

Tracking cyanobacteria blooms: Do different monitoring approaches tell the same story?



Isabella Bertani ^{a,*}, Cara E. Steger ^{a,1,2}, Daniel R. Obenour ^b, Gary L. Fahnenstiel ^{a,c}, Thomas B. Bridgeman ^d, Thomas H. Johengen ^e, Michael J. Sayers ^f, Robert A. Shuchman ^f, Donald Scavia ^a

^a Water Center, Graham Sustainability Institute, University of Michigan, 625 E. Liberty St., Suite 300, Ann Arbor, MI 48104, USA

^b Department of Civil, Construction, & Environmental Engineering, North Carolina State University, Campus Box 7908, Raleigh, NC 27695-7908, USA

^c Great Lakes Research Center, Michigan Technological University, 1400 Townsend Drive, Houghton, MI 49931, USA

^d Department of Environmental Sciences and Lake Erie Center, University of Toledo, 6200 Bayshore Drive, Oregon, OH 43616, USA

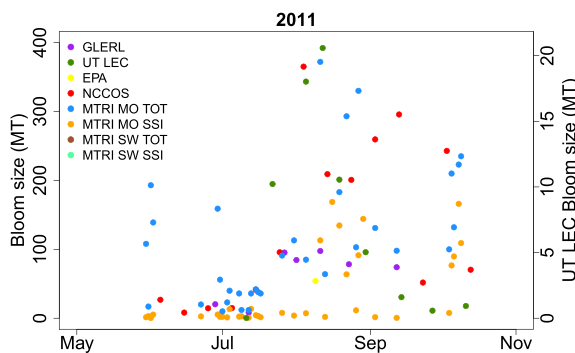
^e Cooperative Institute for Limnology and Ecosystems Research, University of Michigan, 4840 South State St., Ann Arbor, MI 48108, USA

^f Michigan Tech Research Institute, Michigan Technological University, 3600 Green Ct., Suite 100, Ann Arbor, MI 48105, USA

HIGHLIGHTS

- Different monitoring approaches are used to detect cyanobacteria blooms in lakes.
- We assessed the coherence of different monitoring methods in capturing bloom dynamics.
- We modeled relationships between environmental drivers and bloom variability.
- Discrepancies across monitoring methods may influence modeled relationships.
- Integrating multiple survey methods is key to improve bloom detection and modeling.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 2 August 2016

Received in revised form 21 September 2016

Accepted 3 October 2016

Available online xxx

Editor: D. Barcelo

Keywords:

Cyanobacterial harmful algal blooms
Water quality
Long-term time series
Boosted regression trees
Biophysical drivers

ABSTRACT

Cyanobacteria blooms are a major environmental issue worldwide. Our understanding of the biophysical processes driving cyanobacterial proliferation and the ability to develop predictive models that inform resource managers and policy makers rely upon the accurate characterization of bloom dynamics. Models quantifying relationships between bloom severity and environmental drivers are often calibrated to an individual set of bloom observations, and few studies have assessed whether differences among observing platforms could lead to contrasting results in terms of relevant bloom predictors and their estimated influence on bloom severity. The aim of this study was to assess the degree of coherence of different monitoring methods in (1) capturing short- and long-term cyanobacteria bloom dynamics and (2) identifying environmental drivers associated with bloom variability. Using western Lake Erie as a case study, we applied boosted regression tree (BRT) models to long-term time series of cyanobacteria bloom estimates from multiple *in-situ* and remote sensing approaches to quantify the relative influence of physico-chemical and meteorological drivers on bloom variability. Results of BRT models showed remarkable consistency with known ecological requirements of cyanobacteria (e.g., nutrient loading, water temperature, and tributary discharge). However, discrepancies in inter-annual and intra-seasonal bloom dynamics across monitoring approaches led to some inconsistencies in the relative importance, shape, and

* Corresponding author.

E-mail address: ibertani@umich.edu (I. Bertani).

¹ Joint first authors.

² Present address: Graduate Degree Program in Ecology, Natural Resource Ecology Lab, Colorado State University, Campus Delivery A245, Fort Collins, CO 80523-1499, USA.

sign of the modeled relationships between select environmental drivers and bloom severity. This was especially true for variables characterized by high short-term variability, such as wind forcing. These discrepancies might have implications for our understanding of the role of different environmental drivers in regulating bloom dynamics, and subsequently for the development of models capable of informing management and decision making. Our results highlight the need to develop methods to integrate multiple data sources to better characterize bloom spatio-temporal variability and improve our ability to understand and predict cyanobacteria blooms.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

There has been a global increase in cyanobacteria dominance and associated harmful algal blooms (CyanoHABs or CHABs) in aquatic systems over the past two centuries (Paerl and Paul, 2012; Taranu et al., 2015), posing serious threats to the functioning of these ecosystems and to the health of organisms that rely on them, including humans (Codd et al., 2005a; Smith, 2003). CHAB impacts include marked decreases in water transparency, thus promoting the suppression of other primary producers and potentially triggering cascading effects across higher trophic levels, and increases in the extent of hypoxic/anoxic conditions due to algal biomass decomposition (Havens, 2008). The production of toxins by several cyanobacteria taxa also causes severe impairment of freshwater resources, and cases of intoxication of animals and humans due to consumption of contaminated water have been reported in several regions of the world (Carmichael and Boyer, 2016; Chorus and Bartram, 1999; Codd et al., 2005b). These public health risks and decreases in water quality may ultimately also have strong negative economic repercussions on local fisheries, tourism, and recreation industries (Dodds et al., 2009).

Increased anthropogenic nutrient inputs have been identified as a major factor promoting CHABs (Downing et al., 2001; Paerl and Otten, 2013a). Additional drivers thought to enhance cyanobacteria growth include increased water temperatures (Kosten et al., 2012; Paerl and Huisman, 2008), changes in the frequency and timing of extreme weather events (Michalak et al., 2013; Paerl and Huisman, 2009), alterations in hydrologic regime and water residence time (Elliott, 2010), changes in nutrient stoichiometric ratios (Anderson et al., 2002; Baker et al., 2014), and introduction of invasive species (Vanderploeg et al., 2001).

Despite the vast amount of knowledge developed in the past decades on the ecology of cyanobacteria, quantifying the relative influence and interactions of different environmental drivers in regulating bloom dynamics remains a significant challenge (Perovich et al., 2008). This difficulty is due partly to limitations in accuracy and spatio-temporal coverage of monitoring methods typically used to estimate bloom intensity and to model relationships with environmental variables (Bullerjahn et al., 2016; Dale and Murphy, 2014; Ho and Michalak, 2015). CHAB occurrence is commonly tracked via *in-situ* sampling or remote sensing (Bullerjahn et al., 2016; Srivastava et al., 2013). Spatio-temporal coverage of *in-situ* surveys is typically limited by the costs and feasibility of maintaining long-term monitoring programs with adequate numbers of stations and sampling frequency. On the other hand, satellite imagery may incur severe spatio-temporal limitations due to cloud cover, and remote sensing-derived estimates of bloom size may be affected by significant biases due to complex water optical properties and specific saturation constraints of different sensors and retrieval algorithms (Park et al., 2010; Reinart and Kutser, 2006; Shen et al., 2012; Wynne et al., 2013). Extremely dense surface bloom accumulations (surface scums; Fig. S3) may result in satellite underestimation due to signal saturation (Kutser et al., 2006). Furthermore, wind-induced water column mixing can prevent the bloom from rising to the surface, resulting in potential underestimation of bloom intensity from satellite (Wynne et al., 2010).

As a result of these limitations, discrepancies in the characterization of seasonal and inter-annual bloom dynamics often emerge when

analyzing time series of CHAB observations from different monitoring approaches, potentially leading to different conclusions on the main drivers of cyanobacteria bloom development and persistence (Ho and Michalak, 2015). These differences highlight the need to explore the challenges associated with comparing and integrating different CHAB monitoring products to improve our understanding of the uncertainties associated with current bloom tracking and modeling efforts, especially in the light of the increasing need for reliable, science-based recommendations to managers. In this perspective, the integration of bloom measurements from multiple types of monitoring approaches into CHAB modeling efforts has recently been recognized as a key research area to advance our predictive knowledge of CHAB dynamics (Bullerjahn et al., 2016).

The main aim of this work is to assess the degree of coherence of multiple data sources in capturing short- and long-term CHAB dynamics and in identifying key environmental drivers associated with the observed variability in bloom intensity. To this end, we compiled multiple long-term (2002–2013), high-frequency (daily to bi-weekly) datasets of bloom estimates in the western basin of Lake Erie, a Laurentian Great Lake. The Great Lakes are a vital natural resource, containing roughly 20% of the world's surface freshwater. Specifically, Lake Erie provides drinking water to over 11 million people and supports one of the largest freshwater commercial fisheries worldwide. In the 1960s and 70s, Lake Erie experienced intense eutrophication, with large cyanobacterial blooms in the shallow western basin and beyond (Scavia et al., 2014). The Great Lakes Water Quality Agreement (GLWQA) set target loads for total phosphorus (TP) in 1978. Widespread phosphorus load reductions, primarily from point sources, were implemented following this Agreement (Dolan, 1993), and water quality improvements were observed in the 1980s and 90s (DePinto et al., 1986; Makarewicz and Bertram, 1991). However, these trends appear to have reversed since the mid-1990s (Scavia et al., 2014; Watson et al., 2016), with cyanobacteria blooms composed mainly of *Microcystis* spp. increasing in frequency and intensity in the lake's western basin (Bridgeman et al., 2013; Obenour et al., 2014; Stumpf et al., 2012) and causing serious ecological and public health risks.

The specific objectives of this work are (1) to compare the intra-seasonal and inter-annual relative variability in western Lake Erie CHAB estimates derived from different *in-situ* and remote sensing approaches and (2) to assess the degree of coherence across monitoring products in identifying environmental drivers and corresponding critical thresholds associated with CHAB occurrence. We applied boosted regression tree (BRT) analysis to each set of bloom estimates to analyze relationships between bloom size and a suite of candidate physico-chemical and meteorological predictors. We compared modeling results to assess whether different monitoring products lead to identifying different key predictors and/or contrasting functional relationships among variables. Finally, we compared the critical thresholds of environmental drivers identified by the BRT models with results from previous CHAB studies in western Lake Erie and, more generally, with known ecological requirements of cyanobacteria.

2. Materials and methods

The Materials and methods section is organized as follows. We begin by describing the study site (Section 2.1), followed by a description of

the CHAB monitoring products included in this study (Section 2.2). We then provide an account of the environmental variables considered as candidate CHAB predictors (Section 2.3), and we describe the modeling analyses performed on each of the CHAB monitoring products (Section 2.4).

2.1. Study site

The western basin of Lake Erie (Fig. 1) has a surface area of approximately 3000 km². It is relatively shallow (mean depth 7.4 m) and well-mixed, with a water residence time of about 51 days (Michalak et al., 2013). The basin is influenced mainly by the Detroit River to the north and the Maumee River to the south-west. While the Maumee River flow is substantially lower than that of the Detroit River (5% and 90% of the total flow discharged annually into the western basin, respectively), the two rivers contribute a similar portion (45% each) of the annual TP load delivered to the basin from both point and nonpoint sources (IJC, 2014; Maccoux et al., 2016). The even load proportion is due to the substantially higher TP concentrations typically found in the Maumee River, which drains one of the most heavily agricultural watersheds in the Great Lakes region (Richards and Baker, 2002; Richards et al., 2008; Scavia et al., 2016). Compared to the Maumee, Detroit River nutrient concentrations are generally too low to significantly contribute to CHAB formation (IJC, 2014; Scavia et al., 2016). Summer cyanobacteria blooms typically originate and peak in the area directly affected by the Maumee River plume and tend to move eastward later in the season (Wynne and Stumpf, 2015). Recent studies have shown a positive relationship between peak summer CHAB size in the western basin and Maumee River spring phosphorus load (Bertani et al., 2016; Michalak

et al., 2013; Obenour et al., 2014; Stumpf et al., 2012), though seasonal phosphorus load alone may not entirely explain the substantial increase in bloom size observed over the last decade (Obenour et al., 2014). Although several different cyanobacteria taxa can be found in western Lake Erie, *Microcystis* has generally dominated the blooms in recent years (Bridgeman et al., 2012; Brittain et al., 2000; Chaffin et al., 2013).

2.2. CHAB monitoring products in Lake Erie

We compiled eight *in-situ* and remote sensing monitoring products, seven of which provide multiple observations of CHAB size over the course of the CHAB season (Jun–Oct) during 2002–2013. For each sampling date in each dataset we expressed bloom size as total metric tons of chlorophyll-*a* (MT Chl-*a*) over the sampled area, following previously published protocols to calculate bloom size wherever possible (e.g., Stumpf et al., 2012; Bridgeman et al., 2013) and/or applying appropriate conversions as described below and in the Supplementary materials. Comparing absolute bloom sizes derived from different monitoring products is complicated by considerable differences in the location and extent of the sampled areas, efficiencies of sampling methods and analytical techniques, and accuracy of the applied conversion factors. As a result, we limit our analysis to assessing the degree of coherence among monitoring products in capturing intra-seasonal and inter-annual variability in *relative* rather than absolute bloom size.

2.2.1. UT LEC *in-situ* sampling (2002–2013)

The University of Toledo Lake Erie Center (UT LEC) collected *Microcystis* samples at six sites from an area of approximately 340 km² in the vicinity of Maumee Bay (Fig. 1) in the years 2002–2013

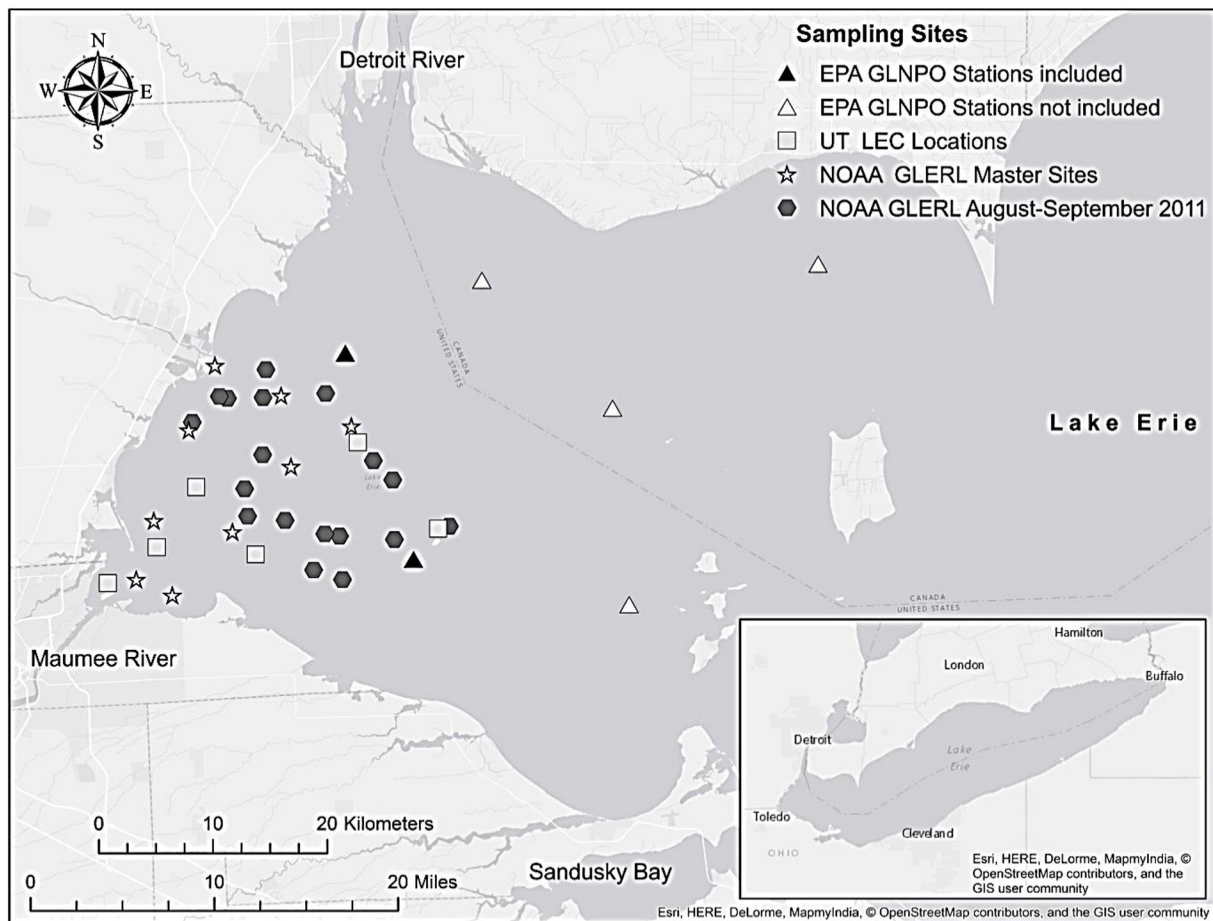


Fig. 1. Map of the western basin of Lake Erie with *in-situ* sampling locations shown. EPA GLNPO: US Environmental Protection Agency's Great Lakes National Program Office; UT LEC: University of Toledo Lake Erie Center; NOAA GLERL: National Oceanic and Atmospheric Administration's Great Lakes Environmental Research Laboratory.

(Bridgeman et al., 2013). Surveys were carried out approximately bi-weekly from May through October and samples were collected with vertical plankton tows over the entire water column. *Microcystis* biovolume (mL/m^2) was determined as reported in Bridgeman et al. (2013). In our analysis, total biovolume was converted to cell dry weight using a relationship specifically derived for *Microcystis* in western Lake Erie (see Fig. 2 in Bridgeman et al., 2013). We then converted cell dry weight to chlorophyll-*a* using the average chlorophyll content of *Microcystis* cells measured in western Lake Erie ($0.006125 \text{ g/g dry wt}$; Chaffin et al., 2012). This value is consistent with measurements of chlorophyll per unit biomass in Lake Erie in 2014 (T. Johengen, pers. comm.) and with average *Microcystis* cell chlorophyll contents reported in the literature (Chen et al., 2011; Long et al., 2001; Wang et al., 2007). To obtain an estimate of total bloom mass for the sampling area in each sampling date, we averaged the values of MT Chl-*a*/ m^2 across the sampling sites, following the approach established in Bridgeman et al. (2013), and multiplied the average by the extent of the sampled area (340 km^2).

2.2.2. NOAA-GLERL in-situ sampling (2008–2013)

The National Oceanic and Atmospheric Administration's Great Lakes Environmental Research Laboratory (NOAA-GLERL, hereafter abbreviated as GLERL) has been monitoring water quality at several sites in the western basin of Lake Erie since 2008. While surveys were carried out monthly in 2008 at 13 sites across a broad portion of the basin, in 2009–2011 nine master stations from an area of approximately 300 km^2 near the Maumee River inflow (Fig. 1) were sampled weekly to bi-weekly from June through September–October of each year. In August–September of 2011, when the bloom moved far beyond the Maumee Bay, samples were collected at stations closer to the central portion of the basin, where the bloom was located. In these cases, we only considered sample sites located within the area covered by the GLERL master

stations or those located outside of that area but which did not extend beyond the region sampled by Bridgeman et al. (2013), so that the size of the total sampled area remained roughly 300 km^2 in all years (Fig. 1). In 2012–2013, four out of the nine master stations were sampled from approximately the same 300 km^2 area. Chlorophyll-*a* concentrations were measured from samples collected with a 1-m Niskin bottle lowered below the water surface. We multiplied chlorophyll values by the depth of the station (from NOAA bathymetry map; <http://www.ngdc.noaa.gov/>) and averaged them across stations for each sampling date, in accordance with the averaging approach used for the UT LEC data (Bridgeman et al., 2013). These averages were then multiplied by an area of 300 km^2 to get total MT chlorophyll for the sampled area.

2.2.3. EPA GLNPO in-situ sampling (2002–2013)

Two cruises per year (April and August) have been carried out since 1983 by the EPA's Great Lakes National Program Office (GLNPO), sampling six offshore stations in the western basin (Fig. 1, <http://www3.epa.gov/greatlakes/monitoring/>). Due to the very limited seasonal sampling frequency, we did not include this product in our modeling analyses. However, chlorophyll concentrations measured during the August cruise were used to obtain a measure of total MT chlorophyll, which was compared to the inter-annual variability of the bloom estimates derived from other data sources. Specifically, we used chlorophyll values measured from samples collected with a Niskin bottle just below the surface at the two EPA stations falling within the area sampled by the other *in-situ* surveys (Fig. 1). We multiplied chlorophyll concentrations by station depth, averaged across stations, and multiplied that average by an area of 300 km^2 to obtain estimates of total MT chlorophyll comparable to those provided by other *in-situ* monitoring products. EPA systematically measured chlorophyll at multiple depths across the water column. While we only used surface chlorophyll values to ensure

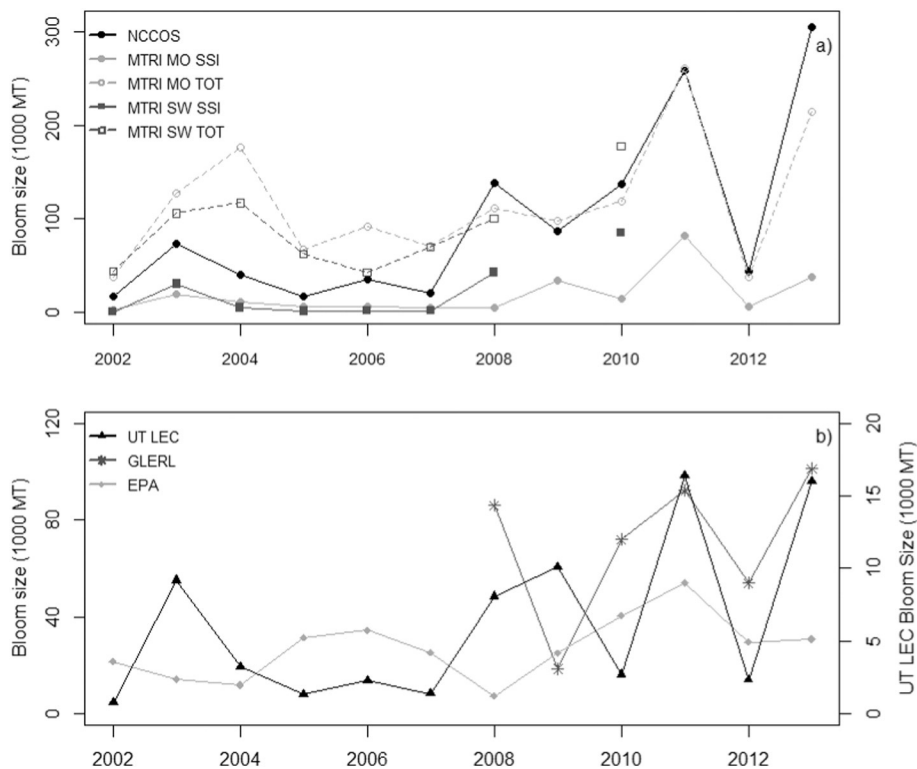


Fig. 2. Annual bloom severity, calculated as the peak 30-day moving average across each year, for the remote sensing (a) and *in-situ* (b) monitoring products. UT LEC is given a separate y-axis for ease of interpretation. SeaWiFS estimates for the year 2009 are omitted due to sensor malfunctions that resulted in only four images available over the whole CHAB season. NCCOS: NOAA's National Centers for Coastal Ocean Science; MTRI MO SSI: Michigan Tech Research Institute MODIS Surface Scum Index; MTRI MO TOT: Michigan Tech Research Institute MODIS Total (chlorophyll-*a* + surface scum); MTRI SW SSI: Michigan Tech Research Institute SeaWiFS Surface Scum Index; MTRI SW TOT: Michigan Tech Research Institute SeaWiFS Total (chlorophyll-*a* + surface scum); UT LEC: University of Toledo Lake Erie Center; GLERL: Great Lakes Environmental Research Laboratory; EPA: Environmental Protection Agency.

consistency with GLERL bloom estimates, we compared EPA bloom size estimates obtained using chlorophyll measurements taken below the surface with estimates obtained by calculating the average chlorophyll value across multiple depths throughout the whole water column. The two sets of bloom size measurements align very closely, showing comparable relative inter-annual variability ($R^2 = 0.95$).

2.2.4. NOAA-NCCOS remote sensing (2002–2013)

CHAB size estimates have been generated by NOAA's National Centers for Coastal Ocean Science (NOAA-NCCOS, hereafter abbreviated as NCCOS) using images from the MEdium-spectral Resolution Imaging Spectrometer (MERIS; 2002–2011) and from the Moderate Resolution Imaging Spectroradiometer (MODIS; 2012–present) (Stumpf et al., 2012; Wynne et al., 2013; Wynne et al., 2010; Wynne et al., 2008). Cyanobacteria biomass is quantified and expressed in terms of a "Cyanobacteria Index" (CI) that is positively related to cyanobacteria abundance in the top portion of the water column (Wynne et al., 2013; Wynne et al., 2010). The CI thus represents the blue-green portion of total surface chlorophyll. Individual cloud-free satellite image pixels are compiled into 10-day composites by dividing the CHAB season (Jun–Oct) into discrete 10-day periods and by taking the highest CI value observed at each cloud-free pixel during each period (Stumpf et al., 2012; Wynne and Stumpf, 2015). These composites are developed to reflect the total biomass of *Microcystis* during each 10-day period.

Bloom intensity is calculated by summing CI values across all pixels within western Lake Erie for each 10-day composite, and annual bloom size is calculated as the maximum 30-day (i.e., three 10-day composites) moving average, according to methods established in Stumpf et al. (2012). One CI corresponds to approximately 1.2×10^{20} cyanobacteria cells (Stumpf et al., 2012), and is equivalent to approximately 4800 metric tons (MT) cyanobacteria dry weight (Obenour et al., 2014). We converted cell dry weight to chlorophyll-*a* using the average chlorophyll content of *Microcystis* cells measured in western Lake Erie (0.006125 g/g dry wt; Chaffin et al., 2012).

2.2.5. MTRI remote sensing (2002–2013)

CHABs in western Lake Erie are mapped by the Michigan Tech Research Institute (MTRI) using MODIS and Sea-Viewing Wide Field-of-View Sensor (SeaWiFS) satellite imagery by using the Color Producing Agent Algorithm (CPA-A) (Shuchman et al., 2013) to first retrieve total chlorophyll-*a* concentration and then classify pixels as CHABs when chlorophyll exceeds 20 $\mu\text{g/L}$ and water temperature is $> 18^\circ\text{C}$ (Sayers et al., 2016; Shuchman, 2014). The CPA-A is a multi-spectral bio-optical retrieval procedure which simultaneously estimates concentrations of the three primary color producing agents (chlorophyll, colored dissolved organic matter, and suspended minerals). This approach allows for total chlorophyll estimation in complex waters, such as those found in Lake Erie (Shuchman et al., 2013), from SeaWiFS and MODIS data.

Surface scum accumulations (Fig. S3) are mapped separately using a Surface Scum Index (SSI) algorithm developed by MTRI, which is an adaptation of the Normalized Difference Vegetation Index (NDVI) (Rouse et al., 1973) and which is also applicable to all of the above mentioned sensors. While the CPA-A algorithm maps total chlorophyll-*a* concentrations for pixels where no scum is present, the SSI is a presence/absence indicator that identifies pixels with surface scum, but does not presently provide a measure of chlorophyll concentration within the scum. Chlorophyll concentrations within a *Microcystis* scum layer typically show high spatial variability (G. Fahnenstiel and T. Bridgeman, pers. comm.), so that given the limited scum sampling data and relatively coarse satellite resolution (~ 1 km), any assumption of the average scum chlorophyll content is a rough approximation. However, to compare the relative variability in surface scum estimates derived from the current version of the SSI algorithm with other available satellite products, we approximated the average chlorophyll content of surface scum to be equivalent to the total MT chlorophyll found in the top

2 m of the water column assuming a concentration equal to the 99th percentile of the distribution of surface chlorophyll concentrations measured in the western basin by NOAA-GLERL over the period 2008–2013 (200 $\mu\text{g/L}$).

Hereafter we refer to the two time series of MTRI surface scum estimates as MO SSI (from MODIS images) and SW SSI (from SeaWiFS images). While we report surface scum estimates obtained from both remote sensing products, only the MO SSI time series was included in our modeling analyses. This is because the SW SSI time series is restricted to a time period when several years had low surface scum occurrence (Fig. 2a), resulting in a relatively low number of data points exhibiting scum values significantly larger than zero.

In addition to the two surface scum time series, we developed two separate time series - MO TOT and SW TOT - where total bloom size was calculated as the sum of the MT chl-*a* calculated for the surface scum pixels and the MT chl-*a* calculated for all non-scum pixels in which chl-*a* > 20 $\mu\text{g/L}$ (Sayers et al., 2016).

In several MTRI MODIS and SeaWiFS images, cloud cover resulted in a varying number of pixels with missing data, possibly leading to CHAB size underestimation. To correct for bias in images with less than one third of pixels missing, we rescaled the observed chlorophyll mass using the ratio of the western basin area (~ 3000 km²) to the total visible area in each image. However, if over one third of an image's pixels were missing, the image was considered unreliable, and removed from this analysis. In 2009, only four SeaWiFS images are available throughout the CHAB season due to severe sensor malfunctions. We therefore omitted 2009 SeaWiFS data from analyses.

2.3. Environmental data and predictor variable development

We considered a suite of candidate physical, chemical, and meteorological predictors that are hypothesized to influence CHAB severity and seasonal dynamics (Table 1). Anthropogenic nutrient enrichment is recognized as one of the major drivers of cyanobacteria blooms in aquatic systems (Brookes and Carey, 2011; Downing et al., 2001; Paerl et al., 2011), and spring TP loading from the Maumee River has been shown to explain a large portion of the inter-annual variability in CHAB size in western Lake Erie (Obenour et al., 2014; Stumpf et al., 2012).

We calculated monthly TP, dissolved reactive phosphorus (DRP), and total nitrogen (TN) loads (MT) from the Maumee River using river nutrient concentration data collected by Heidelberg University's National Center for Water Quality Research (NCWQR, <http://tinyurl.com/jgq2jxp>) and stream flow data from the United States Geological Survey (USGS, <http://www.usgs.gov/water>). Missing nutrient concentrations were imputed as described in Obenour et al. (2014). Cumulative loads from June, April to June, and February to June were included in the models as potential explanatory variables.

Tributary discharge can also influence CHAB development by affecting lake turbidity, local circulation patterns, stratification, and residence time (Mitrovic et al., 2003; Verspagen et al., 2006). Because blooms in the western basin of Lake Erie typically originate and peak in the area influenced by the Maumee River plume (Wynne and Stumpf, 2015), we included Maumee River flow (m³/s) as a candidate predictor.

Lake circulation and mixing regime are affected by wind forcing (Beletsky et al., 2013; Michalak et al., 2013; Wynne et al., 2011). Hourly wind speed (m/s) and wind direction (degrees from true North) were acquired from NOAA's National Buoy Data Center (Buoy Station 45005, <http://www.ndbc.noaa.gov>) (Michalak et al., 2013; Zhou et al., 2015). Wind speed and direction were missing from only 4.7% of the hours in the study period, and were imputed using a linear regression with data from the nearby Buoy Station THL01. After imputation, only 0.6% of the hourly data remained missing. Average daily wind stress was calculated using the drag coefficient determined by Hsu (1974) following the method reported in Wynne et al. (2010). Wind velocity for the northerly and westerly components and wind stress were included in the models.

Table 1
Candidate environmental variables used in the BRT analysis.

Variable	Description	Mean (range)
WaterTemp_d (°C)	Average lake surface temperature for the western basin (from the Great Lakes Surface Environmental Analysis (GLSEA) model), where <i>d</i> is the 2, 8, or 30 days previous	20.0 (8.7–27.5)
Stress_d (Pa)	Average wind stress (calculated following the method reported in Wynne et al., 2010), where <i>d</i> is the 2, 8, or 30 days previous	0.05 (0.00–0.86)
NortherlyWind_d (m/s)	Average northerly wind component (from the north) recorded at National Data Buoy Center's buoy 45005, where <i>d</i> is the 2, 8, or 30 days previous	0.36 (–16.57–10.87)
WesterlyWind_d (m/s)	Average westerly wind component (from the west) recorded at National Data Buoy Center's buoy 45005, where <i>d</i> is the 2, 8, or 30 days previous	0.29 (–13.55–10.26)
Irradiance_d (W/m ²)	Daily average of hourly Climate Forecast System (CFSR) downward shortwave radiation flux at water surface in W/m ² , where <i>d</i> is the 2, 8, or 30 days previous	224 (7–362)
Discharge_d (m ³ /s)	Average daily Maumee river discharge from United States Geological Survey, where <i>d</i> is the 2, 8, or 30 days previous	112.7 (1.6–2217.2)
TP_June (MT)	Sum of Maumee TP daily loads in June	120.6 (5.0–496.8)
TP_JuneApril (MT)	Sum of Maumee TP daily loads in April–June	673.9 (39.6–1516.5)
TP_JuneFeb (MT)	Sum of Maumee TP daily loads in February–June	1343.5 (488.7–2622.3)
DRP_June (MT)	Sum of Maumee DRP daily loads in June	32.3 (0.2–124.5)
DRP_JuneApril (MT)	Sum of Maumee DRP daily loads in April–June	139.1 (7.1–265.5)
DRP_JuneFeb (MT)	Sum of Maumee DRP daily loads in February–June	282.8 (94.6–539.7)
TN_June (MT)	Sum of Maumee TN daily loads in June	3090.4 (72.7–10,115.8)
TN_JuneApril (MT)	Sum of Maumee TN daily loads in April–June	14,069.7 (1092.1–26,281.1)
TN_JuneFeb (MT)	Sum of Maumee TN daily loads in February–June	23,894.1 (6499.0–38,605.0)
Year	Yearly temporal trend	(2002–2013)
Month	Monthly temporal trend	(5–10)

Water temperature may promote cyanobacteria development both directly by affecting algal growth rates, and indirectly by enhancing water column stability (Jöhnk et al., 2008; Paerl and Huisman, 2008). Daily surface water temperatures (°C) were acquired from GLERL's Great Lakes Surface Environmental Analysis (GLSEA) mapping effort and averaged over the entire western basin.

Solar irradiance influences algal photosynthetic and growth rates, and modulates the outcome of competition among phytoplankton taxa with different light requirements (Havens et al., 1998; Huisman et al., 1999; Litchman, 1998). Furthermore, light availability influences *Microcystis* cell buoyancy by controlling carbohydrate production and accumulation through photosynthesis (Ibelings et al., 1991; Kromkamp and Mur, 1984; Wallace and Hamilton, 1999). Hourly short-wave radiation data (W/m²) were acquired from NOAA's National Center for Environmental Prediction's Climate Forecast System Re-Analysis (CFSR, <http://rda.ucar.edu/>), and daily average values were calculated.

For each meteorological and hydrological predictor (wind velocity, wind stress, irradiance, temperature, and river flow), three time-lagged variables were included as candidate predictors in the model. Specifically, we calculated average values for the 2 days, 8 days, and 30 days

preceding each sampling date, to capture potential effects of each variable at different time scales leading up to each bloom observation (Millie et al., 2014) while minimizing collinearity among time-lagged variables and preventing model over-fitting. For example, wind forcing is expected to exert both a short-term (hourly/daily) impact on bloom dynamics by modulating mixing of the water column as well as a longer-term effect by driving lake circulation, currents, and subsequently water/nutrient residence time (Michalak et al., 2013; Zhou et al., 2015).

Monthly and yearly temporal trend components were also included in the models to assess whether any consistent pattern in bloom seasonality emerged across monitoring products (monthly trend) and whether our models would confirm previous findings reporting an increased susceptibility of western Lake Erie to large blooms (yearly trend; Obenour et al., 2014).

2.4. Boosted regression tree analysis

Boosted regression trees (BRT) are a non-parametric machine learning technique that has proven effective in capturing complex biophysical relationships that are difficult to discern when using classical statistical approaches (Bhatt et al., 2013; Elith et al., 2008; Jouffray et al., 2015; Leathwick et al., 2008). BRT combines regression tree algorithms - models that recursively split response variables into homogeneous groups defined by threshold values of explanatory variables (Breiman et al., 1984) - with boosting, a method for combining models to enhance predictive performance (Friedman et al., 2000; Schapire, 2003). Classification and regression tree analysis has been widely applied to identify nonlinear, interactive relationships among variables in ecological datasets, and to quantify critical thresholds of environmental drivers that trigger ecosystem responses (Cleveland et al., 2011; Cottenie, 2005; Fernandez et al., 2006; Hambright et al., 2015), including cyanobacteria blooms (Chen and Mynett, 2004; Huber et al., 2012; Taranu et al., 2015; Wagner and Adrian, 2009). Despite the advantages that regression tree algorithms offer in terms of ability to model discontinuities and interactions, incorporate different types of response and explanatory variables, and effectively select relevant predictors (De'ath and Fabricius, 2000; De'ath, 2002), individual trees tend to show relatively low robustness to small variations in the training dataset, weak predictive performance, and inadequacy in modeling smooth functions (Hastie et al., 2001; Murtaugh, 2009). Boosting algorithms improve these limitations by fitting a large number of trees to the data and then combining them to generate more accurate and robust predictions (De'ath, 2007; Elith et al., 2008; Friedman and Meulman, 2003).

Compared with similar techniques that are based on fitting and averaging results from many trees, such as bagged trees and random forests (Cutler et al., 2007; Prasad et al., 2006), boosting algorithms differ in that they fit trees sequentially to the data through a stage-wise procedure (Elith et al., 2008). Specifically, trees are fitted iteratively to the training data, so that at each step a tree is added that minimizes the overall model prediction error. Each newly added tree is fitted to the residuals of the previous collection of trees, thereby progressively focusing on unexplained variability in the response (Elith et al., 2008; Hastie et al., 2001). At each step the fitted values are calculated as a linear combination of the trees fitted so far, so that the final BRT model can be viewed as an additive model where each tree represents a model term (Friedman et al., 2000).

BRTs have been extensively applied as an exploratory tool to compare the nature and relative importance of functional relationships between variables across different sets of ecological observations (Buston and Elith, 2011; Descy et al., 2016; Segurado et al., 2016; Tisseuil et al., 2012; Walsh and Webb, 2016). We fit BRT models to each set of CHAB observations separately to explore relationships between environmental drivers and bloom size and to assess the degree of coherence in modeling results across monitoring products. A detailed description of model parameterization and evaluation methods is

Table 3

Explanatory variable effects and contributions to each final BRT model. For each variable, the shape of the relationship with the response is shown (+: positive, -: negative, \cap or \cup : hump- or u-shaped) alongside an approximate threshold(s) that defines that relationship. Relative contributions of each variable are given as a percent and they are scaled so that the sum is 100. Only variables with a relative influence $\geq 5\%$ are shown. For each monitoring product, the optimal number of trees used to fit each final BRT model and the associated mean cross-validated R^2 are listed along with their standard deviation. Abbreviations as in Table 1 and Fig. 2.

	UT LEC	GLERL	NCCOS	MTRI MO SSI	MTRI MO TOT	MTRI SW TOT
# Observations	93	78	170	280	280	127
# Trees (SD)	17,990 (7980)	15,367 (9045)	1630 (269)	8281 (1537)	1127 (182)	3338 (568)
CV- R^2 (SD)	38.3% (2.5)	49.7% (2.6)	48.5% (1.6)	18.1% (1.7)	34.4% (1.2)	33.1% (1.9)
TP_JuneApril	-	+ 711	6%	-	-	-
TP_JuneFebruary	-	-	-	+ 1278	8%	-
TP_June	+ 45	8%	+ 95	6%	+ 1631	6%
Year	-	-	+ 1135	6%	+ 45	9%
Month	+ 7.5	6%	+ 8.5	6%	+ 4.6	12%
WaterTemp_30	+ 21	19%	+ 21.5	22%	+ 18	15%
Stress_2	\cup 0.02, 0.04	7%	-	-	-	6%
Stress_8	-	+ 0.03	5%	- 0.08	5%	-
Stress_30	-	-	-	-	-	+ 0.03
NortherlyWind_2	-	-	-	-	-	12%
NortherlyWind_8	-	-	-	-	-	+ 81
NortherlyWind_30	-	+ -0.2	8%	-	-	15%
WesterlyWind_2	-	-	-	-	-	-
WesterlyWind_8	-	-	-	-	-	-
WesterlyWind_30	-	-	-	-	-	-
Irradiance_2	\cap 215, 291	5%	-	\cap 131, 289	5%	-
Irradiance_8	\cap 210, 243	5%	\cap 236, 252	6%	-	-
Irradiance_30	-	-	-	-	-	-
Discharge_2	-	-	-	-	-	-
Discharge_30	- 105	11%	- 33	6%	-	-

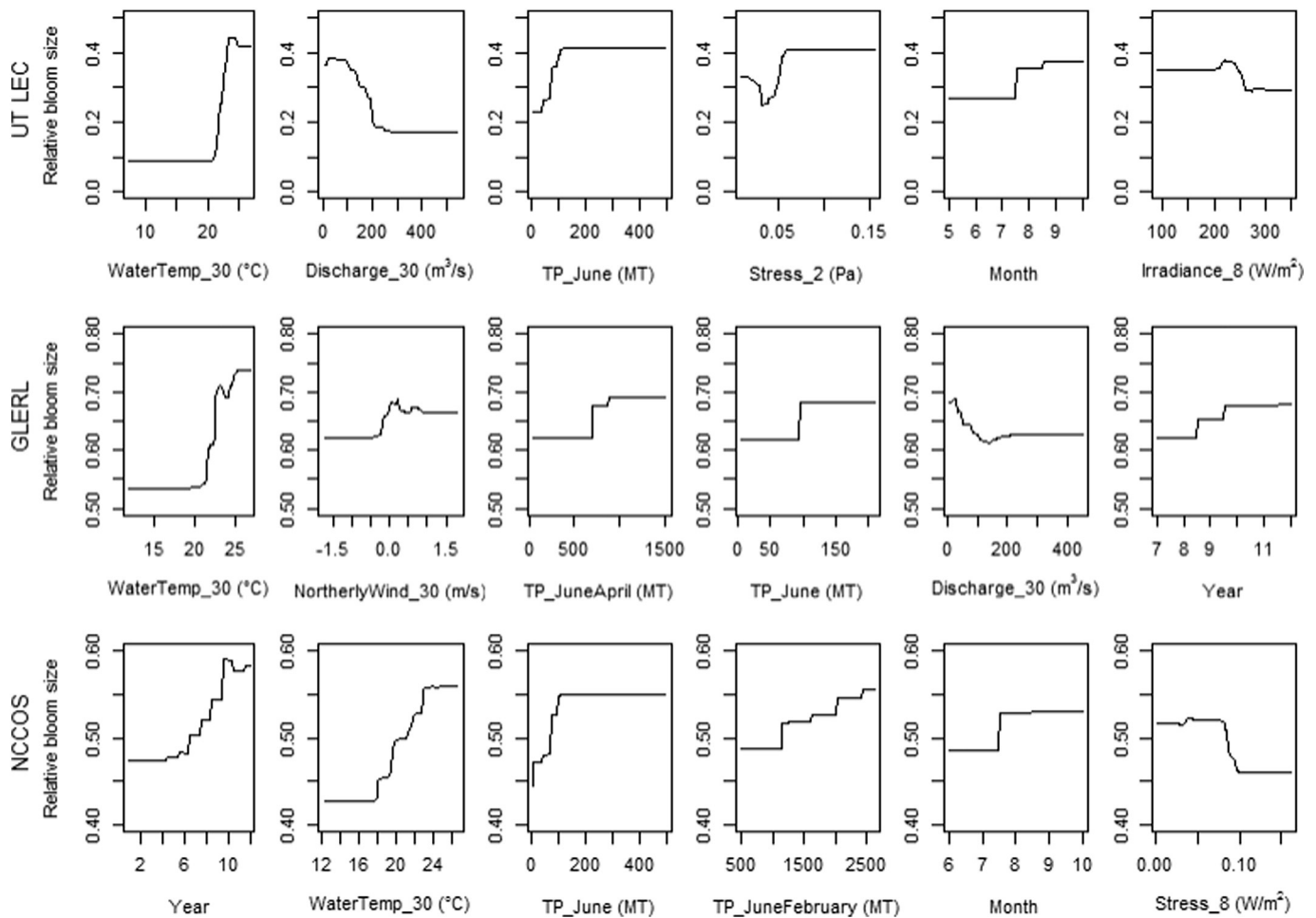


Fig. 3. Partial dependency plots for UT LEC, GLERL, and NCCOS. For each monitoring product, the six most influential variables are shown in decreasing order of relative importance (see Table 3). Bloom size has been scaled to 1 at the maximum value observed by each monitoring product to facilitate comparisons of predictor effect size across models. Abbreviations as in Table 1 and Fig. 2.

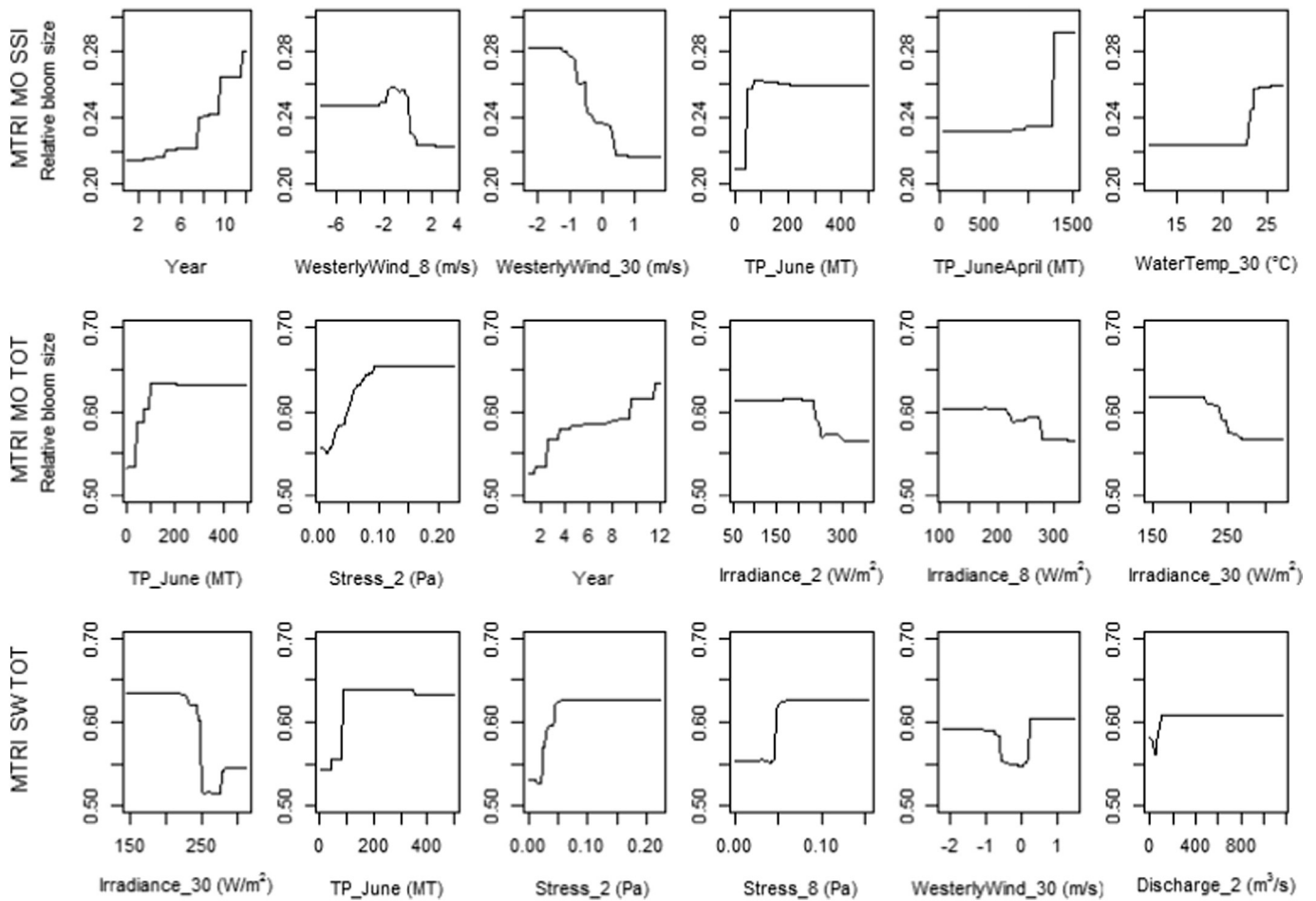


Fig. 4. Partial dependency plots for MTRI MO SSI, MTRI MO TOT, and MTRI SW TOT. For each monitoring product, the six most influential variables are shown in decreasing order of relative importance (see Table 3). Bloom size has been scaled to 1 at the maximum value observed by each monitoring product to facilitate comparisons of predictor effect size across models. Abbreviations as in Table 1 and Fig. 2.

low bloom sizes in the MO SSI model (Fig. 4), while SW TOT showed an u-shaped relationship with WesterlyWind_30 (Fig. 4).

The influence of irradiance was $\geq 5\%$ in all models. According to the MO TOT and SW TOT models, high irradiance values were mostly associated with monotonic decreases in bloom size, with the exception of a somewhat u-shaped relationship found between Irradiance_30 and SW TOT estimates (Fig. 4). The UT LEC, GLERL, and NCCOS models exhibited a hump-shaped relationship of bloom size with irradiance (Figs. 3 and 4).

3.2.3. Temporal trend

The yearly temporal trend was selected as the most important predictor for MO SSI and NCCOS (Table 3). A positive yearly temporal trend in bloom size was evident in all but the UT LEC model, although the trend begins in different years according to different monitoring products (Table 3). The monthly trend had a relative contribution $\geq 5\%$ in the UT LEC and NCCOS models, with a positive threshold in mid-July (Month = 7.5; Table 3).

3.2.4. Two-way interactions

Models fitted to different monitoring products identified different sets of two-way interactions as most influential, with limited consistency across products (Tables S3–S5). A positive interactive effect between WaterTemp_30 and nutrient loading was ranked among the strongest interactions by several models (Tables S3–S5). The NCCOS and MO SSI models showed high relative strength for interactions involving the yearly trend and meteorological predictors, while MO SSI also selected

multiple interactions between wind direction variables and nutrient loading (Tables S3–S5).

4. Discussion

4.1. Coherent patterns across CHAB monitoring products

Most monitoring products exhibit an increase in bloom size in recent years (Fig. 2), and most models show a positive yearly temporal trend, independent of other predictors (Table 3). This is consistent with previous findings suggesting a gradual increase in the lake's susceptibility to CHAB formation over the past decade (Obenour et al., 2014), similar to what has been observed in numerous lakes worldwide (Paerl and Paul, 2012; Taranu et al., 2015). Some of the models also exhibit highly ranked interactions between the yearly trend and multiple meteorological factors, such as a positive interaction between WaterTemp_30 and Year and between Irradiance_2 and Year for the NCCOS model (Fig. S4). This further suggests that the system's response to external drivers might have changed over time, although differences in the interacting factors across models make it difficult to formulate hypotheses on the potential underlying mechanisms.

Most monitoring products also confirm the importance of nutrient loading as a driver of relative variability in CHAB size, although the high correlation among TP, DRP, and TN loads did not allow us to identify a preferred nutrient predictor. Bioassay studies during western Lake Erie blooms indicate that phosphorus is often the limiting nutrient for algal growth, although nitrogen limitation can occur in late summer

(Chaffin et al., 2013), and recent work has highlighted the role of nitrogen in promoting the growth of toxic strains of *Microcystis* in Lake Erie (Gobler et al., 2016; Harke et al., 2016). The role of phosphorus loads in stimulating CHABs in western Lake Erie has been shown by multiple studies (Bertani et al., 2016; Michalak et al., 2013; Obenour et al., 2014; Stumpf et al., 2016; Stumpf et al., 2012; Verhamme et al., in review), and TP and DRP load targets have recently been set for the Maumee River by the United States and Canada under the Great Lakes Water Quality Agreement (GLWQA) Amendment of 2012 (United States and Canada, 2015). Specifically, a spring (Mar–Jul) TP loading target of 860 MT and a corresponding DRP loading target of 186 MT were recommended to significantly reduce CHAB formation based on results from a multiple modeling approach (Scavia et al., 2016). The thresholds identified for the April–June TP load by the GLERL model (711 MT, ~237 MT/month) and for the February–June TP load by the NCCOS model (1135 MT, ~227 MT/month) are comparable to the TP load thresholds determined for the Mar–Jul period by most cyanobacteria models used in the GLWQA effort (125–308 MT/month; Scavia et al., 2016). The higher TP_{JuneApril} and TP_{JuneFebruary} thresholds extracted by the MO SSI model might be due to the fact that this product only records surface scum, a phenomenon that does not appear to be strictly proportional to overall bloom size. In fact, when comparing time series of total bloom size (e.g., MO TOT and NCCOS) with the scum-only MO SSI product, a few years show high total bloom size and relatively large spring TP loads, but comparably low surface scum occurrence (e.g., 2003, 2008, 2010; Figs. 2a and S1).

Despite the positive relationships generally found between bloom size and nutrient loads, the fact that different load time windows were selected as influential predictors by different monitoring products suggests that inconsistencies in inter-annual variability in bloom size might lead to different conclusions on the most critical loading period (Table 3).

Most monitoring products show a positive relationship between bloom size and average water temperature in the preceding 30 days, whereas shorter-lagged temperature variables are never selected as influential predictors (Table 3). The temperature threshold triggering an increase in bloom size is higher for the scum model (~23 °C) than for other models (~18–21 °C), suggesting that higher critical meteorological thresholds might need to be crossed for surface scum to form. Especially high surface water temperatures develop during prolonged low-mixing, calm summer conditions, which are known to enhance scum formation (Ibelings et al., 2003; Soranno, 1997). All models show maximum bloom size when WaterTemp_30 values exceed 23–25 °C, which is consistent with results of a Lake Erie-specific study based on the GLERL time series (Millie et al., 2014) and more generally with observations of *Microcystis* growth rates reaching maxima at temperatures ≥ 25 °C (Butterwick et al., 2005; Reynolds, 2006).

The consistent positive relationship between bloom size and WaterTemp_30 indicates a marked seasonality in bloom development (Wynne and Stumpf, 2015). Ho and Michalak (2015) compared bloom seasonal timing in western Lake Erie using the NCCOS and UT LEC 2002–2011 time series, and they found that the average timing of bloom onset was around mid-July for both monitoring products, which agrees with the positive threshold identified for the monthly trend by the UT LEC and NCCOS models (Monthly = 7.5, i.e. mid-July, Table 3) and with a recent in-depth synthesis of the NCCOS time series (Wynne and Stumpf, 2015). All other models identified the same seasonal timing associated with an increase in bloom size, although the contribution to model fit was below 5% for most models, likely due to the fact that WaterTemp_30 explains a large portion of bloom seasonality.

The positive interaction between WaterTemp_30 and nutrient loading emerging from multiple models agrees with the known ecology of *Microcystis* blooms, which typically peak when high nutrient inputs are combined with favorable physical conditions in late summer (Davis et al., 2009; Taranu et al., 2012).

The negative relationship between bloom size and river discharge exhibited by the two *in-situ* monitoring products (Table 3), whose stations are under the direct influence of the Maumee river plume (Fig. 1), is consistent with the notion of cyanobacteria growth being favored by lower summer flushing rates and higher residence time (Elliott, 2010; Huber et al., 2012; Michalak et al., 2013). The u-shaped relationship found between river flow in the previous two days and SW TOT might be due to occasional high bloom estimates exhibited by this product during high-discharge events resulting in increased water column turbidity and potential bloom size overestimation (see next section). River discharge is used to calculate nutrient loads, so that one might expect a similar effect of these two predictors in the models. However, while we included nutrient loading in the model as cumulative values over different spring periods, river discharge was included as average values over the days preceding each bloom estimate throughout the summer. As a result, while the load variables provide an estimate of the overall amount of nutrients delivered in spring, the discharge variables quantify the effect of intra-seasonal changes in river hydrology on bloom size during the CHAB season. The effect of river hydrology is particularly relevant in the context of potential future climate change-driven shifts in the timing, frequency, and intensity of extreme weather events, such as storms and droughts (Michalak et al., 2013).

4.2. Discrepancies across CHAB monitoring products

Relationships between bloom size and wind stress were less consistent across the various monitoring products. NCCOS estimates exhibit a negative relationship with wind stress, with a decrease at Stress_8 > 0.08 Pa (Table 3). In a previous study relating NCCOS bloom estimates and wind stress in western Lake Erie, Wynne et al. (2010) reported that wind stress > 0.1 Pa resulted in a decrease in remotely sensed bloom size due to mixing of cyanobacteria cells through the water column and subsequent reduction in satellite-detectable near-surface concentrations. The relationship between bloom size and wind stress might therefore be influenced by the inherent constraint that remote sensing only detects near-surface cyanobacteria accumulations, especially in turbid, eutrophic freshwater (Kutser et al., 2008; Wynne et al., 2010). The coupling of *in-situ* surveys with remote sensing approaches may overcome this limitation because samples collected deeper in the water column may be less strongly affected by wind stress. In any case, the relationship between CHABs and wind is expected to be complex, because while *Microcystis* is typically favored by high water column stability, and loses its competitive advantage over non-buoyant phytoplankton under well-mixed conditions (Ibelings et al., 2003; Jöhnk et al., 2008; Visser et al., 1996), relatively short wind-induced mixing events preceding calm conditions might have a positive effect on overall cyanobacteria abundance by enhancing re-suspension of nutrients and overwintering cells (Chaffin et al., 2014; Preston et al., 1980; Verspagen et al., 2005).

The positive relationships with wind stress found for other remote sensing products (i.e., MO TOT and SW TOT) contrast with NCCOS results. Differences in the estimated effect of wind forcing may be due to discrepancies in short-term intra-seasonal bloom dynamics across monitoring products (Fig. S2). One illustrative example is represented by the year 2004. In June and September 2004, MO TOT and SW TOT show short-lived peaks in bloom size that do not appear in the NCCOS product (Fig. S2), resulting in the MO TOT and SW TOT 2004 annual bloom size being larger than that recorded in 2003 (Fig. 2a). Inherent differences in the characteristics of the two remote sensing algorithms (CPA-A for MO and SW TOT products versus CI for NCCOS) likely contribute to these and other similar short-term discrepancies. While NCCOS data quantify only the cyanobacterial component of the phytoplankton assemblage, MO and SW TOT estimate total chlorophyll, which includes other algal groups. Although high chlorophyll levels during the bloom season have generally been dominated by *Microcystis* spp. in recent years (Bridgeman et al., 2013; Millie et al., 2014), the

contribution of eukaryotic phytoplankton to total chlorophyll can be substantial especially in the beginning and the end of the CHAB season (Bridgeman et al., 2012; Millie et al., 2009), thereby leading to higher chlorophyll values estimated by MO TOT and SW TOT relative to NCCOS. Indeed, when comparing the three satellite products in terms of bloom areal extent only (not shown), without quantifying chlorophyll content, some of the observed discrepancies are mitigated (see Shuchman, 2014). Different sensors and algorithms may also respond differently when large sediment plumes affect water turbidity (Park et al., 2010; Wynne et al., 2013), as is often the case in the area influenced by the Maumee River during high discharge or high wind events. For example, changes in the optical properties of water associated with high sediment concentrations may result in occasional overestimation of chlorophyll-*a* by the CPA-A algorithm. Accordingly, select high river flow events (e.g., 2003, 2004, and 2007) or peaks in wind stress (e.g., 2005, 2006) are associated with relatively higher bloom sizes detected by MO TOT and SW TOT compared to NCCOS (Fig. S2). This might also explain the positive relationship found between SW TOT estimates and Discharge₂ > 60 m³/s. MODIS and SeaWiFS bands may also incur signal saturation under high surface scum (Reinart and Kutser, 2006; Wynne et al., 2013), which led to the development of the SSI algorithm. However, chlorophyll concentration estimates for surface scum are provisional (see Materials and methods section), increasing the uncertainty in overall bloom intensity estimates when large scums are present. On the other hand, the NCCOS time series is made of composite estimates obtained by combining images from multiple days, so that temporal variability is reduced, potentially affecting relationships with short-lagged environmental drivers. The procedure adopted to generate the NCCOS composites, based on summing the maximum biomass values recorded at each pixel over 10-day periods, may also be prone to occasional bloom extent overestimation when the bloom is transported to a different region of the basin between subsequent satellite images. The time of the day when satellites pass over the lake (~10 am for MERIS and ~2 pm for MODIS AQUA and SeaWiFS) might also affect detected near-surface bloom size, due to relatively rapid daily vertical migration patterns typically observed for buoyant cyanobacteria (Ibelings et al., 1991; Kromkamp and Mur, 1984; Wynne et al., 2013). On a calm day in western Lake Erie cyanobacteria tend to form surface scum accumulations in mid-day (~11 am) and start to subside in the early afternoon (~2 pm), until no scum is typically present by 6 pm (T. Bridgeman, pers. comm.).

Discrepancies in intra-seasonal bloom dynamics also occur when comparing remote sensing vs. *in-situ* monitoring products (Fig. S2). In the case of western Lake Erie, inconsistencies in bloom temporal dynamics between *in-situ* and satellite estimates have been partly attributed to the proximity of *in-situ* sampling areas to the region of bloom initiation (Ho and Michalak, 2015). More generally, the high spatial patchiness that typically characterizes cyanobacteria blooms is expected to affect CHAB severity estimates obtained from *in-situ* vs. satellite approaches differently. With respect to spatial coverage, satellite products represent an obvious enhancement over *in-situ* sampling, which is typically limited to a few point stations over a restricted portion of the lake. For example, extremely high chlorophyll concentrations measured at one GLERL site near the mouth of the Maumee in 2008 exert a disproportionate influence on overall bloom size estimates. Although more sophisticated spatial integration methods than that used here could improve *in-situ* estimates, the limitations associated with treating measurements from a limited number of stations as representative of a 300 km² area remain salient. The limited spatial coverage of *in-situ* surveys may also result in CHAB size underestimation when the bloom extends beyond or is transported away from the sampling area, such as in 2011 and 2013 in Lake Erie (Ho and Michalak, 2015). On the other hand, some satellite sensors tend to exhibit signal saturation when large patches of extremely high cyanobacteria biomass or sediment are present (Kutser, 2004; Reinart and Kutser, 2006; Wynne et al., 2013), potentially contributing to discrepancies in

the characterization of CHAB seasonal dynamics when compared with *in-situ* measurements.

Some inconsistencies in bloom temporal dynamics occur also across *in-situ* monitoring approaches (Fig. S2). Although UT LEC and GLERL sites cover a similar portion of the lake in terms of areal extent, the somewhat different location of the sampling areas (Fig. 1) might lead to differences in the observed CHAB seasonal dynamics due to bloom horizontal movements and high spatial variability (Ho and Michalak, 2015). Differences in sampling techniques might also affect *in-situ* bloom estimates. For instance, surface cyanobacteria accumulations are better captured through vertical plankton tows (UT LEC) than by sampling below the surface with a Niskin bottle (GLERL). Accordingly, the GLERL product shows less pronounced relative seasonal peaks compared to UT LEC during years with high scum occurrence (e.g., 2011 and 2013; Fig. S2). As mentioned above, the lack of taxonomic specificity of total chlorophyll measurements might also lead to inconsistencies when compared with taxon-specific surveys in cases when *Microcystis* is not the dominant taxon. Methodological differences across *in-situ* sampling approaches often depend on the original intended purpose of the monitoring programs and associated research questions. For example, GLERL surveys were initially designed to provide ground-truth information for satellite algorithm calibration. Similarly, the EPA sampling program in western Lake Erie is part of a broader limnological program initiated to monitor water quality in the offshore waters of the Great Lakes. As a result, EPA stations are spread across the open waters of the western basin, including regions that are typically not impacted by CHABs, and only two stations overlap with the GLERL and UT LEC sampling areas (Fig. 1). Although the original scope of some *in-situ* monitoring programs might limit their application to characterize absolute bloom size, they still provide valuable information on inter-annual variability in relative bloom intensity to validate, integrate, and augment CHAB estimates from other long-term surveys. In this perspective, some of the *in-situ* and remote sensing datasets considered here have been integrated in recent CHAB modeling approaches to support model calibration or validation (Verhamme et al. in review; Obenour et al., 2014).

In general, method-specific limitations in characterizing bloom variability need to be taken into careful consideration when using CHAB monitoring data to infer relationships with environmental drivers. Discrepancies in short-term bloom temporal dynamics across monitoring products may affect the modeled relationships with variables characterized by marked fluctuations on hourly to daily time scales, such as wind, leading to contrasting conclusions on their overall effect on bloom size (Table 3). The development of quantitative approaches to integrate multiple types of bloom observations in CHAB modeling efforts may help overcome the specific limitations of individual monitoring methods and improve bloom characterization and modeling capabilities.

The negative effect of high irradiance levels estimated by some models may suggest that photo-inhibition might play a role in influencing bloom seasonal dynamics (Ibelings and Maberly, 1998). However, while photo-inhibition of surface blooms has been occasionally observed in Lake Erie (Chaffin et al., 2012), *Microcystis* is known to possess several mechanisms to adapt to high levels of irradiance (Paerl and Otten, 2013b; Paerl et al., 1985; Sommaruga et al., 2009), and the relatively high turbidities often observed in western Lake Erie provide further protection against severe photodamage (Chaffin et al., 2012). The different irradiance thresholds identified by different models (Table 3) make it difficult to formulate hypotheses on the irradiance levels that might be detrimental for blooms in western Lake Erie, and more research is needed on the possible influence of high irradiance on bloom dynamics, especially in view of its potential to promote toxic strain dominance in *Microcystis* populations (Paerl and Otten, 2013b).

The scum-only (MO SSI) model explains a lower portion of data variability when compared to other remote sensing products (Table 3). Surface scum formation is often a rapid and transient phenomenon, whose accurate characterization requires an adequately high sampling

frequency. Remote sensing theoretically offers higher temporal resolution than most long-term *in-situ* monitoring programs. However, limitations in image availability or quality due to meteorological conditions often cause multi-day gaps that hinder the characterization of ephemeral surface bloom events. In addition, surface scum occurrence has increased only in the most recent years of the available time period (Fig. 2a), resulting in a limited number of data points on which to train the models. Interestingly, the scum-only model shows the largest contribution of wind direction, including several interactive effects between wind direction and various other predictors (Tables 3 and S3–S5). The horizontal distribution and areal coverage of surface scum might be affected by wind direction in a more direct way than subsurface blooms (Wynne et al., 2011). However, given the limitations outlined above, and considering that the scum concentration estimates are provisional and highly uncertain, extra caution should be taken in interpreting results describing scum dynamics.

In relation to this, it is important to note that there are uncertainties associated with some of the assumptions required to estimate total bloom size from the monitoring data (see Materials and methods section), and that these uncertainties are not explicitly accounted for in the modeling analysis. While we followed established workflows to calculate bloom size (e.g., Stumpf et al., 2012; Bridgeman et al., 2013) wherever possible and focused our analyses on relative bloom size rather than absolute bloom size, enhancements in bloom detection technologies under development in western Lake Erie may help quantify and reduce the uncertainty associated with some of those assumptions in the future. For example, the MTRI SSI algorithm is currently being enhanced to provide quantitative estimates of scum concentrations. A monitoring product capable of quantifying surface scum accumulations is an especially important tool for water resource managers because surface scums directly impede recreational uses due to distasteful visual and odor cues (Fig. S3). A long-term dataset quantifying surface scum areal coverage and concentration also has the potential to improve our ability to understand and model environmental conditions promoting scum development. The increasing deployment of advanced *in-situ* and remote sensing bloom monitoring technologies, such as *in-situ* water quality sensors and aerial hyperspectral sensors, will also offer an unprecedented level of spatio-temporal resolution to augment bloom estimates from more traditional monitoring approaches (Bullerjahn et al., 2016).

5. Conclusions

Different CHAB monitoring products exhibit a high degree of coherence in indicating that high nutrient inputs, high water temperature, and low flushing rates are most conducive to bloom development. The associated thresholds extracted by the BRT models are consistent with *Microcystis* ecology and with recent phosphorus load recommendations derived from a multiple models approach.

However, the influence of environmental drivers characterized by high short-term variability, such as wind forcing, appears less clear, and conflicting results emerge when fitting the same model to different monitoring products. This is likely due to the combination of high spatio-temporal variability of cyanobacteria blooms and method-specific limitations in capturing such variability at the appropriate spatial and temporal scales. Such discrepancies have implications for understanding functional relationships between environmental factors and CHAB formation, and ultimately for developing models capable of informing resource managers with accurate intra-seasonal bloom predictions. While previous studies have provided qualitative assessments of the degree of inter-comparability of different CHAB monitoring approaches in Lake Erie, to our knowledge this is the first study applying a quantitative modeling approach to synthesize and compare multiple time series of CHAB estimates. The monitoring approaches considered in this study are routinely used to

track and forecast CHABs in lakes worldwide, making the implications of our results relevant beyond the study lake.

Our analysis underscores the importance and challenges associated with comparing multiple types of bloom measurements. Scientists and resource managers are still faced with significant uncertainties in characterizing bloom variability, especially at fine temporal scales. Future efforts should focus on implementing rigorous methods to generate bloom size estimates that systematically integrate information from multiple data sources. The development of methods that assimilate data from multiple CHAB monitoring methods has the potential to allow for more accurate bloom detection by leveraging the advantages associated with each monitoring approach while overcoming the respective limitations. Enhancing our ability to quantify bloom distribution and intensity is a key step for both informing resource managers of bloom spatio-temporal impacts and for improving models aimed at advancing our predictive understanding of harmful algal blooms.

Acknowledgements

This work was funded by the University of Michigan Water Center and by the National Science Foundation (NSF) under grant 1313897. We thank Richard Stumpf and Timothy Wynne for providing the NCCOS satellite images. Support for UT LEC data collection was provided over various periods by: The Ohio Sea Grant College Program, Project R/ER-72, under grant NA16RG2252 from the National Sea Grant College Program, the Ohio Lake Erie Commission, Lake Erie Protection Fund Projects #03-19, #0-08 and #SG406-11, USEPA project #GL2009-1-2, the University of Cincinnati, the USDA Natural Resources Conservation Service (MVR&D), and the NSF under grant 1039043.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.scitotenv.2016.10.023>.

References

- Anderson, D.M., Glibert, P.M., Burkholder, J.M., 2002. Harmful algal blooms and eutrophication: nutrient sources, composition, and consequences. *Estuaries* 25, 704–726.
- Auret, L., Aldrich, C., 2011. Empirical comparison of tree ensemble variable importance measures. *Chemom. Intell. Lab. Syst.* 105, 157–170. <http://dx.doi.org/10.1016/j.chemolab.2010.12.004>.
- Baker, D.B., Confesor, R., Ewing, D.E., Johnson, L.T., Kramer, J.W., Merryfield, B.J., 2014. Phosphorus loading to Lake Erie from the Maumee, Sandusky and Cuyahoga rivers: the importance of bioavailability. *J. Great Lakes Res.* 40, 502–517. <http://dx.doi.org/10.1016/j.jglr.2014.05.001>.
- Beletsky, D., Hawley, N., Rao, Y.R., 2013. Modeling summer circulation and thermal structure of Lake Erie. *J. Geophys. Res. Ocean.* 118, 6238–6252. <http://dx.doi.org/10.1002/2013JC008854>.
- Bertani, I., Obenour, D., Steger, C., Stow, C., Gronewold, A., Scavia, D., 2016. Probabilistically assessing the role of nutrient loading in harmful algal bloom formation in western Lake Erie. *J. Great Lakes Res.* <http://dx.doi.org/10.1016/j.jglr.2016.04.002>.
- Bhatt, S., Gething, P.W., Brady, O.J., Messina, J.P., Farlow, A.W., Moyes, C.L., Drake, J.M., Brownstein, J.S., Hoen, A.G., Sankoh, O., Myers, M.F., George, D.B., Jaenisch, T., Wint, G.R.W., Simmons, C.P., Scott, T.W., Farrar, J.J., Hay, S.I., 2013. The global distribution and burden of dengue. *Nature* 496, 504–507. <http://dx.doi.org/10.1038/nature12060>.
- Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.G., 1984. *Classification and Regression Trees*. Chapman and Hall/CRC.
- Bridgeman, T.B., Chaffin, J.D., Kane, D.D., Conroy, J.D., Panek, S.E., Armenio, P.M., 2012. From river to lake: phosphorus partitioning and algal community compositional changes in Western Lake Erie. *J. Great Lakes Res.* 38, 90–97. <http://dx.doi.org/10.1016/j.jglr.2011.09.010>.
- Bridgeman, T.B., Chaffin, J.D., Filbrun, J.E., 2013. A novel method for tracking western Lake Erie *Microcystis* blooms, 2002–2011. *J. Great Lakes Res.* 39, 83–89. <http://dx.doi.org/10.1016/j.jglr.2012.11.004>.
- Brittain, S.M., Wang, J., Babcock-Jackson, L., Carmichael, W.W., Rinehart, K.L., Culver, D.A., 2000. Isolation and characterization of microcystins, cyclic heptapeptide hepatotoxins from a Lake Erie strain of *Microcystis aeruginosa*. *J. Great Lakes Res.* 26, 241–249.
- Brookes, J.D., Carey, C.C., 2011. Resilience to blooms. *Science* 334 (80–), 46–47. <http://dx.doi.org/10.1126/science.1207349>.
- Bullerjahn, G.S., McKay, R.M., Davis, T.W., Baker, D.B., Boyer, G.L., Anglada, L.V.D., Doucette, G.J., Ho, J.C., Irwin, E.G., Kling, C.L., Kudela, R.M., Kurmayer, R., Michalak, A.M., Ortiz, J.D., Otten, T.G., Paerl, H.W., Qin, B., Sohngen, B.L., Stumpf, R.P., Visser, P.M., Wilhelm, S.W., 2016. Global solutions to regional problems: collecting global

- expertise to address the problem of harmful cyanobacterial blooms. A Lake Erie case study. *Harmful Algae* 54, 223–238. <http://dx.doi.org/10.1016/j.hal.2016.01.003>.
- Buston, P.M., Elith, J., 2011. Determinants of reproductive success in dominant pairs of clownfish: a boosted regression tree analysis. *J. Anim. Ecol.* 80, 528–538. <http://dx.doi.org/10.1111/j.1365-2656.2011.01803.x>.
- Butterwick, C., Heaney, S.I., Talling, J.F., 2005. Diversity in the influence of temperature on the growth rates of freshwater algae, and its ecological relevance. *Freshw. Biol.* 50, 291–300. <http://dx.doi.org/10.1111/j.1365-2427.2004.01317.x>.
- Carmichael, W.W., Boyer, G.L., 2016. Health impacts from cyanobacteria harmful algae blooms: implications for the North American Great Lakes. *Harmful Algae* 54, 194–212. <http://dx.doi.org/10.1016/j.hal.2016.02.002>.
- Chaffin, J.D., Bridgeman, T.B., Heckathorn, S.A., Krause, A.E., 2012. Role of suspended sediments and mixing in reducing photoinhibition in the bloom-forming cyanobacterium *Microcystis*. *J. Water Resour. Prot.* 4, 1029–1041. <http://dx.doi.org/10.4236/jwarp.2012.412119>.
- Chaffin, J.D., Bridgeman, T.B., Bade, D.L., 2013. Nitrogen constrains the growth of late summer cyanobacterial blooms in Lake Erie. *Adv. Microbiol.* 16–26.
- Chaffin, J., Sigler, V., Bridgeman, T., 2014. Connecting the blooms: tracking and establishing the origin of the record-breaking Lake Erie *Microcystis* bloom of 2011 using DGGE. *Aquat. Microb. Ecol.* 73, 29–39. <http://dx.doi.org/10.3354/ame01708>.
- Chen, Q., Mynett, A.E., 2004. Predicting *Phaeocystis globosa* bloom in Dutch coastal waters by decision trees and nonlinear piecewise regression. *Ecol. Model.* 176, 277–290. <http://dx.doi.org/10.1016/j.ecolmodel.2003.10.031>.
- Chen, M., Li, J., Dai, X., Sun, Y., Chen, F., 2011. Effect of phosphorus and temperature on chlorophyll a contents and cell sizes of *Scenedesmus obliquus* and *Microcystis aeruginosa*. *Limnology* 12, 187–192. <http://dx.doi.org/10.1007/s10201-010-0336-y>.
- Chorus, I., Bartram, J., 1999. *Toxic Cyanobacteria in Water: a Guide to Their Public Health Consequences, Monitoring and Management*. E & FN Spon, London.
- Cleveland, C.C., Townsend, A.R., Taylor, P., Alvarez-Clare, S., Bustamante, M.M.C., Chuyong, G., Dobrowski, S.Z., Grierson, P., Harms, K.E., Houlton, B.Z., Marklein, A., Parton, W., Porder, S., Reed, S.C., Sierra, C.A., Silver, W.L., Tanner, E.V.J., Wieder, W.R., 2011. Relationships among net primary productivity, nutrients and climate in tropical rain forest: a pan-tropical analysis. *Ecol. Lett.* 14, 939–947. <http://dx.doi.org/10.1111/j.1461-0248.2011.01658.x>.
- Codd, G.A., Lindsay, J., Young, F.M., Morrison, L.F., Metcalf, J.S., 2005a. Harmful cyanobacteria – from mass mortalities to management measures. In: Huisman, J., Matthijs, H., Visser, P.M. (Eds.), *Harmful Cyanobacteria*. Springer, Dordrecht.
- Codd, G.A., Morrison, L.F., Metcalf, J.S., 2005b. Cyanobacterial toxins: risk management for health protection. *Toxicol. Appl. Pharmacol.* 203, 264–272. <http://dx.doi.org/10.1016/j.taap.2004.02.016>.
- Cottenie, K., 2005. Integrating environmental and spatial processes in ecological community dynamics. *Ecol. Lett.* 8, 1175–1182. <http://dx.doi.org/10.1111/j.1461-0248.2005.00820.x>.
- Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. Random forests for classification in ecology. *Ecology* 88, 2783–2792. <http://dx.doi.org/10.1890/07-0539.1>.
- Dale, B., Murphy, M., 2014. A retrospective appraisal of the importance of high-resolution sampling for harmful algal blooms: lessons from long-term phytoplankton monitoring at Sherkin Island, S.W. Ireland. *Harmful Algae* 40, 23–33. <http://dx.doi.org/10.1016/j.hal.2014.09.007>.
- Davis, T.W., Berry, D.L., Boyer, G.L., Gobler, C.J., 2009. The effects of temperature and nutrients on the growth and dynamics of toxic and non-toxic strains of *Microcystis* during cyanobacteria blooms. *Harmful Algae* 8, 715–725. <http://dx.doi.org/10.1016/j.hal.2009.02.004>.
- De'ath, G., 2002. Multivariate regression trees: a new technique for modeling species-environment relationships. *Ecology* 83, 1105–1117.
- De'ath, G., 2007. Boosted trees for ecological modeling and prediction. *Ecology* 88, 243–251. [http://dx.doi.org/10.1890/0012-9658\(2007\)88\[243:BTFFEMA\]2.0.CO;2](http://dx.doi.org/10.1890/0012-9658(2007)88[243:BTFFEMA]2.0.CO;2).
- De'ath, G., Fabricius, K.E., 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology* 81, 3178–3192.
- DePinto, J.V., Young, T.C., McLroy, L.M., 1986. Great Lakes water quality improvement. *Environ. Sci. Technol.* 20, 752–759. <http://dx.doi.org/10.1021/es00150a001>.
- Descy, J.-P., Leprieur, F., Pirlot, S., Leporcq, B., Van Wichelen, J., Peretyatko, A., Teissier, S., Codd, G.A., Triest, L., Vyverman, W., Wilmotte, A., 2016. Identifying the factors determining blooms of cyanobacteria in a set of shallow lakes. *Eco. Inform.* 34, 129–138. <http://dx.doi.org/10.1016/j.ecoinf.2016.05.003>.
- Dodds, W.K., Bouska, W.W., Eitzmann, J.L., Pilger, T.J., Pitts, K.L., Riley, A.J., Schloesser, J.T., Thornbrugh, D.J., 2009. Eutrophication of U.S. freshwaters: analysis of potential economic damages. *Environ. Sci. Technol.* 43, 12–19.
- Dolan, D.M., 1993. Point source loadings of phosphorus to Lake Erie. *J. Great Lakes Res.* 1986–1990.
- Downing, J.A., Watson, S.B., McCauley, E., 2001. Predicting cyanobacteria dominance in lakes. *Can. J. Fish. Aquat. Sci.* 58, 1905–1908. <http://dx.doi.org/10.1139/r01-143>.
- Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. *J. Anim. Ecol.* 77, 802–813. <http://dx.doi.org/10.1111/j.1365-2656.2008.01390.x>.
- Elliott, J.A., 2010. The seasonal sensitivity of cyanobacteria and other phytoplankton to changes in flushing rate and water temperature. *Glob. Chang. Biol.* 16, 864–876. <http://dx.doi.org/10.1111/j.1365-2486.2009.01998.x>.
- Fernandez, D.P., Neff, J.C., Belnap, J., Reynolds, R.L., 2006. Soil respiration in the cold desert environment of the Colorado Plateau (USA): abiotic regulators and thresholds. *Biogeochemistry* 78, 247–265. <http://dx.doi.org/10.1007/s10533-005-4278-0>.
- Friedman, J.H., Meulman, J.J., 2003. Multiple additive regression trees with application in epidemiology. *Stat. Med.* 22, 1365–1381. <http://dx.doi.org/10.1002/sim.1501>.
- Friedman, J., Hastie, T., Tibshirani, R., 2000. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *Ann. Stat.* 28, 337–407.
- Gobler, C.J., Burkholder, J.M., Davis, T.W., Harke, M.J., Johengen, T., Stow, C.A., Van De Waal, D.B., 2016. The dual role of nitrogen supply in controlling the growth and toxicity of cyanobacterial blooms. *Harmful Algae* 54, 87–97. <http://dx.doi.org/10.1016/j.hal.2016.01.010>.
- Hambright, D.K., Beyer, J.E., Easton, J.D., Zamor, R.M., Easton, A.C., Hallidayschult, T.C., 2015. The niche of an invasive marine microbe in a subtropical freshwater impoundment. *ISME J.* 9, 256–264. <http://dx.doi.org/10.1038/ismej.2014.103>.
- Harke, M.J., Davis, T.W., Watson, S.B., Gobler, C.J., 2016. Nutrient-controlled niche differentiation of western Lake Erie cyanobacterial populations revealed via metatranscriptomic surveys. *Environ. Sci. Technol.* 50, 604–615. <http://dx.doi.org/10.1021/acs.est.5b03931>.
- Hastie, T., Tibshirani, R., Friedman, J.H., 2001. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Verlag, New York.
- Havens, K.E., 2008. Cyanobacteria blooms: effects on aquatic ecosystems. In: Hudnell, H.K. (Ed.), *Cyanobacterial Harmful Algal Blooms - State of the Science and Research Needs*. Springer.
- Havens, K.E., Philips, E.J., Cichra, M.F., Li, B.L., 1998. Light availability as a possible regulator of cyanobacteria species composition in a shallow subtropical lake. *Freshw. Biol.* 39, 547–556. <http://dx.doi.org/10.1046/j.1365-2427.1998.00308.x>.
- Ho, J.C., Michalak, A.M., 2015. Challenges in tracking harmful algal blooms: a synthesis of evidence from Lake Erie. *J. Great Lakes Res.* <http://dx.doi.org/10.1016/j.jglr.2015.01.001>.
- Hsu, S.A., 1974. Experimental results of the drag-coefficient estimation for air-coast interfaces. *Bound.-Layer Meteorol.* 6, 505–507.
- Huber, V., Wagner, C., Gerten, D., Adrian, R., 2012. To bloom or not to bloom: contrasting responses of cyanobacteria to recent heat waves explained by critical thresholds of abiotic drivers. *Oecologia* 169, 245–256. <http://dx.doi.org/10.1007/s00442-011-2186-7>.
- Huisman, J., Jonker, R.R., Zonneveld, C., Weissing, F.J., 1999. Competition for light between phytoplankton species: experimental tests of mechanistic theory. *Ecology* 80, 211–222. [http://dx.doi.org/10.1890/0012-9658\(1999\)080\[0211:CFLBPS\]2.0.CO;2](http://dx.doi.org/10.1890/0012-9658(1999)080[0211:CFLBPS]2.0.CO;2).
- Ibelings, B.W., Maberly, S.C., 1998. Photoinhibition and the availability of inorganic carbon restrict photosynthesis by surface blooms of cyanobacteria. *Limnol. Oceanogr.* 43, 408–419. <http://dx.doi.org/10.4319/lo.1998.43.3.0408>.
- Ibelings, B.W., Mur, L.R., Walsby, a.E., 1991. Diurnal changes in buoyancy and vertical-distribution in populations of *Microcystis* in 2 Shallow Lakes. *J. Plankton Res.* 13, 419–436.
- Ibelings, B.W., Vonk, M., Los, H.F.J., Van Der Molen, D.T., Mooij, W.M., 2003. Fuzzy modeling of cyanobacterial surface waterblooms: validation with NOAA-AVHRR satellite images. *Ecol. Appl.* 13, 1456–1472. <http://dx.doi.org/10.1890/01-5345>.
- IJC, 2014. *A Balanced Diet for Lake Erie: Reducing Phosphorus Loadings and Harmful Algal Blooms. Report of the Lake Erie Ecosystem Priority*.
- Jöhnk, K.D., Huisman, J., Sharples, J., Sommeijer, B., Visser, P.M., Stroom, J.M., 2008. Summer heatwaves promote blooms of harmful cyanobacteria. *Glob. Chang. Biol.* 14, 495–512. <http://dx.doi.org/10.1111/j.1365-2486.2007.01510.x>.
- Jouffray, J.-B., Nyström, M., Norström, A., Williams, I., Wedding, L., Kittinger, J., Williams, G., 2015. Identifying multiple coral reef regimes and their drivers across the Hawaiian archipelago. *Philos. Trans. R. Soc. B* 370, 20130268. <http://dx.doi.org/10.1098/rstb.2013.0268>.
- Kosten, S., Huszar, V.L.M., Bécares, E., Costa, L.S., Donk, E., Hansson, L.-A., Jeppesen, E., Kruk, C., Lacerot, G., Mazzeo, N., Meester, L., Moss, B., Lüring, M., Nöges, T., Romo, S., Scheffer, M., 2012. Warmer climates boost cyanobacterial dominance in shallow lakes. *Glob. Chang. Biol.* 18, 118–126. <http://dx.doi.org/10.1111/j.1365-2486.2011.02488.x>.
- Kromkamp, J.C., Mur, L.R., 1984. Buoyant density changes in the cyanobacterium *Microcystis aeruginosa* due to changes in the cellular carbohydrate content. *FEMS Microbiol. Lett.* 25, 105–109. [http://dx.doi.org/10.1016/0378-1097\(84\)90055-7](http://dx.doi.org/10.1016/0378-1097(84)90055-7).
- Kutser, T., 2004. Quantitative detection of chlorophyll in cyanobacterial blooms by satellite remote sensing. *Limnol. Oceanogr.* 49, 2179–2189. <http://dx.doi.org/10.4319/lo.2004.49.6.2179>.
- Kutser, T., Metsamaa, L., Vahtmäe, E., Strömbeck, N., 2006. Suitability of MODIS 250 m resolution band data for quantitative mapping of cyanobacterial blooms. *Proc. Est. Acad. Sci. Biol. Ecol.* 55, 318–328.
- Kutser, T., Metsamaa, L., Dekker, A.G., 2008. Influence of the vertical distribution of cyanobacteria in the water column on the remote sensing signal. *Estuar. Coast. Shelf Sci.* 78, 649–654. <http://dx.doi.org/10.1016/j.ecss.2008.02.024>.
- Leathwick, J.R., Elith, J., Chadderton, W.L., Rowe, D., Hastie, T., 2008. Dispersal, disturbance and the contrasting biogeographies of New Zealand's diadromous and non-diadromous fish species. *J. Biogeogr.* 35, 1481–1497. <http://dx.doi.org/10.1111/j.1365-2699.2008.01887.x>.
- Litchman, E., 1998. Population and community responses of phytoplankton to fluctuating light. *Oecologia* 117, 247–257. <http://dx.doi.org/10.1007/s004420050655>.
- Long, B.M., Jones, G.J., Orr, P.T., 2001. Cellular microcystin content in N-limited *Microcystis aeruginosa* can be predicted from growth rate. *Appl. Environ. Microbiol.* 67, 278–283. <http://dx.doi.org/10.1128/AEM.67.1.278>.
- Maccoux, M., Dove, A., Backus, S., Dolan, D., 2016. Total and soluble reactive phosphorus loadings to Lake Erie. *J. Great Lakes Res.* In press.
- Makarewicz, J.C., Bertram, P., 1991. Evidence for the restoration of the Lake Erie ecosystem. *Bioscience* 41, 216–223.
- Michalak, A.M., Anderson, E.J., Beletsky, D., Bolland, S., Bosch, N.S., Bridgeman, T.B., Chaffin, J.D., Cho, K., Confesor, R., Daloglu, I., Depinto, J.V., Evans, M.A., Fahnenstiel, G.L., He, L., Ho, J.C., Jenkins, L., Johengen, T.H., Kuo, K.C., Laporte, E., Liu, X., McWilliams, M.R., Moore, M.R., Posselt, D.J., Richards, R.P., Scavia, D., Steiner, A.L., Verhamme, E., Wright, D.M., Zagorski, M.A., 2013. Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions.

- Proc. Natl. Acad. Sci. U. S. A. 110, 6448–6452. <http://dx.doi.org/10.1073/pnas.1216006110>.
- Millie, D.F., Fahnenstiel, G.L., Dyble Bressie, J., Pigg, R.J., Rediske, R.R., Klarer, D.M., Tester, P.A., Litaker, R.W., 2009. Late-summer phytoplankton in western Lake Erie (Laurentian Great Lakes): bloom distributions, toxicity, and environmental influences. *Aquat. Ecol.* 43, 915–934. <http://dx.doi.org/10.1007/s10452-009-9238-7>.
- Millie, D.F., Weckman, G.R., Fahnenstiel, G.L., Carrick, H.J., Ardjmand, E., Young, I.I., W.A., Sayers, M.J., Shuchman, R.A., 2014. Using artificial intelligence for CyanoHAB niche modeling: discovery and visualization of *Microcystis*–environmental associations within western Lake Erie. *Can. J. Fish. Aquat. Sci.* 71, 1642–1654. <http://dx.doi.org/10.1139/cjfas-2013-0654>.
- Mitrovic, S.M., Oliver, R.L., Rees, C., Bowling, L.C., Buckley, R.T., 2003. Critical flow velocities for the growth and dominance of *Anabaena circinalis* in some turbid freshwater rivers. *Freshw. Biol.* 48, 164–174. <http://dx.doi.org/10.1046/j.1365-2427.2003.00957.x>.
- Murtaugh, P.A., 2009. Performance of several variable-selection methods applied to real ecological data. *Ecol. Lett.* 12, 1061–1068. <http://dx.doi.org/10.1111/j.1461-0248.2009.01361.x>.
- Nicodemus, K.K., Malley, J.D., 2009. Predictor correlation impacts machine learning algorithms: implications for genomic studies. *Bioinformatics* 25, 1884–1890. <http://dx.doi.org/10.1093/bioinformatics/btp331>.
- Obenour, D., Gronewold, A., Stow, C.A., Scavia, D., 2014. Using a Bayesian hierarchical model to improve Lake Erie cyanobacteria bloom forecasts. *Water Resour. Res.* 50, 7847–7860. <http://dx.doi.org/10.1002/2014WR015616>. Received.
- Paerl, H.W., Huisman, J., 2008. Blooms like it hot. *Science* 320 (80–), 57–58. <http://dx.doi.org/10.1126/science.1155398>.
- Paerl, H.W., Huisman, J., 2009. Climate change: a catalyst for global expansion of harmful cyanobacterial blooms. *Environ. Microbiol. Rep.*
- Paerl, H.W., Otten, T.G., 2013a. Harmful cyanobacterial blooms: causes, consequences, and controls. *Microb. Ecol.* 65, 995–1010. <http://dx.doi.org/10.1007/s00248-012-0159-y>.
- Paerl, H.W., Otten, T.G., 2013b. Blooms bite the hand that feeds them. *Science* 342 (80–), 433–434. <http://dx.doi.org/10.1126/science.1245276>.
- Paerl, H.W., Paul, V.J., 2012. Climate change: links to global expansion of harmful cyanobacteria. *Water Res.* 46, 1349–1363. <http://dx.doi.org/10.1016/j.watres.2011.08.002>.
- Paerl, H.W., Bland, P.T., Bowles, N.D., Haibach, M.E., 1985. Adaptation to high-intensity, low-wavelength light among surface blooms of the cyanobacterium *Microcystis aeruginosa*. *Appl. Environ. Microbiol.* 49, 1046–1052.
- Paerl, H.W., Hall, N.S., Calandrino, E.S., 2011. Controlling harmful cyanobacterial blooms in a world experiencing anthropogenic and climatic-induced change. *Sci. Total Environ.* 409, 1739–1745. <http://dx.doi.org/10.1016/j.scitotenv.2011.02.001>.
- Park, Y.-J., Ruddick, K., Lacroix, G., 2010. Detection of algal blooms in European waters based on satellite chlorophyll data from MERIS and MODIS. *Int. J. Remote Sens.* 31, 6567–6583.
- Perovich, G., Dortch, Q., Goodrich, J., Berger, P.S., Brooks, J., Evens, T.J., Gobler, C.J., Graham, J., Hyde, J., Karner, D., O'Shea, D.K., Paul, V., Paerl, H., Piehler, M., Rosen, B.H., Santelmann, M., Pat Tester, J.W., 2008. Causes, prevention, and mitigation. In: Hudnell, H.K. (Ed.), *Cyanobacterial Harmful Algal Blooms - State of the Science and Research Needs*. Springer.
- Prasad, A.M., Iverson, L.R., Liaw, A., 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems* 9, 181–199. <http://dx.doi.org/10.1007/s10021-005-0054-1>.
- Preston, T., Stewart, W.D.P., Reynolds, C.S., 1980. Bloom-forming cyanobacterium *Microcystis aeruginosa* overwinters on sediment surface. *Nature* 288, 365–367. <http://dx.doi.org/10.1038/288365a0>.
- Reinart, A., Kutser, T., 2006. Comparison of different satellite sensors in detecting cyanobacterial bloom events in the Baltic Sea. *Remote Sens. Environ.* 102, 74–85.
- Reynolds, C., 2006. *Ecology of Phytoplankton*. Cambridge University Press, New York.
- Richards, R.P., Baker, D.B., 2002. Trends in water quality in LEASEQ rivers and streams (northwestern Ohio), 1975–1995. Lake Erie agricultural systems for environmental quality. *J. Environ. Qual.* 31, 90–96. <http://dx.doi.org/10.2134/jeq2002.0090>.
- Richards, R.P., Baker, D.B., Crumrine, J.P., Kramer, J.W., Ewing, D.E., Merryfield, B.J., 2008. Thirty-year trends in suspended sediment in seven Lake Erie tributaries. *J. Environ. Qual.* 37, 1894–1908. <http://dx.doi.org/10.2134/jeq2007.0590>.
- Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., 1973. Monitoring vegetation systems in the Great Plains with ERTS. *Third ERTS Symposium, NASA SP-351 I*, pp. 309–317.
- Sayers, M., Fahnenstiel, G.L., Shuchman, R.A., Whitley, M., 2016. Cyanobacteria blooms in three eutrophic basins of the Great Lakes: a comparative analysis using satellite remote sensing. *Int. J. Remote Sens.* 37, 4148–4171.
- Scavia, D., David Allan, J., Arend, K.K., Bartell, S., Beletsky, D., Bosch, N.S., Brandt, S.B., Briland, R.D., Daloglu, I., DePinto, J.V., Dolan, D.M., Evans, M.A., Farmer, T.M., Goto, D., Han, H., Höök, T.O., Knight, R., Ludsin, S.A., Mason, D., Michalak, A.M., Peter Richards, R., Roberts, J.J., Rucinski, D.K., Rutherford, E., Schwab, D.J., Sesterhenn, T.M., Zhang, H., Zhou, Y., 2014. Assessing and addressing the re-eutrophication of Lake Erie: central basin hypoxia. *J. Great Lakes Res.* 40, 226–246. <http://dx.doi.org/10.1016/j.jglr.2014.02.004>.
- Scavia, D., DePinto, J.V., Bertani, I., 2016. A multi-model approach to evaluating target phosphorus loads for Lake Erie. *J. Great Lakes Res.* <http://dx.doi.org/10.1016/j.jglr.2016.09.007>.
- Schapiro, R.E., 2003. The boosting approach to machine learning - an overview. In: Denison, D.D., Hansen, M.H., Holmes, C.C., Mallick, B., Yu, B. (Eds.), *Nonlinear Estimation and Classification, Lecture Notes in Statistics*. Springer New York, New York, NY <http://dx.doi.org/10.1007/978-0-387-21579-2>.
- Segurado, P., Branco, P., Jauch, E., Neves, R., Ferreira, M.T., 2016. Sensitivity of river fishes to climate change: the role of hydrological stressors on habitat range shifts. *Sci. Total Environ.* 562, 435–445. <http://dx.doi.org/10.1016/j.scitotenv.2016.03.188>.
- Shen, L., Xu, H., Guo, X., 2012. Satellite remote sensing of harmful algal blooms (HABs) and a potential synthesized framework. *Sensors* 12, 7778–7803. <http://dx.doi.org/10.3390/s120607778>.
- Shuchman, R.A., 2014. *Harmful algal bloom mapping in the Great Lakes*. Michigan Tech Research Institute Final Report. July 1, 2011–June 30, 2014. Chicago, IL.
- Shuchman, R.A., Leshkevich, G., Sayers, M.J., Johengen, T., Brooks, C.N., Pozdnyakov, D., 2013. An algorithm to retrieve chlorophyll, dissolved organic carbon, and suspended minerals from great lakes satellite data. *J. Great Lakes Res.* 39, 78–89.
- Smith, V.H., 2003. Eutrophication of freshwater and coastal marine ecosystems: a global problem. *Environ. Sci. Pollut. Res.* 10, 126–139.
- Sommaruga, R., Chen, Y., Liu, Z., 2009. Multiple strategies of bloom-forming microcystis to minimize damage by solar ultraviolet radiation in surface waters. *Microb. Ecol.* 57, 667–674. <http://dx.doi.org/10.1007/s00248-008-9425-4>.
- Soranno, P.A., 1997. Factors affecting the timing of surface scum and epilimnetic blooms of blue-green algae in a eutrophic lake. *Can. J. Fish. Aquat. Sci.* 54, 1965–1975. <http://dx.doi.org/10.1139/cjfas-54-9-1965>.
- Srivastava, A., Singh, S., Ahn, C.Y., Oh, H.M., Asthana, R.K., 2013. Monitoring approaches for a toxic cyanobacterial bloom. *Environ. Sci. Technol.* 47, 8999–9013. <http://dx.doi.org/10.1021/es401245k>.
- Stumpf, R.P., Wynne, T.T., Baker, D.B., Fahnenstiel, G.L., 2012. Interannual variability of cyanobacterial blooms in Lake Erie. *PLoS One* 7, e42444. <http://dx.doi.org/10.1371/journal.pone.0042444>.
- Stumpf, R., Johnson, L., Wynne, T., Baker, D., 2016. Forecasting annual cyanobacterial bloom biomass to inform management decisions in Lake Erie. *J. Great Lakes Res.* <http://dx.doi.org/10.1016/j.jglr.2016.08.006>.
- Taranu, Z.E., Zurawell, R.W., Pick, F., Gregory-Eaves, I., 2012. Predicting cyanobacterial dynamics in the face of global change: the importance of scale and environmental context. *Glob. Chang. Biol.* 18, 3477–3490. <http://dx.doi.org/10.1111/gcb.12015>.
- Taranu, Z.E., Gregory-Eaves, I., Leavitt, P.R., Bunting, L., Buchaca, T., Catalan, J., Domaizon, I., Guilizzoni, P., Lami, A., McGowan, S., Moorhouse, H., Morabito, G., Pick, F.R., Stevenson, M.A., Thompson, P.L., Vinebrooke, R.D., 2015. Acceleration of cyanobacterial dominance in north temperate-subarctic lakes during the Anthropocene. *Ecol. Lett.* <http://dx.doi.org/10.1111/ele.12420>.
- Tisseuil, C., Vrac, M., Grenouillet, G., Wade, A.J., Gevrey, M., Oberdorff, T., Grodwohl, J., Lek, S., 2012. Strengthening the link between climate, hydrological and species distribution modeling to assess the impacts of climate change on freshwater biodiversity. *Sci. Total Environ.* 424, 193–201. <http://dx.doi.org/10.1016/j.scitotenv.2012.02.035>.
- United States, Canada, 2012. *Great Lakes Water Quality Agreement - Annex 4*.
- United States, Canada, 2015. *Recommended Phosphorus Loading Targets for Lake Erie - Annex 4 Objectives and Targets Task Team Final Report to the Nutrients Annex Subcommittee*.
- Vanderploeg, H.A., Liebig, J.R., Carmichael, W.W., Agy, M.A., Johengen, T.H., Fahnenstiel, G.L., Nalepa, T.F., 2001. Zebra mussel (*Dreissena polymorpha*) selective filtration promoted toxic *Microcystis* blooms in Saginaw Bay (Lake Huron) and Lake Erie. *Can. J. Fish. Aquat. Sci.* 58, 1208–1221.
- Verhamme, E., Redder, T., Schlea, D., Grush, J., Bratton, J., DePinto, J., n.d. Development of the Western Lake Erie Ecosystem Model (WLEEM): application to connect phosphorus loads to cyanobacteria biomass. *J. Great Lakes Res.* (in review).
- Verspagen, J.M.H., Snelder, E.O.F.M., Visser, P.M., Jöhnk, K.D., Ibelings, B.W., Mur, L.R., Huisman, J., 2005. Benthic-pelagic coupling in the population dynamics of the harmful cyanobacterium *Microcystis*. *Freshw. Biol.* 50, 854–867. <http://dx.doi.org/10.1111/j.1365-2427.2005.01368.x>.
- Verspagen, J.M.H., Passarge, J., Jöhnk, K.D., Visser, P.M., Boers, P., Laanbroek, H.J., Huisman, J., 2006. Water management strategies against toxic microcystis blooms in the Dutch Delta. *Ecol. Appl.* 16, 313–327.
- Visser, P.M., Ibelings, B.W., Van der Veer, B., Koedood, J., Mur, L.R., 1996. Artificial mixing prevents nuisance blooms of the cyanobacterium *Microcystis* in Lake Nieuwe Meer, the Netherlands. *Freshw. Biol.* 36, 435–450. <http://dx.doi.org/10.1046/j.1365-2427.1996.00093.x>.
- Wagner, C., Adrian, R., 2009. Cyanobacteria dominance: quantifying the effects of climate change. *Limnol. Oceanogr.* 54, 2460–2468. http://dx.doi.org/10.4319/lo.2009.54.6_part_2.2460.
- Wallace, B.B., Hamilton, D.P., 1999. The effect of variations in irradiance on buoyancy regulation in *Microcystis aeruginosa*. *Limnol. Oceanogr.* 44, 273–281. <http://dx.doi.org/10.4319/lo.1999.44.2.0273>.
- Walsh, C.J., Webb, J.A., 2016. Interactive effects of urban stormwater drainage, land clearance, and flow regime on stream macroinvertebrate assemblages across a large metropolitan region. *Freshw. Sci.* 35, 324–339. <http://dx.doi.org/10.1086/685105>.
- Wang, J., Xie, P., Guo, N., 2007. Effects of nonylphenol on the growth and microcystin production of *Microcystis* strains. *Environ. Res.* 103, 70–78. <http://dx.doi.org/10.1016/j.envres.2006.05.013>.
- Watson, S.B., Miller, C., Arhonditsis, G., Boyer, G.L., Carmichael, W., Charlton, M.N., Confesor, R., Depew, D.C., Ludsin, S.A., Matisoff, G., McElmurry, S.P., Ho, T.O., Murray, M.W., Richards, R.P., Rao, Y.R., Steffen, M.M., Wilhelm, S.W., 2016. The re-eutrophication of Lake Erie: harmful algal blooms and hypoxia. *Harmful Algae* 56, 44–66. <http://dx.doi.org/10.1016/j.hal.2016.04.010>.
- Wynne, T., Stumpf, R., 2015. Spatial and temporal patterns in the seasonal distribution of toxic cyanobacteria in Western Lake Erie from 2002–2014. *Toxins (Basel)* 7, 1649–1663. <http://dx.doi.org/10.3390/toxins7051649>.
- Wynne, T.T., Stumpf, R.P., Tomlinson, M.C., Warner, R.A., Tester, P.A., Dyble, J., Fahnenstiel, G.L., 2008. Relating spectral shape to cyanobacterial blooms in the Laurentian Great Lakes. *Int. J. Remote Sens.* 29, 3665–3672. <http://dx.doi.org/10.1080/01431160802007640>.

- Wynne, T.T., Stumpf, R.P., Tomlinson, M.C., Dyble, J., 2010. Characterizing a cyanobacterial bloom in Western Lake Erie using satellite imagery and meteorological data. *Limnol. Oceanogr.* 55, 2025–2036. <http://dx.doi.org/10.4319/lo.2010.55.5.2025>.
- Wynne, T., Stumpf, R., Tomlinson, M.L., Schwab, D., Watabayashi, G., Christensen, J., 2011. Estimating cyanobacterial bloom transport by coupling remotely sensed imagery and a hydrodynamic model. *Ecol. Appl.* 21, 2709–2721.
- Wynne, T.T., Stumpf, R., Briggs, T., 2013. Comparing MODIS and MERIS spectral shapes for cyanobacterial bloom detection. *Int. J. Remote Sens.* 34, 6668–6678. <http://dx.doi.org/10.1080/01431161.2013.804228>.
- Zhou, Y., Michalak, A.M., Beletsky, D., Rao, Y.R., Richards, R.P., 2015. Record-breaking Lake Erie hypoxia during 2012 drought. *Environ. Sci. Technol.* 49, 800–807. <http://dx.doi.org/10.1021/es503981n>.