



A space-time geostatistical model for probabilistic estimation of harmful algal bloom biomass and areal extent

Shiqi Fang^{a,*}, Dario Del Giudice^a, Donald Scavia^b, Caren E. Binding^c, Thomas B. Bridgeman^d, Justin D. Chaffin^e, Mary Anne Evans^f, Joseph Guinness^g, Thomas H. Johengen^h, Daniel R. Obenour^{a,i}

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

^b School for Environment and Sustainability, University of Michigan, 440 Church St., Ann Arbor, MI 48104, USA

^c Water Science and Technology Directorate, Environment and Climate Change Canada, 867 Lakeshore Rd, Burlington, Ontario L7S 1A1, Canada

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

^e F. T. Stone Laboratory and Ohio Sea Grant, The Ohio State University, 878 Bayview Ave, Put-in-Bay, OH 43456, USA

^f U.S. Geological Survey, Great Lakes Science Center, 1451 Green Rd, Ann Arbor, MI 48105, USA

^g Department of Statistics and Data Science, Cornell University, 1178 Comstock Hall, Ithaca, NY 14853, USA

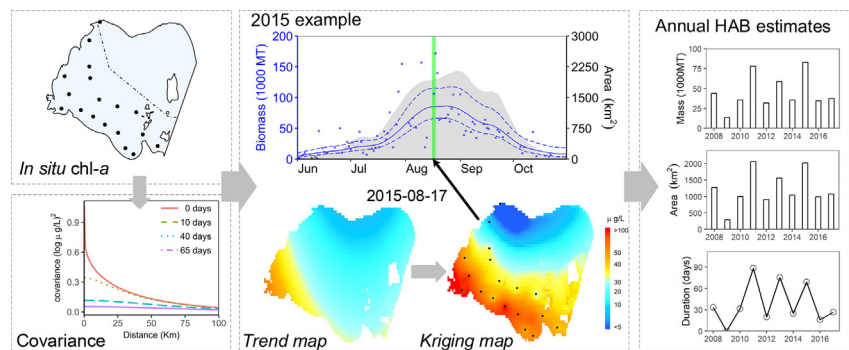
^h Cooperative Institute for Great Lakes Research (CIGLR), University of Michigan, 4840 South State Road, Ann Arbor, MI 48108, USA

ⁱ Center for Geospatial Analytics, North Carolina State University, Campus Box 7106, Raleigh, NC 27695, USA

HIGHLIGHTS

- We develop a space-time geostatistical approach for modeling algal bloom variability.
- Model synthesizes data from an international suite of monitoring programs.
- Trends characterize spatiotemporal patterns and effect of wind mixing on chl-*a*.
- Model simultaneously estimates bloom surface area and overall biomass.
- Results show potential to inform Lake Erie algal bloom monitoring design.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 26 April 2019

Received in revised form 2 August 2019

Accepted 3 August 2019

Available online 8 August 2019

Editor: Daniel Wunderlin

Keywords:

Harmful algal blooms

Space-time geostatistical model

Lake Erie

Probabilistic estimates

Algal biomass and extent

ABSTRACT

Harmful algal blooms (HABs) have been increasing in intensity worldwide, including the western basin of Lake Erie. Substantial efforts have been made to track these blooms using *in situ* sampling and remote sensing. However, such measurements do not fully capture HAB spatial and temporal dynamics due to the limitations of discrete shipboard sampling over large areas and the effects of clouds and winds on remote sensing estimates. To address these limitations, we develop a space-time geostatistical modeling framework for estimating HAB intensity and extent using chlorophyll *a* data sampled during the HAB season (June–October) from 2008 to 2017 by five independent monitoring programs. Based on the Bayesian information criterion for model selection, trend variables explain bloom northerly and easterly expansion from Maumee Bay, wind effects over depth, and variability among sampling methods. Cross validation results demonstrate that space-time kriging explains over half of the variability in daily, location-specific chlorophyll observations, on average. Conditional simulations provide, for the first time, comprehensive estimates of overall bloom biomass (based on depth-integrated concentrations) and surface areal extent with quantified uncertainties. These new estimates are contrasted with previous Lake

* Corresponding author.

E-mail address: sfang6@ncsu.edu (S. Fang).

Erie HAB monitoring studies, and deviations among estimates are explored and discussed. Overall, results highlight the importance of maintaining sufficient monitoring coverage to capture bloom dynamics, as well as the benefits of the proposed approach for synthesizing data from multiple monitoring programs to improve estimation accuracy while reducing uncertainty.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

Harmful algal blooms (HABs) are a growing global threat to ecosystems and human health (Huisman et al., 2018). HABs threaten ecosystems and drinking water through surface scum formation, increased risk of hypoxia, food web alteration, and potential toxin production. Ingestion of cyanobacterial HAB-related toxins, the most well-studied being microcystins, have been linked to numerous health issues, including liver and skin diseases (Paerl and Paul, 2012; Torbick et al., 2018). In August 2014, microcystins contaminated drinking water and prompted a “do not drink” advisory for over 400,000 residents near western Lake Erie (Steffen et al., 2017).

Cyanobacterial HABs have become larger and more persistent over the last decade in western Lake Erie (Bertani et al., 2017; Michalak et al., 2013; Sayers et al., 2019), as they have in many northern lakes (Taranu et al., 2015). To address HABs in Lake Erie, Canada and the United States have committed to “undertake and share research, monitoring and modeling necessary to establish, report on, and assess the management of phosphorus and other nutrients and improve the understanding of relevant issues associated with nutrients and excessive algal blooms” (United States and Canada, 2012). This emphasis on HAB management has motivated development of multiple models that relate bloom size (e.g., biomass) to nutrient loading (Ho and Michalak, 2017; Scavia et al., 2016; Verhamme et al., 2016), and the suite of models is now used for HAB forecasts (NOAA, 2019). These forecasting models rely on robust estimates of nutrient loading and HAB size for calibration and verification. However, unlike the nutrient load, which is determined from intensive daily monitoring of the Maumee River (Jarvie et al., 2017), less intense and disparate monitoring protocols leave large uncertainties and discrepancies among estimates of HAB size (Bertani et al., 2016; Bertani et al., 2017).

In Lake Erie, most large-scale HAB estimates are derived from remote sensing (Sayers et al., 2019; Stumpf et al., 2016a). However, these estimates are limited by cloud cover, the limited spatial and temporal resolution of specific satellite sensors, and algorithm retrieval uncertainties brought about by suspended sediments and adjacency effects (Matthews, 2011). Furthermore, remote sensing only determines near-surface HAB concentrations, such that overall HAB estimation may be biased depending on how meteorological conditions influence the position of the bloom within the water column (Wynne et al., 2018; Wynne et al., 2011). Efforts to overcome these limitations have included the use of composite images (Stumpf et al., 2003; Tomlinson et al., 2004) and rolling averages (Manning et al., 2019) to estimate annual maximum HAB size.

In situ observations have also been used to quantify HAB size expressed, for instance, as algal biomass (Bridgeman et al., 2013; Chaffin et al., 2011; Kane et al., 2014), and to validate remote sensing algorithms (Wynne et al., 2008). However, *in situ* observations are laborious and costly for large spatial and temporal coverage (Page et al., 2018). Fortunately, with several public and private sector groups collecting samples in recent years, temporal and spatial coverage has increased. Bertani et al. (2017) applied a machine learning approach to compare the temporal variability of *in situ* bloom measurements from various sampling programs and how this variability related to potential environmental drivers (e.g., wind). Golnick et al. (2016) explored differences among the sampling methods and analytical procedures of various monitoring programs, showing potential discrepancies among datasets, particularly when surface cyanobacterial scums were present.

While these studies indicate some of the challenges associated with working with data from multiple monitoring programs, there remains the need to synthesize these data within a coherent framework to improve HAB estimation (Ho et al., 2017; U.S. Congress, 2017).

Spatio-temporal correlation analysis and geostatistical modeling is a promising approach for addressing HAB estimation challenges. Geostatistics has been shown to outperform simpler interpolation approaches (Murphy et al., 2010; Varouchakis and Hristopoulos, 2013), while quantifying estimation uncertainty (Chilès and Delfiner, 2012). However, as with all statistical methods, the accuracy and precision of geostatistical estimates are strongly dependent on the quantity and quality of the available sampling data. Geostatistical modeling has been applied to characterize hypoxia (Obenour et al., 2013; Zhou et al., 2014) and phytoplankton distributions (Ludovisi et al., 2005; Wang and Liu, 2005; Zhao and Cai, 2004) in lakes and coastal areas. However, such models have typically focused only on spatial estimation. In contrast, space-time geostatistics incorporates temporal trends and temporal correlation to allow for probabilistic estimation through time (Kyriakidis and Journel, 1999). The utility of space-time approaches has been demonstrated for modeling water quality in river networks (Holbach et al., 2014; Money et al., 2009; Money et al., 2011) and more recently for hypoxia in the Gulf of Mexico (Matli et al., 2018).

In this study, we propose a space-time framework for integrating multiple *in situ* HAB datasets to characterize spatio-temporal chlorophyll *a* (chl-*a*) dynamics in western Lake Erie. While chl-*a* is not specific to cyanobacteria, high summertime chl-*a* concentrations in Lake Erie are indicative of cyanobacterial blooms (Newell et al., 2019). The unique approach developed here simultaneously estimates surface-layer and water column-integrated chl-*a* concentrations to account for the propensity of *Microcystis* to rise in the water column under quiescent conditions (Bosse et al., 2019; Rowe et al., 2016). Specific objectives include (a) develop probabilistic estimates of bloom biomass and areal extent throughout June to October of each year by integrating ten years (2008–2017) of *in situ* chl-*a* data from an international group of sampling programs; (b) assess how different sampling approaches and wind speeds impact chl-*a* biomass estimation; and (c) inform HAB monitoring network design. In addition, we evaluate the new HAB estimates based on their uncertainties (related to sampling coverage) and how they compare with the findings of previous Lake Erie studies.

2. Materials and methods

2.1. Study area and field methods

The western basin of Lake Erie (WBLE, approximately 3000 km²) is relatively shallow compared to the rest of Lake Erie, averaging just 7.4 m in depth, and is thus subject to intermittent stratification and wind-driven mixing (Wynne et al., 2011). In this study, the eastern boundary of the WBLE was defined as a line from Marblehead, Ohio to Point Pelee, Canada (Fig. 1a). Chl-*a* is a common algal surrogate for algal biomass and often used as a reference for HABs (Steinman et al., 2017; Welschmeyer, 1994). We synthesized chl-*a* concentration measurements from five ongoing water quality monitoring programs (Fig. 1b).

The following five *in situ* monitoring programs provide chl-*a* observations over the bloom season (June–October, 2008–2017): NOAA (National Oceanic and Atmospheric Administration, Great Lakes Environmental Research Laboratory), OSU (The Ohio State University,

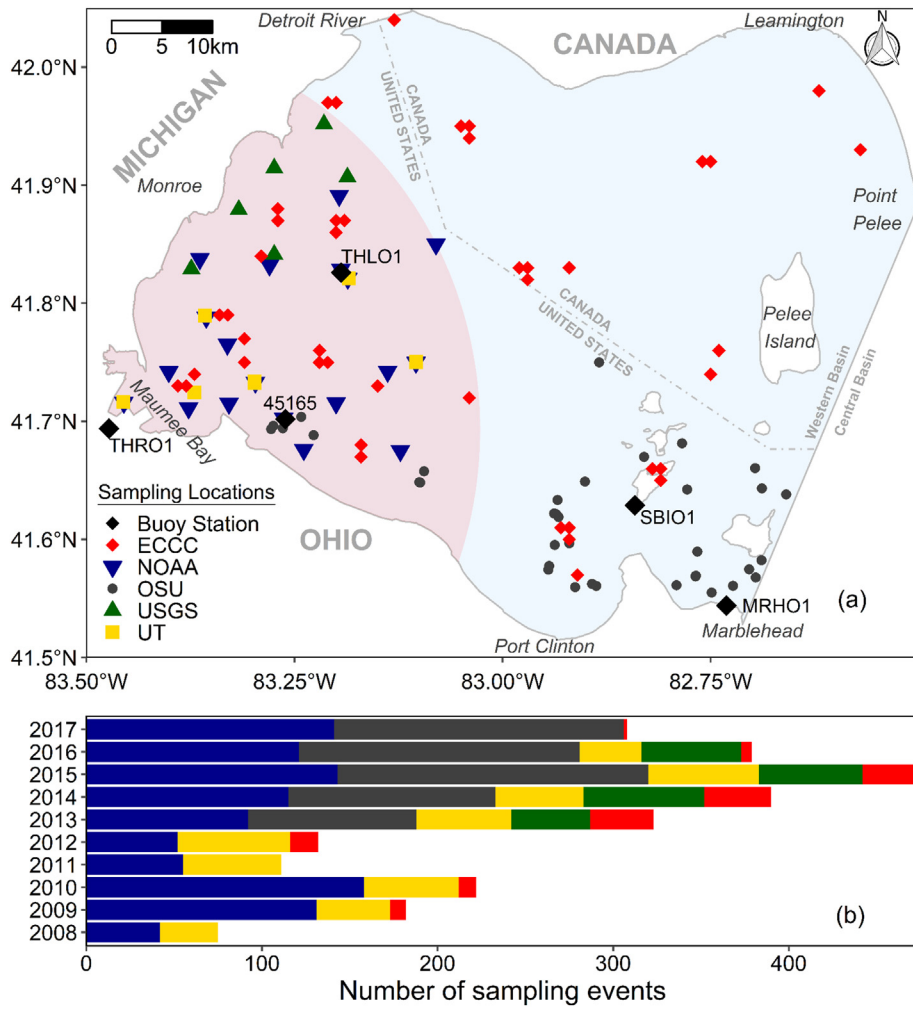


Fig. 1. (a) Map of the western basin of Lake Erie (WBLE) with *in situ* sampling locations shown. Monitoring programs include NOAA (National Oceanic and Atmospheric Administration, Great Lakes Environmental Research Laboratory), OSU (The Ohio State University, Stone Laboratory), USGS (the U.S. Geological Survey, Great Lakes Science Center), UT (University of Toledo, Lake Erie Center), and ECCC (Environment and Climate Change Canada). The near-Maumee region is the pink circle with a radius of 37 km from Buoy Station THRO1. (b) Count of all sampling events from different programs from 2008 to 2017, with colors corresponding to program legend in the map. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Stone Laboratory), USGS (the U.S. Geological Survey, Great Lakes Science Center), UT (University of Toledo, Lake Erie Center), and ECCC (Environment and Climate Change Canada). The programs routinely performed ship-based cruises to collect water samples, where each cruise contained multiple sampling “events” conducted at specific locations and times. NOAA has been monitoring water quality at several sites in WBLE since the early 2000s. The surveys have been carried out between late spring and early fall at intervals of 7 to 14 days. NOAA sampling covers an area of about 600 km² with roughly 7 km between stations. These stations are located within 37 km of the Maumee River mouth, referred to as the “near-Maumee” region in this paper (Fig. 1a pink area). Surface samples were collected with a 0.75 m long vertical Niskin bottle centered at a depth of 0.75 m below the water surface. Since 2015, NOAA has additionally collected bottom samples.

In contrast to NOAA, the other monitoring programs primarily use an “integrated” water-column sampling method (Golnick et al., 2016). Integrated sampling uses a tube sampler that typically collects water from the surface to 1 m above the sediments, or up to 8 m maximum depth. The USGS collects samples in this way, weekly to biweekly from May to October in the vicinity of Monroe, Michigan (Fig. 1a). The spacing between sampling stations is approximately 7 km over an area of approximately 100 km². The UT sampling area partially overlapped with NOAA, and integrated samples were collected biweekly from May to October at a spatial interval of about 8 km. ECCC integrated

sampling covered almost the whole WBLE with sparse measurements (>35 km interval typical spacing) and only conducted large cruises in 2013–2015. OSU focused sampling on Sandusky Bay and the south shoreline of WBLE, conducting sampling more frequently (twice a week or weekly) and at higher spatial resolution (about 4 km interval) than the other programs. The OSU dataset also includes sampling by a volunteer group of charter boat captains. These volunteer samples are integrated over the upper 2 m of the water column and thus treated as surface samples in this study. We also note that USGS occasionally collected samples over the upper 2 m of the water column (at times of strong stratification), and these samples are also treated as surface samples. Each monitoring program had different analytical methods to measure chl-*a* concentrations, but all used filtration to capture phytoplankton on a glass fiber filter and an organic solvent to extract chl-*a* from cells (Golnick et al., 2016).

In addition, we used ancillary data such as bathymetry and wind speed to check the potential chl-*a* response to mixing conditions. Bathymetry data were acquired from NOAA’s National Centers for Environmental Information (<https://www.ngdc.noaa.gov>). Hourly wind speeds from the National Buoy Data Center (<https://www.ndbc.noaa.gov>) for stations SBIO1, THLO1, and THRO1 were adjusted to a consistent 10 m height following the method described by Beletsky et al. (2003). The adjusted point-location data were used to spatially interpolate wind speeds across the WBLE through inverse distance weighting.

2.2. Space-time geostatistical model

In the space-time geostatistical model, observed values of chl-*a* concentration $\mathbf{z}(x, y, t)$ are considered as a realization of a spatio-temporal random field where, x, y, t represent longitude, latitude, and time, respectively. The space-time geostatistical model (Gelfand et al., 2010) can be represented as:

$$\mathbf{z}(x, y, t) = \boldsymbol{\mu}(x, y, t) + \boldsymbol{\eta}(x, y, t) + \boldsymbol{\varepsilon}(x, y, t) \quad (1)$$

Eq. (1) is similar to multiple linear regression (MLR) but includes an additional term $\boldsymbol{\eta}(x, y, t)$ representing spatio-temporally correlated stochasticity. The term $\boldsymbol{\mu}(x, y, t)$ is a deterministic space-time trend function, which can be calculated through $\mathbf{X}\boldsymbol{\beta}$, where \mathbf{X} is a matrix of explanatory variables such as bathymetry and sampling type, and $\boldsymbol{\beta}$ is a vector of estimated regression coefficients. In addition, $\boldsymbol{\varepsilon}(x, y, t)$ is the uncorrelated error term following a Gaussian distribution $N(0, \tau^2)$, where τ^2 represents the spatially uncorrelated residual variance, due to microvariability and measurement error (also called “nugget effect”). To help ensure the residuals were normally distributed, the chl-*a* data were log-transformed (Fig. S1). In this study, $\boldsymbol{\eta}(x, y, t)$ is a Gaussian space-time process with mean zero and a Matérn covariance function (Stein, 2005), parameterized as:

$$\text{Cov}(z(x_i, y_i, t_i), z(x_j, y_j, t_j)) = \begin{cases} \Sigma_{ij} = \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}}(d_{ij})^\nu K_\nu(d_{ij}), d_{ij} > 0 \\ \sigma^2 + \tau^2, d_{ij} = 0 \end{cases} \quad (2)$$

where $\Gamma(\nu)$ is a gamma function, K_ν is a modified Bessel function of the second kind (Weniger and Cížek, 1990), ν is the smoothness parameter, σ^2 is the spatially correlated variance (also known as “partial sill”). Matérn covariance with $\nu = 0.5$ and $\nu \rightarrow \infty$ coincide with the exponential and Gaussian covariance functions, respectively (Gelfand et al., 2010). Also, d_{ij} is the space-time distance calculated as:

$$d_{ij} = \sqrt{\frac{s_{i,j}^2}{\alpha_1^2} + \frac{t_{i,j}^2}{\alpha_2^2}} \quad (3)$$

where, for any pair of observations, $z(x_i, y_i, t_i)$ and $z(x_j, y_j, t_j)$, $s_{i,j}$ is the separation distance (UTM coordinate system), $t_{i,j}$ is the time lag, and α_1 and α_2 represent the space (km) and time (d) scaling parameters, similar to Matli et al. (2018). Thus, day of year is considered as an additional dimension in the space-time coordinate system. The effective spatial (km) and temporal (d) ranges of correlation are determined by the distance or time at which covariance is reduced to 5% of the partial sill (Gelfand et al., 2010).

To efficiently estimate the five unknown covariance parameters ($\theta = (\sigma^2, \alpha_1, \alpha_2, \nu, \tau^2)$) we maximize Vecchia’s likelihood (Vecchia, 1988), which is an approximation of restricted maximum likelihood (Gelfand et al., 2010) with advantages of computational scalability and accuracy (Katzfuss and Guinness, 2017). Once the covariance parameters are estimated, the corresponding best estimate of trend parameters ($\hat{\boldsymbol{\beta}}$) is determined using generalized least squares (Obenour et al., 2012):

$$\hat{\boldsymbol{\beta}} = \mathbf{X}^T \boldsymbol{\Sigma}_{oo}^{-1} \mathbf{X}^{-1} \mathbf{X}^T \boldsymbol{\Sigma}_{oo}^{-1} \mathbf{z} \quad (4)$$

where $\boldsymbol{\Sigma}_{oo}$ is the covariance matrix among observations, populated using Eq. (2).

2.3. BIC-based variable selection

To prevent overparameterization of the model, we used the Bayesian information criterion (BIC) (Schwarz, 1978) to select a subset of trend variables for the deterministic component of Eq. (1). BIC considers both the goodness of fit and the dimensionality (i.e., the number of

variables) of the model, wherein lower BIC scores generally indicate better models. Our model selection framework followed a multistep procedure. For computational efficiency, we first selected the ten best variable subsets (with smallest BIC) by constructing models using conventional MLR. The trend variables corresponding to these ten best subsets were then used to construct ten versions of the geostatistical model, accounting for the spatio-temporal correlation among observations. The overall best model was then selected based on a geostatistical adaptation of BIC (Huang et al., 2007; Obenour et al., 2012).

Candidate trend variables included spatial coordinates computed as distances (easterly and northerly) from Buoy Station THRO1 (Fig. 1a), day of year from June to October, water column depth, and wind speed. Because most sampling cruises were completed by 15:00, we considered wind averaging periods (6, 12, 18, and 24 h) leading up to 15:00. We also considered categorical variables allowing for differences between the two main sampling types (i.e., integrated versus surface sampling) and a potential downward bias in USGS chl-*a* measurements relative to the other program laboratories (Golnick et al., 2016). Additionally, as some studies suggested a threshold wind speed for effective algal mixing in Lake Erie (Rowe et al., 2016; Wynne et al., 2010), we also considered wind as a categorical variable to test whether this representation would outperform the use of wind speed as a continuous variable (described above).

Some higher order trend variables were also considered in the model. Because preliminary data analysis indicated that chl-*a* typically peaks in late summer (Fig. S2), we allowed quadratic temporal trends in the model. Similarly, because remote sensing indicated nonlinear spatial patterns in chl-*a* (Stumpf et al., 2016b), we considered both northerly and easterly quadratic trends in the candidate models. Moreover, two-way interaction terms among depth, wind, and sample type were considered to expand understanding of the relationships among these potential variables.

2.4. Conditional simulation for HAB estimation

The covariance function and BIC-selected variables were used to determine the geostatistical kriging weights across all estimation locations. For a time (t_0), a unique set of weights ($\boldsymbol{\Lambda}_{e,t}$) were determined by solving the following system of linear equations (Matli et al., 2018):

$$\begin{bmatrix} \boldsymbol{\Sigma}_{oo} & \mathbf{X}_o \\ \mathbf{X}_o^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda}_{e,t} \\ -\mathbf{G}_{e,t} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{oe,t} \\ \mathbf{X}_{e,t}^T \end{bmatrix} \quad (5)$$

where $\boldsymbol{\Sigma}_{oo}$ is an $n \times n$ covariance matrix for n observations and $\boldsymbol{\Sigma}_{oe,t}$ is an $n \times m$ covariance matrix among observations and the m estimation locations, with elements determined from Eq. (2). \mathbf{X}_o is the $n \times p$ matrix of p deterministic trend variables associated with the observations, and matrix $\mathbf{X}_{e,t}$ ($m \times p$) includes the same variables for the estimation locations. Finally, $\mathbf{G}_{e,t}$ is a $p \times m$ matrix of Lagrange multipliers used to determine estimation uncertainties (Zhou et al., 2013).

To obtain probabilistic estimates of algal biomass, areal extent, and annual duration, we used conditional simulations, which are spatially consistent Monte Carlo simulations (Chilès and Delfiner, 2012). A conditional simulation is performed at the cost of one unconditional simulation and one conditional expectation (Guinness, 2018). Following Matli et al. (2018), unconditional simulations at the estimation and observation locations and times are given by:

$$\begin{bmatrix} \mathbf{z}_{e,t}^u \\ \mathbf{z}_{o,t}^u \end{bmatrix} = D \left(\begin{bmatrix} \boldsymbol{\Sigma}_{ee} & \boldsymbol{\Sigma}_{oe,t}^T \\ \boldsymbol{\Sigma}_{oe,t} & \boldsymbol{\Sigma}_{oo} \end{bmatrix} \right)^T \mathbf{u} \quad (6)$$

where, $\mathbf{z}_{e,t}^u$ is an $n \times 1$ vector of simulated values at observation locations, $\mathbf{z}_{o,t}^u$ is an $m \times 1$ vector of simulations at estimation grid locations, \mathbf{u} is an $(m + n) \times 1$ vector of random independent samples from the standard normal distribution, and $\boldsymbol{\Sigma}_{ee}$ is the $m \times m$ covariance matrix between

estimation locations. The operator $D(\cdot)$ returns the triangular matrix resulting from Cholesky decomposition (Obenour et al., 2013; Pourahmadi, 1999).

The unconditional simulations were then conditioned to the observed data and deterministic trends using the kriging weights:

$$\mathbf{z}_{e,t}^c = \mathbf{A}_{e,t}^T (\mathbf{z}_o - \mathbf{z}_{o,t}^u) + \mathbf{z}_{e,t}^u \quad (7)$$

where, \mathbf{z}_o is the $n \times 1$ vector of all chl-*a* observations, and the resulting $\mathbf{z}_{e,t}^c$ is the vector of conditionally simulated values across the estimation grid at time t . Conditional simulations were used to reproduce the actual variability (e.g., histogram) and spatial continuity (covariance) of the field of interest (Ersoy et al., 2008). In this study, we performed 1000 simulations of daily chl-*a* concentrations for each sampling type (integrated and surface). Because integrated sampling missed the very bottom of the water column (see Section 2.1), we divided the water column into the sampled layer (upper layer) and the un-sampled layer (bottom layer). A bottom/upper layer chl-*a* concentration ratio was calculated based on available USGS data (when both integrated and bottom samples were collected during the same event), and the ratio (0.87) was used to estimate the chl-*a* concentration in the un-sampled bottom layer.

Total chl-*a* mass was aggregated across the estimation grid, where each grid cell's mass was determined as the (simulated) integrated chl-*a* concentration multiplied by the grid cell area (0.64 km²) and depth. We converted chl-*a* mass to dry weight (referred to as algal biomass) using the average chl-*a* content of *Microcystis* cells measured in western Lake Erie (0.006125 g chl-*a* per g dry weight) (Bertani et al., 2017). For comparison, we calculated "naïve" biomass estimates based on mean observed areal chl-*a* concentration (g/m²) multiplied by the representative sampling area (Bertani et al., 2017; Bridgeman et al., 2013). We used the near-Maumee region as the representative sampling area (1082 km²) since 90% of the data were collected within this region (Fig. 1a).

Surface chl-*a* concentrations were generally representative of the upper 1.5 m water layer, based on the NOAA surface sampling protocol described above. Thus, surface chl-*a* concentrations were used to estimate the bloom areal extent. To determine the areal extent, a threshold chl-*a* concentration is required, and we considered two thresholds (50 µg/L, World Health Organization, 2003; and 18 µg/L, Sayers et al., 2019) in this study. Bloom duration was estimated by counting consecutive days of bloom above a particular concentration threshold and areal extent (e.g., 1000 km²) from June to October.

2.5. Model implementation and performance assessment

Overall, the geostatistical framework, which was implemented in R (R Core Team, 2018), is summarized in Fig. S3. Parameter estimation and variable selection were performed first, followed by kriging and conditional simulation. Model validation was performed using a leave-one-cruise-out cross validation, which is similar to k -fold cross validation (James et al., 2013), and provided a rigorous test of the model's ability to predict across time, in between monitoring cruises. For this validation, we divided observations into $k = 394$ cruises, calibrating the model over $k-1$ cruises and predicting chl-*a* for the cruise left out. We repeated the process for all cruises in order to obtain a complete set of cross-validation kriging predictions. We considered all data sampled on the same day as a "cruise" regardless of which program collected the data. For OSU data, if cruises were conducted within three days, we treated them as the same cruise. Validation performance was characterized by root-mean-square error (RMSE) and the coefficient of determination (R^2), which represents the fraction of variance in the observations (i.e., log chl-*a* concentrations) explained by space-time geostatistical estimates (Taylor, 1990). We used the bootstrap approach from Ohtani (2000) to estimate the standard error of R^2 .

3. Results

3.1. Spatial and temporal trends in observations

The BIC-selected deterministic trend variables, along with their estimated coefficients (β) and standard errors (σ_{β}), are listed in Table 1. All trends are significantly different from zero ($p < 0.05$). Overall, the deterministic component, $\mu(x, y, t)$, of the geostatistical model explains 30.6% of the variance in chl-*a* (log µg/L). The annual intercepts of the deterministic trend (Table S1) account for year-to-year variability in mean chl-*a* concentrations. For example, years 2011 and 2015 have the highest intercept values, reflecting the large blooms in these years. Within years, the deterministic spatial and temporal trends explain from 18.6% to 43.8% of chl-*a* variability, depending on the year. The coefficient for the USGS data adjustment indicates that USGS samples were 25.2% ($1 - \exp(-0.291)$) lower than samples from other programs on the original scale (µg/L), all else being equal, which is generally consistent with the findings of Golnick et al. (2016).

Spatial trends alone explain 13.6% of the variability in chl-*a*, while intra-annual temporal trends alone explain 15.1% of the variability. Spatially, the northing coefficient β_N suggests that higher bloom concentrations occur in the south of the study area. Easterly coefficients β_E and β_{E^2} form a concave upward parabolic trend centered just east of Port Clinton (Fig. 2a). The combination of these spatial trends indicates that Maumee Bay generally has the highest chl-*a* concentrations, and that the north-central portion of WBLE has the lowest concentrations, likely due to the influence of the low-nutrient Detroit River plume (Michalak et al., 2013). Temporal coefficients β_T and β_{T^2} indicate a concave downward parabolic trend, where chl-*a* concentrations typically increase throughout June and July, peak in late August, and decline afterward.

The deterministic trends also included wind, depth, and sample type as covariates. The variable selection procedure identified 6-hour (9:00–15:00 on day of sampling) average wind speed as the most informative wind predictor. Results show that surface chl-*a* samples were typically higher than integrated samples across typical ranges of depth and wind speed, all else being equal (Fig. 3, Fig. S5). Also, both surface and integrated concentrations responded negatively to increasing wind speed, though the response was notably stronger for surface concentrations. Finally, both surface and integrated concentrations also responded negatively to bathymetric depth, at similar rates. Consistent with the interaction parameter between wind and depth (Table 1), at shallower depths the effect of wind mixing was somewhat less strong than at deeper locations. When the wind increased by 1 m/s, the average surface chl-*a* concentration decreased by about 6.2% and the average integrated chl-*a* concentration decreased by about 1.8% (on the original scale). Overall, our model suggests that average surface layer concentrations are 30.1% higher than average integrated chl-*a* concentrations (on the original scale) under average wind conditions (4.7 m/s).

3.2. Covariance of observations

In the geostatistical model, the variance not explained by deterministic trends is represented as spatially correlated and uncorrelated stochasticity, as in Eq. (1). The covariance parameters σ^2 (partial sill) and τ^2 (nugget) were estimated at 0.84 and 0.23 (log µg/L)², respectively. The partial sill is almost four times as large as the nugget, indicating that the majority of the variance in the detrended data is spatially and temporally correlated. The smoothness parameter, ν , is 0.17, and together with the spatial range parameter (54.6 km) and the temporal range parameter (39.1 days), determine how quickly the correlations decrease spatially and temporally (Stein, 2005). Therefore, the effective spatial and temporal correlation ranges were 91.4 km and 65.4 days (Fig. 4a). The spatial correlation range is substantially longer than the typical distance between sampling locations (around 8 km) and the temporal correlation range is considerably longer than weekly or

Table 1

Estimated geostatistical trend coefficients ($\hat{\beta}$) with standard errors ($\sigma_{\hat{\beta}}$). Units are $\log \mu\text{g/L}$ per unit of the predictor variable. Names in italics are used to identify the variables. Colons represent interactions between variables.

Variable (units)	$\hat{\beta}$	$\sigma_{\hat{\beta}}$
Program adjustment, <i>USGS</i> (–)	$-2.91\text{E}-01$	$7.30\text{E}-02$
Northing, <i>N</i> (km)	$-1.69\text{E}-02$	$3.81\text{E}-03$
Easting, <i>E</i> (km)	$-7.09\text{E}-02$	$7.72\text{E}-03$
Easting ² , <i>E</i> ² (km ²)	$6.74\text{E}-04$	$1.00\text{E}-04$
Time, <i>T</i> (day)	$7.73\text{E}-02$	$1.99\text{E}-02$
Time ² , <i>T</i> ² (day ²)	$-1.68\text{E}-04$	$4.42\text{E}-05$
Type (surface), <i>S</i> (–)	$3.99\text{E}-01$	$8.37\text{E}-02$
Depth, <i>D</i> (m)	$-3.83\text{E}-02$	$1.86\text{E}-02$
Wind, <i>W</i> (m/s)	$1.99\text{E}-03$	$6.41\text{E}-03$
Type (surface):Wind, <i>SW</i> (m/s)	$-1.95\text{E}-02$	$3.69\text{E}-03$
Wind:Depth, <i>WD</i> (m ² /s)	$-1.33\text{E}-02$	$5.31\text{E}-03$

biweekly sampling intervals. However, while the reported ranges are the lags at which 95% of the covariance is lost, it is clear from Fig. 4a that there is a substantial reduction in covariance within shorter lags.

The covariance parameters in the geostatistical model reflect the covariance of the data with the deterministic trends removed (i.e., detrended data). For comparison, we also estimated the covariance function for the raw $\log \text{chl-}a$ concentrations without trend removal (Fig. 4b). In this case, the partial sill was $1.60 (\log \mu\text{g/L})^2$ and the effective ranges of spatial and temporal correlation were 148 km and 122 days, respectively (Fig. 4b). Thus, the deterministic trends, which explain large-scale spatial and temporal patterns in bloom development, substantially reduce the portion of the variance explained by the stochastic portion of the model.

Synthesizing the deterministic trends and the space-time covariance function, we generated kriged maps of $\text{chl-}a$ (Fig. 2b). While both the trend maps and kriged maps illustrate the same large-scale temporal and spatial patterns, the kriging maps provide greater spatial detail. As a further example, Fig. S4 provides kriged $\text{chl-}a$ maps covering the period from July 7 to August 17, 2015, at a daily time interval. These example results compare well with remote sensing images in 2015 (NOAA, 2019), which also indicated that the bloom initiated from the area near Port Clinton, OH.

3.3. Predictive performance

Cross validation was used to assess the ability of the model to predict (i.e., geostatistically interpolate) $\text{chl-}a$ concentrations not included within the calibration dataset. In this validation, entire cruises were omitted, in turn, to test the model's ability to predict in both time and space (Section 2.5). Cross validation predictions explained $54.8 \pm 1.3\%$ (mean \pm standard error) of the variability in the observed $\log \text{chl-}a$ concentrations, which is markedly higher than the 30.6% of variability explained by deterministic trends alone (Section 3.1). Furthermore, a plot of observations versus predictions (Fig. S6) demonstrates that model errors are evenly distributed with negligible heteroskedasticity. For years with more frequent sampling (2013–2017), the cross validation predictions explained $60.7 \pm 1.4\%$ of the observed $\log \text{chl-}a$ variability, and about $42.1 \pm 2.8\%$ of the variability for years with less frequent samplings (2008–2012). While our model is based on geostatistics, these R^2 values can be loosely compared with the results of mechanistic phytoplankton models. According to a meta-analysis of phytoplankton models, typical R^2 values are around 40–60% (Arhonditsis and Brett, 2004), similar to what we find here.

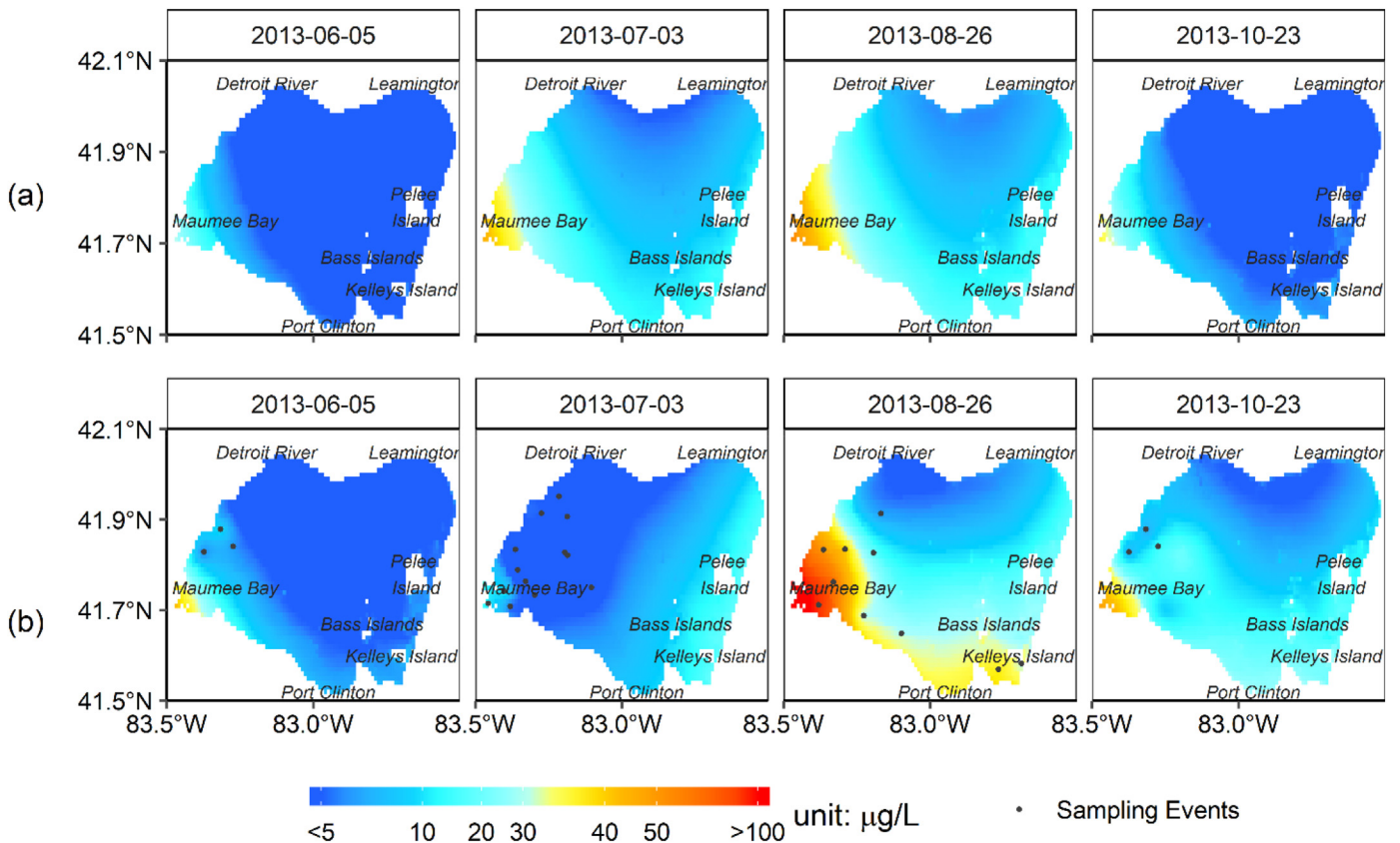


Fig. 2. (a) $\text{Chl-}a$ concentration trend maps for selected dates in 2013 based on the model parameterization provided in Table 1, and (b) associated kriging maps of geostatistically estimated $\text{chl-}a$ concentration. The difference between (a) and (b) indicate how the geostatistical estimates, which are more directly influenced by spatially and temporally proximate observations, differ from the smoother estimates based solely on the deterministic trends. A more extensive set of kriged $\text{chl-}a$ maps is provided in Fig. S4.

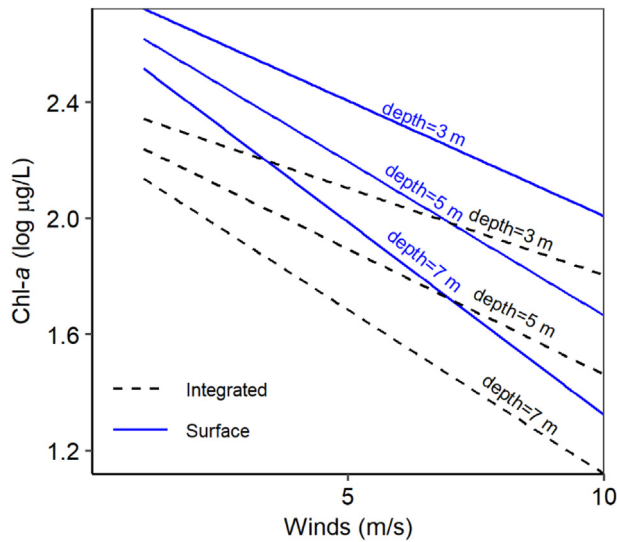


Fig. 3. Modeled chl-*a* concentration trends with wind speed and sample type (integrated or surface) at different bathymetric depths.

3.4. Bloom biomass, extent, and duration estimates

Algal biomass and areal extent time series were estimated probabilistically using conditional simulations from June to October of each year (Fig. 5). Naïve estimates based on simple averages from individual cruises (see Section 2.4) followed similar temporal patterns, but were much noisier, with 47% of such estimates falling outside the 95% confidence intervals determined by conditional simulation (Fig. 5). Confidence intervals were narrower for years with lower blooms and with more observations, especially after 2013. Notably, the years 2011 and 2015 had blooms of similar size, yet the latter had 363 more sampling events than the former (Fig. 1b), which reduced the uncertainties by 37%, on average. Results for the near-Maumee region (defined in Fig. 1a) demonstrated a similar temporal pattern (Fig. S7), but with lower uncertainty, as 90% of samples were collected within this region.

The relationship between uncertainty and sampling intensity can be explored further in terms of coefficient of variation (C_v) (Fig. 6). The average C_v of geostatistically estimated algal biomass in WBLE was reduced by 30% when the total number of monthly sampling events increased from 0 to 100, while for the near-Maumee region C_v reduction reached 35%. For months with very few samples, the uncertainties remained somewhat constrained by the deterministic trends and samples collected in nearby months. Spatially, the uncertainties (C_v) of estimated annual mean log chl-*a* concentrations were substantially smaller on the U.S. side than the Canadian side of WBLE (Fig. S8).

Bloom areal extent showed similar dynamics to the biomass time series ($r^2 > 0.95$, Fig. 5). Further analysis confirmed a strong nonlinear relationship ($p < 0.05$) between daily bloom biomass and areal extent (Fig. S9), enabling one to estimate biomass for a given area or vice versa. For example, the 18 µg/L curve could be used to estimate bloom biomass based on the extent obtained from remote sensing images (Sayers et al., 2016). To further aid in conversion between the surface and depth-integrated HAB measurements, based on the model output, we quantified that the average (integrated) chl-*a* concentration for the whole column was 67% of the concentration in the surface water layer (top 1.5 m), on average. Also, the average ratio of overall biomass in WBLE to biomass in the surface layer was 2.48 (Fig. S10).

Our results can also be aggregated to compare annual maximum algal biomass (30-day mean), areal extent, and duration across the years (Fig. 7). The correlation between algal biomass and extent was very strong, with $r = 0.92$ or 0.99 , depending on whether extent was defined based on chl-*a* concentration > 18 µg/L or > 50 µg/L, respectively. Overall, 2011 and 2015 had both the highest biomass and areal extent

on record (they were not significantly different from each other). The correlation between algal biomass and duration was also strong (e.g., $r = 0.88$ for duration of chl-*a* concentration > 18 µg/L over area > 1000 km²). The HABs of 2011, 2013, and 2015 were the most severe according to the various duration criteria considered (Fig. 7c).

4. Discussion

4.1. HAB variability and comparison to previous studies

Compared to previous studies of HAB variability in lacustrine systems (Bridgeman et al., 2013; Liu et al., 2011; Xu et al., 2017), this study leverages sparse sampling data to provide high temporal resolution (daily) and spatially-resolved algal concentration and biomass estimates with quantified uncertainties. Such probabilistic estimates can inform risk-based HABs management, considering potential ecological, human health, and economic impacts (Brooks et al., 2016; Huisman et al., 2018). Reliable HAB estimates are also important for calibrating and validating forecasting models that link HABs to nutrient loading and other environmental influences (Scavia et al., 2016; Yan et al., 2017). Biomass, in particular, is an important HAB attribute for characterizing seasonal and interannual variability from *in situ* sampling data (Bertani et al., 2017; Bridgeman et al., 2013) and remote sensing

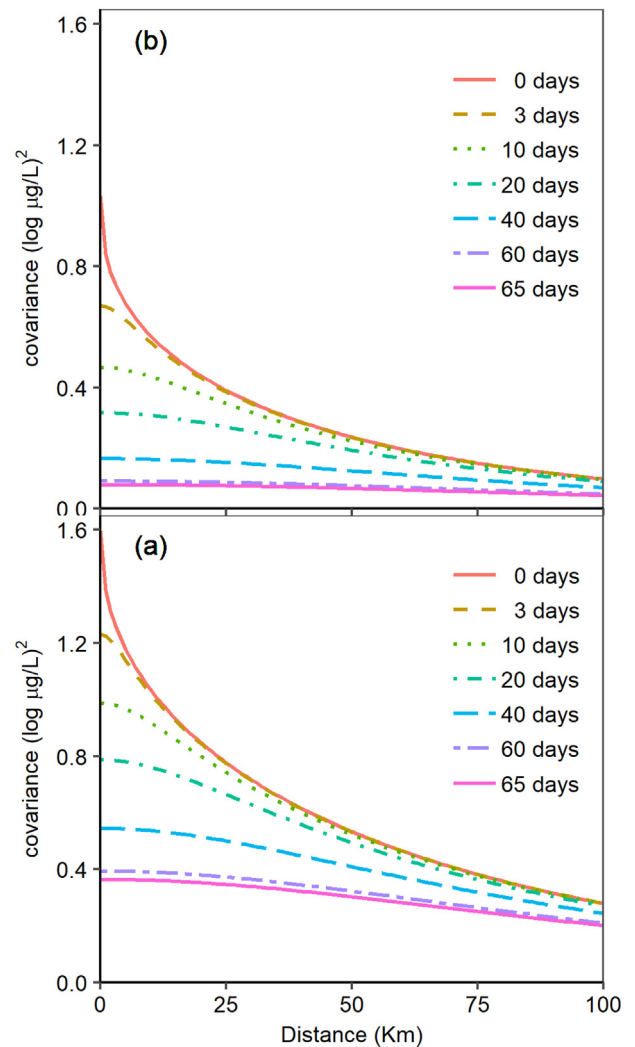


Fig. 4. Fitted space-time covariance function for (a) detrended log chl-*a* concentrations (as used in the geostatistical model) and (b) log chl-*a* concentrations observations.

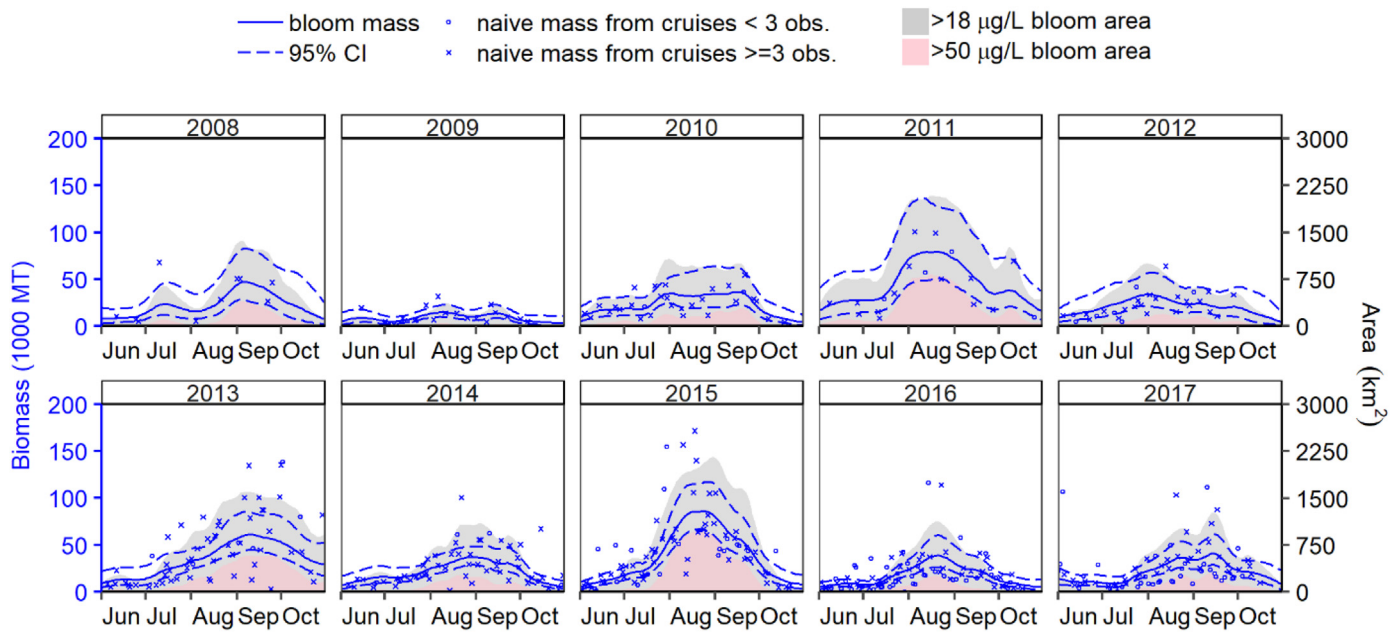


Fig. 5. Time series of geostatistically estimated mean algal biomass (metric tons, MT) with 95% confidence intervals, and “naïve” estimates of algal biomass from individual cruises in WBLE. Also shown are geostatistically estimated bloom areal extent for chl-*a* >18 µg/L and >50 µg/L.

(Sayers et al., 2019; Stumpf et al., 2016b). However, in the absence of a synthesizing model, *in situ* biomass estimates have generally been determined from sparse and noisy estimates of individual cruises or monitoring programs without uncertainty quantification.

Our basin-wide biomass estimates, developed through space-time conditional simulation (Chilès and Delfiner, 2012; Matli et al., 2018) of chl-*a*, can be compared with the relatively noisy “naïve” estimates determined from simple averages of individual cruises (Fig. 5). Naïve estimates suggest that basin-wide biomass regularly fluctuates over an order of magnitude on a time scale of days. This implausible result is likely due to the patchiness of WBLE blooms, such that individual cruises may or may not sample the locations where the bloom is most intense (Bertani et al., 2017; Binding et al., 2019; Vander Woude et al.,

2019). Small cruises (with only one or two samples) produce naïve estimates that are only modestly correlated with the geostatistical estimates ($r^2 = 0.30$), whereas larger cruises produce somewhat more correlated estimates ($r^2 = 0.59$). Consistent with these findings, previously published estimates of annual peak cyanobacteria biomass from an individual monitoring program (Bertani et al., 2017; Bridgeman et al., 2013) have limited correlation ($r^2 \leq 0.41$, Table S2) with the peak biomass estimates of this study and remote sensing studies (Fig. 8). Overall, these results indicate both the importance of maintaining sufficient cruise coverage to capture bloom dynamics (see Section 4.2), and the importance of the space-time geostatistical approach for synthesizing data across multiple programs to improve estimation accuracy while reducing uncertainty.

The geostatistical peak biomass estimates correlate well with previously published remote sensing peak estimates ($r^2 > 0.60$, Table S2). However, there are some instructive deviations. For example, Stumpf et al. (2016b) showed that the 2009 bloom peaked in early September with the highest concentrations near the coast around Leamington, Canada, where no *in situ* observations were collected. Therefore, the geostatistical model likely underestimated chl-*a* mass in this area in that year, resulting in under-estimation of the peak bloom size (Fig. 8). On the other hand, in 2016, the geostatistical estimates indicated the bloom peaked in mid-August when cloud cover limited remote sensing coverage, likely leading to an under-estimation of the bloom by remote sensing (Manning et al., 2019, Fig. 8). Thus, geostatistical model estimates are limited primarily by the spatio-temporal coverage of the *in situ* samples (particularly on the Canadian side of the lake), and remote sensing estimates are sometimes limited by cloud cover. Since geostatistics can also be applied to remote sensing data (Ha et al., 2014; Yang and Hu, 2018), future research could potentially improve HAB estimation by integrating *in situ* and remote sensing data.

It is important to note that some remote sensing studies measure chl-*a* concentrations (Sayers et al., 2019) while others measure cyanobacteria-specific pigments (Stumpf et al., 2016a; Stumpf et al., 2016b). However, our chl-*a*-based estimates of peak bloom biomass correlate similarly well to both types of remote sensing estimates (Fig. 8, Table S2), likely because Lake Erie’s peak bloom has mostly been dominated by cyanobacteria, at least over the past decade (Jankowiak et al., 2019; Michalak et al., 2013). We also note that algal

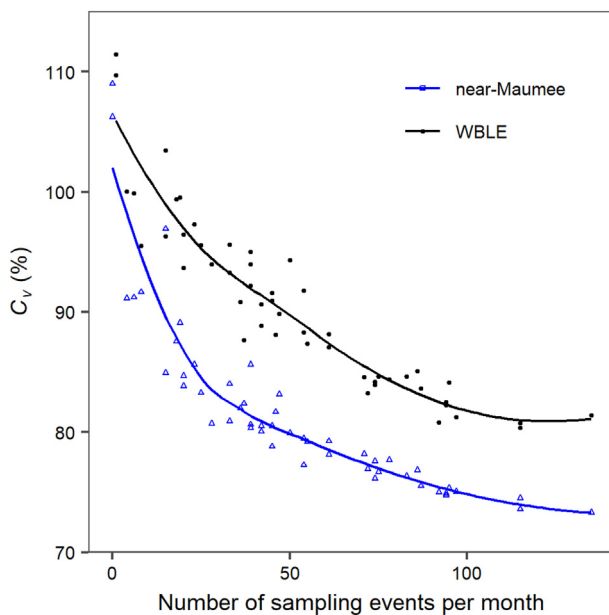


Fig. 6. Relationship between monthly mean coefficient of variation (C_v) for algal biomass estimates (as in Fig. 5) and number of sampling events per month. Results are shown for WBLE and near-Maumee region. Lines are the fitted LOESS curves (span = 0.5).

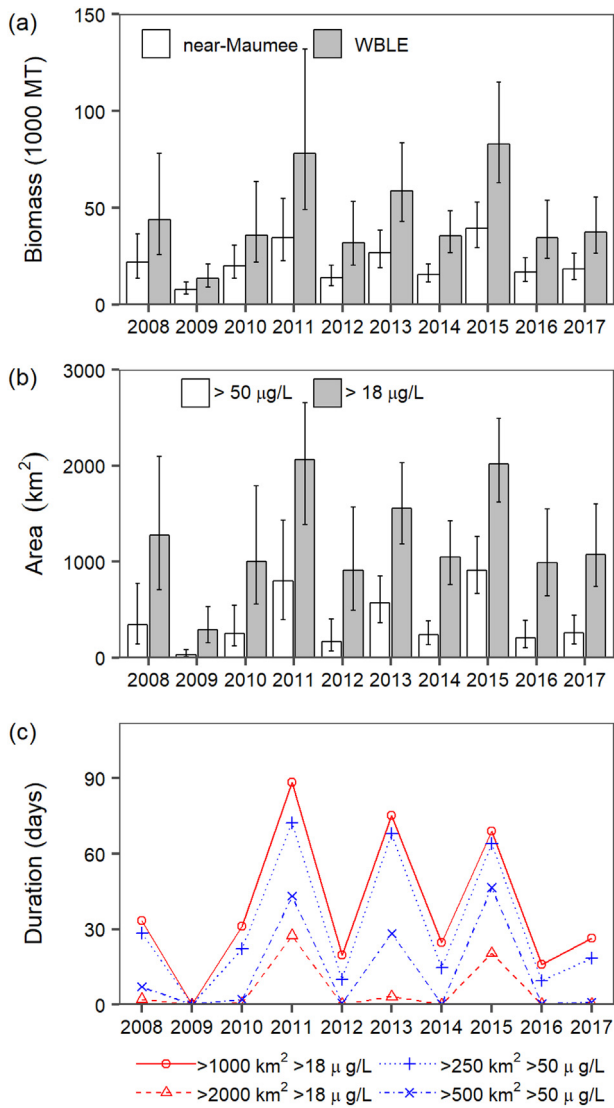


Fig. 7. (a) Annual maximum 30-day mean algal biomass with 95% confidence intervals for WBLE and near-Maumee region; (b) maximum 30-day mean bloom areal extent with 95% intervals for WBLE using alternative bloom concentration criteria; (c) bloom duration for WBLE using multiple criteria.

biomass was determined assuming a constant biomass-to-chl-*a* ratio, consistent with previous studies (Bertani et al., 2017; Conroy et al., 2005), though cyanobacteria contributions to chl-*a* may vary with environmental factors such as light and nutrients (Berry et al., 2017; Chaffin et al., 2018; Nicholls and Dillon, 1978). Thus, while there is general consistency among the peak bloom estimates from this and previous studies, variability in algal community composition may complicate chl-*a*-based HAB estimates (Binding et al., 2018), and there remains a need to better understand how ratios between cyanobacteria biomass and chl-*a* vary over time.

Prior to this study, *in situ* measurements had not been used to quantify HAB areal extent in the Great Lakes, likely due to the sparse spatial coverage of individual sampling programs. However, our composite geostatistical areal extent estimates can be compared with remote sensing areal estimates (Fig. S11). Our annual maximum 30-day extent estimates based on the 18 µg/L threshold are well correlated with the maximum extent estimates of Sayers et al. (2019) who used that same threshold ($r^2 = 0.62$, Table S2); and on average our estimates are just 11% lower. Our maximum extent estimates based on 50 µg/L were well correlated with those of Ho et al. (2017) ($r^2 = 0.65$, Table S2), but their estimates were, on average, twice as high as ours.

Ho et al. (2017) developed extent estimates based on a Cyanobacteria Index threshold of 0.001, which is considered roughly equivalent to 50 µg/L chl-*a* or 10^5 cells/mL (Stumpf et al., 2016b; World Health Organization, 2003). Interestingly, the estimates of Ho et al. (2017) were more similar in magnitude to our extent estimates based on an 18 µg/L threshold. This contrast further indicates the need to better understand relationships among chl-*a*, cyanobacterial biomass, and cell counts, especially given recent interest in developing Lake Erie water quality criteria using cyanobacterial cell counts (Davis et al., 2019).

The geostatistical estimates allow us to explore empirical relationships among biomass and areal extent that could be helpful in future studies and HAB management. The relationship between bloom biomass and extent is found to be nonlinear for the two concentration criteria considered (Fig. S9). For the lower concentration criteria (18 µg/L), areal extent increases rapidly as biomass increases beyond 10^4 MT, but it begins to saturate around 2200 km² (~75% of our WBLE study area) for biomass values above 7.5×10^4 MT. The maximum spatial extent is likely controlled by the low-nutrient Detroit River plume, which limits boom expansion in the northern portion of the study area (Michalak et al., 2013; Stumpf et al., 2016b). On the other hand, for a concentration criterion of 50 µg/L, areal extent appears to have a positive exponential relationship with biomass, suggesting that these smaller areas of particularly high HAB concentration will continue to expand if biomass continues to increase in future years (Kitchens et al., 2018; Obenour et al., 2014).

Bloom duration is another measure of HAB severity, with implications for water supply risk management (Wynne and Stumpf, 2015). Duration is likely related to climate and water column conditions (e.g., temperature, stratification) that control the window of time over which HAB taxa may proliferate (Wells et al., 2015). However, yearly estimates of HAB duration in WBLE have not been reported comprehensively in previous studies. In this study, HAB duration is based on thresholds for chl-*a* concentration and areal extent (Fig. 7). In general, bloom duration estimates were strongly and linearly correlated with

- Geostatistics (30-day max, this study)
- ◇ Remote sensing (30-day max, Stumpf et al. 2016b)
- △ Remote sensing (21-day max, Manning et al. 2019)
- ◆ UT *in situ* (30-day max, Bertani et al. 2017)

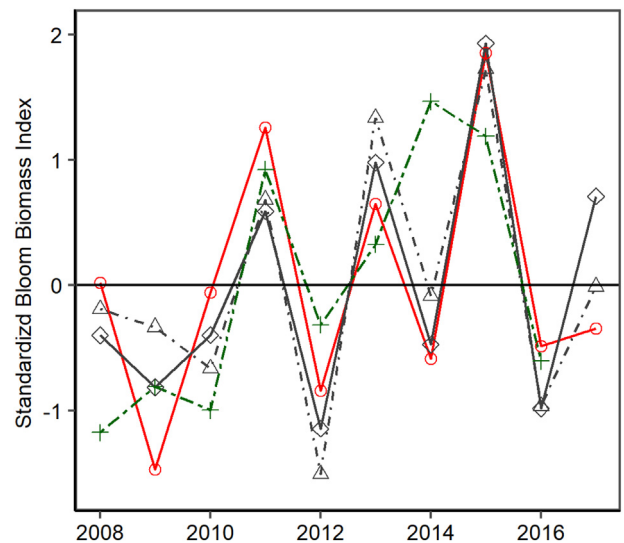


Fig. 8. Standardized WBLE annual biomass index comparison among (1) our maximum 30-day average from geostatistical modeling of integrated chl-*a* concentration, (2) Stumpf et al. (2016b) maximum 30-day average based on surface-layer Cyanobacteria Index from MERIS/MODIS-derived satellite images, (3) Manning et al. (2019) maximum 21-day average based on surface-layer chl-*a* concentration from MODIS-derived images, and (4) Bertani et al. (2017) maximum 30-day average of UT *in situ* measurements (Bridgeman et al., 2013). All estimates are standardized by the mean and standard deviation of their respective datasets.

the maximum 30-day mean bloom biomass estimates (Section 3.4). Thus, reductions in bloom biomass, potentially through nutrient management, are expected to produce comparable reductions in bloom duration, at least for similar climate conditions.

4.2. Wind and depth influence on chl-*a* estimation

In this study, we characterized how wind and depth control the relationship between surface and integrated chl-*a* concentrations (e.g., Fig. 5). This relationship reflects the ability of cyanobacteria to form colonies and regulate their buoyancy, potentially leading to surface layer accumulations under low wind conditions (Chaffin et al., 2012; Oliver et al., 2012). However, while other Lake Erie studies suggested particular wind velocity thresholds for surface accumulation, ranging from 4.9 to 7.7 m/s (Boegman et al., 2008; Wynne et al., 2010; Rowe et al., 2016; Bosse et al., 2019), our model suggests a more continuous effect for the overall WBLE (Fig. 3, S5). Based on statistical variable selection (BIC), a linear trend with wind outperformed stepwise changes based on specific wind thresholds. Furthermore, our results indicate that a 6-hour average (9:00–15:00 on the sampling day) wind speed is a better predictor than longer wind averaging periods, consistent with previous studies demonstrating that blooms respond rapidly to wind-induced mixing (Rowe et al., 2016). While surface chl-*a* concentrations respond most strongly (and negatively) to changes in wind speed, there is a smaller negative relationship between integrated chl-*a* and wind speed. These relationships can potentially improve bloom biomass estimation when only surface measurements are available. For example, our results indicate that at low wind speeds (2 m/s), surface chl-*a* concentrations are 35% higher than integrated concentrations, whereas, for higher wind speeds (8 m/s), surface concentrations are only 17% higher. Thus, the relationships developed in this study can help remove biases in remote sensing bloom estimates caused by severe wind conditions (Bosse et al., 2019; Wynne et al., 2010).

4.3. Sampling guidance

The Harmful Algal Bloom and Hypoxia Research and Control Amendments Act of 2017 (U.S. Congress, 2017) emphasized coordination and integration to prevent unnecessary duplication of effort among programs. Our results provide an assessment of the ongoing monitoring efforts, which could lead toward more efficient and coordinated WBLE monitoring. For example, since 2013, the uncertainty (C_v) of estimated chl-*a* concentration in WBLE decreased due to increased sampling activity from multiple monitoring programs (Fig. 5, S8, S12). In particular, 2015 had the most sampling events and the smallest average C_v . To constrain uncertainties in chl-*a* estimates in WBLE, at least 50 sampling events each month are suggested, yet increasing the total sampling events from 50 to 115 only reduces the C_v by an additional 10% (Fig. 6), based on historical sampling patterns.

Spatially, the average C_v of estimated chl-*a* concentration was generally lowest in the near-Maumee region, and highest on the Canadian side of the lake (Fig. S8, S12). Sparse sampling on the Canadian side increases estimation uncertainties throughout the entire WBLE. As illustrated in Fig. S12, average WBLE chl-*a* uncertainties are substantially reduced at times when cruises are conducted on the Canadian side of the lake. In contrast, individual cruises conducted on the U.S. side do not substantially reduce overall WBLE uncertainties (particularly after 2011, when sampling became more frequent). Thus, managers can consider allocating more sampling resources to the Canadian side to improve basin-wide HAB estimation.

5. Conclusion

This study presents an innovative approach to HAB estimation using space-time geostatistics. We demonstrate how the combination of

deterministic trends and spatio-temporal correlations in chl-*a* concentration can provide realistic estimates throughout time, even on dates when no data are available (Section 3.3). The availability of observational data that are spatially/temporally proximate to the estimation location/time is reflected in the estimation uncertainties (e.g., confidence intervals), which were substantially reduced in later years with more frequent sampling (Fig. 5). In addition, the space-time approach allows us to assess relationships between chl-*a* and environmental covariates (e.g., depth and wind) in a rigorous modeling framework that accounts for the correlation (i.e., lack of statistical independence) among water quality samples (Gelfand et al., 2010; Obenour et al., 2012). We find that wind has an important negative relationship with summer chl-*a* concentration, particularly for surface samples collected at locations with a relatively deep water column. Beyond relationships with environmental covariates, the approach also allows us to account for potential biases related to the different sampling and analytical procedures used by various monitoring programs.

Spatially-aggregated HAB metrics, including depth-integrated biomass and surface areal extent, were estimated probabilistically using space-time conditional simulations. Compared to remote sensing estimates, which are based only on surface concentrations, the depth-integrated biomass estimates provide new information on overall lake productivity. At the same time, the year-to-year variability in our new biomass estimates correlates moderately well with remote sensing-derived estimates, suggesting that both approaches reasonably characterize the temporal variability in WBLE HABs. Deviations between the geostatistical and remote sensing estimates appear to largely reflect data gaps (in the availability of both *in situ* measurements and remote sensing imagery) and suggest that geostatistically integrating *in situ* and remote sensing data could be beneficial future research. Finally, this research, which synthesizes observations from five different U.S. and Canadian monitoring programs, has the potential to guide more coordinated and efficient HAB monitoring. We find that at least 50 samples per month, including samples on the Canadian side of the lake, are needed to prevent WBLE algal biomass and extent estimation uncertainties from ballooning.

Acknowledgments

This project was funded by the National Science Foundation Coastal SEES grant (#OCE-1600012). We would like to thank the Ohio EPA, Ohio Sea Grant, and the Lake Erie Charter Boat Association for supporting Stone Lab sampling during the 2013–2017 period. This research was partially funded by the Great Lakes Restoration Initiative, and funding awarded to Cooperative Institute for Great Lakes Research (CIGLR) through the NOAA Cooperative Agreement with the University of Michigan (NA17OAR4320152). USGS data associated with this publication are available at <https://doi.org/10.5066/P9WY3FDV>. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. This CIGLR contribution number is 1142.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.scitotenv.2019.133776>.

References

- Arhonditsis, G.B., Brett, M.T., 2004. Evaluation of the current state of mechanistic aquatic biogeochemical modeling. *Mar. Ecol. Prog. Ser.* 271, 13–26. <https://doi.org/10.3354/meps271013>.
- Beletsky, D., Schwab, D.J., Roebber, P.J., McCormick, M.J., Miller, G.S., Saylor, J.H., 2003. Modeling wind-driven circulation during the March 1998 sediment resuspension event in Lake Michigan. *J. Geophys. Res. Oceans* 108. <https://doi.org/10.1029/2001JC001159>.
- Berry, M.A., Davis, T.W., Cory, R.M., Duhaime, M.B., Johengen, T.H., Kling, G.W., Marino, J.A., Den Uyl, P.A., Gossiaux, D., Dick, G.J., Deneff, V.J., 2017. Cyanobacterial harmful

- algal blooms are a biological disturbance to Western Lake Erie bacterial communities. *Environ. Microbiol.* 19, 1149–1162. <https://doi.org/10.1111/1462-2920.13640>.
- Bertani, I., Obenour, D.R., Steger, C.E., Stow, C.A., Gronewold, A.D., Scavia, D., 2016. Probabilistically assessing the role of nutrient loading in harmful algal bloom formation in western Lake Erie. *J. Great Lakes Res.* 42, 1184–1192. <https://doi.org/10.1016/j.jglr.2016.04.002>.
- Bertani, I., Steger, C.E., Obenour, D.R., Fahnenstiel, G.L., Bridgeman, T.B., Johengen, T.H., Sayers, M.J., Shuchman, R.A., Scavia, D., 2017. Tracking cyanobacteria blooms: do different monitoring approaches tell the same story? *Sci. Total Environ.* 575, 294–308. <https://doi.org/10.1016/j.scitotenv.2016.10.023>.
- Binding, C.E., Greenberg, T.A., McCullough, G., Watson, S.B., Page, E., 2018. An analysis of satellite-derived chlorophyll and algal bloom indices on Lake Winnipeg. *J. Great Lakes Res.* 44, 436–446. <https://doi.org/10.1016/j.jglr.2018.04.001>.
- Binding, C.E., Zastepa, A., Zeng, C., 2019. The impact of phytoplankton community composition on optical properties and satellite observations of the 2017 western Lake Erie algal bloom. *J. Great Lakes Res.* 45, 573–586. <https://doi.org/10.1016/j.jglr.2018.11.015>.
- Boegman, L., Loewen, M., Hamblin, P., Culver, D., 2008. Vertical mixing and weak stratification over zebra mussel colonies in western Lake Erie. *Limnol. Oceanogr.* 53, 1093–1110. <https://doi.org/10.4319/lo.2008.53.3.1093>.
- Bosse, K.R., Sayers, M.J., Shuchman, R.A., Fahnenstiel, G.L., Ruberg, S.A., Fanslow, D.L., Stuart, D.G., Johengen, T.H., Burtner, A.M., 2019. Spatial-temporal variability of *in situ* cyanobacteria vertical structure in Western Lake Erie: implications for remote sensing observations. *J. Great Lakes Res.* <https://doi.org/10.1016/j.jglr.2019.02.003>.
- 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. <https://doi.org/10.1016/j.jglr.2012.11.004>.
- Brooks, B.W., Lazorchak, J.M., Howard, M.D.A., Johnson, M.V.V., Morton, S.L., Perkins, D.A.K., Reavie, E.D., Scott, G.I., Smith, S.A., Steevens, J.A., 2016. Are harmful algal blooms becoming the greatest inland water quality threat to public health and aquatic ecosystems? *Environ. Toxicol. Chem.* 35, 6–13. <https://doi.org/10.1002/etc.3220>.
- Chaffin, J.D., Bridgeman, T.B., Heckathorn, S.A., Mishra, S., 2011. Assessment of *Microcystis* growth rate potential and nutrient status across a trophic gradient in western Lake Erie. *J. Great Lakes Res.* 37, 92–100. <https://doi.org/10.1016/j.jglr.2010.11.016>.
- 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. <https://doi.org/10.4236/jwarp.2012.412119>.
- Chaffin, J.D., Davis, T.W., Smith, D.J., Baer, M.M., Dick, G.J., 2018. Interactions between nitrogen form, loading rate, and light intensity on *Microcystis* and *Planktothrix* growth and microcystin production. *Harmful Algae* 73, 84–97. <https://doi.org/10.1016/j.hal.2018.02.001>.
- Chilès, J.P., Delfiner, P., 2012. *Geostatistics: Modeling Spatial Uncertainty*. Second edition. John Wiley & Sons, New Jersey.
- Conroy, J.D., Kane, D.D., Dolan, D.M., Edwards, W.J., Charlton, M.N., Culver, D.A., 2005. Temporal trends in Lake Erie plankton biomass: roles of external phosphorus loading and Dreissenid mussels. *J. Great Lakes Res.* 31, 89–110. [https://doi.org/10.1016/S0380-1330\(05\)70307-5](https://doi.org/10.1016/S0380-1330(05)70307-5).
- Davis, T.W., Stumpf, R., Bullerjahn, G.S., McKay, R.M.L., Chaffin, J.D., Bridgeman, T.B., Winslow, C., 2019. Science meets policy: a framework for determining impairment designation criteria for large waterbodies affected by cyanobacterial harmful algal blooms. *Harmful Algae* 81, 59–64. <https://doi.org/10.1016/j.hal.2018.11.016>.
- Ersay, A., Yunsel, T.Y., Atici, Ü., 2008. Geostatistical conditional simulation for the assessment of contaminated land by abandoned heavy metal mining. *Environ. Toxicol.* 23, 96–109. <https://doi.org/10.1002/tox.20314>.
- Gelfand, A.E., Diggle, P.J., Fuentes, M., Guttorp, P., 2010. *Handbook of Spatial Statistics*. CRC press, London.
- Golnick, P.C., Chaffin, J.D., Bridgeman, T.B., Zellner, B.C., Simons, V.E., 2016. A comparison of water sampling and analytical methods in western Lake Erie. *J. Great Lakes Res.* 42, 965–971. <https://doi.org/10.1016/j.jglr.2016.07.031>.
- Guinness, J., 2018. Permutation and grouping methods for sharpening Gaussian process approximations. *Technometrics* 60, 415–429. <https://doi.org/10.1080/00401706.2018.1437476>.
- Ha, N.T.T., Koike, K., Nhuan, M.T., 2014. Improved accuracy of chlorophyll-*a* concentration estimates from MODIS imagery using a two-band ratio algorithm and geostatistics: as applied to the monitoring of eutrophication processes over Tien Yen Bay (northern Vietnam). *Remote Sens.* 6, 421–442. <https://doi.org/10.3390/rs6010421>.
- Ho, J.C., Michalak, A.M., 2017. Phytoplankton blooms in Lake Erie impacted by both long-term and springtime phosphorus loading. *J. Great Lakes Res.* 43, 221–228. <https://doi.org/10.1016/j.jglr.2017.04.001>.
- Ho, J.C., Stumpf, R.P., Bridgeman, T.B., Michalak, A.M., 2017. Using Landsat to extend the historical record of lacustrine phytoplankton blooms: a Lake Erie case study. *Remote Sens. Environ.* 191, 273–285. <https://doi.org/10.1016/j.rse.2016.12.013>.
- Holbach, A., Norra, S., Wang, L., Yijun, Y., Hu, W., Zheng, B., Bi, Y., 2014. Three gorges reservoir: density pump amplification of pollutant transport into tributaries. *Environ. Sci. Technol.* 48, 7798–7806. <https://doi.org/10.1021/es501132k>.
- Huang, H.C., Martinez, F., Mateu, J., Montes, F., 2007. Model comparison and selection for stationary space-time models. *Comput. Stat. Data Anal.* 51, 4577–4596. <https://doi.org/10.1016/j.csda.2006.07.038>.
- Huisman, J., Codd, G.A., Paerl, H.W., Ibelings, B.W., Verspagen, J.M., Visser, P.M., 2018. Cyanobacterial blooms. *Nat. Rev. Microbiol.* 16, 471–483. <https://doi.org/10.1038/s41579-018-0040-1>.
- James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. *An Introduction to Statistical Learning: With Applications in R*. 112. Springer, New York.
- Jankowiak, J., Hattenrath-Lehmann, T., Kramer, B.J., Ladds, M., Gobler, C.J., 2019. Deciphering the effects of nitrogen, phosphorus, and temperature on cyanobacterial bloom intensification, diversity, and toxicity in western Lake Erie. *Limnol. Oceanogr.* 0. <https://doi.org/10.1002/lno.11120>.
- Jarvie, H.P., Johnson, L.T., Sharpley, A.N., Smith, D.R., Baker, D.B., Bruulsema, T.W., Confesor, R., 2017. Increased soluble phosphorus loads to Lake Erie: unintended consequences of conservation practices? *J. Environ. Qual.* 46, 123–132. <https://doi.org/10.2134/jeq2016.07.0248>.
- Kane, D.D., Conroy, J.D., Richards, R.P., Baker, D.B., Culver, D.A., 2014. Re-eutrophication of Lake Erie: correlations between tributary nutrient loads and phytoplankton biomass. *J. Great Lakes Res.* 40, 496–501. <https://doi.org/10.1016/j.jglr.2014.04.004>.
- Katzfuss, M., Guinness, J., 2017. A general framework for Vecchia approximations of Gaussian processes. arXiv preprint arXiv:1708.06302. <https://arxiv.org/abs/1708.06302v4>.
- Kitchens, C.M., Johengen, T.H., Davis, T.W., 2018. Establishing spatial and temporal patterns in *Microcystis* sediment seed stock viability and their relationship to subsequent bloom development in Western Lake Erie. *PLoS One* 13, e0206821. <https://doi.org/10.1371/journal.pone.0206821>.
- Kyriakidis, P.C., Journel, A.G., 1999. Geostatistical space-time models: a review. *Math. Geol.* 31, 651–684. <https://doi.org/10.1023/a:1007528426688>.
- Liu, X., Lu, X., Chen, Y., 2011. The effects of temperature and nutrient ratios on *Microcystis* blooms in Lake Taihu, China: an 11-year investigation. *Harmful Algae* 10, 337–343. <https://doi.org/10.1016/j.hal.2010.12.002>.
- Ludovisi, A., Minozzo, M., Pandolfi, P., Taticchi, M.I., 2005. Modelling the horizontal spatial structure of planktonic community in Lake Trasimeno (Umbria, Italy) using multivariate geostatistical methods. *Ecol. Model.* 181, 247–262. <https://doi.org/10.1016/j.ecolmodel.2004.06.033>.
- Manning, N.F., Wang, Y.-C., Long, C.M., Bertani, I., Sayers, M.J., Bosse, K.R., Shuchman, R.A., Scavia, D., 2019. Extending the forecast model: predicting western Lake Erie harmful algal blooms at multiple spatial scales. *J. Great Lakes Res.* <https://doi.org/10.1016/j.jglr.2019.03.004>.
- Matil, V.R.R., Fang, S., Guinness, J., Rabalais, N.N., Craig, J.K., Obenour, D.R., 2018. A space-time geostatistical assessment of hypoxia in the northern Gulf of Mexico. *Environ. Sci. Technol.* 52, 12484–12493. <https://doi.org/10.1021/acs.est.8b03474>.
- Matthews, M.W., 2011. A current review of empirical procedures of remote sensing in inland and near-coastal transitional waters. *Int. J. Remote Sens.* 32, 6855–6899. <https://doi.org/10.1080/01431161.2010.512947>.
- Michalak, A.M., Anderson, E.J., Beletsky, D., Boland, 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.* 110, 6448–6452. <https://doi.org/10.1073/pnas.1216006110>.
- Money, E.S., Carter, G.P., Serre, M.L., 2009. Modern space/time geostatistics using river distances: data integration of turbidity and *E. coli* measurements to assess fecal contamination along the Raritan River in New Jersey. *Environ. Sci. Technol.* 43, 3736–3742. <https://doi.org/10.1021/es803236j>.
- Money, E.S., Sackett, D.K., Aday, D.D., Serre, M.L., 2011. Using river distance and existing hydrography data can improve the geostatistical estimation of fish tissue mercury at unsampled locations. *Environ. Sci. Technol.* 45, 7746–7753. <https://doi.org/10.1021/es2003827>.
- Murphy, R.R., Curriero, F.C., Ball, W.P., 2010. Comparison of spatial interpolation methods for water quality evaluation in the Chesapeake Bay. *J. Environ. Eng. ASCE* 136, 160–171. [https://doi.org/10.1061/\(ASCE\)Ee.1943-7870.0000121](https://doi.org/10.1061/(ASCE)Ee.1943-7870.0000121).
- Newell, S.E., Davis, T.W., Johengen, T.H., Gossiaux, D., Burtner, A., Palladino, D., McCarthy, M.J., 2019. Reduced forms of nitrogen are a driver of non-nitrogen-fixing harmful cyanobacterial blooms and toxicity in Lake Erie. *Harmful Algae* 81, 86–93. <https://doi.org/10.1016/j.hal.2018.11.003>.
- Nicholls, K.H., Dillon, P.J., 1978. An evaluation of phosphorus-chlorophyll-phytoplankton relationships for lakes. *Internationale Revue der gesamten Hydrobiologie und Hydrographie* 63, 141–154. <https://doi.org/10.1002/iroh.19780630203>.
- NOAA, 2019. Harmful algal blooms in Lake Erie - experimental and operational HAB bulletin archive. Available at: https://www.glerl.noaa.gov/res/HABs_and_Hypoxia/lakeErieHABArchive/, Accessed date: 26 June 2019.
- Obenour, D.R., Michalak, A.M., Zhou, Y., Scavia, D., 2012. Quantifying the impacts of stratification and nutrient loading on hypoxia in the northern Gulf of Mexico. *Environ. Sci. Technol.* 46, 5489–5496. <https://doi.org/10.1021/es204481a>.
- Obenour, D.R., Scavia, D., Rabalais, N.N., Turner, R.E., Michalak, A.M., 2013. Retrospective analysis of midsummer hypoxic area and volume in the northern Gulf of Mexico, 1985–2011. *Environ. Sci. Technol.* 47, 9808–9815. <https://doi.org/10.1021/es400983g>.
- Obenour, D.R., Gronewold, A.D., Stow, C.A., Scavia, D., 2014. Using a Bayesian hierarchical model to improve Lake Erie cyanobacteria bloom forecasts. *Water Resour. Res.* 50, 7847–7860. <https://doi.org/10.1002/2014wr015616>.
- Ohtani, K., 2000. Bootstrapping R2 and adjusted R2 in regression analysis. *Econ. Model.* 17, 473–483. [https://doi.org/10.1016/S0264-9993\(99\)00034-6](https://doi.org/10.1016/S0264-9993(99)00034-6).
- Oliver, R.L., Hamilton, D.P., Brookes, J.D., Ganf, G.G., 2012. Ecology of Cyanobacteria II: Their Diversity in Space and Time. Whitton, B.A. (ed), pp. 155–194, Springer Netherlands, Dordrecht.
- Paerl, H.W., Paul, V.J., 2012. Climate change: links to global expansion of harmful cyanobacteria. *Water Res.* 46, 1349–1363. <https://doi.org/10.1016/j.watres.2011.08.002>.
- Page, B.P., Kumar, A., Mishra, D.R., 2018. A novel cross-satellite based assessment of the spatio-temporal development of a cyanobacterial harmful algal bloom. *Int. J. Appl. Earth Obs. Geoinf.* 66, 69–81. <https://doi.org/10.1016/j.jag.2017.11.003>.
- Pourahmadi, M., 1999. Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation. *Biometrika* 86, 677–690. <https://doi.org/10.1093/biomet/86.3.677>.

- R Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, V, Austria Available at. <http://www.R-project.org>.
- Rowe, M.D., Anderson, E.J., Wynne, T.T., Stumpf, R.P., Fanslow, D.L., Kijanka, K., Vanderploeg, H.A., Strickler, J.R., Davis, T.W., 2016. Vertical distribution of buoyant *Microcystis* blooms in a Lagrangian particle tracking model for short-term forecasts in Lake Erie. *J. Geophys. Res. Oceans* 121, 5296–5314. <https://doi.org/10.1002/2016jc011720>.
- 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. <https://doi.org/10.1080/01431161.2016.1207265>.
- Sayers, M.J., Grimm, A.G., Shuchman, R.A., Bosse, K.R., Fahnenstiel, G.L., Ruberg, S.A., Leshkevich, G.A., 2019. Satellite monitoring of harmful algal blooms in the Western Basin of Lake Erie: a 20-year time-series. *J. Great Lakes Res.* <https://doi.org/10.1016/j.jglr.2019.01.005>.
- Scavia, D., DePinto, J.V., Bertani, I., 2016. A multi-model approach to evaluating target phosphorus loads for Lake Erie. *J. Great Lakes Res.* 42, 1139–1150. <https://doi.org/10.1016/j.jglr.2016.09.007>.
- Schwarz, G., 1978. Estimating the dimension of a model. *Ann. Stat.* 6, 461–464. <https://doi.org/10.1214/aos/1176344136>.
- Steffen, M.M., Davis, T.W., McKay, R.M.L., Bullerjahn, G.S., Krausfeldt, L.E., Stough, J.M.A., Neitzey, M.L., Gilbert, N.E., Boyer, G.L., Johengen, T.H., Gossiaux, D.C., Burtner, A.M., Palladino, D., Rowe, M.D., Dick, G.J., Meyer, K.A., Levy, S., Boone, B.E., Stumpf, R.P., Wynne, T.T., Zimba, P.V., Gutierrez, D., Wilhelm, S.W., 2017. Ecophysiological examination of the Lake Erie *Microcystis* bloom in 2014: linkages between biology and the water supply shutdown of Toledo, OH. *Environ. Sci. Technol.* 51, 6745–6755. <https://doi.org/10.1021/acs.est.7b00856>.
- Stein, M.L., 2005. Space-time covariance functions. *J. Am. Stat. Assoc.* 100, 310–321. <https://doi.org/10.1198/016214504000000854>.
- Steinman, A.D., Lamberti, G.A., Leavitt, P.R., Uzarski, D.G., 2017. In: Hauer, F.R., Lamberti, G.A. (Eds.), *Methods in Stream Ecology*, Third edition vol. Volume 1. Academic Press, Boston, pp. 223–241.
- Stumpf, R.P., Culver, M.E., Tester, P.A., Tomlinson, M., Kirkpatrick, G.J., Pederson, B.A., Truby, E., Ransibrahmanakul, V., Soracco, M., 2003. Monitoring *Karenia brevis* blooms in the Gulf of Mexico using satellite ocean color imagery and other data. *Harmful Algae* 2, 147–160. [https://doi.org/10.1016/S1568-9883\(02\)00083-5](https://doi.org/10.1016/S1568-9883(02)00083-5).
- Stumpf, R.P., Davis, T.W., Wynne, T.T., Graham, J.L., Loftin, K.A., Johengen, T.H., Gossiaux, D., Palladino, D., Burtner, A., 2016a. Challenges for mapping cyanotoxin patterns from remote sensing of cyanobacteria. *Harmful Algae* 54, 160–173. <https://doi.org/10.1016/j.hal.2016.01.005>.
- Stumpf, R.P., Johnson, L.T., Wynne, T.T., Baker, D.B., 2016b. Forecasting annual cyanobacterial bloom biomass to inform management decisions in Lake Erie. *J. Great Lakes Res.* 42, 1174–1183. <https://doi.org/10.1016/j.jglr.2016.08.006>.
- 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.* 18, 375–384. <https://doi.org/10.1111/ele.12420>.
- Taylor, R., 1990. Interpretation of the correlation coefficient: a basic review. *J. Diagn. Med. Son.* 6, 35–39. <https://doi.org/10.1177/875647939000600106>.
- Tomlinson, M.C., Stumpf, R.P., Ransibrahmanakul, V., Truby, E.W., Kirkpatrick, G.J., Pederson, B.A., Vargo, G.A., Heil, C.A., 2004. Evaluation of the use of SeaWiFS imagery for detecting *Karenia brevis* harmful algal blooms in the eastern Gulf of Mexico. *Remote Sens. Environ.* 91, 293–303. <https://doi.org/10.1016/j.rse.2004.02.014>.
- Torbick, N., Ziniti, B., Stommel, E., Linder, E., Andrew, A., Caller, T., Haney, J., Bradley, W., Henegan, P.L., Shi, X., 2018. Assessing cyanobacterial harmful algal blooms as risk factors for amyotrophic lateral sclerosis. *Neurotox. Res.* 33, 199–212. <https://doi.org/10.1007/s12640-017-9740-y>.
- U.S. Congress, 2017. Harmful Algal Bloom and Hypoxia Research and Control Amendments Act of 2017. Available at. <https://www.congress.gov/bill/115th-congress/senate-bill/1057> (Pub. S, 1057).
- United States and Canada, 2012. The 2012 Great Lakes Water Quality Agreement – Annex 4. Available at. <http://www.ec.gc.ca/grandslacs-greatlakes/default.asp?lang=En&nav=A1C62826> (Accessed: 2019/06/26).
- Vander Woude, A., Ruberg, S., Johengen, T., Miller, R., Stuart, D., 2019. Spatial and temporal scales of variability of cyanobacteria harmful algal blooms from NOAA GLERL airborne hyperspectral imagery. *J. Great Lakes Res.* 45, 536–546. <https://doi.org/10.1016/j.jglr.2019.02.006>.
- Varouchakis, E.A., Hristopoulos, D.T., 2013. Comparison of stochastic and deterministic methods for mapping groundwater level spatial variability in sparsely monitored basins. *Environ. Monit. Assess.* 185, 1–19. <https://doi.org/10.1007/s10661-012-2527-y>.
- Vecchia, A.V., 1988. Estimation and model identification for continuous spatial processes. *J. R. Stat. Soc. Ser. B Methodol.* 50, 297–312. <https://doi.org/10.1111/j.2517-6161.1988.tb01729.x>.
- Verhamme, E.M., Redder, T.M., Schlea, D.A., Grush, J., Bratton, J.F., DePinto, J.V., 2016. Development of the Western Lake Erie ecosystem model (WLEEM): application to connect phosphorus loads to cyanobacteria biomass. *J. Great Lakes Res.* 42, 1193–1205. <https://doi.org/10.1016/j.jglr.2016.09.006>.
- Wang, X.J., Liu, R.M., 2005. Spatial analysis and eutrophication assessment for chlorophyll *a* in Taihu Lake. *Environ. Monit. Assess.* 101, 167–174. <https://doi.org/10.1007/s10661-005-9154-9>.
- Wells, M.L., Trainer, V.L., Smayda, T.J., Karlson, B.S.O., Trick, C.G., Kudela, R.M., Ishikawa, A., Bernard, S., Wulff, A., Anderson, D.M., Cochlan, W.P., 2015. Harmful algal blooms and climate change: learning from the past and present to forecast the future. *Harmful Algae* 49, 68–93. <https://doi.org/10.1016/j.hal.2015.07.009>.
- Welschmeyer, N.A., 1994. Fluorometric analysis of chlorophyll *a* in the presence of chlorophyll *b* and pheopigments. *Limnol. Oceanogr.* 39, 1985–1992. <https://doi.org/10.4319/lo.1994.39.8.1985>.
- Weniger, E.J., Čížek, J., 1990. Rational approximations for the modified Bessel function of the second kind. *Comput. Phys. Commun.* 59, 471–493. [https://doi.org/10.1016/0010-4655\(90\)90089-j](https://doi.org/10.1016/0010-4655(90)90089-j).
- World Health Organization, 2003. *Guidelines for Safe Recreational Water Environments: Coastal and Fresh Waters*. vol. Volume 1. World Health Organization, Geneva.
- Wynne, T., Meredith, A., Briggs, T., Litaker, W., Stumpf, R., 2018. Harmful Algal Bloom Forecasting Branch Ocean Color Satellite Imagery Processing Guidelines. NOAA Technical Memorandum NOS NCCOS. vol. 252 p. 3.
- Wynne, T.T., Stumpf, R.P., 2015. Spatial and temporal patterns in the seasonal distribution of toxic cyanobacteria in Western Lake Erie from 2002–2014. *Toxins* 7, 1649. <https://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. <https://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. <https://doi.org/10.4319/lo.2010.55.5.2025>.
- Wynne, T.T., Stumpf, R.P., Tomlinson, M.C., Schwab, D.J., Watabayashi, G.Y., Christensen, J.D., 2011. Estimating cyanobacterial bloom transport by coupling remotely sensed imagery and a hydrodynamic model. *Ecol. Appl.* 21, 2709–2721. <https://doi.org/10.1890/10-1454.1>.
- Xu, H., Paerl, H.W., Zhu, G., Qin, B., Hall, N.S., Zhu, M., 2017. Long-term nutrient trends and harmful cyanobacterial bloom potential in hypertrophic Lake Taihu, China. *Hydrobiologia* 787, 229–242. <https://doi.org/10.1007/s10750-016-2967-4>.
- Yan, X., Xu, X., Wang, M., Wang, G., Wu, S., Li, Z., Sun, H., Shi, A., Yang, Y., 2017. Climate warming and cyanobacteria blooms: looks at their relationships from a new perspective. *Water Res.* 125, 449–457. <https://doi.org/10.1016/j.watres.2017.09.008>.
- Yang, J., Hu, M., 2018. Filling the missing data gaps of daily MODIS AOD using spatiotemporal interpolation. *Sci. Total Environ.* 633, 677–683. <https://doi.org/10.1016/j.scitotenv.2018.03.202>.
- Zhao, B., Cai, Q., 2004. Geostatistical analysis of chlorophyll *a* in freshwater ecosystems. *J. Freshw. Ecol.* 19, 613–621. <https://doi.org/10.1080/02705060.2004.9664742>.
- Zhou, Y., Obenour, D.R., Scavia, D., Johengen, T.H., Michalak, A.M., 2013. Spatial and temporal trends in Lake Erie hypoxia, 1987–2007. *Environ. Sci. Technol.* 47, 899–905. <https://doi.org/10.1021/es303401b>.
- Zhou, Y., Scavia, D., Michalak, A.M., 2014. Nutrient loading and meteorological conditions explain interannual variability of hypoxia in Chesapeake Bay. *Limnol. Oceanogr.* 59, 373–384. <https://doi.org/10.4319/lo.2014.59.2.0373>.