Author Manuscript

Ten-year Hindcast Assessment of an Improved Probabilistic Forecast System for Cyanotoxin (Microcystins) Risk Level in Lake Erie

Qianqian Liu[1,2+], Mark D. Rowe[3], Richard P. Stumpf[4], Reagan Errera[3], Casey Godwin[5], Justin D. Chaffin[6], Eric J. Anderson[7], Tongyao Pu[5]

[1]*Department of Physics and Physical Oceanography, University of North Carolina Wilmington, Wilmington, NC 28403, USA*

[2]*Center for Marine Science, University of North Carolina Wilmington, Wilmington, NC 28409, USA*

[3]*Great Lakes Environmental Research Laboratory, National Oceanic and Atmospheric Administration, Ann Arbor, MI 48018, USA*

[4]*National Centers for Coastal Ocean Science, National Oceanic and Atmospheric Administration, Silver Spring, MD 20910, USA*

[5]*Cooperative Institute for Great Lakes Research (CIGLR), University of Michigan, Ann Arbor, MI 48018, USA*

[6]*Stone Laboratory and Ohio Sea Grant, the Ohio State University, Put-In-Bay, OH 43456, USA*

[7]Civil and Environmental Engineering, Colorado School of Mines, Golden, CO, USA

[+]Corresponding author: Qianqian Liu (liuq@uncw.edu)

Key Points:

- We improved the microcystin forecast system for harmful algal blooms in western Lake Erie.
- The system combines satellite remote sensing, in situ observations, and numerical modeling.
- Cross-validation showed reasonable skill over regions including surface water, public water system plant intake sites, and bottom waters.

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1029/2024WR038952.

This article is protected by copyright. All rights reserved.

Abstract:

Toxic harmful algal blooms produce public health hazards in freshwater systems around the world. There is a need for forecast systems that can mitigate risk of public exposure to toxins. We improved an approach to predict the spatially- and temporally-resolved probability of microcystins (MCs) exceeding a threshold level (6 μg L^{-1}) in western Lake Erie. This approach combines a 5-day chlorophyll-a forecast model, a weekly-updated regression model predicting MCs from chlorophyll-a, and an empirical relationship between predicted MCs and observed probability of MCs exceeding the threshold calibrated over a hindcast period. We included additional years in the database for calibration and assessment, applied an empirical bias adjustment to MODIS for consistency with Sentinel-3 satellite imagery, and applied a robust Siegel regression method. Cross-validation showed reasonable skill over regions including surface water, public water system plant intake sites, and bottom waters. The forecast also presented useful skill when assessed against two intensive sampling events of *Microcystis* blooms in western Lake Erie in 2018 and 2019. Our results provide a comprehensive assessment of a novel method to forecast MC risk, which may be recalibrated and applied to other systems affected by toxic cyanobacterial blooms, where a similar relationship exists between chlorophyll and toxin concentrations at toxin levels relevant to advisory levels.

Keywords: Lake Erie; cyanobacterial harmful algal bloom; microcystins; probabilistic forecast; cross-validation; skill assessment

Author Manuscript

This article is protected by copyright. All rights reserved.

1. Introduction

Harmful Algal Blooms (HABs) are among the most severe and complex issues threatening coastal ecosystems worldwide (Feng et al., 2024). The most frequent and severe blooms in freshwater ecosystems are caused by cyanobacteria. Predicting cyanobacterial HABs (CHABs) and associated toxins accurately is crucial to mitigate their effects on both the environment and public health. However, the complex association between environmental factors, cyanobacteria population densities and toxin concentrations makes accurate prediction challenging (Wells et al., 2015; Zahir et al., 2024). Lake Erie, as the most productive, shallow and warm of the Laurentian Great Lakes of North America (Figure 1X), has experienced a re-occurrence of CHABs since the early 2000s, after a period of eutrophication in the first half of the 20th century and a period of recovery after the passage of the Clean Water Act in the 1970s (De Pinto et al., 1986; Davis et al., 2019). In this study, by further developing a toxin forecast system for Lake Erie, we provide an approach that may be applied to other systems affected by toxic cyanobacterial blooms.

In Lake Erie, the most severe blooms originate in the shallow western basin, which receives runoff from a large agricultural watershed through the Maumee River. HABs in Lake Erie are dominated by *Microcystis aeruginosa*, which produces microcystins (MCs), a class of cyclic peptides (Rinta-Kanto et al., 2009a, b). MCs are hepatotoxins that can threaten human health through contact with contaminated water, drinking water, or potentially by inhalation of aerosols (Olson et al., 2020; Lad et al., 2022; May et al., 2018). In May 2019, the United States Environmental Protection Agency (USEPA) reviewed US state and international guidelines for advisory levels of MCs in recreational waters; advisory levels ranged from 4 to 20 μg L^{-1} in various jurisdictions, and USEPA recommended an advisory level of 8 μg L^{-1} for contact and recreational use (U.S. EPA 2019). Meanwhile, the World Health Organization has a revised less stringent, short-term contact level of 24 μg L^{-1} for recreational water (WHO, 2020). Western Lake Erie is a source of drinking water to coastal communities. Treatment methods can remove toxins from drinking water, but enhanced treatment may be needed with increasing toxin levels in source water; a multi-barrier system is recommended to protect drinking water from cyanotoxins, including monitoring and early-warning systems (He et al., 2016).

Several methods have been used to monitor CHABs in Lake Erie, including satellite imagery, airborne hyperspectral imaging, and in-situ grab samples to quantify CHAB abundance

This article is protected by copyright. All rights reserved.

Author Manuscript

(Kutser, 2009; Boegehold et al., 2023; Wynne et al., 2022). The sampling program by the National Oceanic and Atmospheric Administration (NOAA) Great Lakes Environmental Research Laboratory (GLERL) and the Cooperative Institute of Great Lakes Research (CIGLR) and in western Lake Erie has been collecting samples weekly to bi-weekly at eight western Lake Erie stations (Figure 1) from June to September since 2008, covering the peak growing season for cyanobacteria (https://accession.nodc.noaa.go v/0187718). In addition, multiple agencies and research groups sample and report MC concentrations (based on the ELISA quantification method) at additional locations and dates in western Lake Erie (Figure 1).

Forecast models have been used to predict CHABs biomass in Lake Erie, including the statistical forecast of annual maximum bloom extent using the Maumee River discharge or total bioavailable phosphorus (Stumpf et al., 2016a), and the short-term, 5-day, Lake Erie HAB Forecast, which predicts HAB position daily over a 5-day forecast horizon using a numerical model initialized from satellite imagery (Rowe et al., 2016). NOAA is currently operating the Lake Erie HAB Forecast driven by currents from the Lake Erie Operational Forecast System (LEOFS; Kelley et al., 2018) hydrodynamic model and Lagrangian particle dispersion model (Rowe et al., 2016), which provides information on HAB transport and vertical mixing with buoyancy (formation of surface scums versus deep mixing). Recent studies have assessed alternative approaches to short-term forecasts of HABs in western Lake Erie (Zhou et al., 2023a; Soontiens et al., 2019).

In contrast, few studies have focused on the forecast of HAB-generated MCs in lakes. Even though cyanobacteria cells are the source of MCs, predicting MC concentrations associated with HABs is challenging due to the complex association between cyanobacteria population densities and MC concentrations. Not all strains of *Microcystis* carry the full set of genes required to produce MCs (Yancey et al., 2022; Dick et al., 2021), and MC production among the MC-producing strains has been suggested to be controlled by interaction among temperature, light levels, nitrogen availability, and environmental variables that induce oxidative stress (Hellweger et al. 2022). The relationship between MC and chlorophyll-a concentrations (MC:chla) is complicated by the varying abundance of MC-producing species and strains in the phytoplankton community (Chaffin et al., 2021). Several studies have applied machine learning (Chan et al., 2007; Jiang et al., 2016) or mechanistic models (Bte Sukarji et al., 2022; Recknagel et al., 2017) to forecast MCs in lakes smaller than Lake Erie, but these methods did not include a

This article is protected by copyright. All rights reserved.

Author Manuscript

spatial prediction. Spatial prediction is an important component of a Lake Erie forecast owing to the spatial extent of the lake, multiple locations of interest such as drinking water intakes, and the importance of transport by currents contributing to risk at specific locations (Rowe et al., 2016; Steffen et al., 2014). Previous studies did not find satellite-derived cyanobacterial abundance to be a strong predictor of the spatial distribution of cyanobacterial toxins (Stumpf et al., 2016b), although more recent studies found improved predictions of MCs from chlorophyll-a when regression models were updated with the most recent sampling data, and when concentration of MCs was near advisory levels (Liu et al., 2020; Qian et al., 2021).

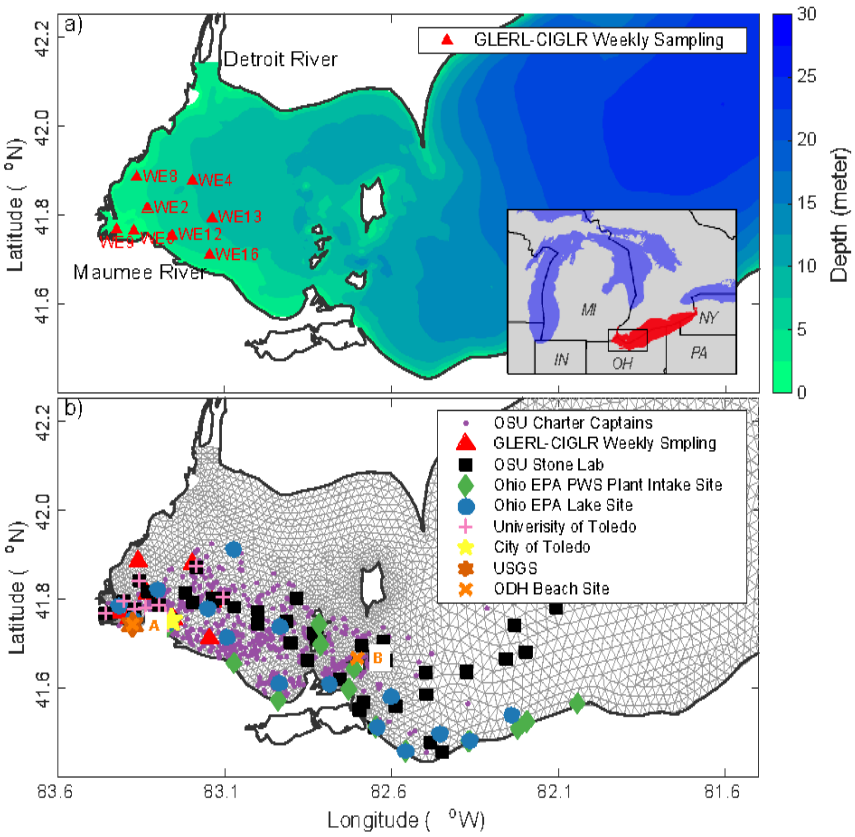

Figure 1. a) Bathymetry of western Lake Erie (in m) and locations of stations sampled by GLERL and CIGLR. b) LEOFS (FVCOM) model grid of the western portion of Lake Erie, and locations of all MC data used to calibrate and assess the forecast system from 2014 to 2023 including the Ohio Department of Health (ODH) Beach sites and the public water system (PWS) intake sites by Ohio EPA. Site A represents the Maumee Bay State Park (Erie), and site B represents Kelleys Island State Park. The spatial resolution of the LEOFS (FVCOM) grid is ~ 2 km in the central basin, 1.5 km in the western basin, and 0.5 km in Maumee Bay and the islands.

This article is protected by copyright. All rights reserved.

Author Manuscript

The inset map in panel a) shows the location of western Lake Erie between the US states of Michigan (MI) and Ohio (OH).

We previously developed an approach to predict the *probability of exceeding* public health advisory levels of microcystins (MCs) in Lake Erie's western basin (Liu et al., 2020). This approach (Figure 2) combines satellite-derived cyanobacterial biomass index, in-situ chlorophyll-a and MC concentration observations, Lake Erie HAB Forecast, and statistical models. The forecast assumes a constant MC:chla relationship over short-term (5-day) periods. Previous studies have shown that this assumption is weak at low MC and chlorophyll-a concentrations (i.e., MC < 1.6 μg L^{-1}, chlorophyll-a < 20 μg L^{-1}), but may hold sufficiently to make valuable predictions at toxin concentrations near public health advisory levels (Liu et al., 2020; Chaffin et al., 2021). The MC:chla ratio can range from 0.05 to 0.55, but most of the variation occurred at chlorophyll-a concentrations less than 20 μg L^{-1} (Chaffin et al., 2021), therefore, a sample with a high MC:chla ratio and low chlorophyll-a concentration has a low probability of exceeding 6 μg L^{-1}. The final product of the system developed by Liu et al. (2020) was the probability of MC concentration exceeding the toxin threshold, based on an exceedance probability model trained over a limited period of time from 2014 to 2016. With the accumulation of observations, there is a need to extend the training dataset to more recent years, adjust for bias between different satellite periods of record, assess alternative MC:chla regression methods, and revise the exceedance probability function to reduce bias.

The forecast was calibrated using a multi-year hindcast period, requiring consistent satellite data. However, various satellites, including the Medium Resolution Imaging Spectrometer from 2002-2011 (MERIS; Wynne and Stumpf, 2015; Rowe et al., 2016), Moderate Resolution Imaging Spectroradiometer (MODIS) from 2012 to 2016, and Sentinel-3 with Ocean Land Color Instrument (OLCI) sensor from 2017 to the present, have been used to detect HABs in Lake Erie. After MERIS was discontinued in 2012, MODIS sensors, while noisy and less sensitive, were the only suitable option (Wynne et al., 2013) until OLCI. To ensure consistency between different satellite data, empirical bias adjustment is needed for MODIS data (with details shown in the Supporting Information).

In this study, we enhanced the probabilistic MC forecast by extending the training dataset to cover 2014 to 2023, testing different regression methods, bias-adjusting MODIS data used,

This article is protected by copyright. All rights reserved.

Author Manuscript

and revising the exceedance probability. We also offered a more comprehensive assessment of the MC risk forecast in Lake Erie, an approach that may apply to other regions affected by HABs after recalibration with local data.

2. Materials and methods:

2.1 Predictive model

The probabilistic MC forecast system for Lake Erie (Liu et al., 2020) produces the probability of the MC concentration exceeding a threshold concentration of MCs in the western basin of Lake Erie (Figure 2). The system builds on the Lake Erie HAB Forecast system (M1 in Figure 2) to predict the CHAB chlorophyll-a concentration and movement (D1), initialized from satellite-derived CHAB concentration. The satellite-derived CHAB concentration was from MODIS imagery from 2014 to 2016 and OLCI imagery during and after 2017. Future forecast applications will use OLCI. Potential bias would be introduced in D1 by calibrating the model using MODIS for part of the calibration hindcast period and making forecasts using OLCI. To bias-adjust MODIS for consistency with OLCI images, we used paired MODIS and OLCI images on the same days in 2017 to develop a regression model (a hockey stick relationship) relating CI_{MODIS} to CI_{OLCI}. The regression model generated by the comparison was used to bias-adjust MODIS for the initialization of HAB Forecast runs during the training years. Please see the Supporting Information for detailed information about the generation of the hockey stick relationship.

The predicted CHAB chlorophyll-a concentration is then converted into MC concentration (D2) through a regression model (M2). The original forecast system (Liu et al., 2020) employed a least-squares linear regression of MC concentration on chlorophyll-a concentration, using observations from the most recent sampling event (M2 in Figure 2). By updating the MC:chla relationship weekly, the model accounts for the varying toxin content of the phytoplankton community (Liu et al., 2020; Chaffin et al., 2021; Song et al., 2021). The number of paired samples of MC and chlorophyll-a varies throughout the season due to the sampling program; the average number of pairs is 10 per week. To enhance the model, we employed a "robust" Siegel Regression technique using the R package *RobustLinearReg* (R Core Team, 2021; Theil, 1992). The Siegel regression method is based on the median of slopes between all data pairs. In contrast with the least squares linear regression, Siegel regression is

This article is protected by copyright. All rights reserved.

Author Manuscript

less sensitive to outliers. In our assessments, Siegel regression produced slightly greater overall skill than the similar Thiel-Sen regression method, and was less sensitive to outliers than least-squares regression. For brevity, we only report results for Siegel regression.

Uncertainties introduced by M1 and M2 in Figure 2 are expressed as probability (D3) through a statistical model (M3), which is based on modeled and observed MC matchups from a hindcast period. For the hindcast assessment, we used an extensive data set of MC observations from multiple agencies and researchers, covering all HAB-affected areas of Lake Erie (Figure 1). To reduce bias introduced by assuming probability approaches unity at high predicted MC values (Liu et al., 2020; their Fig. 11), we replaced the polynomial fit in M3 in the original forecast system with a spline fit using the R-package smooth.spline. We further assumed that the probability does not increase beyond that of the highest binned interval at higher predicted MC values for which observations were too limited to calculate a proportion.

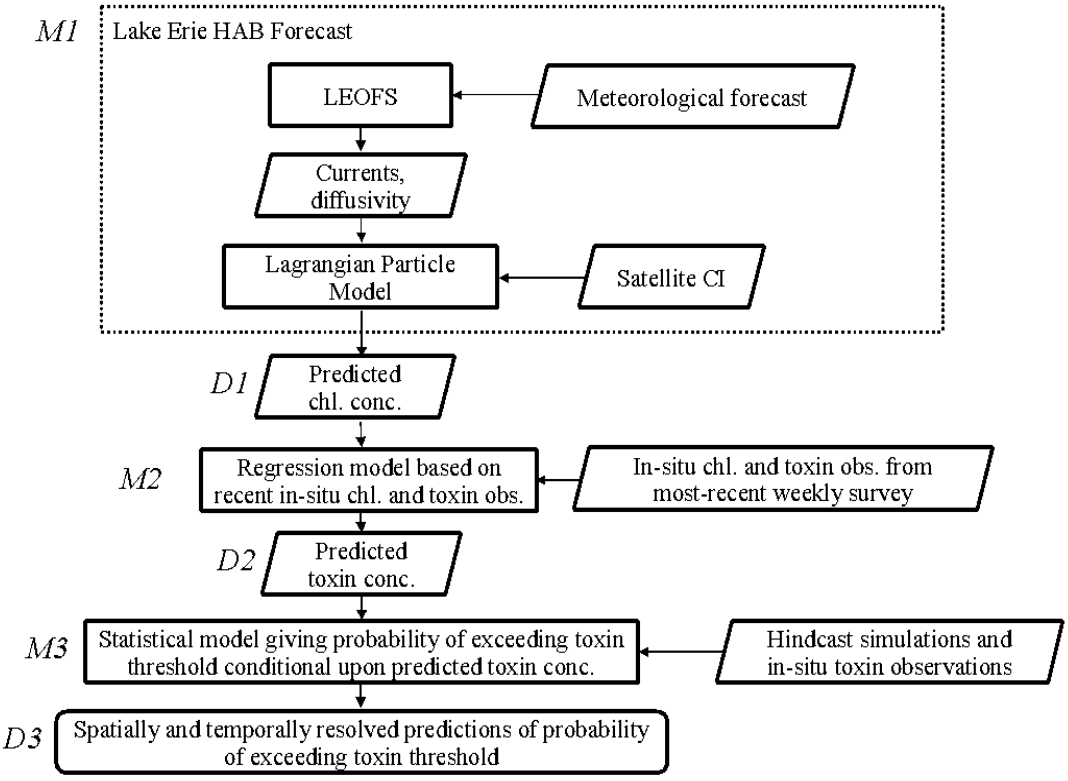

Figure 2. The structure of the probabilistic toxin forecast model, modified from Liu et al. (2020). The Lake Erie Operational Forecast System (LEOFS) provides forecasted currents and vertical

This article is protected by copyright. All rights reserved.

turbulent diffusivity in Lake Erie. Model components (M1,2,3) and model-produced data (D1,2) are labeled for reference in the text.

2.2 In-situ Observations

To develop the MC:chla regression (Figure 2, M2), we used paired observations of MCs and chlorophyll-a (from the same grab sample) from weekly to bi-weekly water samples collected at eight stations by GLERL-CIGLR (Figure 1; https://accession.nodc.noaa.gov/0187718). The GLERL-CIGLR dataset has produced the longest-running record of paired MC and chlorophyll-a concentrations (since 2008), whereas the other data sources listed in Figure 1 only began to measure both chlorophyll-a and MCs more recently (since ~2013). Samples were collected at the surface and bottom, and analyzed using the methods provided in Boegehold et al. (2023). We used surface and bottom samples in our analysis, in contrast to the use of surface samples only by Liu et al. (2020). For assessment of the MC forecast and development of the probability function (Figure 2; M3), we used observations of MC toxin concentrations from multiple agencies and researchers, including GLERL-CIGLR weekly sampling, Ohio EPA, Ohio State University Stone Lab, Stone Lab Charter boat captain sampling program, and USGS. In addition, we obtained beach monitoring data from the Ohio Department of Health (Maumee Bay and Kelley's Island State Parks). To obtain spatially detailed measurements of HABs and MC concentration in Lake Erie on two days, a binational collaboration by the USA and Canada conducted two intensive "HABs Grab" sampling events of *Microcystis* blooms in the western basin of Lake Erie on 9 August 2018 and 7 August 2019 (Chaffin et al., 2021). The program collected 100 and 172 grab samples within a six-hour window in 2018 and 2019, respectively, and analyzed MC concentrations using the ELISA method (Chaffin et al., 2021). Due to the high spatial resolution, the HABs Grab locations are not shown in Figure 2b, and will be shown in the corresponding assessment figure. Because the detection limit and the reporting limit varied across samples and data sources (ranging from 0.10 µg L^{-1} to 0.30 µg L^{-1}), we replaced below detection results (coded as "bdl", "nd", or "<" in the original dataset) with 0 in the regression for step M2 and in skill assessment. The GLERL-CIGLR program reported particulate and dissolved MCs separately, which we combined to represent total MCs for use in calibration and assessment. Dissolved MCs were typically a small proportion of total MCs when MC

This article is protected by copyright. All rights reserved.

Author Manuscript

concentrations were near the advisory level (Liu et al., 2020). All other programs reported total MCs (particulate and dissolved combined).

2.3 Model skill statistics

To assess the skill of the Lake Erie HAB Forecast (M1), the following metrics for binary prediction of HAB/NO-HAB events (cyanobacterial chlorophyll-a > 23 μg L^{-1}) were used (Rowe et al., 2016): a) rate of accurate HAB occurrence, $a/(a+c)$, and rate of accurate No-HAB prediction, $d/(b+d)$, where a represents correctly predicted events (hits); b represents false events (false positive); c represents false negatives (misses), and d represents correct nonevents. b) Pierce skill score (PSS) gives the hit rate minus the false positive rate: $PSS = (ad - bc)/(b+d)(a+c)$. PSS values range from -1.0 to 1.0, with positive values indicating that the hit rate is greater than the false positive rate, therefore the model has greater skill than a random forecast or constant HAB or no-HAB prediction (Hogan and Mason, 2012).

As a benchmark for HAB Forecast skill calculation, we made a "persistence" forecast in which chlorophyll concentrations were unchanged from the satellite image that was used to initialize the model, which represents the best available information to a forecast user in the absence of a useful model (Rowe et al., 2016). The difference in PSS between the HAB Forecast model and the persistence forecast (PSS model – PSS persistence) was used to test whether the HAB Forecast model had greater skill than the benchmark persistence forecast.

We carried out K-fold cross-validation (Geisser, 1975; Stone, 1974) to assess the skill of the forecast system. The data set comprises 10 years (2014-2023) of hindcasts, and corresponding observations. In cross-validation the data were divided into 10 subsets, each consisting of a single year of data, and for each subset the spline fit was calibrated based on data from the other 9 subsets (years) so that the leave-out subset was not used in (or autocorrelated within) the calibration. The model was tested against the leave-out subset by matching a predicted probability of exceeding the MC threshold with each in-lake observation of MC concentration at the location, time, and depth of the observation. This procedure generated 9 leave-out validation datasets (one for each year). Considering the strong interannual variability in HABs, we combined the 9 leave-out datasets to produce the skill metrics across all years. We refer to this method as "leave-out-one-year cross-validation". We used the reliability diagram to assess the skill of the toxin forecast system. The reliability diagram was created by binning the

This article is protected by copyright. All rights reserved.

Author Manuscript

paired predictions and observations within ranges of predicted probability of exceeding the MC threshold on the x-axis, then the proportion of observed MC concentrations exceeding the threshold value was calculated for each bin, and plotted on the y-axis, with a binomial confidence interval. The reliability diagram can tell the user how closely the forecast probability corresponds to the actual chance of observing the event (Hartmann et al., 2002). Using the same data set as the reliability diagram, the interpretation of model skill was further visualized using a box and whisker plot, visualizing the distribution of the observed toxin concentrations against binned intervals of the forecast probability.

3. Results

3.1 Application of MODIS-OLCI conversion formula

We applied a hockey stick relationship to CI_{MODIS} imagery to avoid bias introduced by the transition from MODIS to OLCI imagery in M1. For detailed information, please see the Supporting Information. The skill of the Lake Erie HAB Forecast initialized with MODIS imagery was improved after bias-adjustment of MODIS imagery. Without bias adjustment, the Lake Erie HAB Forecast often performed worse than persistence (PSS<0) (Figure 3a). With the bias adjustment, the model was always more accurate than the persistence forecast (PSS > 0, Figure 3b).

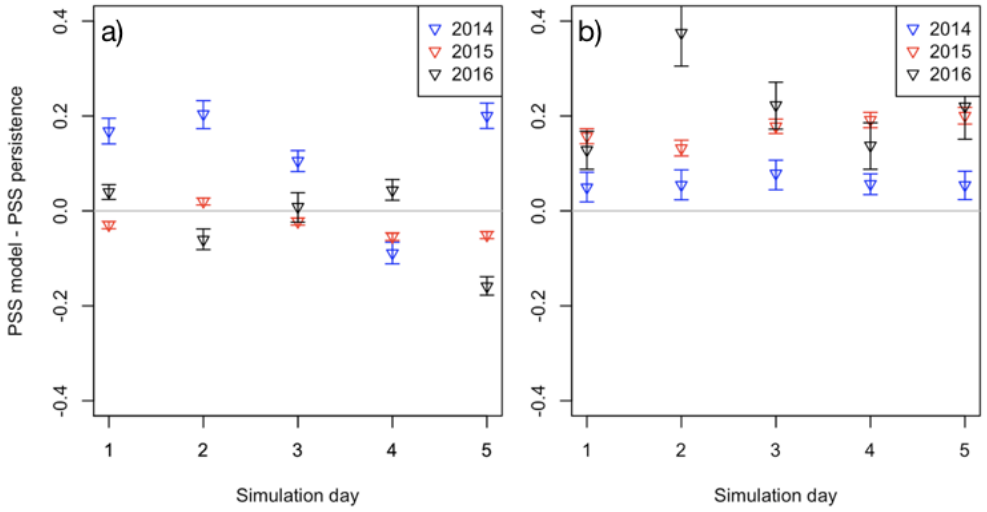

Figure 3. Pierce Skill Score (PSS) model – PSS persistence for a) the HAB Forecast initialized from original MODIS images and b) the HAB Forecast initialized from adjusted MODIS images. Positive values indicate that HAB Forecast has greater skill than the persistence forecast. Error

This article is protected by copyright. All rights reserved.

Author Manuscript

bars represent the 95% bootstrap confidence interval on the difference in PSS for hindcast simulations.

Having applied the MODIS bias adjustment to the hindcast years, replaced the linear regression in M2 with the more robust Siegel regression method, and included the additional years from 2017 to 2023, we updated the modeled and observed MC matchups and generated a spline fit to the observed proportion of observations exceeding 6 µg L^{-1} MCs over binned intervals of predicted toxin concentration (Figure 4). We assumed that the probability of exceeding the MC threshold was equal to the proportion exceeding the threshold in the highest binned interval for which sufficient observations were available to calculate the proportion (horizontal straight-line extension of the orange spline fit at high predicted MCs, >40 µg/L, in Figure 4). This assumption was necessary because observations become increasingly rare as MC concentrations increase. The spline fit function served as M3 (Figure 2) when running the forecast.

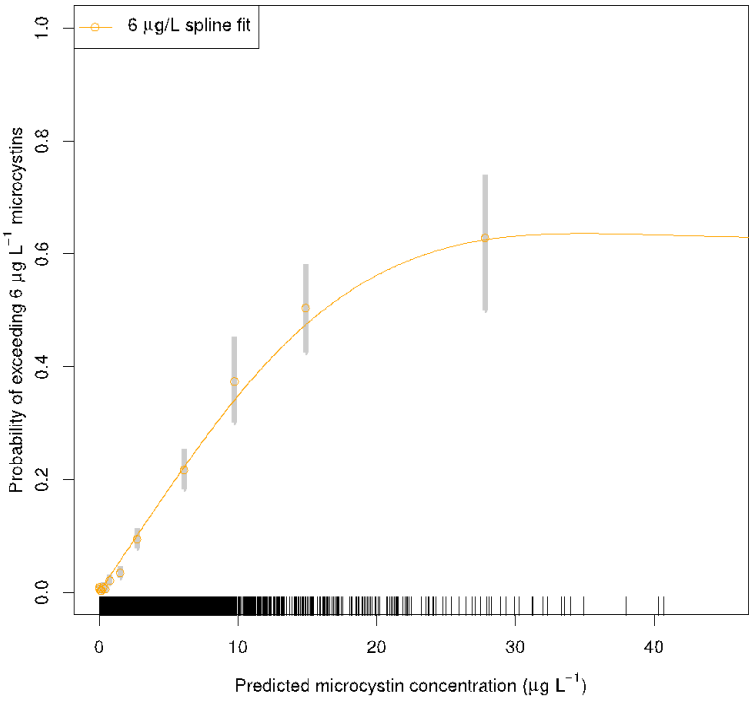

Figure 4. Proportion of observed MCs exceeding the 6 µg L^{-1} threshold for binned intervals of predicted MCs (symbol) with 95% binomial confidence interval (vertical gray lines). The rug on the horizontal axis indicates the locations of predicted MC values. Spline fit (line) used to predict

This article is protected by copyright. All rights reserved.

Author Manuscript

the probability of exceeding 6 μg L^{-1} MCs (M3 in Figure 2) from model-predicted MC concentration.

The results of the leave-out-one-year cross-validation are shown in Figure 5. The forecast system's reliability is high for forecast days 0-2 (Figure 5b), and decreases slightly for days 3-5 (Figure 5d). This is because, with the increase of forecast horizon, the assumption that the MC:chla relationship remains constant during the forecast period becomes weaker, leading to reduced model performance. Additionally, the increased uncertainties inherent in the HAB forecast system also weaken the model's performance as forecast horizon increases. Confidence intervals of the observed proportion exceeding the MC threshold include the 1:1 line in most cases, indicating little bias in the forecast. The positive slopes of the quantile regression lines in the whisker plots (Figures 5a, 5c) show that with increasing forecast probability, the fraction of observed toxin concentration exceeding the threshold increases, indicating forecast skill.

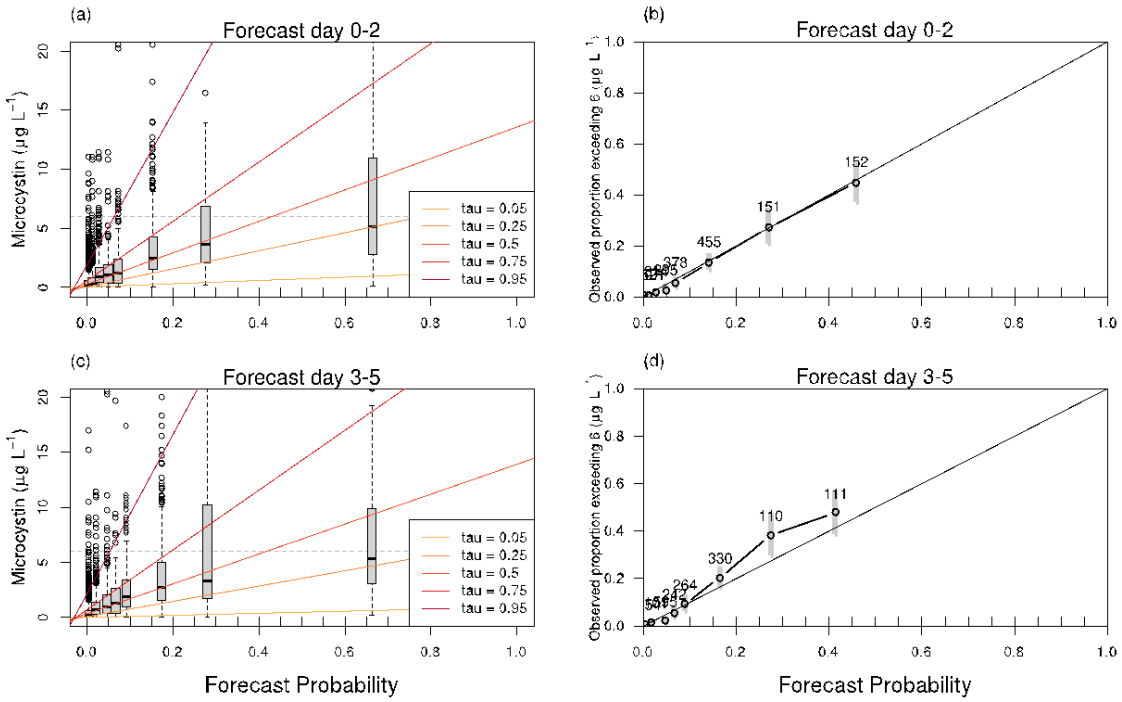

Figure 5. Model skill for data outside the calibration set, based on the 10-fold leave-out-one-year cross-validation hindcast assessment for the period of 2014-2023 using whisker plots (a, c) and reliability diagrams (b, d) for nearer-term forecasts (day 0-2) and longer-term forecasts (day 3-5). Forecast probabilities are plotted at eight quantiles of all predictions for which sample data

This article is protected by copyright. All rights reserved.

Author Manuscript

are available. The boxes in the whisker plots represent the interquartile ranges with the middle line representing the median, the whiskers extend to the most extreme data point which is no more than 1.5 times the interquartile range from the box, and values beyond the whiskers are plotted as symbols (circles). The horizontal dashed line represents the threshold MC concentration of 6 μg L^{-1}. The colored lines (tau) represent the quantile regressions. The gray bars in the reliability diagrams indicate the 95% binomial confidence interval on the estimate of the proportion, and the numbers next to the gray bars represent the numbers of samples used in each probability bin. In a perfect forecast, points would fall on the 1:1 line of the reliability diagram.

While the probability function in Figure 4 was calibrated for the threshold MC concentration of 6 μg L^{-1}, the data in Figure 5 can be used to define the observed probability of exceeding other threshold concentrations, including the public health advisory level of 8 μg L^{-1}. Although the analytical precision of the ELISA method may be insufficient to resolve the difference between 6 and 8 μg L^{-1} MC concentrations (Qian et al., 2015), for the purpose of presenting the forecast to the public, we defined four MC risk levels, with the quantified probability of exceeding 6 and 8 μg L^{-1} levels because these levels were former or current advisory levels. In Table 1, we summarized the proportion of observed toxins exceeding 6 μg L^{-1} (the Ohio advisory level before 2019) and 8 μg L^{-1} (the USEPA recommended advisory level in 2019) for MC risk levels using 2014-2023 dataset. We defined the bounds of the risk level categories so that the low risk category had an observed proportion exceeding 6 and 8 μg L^{-1} < 0.1, and the higher risk categories had sufficient observations to provide non-overlapping confidence intervals, with the exception that the medium and high risk categories have effectively the same proportion exceeding 8 μg L^{-1}.

In Figure 6, we further examined the model's skill for various locations of interest to stakeholder groups, including the PWS plant intake sites in Figure 1b (Figures 6a, b), bottom water samples (Figures 6c, d), and ODH Beach data (Figures 6e, f) for forecast days 0-5. The forecast system has good reliability for the PWS sites and bottom waters. However, for the beach sites the predicted probability was considerably lower than the observed proportion exceeding the MC threshold.

This article is protected by copyright. All rights reserved.

Author Manuscript

Author Manuscript

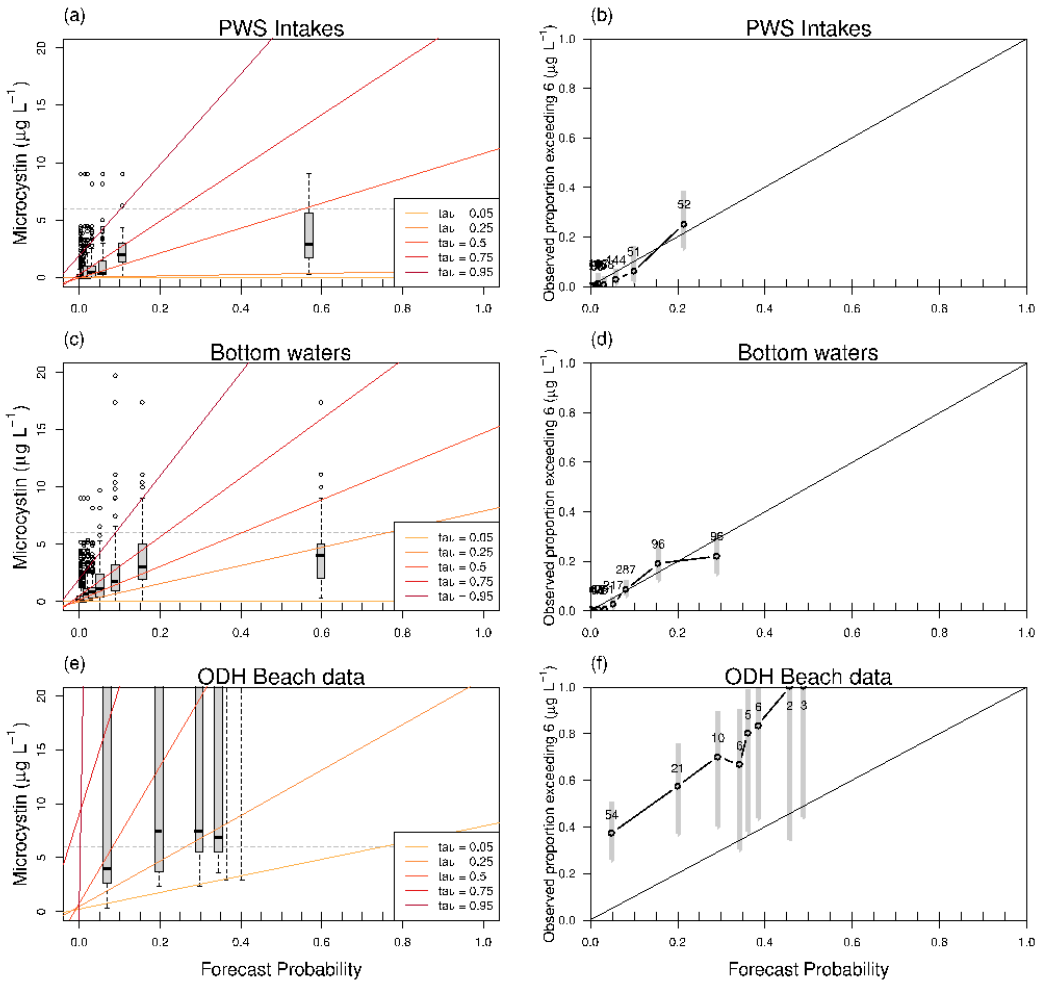

Figure 6. Box and whisker plots and reliability diagrams to examine the forecast system's skill for the PWS intake sites (a, b), all bottom water samples from Figure 1b stations including USGS station and GLERL-CIGLR stations (c, d) and ODH Beach data (e, f) for forecast days 0-5. The forecast probability ranges from 0.0 to 1.0, representing 0% to 100% probability.

3.2 Hindcast assessment using intensive sampling dates

We ran hindcast simulations for the intensive HABs Grab field sampling dates in 2018 and 2019, utilizing Siegel regression models for the most recent CIGLR/GLERL sampling dates preceding the 2018 and 2019 intensive sampling dates (Figure 7). Maximum MC and chlorophyll-a concentrations were lower on the intensive sampling date in 2018 (maximum observed concentrations were 6.38 µg L^{-1}, 75.20 µg L^{-1}, respectively) than in 2019 (maximum observed concentrations were 45.56 µg L^{-1}, 196.16 µg L^{-1}, respectively). Figures 8a and 8b show the modeled chlorophyll-a concentration and the MC risk levels for the 2019 intensive sampling

This article is protected by copyright. All rights reserved.

Author Manuscript

date. In the hindcast simulation, the HAB forecast was initialized from a satellite image on the same day and MC:chla regression was from two days prior. The model-predicted HAB area showed reasonable agreement with HAB area indicated by in-situ observations, with some disagreement occurring due to the mismatch between satellite-predicted and observed chlorophyll-a (Figure 8a). The observed MC concentrations greater than 6 μg L^{-1} were associated with higher MC risk levels, indicating a potentially useful forecast (Figure 8b).

To examine the sensitivity of the MC forecast to the HAB forecast initialization date, we carried out another hindcast for the 2019 HABs Grab date initialized from a satellite image four days prior (Figures 8c,d). The longer forecast horizon led to less agreement between the observed and modeled HAB area. Most (80%) of the MC concentrations exceeding 6 μg L^{-1} were associated with high and medium-high MC risk levels, indicating the forecast still has skill.

On the 2018 intensive sampling date, both modeled and observed chlorophyll concentrations were relatively low (Figure 8e,f). The model hindcast probability of exceeding 6 μg L^{-1} was low for all stations (100%). Consistent with the low predicted probability of exceedance, the observations showed that only one station close to Port Clinton, OH had MCs exceeding 6 μg L^{-1}.

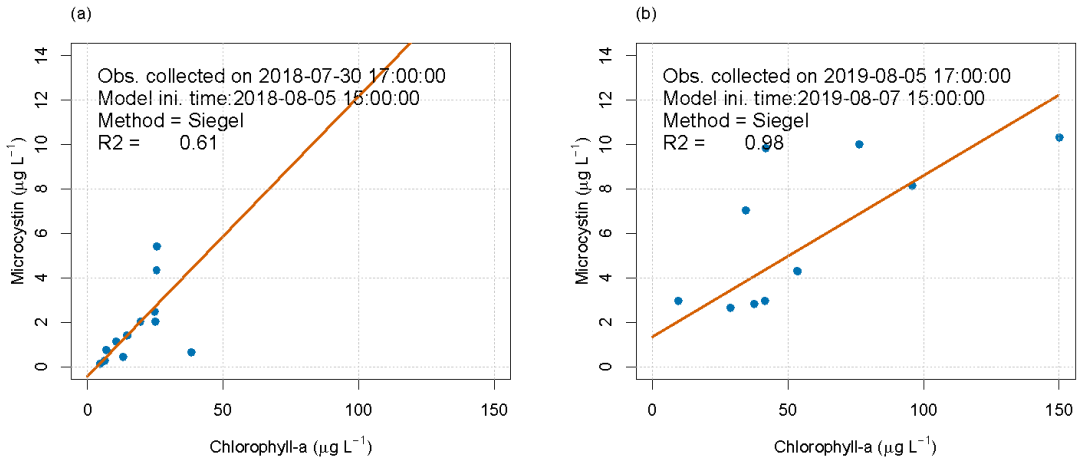

Figure 7. Siegel regression between chlorophyll-a and MCs for the most recent CIGLR/GLERL sampling dates over the western basin of Lake Erie preceding 2018 and 2019 intensive sampling dates (Figure 8), collected on 30 July 2018 and 5 August 2019, respectively.

This article is protected by copyright. All rights reserved.

Table 1. The binned intervals of forecast probability used to define MC risk levels with the observed proportion of grab-sample observations exceeding the threshold MC levels of 6 μg L^{-1} and 8 μg L^{-1} for 2014-2023 dataset. The fractions inside the parentheses represent the 95% binomial confidence intervals of the proportion.

Forecast Probability Bin range (based on 6 μg L^{-1})	**MC Risk Level**	**Obs. Proportion Exceeding 6 μg L^{-1}**	**Obs. Proportion Exceeding 8 μg L^{-1}**
0-0.099	low	0.014 (0.012 - 0.016)	0.0092 (0.0076-0.011)
0.099-0.23	medium-low	0.17 (0.14-0.20)	0.16 (0.14-0.19)
0.23-0.33	medium-high	0.31 (0.26-0.37)	0.35 (0.29-0.41)
0.33 - 1.0	high	0.46 (0.40-0.52)	0.31 (0.22-0.42)

To illustrate how a user of the forecast could reduce their risk of MC exposure, we conducted a comparative analysis of the likelihood of encountering toxin levels exceeding 8 μg L^{-1}. The analysis compares a scenario in which the user avoided elevated risk areas of the lake identified by the forecast, in comparison to a scenario in which the user had no forecast risk information, and therefore was equally likely to encounter any of the MC observations on that date. To estimate the relative risk for each scenario, we calculated the proportion of HABs Grab observations >8 μg L^{-1} in areas identified as lower risk by the forecast compared to the proportion >8 μg L^{-1} in the full HABs Grab data set on a given date. If a user chose to avoid the two highest risk categories (orange and red regions of the forecast map) according to the forecast initiated on 2019-08-07, their observed frequency of encountering MC concentration >8 μg L^{-1} would decrease from 15.4 % to 0 % (Figure 8a,b). By applying the same criteria to the less accurate forecast initiated four days earlier, the user's observed frequency of encountering MC concentrations > 8 μg L^{-1} would drop from 14.7 % to 4.2 % (Figure 8c,d). Applying the same criterion to the 2018 forecast initiated four days in advance, the user would feel free to use all of Lake Erie, and the observed frequency of encountering MC concentrations > 8 μg L^{-1} was 0 % (same as observation) (Figure 8e,f)

If a user chose to avoid the three highest risk categories, using the forecast initiated on August 7, 2019, the user would reduce their frequency of encountering MC concentrations > 8 μg L^{-1} from 15.4 % to 0 % (Figure 8a,b). Using the same criteria applied to the less accurate

This article is protected by copyright. All rights reserved.

Author Manuscript

forecast initiated four days earlier, the user would reduce their frequency of encountering MC concentrations > 8 µg L^{-1} from 14.7% to 0 % (Figure 8c,d). Applying the same criterion to the 2018 forecast initiated four days in advance, the user could visit all of Lake Erie, and the observed frequency of encountering MC concentrations > 8 µg L^{-1} was 0 % (same as observation) (Figure 8e,f).

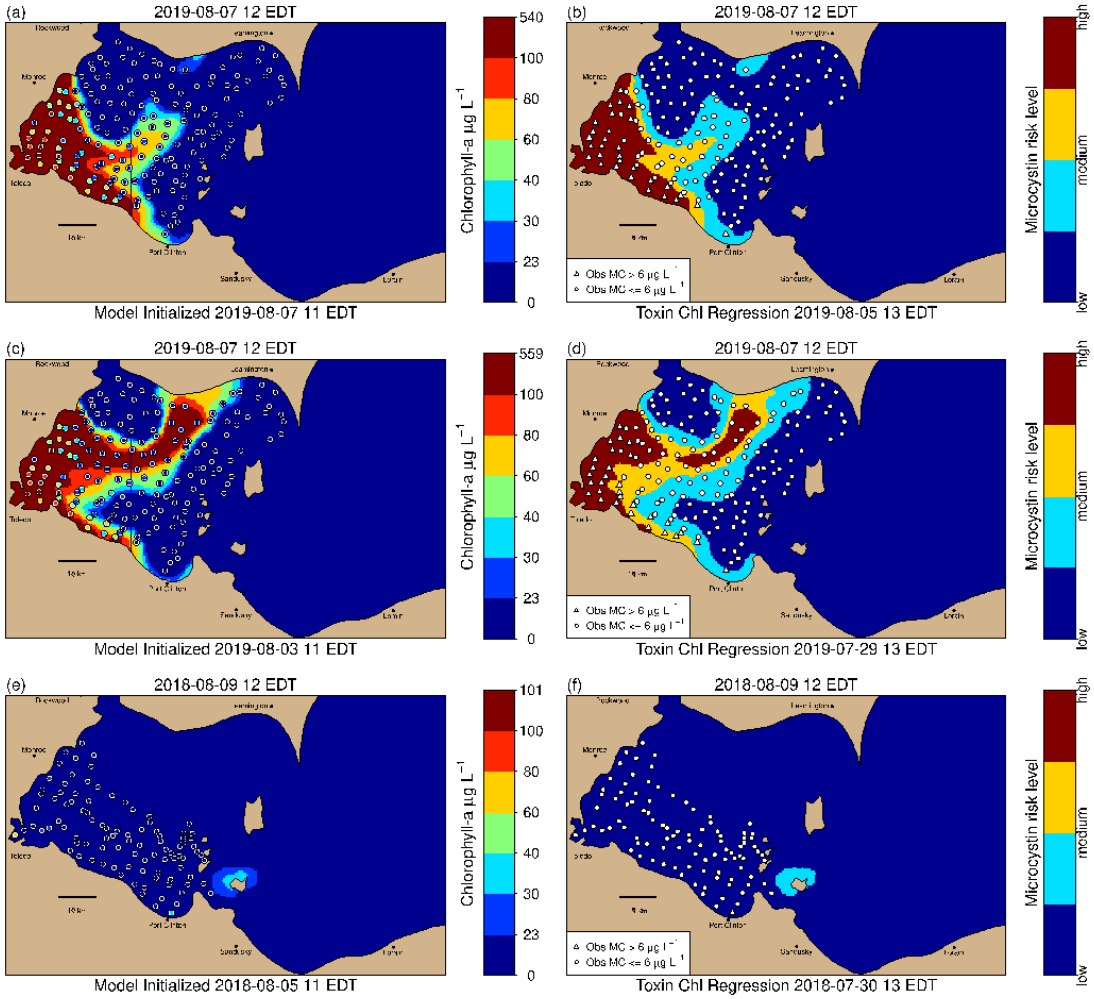

Figure 8. The model predicted chlorophyll concentrations and probability of MCs exceeding 6 µg L^{-1} for a-d) 2019 and e-f) 2018 intensive sampling dates in August. Colored dots in subplots a), c) and e) represent 2019 in-situ chlorophyll concentration observations; triangles in b), d) and f) represent stations with observed MCs exceeding 6 µg L^{-1}, and dots represent stations with observed MCs <= 6 µg L^{-1}. In a) and b), the HAB forecast was initialized from a satellite image on the same day and MC:chla regression from two days prior; in c) and d), the HAB forecast was initialized from a satellite four days prior and MC:chla regression from ten days prior; e) and f)

This article is protected by copyright. All rights reserved.

Author Manuscript

are for the 2018 intensive sampling date, and the HAB forecast was initialized from satellite image four days prior and MC:chla regression was from eleven days prior.

4. Discussion

Our novel approach combines remote sensing, water quality sampling, hydrodynamic forecasting, and a Lagrangian particle dispersion model, to predict the probability of MC exceeding a threshold value. In comparison to our previous approach (Liu et al, 2020), we expanded the hindcast period to cover 2014 to 2023, and improved the model skill and calibration by including bias adjustment of the satellite imagery over the hindcast period, replacing the linear regression between chlorophyll and MCs by a more robust Siegel regression model, and by applying a spline fit to provide a function to relate the probability of exceeding an MC threshold value to model-predicted MC concentration. We also conducted a more rigorous skill assessment of the forecast system using leave-out-one-year cross-validation over a 10-year hindcast period, and two intensive sampling dates in 2018 and 2019, highlighting our system's potential application to reduce a user's risk of exposure to MC toxins in Lake Erie. In addition, our assessment showed a useful level of skill for surface water, PWS intake sites, and bottom water where PWS intakes are usually located, in contrast to the forecast system of Liu et al. (2020), which was calibrated and assessed using surface observations only. Because the relationship between model-predicted MCs and the probability of exceeding the threshold MC concentration in grab samples (Figure 4) incorporates uncertainty in model-predicted chlorophyll-a concentrations, the natural variability in the MC:chla relationship, and the spatial scale mismatch between modeled or satellite-derived values (~km) and grab samples (< 1m), the maximum probability of exceedance the model can predict is ~0.6. This limitation reflects the reality that it is not currently possible to predict MC threshold exceedances with absolute certainty. The forecast cannot be recommended for ODH Beach sites due to bias revealed in the assessment. Although the number of observations and sites were limited for the beach data, the bias may result from a different sampling methodology in the beach program, versus open water sampling programs, or mechanisms of localized concentration of *Microcystis* colonies in downwind beach areas not represented in the model.

Our method relies on MCs and chlorophyll-a data from grab samples collected in Lake Erie, which we used to develop a weekly-updated regression of MC on chlorophyll-a. Past

This article is protected by copyright. All rights reserved.

Author Manuscript

research has shown that the MC:chla ratio decreases throughout the bloom season (Gobler et al. 2016), which is likely due to an interaction between nitrogen availability, oxidative stress, and a shift in the *Microcystis* gene pool from strains that can produce MCs to strains that cannot (Paerl and Otten 2013; Hellweger et al. 2022; Yancey et al. 2022). While the MC:chla appears to decrease throughout the bloom season every summer, the maximum ratio and rate of decrease varies from year to year (Gobler et al. 2016; Liu et al. 2020). For example, MC:chla was 0.31-0.4 in late July and early August 2014 whereas the ratio was less than 0.15 during the same time period in 2015 (Liu et al. 2020). While the MC:chla relationship varies over time, our skill assessment shows that skillful predictions can be made by updating the MC:chla relationship using weekly monitoring data. A limitation of our approach is that the forecast relies on monitoring locations that are representative of MC levels found in the bloom. For example, if a toxic bloom formed in an area not covered by the monitoring stations, the forecast would likely have little skill. An additional limitation is that MC levels can increase rapidly when a toxic bloom initiates (Liu et al., 2020), and weekly updating of the MC:chla relationship may not be sufficiently frequent to capture these cases. However, since our skill assessment covered the full HAB season, a ten-year period, and the spatial area in which HABs typically occur in Lake Erie, these limitations should be represented in probabilities of exceedance associated with the risk levels (Table 1). Greater spatial and temporal coverage of monitoring would likely improve the skill of our forecast and reduce the uncertainty of predictions.

Zhou et al. (2023b) assessed spatial-temporal forecasts of MCs in Lake Erie using a method similar to ours, but with key differences. Zhou et al. (2023b) initialized their model with spatial maps of MC concentrations interpolated from grab-sample observations from the same data sources we used, while we initialized our model with satellite-derived cyanobacterial chlorophyll-a and an MC:chla regression updated weekly based on observations. In addition, Zhou et al. (2023b) used an Eulerian tracer model (Zhou et al., 2023a) and incorporated measured rates of MC production (Chaffin et al., 2022), and the incorporation of MC production rates improved model accuracy by 10% compared to models without MC production rates, especially at higher MC concentrations. While some of these features could be assessed in future versions of our method, we chose to assess a forecast based on input data that are already being routinely produced by NOAA, specifically, the satellite-derived cyanobacterial index imagery, a Lagrangian-based HAB forecast, and weekly paired samples of chlorophyll-a and MCs.

This article is protected by copyright. All rights reserved.

Author Manuscript

Information on the rate of production of MC (Zhou et al., 2023b), or information on genes associated with MC production (Dick et al., 2021) could be considered for incorporation into future versions of a forecast.

The strategy of using a "dual" model strategy–satellite for chlorophyll and field data for conversion of chlorophyll to MC–allows the method to be adaptable to other areas. The satellite-based cyanobacteria determination is used in the Cyanobacterial Assessment Network (CyAN), a U.S. national monitoring program for nearly 2000 large lakes (Schaeffer et al., 2022). Accordingly, our approach has the potential to be applied to other systems suffering from HABs, where a strong correlation between chlorophyll and MC concentrations exists at toxin levels relevant to advisory levels. For example, Izydorczyk et al. (2009) found a positive correlation between the chlorophyll-a concentration from cyanobacteria and the concentration of intracellular MCs in the drinking water intake point at the Bronislawow Bay in Sulejow Reservoir (Poland), which makes the measurement of chlorophyll-a concentration an effective early warning system for cyanobacteria in that particular region. Sakai et al. (2013) found that the occurrence and distribution of MCs in Lake Taihu, the third largest lake in China, is highly correlated with chlorophyll-a concentration. Hayes and Vanni (2018) found the studied lakes and reservoirs in Ohio, which have small watershed areas relative to lake surface areas, had elevated MC concentrations when phytoplankton biomass was high. While further research into the biological mechanisms of HAB species composition and MC production may improve MC prediction, our approach provides a framework for MC forecasting that may be applicable across systems with sufficient observational data.

Authorship contribution statement

Qianqian Liu: Writing – original draft, Writing – review & editing, Methodology, Conceptualization, Investigation, Formal analysis, Validation, Visualization. **Mark D. Rowe:** Methodology, Conceptualization, Data curation, Project administration, Resources, Supervision, Writing – review & editing. **Richard P. Stumpf:** Methodology, Writing – review & editing. **Reagan Errera:** Data curation, Writing – review & editing. **Casey Godwin:** Data curation,

This article is protected by copyright. All rights reserved.

Author Manuscript

Writing – review & editing. **Justin D. Chaffin:** Data curation, Writing – review & editing. **Eric J. Anderson:** Conceptualization, Writing – review & editing. **Tongyao Pu:** Validation, Visualization, Writing – review & editing.

Acknowledgments

This work was supported by funding from the Great Lakes Restoration Initiative, through US EPA. Funding was awarded to the University of North Carolina at Wilmington through NOAA (1305M320PNRMA0357). The HABs Grab was funded by National Oceanic and Atmospheric Administration's National Centers for Coastal Ocean Science under award NA17NOS4780186 to the Ohio State University. Funding was awarded to CIGLR through the NOAA Cooperative Agreement NA17OAR4320152 with the University of Michigan. This is CIGLR publication #_____ and GLERL publication #_____.

Data availability statement

Software used: R 4.2.2

Data availability: the scripts for toxin forecast and model assessment, and example data are published on Zenodo (Great Lakes Environmental Research Laboratory, 2024; https://doi.org/10.5281/zenodo.13685470).

Reference:

Boegehold, Anna G., Ashley M. Burtner, Andrew C. Camilleri, Glenn Carter, Paul DenUyl, David Fanslow, Deanna Fyffe Semenyuk et al. 2023. "Routine monitoring of western Lake Erie to track water quality changes associated with cyanobacterial harmful algal blooms." *Earth System Science Data Discussions* 2023: 1-39.

This article is protected by copyright. All rights reserved.

Author Manuscript

Bte Sukarji, N.H.; He, Y.; Te, S.H.; Gin, K.Y.-H. 2022. "Application of a Mechanistic Model for the Prediction of Microcystin Production by Microcystis in Lab Cultures and Tropical Lake." *Toxins, 14*, 103.

Chaffin, Justin D. et al. 2021. "The Lake Erie HABs Grab: A Binational Collaboration to Characterize the Western Basin Cyanobacterial Harmful Algal Blooms at an Unprecedented High-Resolution Spatial Scale." *Harmful Algae* 108: 102080.

Chaffin, Justin D. et al. 2022. "Quantification of Microcystin Production and Biodegradation Rates in the Western Basin of Lake Erie." *Limnology and Oceanography*.

Davis, Timothy W. et al. 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.

De Pinto, J.V.; Young, T.C.; McIlroy, L.M. 1986. "Great Lakes Water Quality Improvement." *Environmental science & technology* 20, 752–759.

Dick, Gregory J. et al. 2021. "The Genetic and Ecophysiological Diversity of *Microcystis*." *Environmental Microbiology* 23(12): 7278–7313.

Feng, L., Wang, Y., Hou, X., Qin, B., Kuster, T., Qu, F., Chen, N., Paerl, H.W. and Zheng, C., 2024. Harmful algal blooms in inland waters. *Nature Reviews Earth & Environment*, 5(9), pp.631-644.

Geisser, Seymour. 1975. "The Predictive Sample Reuse Method with Applications." *Journal of the American Statistical Association* 70(350): 320–28.

Gobler, Christopher J. et al. 2016. "The Dual Role of Nitrogen Supply in Controlling the Growth and Toxicity of Cyanobacterial Blooms." *Harmful Algae* 54: 87–97.

Great Lakes Environmental Research Laboratory. 2024. "Ten-year hindcast assessment of an improved probabilistic forecast system for cyanotoxin (microcystins) risk level in Lake Erie." *Zenodo*. https://doi.org/10.5281/zenodo.13685470.

Hartmann, Holly C., Thomas C. Pagano, S. Sorooshian, and R. Bales. 2002. "Confidence Builders: Evaluating Seasonal Climate Forecasts from User Perspectives." *Bulletin of the American Meteorological Society* 83(5): 683–98.

Hayes, Nicole M., and Michael J. Vanni. 2018. "Microcystin Concentrations Can Be Predicted with Phytoplankton Biomass and Watershed Morphology." *Inland Waters* 8(3): 273–83.

He, Xuexiang, Yen-Ling Liu, Amanda Conklin, Judy Westrick, Linda K. Weavers, Dionysios D. Dionysiou, John J. Lenhart, Paula J. Mouser, David Szlag, and Harold W. Walker. 2016. "Toxic cyanobacteria and drinking water: Impacts, detection, and treatment." *Harmful algae* 54: 174-193.

This article is protected by copyright. All rights reserved.

Author Manuscript

Hellweger, Ferdi L. et al. 2022. “Models Predict Planned Phosphorus Load Reduction Will Make Lake Erie More Toxic.” *Science* 376(6596): 1001–5.

Izydorczyk, Katarzyna et al. 2009. “Establishment of an Alert Level Framework for Cyanobacteria in Drinking Water Resources by Using the Algae Online Analyser for Monitoring Cyanobacterial Chlorophyll a.” *Water Research* 43(4): 989–96.

Jiang, P.; Liu, X.; Zhang, J.; Yuan, X. 2016. “A Framework Based on Hidden Markov Model with Adaptive Weighting for Microcystin Forecasting and Early-Warning.” *Decision Support Systems, 84*, 89–103.

Kelley, J.G.W., Chen, Y., Anderson, E.J., Lang, G.A. and Xu, J., 2018. “Upgrade of NOS Lake Erie Operational Forecast System (LEOFS) to FVCOM: Model development and hindcast skill assessment.”

Kutser, Tiit. 2009. “Passive Optical Remote Sensing of Cyanobacteria and Other Intense Phytoplankton Blooms in Coastal and Inland Waters.” *International Journal of Remote Sensing* 30(17): 4401–25.

Lad, Apurva, Joshua D. Breidenbach, Robin C. Su, Jordan Murray, Rebecca Kuang, Alison Mascarenhas, John Najjar et al. 2022. "As we drink and breathe: Adverse health effects of microcystins and other harmful algal bloom toxins in the liver, gut, lungs and beyond." *Life* 12, no. 3: 418.

Liu, Qianqian et al. 2020. “Probabilistic Forecast of Microcystin Toxin Using Satellite Remote Sensing, in Situ Observations and Numerical Modeling.” *Environmental Modelling & Software* 128: 104705.

May, N.W.; Olson, N.E.; Panas, M.; Axson, J.L.; Tirella, P.S.; Kirpes, R.M.; Craig, R.L.; Gunsch, M.J.; China, S.; Laskin, A.; et al. 2018. “Aerosol Emissions from Great Lakes Harmful Algal Blooms.” *Environmental science & technology 52*, 397–405.

Olson, Nicole E., Madeline E. Cooke, Jia H. Shi, Johnna A. Birbeck, Judy A. Westrick, and Andrew P. Ault. 2020. "Harmful algal bloom toxins in aerosol generated from inland lake water." *Environmental science & technology* 54, no. 8: 4769-4780.

Paerl, Hans W. et al. 2016. “Mitigating Cyanobacterial Harmful Algal Blooms in Aquatic Ecosystems Impacted by Climate Change and Anthropogenic Nutrients.” *Harmful Algae* 54: 213–22.

Qian, S.S.; Chaffin, J.D.; DuFour, M.R.; Sherman, J.J.; Golnick, P.C.; Collier, C.D.; Nummer, S.A.; Margida, M.G. 2015. “Quantifying and Reducing Uncertainty in Estimated Microcystin Concentrations from the ELISA Method”. *Environmental science & technology* 49, 14221–14229.

Qian, Song, S., Craig A. Stow, Freya E. Rowland, Qianqian Liu, Mark D. Rowe, Eric J. Anderson, Richard P. Stumpf, and Thomas H. Johengen. 2021. “Chlorophyll a as an

This article is protected by copyright. All rights reserved.

Author Manuscript

indicator of microcystin: Short-term forecasting and risk assessment in Lake Erie." *Ecological Indicators* 130: 108055.

R Core Team. 2021. "R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/."

Recknagel, F.; Orr, P.T.; Bartkow, M.; Swanepoel, A.; Cao, H. 2017. "Early Warning of Limit-Exceeding Concentrations of Cyanobacteria and Cyanotoxins in Drinking Water Reservoirs by Inferential Modelling." *Harmful Algae*, *69*, 18–27.

Rinta-Kanto, Johanna M., Elisabeth A. Konopko, et al. 2009a. "Lake Erie Microcystis: Relationship between Microcystin Production, Dynamics of Genotypes and Environmental Parameters in a Large Lake." *Harmful Algae* 8(5): 665–73.

Rinta-Kanto, Johanna M., Matthew A. Saxton, et al. 2009b. "The Diversity and Distribution of Toxigenic Microcystis Spp. in Present Day and Archived Pelagic and Sediment Samples from Lake Erie." *Harmful Algae* 8(3): 385–94.

Rowe, M. D.; Anderson, E. J.; Wynne, T. T.; Stumpf, R. P.; Fanslow, D. L.; Kijanka, K.; Vanderploeg, H. A.; Strickler, J. R. and Davis, T. W. 2016. "Vertical Distribution of Buoyant *Microcystis* Blooms in a Lagrangian Particle Tracking Model for Short-Term Forecasts in Lake Erie." *Journal of Geophysical Research: Oceans* 175(4449): 238.

Sakai H, Hao A, Iseri Y, Wang S, Kuba T, Zhang Z, Katayama H. Occurrence and distribution of microcystins in Lake Taihu, China. S*cientificWorldJournal*. 2013 Jun 16;2013:838176. doi: 10.1155/2013/838176. PMID: 23853542.

Schaeffer, B., Urquhart E., Coffer, M., Salls, W., Stumpf, R.P., Loftin, K., Werdell, P. 2022. "Satellites quantify the spatial extent of cyanobacterial blooms across the United States at multiple scales.". *Ecological Indicators*, 140:108990, doi: 10.1016/j.ecolind.2022.108990

Soontiens, N., Binding, C., Fortin, V., Mackay, M., & Rao, Y. R. 2019. "Algal bloom transport in Lake Erie using remote sensing and hydrodynamic modelling: sensitivity to buoyancy velocity and initial vertical distribution." *Journal of Great Lakes Research*, 45(3), 556-572.

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.; et al. 2014. "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.

Stone, M. 1974. "Cross-Validatory Choice and Assessment of Statistical Predictions." *Journal of the Royal Statistical Society* 36(2): 111–47.

Stumpf, Richard P., Laura T. Johnson, Timothy T. Wynne, and David B. Baker. 2016a. "Forecasting Annual Cyanobacterial Bloom Biomass to Inform Management Decisions in Lake Erie." *Journal of Great Lakes Research* 42(6): 1174–83.

This article is protected by copyright. All rights reserved.

Author Manuscript

Stumpf, Richard P., Timothy W. Davis, et al. 2016b. “Challenges for Mapping Cyanotoxin Patterns from Remote Sensing of Cyanobacteria.” *Harmful Algae* 54: 160–73.

Theil, H. 1992. “A Rank-Invariant Method of Linear and Polynomial Regression Analysis.” In *Henri Theil’s Contributions to Economics and Econometrics*; Raj, B., Koerts, J., Eds.; Advanced Studies in Theoretical and Applied Econometrics; Springer Netherlands: Dordrecht. Vol. 23, pp. 345–381 ISBN 978-94-010-5124-8.

U.S. EPA. 2019. “Recommendations for Cyanobacteria and Cyanotoxin Monitoring in Recreational Waters.”

Wells, M.L., Trainer, V.L., Smayda, T.J., Karlson, B.S., Trick, C.G., Kudela, R.M., Ishikawa, A., Bernard, S., Wulff, A., Anderson, D.M. and Cochlan, W.P., 2015. Harmful algal blooms and climate change: Learning from the past and present to forecast the future. *Harmful algae*, 49, pp.68-93.

WHO, 2020. “Cyanobacterial toxins: microcystins. Background document for development of WHO Guidelines for drinking-water quality and Guidelines for safe recreational water environments”. Geneva: World Health Organization; 2020 (WHO/HEP/ECH/WSH/2020.6). Licence: CC BY-NCSA 3.0 IGO.

Wynne, T. T., R. P. Stumpf, and T. O. Briggs. 2013. “Comparing MODIS and MERIS Spectral Shapes for Cyanobacterial Bloom Detection.” *International Journal of Remote Sensing* 34(19): 6668–78.

Wynne, T. T., and R. P. Stumpf. 2015. “Spatial and Temporal Patterns in the Seasonal Distribution of Toxic Cyanobacteria in Western Lake Erie from 2002–2014.” *Toxins* 7(5): 1649–63.

Wynne, T. T., R. P. Stumpf, K. L. Pokrzywinski, R. W. Litaker, B. T. De Stasio, and R. R. Hood. 2022. "Cyanobacterial Bloom Phenology in Green Bay Using MERIS Satellite Data and Comparisons with Western Lake Erie and Saginaw Bay." *Water* 14, no. 17: 2636.

Yancey, Colleen E. et al. 2022. “Metagenomic and Metatranscriptomic Insights into Population Diversity of *Microcystis* Blooms: Spatial and Temporal Dynamics of *mcy* Genotypes, Including a Partial Operon That Can Be Abundant and Expressed.” *Applied and Environmental Microbiology* 88(9): e02464-21.

Zahir, M., Su, Y., Shahzad, M.I., Ayub, G., Rehman, S.U. and Ijaz, J., 2024. A review on monitoring, forecasting, and early warning of harmful algal bloom. *Aquaculture*, p.741351.

Zhou, X., Rowe, M., Liu, Q., & Xue, P. 2023a. “Comparison of Eulerian and Lagrangian transport models for harmful algal bloom forecasts in Lake Erie.” *Environmental Modelling & Software*, 162, 105641.

This article is protected by copyright. All rights reserved.

Author Manuscript

Zhou, X.; Chaffin, J.D.; Bratton, J.F.; Verhamme, E.M.; Xue, P. 2023b. “Forecasting Microcystin Concentrations in Lake Erie Using an Eulerian Tracer Model.” *Journal of Great Lakes Research, 49,* 1029–1044.

Chan, Wai Sum, Friedrich Recknagel, Hongqing Cao, and Ho-Dong Park. 2007. “Elucidation and Short-Term Forecasting of Microcystin Concentrations in Lake Suwa (Japan) by Means of Artificial Neural Networks and Evolutionary Algorithms.” Water Research 41 (10): 2247–55. https://doi.org/10.1016/j.watres.2007.02.001.

Hogan, R.J., Mason, I.B., 2012. “Deterministic forecasts of binary events. In: Forecast Verification: A Practitioner’s Guide in Atmospheric Science”. second ed. John Wiley & Sons, Ltd., West Sussex, U. K, pp. 31–59.

Author Manuscript

This article is protected by copyright. All rights reserved.