Evaluation of cyanobacteria cell count detection derived from MERIS imagery across the eastern USA

Ross S. Lunetta a, Blake A. Schaeffer b,⁎, Richard P. Stumpf c, Darryl Keith d, Scott A. Jacobs e, Mark S. Murphy f

a US EPA National Exposure Research Laboratory, 109 T. W. Alexander Drive, Mail Code: E243-05, Research Triangle Park, NC 27709, United States
b US EPA National Health and Environmental Effects Research Laboratory, Gulf Ecology Division, 1 Sabine Island Drive, Gulf Breeze, FL 32561, United States
c NOAA National Oceans Service, Center for Coastal Monitoring and Assessment, 1305 East West Highway, Silver Spring, MD 20910, United States
d US EPA National Health and Environmental Effects Research Laboratory, Atlantic Ecology Division, 27 Tarzwell Drive, Narragansett, RI 02882, United States
e US EPA National Risk Management Research Laboratory, Land Remediation and Pollution Control Division, 26 Martin Luther King Drive, Cincinnati, OH 45268, United States
f US EPA National Exposure Research Laboratory, 109 T. W. Alexander Drive, Mail Code: E243-05, Research Triangle Park, NC 27709, United States

A R T I C L E   I N F O
Article history:
Received 27 August 2013
Accepted 9 June 2014
Available online 7 July 2014

Keywords:
Cyanobacteria monitoring
Harmful algal blooms
Algorithm validation
MERIS
Sentinel-3
Water quality

A B S T R A C T
Inland waters across the United States (US) are at potential risk for increased outbreaks of toxic cyanobacteria blooms events resulting from elevated water temperatures and extreme hydrologic events attributable to climate change and increased nutrient loadings associated with intensive agricultural practices. Current monitoring efforts are limited in scope due to resource limitations, analytical complexity, and data integration efforts. The goals of this study were to validate an algorithm for satellite imagery that could potentially be used to monitor surface cyanobacteria events in near real-time to provide a compressive monitoring capability for freshwater lakes (>100 ha). The algorithm incorporated narrow spectral bands specific to the European Space Agency’s (ESA’s) MEdium Resolution Imaging Spectrometer (MERIS) instrument that were optimally oriented at phytoplankton pigment absorption features including phycocyanin at 620 nm. A validation of derived cyanobacteria cell counts was performed using available in situ data assembled from existing monitoring programs across eight states in the eastern US over a 39-month period (2009–2012). Results indicated that MERIS provided robust estimates for low (10,000–109,000 cells/mL) and very high (>1,000,000 cells/mL) cell enumeration ranges (approximately 90% and 83%, respectively). However, the results for two intermediate ranges (110,000–299,000 cells/mL and 300,000–1,000,000 cells/mL) were substandard, at approximately 28% and 40%, respectively. The confusion associated with intermediate cyanobacteria cell count ranges was largely attributed to the lack of available taxonomic data and distinction of natural counting units for the in situ measurements that would have facilitated conversions between cell counts and cell volumes. The results of this study document the potential for using MERIS-derived cyanobacteria cell count estimates to monitor freshwater lakes (>100 ha) across the eastern US.

⁎ Corresponding author. Tel.: +1 919 541 5571.
E-mail address: schaeffer.blake@epa.gov (B.A. Schaeffer).

1. Introduction
Cyanobacteria and their cyanotoxins are unregulated contaminants. However, cyanotoxins are included in the US Environmental Protection Agency (USEPA) Safe Drinking Water Act “Contaminant Candidate List” (USEPA, 2013a). Cyanobacteria blooms occur worldwide and are associated with human respiratory irritation, taste and odor of potable water, and human illness as a result of ingestion or skin exposure during recreational activities. The term bloom is defined here as anytime occurrence may result in negative environmental or health consequences (Smayda, 1997). Pets, domestic animals, and wildlife are also affected by exposure to cyanotoxins, with deaths reported annually (Backer, 2002). Cyanotoxins can be found in water bodies used for drinking, aquaculture, crop irrigation, and recreation (Stewart, Webb, Schluter, & Shaw, 2006). It has been hypothesized that cyanotoxins could even be transferred to crops designated for human consumption via spray irrigation (Hunter, Tyler, Carvalho, Codd, & Maberly, 2010). Cyanobacteria blooms can be aesthetically unappealing, which is enhanced by their tendency to concentrate along shorelines where they are encountered frequently by the public (Chorus, Salas, & Bartram, 2000). Increasing frequency, duration and magnitude of blooms within some systems has prompted management actions. Postulated causes of these blooms include excessive nutrient loads, introduction of invasive species (Budd, Drummer, Nalepa, & Fahnenstiel, 2001), and increasing temperatures from climate change and variability, where warmer surface waters favor cyanobacteria growth (Paerl & Huisman, 2008). Alterations in land-cover (e.g., urbanization) and changes in land-use practices, such as intensive agricultural practices and biofuel mandates (e.g., 2007 Energy Independence and Security Act) can
result in watersheds experiencing additional sediment loading and nutrient delivery, which can influence cyanobacteria growth (Lunetta, Shao, Ediriwickrema, & Lyon, 2010; Michalak et al., 2013; USEPA, 2011).

Harmful algal blooms exact a cost in freshwater degradation of approximately $2.2 billion annually in the United States (US) (Dodds et al., 2009). Costs include loss of recreational water usage, declines in waterfront real estate value, spending on recovery of biodiversity, and drinking water treatment. The greatest economic losses were in the decline of real estate value and income resulting from recreational use. Accordingly, there is a critical need to determine the actual underlying causes of these blooms. Despite ecosystem, economic, and public health concerns, cyanotoxins are typically assessed infrequently due to the cost, required expertise, and time requirements for complex analysis. Likewise, measuring associated water quality parameters is time and cost intensive, often requiring a field team to spend a majority of the day on a boat collecting samples. Data from discrete stations typically provide incomplete spatial and temporal coverage of a water body. While field sampling is certainly necessary and will continue to be so, managers need other options to improve monitoring, reduce costs, and minimize potential exposures associated with sample collection.

The incidence of cyanobacteria blooms has been increasing in the US (Hudnell, 2010) and extends across varied habitats from smaller eutrophic prairie ponds to larger more oligotrophic regions of the Laurentian Great Lakes (Fristachi & Sinclair, 2008). Microcystin concentrations are moving northward along a gradient of increasing lake trophic status across the Midwest (Graham, Jones, Jones, Downing, & Clevenger, 2004). Stumpf, Wynne, Baker, and Fahrenstiell (2012) determined that in 2011 Lake Erie experienced a cyanobacteria bloom event three times greater than any previously observed event as a result of unusual high runoff. Michalak et al. (2013) further link the outbreak to increased phosphorous loading associated with long-term agricultural land-use practices (expanded corn production) coupled with meteorological conditions in spring of 2011 that were consistent with climate trends predicted for future conditions. Climate change is a potential catalyst for further bloom development as cyanobacteria often out compete other phytoplankton taxa at higher temperatures (Paerl & Huisman, 2008). However, there are no comprehensive data sources documenting the occurrences of cyanobacteria outbreaks across the US (Hudnell, 2010).

1.1. Cyanobacteria monitoring

Cyanobacteria blooms have been documented across the US, and numerous states have issued health advisories or closed recreational areas due to potential risks from exposure. Cyanobacteria and their toxins are addressed differently by each state. A few states have toxin monitoring programs in place, while others only conduct event-based responses, and some provide public education focused on human and animal protection from toxin exposure (Graham, Loftin, & Kamman, 2009). Typically, monitoring programs include visual assessments and water sampling. Many use the World Health Organization (WHO) guidelines, with a three-level approach using chlorophyll-a (Chl-a), cell counts, and microcystin-LR measurements during recreational activities (Chorus & Bartram, 1999). It is important to emphasize that the cell count range categories used by the WHO are based on “relative probabilities of acute health effects” and do not directly correspond to the ranges used in this study. In addition, taxonomy is typically not a standard part of these measurements. This allows for a series of alert levels where various data sources are the foundation of advisory postings or water body closings. Other states, such as Oklahoma and Massachusetts, have developed state-specific guidelines to establish protective levels. Oklahoma has legislation limiting exposure to freshwater algae, where warnings are issued to lake users if algae cell counts exceed 100,000 cells/mL and microcystin concentrations exceed 20 μg/L. Massachusetts has established guidelines for issuing an advisory against contact with water when counts exceed 70,000 cells/mL or microcystin concentrations exceed 14 μg/L.

States have experienced many challenges in the development of monitoring programs for cyanobacteria toxins. Field based monitoring programs may be insufficient to provide timely warnings of cyanobacteria bloom development across large geographic areas because estimating cyanobacteria concentrations is a labor intensive and time consuming endeavor, involving water sample collections, laboratory analysis, and the visual identification and enumeration of phytoplankton species. Further, the recognition of cyanobacteria blooms in the field are typically limited to visual assessments of water discoloration or the identification of floating scums and mats, which are patchy and transient in space and time. Finally, there are difficulties in developing appropriate sampling designs that address large areas using fewer resources or affordable detection methods; and development of guidance based on taxonomic identification and/or cell counts can be overly conservative in protecting public health compared to states that identify and measure toxins. In light of these challenges, new rapid tools are needed to assist states in development of efficient and effective cyanobacteria monitoring programs. Satellite remote sensing provides an opportunity for such a tool. Although this technology is not capable of detecting cyanobacteria toxins based on optical properties of the water column, it can be used to derive cyanobacteria concentration information.

1.2. Remote sensing approaches

Studies have shown that satellite data can detect and quantify cyanobacteria blooms in lakes and estuaries (Kahru, Savchuk, & Elmqren, 2007; Simis, Peters, & Gons, 2005; Alikas, Kangro, & Reinart, 2010; Wynne, Stumpf, Tomlinson, & Dybleh, 2010; Gómez, Alonso, & García, 2011; Matthews, Bernard, & Robertson, 2012; Duan, Ma, & Hu, 2012; see also review by Kutser, 2009). These include a variety of methods and approaches including identifying blooms and scums (Kahru et al., 2007), estimates of phycocyanin (Ruiz-Verdú, Simis, Hoyos, Gons, & Peña-Martínez, 2008) or the presence of phycocyanin (Matthews et al., 2012; Simis et al., 2005), biomass as chlorophyll-a (Matthews et al., 2012), and cell count concentrations (Hunter et al., 2010; Wynne et al., 2010). Various satellites have been examined such as the OCM (Dash et al., 2011), Landsat (Vincent, Qin, Mckay, & Miner, 2004), AVHRR (Kahru et al., 2007), MODIS (Becker, Sultan, Boyer, Twiss, & Konopko, 2009; Wynne et al., 2013), and MERIS (Binding, Greenberg, Jerome, Bukata, & Letourneau, 2011; Matthews, Bernard, & Winter, 2010; Ruiz-Verdú et al., 2008; Wynne et al., 2008). In the US, the longest running monitoring program for cyanobacteria is for Lake Erie. The standard algorithm for that program is the “cyanobacterial index” (CI) (Wynne et al., 2008), which uses a second derivative centered on the MERIS fluorescence band at 681 nm or the MODIS 678 nm band (Wynne et al., 2010; Wynne et al., 2013; Wynne et al., 2008). The second derivative functions with minimal or poor atmospheric correction (Philpot, 1991), eliminating a significant challenge to near daily application in highly turbid water. In fact, several applications of second derivative functions do not use any atmospheric correction (Gower, King, & Goncalves, 2008; Hu, 2009). The resulting “Lake Erie HAB Bulletin” product is used by hundreds of people including many state and city managers (Wynne et al., 2013). With the demonstrated application of this analysis, and need for monitoring in many lakes in the US, an evaluation of its ability to detect cyanobacteria blooms in other lakes is warranted. It is acknowledged there are issues in the comparison of two measured quantities with no direct overlap; optical signals of light absorption and scattering by pigments in cells, and cell counts without biovolume or biomass information. The ratio of pigmentation and biomass is not constant, and neither is the ratio of cell number to biomass. The study objective was to evaluate whether a cyanobacteria algorithm based on Wynne et al. (2008) can be applied.
to MERIS data to identify surface cyanobacteria bloom events in freshwater lake systems for applications throughout the eastern US.

Potential benefits associated with remote sensing tools include enabling policymakers and environmental managers to assess the sustainability of watershed ecosystems, and the services they provide, under current and future land-use practices. The most direct way to ensure that management practices are achieving sustainability is to monitor the environment on a synoptic scale. Sustainable practices have a significant impact when widely implemented and relevant at a national scale. Satellite technology could allow for the development of cyanobacteria early warning indicators. Rapid detection of potential blooms is essential for protecting the general population from exposure, including magnitude and duration of contact to the stressor. The National Research Council report on “Exposure Science in the 21st Century” suggested that assessing and mitigating toxic exposures effectively requires rapid measurement of the stressor on diverse geographic, temporal, and biological scales and an enhanced infrastructure for rapid deployment of resources to address imminent threats (NRC, 2012). Predictive tools allow forecasting, prevention, and mitigation of the potential effects. The NRC report specifically calls for innovative and expedient assessment approaches that strategically use diverse information, such as satellite remote sensing, for the identification and quantification

---

**Study Areas and Measurement Locations**

**Fig. 1.** Study area location map indicating the field sampling locations and MERIS scene boundaries.
of relevant exposures that may pose a threat to ecosystems or human health.

2. Methods

2.1. Study site description

In situ measurement data were assembled for lakes in eight states across New England, Ohio, and Florida (Fig. 1) to support the evaluation. MERIS high resolution (300 m) imagery acquisitions were matched with available field measurement data to support the validation of the cyanobacteria algorithm to assess the potential for regional scale performance across the eastern US.

The New England lakes used in this study were part of the Northern Appalachians ecoregion. There were 5226 lakes in the Northern Appalachians included in the National Lakes Assessment Program (NLAP) inventory, where 54% were man-made reservoirs. Based on the 1992 National Land Cover Dataset, the distribution of land cover was 69% forested, 17% planted/cultivated, and 14% characterized as other types. Agricultural and urban runoff from watersheds or drainage basins were the primary source of excess nitrogen and phosphorus to New England Lakes which exacerbated algae blooms. Lake Champlain, VT was the largest lake in the Northern Appalachians ecoregion and one of the largest in the United States. While primarily a recreational lake, it also served as a source of drinking water and a site for the disposal of municipal wastes (USEPA, 2009). In 2009, monitoring results indicated that toxic cyanobacteria blooms continued to be prevalent in two primary areas in the lake (Missisquoi Bay and St. Albans Bay) and at several locations in the northern region of the lake. Transient algal accumulations were found at scattered sites throughout the lake. During 2008–2009, microcystin concentrations ranged from 0.01 to > 94 μg/L, with highest concentrations in Missisquoi Bay (Watzin, Fuller, Bronson, Gorney, & Schuster, 2011).

Several ecoregions intersect in Ohio including the Eastern Corn Belt Plains, the Western Allegheny Plateau, and the Erie Drift Plain. The trophic state of these lakes, as based on chlorophyll-a concentrations, was 68% hypereutrophic or eutrophic. Averaging the percentage for all basins sampled in Ohio for major land use types indicates that 22% of land was developed, 27% was forested upland, and 42% was planted/cultivated. Impacts to surface water in this region were primarily from agricultural and urban sources. Grand Lake St. Mary's in northwestern Ohio was a man-made reservoir that has been plagued in recent years with persistent harmful algal blooms. The reservoir was a source for drinking water with 66% agriculture land use designation, which resulted in high concentrations of nutrients in the tributaries that feed into the lake. Ohio has reported high levels of microcystin concentrations (sometimes exceeding 100 ppb) in Grand Lake St. Mary's that have remained high into the late fall and early winter (USEPA, 2009).

The State of Florida was located in the Coastal Plains ecoregion, of which 69% of lakes were constructed reservoirs within the ecoregion. Based on the 1992 National Land Cover, the ecoregion land cover was 39% forested, 30% planted/cultivated, and 16% wetlands with the remaining 15% of land in other types of cover (USEPA, 2009). Anthropogenic nutrient sources included agriculture, phosphate mining and point sources and have been suggested to increase the occurrence of cyanobacteria from historical levels in central Florida lakes (Paerl, 2008; Riedinger-Whitmore et al., 2005).

The temporal and spatial stability of cyanobacteria blooms in a particular ecosystem depends on the extent that different environmental factors influence bloom dynamics. In addition to nutrient supply, these factors include wind velocity, rainfall amount and intensity, and the solar radiation (Soranno, 1997). Because these factors vary widely, cyanobacterial blooms display a range of temporal dynamics. Some freshwater bodies have seasonal blooms that begin during the summer and last into fall, some have persistent blooms that cover all four seasons while others have blooms that occur just for days or weeks. The NLAP assessment data was based on a probabilistic sampling scheme and did not include methods for collecting data on those factors favorable to bloom development and sustainability (such as wind velocity, solar radiation, and rainfall), therefore no assessment was conducted on the temporal and spatial stability of cyanobacteria blooms from the these datasets.

2.2. Cyanobacteria algorithm

Data were processed to Rayleigh-corrected reflectance (R) using NASA’s SeaWiFS Data Analysis System (SeaDAS) standard l2gen software (“rhou” dimensionless product) and mapped to standard UTM projection with 300 m pixels using nearest neighbor resampling. The “rhou” product includes correction for sun angle, but does not involve a correction for aerosols. Each MERIS image was masked for land with the standard 250 m landmask distributed with SeaDAS and for clouds using the cloud albedo, excluding bright reflectance water that has strong spectral variation (high reflectance in the near-infrared and green and low in the red and blue) to avoid masking intense blooms that would otherwise be flagged as clouds. Clouds were flagged using threshold algorithms, corrected for turbid water. The cyanobacteria detection algorithm used second derivative spectral shapes (SS) on the reflectance spectra with an equation:

\[
SS(\lambda) = R(\lambda) - R(\lambda^-) = \left\{ R(\lambda^-) - R(\lambda^+) \right\} \times \frac{(\lambda - \lambda^-)}{(\lambda - \lambda^+)}. \tag{1}
\]

Phlipot (1991) demonstrated that this form was insensitive to poor atmospheric correction. Wynne et al. (2008) used the spectral shape (SS) around the 681 nm, such that \( \lambda = 681 \) nm, \( \lambda^+ = 709 \) nm, and \( \lambda^- = 665 \) nm. They derived a Cyanobacteria Index (CI) = −SS(681). SS(681) is centered on the 681 nm band, which covers peaks in both chlorophyll absorption and chlorophyll fluorescence. In clearer water with non-cyanobacterial blooms, a fluorescence peak is observed at 681 nm. When there is negligible fluorescence, as is the case in cyanobacteria blooms (Seppälä et al., 2007; Wynne et al., 2008), there is no fluorescence peak. Instead the strong chlorophyll absorption leads to a dip or “valley” at 681 nm resulting in negative SS(681) (Binding et al., 2011). As CI = −SS(681), increased chlorophyll absorption leads to increased CI. The CI spectral shape method has proved effective in Lake Erie, where the summer blooms are primarily composed of Microcystis spp. and provided a robust estimate of cyanobacteria cell counts (Wynne et al., 2010). The CI can flag other blooms including chlorophytes (Wynne et al., 2013) and some dense dinoflagellate blooms (Stumpf, unpublished data) causing false positives. To address the false positive issue, a test had been examined to separate cyanobacteria from other blooms using a derivative that includes the 620 nm band, which is sensitive to phycocyanin, namely the SS(665) with \( \lambda = 665 \) nm, \( \lambda^+ = 681 \) nm, and \( \lambda^- = 620 \) nm. This condition has been applied by Matthews et al. (2012, Eqs. 3–4) to separate cyanobacteria from other blooms in African lakes. The resultant, multiple shape algorithm, termed the CI-multi uses the CI for the biomass estimate, but uses the spectral shape around the 665 nm band as an exclusion criterion. Elevated phycocyanin absorption is presumed to depress the reflectance at 620 nm (e.g., Simis et al., 2005), causing SS(665) to change from negative to positive. Accordingly, when the CI value SS(665) < 0, cyanobacteria are presumed absent, when SS(665) > 0, cyanobacteria are presumed present.

2.3. Algorithm validation

The cyanobacteria cell count reference data used to support MERIS-derived estimates came from several monitoring programs in Ohio, Florida, and six New England states. Participating agencies included the Ohio EPA (Ohio EPA, 2012), St. Johns River Water Management
District (Abbott et al., 2009), and the US EPA’s National Lakes Assessment Program (NLAP, USEPA, 2007a; USEPA, 2007b). The New England data sets also included data collected from EPA’s Region I Office (USEPA, 2009), Rhode Island Department of Environmental Management (ESS Group, Inc., 2011), the Massachusetts Departments of Environmental Protection and Public Health (Beskenis, 2014), the New Hampshire Department of Environmental Services and the University of New Hampshire (NH Department of Environmental Services, 2005a,b), Vermont Department of Environmental Conservation (VT Department of Environmental Conservation, 2014), and the Lake Chaplain Basin Program (Watzin et al., 2011). Databases provided to the USEPA included site identification numbers, collection dates, geographic coordinates, and cyanobacteria cell count numbers, which ranged from zero to approximately 2.0E + 7 (Fig. 2). All 10 data sources used different methodologies for field sampling and cell count enumerations. The protocols for the NLAP are described here, since it was the largest and most comprehensive database with sampled points throughout the US. Details regarding the other sampling programs were found in the associated references mentioned previously. NLAP water samples were collected with an integrated sampler from the upper two meters, or less, of the water column. The sampler was rinsed in triplicate by submerging it vertically into the water column. Phytoplankton samples were preserved with Lugol’s solution (USEPA, 2007a). Cell counts included 300 natural counts using a magnification of 1000 x or higher. Natural counting units were defined as one for each colony, filament, or cell, regardless if colonial, filamentous, or unicellular (USEPA, 2007b). Picoplankton were not counted throughout the sample set, in addition chlorophyll-a and phycocyanin fluorescence was lost after preservation with Lugol’s, limiting confirmation of picoplankton counts. Only the Ohio database included cyanobacteria taxonomic information, but none of these data were paired with MERIS estimates. Of the available 6014 in situ samples, 3946 (66%) were excluded from the analysis due to missing data values or sample collection from > 2.0 m depth. Samples with missing data included 41 samples (12%) in Florida, 3 samples (4.0%) in Ohio, and 238 samples (4%) in New England flagged with no data. Additionally, 3664 samples (69%) in New England were flagged as out of depth range (>2.0 m), leaving a balance of 2068 samples to support algorithm validation. The majority of samples only contained a single surface measurement within <2.0 m depth. For a consistent approach across all monitoring programs, it was decided only surface (<2.0 m) would be used for this effort.

Cyanobacteria cell count ranges (cells/mL) were classified into Low (10,000–109,999), Medium (110,000–299,999), High (300,000–1,000,000), and Very High (>1,000,000) values (Table 1). The count ranges were established as a compromise to correspond with ranges of concern to protect ecosystem services and public health, and to best fit observed data distributions associated with cyanobacteria occurrences in the eastern US. Fig. 2 illustrates the frequency distribution of the in situ cell counts data for values > 10,000 cells/mL (n = 939) and all data values (n = 2068), respectively.

A total of 1275 MERIS images were acquired for three study areas including six New England states, Ohio, and Florida over the 39-month study period beginning in February 2009 through April 2012 (Fig. 1). The MERIS CI-multi algorithm outputs were converted to cyanobacteria cell count values ranging from approximately 10,000–3.0 M (cells/mL). MERIS data were converted to cyanobacteria cell count values using the following equation (Wynne et al., 2010):

\[
\text{Cyano Abundance (cells/mL)} = \text{CI-multi} \times 1.0E + 8
\]  

The MERIS-derived cell count values were then matched with approximately 2068 independent in situ measurements taken from lakes across the study area. Out of the 2068 available samples only 579 (28%) were matched with MERIS-derived cyanobacteria pixels using the ± 7-days window with the majority (82%) corresponding to “Low” cell count range. Additionally, clouds and land excluded 17% and 34% of the available data matches, respectively. Virtually all of the MERIS correspondence pixels classified as land were located in Florida and New England, where the sampling designs did not incorporate fixed locations, thus a high potential for erroneous coordinates. Visual analysis of a subset of land pixels demonstrated that their occurrence was closely tied to shorelines and very small lakes, both of which would be difficult for MERIS to resolve.

Three separate analyses were conducted to pair the observed measurement data with the MERIS predicted values including: (i) 300 m direct correspondence (single pixel); (ii) 900 m average correspondence (3 x 3 pixels); and (iii) 900 m closest value correspondence (450 m buffer). These three analyses were performed to maximize the number of match-ups between available in-situ data and MERIS derived cyanobacteria cell counts for algorithm evaluation. All correspondence analyses were performed for < 1-day, ± 1-day, ± 3-day, ± 5-day, and ± 7-day temporal windows. The 300 m analysis was accomplished using a simple point-on-pixel overlay operation between in situ measurements and the corresponding 300 m MERIS pixel. A semi-automated GIS routine was used to assign the values from corresponding MERIS pixels within the various temporal windows to each in situ measurement. This process resulted in the addition of as many as 15 fields (± 7-day) to derive the value of the corresponding MERIS pixel on each available date within a given temporal window. Output from this process was merged.

![Fig. 2. Histogram illustrating the distribution of in situ cell count values for > 10,000 cells/mL (n = 939) for cyanobacteria measurement data from lakes in Ohio, Florida, and New England (2009–2012). Gray lines indicate break points.](image-url)
Corresponding values were calculated using normalized matrix values to mitigate disproportionate cell count values. The Kappa value was calculated using a normalized matrix due to the disproportional number of samples (n = 1883) in the “Low” cell count category to provide the most representative coefficient value. The 900 m average correspondence analysis incorporated MERIS cell count values and followed the same 300 m analytical procedures, except the 900 m average correspondence analysis was performed at multiple spatial scales (300 × 300 m and 900 × 900 m) using ± 1-day, ± 3-day, ± 5-day and ± 7-day temporal windows. The 900 m analysis included both an average MERIS-derived value (a) and closest correspondence pixel value (b). Kappa coefficient values were calculated based on a normalized matrix using the Margit and Kappa programs (Congalton, 1991). The normalized matrix was calculated using Table 2. The Kappa value was calculated using a normalized matrix due to the disproportional number of samples (n = 1883) in the “Low” cell count category to provide the most representative coefficient value. The 900 m average correspondence analysis incorporated MERIS cell count values and followed the same 300 m analytical procedures, except the MERIS imagery was averaged to a 900 m cell prior to performing the point-on-pixel overlay. The pixels were aggregated to 900 m using a single geospatial layer and transposed to produce a unique record for each date match between the in situ and MERIS data. Records containing no data, land, or cloud values were removed from the database.

Geospatial output was exported to Microsoft Access, where the multi-temporal tables (e.g., ± 1-day, ± 3-day, etc.) were created and the three study locations merged. Records were then classified into bins based on in situ measurements representing Low to Very High cell count ranges (Table 1). Values < 10,000 were classified as Not Detectable because they were classified as absent in the MERIS products and were not included in the subsequent validation process. The in situ measurements were then plotted versus MERIS-derived values (cells/mL) and data were analyzed using correspondence tables, histograms, and confusion matrices. Kappa coefficient values were calculated based on a normalized matrix using the Margit and Kappa programs (Congalton, 1991). The normalized matrix was calculated using Table 2. The Kappa value was calculated using a normalized matrix due to the disproportional number of samples (n = 1883) in the “Low” cell count category to provide the most representative coefficient value. The 900 m average correspondence analysis incorporated MERIS cell count values and followed the same 300 m analytical procedures, except the MERIS imagery was averaged to a 900 m cell prior to performing the point-on-pixel overlay. The pixels were aggregated to 900 m using a single geospatial layer and transposed to produce a unique record for each date match between the in situ and MERIS data. Records containing no data, land, or cloud values were removed from the database.

### Table 1

<table>
<thead>
<tr>
<th>Density range (cells/mL)</th>
<th>Correspondence (%)</th>
<th>Weighted average (%)</th>
<th>Total weighted correspondence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>300 m</td>
<td>900 m²</td>
<td>900 m³</td>
</tr>
<tr>
<td></td>
<td>300 m</td>
<td>900 m²</td>
<td>900 m³</td>
</tr>
<tr>
<td>≤ 1-day</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not detectable</td>
<td>(&lt;10,000)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Low</td>
<td>(10,000–109,999)</td>
<td>76 83 90</td>
<td>98 109 118</td>
</tr>
<tr>
<td>Medium</td>
<td>(110,000–299,999)</td>
<td>19 30 15</td>
<td>21 34 36</td>
</tr>
<tr>
<td>High</td>
<td>(300,000–1,000,000)</td>
<td>50 43 29</td>
<td>38 56 65</td>
</tr>
<tr>
<td>Very high</td>
<td>(&gt;1,000,000)</td>
<td>50 50 50</td>
<td>50 50 50</td>
</tr>
<tr>
<td>Total weighted correspondance</td>
<td>67 74 76</td>
<td>(73) 114 151 172</td>
<td></td>
</tr>
<tr>
<td>± 1-days</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not detectable</td>
<td>(&lt;10,000)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Low</td>
<td>(10,000–109,999)</td>
<td>75 85 90</td>
<td>85 109 118</td>
</tr>
<tr>
<td>Medium</td>
<td>(110,000–299,999)</td>
<td>19 21 18</td>
<td>19 34 36</td>
</tr>
<tr>
<td>High</td>
<td>(300,000–1,000,000)</td>
<td>63 35 27</td>
<td>36 56 65</td>
</tr>
<tr>
<td>Very high</td>
<td>(&gt;1,000,000)</td>
<td>79 64 73</td>
<td>72 90 94</td>
</tr>
<tr>
<td>Total weighted correspondance</td>
<td>68 74 76</td>
<td>(73) 276 396 466</td>
<td></td>
</tr>
<tr>
<td>± 3-days</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not detectable</td>
<td>(&lt;10,000)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Low</td>
<td>(10,000–109,999)</td>
<td>77 84 90</td>
<td>85 109 118</td>
</tr>
<tr>
<td>Medium</td>
<td>(110,000–299,999)</td>
<td>20 17 19</td>
<td>19 34 36</td>
</tr>
<tr>
<td>High</td>
<td>(300,000–1,000,000)</td>
<td>54 30 28</td>
<td>34 56 65</td>
</tr>
<tr>
<td>Very high</td>
<td>(&gt;1,000,000)</td>
<td>76 66 86</td>
<td>76 90 94</td>
</tr>
<tr>
<td>Total weighted correspondance</td>
<td>69 73 77</td>
<td>(73) 747 957 1187</td>
<td></td>
</tr>
<tr>
<td>± 5-days</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not detectable</td>
<td>(&lt;10,000)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Low</td>
<td>(10,000–109,999)</td>
<td>79 84 90</td>
<td>85 109 118</td>
</tr>
<tr>
<td>Medium</td>
<td>(110,000–299,999)</td>
<td>17 17 17</td>
<td>17 34 36</td>
</tr>
<tr>
<td>High</td>
<td>(300,000–1,000,000)</td>
<td>59 34 33</td>
<td>39 56 65</td>
</tr>
<tr>
<td>Very high</td>
<td>(&gt;1,000,000)</td>
<td>77 67 83</td>
<td>75 90 94</td>
</tr>
<tr>
<td>Total weighted correspondance</td>
<td>70 73 77</td>
<td>(74) 1221 1517 1907</td>
<td></td>
</tr>
<tr>
<td>± 7-days</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not detectable</td>
<td>(&lt;10,000)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Low</td>
<td>(10,000–109,999)</td>
<td>80 84 90</td>
<td>85 109 118</td>
</tr>
<tr>
<td>Medium</td>
<td>(110,000–299,999)</td>
<td>16 17 17</td>
<td>17 34 36</td>
</tr>
<tr>
<td>High</td>
<td>(300,000–1,000,000)</td>
<td>58 35 35</td>
<td>40 56 65</td>
</tr>
<tr>
<td>Very high</td>
<td>(&gt;1,000,000)</td>
<td>75 65 83</td>
<td>74 90 94</td>
</tr>
<tr>
<td>Total weighted correspondance</td>
<td>70 72 77</td>
<td>(74) 1623 2010 2536</td>
<td></td>
</tr>
</tbody>
</table>

### Table 2

<table>
<thead>
<tr>
<th>Field measurement data (observed)</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
<th>Very high</th>
<th>Row totals</th>
<th>% Correct</th>
<th>% Omission</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>1701</td>
<td>37</td>
<td>116</td>
<td>29</td>
<td>1883</td>
<td>90</td>
<td>10</td>
</tr>
<tr>
<td>Medium</td>
<td>239</td>
<td>59</td>
<td>43</td>
<td>6</td>
<td>1347</td>
<td>17</td>
<td>83</td>
</tr>
<tr>
<td>High</td>
<td>63</td>
<td>3</td>
<td>39</td>
<td>8</td>
<td>113</td>
<td>35</td>
<td>65</td>
</tr>
<tr>
<td>Very high</td>
<td>27</td>
<td>0</td>
<td>5</td>
<td>161</td>
<td>193</td>
<td>83</td>
<td>17</td>
</tr>
<tr>
<td>Column totals</td>
<td>2030</td>
<td>99</td>
<td>203</td>
<td>204</td>
<td>2536</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Correct</td>
<td>84</td>
<td>60</td>
<td>19</td>
<td>79</td>
<td>77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Commission</td>
<td>16</td>
<td>40</td>
<td>81</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Kappa = 0.57
mean statistical function that ignored ‘no data’ values in the calculation. This process ensured that no masked pixels were included in the averaging of MERIS pixels. Therefore, samples that fell on land due to imprecise coordinates from in situ records would not be incorporated in the 900 m average correspondence analysis, because there was no pixel there for the sample to intersect.

The 900 m closest correspondence analysis differs from the previous two analyses because it was not reliant on direct spatial correspondence with MERIS pixels. This vector-based approach first converted MERIS imagery to points, and then selected the points that fell within a 450 m buffer radius of each in situ measurement location. This was accomplished by applying a 450 m one-to-many spatial join to produce a unique record across multiple temporal windows. The 900 m closest correspondence analysis was used to search around an in situ measurement within the 450 m buffer to find the MERIS value that was closest to the in situ cell count, or with an intersecting pixel within the same cell count category if possible. If none of the pixels within the 450 m buffer were within the same cell count category as the in situ sample, then the search would use the closest value pixel within the 450 m buffer. These data were then analyzed statistically in the same manner as the previous two analyses.

3. Results

Example spectra of MERIS ρhō_s (Fig. 4) were representative of not-detectable, Medium, High, and Very High count ranges where the CI used 665, 681, and 709 nm wavebands. The 760 nm waveband was removed because it occurred at the O₂ absorption for atmospheric correction. Calculated CI for not-detectable, Medium, and Very High count ranges were — 0.0008, 0.0013, and 0.0126 respectively. The CI for the High spectra with 306,000 cells/mL was 0.0030 and the CI for the High spectra with 510,000 cells/mL was 0.0051. The ρhō_s corrects only for atmospheric Rayleigh radiance, so that the ρhō_s reflectance are higher than expected for the observed water type. Standard atmospheric corrections for turbid water tends to overcorrect the blue bands, leading to frequent negative reflectance in blooms (Wynne et al., 2010). The use of the spectral shape on the red and NIR ρhō_s avoids errors caused by atmospheric correction (Gower et al., 2008; Hu, 2009; Philpot, 1991; Stumpf & Werdell, 2010).

There were 1960 paired observations above the not-detectable limit (10,000 cells/mL) from in situ measures and the MERIS imagery between February 2009 and April 2012. Paired observations included any match-up within ±/−7 days of the satellite overpass. There was a significant positive correlation between MERIS cyanobacteria cell counts and in situ cyanobacteria cell counts (Fig. 5, slope = 0.94, R² = 0.87, p < 0.0001, RMSE = 225,369 cells/mL). Due to the high RMSE additional correspondence comparisons were undertaken.

Table 1 summarizes validation results across the eastern US for inland lakes located in New England (RI, MA, NH, ME, NY, and VT), Florida, and Ohio (Fig. 1). Statistics for lake color, turbidity, and chlorophyll-a concentrations were summarized to demonstrate the range of optically active constituents across lakes (Table 3). Correspondence comparisons (MERIS-derived estimates vs. in situ measurement data) were performed for multiple temporal windows including ≤1-day, and ±1, 3, 5, and 7 days. With the exception of ≤1-day, all temporal windows contained a sufficient number of samples to support a robust data evaluation. For the ≤1-day window, sample sizes for the Medium, High, and Very High cell count ranges were insufficient for meaningful analysis. Similarly, the ±1-day window had too small a sample size for the High cell count range. All other temporal ranges had sufficient sampling sizes to support a correspondence based assessment of performance. The correspondence values for ±1, ±3, ±5, and ±7 days, were very similar. Both the Low and Very High data range values had the highest

![Fig. 3](image3.png) Histogram illustrating the distribution of all available in situ data values (n = 2068) for the five cyanobacteria cell count classifications from lakes in Ohio, Florida, and New England for 2009–2012. Includes all available data values for sampling depths ≤2.0 m (blank and zero data values omitted). To clarify, the ‘Not Detectable’ classification refers to the MERIS algorithm result.

![Fig. 4](image4.png) Example MERIS spectra of not-detectable, Medium, High, and Very High count ranges. The top two spectra were taken from a different scene with more aerosol than the scene of the other spectra. The two spectra with wider lines have essentially the same CI value.

![Fig. 5](image5.png) Regression of measured vs. MERIS estimated cell counts for all data with match-up to MERIS within ±/−7 days.
correspondence (90% and 86%, respectively) using the closest correspondence approach (900 m), the Medium range was relatively insensitive to the approach used (15–19%), and the High range performed best using the pixel-wise approach (54–63%). The relative correspondence between the field measurement values and MERIS-derived cell counts can be characterized based on a correspondence range for any given cell count range metric (e.g., Very High = 76%–86%), highest correspondence value (e.g., Very High = 86%) or weighted average correspondence (e.g., Very High = 75.5%). Each metric can provide valuable information, but they are probably best used in combination to evaluate performance. It is important to note that these correspondence values incorporate the inaccuracies (errors) associated with both the reference data (in situ cell counts) and MERIS-derived estimates (Congalton, Iliamis, Knight, Konst, & Mace, 2001). Accordingly, regardless of the correspondence metric applied, the values represent conservative estimates for algorithm performance.

There was an elevated performance for the high value range 300 m pixel-wise analysis (Table 1). However, the other cell count ranges typically performed better in the 900 m analysis. Visual analysis of a large subset of records indicated that the elevated misclassifications associated with the 900 m analyses were attributable to shoreline (land) pixels that were included in the 900 m closest correspondence analysis but were flagged as land in the 300 m pixel-wise correspondence analysis and excluded.

The results of our analysis indicate that the algorithm can estimate cyanobacteria cell counts for inland lakes in the eastern US at Low (10,000–109,999) and Very High (>1,000,000) correspondence levels of approximately 90% and 83%, respectively. The intermediate cell count estimates (110,000–299,999 and 300,000–1,000,000) were below acceptable performance levels at approximately 28% and 40%, respectively (Table 1). Table 2 provides a detailed analysis of confusion between cell count ranges including the percent correct correspondence, and the percentages for both commission (Type I errors) and omission (Type II errors). The Low range had the best performance with an accuracy of 90%, only 10% omission, and moderate commission errors (16%). The Medium and High cell count ranges both had high omission (83% and 65%, respectively) and high commission errors (40% and 81%, respectively). The errors associated with the intermediate count ranges were mostly a result of confusion with the Low count range. The Very High cell count range has a correspondence accuracy of 83% with omission (17%) and commission (21%) errors nearly equally distributed. Overall performance resulted in a Kappa coefficient of 0.57. This can be interpreted as performing 57% better than random assignments across all cell count classes.

To best interpret the correspondence results, a thorough analysis of the characteristics associated with the in situ reference data is presented in Fig. 1. The frequency distribution of the in situ cell count data indicated that 54.6% of the samples (n = 1129) were below the minimal MERIS detection limit (10,000). 32.5% (n = 672) were in the Low range, 5.8% (n = 121) were in the Medium range, only 2.2% (n = 45) in the High range, and 4.8% (n = 101) in the Very High range. Wynne et al. (2010), Wynne et al. (2013) suggested that detection was possible to 35,000 cells/mL, but had insufficient field data to provide more resolution. MERIS has the sensitivity to detect variations of CI to 0.0001, which corresponds to 10,000 cells/mL. The frequency distribution of the in situ cell count data indicated that 54.6% of the samples (n = 1129) were below the minimal MERIS detection limit (10,000). 32.5% (n = 672) were in the Low range, 5.8% (n = 121) were in the Medium range, only 2.2% (n = 45) in the High range, and 4.8% (n = 101) in the Very High range. Wynne et al. (2010), Wynne et al. (2013) suggested that detection was possible to 35,000 cells/mL, but had insufficient field data to provide more resolution. MERIS has the sensitivity to detect variations of CI to 0.0001, which corresponds to 10,000 cells/mL.

4. Discussion

The high incidence (45%) of inland lakes with cyanobacteria cell counts >10,000 cells/mL confirms that occurrences of cyanobacteria blooms are now common events across the eastern US. However, due to the absence of a comprehensive monitoring program, it is not known if this represents an increased incidence or simply reflects...
long-term baseline conditions (Fristachi & Sinclair, 2008). However, there is widespread agreement among scientists and managers that cyanobacteria blooms are increasing (Hudnell, 2010) including one recent study documenting a record setting bloom event in Lake Erie in 2011 that was attributed to climatic and nutrient mediated changes consistent with anticipated future trends (Michalak et al., 2013). Others have also documented troubling trends in environmental perturbations including water temperature increases attributable to climate change (Paerl & Paul, 2012) and increased nutrient loadings resulting from intensive agricultural practices (USEPA, 2013b), these trends are indicative of potential risk factors for increased cyanobacteria bloom occurrences.

The results presented in Table 1 provide an interesting insight concerning the duration of cyanobacteria bloom events. Very little variability in the correspondence results was observed for temporal windows ranging from 3 to 15 days in duration. This result indicates that the events tend to be of relatively long duration and are typically stable for extended periods of time. No attempt was made to go beyond the maximum 15-day temporal window presented here. Thus, no additional insights concerning viable maximum temporal windows for derived product calibration and validation efforts can be provided.

This study has validated the robust retrieval capability of cyanobacteria cell counts for inland lakes in the eastern US that coincide with Low (10,000–109,999) and Very High (>1,000,000) values. The confusion with the intermediate cell count ranges is largely attributed to confusion with the low cell count range (Table 1). However, this association needs to be tempered due to the skewed nature of the in situ reference data that was dominated (72%) by low range values (n = 672). Intermediate cell count estimates (110,000–299,999 and 300,000–1,000,000) were below acceptable performance levels. A plausible explanation for the substandard performance for the intermediate “Medium and High” cell count values is related to the cell count measurements used for cyanobacteria monitoring. First, the uncertainty associated with the cell count laboratory procedures is unknown and the distinction of natural counting units was limited. Also, we hypothesize that the algorithm was measuring the intensity of phycocyanin and Chl-a absorption features (λ = 620 and 665 nm) to derive the cyanobacteria cell count estimates. As previously mentioned, there were issues in the comparison of two measured quantities with no direct overlap; optical signals of light absorption and scattering by pigments in cells, and cell counts without biovolume or biomass information. The ratio of pigmentation and biomass is not constant, and neither is the ratio of cell number to biomass. Spatial variability can be a factor in mis-match between satellite and field measurements. Cyanobacteria blooms have extreme spatial variability as documented by Wyme et al. (2010); thus a 300 m pixel (9.0 ha) can contain a wide range of cyanobacteria cell densities. Additionally, while the CI second derivative approach, centered on the MERIS fluorescence band at 681 nm, is far less sensitive to poor atmospheric correction, thin clouds can attenuate the CI values, which would lead to underestimates from satellite. Over the six state geographic extent of the study area, we would expect that the cyanobacteria measurement data would incorporate a number of different cyanobacteria taxa. Importantly, the biovolume (BV) impacts on reflectance as determined by the quantity of pigment and scattering (cross-sectional area) of individual species can differ by an order of magnitude (e.g., Microcystis sp. (small) = 19 µm² vs. Anabaena circinalis = 208 µm²) and even Microcystis sp. (large) BVs exhibit large differences (93 µm²). Accordingly, cyanobacteria cell counts should optimally be normalized to BV (ml/L) before interspecies comparisons or analysis is performed (MEMH, 2009). Biovolumes are not currently standard and pigment measurements are rare. However, for cyanobacteria remote sensing estimates, phycocyanin and Chl-a concentrations would likely be more closely related to BV’s than cell counts for heterogeneous cyanobacteria taxa. Of the available 6000 monitoring samples available over the study period, only 34% were suitable for validation analysis. Open and effective discussions and forums between scientists, stakeholders, policy makers, and environmental managers would be required to modify existing field sampling designs for providing high quality reference data for future validation efforts of satellite sensor observations (Schaefler et al., 2013).

The results of this study show that MERIS products can be used to monitor the onset of cyanobacteria bloom events and potentially provide a capability for monitoring outbreaks of concern in freshwater lakes. MERIS, and in the future Sentinel-3, provide a potential for retrieving data for small lakes at 1–3 day intervals. This represents a major advancement over current in situ based monitoring programs by providing the capability for near-real time products over large geographic areas (regional–national) at high temporal intervals. Implementation of an operational satellite cyanobacteria bloom monitoring program would have the benefit of the European Space Agency’s (ESA) new Ocean Land Colour Imager (OLCI) on the Sentinel-3 satellite scheduled for launch and operational service beginning in 2015. The OLCI was designed to complement MERIS by providing operational system enhancements including improved signal to noise ratio (SNR), additional bands for improved atmospheric corrections, and pointing to reduce sun glint (Table 4). Additionally, ESA plans include the eventual deployment of two OLCI sensors that would provide a 1–2 day revisit capability for most US locations of interest.

An additional application would be to use the robust detection capability for Low cyanobacteria cell counts to trigger predictive model implementation coinciding with the onset of cyanobacteria bloom events to forecast the potential for intensification. Subsequent OLCI-derived cell counts could then be used to generate alarm products for lakes

### Table 4: Technical characteristics of the ENVISAT Medium Resolution Imaging Spectrometer (MERIS) and Sentinel-3 Ocean and Land Colour Instrument (OLCI) systems (Donlon et al., 2012).

<table>
<thead>
<tr>
<th>Channel</th>
<th>Wavelength (nm)</th>
<th>Band Width (nm)</th>
<th>SNR</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>412.5</td>
<td>10</td>
<td>468:1</td>
<td>Yellow substance</td>
</tr>
<tr>
<td>2</td>
<td>442.5</td>
<td>10</td>
<td>413:1</td>
<td>Chl-a absorption max</td>
</tr>
<tr>
<td>3</td>
<td>490</td>
<td>10</td>
<td>355:1</td>
<td>Accessory pigments</td>
</tr>
<tr>
<td>4</td>
<td>510</td>
<td>10</td>
<td>306:1</td>
<td>Sediment/turbidity</td>
</tr>
<tr>
<td>5</td>
<td>560</td>
<td>10</td>
<td>289:1</td>
<td>Chl-a minimum</td>
</tr>
<tr>
<td>6</td>
<td>620</td>
<td>10</td>
<td>196:1</td>
<td>Phycocyanin/sediment</td>
</tr>
<tr>
<td>7</td>
<td>665</td>
<td>10</td>
<td>177:1</td>
<td>Chl-a absorption max</td>
</tr>
<tr>
<td>8</td>
<td>681.3</td>
<td>7.5</td>
<td>147:1</td>
<td>Chl-a fluorescence peak</td>
</tr>
<tr>
<td>9</td>
<td>709</td>
<td>9</td>
<td>158:1</td>
<td>Chl-a baseline</td>
</tr>
<tr>
<td>10</td>
<td>753.8</td>
<td>7.5</td>
<td>101:1</td>
<td>O2absorption/clouds</td>
</tr>
<tr>
<td>11</td>
<td>760.6</td>
<td>3.75</td>
<td>375:1</td>
<td>O2absorption/aerosols</td>
</tr>
<tr>
<td>12</td>
<td>775</td>
<td>14</td>
<td>157:1</td>
<td>Atmospheric corrections</td>
</tr>
<tr>
<td>13</td>
<td>865</td>
<td>20</td>
<td>114:1</td>
<td>Atmospheric corrections</td>
</tr>
<tr>
<td>14</td>
<td>885</td>
<td>10</td>
<td>113:1</td>
<td>Water vapor reference</td>
</tr>
<tr>
<td>15</td>
<td>900</td>
<td>10</td>
<td>51:1</td>
<td>Water vapor absorption</td>
</tr>
</tbody>
</table>

### OLCI (2–3 days, 1 satellite; 1–3 days, 2 satellites)

<table>
<thead>
<tr>
<th>Channel</th>
<th>Wavelength (nm)</th>
<th>Band Width (nm)</th>
<th>SNR</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>400</td>
<td>15</td>
<td>2,188:1</td>
<td>Aerosol correction</td>
</tr>
<tr>
<td>2</td>
<td>412.5</td>
<td>10</td>
<td>2,061:1</td>
<td>Chl-a fluorescence</td>
</tr>
<tr>
<td>3</td>
<td>442.5</td>
<td>10</td>
<td>1,811:1</td>
<td>Chl-a fluorescence</td>
</tr>
<tr>
<td>4</td>
<td>442</td>
<td>10</td>
<td>1,341:1</td>
<td>Chl-a absorption max</td>
</tr>
<tr>
<td>5</td>
<td>510</td>
<td>10</td>
<td>1,488:1</td>
<td>O2absorption/clouds</td>
</tr>
<tr>
<td>6</td>
<td>560</td>
<td>10</td>
<td>1,280:1</td>
<td>O2absorption/aerosols</td>
</tr>
<tr>
<td>7</td>
<td>620</td>
<td>10</td>
<td>997:1</td>
<td>Atmospheric correction</td>
</tr>
<tr>
<td>8</td>
<td>665</td>
<td>10</td>
<td>883:1</td>
<td>Atmospheric correction</td>
</tr>
<tr>
<td>9</td>
<td>673.75</td>
<td>7.5</td>
<td>707:1</td>
<td>Chl-a fluorescence</td>
</tr>
<tr>
<td>10</td>
<td>681.25</td>
<td>7.5</td>
<td>745:1</td>
<td>Atmospheric correction</td>
</tr>
<tr>
<td>11</td>
<td>709.75</td>
<td>10</td>
<td>785:1</td>
<td>Atmospheric correction</td>
</tr>
<tr>
<td>12</td>
<td>753.75</td>
<td>7</td>
<td>605:1</td>
<td>Atmospheric correction</td>
</tr>
<tr>
<td>13</td>
<td>761.25</td>
<td>2.5</td>
<td>232:1</td>
<td>Atmospheric correction</td>
</tr>
<tr>
<td>14</td>
<td>764.375</td>
<td>3.75</td>
<td>305:1</td>
<td>Atmospheric correction</td>
</tr>
<tr>
<td>15</td>
<td>767.5</td>
<td>2.5</td>
<td>330:1</td>
<td>Atmospheric correction</td>
</tr>
<tr>
<td>16</td>
<td>778.75</td>
<td>15</td>
<td>812:1</td>
<td>Atmospheric correction</td>
</tr>
<tr>
<td>17</td>
<td>865</td>
<td>20</td>
<td>666:1</td>
<td>Atmospheric correction</td>
</tr>
<tr>
<td>18</td>
<td>885</td>
<td>10</td>
<td>395:1</td>
<td>Atmospheric correction</td>
</tr>
<tr>
<td>19</td>
<td>900</td>
<td>10</td>
<td>308:1</td>
<td>Atmospheric correction</td>
</tr>
<tr>
<td>20</td>
<td>940</td>
<td>20</td>
<td>203:1</td>
<td>Atmospheric correction</td>
</tr>
<tr>
<td>21</td>
<td>1020</td>
<td>40</td>
<td>152:1</td>
<td>Aerosol correction</td>
</tr>
</tbody>
</table>
exhibiting Very High cell counts that could be applied to validate modeling predictions by evaluating the percentage of correct predictions and characterizing Type I and II errors. The application of a predictive model would potentially provide valuable forecast information to local managers, decision makers and scientists for directing limited field monitoring resources and to provide public recreational advisories. A robust modeling capability would provide a forecasting capability and would serve to fill the technological gap for the middle range cyanobacteria cell counts.

5. Conclusions

The results of this study document the potential for using MERIS-derived cyanobacteria cell count estimates to monitor freshwater lakes (> 100 ha) across the eastern US. The capability to accurately estimate Low and Very High cell count ranges represent a significant capability for the development of an operational monitoring program. The substandard performance for intermediate cell count ranges could be attributed to the possibility that the algorithm does not have sensitivity to the presence of cyanobacteria in this range, the lack of available in situ cyanobacteria taxonomic data, or the distinction of natural counting units to support the conversion of cell counts to cell volumes. Because the algorithm functions by detecting the impacts of absorption by phycocyanin and Chlorophyll-a on the spectral curvature (CI-mutli), correlations would optimally be compared with cell volume (surface area) or pigment concentrations versus cell counts. The potential development of a cyanobacteria predictive modeling capability for integration with future OLCI-derived cell count estimates may provide a good solution for an operational monitoring capability that could support cyanobacteria bloom predictions for the eastern US. This study was a first attempt to use existing monitoring programs across multiple states and agencies to validate remote sensing cyanobacteria cell counts. Finally, to support future validation efforts of satellite sensor observations of freshwater water quality parameters, modeling and standardization of existing algal field sampling and cytological enumeration protocols are needed to provide high quality reference data.

Acknowledgments

The research described in this article was partially funded and conducted by the USEPA’s Office of Research and Development (ORD) under a Pathfinder Innovation Project II (PPIP) grant and by the NASA Applied Science Program announcement NNH08ZDA001N (Human Health and Air Quality) under contract NNH09AL531. It has been subject to the US EPA programmatic review and has been approved for publication. Mention of any trade names or commercial products does not constitute endorsement or recommendation for use. The authors would like to acknowledge the assistance of Keith Loftin (USGS), Robyn Comny (USEPA), James Hagy (USEPA), Bryan Milstead (USEPA), Lesley Vazquez-Coriano (USEPA), and Jan Kurtz (USEPA), Tim Wynne (NOAA) and Shelly Tomlinson (NOAA), Rick Wilson (Ohio EPA), Russ Gibson (Ohio EPA), Dan Glomski (Ohio EPA), and Travis Briggs (CSS). The Florida data used in the study were collected and provided by the St. Johns River Water Management District (WMID) Division of Water Resources, Environmental Sciences Bureau by Rolland Fulton, John Hendrickson, Robert Burks, and Erich Marzolf all with the Suwannee River WMWD.

References

Michaik, A.M., et al. (2013). Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with future conditions (early ed.). PNAS.


