6-2019

Determining remote sensing spatial resolution requirements for the monitoring of harmful algal blooms in the Great Lakes

John Lekki
NASA Glenn Research Center

Eric Deutsch
NASA Glenn Research Center

Michael Sayers
Michigan Technological University, mjsayers@mtu.edu

Karl Bosse
Michigan Technological University, krbosse@mtu.edu

Robert Anderson
NASA Glenn Research Center

See next page for additional authors

Journal of Great Lakes Research 45 (2019) 434–443

Contents lists available at ScienceDirect

Journal of Great Lakes Research

journal homepage: www.elsevier.com/locate/jglr

Determining remote sensing spatial resolution requirements for the monitoring of harmful algal blooms in the Great Lakes

John Lekki [a,*], Eric Deutsch [a], Mike Sayers [b], Karl Bosse [b], Robert Anderson [a], Roger Tokars [a], Reid Sawtell [b]

[a] NASA Glenn Research Center, 21000 Brookpark Rd., Cleveland, OH 44135, United States of America
[b] Michigan Tech Research Institute, 3600 Green Ct., Ste. 100, Ann Arbor, MI 48105, United States of America

ARTICLE INFO

Article history:
Received 25 May 2018
Accepted 15 March 2019
Available online 2 April 2019

Communicated by George Leshkevich

Keywords:
Remote sensing
Hyperspectral
Cyanobacteria
Harmful algal blooms
Airborne
Spatial resolution

ABSTRACT

Harmful algal blooms (HABs) have become a major health and environmental concern in the Great Lakes. In 2014, severe HABs prompted the State of Ohio to request NASA Glenn Research Center (GRC) to assist with monitoring algal blooms in Lake Erie. The most notable species of HAB is *Microcystis aeruginosa*, a hepatotoxin producing cyanobacteria that is responsible for liver complications for humans and other fauna that come in contact with these blooms. NASA GRC conducts semiweekly flights in order to gather up-to-date imagery regarding the blooms' spatial extents and concentrations. Airborne hyperspectral imagery is collected using two hyperspectral imagers, HSI-2 and HSI-3. Hyperspectral imagery is necessary in order to conduct experiments on differentiation of algal bloom types based on their spectral reflectance. In this analysis, imagery from September 19, 2016 was utilized to study the subpixel variability within the footprint of arbitrary sized pixels using several analysis techniques. This particular data set is utilized because it represents a worst case scenario where there is significant potential for public health concern due to high concentrations of microcystin toxin found in the water on this day and the concurrent observational challenges to accurately measure the algal bloom concentration variability with a remote sensing system due to the blooms high spatial variability. It has been determined that the optimal spatial resolution to monitor algal blooms in the Great Lakes is at most 50 m, and for much lower error 25 m, thus allowing for greater ease in identifying high concentration blooms near the surface. This resolution provides the best sensitivity to high concentration areas that are of significant importance in regard to human health and ecological damage.

Introduction

Harmful algal blooms (HABs) are becoming a major concern worldwide (Hallegraeff, 1993; Anderson et al., 2012). They are of very significant concern in the Laurentian Great Lakes. The Great Lakes are a vital resource, providing drinking water to approximately 40 million people in the United States and Canada (Gronewold et al., 2013) and a home to fisheries that are worth more than $7 billion dollars annually for the two countries (Great Lakes Fishery Commission, 2014). These lakes are also high traffic areas for marine transport and provide for many recreational activities. Harmful algal blooms are particularly severe in the relatively shallow western basin of Lake Erie from which three million people receive drinking water (Wynne and Stumpf, 2015).

Understanding these HABs is necessary to provide the basis for effective measures to mitigate impacts and eventually curtail these growths in the future (Sinha et al., 2017; Anderson et al., 2012). Frequent monitoring and tracking of the bloom species, spatial extent, and concentration over long periods of time is necessary to identify trends and factors that contribute to a bloom. Techniques for differentiating between algal types are vital, particularly in distinguishing between nuisance and harmful blooms (Ohio Environmental Protection Agency, n.d). Nuisance blooms are those that pose no health risk to humans. In contrast *Microcystis aeruginosa*, which is commonly found in the western basin of Lake Erie, is a cyanobacteria that produces the hepatotoxin microcystin (Hitzfeld et al., 2000). This toxin infiltrates the liver and is responsible for acute and chronic illness in humans and wildlife (Butler et al., 2009). Mapping the spatial extents and concentrations of algal blooms is important for characterizing the larger macrostructures in blooms since algal growths do not develop as regions of homogenous unicellular concentrations, but rather as colonies or strands and filament-like structures in the water.

Abbreviations: L_λ, spectral radiance (W/m^2 sr nm); L_u, upwelling spectral radiance (W/m^2 sr nm) measured just over water; L_a, upwelling spectral radiance (W/m^2 sr nm) measured from the aircraft; E_d, downward spectral irradiance (W/m^2 nm) measured just over water; E_a, downward spectral irradiance (W/m^2 nm) measured from the aircraft; R_{RS}, remote sensing reflectance (sr^{-1}) measured just over water; ρ_a, remote sensing reflectance (non dimensional); R(), non dimensional remote sensing reflectance for a particular band (non dimensional); θ_s, solar zenith angle (degrees); SS, spectral shape (non dimensional); CI, cyanobacteria index (non dimensional).

* Corresponding author.
E-mail address: John.D.Lekki@nasa.gov (J. Lekki).

https://doi.org/10.1016/j.jglr.2019.03.014

The use of remote sensing and, specifically hyperspectral remote sensing, is a key component to adequately monitor and research these algal blooms (Kutser et al., 2006). In situ data collection is fundamentally necessary for studying HABs, but airborne remote sensing significantly augments in situ measurements for obtaining high resolution measurements over larger areas, which is important when studying HABs due to the sheer scale of the blooms. Airborne data collection allows researchers to collect imagery during ideal weather conditions; and, if cloudy weather cannot be avoided, flying under the clouds is possible and has been done in the past. Flying closer to the surface than an orbiting satellite allows for higher spatial resolution and flying at different altitudes can allow for changing the balance between spatial resolution and imaged swath width as desired. For satellites, there is always a tradeoff between resolution, swath width and repeat time. These trades are made during the design phase and cannot be altered after launch. With that in mind, it is clear that maximizing spatial resolution for research isn't always optimal.

The analysis in this paper has been focused on determining what spatial resolution would be optimal for obtaining maximum coverage of HABs while not losing key details necessary for adequately monitoring bloom concentrations. The benefits of determining the optimal spatial resolution are significant. First, monitoring algal growths at this optimal resolution will elucidate the presence and structure of these growths and provide for the best information on the bloom. Second, if the spatial resolution is only constrained to the necessary spatial resolution for adequate observation (and no further), then more area can be observed and shorter repeat times can be obtained from orbiting remote sensing instruments. Third, the cost of developing, sourcing components, and maintaining remote sensing systems may be reduced when building to lower technical specifications.

The data set chosen for analysis in this paper represents a worst case scenario that is insightful for determining remote sensing resolution requirements for a system that is tasked with monitoring harmful algal blooms. While NASA Glenn has acquired many hyperspectral data sets of potentially harmful algal blooms in 2014, 2015 and 2016, this particular data set was analyzed because the bloom had very high spatial variability. There were areas of scum, which can have the most significant human health concerns in a toxin producing bloom, as well as much lower concentration areas that do not have the same level of concern for potential health impacts (World Health Organization, 2003). Water sampling data were also collected on the same day. At both locations scums were reported and samples were obtained both outside and inside the scum (H. Vanderploeg, personal communication). Laboratory analysis from scum samples at the two locations measured 1.62 and 0.5 μg/l of particulate microcystin, 0.24 and 0.1 μg/l of dissolved microcystin and extracted phycocyanin was 109 and 16.75 μg/l. Outside of the scum at these two locations the reported values were much lower than the scum values with 0.1 and 0.13 μg/l of particulate microcystin, 0.27 μg/l and below detection limit of dissolved microcystin, and extracted phycocyanin was 42.4 and 1.14 μg/l. This sampling confirms that this was patchy harmful algal bloom. This condition is of significant relevance for a remote sensing system being utilized for water quality monitoring because the human health concerns associated with the highest concentrations drive the need for accurate measurements of concentration variability. If a remote sensing system does not have the resolution to accurately measure the variability, then multiple concentrations will be averaged together and eventually the information on the presence of very high concentrations, the peaks of the signal, may be lost. It should be noted that the size of the high concentration areas may be different for different algal types and surface conditions, and so this analysis is not meant to represent typical situations but instead address a worst case scenario for a microcystis bloom in the western basin of Lake Erie. In the following sections, the process that has been utilized for obtaining remote sensing and in situ data to address this question is described and is followed by an analysis of these data to identify a minimum spatial resolution for accurately monitoring these blooms.

Data collection and processing

Data collection

In 2014, 2015, 2016, and 2017 the NASA Glenn Research Center used the S-3 Viking and the DHC-6 Twin Otter as platforms for hyperspectral sensors in order to study remote sensing algorithms for HAB detection and discrimination (Lekki et al., 2017). Depending on location and weather, the aircraft typically fly over regions of interest at either 3000 ft, 8500 ft, or 24,000 ft. From late July to early August 2016, the NASA GRC has been flying the HSI-2. The specifications of HSI-2 are shown in Table 1.

In August 2016, the researchers began using newer HSI-3.0 imager that has specifications detailed in Table 2, often while simultaneously using the HSI-2.

The increase in horizontal pixels between these two sensors, in addition to the much wider field of view, allows for much wider swaths to be imaged during a given flyover. In addition to a downward-looking hyperspectral imager, the aircraft is also equipped with an upward-looking spectroradiometer for measuring downwelling irradiance at the aircraft. This radiometer is particularly useful when operating under clouds.

The research team conducted approximately one flight per week throughout the summer and early fall of 2016. Coinciding with these flights, collaborating researchers would go out on the water to obtain validation data at designated locations consisting of ground-based readings using a hand held spectroradiometer, fluoroprobe measurements and water samples. The water validation of the airborne measurements was extensive and is detailed in Lekki et al. (2017).

Hyperspectral data pre-processing

The hyperspectral imaging system is activated when the aircraft is over a target of interest. With the imager on, at each position along a track, a two-dimensional image is acquired with spectral information along one dimension and cross-swath information along the other dimension. Stacked up, the result is a 3-dimensional data stack with two linear dimensions (across the swath and along the track) and one spectral dimension. Along with each image in the stack, ancillary data is available including a time stamp and GPS coordinates, along with altitude, roll, pitch and yaw. The upward-looking spectroradiometer data is acquired independently from the hyperspectral images.

Before the data can be utilized for analysis, each data stack is processed using routines written using IDL and ENVI (Exelis Visual Information Solutions, Boulder, Colorado). Each hyperspectral data stack must be converted to units of radiance using calibration images based on pre-flight data acquired in the lab using a calibrated Integrating Sphere. The wavelength scale is adjusted, if necessary, based on the location of specific solar spectral features. This adjustment is necessary because the instrument gets much colder in flight than the temperature at which it is calibrated, and this causes thermal contraction of the instrument which leads to a shift in the wavelength measurement. Following this, each pixel in the track is provided with a longitude and latitude using the ancillary information (GPS, altitude, roll, pitch, and yaw) mentioned above. The data in the track may then be georeferenced using built-in ENVI routines.

Table 1
HSI-2 technical specifications.

Wavelength range, nm	400 nm–900 nm
Spectral sampling resolution	1 nm
Radiometric resolution	14 bit
Horizontal pixels	658
Field of view (FOV)	12°
Optical spectral resolution	2.5 nm

Table 2
HSI-3 technical specifications.

Wavelength range, nm	400 nm–900 nm
Spectral sampling resolution	2 nm
Radiometric resolution	14 bit
Horizontal pixels	1032
Field of view (FOV)	72°
Optical spectral resolution	4 nm

Hyperspectral data processing

There are many ways to process the hyperspectral data and analyze the spatial variability in terms of a number of criteria. For instance the data can be examined as a color image. While the water in the true-color image (Fig. 1a) does have a blue green appearance, it is extremely difficult to visually discern areas of relatively high algal concentration from areas of relatively low concentration. From true-color imagery there often is an impression that there is very little change in algal concentration when there actually is a very large difference. However, in a color-infrared image (Fig. 1b) the contrast between high and low concentration can be readily seen. It is easy to see areas of higher density due to the higher reflectance in the 709 nm band which shows up as a strong red color in this particular image.

Several factors played into the decision about the best image to examine in order to adequately observe spatial variability. If an image of the scene was a true-color image, it is likely that there would be very little variability identified. On the other hand, by looking at only the 709 nm band, there would be significantly higher frequency spatial variability but there would be difficulty directly relating it to concentration. In this study we have chosen to use the spatial variability of the cyanobacterial index image (CI) (Wynne et al., 2008; Wynne et al., 2010; Wynne et al., 2013; Wynne and Stumpf, 2015). The Cyanobacterial Index is a calculation that quantifies the curvature of the chlorophyll *a* spectral reflectance around 681 nm.

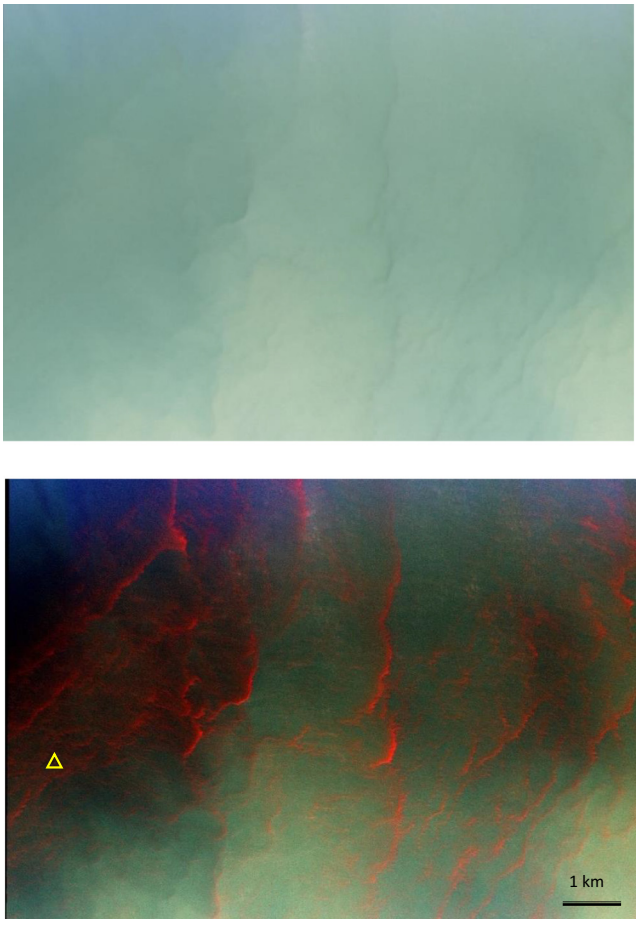

Fig. 1. Images of the algal bloom studied taken at 10:30 AM on September 19, 2016. a) A true-color image of the bloom b) color-infrared image of the bloom using 490 nm for blue, 550 nm for green and 709 nm for red. The yellow triangle shows the location of Toledo Light #2.

The cyanobacterial index image was chosen because it is commonly utilized in the NOAA HAB bulletin to report on harmful algal blooms in the western basin of Lake Erie. It has also been documented that the cyanobacterial index has a relationship to cell count concentration in the water (Wynne et al., 2010; Lunetta et al., 2015). By approaching algal bloom hazards primarily from the perspective of a human health issue, variability of concentration is the metric we are most interested in. This perspective made the CI image the more appropriate choice for studying spatial variability of algal blooms.

The CI image used for this study was obtained through the following process. Hyperspectral radiance imagery was obtained from the HSI3 flown at an altitude of 24,000 ft at 10:30 on September 19, 2016. This particular data set is well suited for this study because the high altitude, coupled with the wide field of view of the instrument, produced an image that is approximately 11 km by 11 km and there were concurrent water samples obtained within the image area. With high spatial resolution over a very large area it is possible to study spatial variability by emulating the spatial specifications of common satellite sensors and still have many pixels for an adequate sample size. For instance, if the many 2.4 m resolution HSI3 pixels are combined together to synthetically produce 1 km^2 size pixels, there will still be at least 100 synthetic pixels in the approximately 120 km^2 image, which are enough to have a statistically powerful sample size.

The data are being interpreted in terms of at-sensor-reflectance. In this case non-dimensional remote sensing reflectance is calculated using the spectral irradiance measured at the aircraft, E_a, and the Hyperspectral Radiance measured at the aircraft, L_a, and the solar zenith angle θ_s using the following equation (Gordon and Wang, 1994; Mobley et al., 2016).

$$\rho_a = \frac{\pi L_a}{E_a * \cos\theta_s} \quad (1)$$

It should be noted that atmospheric correction was not applied to this data set. The CI is relatively insensitive to atmospheric effects because the method is based on determining spectral shape for bands that are relatively close in wavelength (Wynne et al., 2010; Campbell and Esaias, 1983; Philpot, 1991). The reflectance is obtained for the bands shown in Table 3 by averaging the hyperspectral bands across the defined MERIS band wavelength ranges to closely match the MERIS bands that were used to obtain cell concentration relationships in Wynne et al. Table 3 shows the bands that were created from the hyperspectral data on the corresponding MERIS band:

Table 3
Corresponding MERIS band to hyperspectral imager synthetic bands.

MERIS band	Hyperspectral band center wavelength (nm)	Hyperspectral wavelength range (nm)
7	665	660–670
8	681	677–685
9	709	704–714

The CI image was then calculated using the following equations.

$$SS = R(681) - R(665) - [R(709) - R(665)] \times \left(\frac{681-665}{709-665}\right) \quad (2)$$

$$CI = -SS_{(681)} \quad (3)$$

$$CI = R(681) - R(665) - [R(709) - R(665)] \times \left(\frac{681-665}{709-665}\right) \times -1 \quad (4)$$

where:

R(665) = reflectance (665 nm)
SS = spectral shape
R(681) = reflectance (681 nm)
CI = chlorophyll-*a* index
R(709) = reflectance (709 nm).

This CI data set was then utilized for examining spatial variability at multiple synthetic resolutions.

In situ radiometry

The in situ (on water) spectrometer system was mounted to the structure of Toledo Light #2 (Fig. 1b) and it was designed to simultaneously measure the downwelling spectral plane irradiance (Ed, W/m^2/nm) and upwelling spectral radiance (Lu, W/m^2 sr nm) in order to derive remote sensing reflectance (Rrs, sr^{-1}) as the ratio of Lu/Ed. It should be noted that the diffuse sky surface reflectance was not measured or removed from these measurements. Irradiance and radiance measurements were made with two independent Ocean Optics STS spectroradiometers (https://oceanoptics.com/product-category/sts-series/) controlled by a Raspberry Pi (https://www.raspberrypi.org/) computer. The STS radiometers have bands that are sampled every 0.5 nm and have optical resolution of 1.5 nm from 380 to 850 nm. Irradiance was measured with a cosine collector foreoptic connected to a fiber optic cable which plugged into one of the radiometers. Upwelling radiance was measured with a 25 degree field-of-view (FOV) foreoptic also connected to a fiber optic cable plugged into the other spectroradiometer. The upwelling radiance sensor was pointed 40° off-nadir and oriented in the southwest direction so as to minimize the water surface specular reflection (Mobley, 1999) in the morning hours when airborne flights typically occurred. Both the irradiance and radiance spectrometers were relatively calibrated against an ASD Fieldspec III spectroradiometer recently calibrated by the manufacturer. The radiometer package was mounted at a height of approximately five meters above the water surface. With the 25 degree FOV pointed at 40° off nadir at a height of five meters, the ground sample resolution was an ellipse with a major axis of 3.9 m and a minor axis of 2.2 m and a total viewing area of 6.8 m^2.

Spectral profiles were measured at one minute intervals for the daylight period. The radiometer system was powered by battery recharged with solar panels. In cases of prolonged (i.e. days) cloud cover the radiometer did not record spectra due to insufficient power. For the September 19, 2016 study date the radiometer system collected 175 spectra (just under 3 h) beginning at 11:33 and ending at 14:27 EDT. The CI was calculated for all valid spectra following Eq. (4) where the STS bands were resampled to the specifications detailed in Table 3.

Since the ground sample area of in situ radiometer package is fixed as well as the sampling rate (i.e. 1 min) the distance or size of the water masses passing through the instrument field-of-view can be calculated using an estimate of the current speed. Currents were acquired from the Great Lakes Observing System (GLOS) data portal (https://www.glos.us/) for September 19, 2016 at the Toledo Light #2. This current data is derived from the Great Lakes Coastal Forecast System (GLCFS) which provide direction and velocity (m/s) at hourly intervals for any selectable location.

Patches of high CI (Table 4) were identified from the spectra. Contiguous patches were defined as a series of spectra where CI was above the high (0.004) threshold. Patch size was estimated using the following equation:

$$Distance\ (m) = Rate\ (m/s) \times Time\ (s) \quad (5)$$

where the rate is the current speed and Time is the number of seconds a contiguous patch was present in the instrument FOV. This in situ data and the remote sensing data discussed earlier is used in the subsequent section for the spatial analysis of this algal bloom.

Spatial analysis

For comparison, the spatial variability of the CI data set was analyzed in multiple ways. The first method was to resample the imagery into lower resolution images to obtain a visual understanding of how the lower resolution impacts the image clarity (Cushnie, 2004). The second method was to perform a neighborhood variance analysis of the image at different resolutions to understand how variance in the image changes with resolution (Woodcock and Strahler, 1987). The third method was to perform visual inspection of the image, manually measure the size of features at hundreds of locations and then examine the statistics of those measurements (Korpela, 2004). The fourth method was to look at the variability of a pixel centered in the image at various scales (Atkinson and Aplin, 2004). The final approach was to create an arbitrary transect across the image to see how wide features are at the medium and high CI thresholds on this one dimensional transect. By using these methods together, we can obtain an understanding of the spatial size of cyanobacterial colonies in the western basin of Lake Erie.

Degradation of imagery to lower spatial resolutions

In Fig. 2, the CI image of the algal bloom on Sept 19, 2016 is shown where it has been synthetically resampled to lower resolutions. In Fig. 2a–f, the image is shown after being resampled from 2.4 m to 31 m, 50 m, 100 m, 200 m, 300 m, and 400 m, respectively. While the loss in detail at lower resolution is difficult to discern from these small images, they can be used to see where larger structures start to lose their consistency. In both the 31 m and 50 m images (Fig. 2a and b) it can be readily seen that a lot of the detail in the image is retained. At 100 m resolution (Fig. 2c), many smaller details are noticeably blurred and some detail/information has been lost. At much lower resolutions 200, 300, and 400 m it is clear that the image has been seriously degraded. From the imagery in Fig. 2, it appears that a resolution that does not result in too much loss of information is in the range of 50 m.

Global subpixel standard deviation analysis

The global subpixel standard deviation was computed for the CI image to understand the variability within the image. This was calculated by replacing each pixel in the CI image with the standard deviation of an N × N box of pixels around a target pixel. This is done for all of the pixels in the image except for the pixels along the edge of the image. The

Table 4
Categorization of CI corresponding to CI values and Cell concentration from Wynne et al. (2010).

Category	CI values	Approximate cell count/mL
Low CI	0–0.0004	1 to 300,000
Medium	0.0004–0.004	300,000 to 500,000
High	>0.004	>500,000

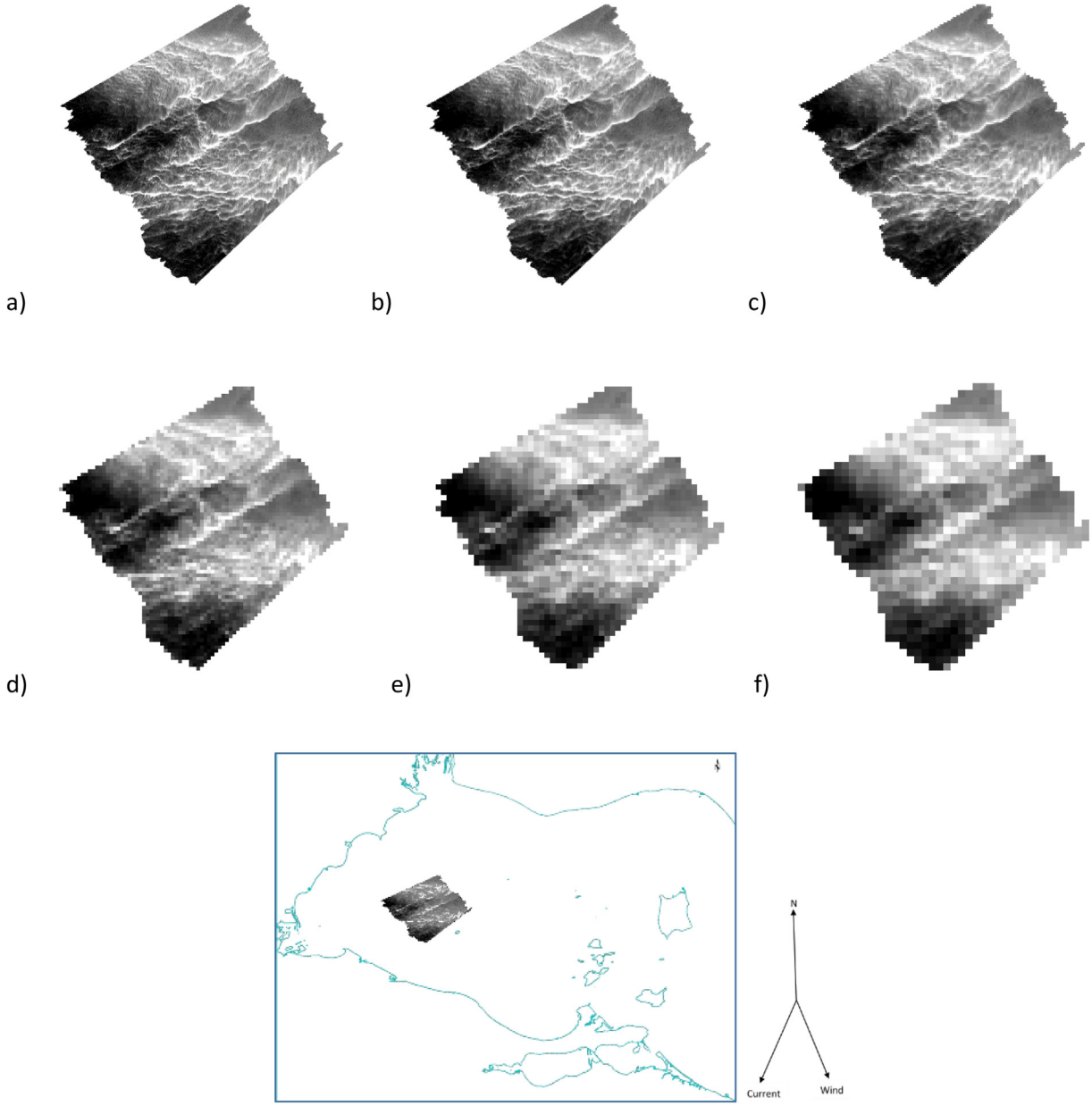

Fig. 2. Resampled imagery at a) 31 m, b) 50 m, and c) 100 m, d) 200 m, e) 300 m, f) 400 m respectively The image is approximately an 11 km × 11 km area. A map showing the location and size of the bloom in the western basin of Lake Erie is shown in the inset along with arrows showing North, the direction the current is moving toward and the direction the wind is moving toward.

total of all the standard deviation values obtained in this way is recorded for each image. The total subpixel standard deviation was computed for larger neighborhood sizes to understand the variability within larger pixel sizes (Fig. 3).

From the graph in Fig. 3, the global subpixel standard deviation within the neighborhoods reaches a peak at a neighborhood size of 41 pixels and drops off after that. A neighborhood of 41 × 41 is approximately 98 m × 98 m in size. In this analysis, we find that the peak total standard deviation using this metric is 98 m and the point at which a lower variability is found is at 7 pixels or 26 meter resolution.

Determination of high concentration colony patch widths using visual inspection

Visual inspection of the widths of high concentration colony patches in the CI image was also used to examine the scale of the algal blooms. As the blooms look continuous, ranging from a very low concentration

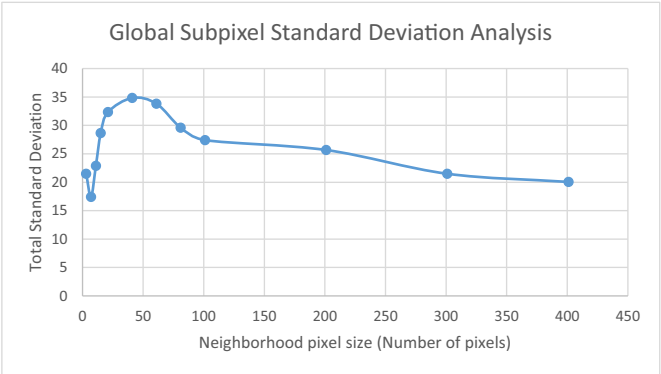

Fig. 3. Plot of the global subpixel variability for the original CI image and subsequent images with coarser synthetic resolutions.

to a very high concentration, we utilized thresholds for Medium and High concentrations as defined in Wynne et al. (2010) to determine the colony size. While these images are similar to the color infrared image (Fig. 1b), Wynne provides links between CI values and cell concentration approximations that can in turn be used to visualize the severity of the bloom. The thresholds provided in that paper are shown in Table 4.

The size and the form of the blooms have been color coded into "medium" and "high" categories (Fig. 4) according to the CI thresholds specified in Table 4. Inspection of the size of the 'High' and 'Medium' concentration features provides further insight into the size of these features. Note that the shape of the dense bloom patches is very long, so we have focused on the width of these patches, which directly relates to their spatial detectability in remote sensing systems when a given pixel is likely to be a mixture of two concentrations.

The visual inspection process was performed by using the both the 'Medium' and 'High' CI image. Hundreds of vector lines were drawn perpendicular to the length axis of any algal community structures and is illustrated in Fig. 5. The length of each vector line was recorded and tabulated. From this record of vector line lengths, a histogram was created for each of the concentrations areas. The high concentration vectors were binned in 10 meter increments while the vectors across the medium concentration features were binned in 25 meter increments.

Because the vectors were manually drawn, there is inherent human error in the measurements. The spacing on the vectors is not completely uniform and not all of the features are covered by vectors in the medium concentration image. However, 100 s of vectors have been drawn across both wider and thinner features and so the numbers of vectors drawn should help to alleviate the effects of inevitable human error.

Results from inspection of feature sizes

The following results show the accumulated histograms for the vectors manually drawn across the high CI regions and the medium CI regions. The histograms of the medium and high threshold vector measurements are shown respectively in Fig. 6. The histogram for the medium image show that 75% of the vectors going across the features are of 75 m or less in size and the peak bin is at 75 m. The data for the high concentration threshold image show that 75% of the features are 30 m or less in width while the peak is in the 30 meter bin. It should also be noted here that the shape of the histograms are very similar to the local standard deviation neighborhood analysis in Fig. 3, and the peak of that analysis lies between the peaks of these two histograms. From this data set, it can be inferred that the resolution for a system that is matched to the feature sizes of this high and medium concentration bloom image would be in the range of 30 to 75 m.

Transect information

To gain further understanding of the bloom features, a plot of the CI values along a linear transect (Murphy et al., 2008) was made and is shown in Fig. 7. In this plot it is easy to see how the CI varies along the transect, and as the CI is directly related to concentration, Fig. 7 also yields the relative concentration of the algal bloom. The tendency of cyanobacteria to form dense patches is clearly evident. Only four of the locations along the 9 km transect have CI values that are above 0.004 and hence concentrations above 500,000 cells/mL.

While there are very few locations that cross the High concentration threshold, there are many that cross the Medium concentration threshold. Most of those are much smaller in width than the very high peaks.

For each selected peak the widths for High and Medium concentration are: a) 53 m High, 312 m Medium, b) 94 m High, 256 m Medium, c) 10 m High, 160 m Medium, d) 33 m High, 185 m Medium, e) 90 m Medium.

In four locations where the CI is high, the widths are 10, 33, 53, and 94 m (Fig. 7). While this is just one line of data from the image it is apparent that the values would not be unexpected given the histogram in Fig. 6. Likewise, the corresponding widths where the CI is above medium is 160 m, 185 m, 312 m, and 256 m. This set is on the higher end of the medium histogram in Fig. 8, and this should be expected as these are the locations on the transect that have the highest peak concentration while there are many smaller peaks in the transect that have not been inspected.

There is a pattern that should also be noted from the Fig. 7 plot. The shapes of CI peaks are asymmetrical. Based on the wind directions of 150° and current direction of between 208 and 167°, which is travelling in approximately the direction of the transect in Fig. 4b, we see that the CI peaks start low (right side of graphs) and then there is a steep increase followed by a more gradual decrease. This indicates the forming of a strong frontal feature of cyanobacteria biomass in the direction of the current and wind in these conditions.

A further check on the histograms obtained through manual inspection, Fig. 6, can be made by examining the transect plot, Fig. 7. The widths of the peaks where the CI is greater than the medium threshold may be recorded and the statistics of these peak widths can be examined in the histogram in Fig. 8. Here the peak is in the 50 meter bin instead of the 75 meter bin. This is probably because more small features are captured through an automated analysis than through

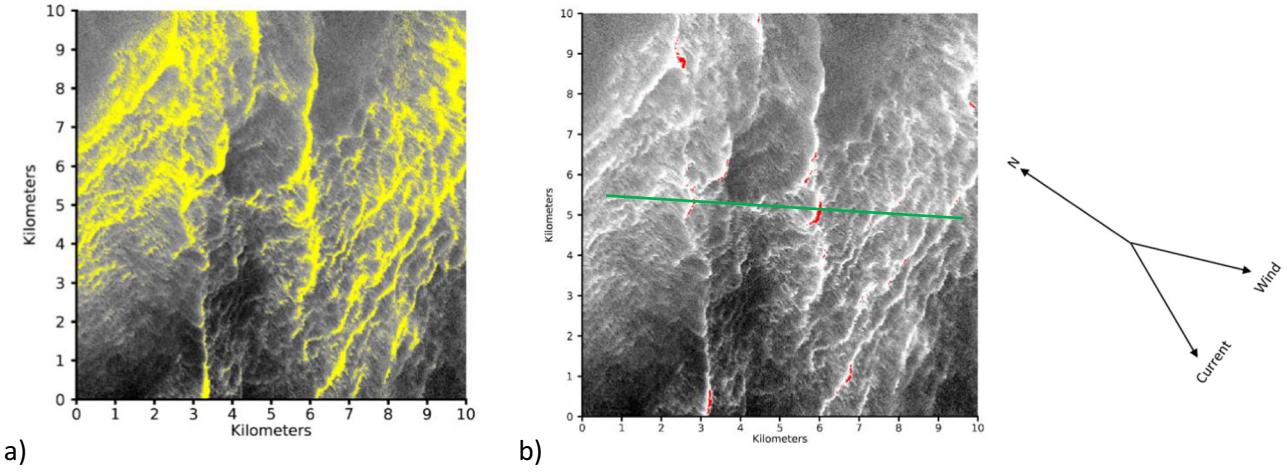

Fig. 4. a) CI image of the algal bloom with yellow representing areas greater than a Medium CI lower threshold. b) CI image of the algal bloom with red representing areas greater than a High CI threshold. The green line represents the transect that is analyzed. It should be noted that the transect is approximately in the direction of the wind.

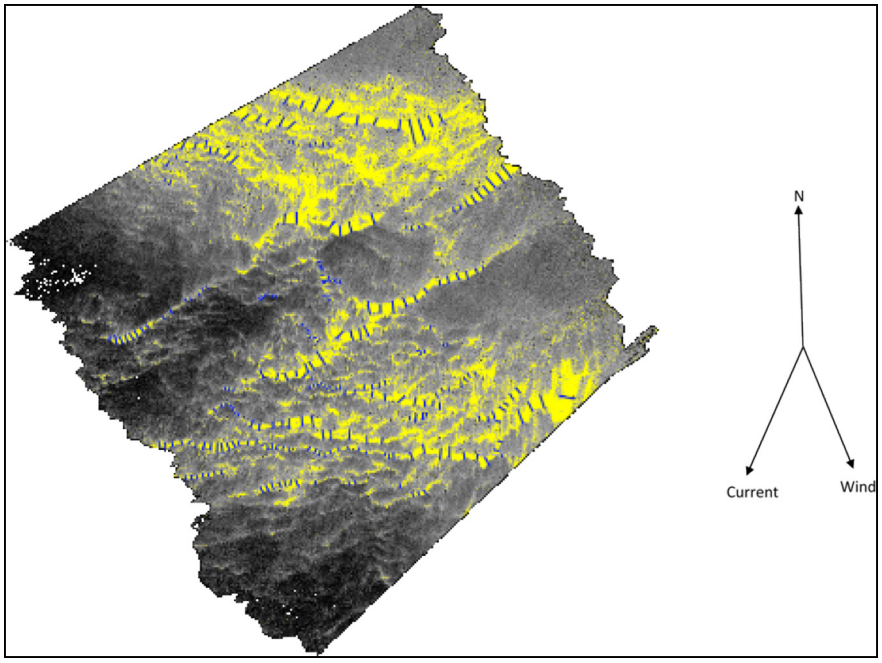

Fig. 5. Image with inspection vector lines (blue hashes) added across the 'Medium' (yellow) concentration features.

inspection. This data is just one line in the image, so it cannot be given too much weight, but it does support the conclusion that the manual inspection histograms shown earlier are likely weighted slightly toward a larger peak feature size. Based on the results shown in the histogram, the peak for the medium features is at 50 m and 75% of the features can be classified as above the 25 meter bin.

From the transect data we can also look at the highest peak, and resample the data to be equivalent to a coarser resolution. This will allow for a comparison between the peak value from native data and what would be measured with lower resolution systems. The largest peak in Fig. 7 (at approximately 5325 m along the transect) was used to make this comparison. The results from this comparison are shown in Table 5. The CI peak value in the measurement was 0.0082 and decreased as the synthetic resolution decreased. In the 25 meter resolution image the CI value was within 8% of what was measured. At 50 meter resolution it was within 22% and at 75 m it is within 35%.

To put this table into context in terms of cell counts we can utilize the relationship of Cyano Abundance to CI to further evaluate the differences (Lunetta et al., 2015; Wynne et al., 2010) shown in this equation

$$Cyano\ Abundance\left(\frac{cells}{mL}\right) = CI * 1.0E + 8.$$

where the CI is showing that the peak cyano abundance in this area is approximately 820,000 cells/mL, if the resolution of the sensor were 100 m the CI would only indicate a peak cyano abundance of approximately 470,000 cells/mL. This is a very significant difference that would simply be a result of the differing resolutions of remote sensing systems.

We can also utilize Table 5 information to estimate an optimal resolution for remote sensing systems that are built for monitoring HABs in the Great Lakes. In keeping with the defined requirement for an

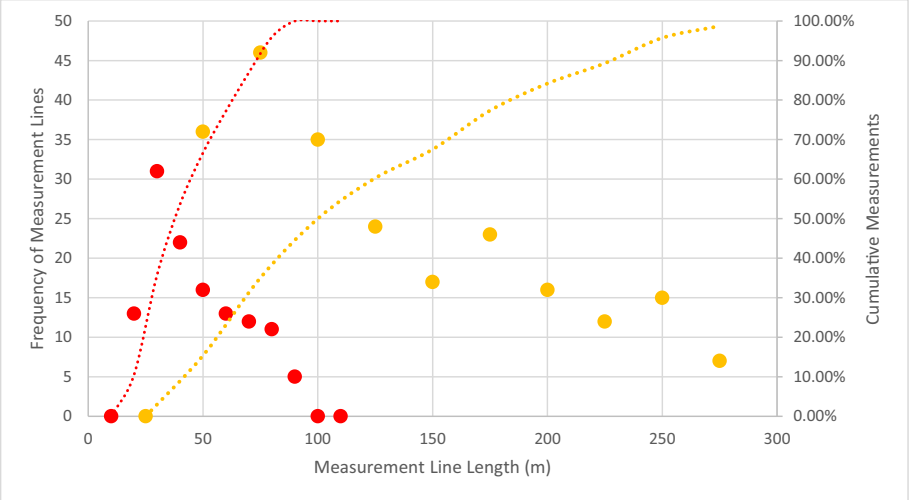

Fig. 6. Plot of the number of lines in a given length bin for the areas above medium concentration (yellow dots) and from inspection of areas above High concentration (red dots) over entire image. The cumulative measurements below a given length are shown by the dotted lines (red for high concentration and yellow for medium).

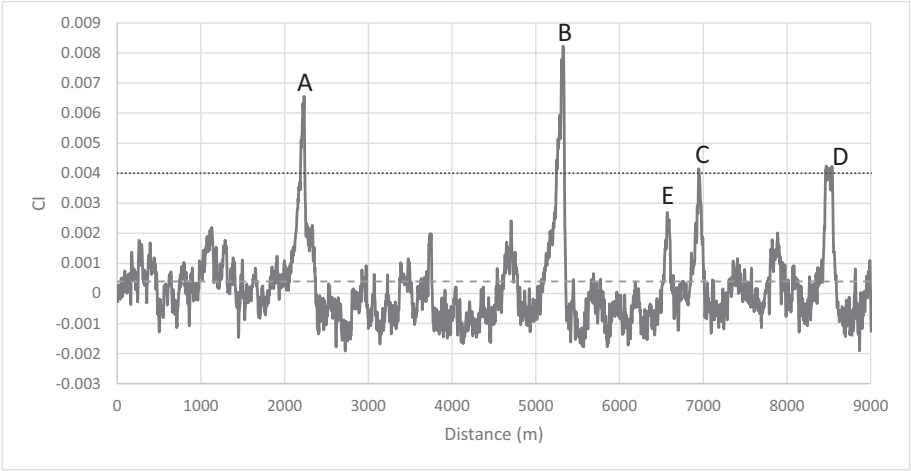

Fig. 7. Plot of the CI transect through the bloom. The transect location is shown as a green line in Fig. 4. The high threshold is shown as a dotted line and the medium threshold is shown as a dashed line. For each of the major peaks the widths for High and Medium concentration are: A) 53 m High, 312 m Medium, B) 94 m High, 256 m Medium, C) 33 m High, 185 m Medium, D) 10 m High, 160 m Medium, E) 90 m Medium.

estimate that is within 25% of what the peak subpixel CI actually is, the resolution would need to be approximately 50 m. In order to get within 8% the resolution would have to be 25 m.

Analysis of in situ data

An on water/in situ transect was likewise obtained by looking at a point measurement within this scene at Toledo light 2 to estimate the variability in the water. From the minute interval time-series, radiometric measurements at Toledo Light #2, patches of high CI were quantified (Fig. 9). CI values were quite variable throughout the time-series with the largest values in the later part of the series after 120 min had passed. High CI values were observed in the 140–180 minute period. These results highlight the highly dynamic nature of cyanobacteria formations on very short time and spatial scales in Lake Erie. Using currents along with the radiometer time-series data, patch sizes were calculated. Using a CI threshold of 0.004, there were 8 high CI patches identified ranging from 1 to 16 min in duration. These patches ranged in size from around 1.3–22.2 m in width, with an average of 5.2 m.

In Fig. 9 we also note that the shape of the high concentration features is asymmetrical. In this figure patches of cyanobacteria are travelling past the light #2, so the frontal edge of each peak is on the left and the trailing edge is on the right. For example, the leading edge of the very high peak in Fig. 9 starts at 118 min (which is the front) and the trailing edge falls off at approximately 129 min. For this peak, the CI at the front rises very fast over 3 min (118 to 121 min) and then falls back to the lower value over approximately 8 min (121 to 129 min).

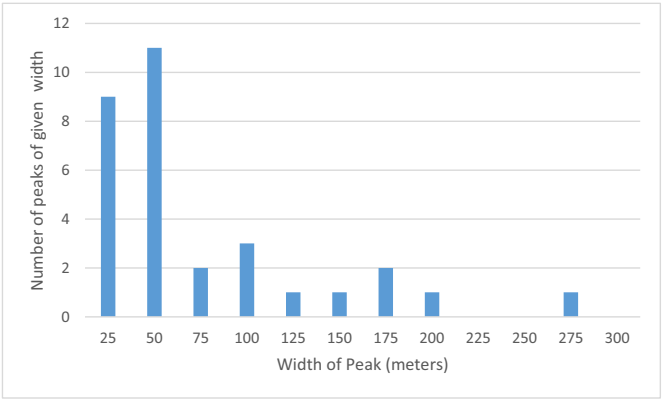

Fig. 8. Histogram of the width of the concentration peaks in the transect shown in (Fig. 7).

Discussion

From the data and results obtained in this work, a range of potential resolutions has been examined. In the synthetic generation of lower resolution CI images of the algal bloom, there was a significant visual reduction in bloom visibility in the 50 m image in comparison with the 100 m image. In the global subpixel standard deviation analysis, the minimum variability was for 26 meter pixels and the variability increased for larger pixel sizes up to a peak at 98 meter pixels. In the manual inspection analysis, we put a threshold value of 75% on the number of size features that the imaging system resolution should be equivalent to. With this threshold, we found that from the manual inspection of the high concentration and medium concentration images, a resolution of 30 m and 75 m would respectively be required.

To confirm that the manual inspection information is reasonable, a single cross-image transect was analyzed. In this transect we found that the high and low concentration peaks were probably shifted too high in the manual inspection. When we performed an analysis of the width of the peaks in the transect, we found many smaller features and can conclude that the manual inspection probably misses many smaller features. With the transect information we find that the 25 meter threshold covers 75% of the features in the transect. This is a small subset of the overall data, but it does support that all of the previous measurements are reasonable.

The in situ water measurements from the light #2 data provided a time/distance series that had many similarities to the data from the

Table 5
Summary of the results of the comparison of the native peak value to those obtained from binning to successively coarser resolutions.

Pixel size	CI peak value	Percent of measured peak value	Percent delta
2.4	0.0082	100	
25	0.0076	92	8
50	0.0064	78	22
75	0.0054	65	35
100	0.0047	57	43
125	0.0041	50	50
150	0.0037	45	55
200	0.0031	37	63
300	0.0027	32	68
400	0.0019	23	77

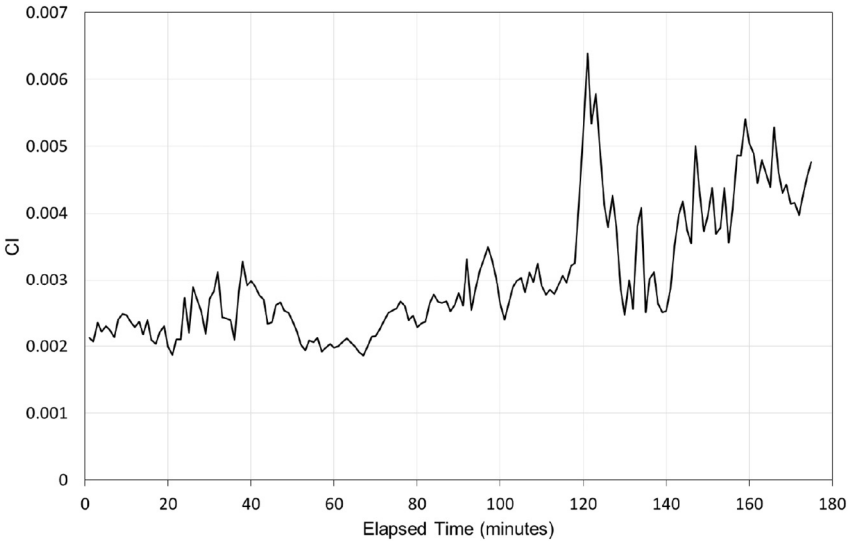

Fig. 9. Toledo Light #2 radiometer derived time-series observations of CI.

transect. From the water measurements, the patches were at most 22 m in width where it was above a CI of 0.004 (high level). This is in the range of what was measured using the remote sensing data but was on the lower end of that range. This is not unexpected as the on-water measurement is only a limited area of the bloom. It is estimated that only about 175 m of the bloom passed by the radiometer during the 175 min of measurements in Fig. 9 while the HSI transect was 9000 m in length. Thus the radiometer observed only 2% of the distance shown by the transect. There was also asymmetry in the direction of travel noted in both. Finally, the CI values from the water measurement were within the range of the transect values. While they were a little higher than the data shown from the HSI, it is difficult to make direct comparisons as the data were taken at a minimum of 1 hour difference in time and also Toledo Light#2 disturbs the flow of cyanobacteria causing build ups on the edges of the light where the sensor is located.

Finally, the potential for significant difference between the value measured and a peak that could be within a pixel at lower resolutions was examined using the transect data. From this we find that there is the potential for a variation within a pixel of up to 25% with a resolution of 50 m. Based upon the preponderance of evidence, we find that the resolution of a remote sensing system for HAB study and monitoring in the great lakes should be at most 50 m in resolution. In addition, we find that there is a significant potential for improvement by reducing the resolution to 25 m, because the potential for error within a pixel drops from 25% to 8%. At the minimum threshold for the High concentration, 500,000 cells/mL, this 8% error would mean a potential of ± 40,000 cells/mL. In the 25% case the error would be 125,000 cells/mL.

Summary

Hyperspectral imagery was utilized to study the spatial variability of harmful algal blooms in the western basin of Lake Erie. To acquire a data set where resolution could be resampled to coarser resolutions, a hyperspectral imager with a wide field of view was flown at an altitude of 24,000 ft to obtain an 11 km by 11 km hyperspectral image of a very well-developed algal bloom on a day with very low wind. This represents some of the most extreme conditions for observing very high concentration changes in very short distances. This data set was analyzed for statistical variability using several corroborating techniques that point to a need for a remote sensing resolution of 25 m in order to limit subpixel variation to within 10% for the algal bloom studied. To further substantiate this finding, an in situ data set was obtained on the same day and in the same location that corroborates the findings from the remote sensing data set.

The airborne hyperspectral data has been shown to be effective to provide multi scale assessments of cyanobacteria spatial distributions in an inland lake. The HSI data obtained showed strong similarities to the on water optical measurements. It has also been shown that the data can be used to approximate multiple satellite spatial resolutions (and spectral bands) to estimate anticipated performance of future remote sensing systems.

Acknowledgements

This research was made possible through funding from the Earth Science Division of the NASA Science Mission Directorate.

References

Anderson, Donald M., Cembella, Allan D., Hallegraeff, Gustaaf M., 2012. Progress in understanding harmful algal blooms (HABs): paradigm shifts and new technologies for research, monitoring and management. Annu. Rev. Mar. Sci. 4, 143–176.

Atkinson, P. M., and Aplin, P. 2004. "Spatial variation in land cover and choice of spatial resolution for remote sensing." International Journal of Remote Sensing 25: 3687–3702.

Butler, Ned, James C. Carlisle, Regina Linville, and Barbara Washburn. 2009. Microcystins: A Brief Overview of Their Toxicity and Effects, With Special Reference to Fish, Wildlife, and Livestock. Office of Environmental Health Hazard Assessment, California Environmental Protection Agency.

Campbell, J.W., Esaias, W.E., 1983. Basis fo spectral curvature algorithms in remote sensing of chlorophyll. Appl. Opt. 22, 1084–1093.

Cushnie, J.L., 2004. The interactive effect of spatial resolution and degree of internal variability within land-cover types on classification accuracies. Int. J. Remote Sens. 8, 15–29.

Gordon, H.R., Wang, M., 1994. Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with SeaWiFS: a preliminary algorithm. Appl. Opt. 33 (3), 443–452.

Great Lakes Fishery Commission. 2014. Great Lakes Fishery Commission Fact Sheet. Accessed December 19, 2018. http://www.glfc.org/pubs/factsheets/FACT%201_14-0912_HR.pdf.

Gronewold, Andrew D., Fortin, Vincent, Lofgren, Brent, Clites, Anne, Stow, Craig A., Quinn, Frank, 2013. Coasts, water levels, and climate change: a Great Lakes perspective. Climate Change 697–711.

Hallegraeff, G.M., 1993. A review of harmful algal blooms and their apparent global increase. Phycologia 32 (2), 79–99.

Hitzfeld, B.C., Hoger, S.J., Dietrich, D.R., 2000. Cyanobacterial toxins: removal during drinking water treatment, and human risk assessment. Environ. Health Perspect. 108, 113–122.

Korpela, Ilkka, 2004. Individual tree measurements by means of digital aerial photogrammetry. Silva Fennica Monographs. 3, 93.

Kutser, T., Metsamaa, L., Strombeck, N., Vahtmae, E., 2006. Monitoring cyanobacterial blooms by satellite remote sensing. Estuar. Coast. Shelf Sci. 67, 303–312.

Lekki, John, Anderson, Robert, Avouris, Dulcinea, Becker, Richard, Churnside, James, Cline, Michael, 2017. Airborne hyperspectral sensing of monitoring harmful algal blooms in the Great Lakes region: system calibration and validation. NASA Technical Memorandum, Glenn Research Center. National Aeronautics and Space Administration, Cleveland, p. 38.

Lunetta, R.S., Schaeffer, B.A., Stumpf, R.P., Keith, D., Jacobs, S.A., Murphy, M.S., 2015. Evaluation of cyanobacteria cell count detection derived from MERIS imagery across the eastern USA. Remote Sens. Environ. 157, 24–34.
Mobley, C.D., 1999. Estimation of the remote-sensing reflectance from above-surface measurements. Appl. Opt. 38 (36), 7442–7455.
Mobley, C. D., J. Werdell, B. Franz, Z. Ahmad, and S. Bailey. 2016. Atmospheric correction for satellite ocean color radiometry. NASA Technical Memorandum, Goddard Space Flight Center, Greenbelt: National Aeronautics and Space Administration.
Murphy, R., Tolhurst, T., Chapman, M., Underwood, A., 2008. Spatial variation of chlorophyll on estuarine mudflats determined by field-based remote sensing. Mar. Ecol. Prog. Ser. 365, 45–55.
Ohio Environmental Protection Agency. n.d. Visual identification of cyanobacteria blooms. Accessed April 30, 2018. http://epa.ohio.gov/portals/28/Documents/HAB/BloomCharacterizationGuide-DRAFT.pdf.
Philpot, W.D., 1991. The derivative ratio algorithm: avoiding atmospheric effects in remote sensing. IEEE Trans. Geosci. Remote Sens. 29, 350–357.
Sinha, E., Michalak, A.M., Balaji, V., 2017. Eutrophication will increase during the 21st century as a result of precipitation changes. Science 357 (6349), 405–408.
Vanderploeg, Hank, 2016. Personal Correspondance. NOAA Great Lakes Environmental Research Lab.
Woodcock, Curtis, Strahler, Alan, 1987. The factor of scale in remote sensing. Remote Sens. Environ. 311–332.
World Health Organization. 2003. Guidelines for Safe Recreational Water Environments Volume 1: Coastal and Fresh Waters. Geneva: World Health Organization.
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 7, 1649–16663.
Wynne, T.T., Stumpf, R.P., Tomlinson, M.C., Warner, R.A., Tester, P.A., Dyble, J., Fahnestiel, G.L., 2008. Relating spectral shape to cyanobacterial blooms in the Laurentian Great Lakes. Int. J. Remote Sens. 29, 3665–3672.
Wynne, Timothy T., Stumpf, Richard P., Tomlinson, Michelle C., Dyble, Julieanne, 2010. Charecterizing a cyanobacterial bloom in western Lake Erie using satellite imagery and meteorological data. Limnol. Oceanogr. 2025–2036.
Wynne, T.T., Stumpf, R.P., Tomlinson, M.C., Fahnenstiel, G.L., Dyble, J., Schwab, D.J., Joshi, S.J., 2013. Evolution of a cyanobacterial bloom forecast system in western Lake Erie: development and initial evaluation. J. Great Lakes Res. 39, 90–99.