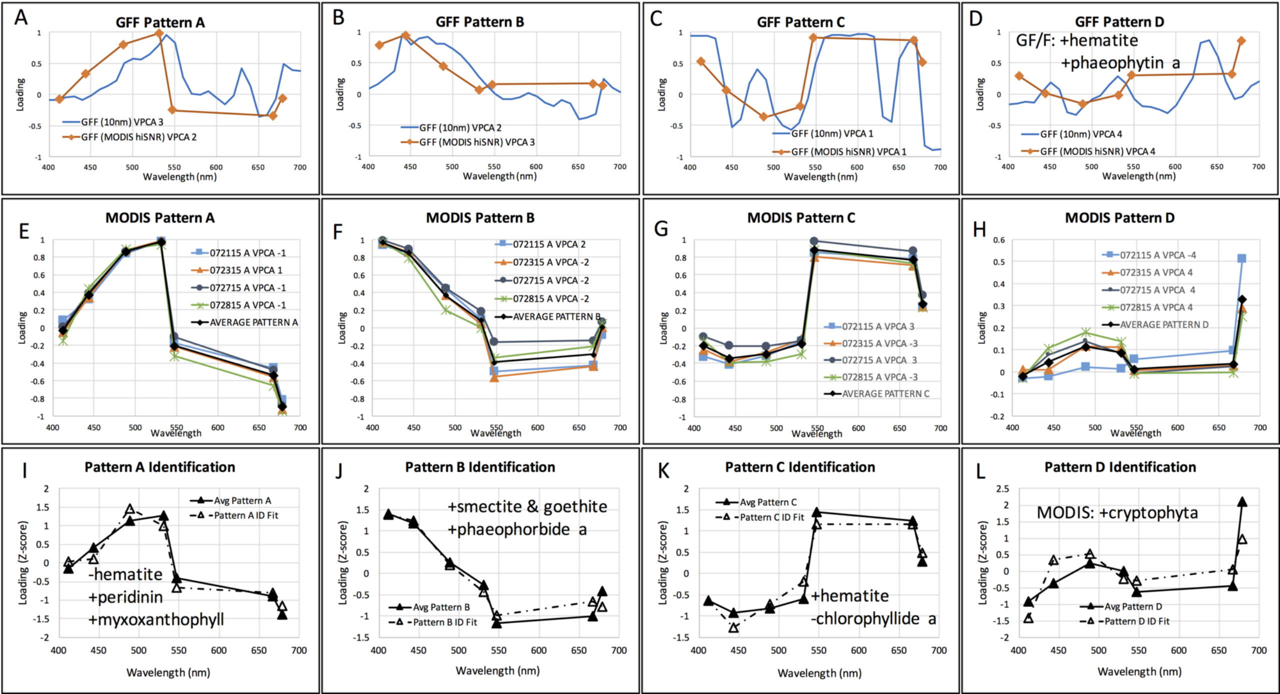
Comparison of myxoxanthophyll & peridinin pigment signatures


B

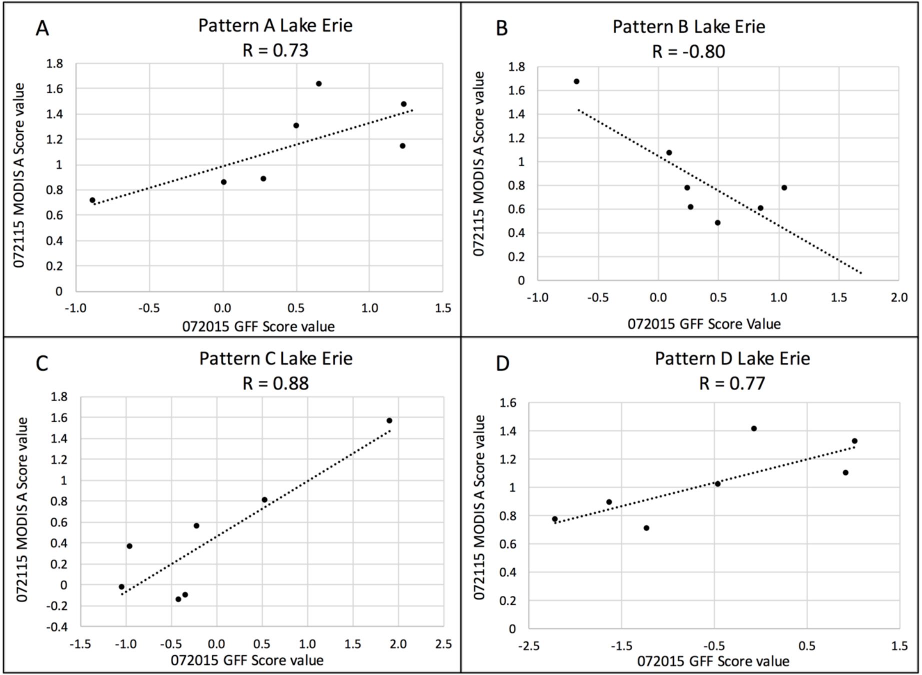
Comparison of goethite, hematite & smectite pigment signatures

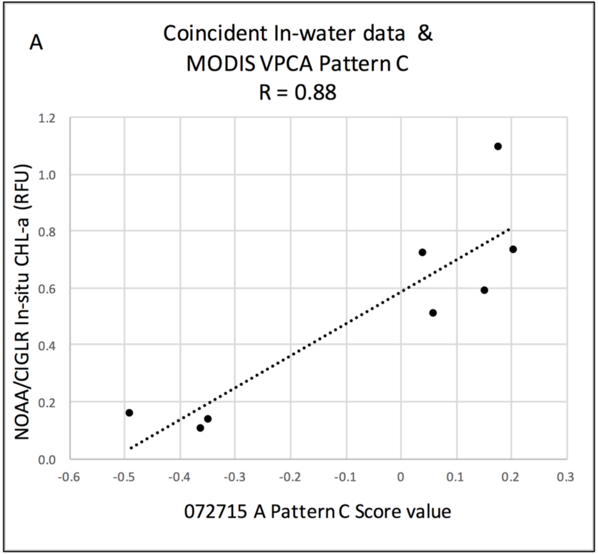

C

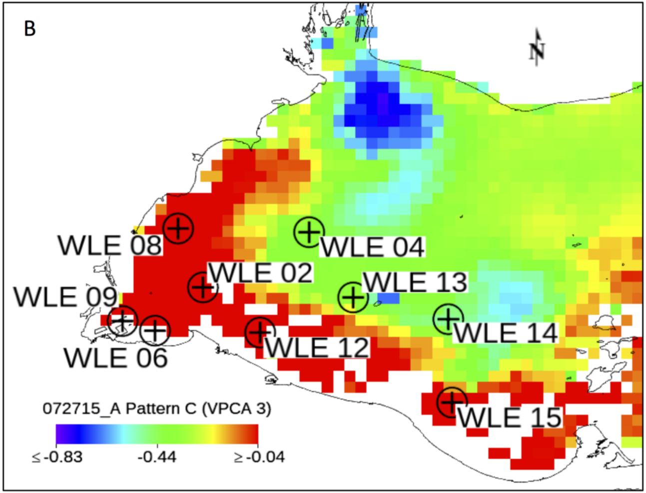
Comparison of phaeophytin a, chlorophyllide a, phaeophorbide a, & chl a pigment signatures

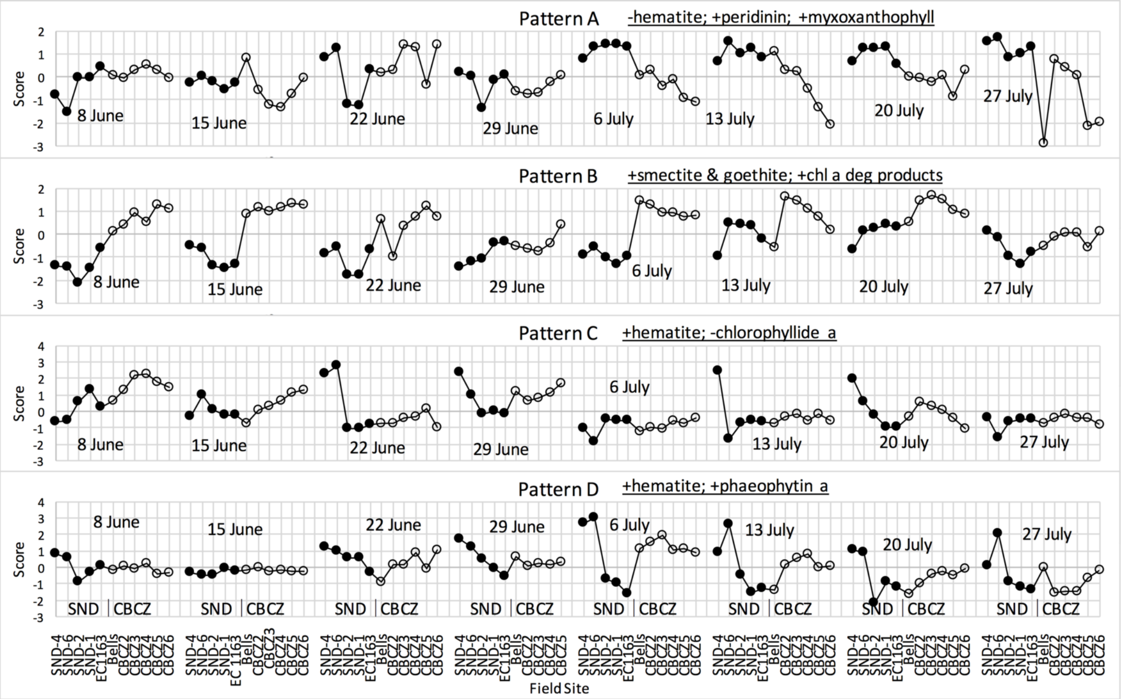


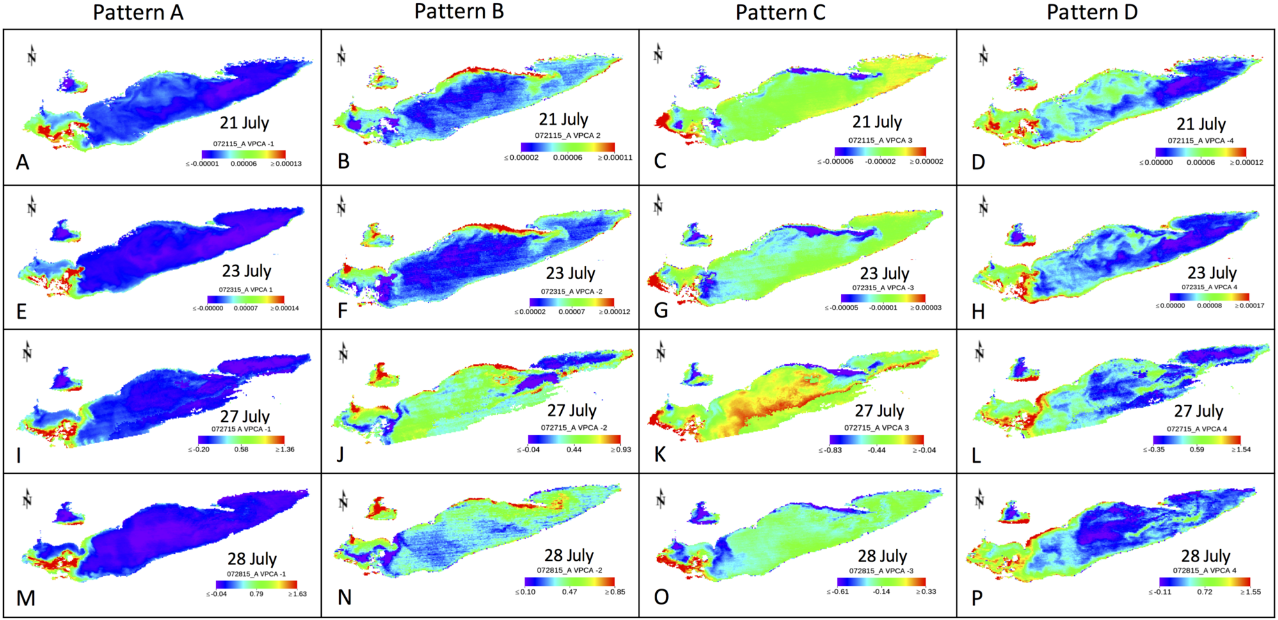
A
GFF Pattern A
GFF (10nm) VPCA 3
GFF (MODIS hiSNR) VPCA 2
B
GFF Pattern B
GFF (10nm) VPCA 2
GFF (MODIS hiSNR) VPCA 3
C
GFF Pattern C
GFF (10nm) VPCA 1
GFF (MODIS hiSNR) VPCA 1
D
GFF Pattern D
GF/F: +hematite +phaeophytin a
GFF (10nm) VPCA 4
GFF (MODIS hiSNR) VPCA 4
Loading
Wavelength (nm)
E
MODIS Pattern A
072115 A VPCA -1
072315 A VPCA 1
072715 A VPCA -1
072815 A VPCA -1
AVERAGE PATTERN A
F
MODIS Pattern B
072115 A VPCA 2
072315 A VPCA -2
072715 A VPCA -2
072815 A VPCA -2
AVERAGE PATTERN B
G
MODIS Pattern C
072115 A VPCA 3
072315 A VPCA -3
072715 A VPCA 3
072815 A VPCA -3
AVERAGE PATTERN C
H
MODIS Pattern D
072115 A VPCA -4
072315 A VPCA 4
072715 A VPCA 4
072815 A VPCA 4
AVERAGE PATTERN D
Wavelength
I
Pattern A Identification
Avg Pattern A
Pattern A ID Fit
-hematite
+peridinin
+myxoxanthophyll
J
Pattern B Identification
+smectite & goethite
+phaeophorbide a
Avg Pattern B
Pattern B ID Fit
K
Pattern C Identification
Avg Pattern C
Pattern C ID Fit
+hematite
-chlorophyllide a
L
Pattern D Identification
MODIS: +cryptophyta
Avg Pattern D
Pattern D ID Fit
Loading (Z-score)
Wavelength (nm)
400
450
500
550
600
650
700


A
Pattern A Lake Erie
R = 0.73
072115 MODIS A Score value
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-1.0
-0.5
0.0
0.5
1.0
1.5
072015 GFF Score value
B
Pattern B Lake Erie
R = -0.80
072115 MODIS A Score value
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
072015 GFF Score Value
C
Pattern C Lake Erie
R = 0.88
072115 MODIS A Score value
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
2.5
072015 GFF Score value
D
Pattern D Lake Erie
R = 0.77
072115 MODIS A Score value
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-2.5
-1.5
-0.5
0.5
1.5
072015 GFF Score value

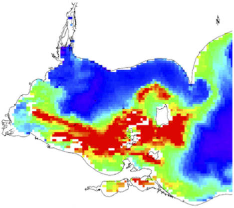

A
Coincident In-water data &
MODIS VPCA Pattern C
R = 0.88
NOAA/CIGLR In-situ CHL-a (RFU)
0.0
0.2
0.4
0.6
0.8
1.0
1.2
-0.6
-0.5
-0.4
-0.3
-0.2
-0.1
0
0.1
0.2
0.3
072715 A Pattern C Score value


B
N
WLE 08
WLE 04
WLE 02
WLE 13
WLE 09
WLE 14
WLE 06
WLE 12
WLE 15
072715_A Pattern C (VPCA 3)
≤-0.83
-0.44
≥-0.04


Pattern A
-hematite; +peridinin; +myxoxanthophyll
Pattern B
+smectite & goethite; +chl a deg products
Pattern C
+hematite; -chlorophyllide a
Pattern D
+hematite; +phaeophytin a
Score
2
1
0
-1
-2
-3
4
3
8 June
15 June
22 June
29 June
6 July
13 July
20 July
27 July
SND
CBCZ
SND-4
SND-6
SND-2
SND-1
EC1163
EC 1163
Bells
CBCZ2
CBCZ3
CBCZ4
CBCZ5
CBCZ6
Field Site

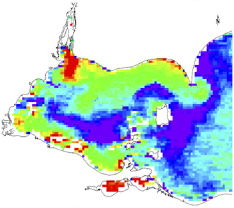

Pattern A
Pattern B
Pattern C
Pattern D
N
21 July
072115_A VPCA -1
≤-0.00001
0.00006
≥0.00013
A
21 July
072115_A VPCA 2
≤0.00002
0.00006
≥0.00011
B
21 July
072115_A VPCA 3
≤-0.00006
-0.00002
≥0.00002
C
21 July
072115_A VPCA -4
≤0.00000
0.00006
≥0.00012
D
23 July
072315_A VPCA 1
≤-0.00000
0.00007
≥0.00014
E
23 July
072315_A VPCA -2
≤0.00002
0.00007
≥0.00012
F
23 July
072315_A VPCA -3
≤-0.00005
-0.00001
≥0.00003
G
23 July
072315_A VPCA 4
≤0.00000
0.00008
≥0.00017
H
27 July
072715 A VPCA -1
≤-0.20
0.58
≥1.36
I
27 July
072715_A VPCA -2
≤-0.04
0.44
≥0.93
J
27 July
072715_A VPCA 3
≤-0.83
-0.44
≥-0.04
K
27 July
072715_A VPCA 4
≤-0.35
0.59
≥1.54
L
28 July
072815_A VPCA -1
≤-0.04
0.79
≥1.63
M
28 July
072815_A VPCA -2
≤0.10
0.47
≥0.85
N
28 July
072815_A VPCA -3
≤-0.61
-0.14
≥0.33
O
28 July
072815_A VPCA 4
≤-0.11
0.72
≥1.55
P

Pattern A



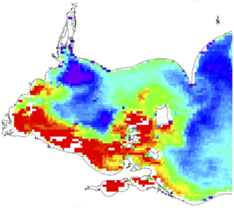
+

Pattern B



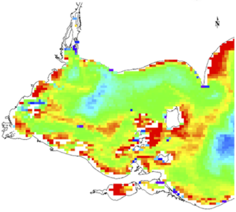
+

Pattern C



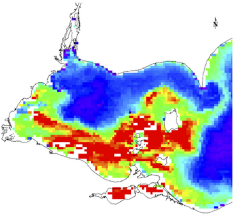
+

Pattern D



=

Constructed CI



072815 A
NOAA CI