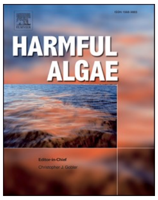
Harmful Algae 139 (2024) 102731

Contents lists available at ScienceDirect

Harmful Algae

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



Defining algal bloom phenology in Lake Erie

Timothy J. Maguire [a,*], Alain Isabwe [b], Craig A. Stow [c], Casey M. Godwin [b]

[a] Academy of Natural Science, Drexel University, Philadelphia, PA, USA
[b] Cooperative Institute for Great Lakes Research, University of Michigan, Ann Arbor, MI, USA
[c] Great Lakes Environmental Research Laboratory, National Oceanic and Atmospheric Administration, Ann Arbor, MI, USA

ARTICLE INFO

Keywords:
Chlorophyll
Microcystin
Phenology
Harmful algal bloom
Lake
Monitoring

ABSTRACT

Elucidating the impact of global climate change on aquatic ecosystems, particularly through phenological shifts in primary producers, is critical for understanding ecological resilience. Here, we focus on the phenological shifts in chlorophyll as a proxy for algae biomass and primary production in aquatic ecosystems, specifically in Lake Erie as well as concentrations of the toxin microcystin. By tracking temporal changes in each, we identified key phenological phases important to estimate duration, magnitude, and intensity of harmful algal blooms (HABs). Determining which influential biotic and abiotic factors such as temperature, wind speed, nutrient availability, and climate change is most important, is a long-term management need for Lake Erie, which can be explored using our methodology. Our novel statistical framework employing Bayesian generalized additive mixed models described seasonal chlorophyll and particulate microcystin concentration from Lake Erie and our simple geometric method identified the start, peak, and end of algal blooms. This research enhances our understanding of the ecological effects of nutrient pollution on aquatic ecosystems and provides a repeatable method for determining phenological events without the need for user defined cutoffs which aids in the management and mitigation of HABs, safeguarding water quality in regions dependent on lakes for drinking water.

1. Introduction

Articulating the timing of cyclical biological events is vital for understanding ecological dynamics and ecosystem resilience (Meng, 2021; Roslin et al., 2021). As global temperatures rise, aquatic ecosystems are experiencing subtle variations in the timing of physical events (Anderson et al., 2021; Hrycik et al., 2024; Visser and Both, 2005; Wang et al., 2024; Woolway et al., 2021). Additionally, changes in seasonal temperatures are expected to extend the growing season leading to phenological mismatches between physical and biological events (Woolway et al., 2021). These variations exert major ecological effects on aquatic biota, mainly by determining the availability of light, nutrients, carbon and oxygen to organisms (Asch et al., 2019; Woolway et al., 2021). Although not fully explored, phenological shifts of aquatic primary producers may offer critical insights into how climate change is restructuring the foundational productivity cycles that fuel aquatic food webs. Phenological shifts in chlorophyll — a measure of algae biomass— can indicate changes in the duration, magnitude, and intensity of harmful algal blooms (HABs) in aquatic ecosystems. These shifts arise from community responses to both local and climatic forcing (Greve, 2003; Ho et al., 2015; Inouye, 2022; Woods et al., 2022).

Tracking chlorophyll temporal changes has been instrumental in identifying phenologically- important phases across lakes. For example, Adams et al. (2021) identified periods of chlorophyll increase and linked these trends to changes in environmental stressors in time series chlorophyll data in lakes ranging across three climate zones. Temperature (Kim et al., 2024; Shi et al., 2019), wind speed, nutrient availability (Shi et al., 2019), climate change (Moe et al., 2022), species invasion (Rohwer et al., 2023), and sensitivity to food web changes (Bailey and Hood, 2024) are examples of biotic and abiotic factors that primarily influence the initiation, duration, and termination of phenological events in aquatic ecosystems. However, a rigorous numerical framework that identifies aquatic phenological events that can be applied in any waterbody and is flexible enough to work with limited spatial and temporal data is still required as managing water quality is globally important.

Lake Erie, a drinking water source for ~11 million people (Miller et al., 2017), has a history of water quality problems due to the regular occurrence of HABs resulting from excessive nutrient loading (Boegehold et al., 2023). Ongoing HAB events in Lake Erie have led to

---

* Corresponding author.
E-mail address: TJM438@DREXEL.EDU (T.J. Maguire).

https://doi.org/10.1016/j.hal.2024.102731
Received 20 June 2024; Received in revised form 6 September 2024; Accepted 25 September 2024

economic losses via impacts to property values, commercial fishing, recreation, and cultural services (Wolf et al., 2022). The 2014 Lake Erie algal bloom caused the city of Toledo to restrict its drinking water system to a cost of ~65 million dollars (Bingham et al., 2015). Consequently, municipal drinking water systems must surveil water quality regularly and quickly employ remediation technologies (He et al., 2016). This surveillance is complicated as algal blooms in Lake Erie can exhibit a range of toxin concentrations relative to their chlorophyll (Rinta-Kanto et al., 2009). Decoupling Lake Erie toxic and non-toxic phenological patterns would facilitate managers to concentrate their remediation efforts and constrain costs within defined temporal windows.

Defining a temporal window that identifies the start and end of a bloom, and the environmental conditions that favor this window, has been separately addressed for chlorophyll concentration, cyanobacteria abundance, and toxin phenological patterns (Beltran-Perez and Waniek, 2021; Kim et al., 2024; Scharfe and Wiltshire, 2019). A primary method for this is the detection of cardinal dates of the onset, peak and end of a phenology in a time series using a Weibull distribution function (Rolinski et al., 2007). This approach ensures that the fitted phenology is unimodal and has lower predicted concentrations at the beginning and end of the season. This function has been fit to time series observations to identify key phenological transition dates (Beltran-Perez and Waniek, 2021; Kim et al., 2024). Typically, the onset is the date when the model fit reaches a specific threshold concentration. The peak and end dates normally correspond to the date of the maximum value of the fit and to a lower threshold, respectively. The scale parameters of the Weibull function can be optimized to provide the best fit to the time series data, identifying the key transition dates in a repeatable manner (Rolinski et al., 2007). Although the Weibull function provides a flexible phenomenological model to extract biologically meaningful transition dates, and despite recent improvements (Belitz et al., 2020), conceptual and practical issues limit the utility of this estimator, which can be problematic for making generalizations from sparsely sampled data and less appropriate to algal communities that exhibit generation times of days, resulting in substantial changes within weeks (Iler et al., 2021).

Previous methods of estimating algal bloom start and end include chlorophyll concentrations exceeding and subsequently dropping below an investigator-defined concentration, conditions relative to an investigator-set percentage of peak conditions, or deviations from an investigator-defined background concentration (Ho and Michalak, 2015; Palmer et al., 2015; Shi et al., 2019). These approaches are difficult when applied to a large basin like western Lake Erie. Chlorophyll conditions in Lake Erie are heterogeneously distributed spatially with concentrations differing from near-shore to offshore (Rowland et al., 2019). Because of this spatial heterogeneity, low concentration sites, that exhibit bloom patterns yet may never exceed the user defined cutoff for start of bloom, are difficult to incorporate into basin wide estimates. The opposite condition also exists where high concentration sites are always observed above the cutoff and again become difficult to incorporate into estimates of lake phenology. Additionally, chlorophyll concentrations are stochastic and observations exceeding a defined limit may be observed, preceding potentially by weeks, the high concentration event defined as a bloom or the seasonal peak. A method that can be consistently applied to a broad spectrum of concentration ranges, be flexible enough to incorporate all possible observations, and be used broadly in any waterbody is needed.

Bayesian generalized additive mixed models are robust to zero-inflated and censored data (Bailey and Hood, 2024; Murphy et al., 2019). These models allow information to be shared across different levels of a model and efficiently incorporate the temporal correlation inherent in the data. They can be useful to incorporate below detection limit observations, and random effects on year and site to accommodate high and low concentration conditions while still fitting seasonally driven patterns. With these models, investigators can avoid independently fitting curves according to structures in their dataset or study design, instead, parameters defining the annual onset, peak, and duration can be fitted hierarchically. Here, we define a novel geometric method of determining chlorophyll and microcystin bloom start, peak, and end that can accommodate the spatial distribution of heterogenous maximum chlorophyll and particulate microcystin concentrations in western Lake Erie. Our method leverages the information sharing capacity of hierarchical Bayesian approaches that enable us to fit seasonal patterns to sites with differing numbers of observations.

2. Material and methods

Chlorophyll (μg/L, $n = 1182$) and particulate microcystin (microcystin, μg/L, $n = 1182$) concentration data were retrieved from the National Oceanographic and Atmospheric Administration (NOAA) National Centers for Environmental Information OneStop web application (Cooperative Institute for Great Lakes Research University of Michigan and NOAA Great Lakes Environmental Research Laboratory, 2019) (Fig. 1). Observations retrieved were made at 8 sites across western Lake Erie from 2012 to 2021 (Boegehold et al., 2023). Samples were collected approximately weekly from May to October, however the frequency and sampling period varied by year and site. Of the 1182 paired observations of chlorophyll and particulate microcystin, 439 particulate microcystin observations were reported as below the detection limit (0.1 μg/L).

To define phenological events of bloom start, end, and peak from the chlorophyll and microcystin data we fit an overall pattern and by-site annual cyclical pattern (broadly defined as low concentrations, bloom to a peak, and descending concentrations to an end).

We independently fit expected chlorophyll and microcystin concentrations (*E(concentration)*) to a smoothed spline of day of year (DOY) in a generalized additive model (GAM, Eq. (1)). GAMs allow us to use the sum of individual smoothers (f_j) applied to predictors (x_j), individual basis functions on each smoother, intercept term (α), and appropriate link functions (g) to fit models to the concentration data (Berhane and

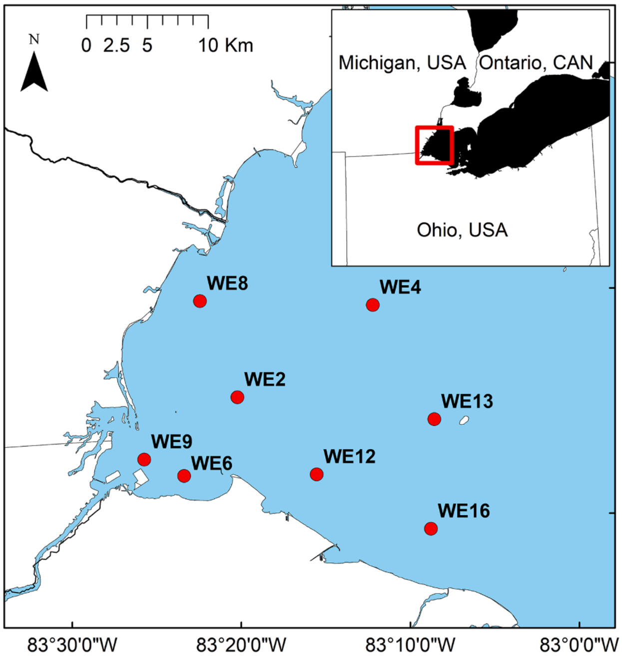


Fig. 1. Chlorophyll (μg/L) and particulate microcystin (μg/L) concentration data were retrieved from the National Oceanographic and Atmospheric Administration (NOAA) National Centers for Environmental Information OneStop web application (data.noaa.gov/onestop). These data were collected at 8 sites across western Lake Erie from 2012 to 2021 (Cooperative Institute for Great Lakes Research University of Michigan and NOAA Great Lakes Environmental Research Laboratory, 2019). Sampling frequency was approximately weekly from May to October, however this frequency varied by year and site.

Tibshirani, 1998; Ruppert et al., 2003). Because the concentration data are structured by site and year, we fit these models as hierarchical GAMs (HGAM) (Pedersen et al., 2019). The HGAM was defined for chlorophyll and microcystin as a smooth on DOY with a cyclic cubic regression basis function (f_a), a smooth on each site's DOY with a cyclic cubic regression basis function (f_b), a random effect smooth on year within site (f_c), and a random effect smooth on site (f_d). Cyclic regression basis functions were used based on the assumption that concentrations will annually increase from and later return to a site-specific baseline concentration. A random effect for year nested within site enabled a method to capture interannual variability in concentrations applied to the site-specific spline, the random effect for each site similarly incorporates a measure of the inter-site variability in the overall western Lake Erie spline.

$$E(Concentration) = g^{-1}\left(\alpha + \sum_{j=1}^{j} f_j(x_j)\right)$$

$$\begin{aligned} f_j(x_i) = {} & f_a(DOY), \\ & f_b(DOY,\ by\ Site), \\ & f_c(Site,\ Year), \\ & f_d(Site) \end{aligned} \tag{1}$$

The Bayesian framework allows for the below detection limit microcystin observations to be fit as distributions between the left censor of zero and the right censor of the detection limit. These models were fit within a Bayesian framework assuming that our response variable error is Gamma distributed (log-link function) because concentration data are left censored at zero. We fit models using the "brms" R package (Bürkner, 2017) to compile code run in Stan (Carpenter et al., 2017). Both HGAMs were fit with three Markov chains, burn-in of 30, 000 iterations, thin of two, and a total of 5000 effective samples per chain.

The splines (Eq. (1), f_a and f_b) fit in the HGAM will follow the concave-down pattern inherited from the seasonal start-peak-end temporal pattern in the concentration data. We used a geometric methodology to define date estimates associated with start and end of the chlorophyll and microcystin bloom in western Lake Erie based in these splines (Fig. 2). Our method repeatably produces dates for these start and end events, independent of any user defined threshold, by extracting the 15,000 samples of conditional smooths f_a and f_b from the Markov chains and plotting these smoothed concentrations per DOY per site from each sample. A line (AB) is drawn between the start of the annual concentration pattern and the peak of the conditional smooth. A second line (CD) is drawn perpendicular to AB where the length of CD is maximized. The DOY associated with point C is designated as bloom start for each of the 15,000 samples. The same operation is mirrored to define bloom end.

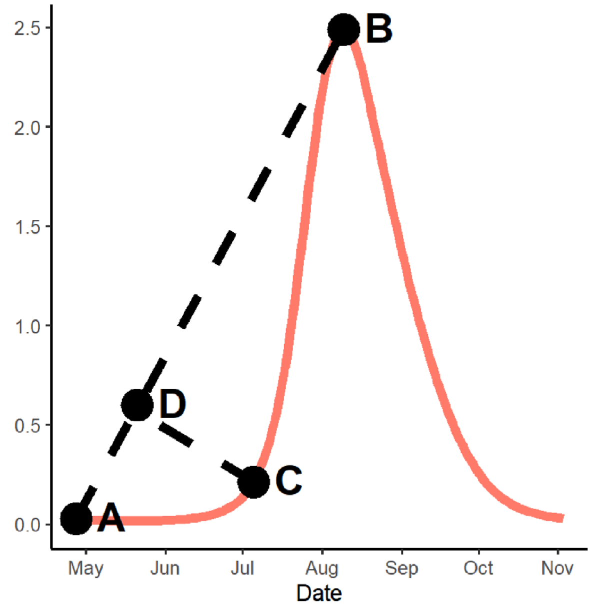


Fig. 2. A geometric method for determining the start and stop dates of blooms was applied to each seasonal trend posterior sample. The method draws a line between the start of the trend and the peak (AB) and a perpendicular line (CD) from the trend line to AB where the length of CD is maximized. Point C is interpreted as bloom start, this process is mirrored to estimate bloom end. This method was applied to each of the 15,000 posterior samples of our modeled seasonality to generate a distribution of start and end dates.

3. Results

The hierarchical Bayesian generalized additive models efficiently incorporated the temporal correlation inherent in the data, provided a framework that incorporated below detection limit observations, and used random effects on year and site to accommodate high and low concentration conditions while still fitting seasonally driven patterns. These models allowed for the extraction of the overall seasonality of both chlorophyll and microcystin across all western Lake Erie sites (Fig. 3a) and the seasonality of individual sites (Supplemental Figure 1a-1 h). Western Lake Erie's overall concentration patterns for chlorophyll and microcystin start low through June, peak in August, and dissipate in October (Table S1). Our geometric method defining start, peak, and stop of this seasonality enabled estimates of bloom phenological events consistently irrespective of concentrations.

The median western Lake Erie chlorophyll bloom start date occurred on July 9 (95 % predictive interval; June 26 – July 21). The median date for chlorophyll bloom peak was August 24 (95 % PI; August 21 – August 27), while the median bloom end estimate was September 29 (95 % PI; September 26 – October 7) (Fig. 4). The median particulate microcystin bloom start date for overall western Lake Erie was July 5 (95 % PI; July 1 – July 8). The median peak date for particulate microcystin occurred on August 10 (95 % PI; August 4 – August 16) and the median end of the microcystin bloom in western Lake Erie overall was September 23 (95 % PI; September 3 – October 7).

Median dates for start-peak-end chlorophyll and particulate microcystin events were estimated by site (Fig. 4, Table S2). Individual site chlorophyll bloom start-date estimates occurred between July 6 (WE8, 95 % PI; June 29 – July 15) and July 9 (WE 16, 95 % PI; June 28 – July 22). Peak chlorophyll *by* site ranged from August 22 (WE9, 95 % PI; August 17 - August 25) and August 27 (WE13, 95 % PI; August 23 – September 8). Chlorophyll bloom end estimates by site were between September 26 (WE6, 95 % PI; September 24, October 1) and October 3 (WE13, 95 % PI; September 27, October 24). Particulate microcystin bloom starts by site from July 5 (WE12, 95 % PI; June 30 – July 8) and July 12 (WE16, 95 % PI; July 7, July 18). By site particulate microcystin median bloom peak occurred between August 10 (WE4, 95 % PI; August 5 – August 14) and August 20 (WE9, 95 % PI; August 12 – August 24). End bloom median estimates for particulate microcystin were September 19 (WE9, 95 % PI; September 15 – September 24) and September 26 (WE8, 95 % PI; September 19 - October 7). Broadly, the start date for both the chlorophyll bloom and the microcystin bloom are similar across all sites, while end bloom date estimates vary across sites with heterogeneous uncertainty by site (Fig. 4). The timing of peak events of each modeled bloom does appear to correspond to site distance from western Lake Erie's major nutrient source (Maumee River, Fig. 1 and 4).

4. Discussion

Using hierarchical Bayesian generalized additive models enabled the fit of multi-station time-series dependent data. The extracted repeated annual cycles (estimated via fitting hierarchical splines) were extracted and used to describe timing of phenological events and support subsequent hypothesis testing within an economically and environmentally

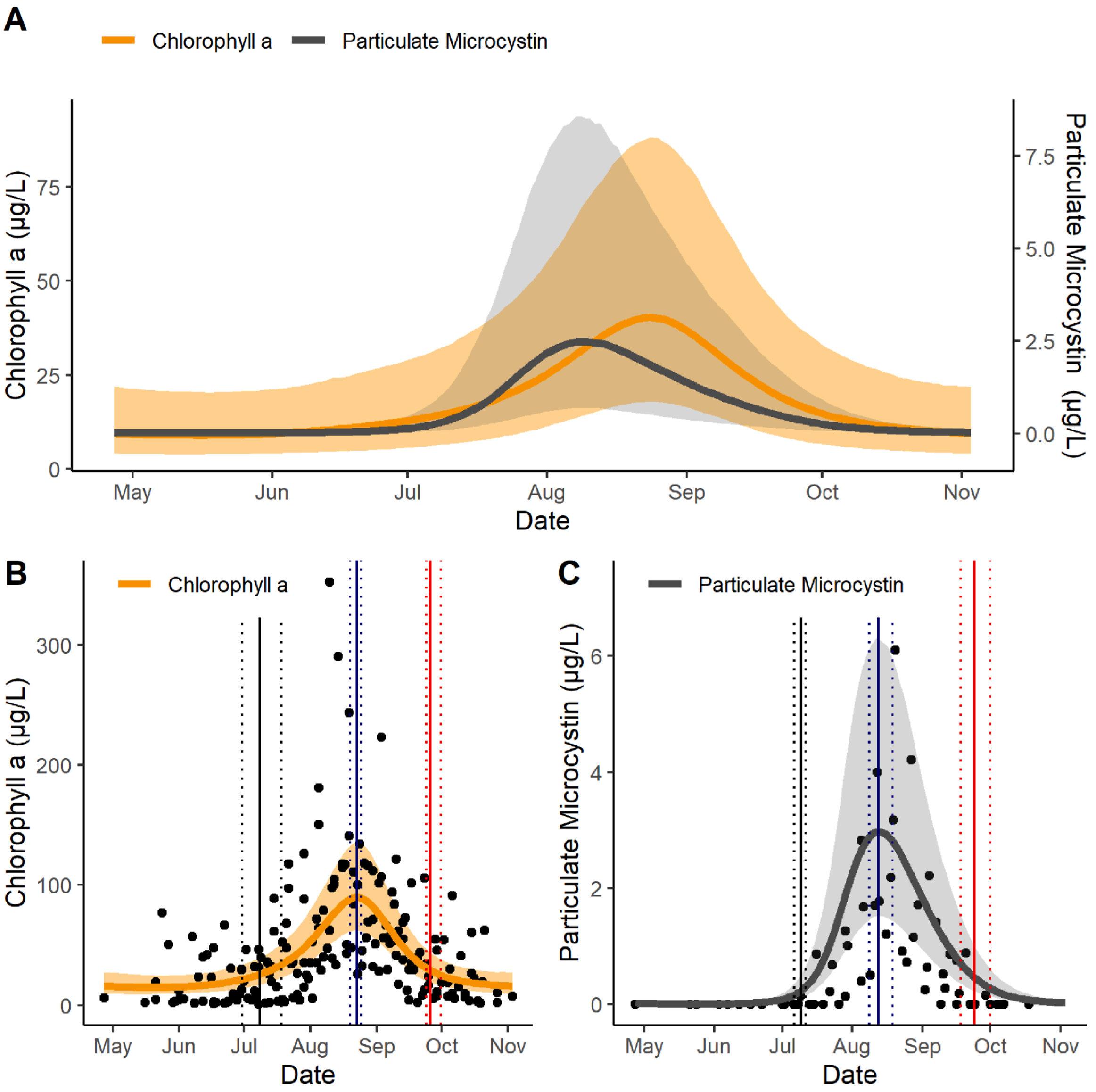


Fig. 3. (A) Western Lake Erie overall pattern of seasonal chlorophyll and particulate microcystin (μg/L) were estimated by Bayesian hierarchical general additive models (with associated 95 % predictive interval). The peak in particulate microcystin concentration precedes the chlorophyll peak. (B) 182 observations of chlorophyll were collected at site WE6, our method of defining the seasonal trend (orange line) and the phenological events of bloom start, peak, and end (vertical lines, 95 % predictive intervals as dashed lines) adequately describe the observable pattern. (C) 61 observations of particulate microcystin were collected at site WE16, 26 of these observations were below the detection limit (BDL). Despite the small number of samples and BDLs our methodology works to define trend and phenological events equally as sites with more data (B).

sensitive water body. This methodology also incorporated data collected at irregular intervals, while other time-series techniques require data collected at specified frequencies or missing values to be imputed to gap-fill (Amorim et al., 2021). Additionally, the lack of rigid data collection frequency requirements allowed for fitting models against time-series with missing data. This type of additive model also was insensitive to differing sample numbers per site and the hierarchical structure helped fit representative splines both through individual sites and the annual growing season of the whole western Lake Erie. Longitudinal datasets with observations below detection limits are common in ecology and environmental monitoring and methods which substitute user defined values (e.g., the detection limit or half the detection limit) fail to incorporate all the information about the data and its structure (Hall Jr et al., 2020; Lawson et al., 2020). Here, the generalized additive models were able to fit posterior distributions to the below detection limit observations that were truncated at 0 and at the detection limit, allowing our model to use all the observations and information in the dataset. The methods here also fit splines to time-series chlorophyll and microcystin concentration data, which varied in across sites from low to high concentration. Our hierarchical Bayesian generalized additive models were adept at overcoming each of these common water quality dataset challenges and produced overall western Lake Erie and by site chlorophyll and microcystin annual bloom patterns.

Seasonal patterns based on the observations were fit for both the overall pattern across western Lake Erie (Fig. 3a) and the individual sites (Supplemental Figure 1a-1 h). Estimates at these two extents are important because modeling efforts to determine the causes of Lake Erie blooms are fit against both spatial scales (Del Giudice et al., 2021; Hellweger et al., 2022). Del Giudice et al. (2021) created a mechanistic model of chlorophyll as a function of nitrogen and phosphorus concentrations, this model was applied to 271 km^2 of Lake Erie's western basin adjacent the Maumee River terminus and employed a fixed nominal chlorophyll bloom initiation date of July 24. Our estimate for the overall western Lake Erie chlorophyll bloom start date is July 9 (95 % PI;

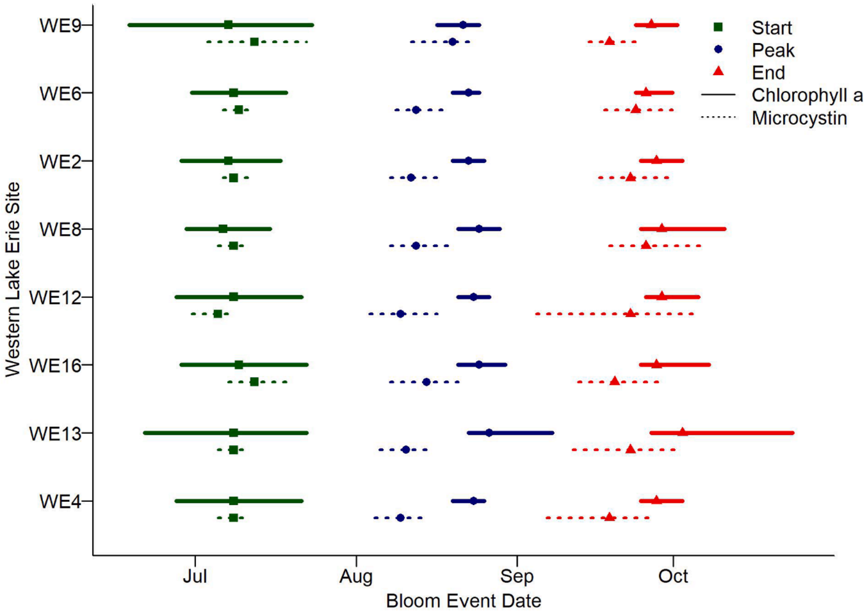


Fig. 4. Bloom phenology derived from the geometric pattern in seasonal concentrations were estimated by site as start, peak, and stop (with associated 95 % predictive intervals) from two independent models of microcystin and chlorophyll concentrations. Particulate microcystin concentrations peak prior to chlorophyll at each site except for WE9. Sites are presented in diminishing distance from the Maumee River (WE9 & WE6 closest, WE13 and WE4 furthest). The difference between microcystin bloom stop and chlorophyll bloom stop diminishes with proximity to the Maumee River.

June 26 – July 21), which would potentially change the behavior of mechanistic models at this scale. Additionally, since our estimate is a posterior distribution, its adoption in further Bayesian predictive models should be seamless. Fang et al. (2019) also generated estimates at this larger spatial scale for algal biomass and spatial extent of the bloom but fit their state-space approach to individual stations. Our splines fit by station would integrate succinctly in this modeling framework as distributions of daily concentrations of chlorophyll.

Our method produced sitewise annual growing season trends that fit the data well and enabled estimates of start-peak-end bloom events (Fig. 3b,c). Both particulate microcystin and chlorophyll concentrations have low concentration sites that would never exceed a predefined threshold of "bloom" and high concentrations sites whose initial values would begin the season exceeding the threshold. Also, the stochastic appearance of anomalously high individual observations of both analytes makes demarcation of seasonality as a percent of the peak difficult. Inherently ambiguous and bespoke site conditions make it necessary to estimate seasonal trends fit to chlorophyll and microcystin concentrations irrespective of their magnitude. Our method generates repeatable, user independent estimates that can be used for similar biological compound seasonal estimates.

Several factors influence the width of the uncertainty about start, peak, and end estimates. The overall pattern of chlorophyll concentrations (Fig. 3A) has wider uncertainty associated with start and end times because this pattern is fit against both high and low concentration sites. This heterogeneity of start and end conditions forces the model to fit higher uncertainty at the extremes resulting in "fatter tails". Some individual sites also have wide uncertainties about start and end estimates. These generally results from sites where the difference between the peak concentration and the early and late sample concentrations are small. Site WE9 and WE13 (Supplemental Figure 1H and 1B, respectively) both have wide uncertainty in start dates because of this diminished difference between peak and early samples. Microcystin patterns are less susceptible to these issues as the samples collected in the beginning and end of the sampling period are often below detection limit. The wide range of uncertainty in start, peak, and end across sites (Fig. 4) reflect a strength of our methods because Bayesian methods fit posterior distributions reflecting the true heterogeneity in environmental data. Honest accounting of uncertainty while confronted with challenging data is an important aspect of quantitative ecology.

Our overall western Lake Erie fitted seasonal pattern (Fig. 3a) showed differences between the start-peak-end phenological patterns of chlorophyll and particulate microcystin. The overall estimate of microcystin bloom starts an average of 3.3 days prior to the overall chlorophyll bloom, microcystin also peaks an average of 14.3 days prior to the chlorophyll bloom peak and ends an average of 6.2 days prior to the chlorophyll bloom. However, earlier start and end dates were weakly supported as 71 % and 84 % of the differences in posterior samples were greater than zero. The early shift for the peak microcystin date was strongly supported with >99 % of differences in posterior samples greater than zero. This early shift for bloom phenological events of microcystin concentration compared to chlorophyll concentrations is immediately applicable to eDNA studies characterizing the microbial population of "toxin" and "non-toxin" algal blooms in western Lake Erie (Wang et al., 2024). Here we cannot ascribe a driver or explanation as to why microcystin phenology has this early shift, however, models incorporating environmental covariates may explain conditions which cause the earlier microcystin peak.

A strength of our methodology was the information sharing that occurred among sites and within sites by fitted site-specific splines and the overall pattern. Because of this interconnection our site-wise phenological patterns informed the overall pattern but were still free to fit site-specific estimates that differed from the overall pattern (Fig. 3b and 3c, Fig. 4). The phenomenological event that best shows this disconnection is the start of bloom estimates for both chlorophyll and microcystin, while in the overall pattern microcystin precedes chlorophyll start, only site WE12 has similar weak evidence for microcystin early bloom start with 71 % of posterior differences exceeding zero (Fig. 4, Table S2). The remainder of the sites have no evidence of a separation in microcystin and chlorophyll bloom start. The shift to earlier peak microcystin is consistent across sites with >99 % of posterior differences exceeding zero, except for site WE9 where 81 % of posterior differences exceed zero (Fig. 4). The bloom end timing is more nuanced for the differences between the blooms. Four sites (WE4, WE9, WE13, and WE16) had microcystin end dates preceded chlorophyll with >95 % posterior differences exceeded zero, three sites (WE2, WE6, and

WE12) >80 % posterior differences exceeded zero, and at WE8 74 % posterior differences exceeded zero. There does not appear to be spatial pattern in the dates of the phenological start or end of microcystin and chlorophyll blooms based on sampling site location relative to nutrient inputs (Maumee River) or water inputs (Detroit River).

Seasonal timing of phenological start and end events are not only insensitive to spatial positions relative to riverine nutrient sources but are also aligned between sites whose concentrations of chlorophyll and microcystin are markedly different, e.g., site WE6 has a mean annual chlorophyll concentration of 46.66 ug/l while WE4 has a mean annual concentration of 14.63 yet their chlorophyl bloom start and end dates are indistinguishable. Future analysis may provide insights into the factors responsible for this timing of events.

Deterministic models that couple biological and physical processes of blooms in western Lake Erie have identified temperature as a driver of bloom onset (Del Giudice et al., 2021). Additionally, here we explicitly hold year as a random variable (Eq. (1)), a model approach that applies year as a fixed effect may also gain insights. Our approach was anchored in defining broad patterns in seasonality from the available observations. Superimposing the growing season trends of chlorophyll and particulate microcystin in western Lake Erie (Fig. 3a) should aid in defining experimental designs and resource allocation as scientists and water managers can plan field operations targeting specific bloom events.

5. Conclusion

Our approach has several advantages over existing phenological estimators. The method presented here can be used in waterbodies with a range of concentrations. This flexibility was required here as within western Lake Erie there is a gradient from very high concentrations inshore to lower concentrations offshore, this variety of concentration magnitude decreases the usefulness of a user defined exceedance value. By using the conditional smooths of DOY per site our method also is insensitive to stochastic observations of unusually high concentrations early or late in the seasonal cycle. Our geometric method is simple, broadly applicable, and can be consistently applied across investigators rather than diverse user defined percentages from the peak. Historic longitudinal datasets of water quality are easily incorporated into this malleable method sequence and are particularly useful for those trying to tell new stories with old data. Our code and data are provided at doi.org/10.5281/zenodo.13625365.

CRediT authorship contribution statement

Timothy J. Maguire: Writing – original draft, Visualization, Methodology, Formal analysis, Data curation, Conceptualization. **Alain Isabwe:** Writing – original draft, Validation, Methodology, Investigation, Conceptualization. **Craig A. Stow:** Writing – review & editing, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Casey M. Godwin:** Writing – review & editing, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding Sources

Funding was awarded to CIGLR through the NOAA Cooperative Agreement with the University of Michigan (NA17OAR4320152 and NA22OAR4320150). Funding was provided by the Great Lakes Restoration Initiative.

Acknowledgments

The authors wish to acknowledge the contributions of CIGLR and GLERL laboratory, field staff, and vessel operators and crews whose efforts in sample collection have made our methods possible. Additionally, we acknowledge the work of the data repository managers who make these data publicly available. This publication is contribution no. 1251 from CIGLR and contribution no. 2058 from NOAA GLERL. Data and code development as part of this manuscript are available through; Maguire, T., Isabwe, A., Stow, C., & Godwin, C. (2024). Defining algal bloom phenology in Lake Erie. In Harmful Algae. Zenodo. https://doi.org/10.5281/zenodo.13625365.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.hal.2024.102731.

Data availability

Data and code are provided on Zenodo with a DOI, this is still editable but will be published if the manuscript is accepted.

References

Adams, H., Ye, J., Persaud, B., Slowinski, S., Kheyrollah Pour, H., Van Cappellen, P., 2021. Chlorophyll-a growth rates and related environmental variables in global temperate and cold-temperate lakes. Earth Syst. Sci. Data Discuss. 2021, 1–30.

Amorim, F.de L.L.de, Rick, J., Lohmann, G., Wiltshire, K.H., 2021. Evaluation of machine learning predictions of a highly resolved time series of chlorophyll-a concentration. Appl. Sci. 11, 7208.

Anderson, E.J., Stow, C.A., Gronewold, A.D., Mason, L.A., McCormick, M.J., Qian, S.S., Ruberg, S.A., Beadle, K., Constant, S.A., Hawley, N., 2021. Seasonal overturn and stratification changes drive deep-water warming in one of Earth's largest lakes. Nat. Commun. 12, 1688.

Asch, R.G., Stock, C.A., Sarmiento, J.L., 2019. Climate change impacts on mismatches between phytoplankton blooms and fish spawning phenology. Glob. Chang. Biol. 25, 2544–2559.

Bailey, J., Hood, J.M., 2024. Biotic and thermal drivers alter zooplankton phenology in western Lake Erie. Limnol. Oceanogr. Lett.

Belitz, M.W., Larsen, E.A., Ries, L., Guralnick, R.P., 2020. The accuracy of phenology estimators for use with sparsely sampled presence-only observations. Methods Ecol. Evol. 11, 1273–1285.

Beltran-Perez, O.D., Waniek, J.J., 2021. Environmental window of cyanobacteria bloom occurrence. J. Mar. Syst. 224, 103618.

Berhane, K., Tibshirani, R.J., 1998. Generalized additive models for longitudinal data. Can. J. Stat. 26, 517–535.

Bingham, M., Sinha, S.K., Lupi, F., 2015. Economic benefits of reducing harmful algal blooms in Lake Erie. Environ. Consult. Technol. Inc., Rep. 66.

Boegehold, A.G., Burtner, A.M., Camilleri, A.C., Carter, G., DenUyl, P., Fanslow, D., Fyffe Semenyuk, D., Godwin, C.M., Gossiaux, D., Johengen, T.H., 2023. Routine monitoring of western Lake Erie to track water quality changes associated with cyanobacterial harmful algal blooms. Earth Syst. Sci. Data Discuss. 2023, 1–39.

Bürkner, P.C., 2017. brms: an R package for Bayesian multilevel models using Stan. J. Stat. Softw. 80, 1–28.

Carpenter, B., Gelman, A., Hoffman, M.D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M.A., Guo, J., Li, P., Riddell, A., 2017. Stan: a probabilistic programming language. J. Stat. Softw. 76.

Cooperative Institute for Great Lakes Research University of Michigan, NOAA Great Lakes Environmental Research Laboratory, 2019. Physical, chemical, and biological water quality monitoring data to support detection of Harmful Algal Blooms (HABs) in western Lake Erie, collected by great lakes environmental research laboratory and the cooperative institute for great Lakes Research [WWW Document]. 10.25921/11da-3x54.

Del Giudice, D., Fang, S., Scavia, D., Davis, T.W., Evans, M.A., Obenour, D.R., 2021. Elucidating controls on cyanobacteria bloom timing and intensity via Bayesian mechanistic modeling. Sci. Total Environ. 755, 142487.

Fang, S., Del Giudice, D., Scavia, D., Binding, C.E., Bridgeman, T.B., Chaffin, J.D., Evans, M.A., Guinness, J., Johengen, T.H., Obenour, D.R., 2019. A space-time geostatistical model for probabilistic estimation of harmful algal bloom biomass and areal extent. Sci. Total Environ. 695, 133776. https://doi.org/10.1016/j.scitotenv.2019.133776.

Greve, W., 2003. Aquatic plants and animals. Phenol. An Integr. Environ. Sci. 385–403.

Hall Jr, L.W., Perry, E., Anderson, R.D., 2020. A comparison of different statistical methods for addressing censored left data in temporal trends analysis of pyrethroids in a California stream. Arch. Environ. Contam. Toxicol. 79, 508–523.

He, X., Liu, Y.L., Conklin, A., Westrick, J., Weavers, L.K., Dionysiou, D.D., Lenhart, J.J., Mouser, P.J., Szlag, D., Walker, H.W., 2016. Toxic cyanobacteria and drinking water: impacts, detection, and treatment. Harmful Algae 54, 174–193.

Hellweger, F.L., Martin, R.M., Eigemann, F., Smith, D.J., Dick, G.J., Wilhelm, S.W., 2022. Models predict planned phosphorus load reduction will make Lake Erie more toxic. Science (80-.) 376, 1001–1005.

Ho, J.C., Michalak, A.M., 2015. Challenges in tracking harmful algal blooms: a synthesis of evidence from Lake Erie. J. Great Lakes Res. 41, 317–325.

Ho, J.C., Michalak, A.M., Chmura, H.E., Kharouba, H.M., Ashander, J., Ehlman, S.M., Rivest, E.B., Yang, L.H., Shi, K., Zhang, Y.Y., Zhang, Y.Y., Li, N., Qin, B., Zhu, G., Zhou, Y., Bingham, M., Sinha, S.K., Lupi, F., Qian, S.S., Stow, C.A., Rowland, F.E., Liu, Q., Rowe, M.D., Anderson, E.J., Stumpf, R.P., Johengen, T.H., Palmer, S.C.J., Odermatt, D., Hunter, P.D., Brockmann, C., Presing, M., Balzter, H., Tóth, V.R., Rinta-Kanto, J.M., Konopko, E.A., DeBruyn, J.M., Bourbonniere, R.A., Boyer, G.L., Wilhelm, S.W., Miller, T.R., Beversdorf, L.J., Weirich, C.A., Bartlett, S.L., He, X., Liu, Y.L., Conklin, A., Westrick, J., Weavers, L.K., Dionysiou, D.D., Lenhart, J.J., Mouser, P.J., Szlag, D., Walker, H.W., Iler, A.M., Humphrey, P.T., Ogilvie, J.E., CaraDonna, P.J., Zhou, B., Shi, K., Wang, W., Zhang, D., Qin, B., Zhang, Y.Y., Dong, B., Shang, M., Roslin, T., Antao, L., Hällfors, M., Meyke, E., Lo, C., Tikhonov, G., Delgado, M.del M., Gurarie, E., Abadonova, M., Abduraimov, O., 2015. Challenges in tracking harmful algal blooms: a synthesis of evidence from Lake Erie. Harmful Algae 8, 108055.

Hrycik, A.R., Isles, P.D.F., Pierson, D.C., Stockwell, J.D., 2024. Winter/spring runoff is earlier, more protracted, and increasing in volume in the. Laurentian Great Lakes Basin. Water Resour. Res. 60 e2023WR035773.

Iler, A.M., Humphrey, P.T., Ogilvie, J.E., CaraDonna, P.J., 2021. Conceptual and practical issues limit the utility of statistical estimators of phenological events. Ecosphere 12, e03828.

Inouye, D.W., 2022. Climate change and phenology. Wiley Interdiscip. Rev. Clim. Chang. 13, e764.

Kim, H.G., Cha, Y., Cho, K.H., 2024. Projected climate change impact on cyanobacterial bloom phenology in temperate rivers based on temperature dependency. Water Res 249, 120928.

Lawson, S.J., Law, C.S., Harvey, M.J., Bell, T.G., Walker, C.F., De Bruyn, W.J., Saltzman, E.S., 2020. Methanethiol, dimethyl sulfide and acetone over biologically productive waters in the southwest Pacific Ocean. Atmos. Chem. Phys. 20, 3061–3078.

Meng, L., 2021. Green with phenology. Science (80) 374, 1065–1066.

Miller, T.R., Beversdorf, L.J., Weirich, C.A., Bartlett, S.L., 2017. Cyanobacterial toxins of the Laurentian Great Lakes, their toxicological effects, and numerical limits in drinking water. Mar. Drugs 15, 160.

Moe, S.J., Hobæk, A., Persson, J., Skjelbred, B., Løvik, J.E., 2022. Shifted dynamics of plankton communities in a restored lake: exploring the effects of climate change on phenology through four decades. Clim. Res. 86, 125–143.

Murphy, R.R., Perry, E., Harcum, J., Keisman, J., 2019. A generalized additive model approach to evaluating water quality: chesapeake Bay case study. Environ. Model. Softw. 118, 1–13.

Palmer, S.C.J., Odermatt, D., Hunter, P.D., Brockmann, C., Presing, M., Balzter, H., Tóth, V.R., 2015. Satellite remote sensing of phytoplankton phenology in Lake Balaton using 10 years of MERIS observations. Remote Sens. Environ. 158, 441–452.

Pedersen, E.J., Miller, D.L., Simpson, G.L., Ross, N., 2019. Hierarchical generalized additive models in ecology: an introduction with mgcv. PeerJ 7, e6876.

Rinta-Kanto, J.M., Konopko, E.A., DeBruyn, J.M., Bourbonniere, R.A., Boyer, G.L., Wilhelm, S.W., 2009. Lake Erie Microcystis: relationship between microcystin production, dynamics of genotypes and environmental parameters in a large lake. Harmful Algae 8, 665–673.

Rohwer, R.R., Hale, R.J., Vander Zanden, M.J., Miller, T.R., McMahon, K.D., 2023. Species invasions shift microbial phenology in a two-decade freshwater time series. Proc. Natl. Acad. Sci. 120, e2211796120.

Rolinski, S., Horn, H., Petzoldt, T., Paul, L., 2007. Identifying cardinal dates in phytoplankton time series to enable the analysis of long-term trends. Oecologia 153, 997–1008.

Roslin, T., Antao, L., Hällfors, M., Meyke, E., Lo, C., Tikhonov, G., Delgado, M.del M., Gurarie, E., Abadonova, M., Abduraimov, O., 2021. Phenological shifts of abiotic events, producers and consumers across a continent. Nat. Clim. Chang. 11, 241–248.

Rowland, F.E., Stow, C.A., Johengen, T.H., Burtner, A.M., Palladino, D., Gossiaux, D.C., Davis, T.W., Johnson, L.T., Ruberg, S., 2019. Recent patterns in Lake Erie phosphorus and chlorophyll a concentrations in response to changing loads. Environ. Sci. Technol. 54, 835–841.

Ruppert, D., Wand, M.P., Carroll, R.J., 2003. Semiparametric regression. Cambridge university press.

Scharfe, M., Wiltshire, K.H., 2019. Modeling of intra-annual abundance distributions: constancy and variation in the phenology of marine phytoplankton species over five decades at Helgoland Roads (North Sea). Ecol. Modell. 404, 46–60.

Shi, K., Zhang, Yunlin, Zhang, Yibo, Li, N., Qin, B., Zhu, G., Zhou, Y., 2019. Phenology of phytoplankton blooms in a trophic lake observed from long-term MODIS data. Environ. Sci. Technol. 53, 2324–2331.

Visser, M.E., Both, C., 2005. Shifts in phenology due to global climate change: the need for a yardstick. Proc. R. Soc. B Biol. Sci. 272, 2561–2569.

Wang, H., Liu, J., Klaar, M., Chen, A., Gudmundsson, L., Holden, J., 2024a. Anthropogenic climate change has influenced global river flow seasonality. Science (80-.) 383, 1009–1014.

Wang, N., Mark, N., Launer, N., Hirtler, A., Weston, C., Cleckner, L., Faehndrich, C., LaGorga, L., Xia, L., Pyrek, D., 2024b. Harmful algal blooms in Cayuga lake, NY: from microbiome analysis to eDNA monitoring. J. Environ. Manage. 354, 120128.

Wolf, D., Gopalakrishnan, S., Klaiber, H.A., 2022. Staying afloat: the effect of algae contamination on Lake Erie housing prices. Am. J. Agric. Econ. 104, 1701–1723.

Woods, T., Kaz, A., Giam, X., 2022. Phenology in freshwaters: a review and recommendations for future research. Ecography (Cop.) 2022, e05564.

Woolway, R.I., Jennings, E., Shatwell, T., Golub, M., Pierson, D.C., Maberly, S.C., 2021. Lake heatwaves under climate change. Nature 589, 402–407.