Science of the Total Environment 856 (2023) 158959

Contents lists available at ScienceDirect

Science of the Total Environment

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

Advancing freshwater ecological forecasts: Harmful algal blooms in Lake Erie

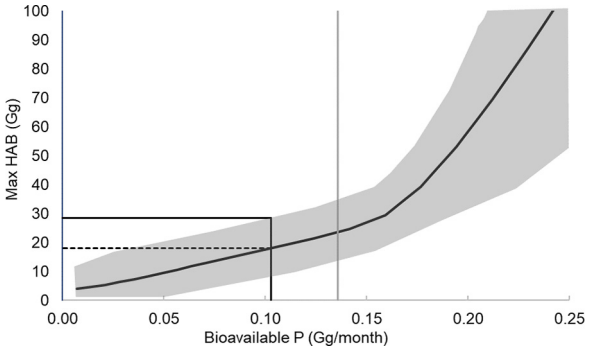
Donald Scavia [a,*], Yu-Chen Wang [a], Daniel R. Obenour [b]

[a] School for Environment and Sustainability, University of Michigan, Ann Arbor, MI 48103, USA
[b] Department of Civil, Construction & Environmental Engineering, NC State University, Raleigh, NC 27695, USA

HIGHLIGHTS

- Forecasting harmful algal blooms with quantified uncertainty is important.
- A new model is proposed that addresses earlier shortcomings.
- Phosphorus internal recycling is an important driver.
- Updated scenarios suggest current load reduction targets may be too stringent.
- Forecasting best practices are evaluated.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Lotfi Aleya

Keywords:
Bayesian
Lake Erie
Ecological forecasts
Harmful algal blooms

ABSTRACT

Ecological models help provide forecasts of ecosystem responses to natural and anthropogenic stresses. However, their ability to create reliable predictions requires forecasts with track records sufficiently long to build confidence, skill assessments, and treating uncertainty quantitatively. We use Lake Erie harmful algal blooms as a case study to help formalize ecological forecasting. Key challenges for models include uncertainty in the deterministic structure of the load-bloom relationship and the need to assess alternative drivers (e.g., biologically available phosphorus load, spring load, longer term cumulative load) with a larger dataset. We enhanced a Bayesian model considering new information and an expanded data set, test it through cross validation and blind forecasts, quantify and discuss its uncertainties, and apply it for assessing historical and future scenarios. Allowing a segmented relationship between bloom size and spring load indicates that loading above 0.15 Gg/month will have a substantially higher marginal impact on bloom size. The new model explains 84 % of interannual variability (9.09 Gg RMSE) when calibrated to the 19-year data set and 66 % of variability in cross validation (12.58 Gg RMSE). Blind forecasts explain 84 % of HAB variability between 2014 and 2020, which is substantially better than the actual forecast track record ($R^2 = 0.32$) over this same period. Because of internal phosphorus recycling, represented by the long-term cumulative load, it could take over a decade for HABs to fully respond to loading reductions, depending on the pace of those reductions. Thus, the desired speed and endpoint of the lake's recovery should be considered when updating and adaptively managing load reduction targets.
Results are discussed in the context of ecological forecasting best pactices: incorporate new knowledge and data in model construction; account for multiple sources of uncertainty; evaluate predictive skill through validation and hindcasting; and answer management questions related to both short-term forecasts and long-term scenarios.

* Corresponding author.
E-mail address: scavia@umich.edu (D. Scavia).

http://dx.doi.org/10.1016/j.scitotenv.2022.158959
Received 1 June 2022; Received in revised form 16 September 2022; Accepted 19 September 2022
Available online 22 September 2022

1. Introduction

Ecological models can address the need for accurate and reliable forecasts of ecosystem responses to natural and anthropogenic stresses (Coreau et al., 2009; Luo et al., 2011; Payne et al., 2017; Ross et al., 2020). These models strengthen linkages between resource management and research, and can expand knowledge of underlying system dynamics (Testa et al., 2017; Dietze et al., 2018). Applying forecasts to aid decision making, particularly within adaptive management (Walters and Holling, 1990; Westgate et al., 2013), has generated interest and support within academia and among government agencies in the United States (CENR, 2001). For example, the National Oceanic and Atmospheric Administration (NOAA) promotes forecasts for harmful algal blooms (HABs), hypoxia, fisheries, and pathogens (Valette-Silver and Scavia, 2003; NOAA, 2020). Other US agencies sponsor similar efforts (Bradford et al., 2020; NASA, 2020) and the Ecological Forecasting Initiative (EFI, 2020) was created to share experiences within the broader forecasting community.

However, a lack of consensus on the ability of these models to create reliable, policy-relevant predictions (Beckage et al., 2011; Schindler and Hilborn, 2015) is due partly to insufficient forecast track records and skill assessments (Johnson-Bice et al., 2020). In some cases, this is because predictive time frames are too long, as in climate-related forecasting (Dietze et al., 2018), or because protocols do not exist for integrating new observations (White et al., 2019). While quantifying uncertainty is becoming more common (Harwood and Stokes, 2003; Clark, 2005; Baker et al., 2014; Gimenez et al., 2014), comprehensive treatments of uncertainty are often lacking (Dietze et al., 2018), such that the true risk of an adverse outcome is unknown (Pappenberger and Beven, 2006; Raftery, 2016).

At the same time, it may be prudent to begin developing management-relevant forecasts even before models are perfected, as it is important to learn by doing (Dietze et al., 2018). We have been doing that for Gulf of Mexico hypoxia since 2002, Chesapeake Bay hypoxia since 2007, and Lake Erie HABs since 2009 (Scavia et al., 2021c). These annual public forecasts (e.g., Press Releases, 2021) and policy scenarios (e.g., Scavia et al., 2016, 2017a, 2017b, 2019, 2021a, 2021b; Scavia and Donnelly, 2007; Testa et al., 2017) have also enabled continuous model evaluations and improvements.

We recently used a Chesapeake Bay case study (Scavia et al., 2021b) to illustrate advances in formalizing hypoxia forecasts, one of the most mature examples of operational ecological forecasting (e.g., Scavia et al., 2003, 2006; Testa et al., 2017, VIMS, 2020, Bever et al., 2021, Katin et al., 2019; North Carolina Sea Grant, 2020). HAB forecasts are also becoming more common (Stumpf et al., 2012; Obenour et al., 2014; Verhamme et al., 2016; Harre and Campbell, 2014; Jewett et al., 2007; Davidson et al., 2013; Zhang et al., 2013; Wang and Boegman, 2021) and necessary (Treuer et al., 2021).

In this paper, we use a Lake Erie case study to add to the formalization of the HAB forecasting process. Lake Erie has a relatively detailed 19-year HAB data set, and we have been making seasonal forecasts for the lake since 2014. In addition, our model has been used along with several others in operational forecasts and scenario ensembles by government agencies (e.g., https://tinyurl.com/2psfg85d, Scavia et al., 2016).

Key challenges for existing HAB forecast models include uncertainty in the structure of the load-bloom relationship (e.g., linear, exponential) and the need to assess new alternative drivers (e.g., biologically available loads, spring loads, longer term cumulative loads) as both datasets and scientific knowledge advance. Herein, we develop a new Bayesian HAB forecast model building on earlier efforts (Obenour et al., 2014; Bertani et al., 2016; Scavia et al., 2021a; Ho and Michalak, 2017), new information from recent studies, and an expanded data set. We test it through leave-one-out cross validation and simulated blind forecasting, and we apply it to potential management scenarios. We then discuss the results in the context of best practices, such as incorporation of new knowledge, model testing and uncertainty quantification, and informing policy and adaptive management.

2. Methods

2.1. Study site and management context

Lake Erie is the warmest and most eutrophic of the Laurentian Great Lakes, the largest freshwater system in the world. It has been affected by HABs, especially in its western basin (Bullerjahn et al., 2016; Michalak et al., 2013; Jankowiak et al., 2019; Watson et al., 2016; Bridgeman et al., 2013), which is large (~6500 km^2) and shallow (average depth ~7 m). It has a water residence time of 20–40 days and is subject to intermittent stratification and wind-driven mixing (Wynne et al., 2011). It provides a wide range of ecosystem services important to the economies of Ohio, Michigan, and southern Ontario (Allan et al., 2017). Between its two major tributaries, the Detroit River's average discharge (5324 m^3/s) is much higher than the Maumee River's (150 m^3/s), but they deliver comparable total phosphorus (TP) loads (41 % and 48 % of the total annual load to the western basin, respectively). The lower concentration and higher flow from the Detroit River tends to dilute and deflect the HAB, whereas the Maumee River's much higher TP concentration makes it a strong driver of bloom development (Michalak et al., 2013; Scavia et al., 2016).

In response to increased HABs, hypoxia, and nearshore algae growth (Scavia et al., 2014; Watson et al., 2016), and guided by an ensemble of lake models (Scavia et al., 2016), Canada and the United States revised the Great Lakes Water Quality Agreement (GLWQA) to include a 40 % reduction in Lake Erie's annual TP load from 2008 levels (GLWQA, 2016). For HABs, the Maumee River has an additional target of reducing the March–July TP and dissolved reactive phosphorus (DRP) loads by 40 % to reduce HAB extent to the mild 2012 extent in 9 out of 10 years.

2.2. HAB model

Obenour et al. (2014) developed a model to predict annual maximum HAB biomass as a function of TP load from the Maumee River. Within a Bayesian framework, this model provided estimates of the critical loading period, a loading threshold below which blooms are de minimis, and a time term representing how this threshold has declined over time. It used a novel, hierarchical approach for integrating two independent sets of bloom estimates, allowing a comparison of measurement, prediction, and model parameter uncertainties. Bertani et al. (2016) expanded the calibration period and improved the model by replacing TP load with an estimate of bioavailable P (BioP) defined as DRP plus a fraction of non-DRP P (TP-DRP) (referred to hereafter as “particulate P” for convenience) with that fraction estimated within the Bayesian framework. Stumpf et al. (2012) developed an exponential regression model driven by spring TP, and then updated it using their estimate of spring BioP (Stumpf et al., 2016). Ho and Michalak (2017) developed a linear model that used the current spring DRP load plus the cumulative DRP load from the previous nine years, a representation of internal P storage and recycling capacity. The cumulative loading term provided an alternative to the linear time term of Obenour et al. (2014) and Bertani et al. (2016). Scavia et al. (2021a) added a third independent set of HAB estimates, expanded the calibration through 2017, and used the Bertani et al. (2016), Obenour et al. (2014), Stumpf et al. (2016), and Ho and Michalak (2017) models to explore the relative impacts of historical, current, and future loads on HAB projections and associated uncertainties. A common feature among these four models was underestimating the relatively large blooms of 2011, 2013, and 2015 (Fig. 3 in Scavia et al., 2021a).

Here, we propose a new segmented linear model with slopes representing the response to spring loads before and after a model-determined change point. This formulation addresses the potential nonlinear relationship between load and bloom size, considering the various linear, threshold, and exponential relationships that have been used previously (Stumpf et al., 2012; Obenour et al., 2014; Ho and Michalak, 2017). We also adopt the cumulative load feature of Ho and Michalak (2017) because preliminary analysis indicates it is a strong predictor for our new model and updated data set, and because it has a more mechanistic basis. We used Bayesian

estimation to identify (with quantified uncertainty) the time period for accumulating the BioP load along with other model parameters. We explored other potential drivers, such as temperature, wind, and blooms timing, but found no stronger predictors (Supporting Information).

The new model (before and after the change point, W_{cp}) is:

$$Z = \beta_0 + \beta_w W + \beta_c C + e + m \qquad \text{for } W < W_{cp}$$

$$Z = \beta_0 + \beta_w W + \beta_{w2}(W - W_{cp}) + \beta_c C + e + m \qquad \text{for } W > W_{cp}$$

where (Z, Gg) is the western basin summer HAB biomass, and W (Gg/month) is the weighted average sum of monthly BioP loads (Obenour et al., 2014), hereafter referred to as the spring BioP load for brevity. The estimated weight parameter (t_w) is constrained to the range [1–6] for models that consider loads from January through June and to the range [1–7] for models that use loads from January through July (Supporting Information Eq. S1). For example, for models that consider loads through June, if $t_w = 4.6$, then loads for January to March all receive weights of zero, April (i.e., month 4) receives a weight of 0.4, and May and June both receive a weight of one.

Parameter β_w is the slope (bloom mass per spring BioP load, Gg-HAB/Gg-load/month) before the change point, and ($\beta_w + \beta_{w2}$) is the total slope after the change point. C is the monthly average cumulative BioP load, with the period of accumulation estimated as the previous t_c years (Supporting Information Eq. S2), and β_c is the slope with respect to the cumulative load. BioP loads are defined as DRP + θ × (TP - DRP) with θ estimated with the other model parameters and reflecting both the bioavailability and the potential for load changes downstream of the monitoring station at Waterville (Baker et al. 2014). Finally, *e* is model prediction error (i.e., a year-specific random effect) and *m* is measurement error across the three HAB extent estimates.

2.3. Phosphorus loads

Daily TP and DRP loads for 1974–2020 were obtained from Heidelberg University's National Center for Water Quality Research (L. Johnson, personal communication, https://ncwqr.org/) for the Maumee River at the Waterville station, which is 42 km upstream from the lake, as described in Stumpf et al. (2016). Missing data were estimated with linear interpolation, averaging most recent data, or with a flow/concentration regression with the most recent data, depending on how many data points were missing. Daily loads were aggregated to monthly values and used for model calibration, forecasts, hindcasts, and pseudo-blind forecasts.

2.4. HAB extent

HAB extent estimates for 2002–2020 are from three independent data sets: two satellite-derived (Stumpf et al., 2016; Manning et al., 2019) and a geostatistical synthesis of in situ observations (Fang et al., 2019). Because the Stumpf estimates are lake-wide, whereas the other two estimates cover only the western basin, Stumpf estimates were clipped to the western basin as described by Fang et al. (2019). Satellite-based estimates are based on cyanobacteria-specific spectra (Stumpf et al., 2016) and chlorophyll *a* from MODIS imagery (Sayers et al., 2016; Manning et al., 2019) which is mostly cyanobacteria for chlorophyll concentrations above the 18 μg/l threshold that we used (Sayers et al., 2016). Both are constrained to near-surface observations. The in situ chlorophyll observations above 18 μg/l (Fang et al., 2019) are sparser but allow for full water-column estimates. Because of the differences among these approaches and because we are most interested in inter-annual variability, we scaled the Manning and Stumpf estimates to the Fang estimates using the ratio of their averages, as described in Fang et al. (2019). HABs are thus reported in units of biomass (Gg), estimated from 3 consecutive maximum 10-day composites (Stumpf), the maximum 21-day rolling average (Manning), or the 30-day rolling average (Fang).

2.5. Model calibration and validation

The model was calibrated within a Bayesian framework, which readily provides data-driven inference and uncertainty quantification for all parameters in a non-linear model (Lunn et al., 2013). The framework also allows for non-standard error distributions, such as the gamma distribution (Obenour et al., 2014), and for partitioning variance across different levels (Qian et al., 2010) to help differentiate between observation and prediction error. We find this framework to be effective for seasonal forecasting and testing hypotheses about key HAB drivers while rigorously accounting for uncertainty. At the same time, we acknowledge that shorter-term forecasting may benefit from other approaches, such as machine learning algorithms that take advantage of high-frequency data (e.g., Shamshirband et al., 2019). The Bayesian posterior parameter distribution is developed through conventional Markov Chain Monte Carlo (MCMC) sampling in WinBUGS (Lunn et al., 2000) interfaced with the R package, R2WinBUGS (R Core Team, 2015; Sturtz et al., 2005). Three parallel MCMC chains were performed with 60,000 samples and the first 10,000 samples were discarded as a burn-in period. This resulted in R-hat statistics smaller than 1.1, which indicates that model convergence was achieved (Brooks and Gelman, 1998).

To test model robustness, we performed a leave-one-year-out cross validation (CV) where the observations for each year are predicted after removing those observations from the calibration data set and recalibrating the model to the reduced data set. The model's CV performance is a measure of how well it will perform when predicting "out-of-sample" (i.e., future) conditions (Elsner and Schmertmann, 1994; Chatfield, 2006). We also conducted pseudo-blind forecasts to test the model when calibrated only to the observations from previous years. For each year, we calibrated the model with observations from 2002 to the year preceding the forecast year. This was repeated for 2009 through 2019, years with three sets of HAB estimates.

2.6. Scenarios

We developed response curves with and without interannual variability in discharge ("discharge uncertainty") to predict HAB extent under different spring loads and associated cumulative BioP. The advantage of these scenarios is that they directly address key management concerns with uncertainty in the context of what can be controlled (phosphorus concentration) vs what cannot be controlled (hydrologic variability). In the Monte Carlo analysis accounting for discharge uncertainty, each of 1000 values were drawn with replacement from the 1975–2020 historical monthly-weighted spring discharges (calculated as above for spring loads). Otherwise, the mean discharge of the 1000 samples was used. For the response curves, spring loads were calculated as discharge times a range of spring BioP concentrations bracketing the 2008 flow-weighted mean concentration. For the Monte Carlo analysis, 5000 samples were also drawn from the model's posterior distribution for the cases with and without interannual discharge uncertainty. For these hypothetical response curves (e.g., projecting future HABs after spring loads were reduced), cumulative BioP is the value after it equilibrates with the reduced spring load, assuming annual TP and DRP loads were reduced by the same percent. The cumulative BioP for the response curve representing 2008 and 2015–2020 average conditions are based on observed TP and DRP loads and model parameters, and the HAB predictions were based on the posterior parameter distribution.

3. Results and discussion

3.1. HAB model calibration

The calibrated model (Table 1, Fig. 1) performed well, explaining 84 % of the interannual variability ($R^2 = 0.84$, RMSE = 9.09 Gg, percent bias (Pb) = 7.0) with a leave-one-year-out CV R^2 of 0.66 (RMSE = 12.58 Gg). This is similar to the Ho and Michalak (2017) model ($R^2 = 0.84$), and a

Table 1
Model parameter estimates with 95 % credible intervals from Bayesian posterior distribution.

Values	Mean	High	Low	Units
β0	−32.99	2.79	−63.80	Gg
βw	155.44	282.40	16.25	Gg/Gg/month
βw2	988.12	3266.00	38.28	Gg/Gg/month
βc	570.28	947.20	153.20	Gg/Gg/month
tw	3.53	4.59	2.49	Months
tc	9.07	16.02	4.58	Years
Wcp	0.15	0.24	0.03	Gg/month
σe	7.94	16.19	0.97	Gg
σm	10.52	14.82	7.65	Gg
θ	0.29	0.53	0.20	

significant improvement over the Bertani et al. (2016), Obenour et al. (2014), and Stumpf et al. (2016) models that explained 78 %, 76 %, and 63 % of variability, respectively, when each model was recalibrated to a common 2002–2017 HAB data set (Scavia et al., 2021a). It was also a significant improvement over Ho and Michalak's (2017) original implementation calibrated with satellite-based HAB area estimates over a similar time frame (2001–2015, R^2 = 0.78).

The estimated fraction of particulate P that is bioavailable (θ = 0.29, 95 % CI: 0.20–0.53, Table 1) is similar to that summarized from the literature (0.20–0.42) by Bertani et al. (2016). Standard deviations for prediction error (σ_e) and measurement error (σ_m) were of similar magnitude (Table 1). The model identified the relevant spring loading period as including 47 % of the March load plus April through July loads (t_w = 3.53). The cumulative BioP load is expected to include 7 % of the 10th prior year load plus all subsequent annual loads through December of the year before the forecast (t_c = 9.07), similar to the 9-year period used by Ho and Michalak (2017). However, there is considerable uncertainty around this parameter estimate (Table 1), such that it could take anywhere from 5 to 16 years for the cumulative load to come into equilibrium with a sudden reduction in nutrient loading.

The model also indicates a spring BioP load change point (W_{cp}) of around 0.15 Gg/month. While the mean 2015–2020 spring BioP load (0.13 Gg/month) was within the 95 % credible interval (0.03–0.24), the GLWQA target load, represented as BioP (0.061 Gg/month), was considerably lower. Based on posterior distributions (Table 1), the rate of growth in bloom size per unit load above this change point is roughly seven times higher than below the change point. The higher slope improves the model's ability to capture higher bloom extents, which is important when forecasting annual bloom extents. Also, the slope for the cumulative loading effect (β_c = 570 Gg/Gg/month) suggests cumulative load is relatively influential when the spring load is below the change point (i.e., β_w is only 155 Gg/Gg/month), but plays a smaller relative role when the spring load is above the change point ($\beta_w + \beta_{w2}$ = 1144 Gg/Gg/month). Thus, if seasonal loads are low, then the internal loading surrogate (cumulative load) will account for a larger portion of the boom magnitude, as suggested by Matisoff et al. (2016).

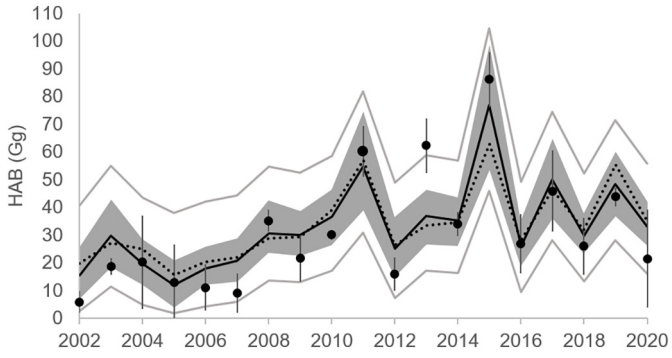


Fig. 1. Calibrated HAB model. Mean and standard deviation of observations (boxes with error bars), prediction (solid line) with 95 % credible intervals (reflecting parameter uncertainties, shaded grey) and 95 % prediction intervals (reflecting parameter and predictive residual uncertainties, grey line). Dotted line is prediction for the model calibrated with loads through June.

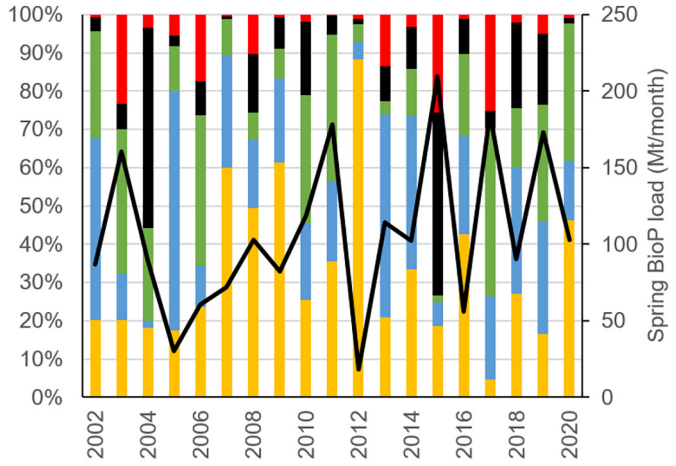


Fig. 2. Distribution of the March–July bioavailable P load by month (March = yellow; April = blue, May = green; June = black; July = red). Spring BioP load (black line).

The new model tracks the observations well, especially when considering HAB measurement uncertainty, and its 95 % predictive intervals capture all but the 2013 bloom (Fig. 1). Note that Fig. 1 also provides 95 % credible intervals (Lunn et al., 2013), which reflect the uncertainty in the mean bloom size (if averaged across multiple years). When analyzing errors, it is worth noting that the 2013 BioP monthly loading distribution was different from most other years (Fig. 2). For example, the May load was unusually small, while June and July loads were substantial.

Based on the calibrated model parameters, characterizing relevant loading periods and bioavailability, temporal patterns in critical loading inputs can be assessed. The spring BioP load increased since 1985, with several peaks after 2010, and interannual variability driven largely by discharge (Fig. 3). The cumulative BioP load decreased between 1985 and 1997, likely driven by the tail end of major point source controls (Maccoux et al., 2016). It then increased between 2000 and 2012 due to agricultural practices (Smith et al., 2015), climate (Stow et al., 2015), or both (Jarvie et al., 2017; Daloğlu et al., 2012), and stabilized after 2012.

3.2. Pseudo-blind forecasts

To further assess model performance, we generated blind forecasts for years with three sets of HAB estimates (2009–2019). These forecasts explained 54 % of HAB variability and all but one of the observations (2013) fell within the 95 % prediction intervals (Fig. 4). Actual annual forecasts, made with observations and models that evolved over time,

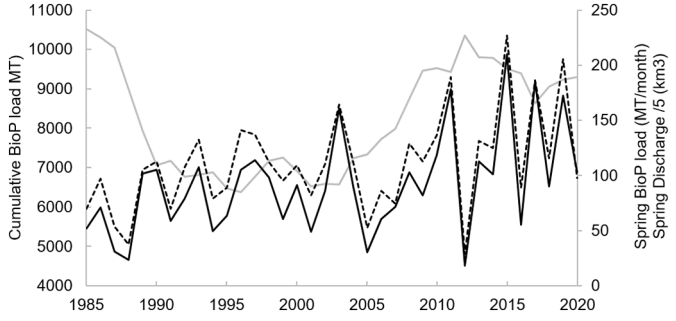


Fig. 3. Weighted spring bioavailable P load (Gg/month, black line), cumulative BioP load (Gg, grey line), and weighted spring discharge (km^3, dashed black line) from 1985 through 2020.

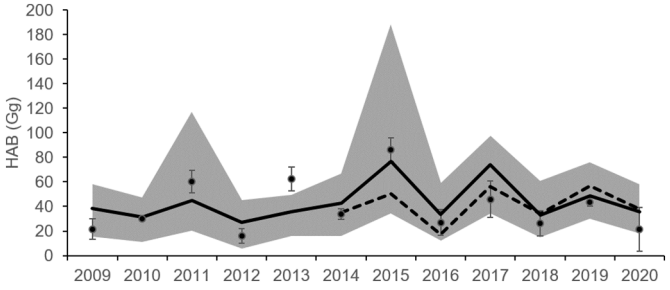


Fig. 4. Pseudo-blind forecasts for 2009–2020 using the new model. Observed mean and standard deviation (dots and error bars), pseudo-blind forecast (black line) with 95 % predictive interval accounting for parameter and prediction error. Actual annual forecasts (dashed line).

explained 32 % of the 2014–2020 variability (Scavia et al., 2021a, 2021b, 2021c). Blind forecasts from the new model explained 84 % of the variability for those same years.

4. Applications

A key ecological forecasting challenge is making accurate forecasts with quantified uncertainty that matter to decision makers. Earlier versions of this model provided both annual forecasts (Scavia et al., 2021a, 2021b, 2021c; NOAA, 2021) and scenarios to guide longer-term policies (Bertani et al., 2016; Scavia et al., 2016, 2021a). Here we provide an analysis related to applying the model in these contexts.

4.1. Annual forecasts

In response to stakeholder needs, NOAA and its partners produce ensemble annual forecasts based on loads only through June (NOAA, 2021). As expected, the model with loads through July performed better (R^2 = 0.84, RMSE = 9.09, Pb = 7.0) than one including loads through June (R^2 = 0.74, RMSE = 11.63, Pb = 8.6, Fig. 1). Models calibrated through the first and second weeks of July (Table S1) resulted in intermediate R^2 values of 0.78 and 0.83 (RMSE = 10.76 and 9.42), respectively. To take advantage of the improved performance, but keep the annual forecasts around the end of June, we developed regression models to predict TP and DRP (and thus BioP) loads based on precipitation during the last half of June and the first 7 days in July (Figs. S1, S2). This approach could potentially leverage 7-day precipitation forecasts from NOAA's Weather Prediction Center (2021). Adding the predicted total July loads to measured loads through June resulted in an R^2 of 0.83 (RMSE = 9.1), comparable to using the measured July loads.

4.2. Hindcasts

Hindcasts based on measured historical loads produced relatively low HABs prior to the 2000s (Fig. S3), consistent with satellite-based observations showing that HABs were not likely a nuisance in the 1980s and early 1990s (Ho and Michalak, 2017). In fact, hindcast HABs decline by about 33 % (from 30 to 20 Gg, Fig. S3) from the mid-1980s to the mid-1990s. Thus, the decreasing cumulative load between 1985 and 1995 appears to have more than compensated for the increases in spring BioP load over this same time period (Fig. 3). In this way, the cumulative load is acting as a capacitor to changes in HAB severity.

4.3. Scenarios

The GLWQA called for a 40 % reduction in March–July Maumee River TP and DRP loads (relative to a 2008 baseline) to keep HABs below the mild extent measured in 2012 in nine out of ten years (GLWQA, 2016). Putting that HAB goal and loading targets in the context of our model calculations, our average 2012 HAB estimate is 16.0 Gg, compared to the 2015–2020 average of 41.7 Gg. Using the model's estimates of spring BioP load, a 40 % reduction from the 2008 spring load (0.103 Gg/month) would be a load of 0.062 Gg/month, down 55 % from the 2015–2020 average spring load of 0.134 Gg/month.

Ignoring changes in cumulative load can result in different HAB predictions. To illustrate this, we constructed two HAB response curves based on different assumptions regarding the cumulative load. One curve (Fig. 5a) assumes that the cumulative BioP load is the current 2015–2020 average of 0.088 Gg/month, while the other (Fig. 5b) assumes the cumulative load has equilibrated with the 2008 spring load (0.073 Gg/month). For a spring load equal to the 2008 load (0.103 Gg/month), the first predicts an average HAB response of 32.5 Gg and below 44.0 Gg in 9 out of 10 years, while the second predicts an average HAB extent of 25.0 Gg and below 36.5 Gg in 9 out of 10 years. The second curve would be applicable about nine years (t_c = 9.07) after the loads are reduced to 2008 levels (assuming the reduction is immediate). A more realistic assumption of achieving the load reduction in increments of 10 % per year stretches the delay to 18 years (15–22 years considering the uncertainty in parameter t_c).

Of course, nutrient loading also varies from year to year based on hydro-climatological variability. To explore this issue in more detail, we developed response curves with and without consideration of discharge uncertainty (Fig. 6). These curves assume that sufficient time has passed for the cumulative BioP load to come into equilibrium with the reduced seasonal loads. The 90th percentile predictions including discharge variability provide a more realistic assessment of the loading reductions required to achieve the GLWQA HAB reduction target in 9 out of 10 years. From these curves, a spring BioP load of 0.063 Gg/month would meet the 16.0 Gg HAB goal 90 % of the time. This 39 % reduction from the 2008

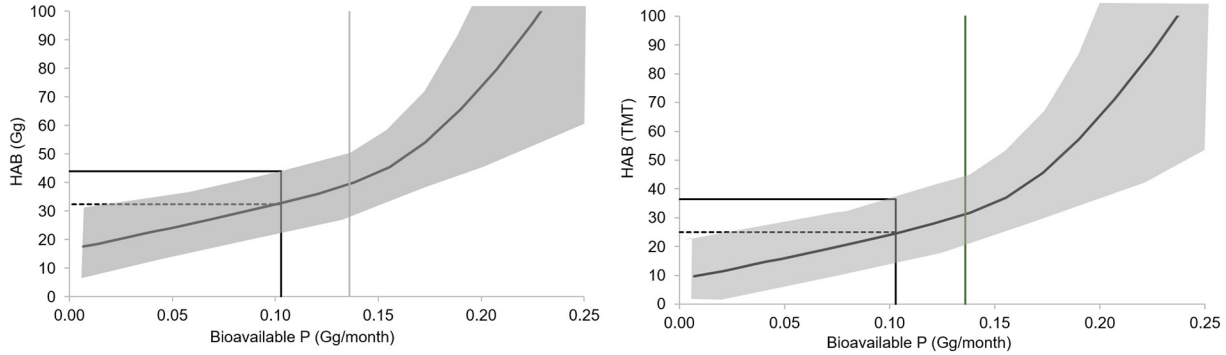


Fig. 5. Mean response (black line) and 80 % prediction intervals (grey shade) for HAB extent as a function of spring BioP load assuming: (a, left) the cumulative BioP load is the same as the 2015–2020 mean and (b, right) the cumulative load is at 2008 levels. Vertical grey lines represent current loads. Also shown are predictions for the mean HAB extent (dashed) and for HAB extent in 9 out of 10 years (solid) under the 2008 load.

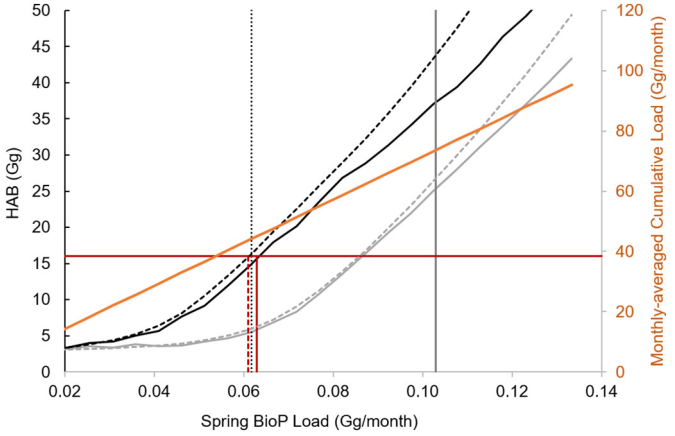


Fig. 6. HAB extent response to spring BioP load with (dashed lines) and without (solid lines) discharge variability. Results include 90th percentile HAB (black) and mean HAB extent (grey) as a function of spring BioP load (Gg/month), assuming that the cumulative BioP load is reduced proportionally (orange). Vertical lines: 2008 load (grey) and 40 % reduction from the 2008 load (dotted black). GLWQA target of below 16.0 Gg in 9 out of 10 years (horizontal red line). Spring BioP load required to meet the GLWQA goal in 9 out of 10 years with (dashed vertical red line) and without (solid vertical red line) discharge variability.

spring load would also result in an average HAB size of 5.8 Gg. Those values are a 41 % reduction to 0.061 Gg/month when hydro-climatological variability is considered. For higher loads (e.g., above 0.15 Gg/month), a growing proportion of the discharge distribution is above the change point, skewing the HAB distribution to higher values and increasing the uncertainty. However, this has little impact on these scenarios that focus on lower loads.

5. The long-term loading hypothesis

Including the cumulative load as a representation of internal recycling is critical. Based on extensive field measurements in Lake Erie's western basin, Matisoff et al. (2016) suggested that, while internal recycling is unlikely to trigger HABs by itself, it is sufficiently large to cause blooms when combined with external loads. They also suggest it may be responsible for delayed response to decreased external load. Gibbons and Bridgeman (2020), using similar methods, showed that when anoxic conditions are accompanied by elevated water temperature, western basin P loads from lake sediments are comparable to external loads from the Maumee River. The cumulative, or legacy, load formulation proposed by Ho and Michalak (2017) and adapted here assumes that loads from recent years drive algal production and sedimentation that, in turn, fuels sediment metabolism (e.g., sediment oxygen demand, SOD, and nutrient recycling). This connection between external load and internal nutrient recycling has been demonstrated, for example, in a chain of four Wisconsin lakes (Robertson and Diebel, 2020), a set of 16 Danish lakes (Jensen et al., 2006), and long-term recovery dynamics in Shagawa Lake, Minnesota (Chapra and Canale, 1991). These studies indicate that it takes a decade or longer for internal loading to fully respond to changes in external loading. For Lake Erie, Burns (1976) reported that 93 % of its external P load is retained and Zhang et al. (2016) simulated a net accumulation of P in its sediments. Further, Ho and Michalak's 9-year accumulation and our Bayesian estimate of 9.07 years (see Table 1) suggest multi-decadal recovery times that are consistent with other estimates for Lake Erie (Watson et al., 2016) and elsewhere (e.g., Jeppesen et al., 2007; Phillips et al., 2005). Thus, the nine-year averaging period for determining legacy loads is generally consistent with previous literature on lake response times. At the same time, as pointed out by an external reviewer, Lake Erie is much larger and subject to unique hydrodynamics, such that additional research will be important to further test the long-term loading hypothesis.

6. Great lakes water quality agreement policy implications

The binational Great Lakes Water Quality Agreement set a goal of reducing the summer maximum HAB extent to be less than the mild bloom of 2012 in 9 out of 10 years (GLWQA, 2016). Based on a suite of existing models calibrated to 2002–2015 observations (Scavia et al., 2016) and public comment, the agreement recommended a 40 % reduction from the 2008 loads. Our results (Fig. 6) are consistent with that target, and suggest that within that loading range hydro-climatological variability only requires slightly larger reductions. The model also predicts that meeting the goal in 9 out of 10 years results in an average HAB of 5.8 Gg, substantially below the 16 Gg goal. Given the high fiscal and social costs (e.g., Palm-Forster et al., 2016; Liu et al., 2020a, 2020b) of the management actions required to meet the goal (Martin et al., 2021; Scavia et al., 2017a, 2017b), surpassing the target by that much on average might suggest considering a slightly less stringent loading target.

At the same time, all of the recent studies (including this one) indicate the need to reduce loadings far below the 2015–2020 average loads. For example, meeting the target suggested in this study still requires a 55 % reduction from the 2015–2020 average load. Experiences in the Lake Erie watershed (Martin et al., 2021) and elsewhere (e.g., Gulf of Mexico: Stackpoole et al., 2021; Chesapeake Bay: Sabo et al., 2022) have shown achieving measurable load reductions can take years to decades after watershed actions are taken. This underscores the value of the GLWQA adaptive management plan where actions are taken, loads and HABs monitored, models are improved, and plans adjusted as one titrates toward the goal. So, it is important to continue (and likely expand) nutrient reduction efforts while monitoring progresses.

7. Addressing ecological forecasting challenges

As applied ecological forecasts becomes more common, decision makers and forecasters should ensure that model development, testing, and application becomes more formalized. While the variety of model types and applications argues against rigid rules, evaluating results within a set of best practices provides more robust models and management guidance. Here we summarize our results within a set of those best practices.

7.1. Incorporating new knowledge and data in model construction

The Bayesian forecasting model has evolved over time (Obenour et al., 2014; Bertani et al., 2016), and it has been tested against alternative model structures (Scavia et al., 2021a). During that evolution, this model (along with others) has been used for annual HAB forecasts (NOAA, 2021) and policy scenarios that supported binational environmental decisions (Scavia et al., 2016). The near-term forecasts made annually since 2009 follow the iterative forecasting approach suggested by Dietze et al. (2018) as a way to test skill and improve the model over time. Our new model incorporates additional drivers (e.g., cumulative load), and recognizing that previous models did not adequately capture large blooms, introduces a segmented (change-point) response to the spring BioP load.

The segmented load-bloom formulation allows for a change in the marginal rate of bloom expansion above some threshold in loading (the change point), which is not available in a purely linear model (Ho and Michalak, 2017) and which is rigidly pre-specified in an exponential relationship (Stumpf et al., 2012). Thus, the segmented model addresses a structural uncertainty in the load-bloom relationship, reflected by the uncertainty in the two additional model parameters (change point, and slope adjustment above the change point). While it may not have been practical to include such a relationship in earlier iterations of the model (Obenour et al., 2014, Bertani et al., 2016), the incorporation of an expanded dataset helps make this possible. In general, as more data become available, HAB forecasting models can support more complex relationships with additional parameters, but to balance model complexity with increasing parameter uncertainty (Dietze et al., 2018), that uncertainty should be rigorously quantified, as demonstrated here through Bayesian inference.

Incorporating new information (e.g., Ho and Michalak, 2017) also provides insights into historical trends and potential delays in response to reduced loads. For example, we show that the decreasing cumulative load in the 1980s and 1990s likely delayed the HAB response to increasing spring loads during that period. After increasing between 1997 and 2010, the cumulative load remained fairly constant, and this new regime associated with high phosphorus recycling, promoted the emergence of HABs in the lake. We also show that the influence of the cumulative load will not only delay responses to load reductions, but when combined with the current loading target, may eventually lead to HAB extents well below the intended goal.

7.2. Evaluate forecast performance through validation and blind forecasts

Dietze et al. (2018) encouraged avoiding waiting until models are "good enough" before starting to provide forecasts that are useful to environmental management, and that allow us to learn by doing. For the Lake Erie HAB model, our leave-one-year-out cross-validation, a measure of how the model would work when forecasting outside the calibration data set, explained 65 % of the interannual variability between 2002 and 2020, compared to 84 % with the fully calibrated model. When making blind forecasts of previous events with a model calibrated through the prior year, the new model explained 54 % of HAB variability between 2009 and 2020. However, the new model explained 84 % of the interannual variability, compared to 32 % explained with actual forecasts made between 2014 and 2020. Building similar track records and conducting routine improvements will become increasingly possible as data on relevant drivers and state variables continue to grow.

7.3. Account for multiple sources of uncertainty

Because forecasts, especially ecological ones, are not perfect they should be presented as probability distributions that recognize their uncertainties (Dietze et al., 2018). The Bayesian framework is an appropriate approach; however, an honest accounting of the prediction uncertainty requires an accounting of all its sources. Not doing so, can result in falsely overconfident predictions. Our model accounts for parameter uncertainty, prediction error, and measurement error (Fig. 1), as well as discharge uncertainty (Fig. 6). Compared to many other HAB forecasting models, we provide a relatively comprehensive treatment of uncertainties in the model formulation, including the relevant loading windows (through parameters t_w and t_c), phosphorus bioavailability (θ), and changes in the load-bloom response slope (W_{cp} and b_{w2}).

Discharge (i.e., hydro-meteorological) variability can be a driving source of load uncertainty. While measured loads can be used for annual forecasts, it is important to account for discharge uncertainty in longer-term scenarios, especially when comparing to criteria based on allowable exceedances (e.g., HAB target must be met 9 out of 10 years). In our case, discharge uncertainty increased the average width of the 80 % prediction intervals by 3.5 Gg, from 19.6 to 23.1 Gg, for spring loads sufficiently below the 0.15 Gg/month change point. Because our load-reduction scenarios were well below the change point, accounting for discharge uncertainty only increased the needed load reduction from 39 % to 41 %. However, when loads regularly exceed the change point, as in recent years (2015–2020), the contribution of discharge uncertainty to the 80 % credible intervals increases to about 21 Gg (from 3 Gg).

While discharge uncertainty is a non-reducible property of the natural system, HAB measurement error and model prediction error can be potentially reduced through additional monitoring and research. Both sources of error were found to be substantial in this analysis (Table 1). Measurement error can be reduced through enhanced monitoring with high-frequency sensors and remote sensing to reduce spatial and temporal gaps. It can also be reduced through sophisticated data integration techniques that generate HAB extent estimates with quantified uncertainties (Fang et al., 2019). As three-dimensional ecological models improve (e.g., Wang and Boegman, 2021; Verhamme et al., 2016; Bocaniov et al., 2016; Zhang et al., 2016), they can also integrate point estimates across time and space to generate continuous time series. And these approaches can potentially be integrated. For example, Matli et al. (2018, 2020) incorporated output from a three-dimensional biogeochemical model as covariates in a space–time geostatistical analysis of Gulf of Mexico hypoxia, reducing uncertainties by 11–40 % compared to using measurement alone.

Predictive model error comes from incompletely representing system dynamics and drivers. In our case, replacing the time term (e.g., Obenour et al., 2014) with cumulative BioP load (adapted from Ho and Michalak, 2017) reduced the predictive error. In addition, expanding the "spring" loading time frame to include July loads and moving to the segmented model are other enhancements that reduced predictive error. Further enhancements to the model formulation may also be warranted, but should be assessed through a consideration of parameter uncertainties and validation exercises to prevent overfitting.

Given that the new model explains 84 % of the inter-annual HAB variability, assessing the impact of secondary variables, such as temperature and wind, may be better accommodated through more mechanistic biogeochemical models. But while these models help understand dynamics and have been used for scenarios (Wang and Boegman, 2021; Bocaniov et al., 2016; Lam et al., 2008; Zhang et al., 2016; Verhamme et al., 2016) and very short-term (e.g., daily) forecasts (Rowe et al., 2016, Wynne et al., 2013, Liu et al., 2020a, 2020b), they require drivers that are not reliably forecasted at seasonal time scales, making them less effective in seasonal forecasts. HAB models of intermediate complexity that combine the strengths of data assimilation with relatively simple ecological representations (e.g., Del Giudice et al., 2021; Katin et al., 2022) are alternatives that can also explain additional variability while still characterizing uncertainties. At the end of the day, all ecological forecast models will need to strike a balance between performance, parsimony, and the availability of reliable model inputs.

7.4. Apply model to answer management questions with short-term forecasts and scenarios

Constructing and testing ecological models, especially against new data, can help advance ecological science by providing more rigorous tests of hypotheses (Dietze et al., 2018); but applying those models in the context of environmental decision making is what makes the science socially relevant. We used the revised model to address key management questions. We demonstrated that including loads through July is both more consistent with current policy and improves annual forecast performance. However, to support the demand for earlier forecasts, we provided regression models to predict July loads from early July precipitation. Using these predicted loads for 2002–2020 did not degrade the predictions.

We reevaluated the GLWQA loading targets and showed that HAB projections depend not only on spring load reductions, but also on how the cumulative load responds. Similar to Ho and Michalak (2017), we showed that if spring loads were held consistently lower, the cumulative load will decline slowly such that the full HAB response is delayed by 5–16 years with an immediate load reduction, 15–22 years if the reduction was 10 % of the target per year.

Our projections also showed that once the cumulative load equilibrates with the GLWQA target spring loads, the "9 out of 10" goal is met, but the average HAB extent would be substantially lower than the policy goal. The high fiscal and social cost of load reduction management actions underscores the need for adaptive management where loads and HABs are monitored after actions are taken, models are updated, and strategies are adjusted as needed over time.

8. Model limitations and future directions

There are advantages and disadvantages to using simple models for forecasting (e.g., Glibert et al., 2010). Key advantages include the reliance on

data that are relatively easy to access (e.g., nutrient loads and HAB extents), the ability to quantify uncertainty from multiple sources (e.g., parameter, prediction, and measurement errors), and the relative ease in testing new relationships (e.g., the explanatory power of alternative drivers). For seasonal forecasting, they also have the advantage of being driven by properties that can be observed in time for making the forecasts (e.g., nutrient loads). A key disadvantage is the difficulty in integrating and testing more complex relationships like short term effects of winds, temperature, water movements, and food-web dynamics. More complex three-dimensional ecological models can be useful for these explorations, but some of their drivers also require forecasts (e.g., wind) that are, so far, only good for a few days, not seasons.

In light of the best practices outlined above, there are potential improvements to this simple model that could strike a balance between enhancing its mechanistic foundation and maintaining the advantages of simpler models. By incorporating new data and accounting for multiple sources of uncertainty, it may be possible to improve predictive skill and answer management questions related to seasonal and long-term forecasts. For example, while we suggest that the legacy loading impact on HAB extent has a solid mechanistic foundation, the modest over-predictions of small blooms in recent years (e.g., 2020) could potentially be improved with refinements and testing as new data become available. Alternative weighting functions (rather than the simple uniform/rectangular weighing used here) could be explored to give more weight to more recent loads. Such weighing could potentially be informed by sediment diagenesis models or other mechanistic considerations. Alternative formulations to our segmented linear model to simulate the change in HAB response to nutrient load changes could also be tested as more extreme HAB extent cases occur. These tests and others would also benefit from a reconciliation of HAB extent measurement methodologies. Several remote-sensing and in-situ methods are currently available, and various models are being developed based on different measures, which complicate comparisons and the development of ensemble forecasts.

Credit authorship contribution statement

Donald Scavia: Conceptualization; Formal analysis; Supervision; Project Administration; Resources; Validation; Roles/Writing – original draft.

Yu-Chen Wang: Data curation; Formal analysis; Methodology; Writing – review & editing.

Daniel Obenour: Conceptualization; Formal analysis; Methodology; Resources; Supervision; Validation, Writing – review & editing.

Data availability

Access to data and code is provided as a link in Supplemental Information.

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.

Acknowledgements

We thank Isabella Bertani for her continuing advice. Shiqi Fang (North Carolina State University), Michael Sayers and Karl Bosse (Michigan Technological University), and Richard Stumpf (NOAA/NCCOS) provided HAB extent estimates. Laura Johnson (Heidelberg University) provided Maumee River loading data. Asmita Murumkar and Margaret Kalcic (Ohio State University) provided historical watershed total precipitation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2022.158959.

References

Allan, J.D., Manning, N.F., Smith, S.D., Dickinson, C.E., Joseph, C.A., Pearsall, D.R., 2017. Ecosystem services of Lake Erie: spatial distribution and concordance of multiple services. J. Great Lakes Res. 43 (4), 678–688.

Baker, D.B., Ewing, D.E., Johnson, L.T., Kramer, J.W., Merryfield, B.J., Confesor, R.B., Richards, R.P., Roerdink, A.A., 2014. Lagrangian analysis of the transport and processing of agricultural runoff in the lower Maumee River and Maumee Bay. J. Great Lakes Res. 40, 479–495.

Beckage, B., Gross, L.J., Kauffman, S., 2011. The limits to prediction in ecological systems. Ecosphere 2, 125.

Bertani, I., Obenour, D.R., Steger, C.E., Stow, C.A., Gronewold, A.D., Scavia, D., 2016. Probabilistically assessing the role of nutrient loading in harmful algal bloom formation in western Lake Erie. J. Great Lakes Res. 42 (1184), 1192.

Bever, A.J., Friedrichs, M.A.M., St-Laurent, P., 2021. Real-time environmental forecasts of the Chesapeake Bay: model setup, improvements, and online visualization. Environ. Model. Softw. 140, 105036.

Bocaniov, S.A., Keon, L.F., Rao, Y.R., Schwab, D.J., Scavia, D., 2016. Simulating the effect of nutrient reduction on hypoxia in a large lake (Lake Erie, USA-Canada) with a three-dimensional lake model. J. Great Lakes Res. 42, 1228–1240.

Bradford, J.B., Weltzin, J.F., McCormick, M., Baron, J., Bowen, Z., Bristol, S., Carlisle, D., Crimmins, T., Cross, P., DeVivo, J., Dietze, M., Freeman, M., Goldberg, J., Hooten, M., Hsu, L., Jenni, K., Keisman, J., Kennen, J., Lee, K., Lesmes, D., Loftin, K., Miller, B.W., Murdoch, P., Newman, J., Prentice, K.L., Rangwala, I., Read, J., Sieracki, J., Sofaer, H., Thur, S., Toevs, G., Werner, F., White, C.L., White, T., Wiltermuth, M., 2020. Ecological Forecasting—21st Century Science for 21st Century Management: U.S. Geological Survey Open-file Report 2020–1073. 54 p. https://doi.org/10.3133/ofr20201073.

Bridgeman, T.B., Chaffin, J.D., Filbrun, J.E., 2013. A novel method for tracking western Lake Erie Microcystis blooms, 2002–2011. J. Gt. Lakes Res. 39, 83–89. https://doi.org/10.1016/j.jglr.2012.11.004.

Brooks, S.P., Gelman, A., 1998. General methods for monitoring convergence of iterative simulations. J. Comput. Graph. Stat. 7 (4), 434–455. https://doi.org/10.1080/10618600.1998.10474787.

Bullerjahn, G.S., McKay, R.M., Davis, T.W., Baker, D.B., Boyer, G.L., D'Anglada, L.V., Doucette, G.J., Ho, J.C., Irwin, E.G., Kling, C.L., Kudela, R.M., Kurmayer, R., Michalak, A.M., Ortiz, J.D., Otten, T.G., Paerl, H.W., Qin, B., Sohngen, B.L., Stumpf, R.P., Visser, P.M., Wilhelm, S.W., 2016. Global solutions to regional problems: collecting global expertise to address the problem of harmful cyanobacterial blooms. A Lake Erie case study. Harmful Algae, Global Expansion of Harmful Cyanobacterial Blooms: Diversity, ecology, causes, and controls 54, 223–238. https://doi.org/10.1016/j.hal.2016.01.003.

Burns, N.M., 1976. Temperature, oxygen, and nutrient distribution patterns in Lake Erie, 1970. J. Fish. Res. Board Can. 33, 485–511. https://doi.org/10.1139/f76-068.

CENR, 2001. Ecological Forecasting: Agenda for the Future. Committee on Environment and Natural Resources, National Science and Technology Council, Washington DC.

Chapra, S.C., Canale, R.P., 1991. Long-term phenomenological model of phosphorus and oxygen for stratified lakes. Wat. Re. 25, 707–715.

Chatfield, C., 2006. Model Uncertainty. Encyclopedia of Environmetrics. John Wiley, Hoboken, N. J.

Clark, J.S., 2005. Why environmental scientists are becoming Bayesians. Ecol. Lett. 8, 2–14.

Coreau, A., Pinay, G., Thompson, J.D., Cheptou, P.-O., Mermet, L., 2009. The rise of research on futures in ecology: rebalancing scenarios and predictions. Ecol. Lett. 12, 1277–1286.

Daloğlu, I., Cho, K.H., Scavia, D., 2012. Evaluating causes of trends in long-term dissolved reactive phosphorus loads to Lake Erie. Environ. Sci. Technol. 46, 10660–10666.

Davidson, K., Anderson, D.M., Mateus, M., Reguera, B., Silke, J., Sourisseau, M., Maguire, J., 2013. Forecasting the risk of harmful algal blooms. Harmful Algae 42, 1–7.

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.

Dietze, M.C., Fox, A., Beck-Johnson, L.M., Betancourt, J.L., Hooten, M.B., Jarnevich, C.S., Keitt, T.H., Kenney, M.A., Laney, C.M., Larsen, L.G., Loescher, H.W., Lunch, C.K., Pijanowski, B.C., Randerson, J.T., Read, E.K., Tredennick, A.T., Vargas, R., Weathers, K.C., White, E.P., 2018. Iterative near-term ecological forecasting: needs, opportunities, and challenges. Proc. Natl. Acad. Sci. 115 (7), 1424–1432.

EFI, 2020. Forecasts to understand, manage, and conserve ecosystems. Ecological Forecasting Initiative Webpage. Accessed November 2020. https://ecoforecast.org.

Elsner, J.B., Schmertmann, C.P., 1994. Assessing forecast skill through cross validation. Weather Forecast. 9 (4), 619–624.

Fang, S., Giudice, D.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. https://doi.org/10.1016/j.scitotenv.2019.133776.

Gibbons, K.J., Bridgeman, T.B., 2020. Effect of temperature on phosphorus flux from anoxic western Lake Erie sediments. Water Res. 182, 116022. https://doi.org/10.1016/j.watres.2020.116022.

Gimenez, O., Buckland, S.T., Morgan, B.J.T., Bez, N., Bertrand, S., Choquet, R., Dray, S., Etienne, M.-P., Fewster, R., Gosselin, F., Merigot, B., Monestiez, P., Morales, J.M., Mortier, F., Munoz, F., Ovaskainen, O., Pavoine, S., Pradel, R., Schurr, F.M., Thomas, L., Thuiller, W., Trenkel, V., de Valpine, P., Rexstad, E., 2014. Statistical ecology comes of age. Biol. Lett. 10 (12), 20140698.

Glibert, P.M., Icarus Allen, J., Bouwman, A.F., Brown, C.W., Flynn, K.J., Lewitus, A.J., Madden, C.J., 2010. Modeling of HABs and eutrophication: status, advances, challenges. J. Mar. Syst. 83, 262–275.

GLWQA (Great Lakes Water Quality Agreement), 2016. The United States and Canada Adopt Phosphorus Load Reduction Targets to Combat Lake Erie Algal Blooms. https://binational.net/2016/02/22/finalptargets-ciblesfinalesdep.

Harre, L.B., Campbell, L., 2014. Predicting harmful algal blooms: a case study with Dinophysis ovum in the Gulf of Mexico. J. Plankton Res. 36 (6), 1434–1445. https://doi.org/10.1093/plankt/fbu070.

Harwood, J., Stokes, K., 2003. Coping with uncertainty in ecological advice: lessons from fisheries. Trends in Ecology and Evolution 18 (12), 617–622.

Ho, J.C., Michalak, A.M., 2017. Phytoplankton blooms in Lake Erie impacted by both long-term and springtime phosphorus loading. J. Great Lakes Res. 43 (3), 221–228.

Jankowiak, J., Hattenrath-Lehmann, T., Kramer, B.J., Ladds, M., Gobler, C.J., 2019. Deciphering the effects of nitrogen, phosphorus, and temperature on cyanobacterial bloom intensification, diversity, and toxicity in western Lake Erie. Limnol. Oceanogr. 64, 1347–1370. https://doi.org/10.1002/lno.11120.

Jarvie, H.P., Johnson, L.T., Sharpley, A.N., Smith, D.R., Baker, D.B., Bruulsema, T.W., Confesor, R., 2017. Increased soluble phosphorus loading to Lake Erie: unintended consequences of conservation practices? J. Environ. Qual. 46, 123–132. https://doi.org/10.2134/jeq2016.07.0248.

Jensen, J.P., Pedersen, A.R., Jeppesen, E., Sondergaard, M., 2006. An empirical model describing the seasonal dynamics of phosphorus in 16 shallow eutrophic lakes after external loading reduction. Limnol. Oceanogr. 51, 791–800.

Jeppesen, E., Søndergaard, M., Meerhoff, M., Lauridsen, T.L., Jensen, J.P., 2007. Shallow lake restoration by nutrient loading reduction - some recent findings and challenges ahead. Hydrobiologia 584, 239–252. https://doi.org/10.1007/s10750-007-0596-7.

Jewett, E.B., Lopez, C.B., Dortch, Q., Etheridge, S.M., 2007. National Assessment of Efforts to Predict and Respond to Harmful Algal Blooms in U.S. Waters. Interagency Working Group on Harmful Algal Blooms, Hypoxia, and Human Health of the Joint Subcommittee on Ocean Science and Technology. Office of Science and Technology Policy, Washington, DC (61 pp).

Johnson-Bice, S.M., Ferguson, J.M., Erb, J.D., Gable, T.D., Windels, S.K., 2020. Ecological forecasts reveal limitations of common model selection methods: predicting changes in beaver colony densities. Ecol. Appl. e02198.

Katin, A., Del Giudice, D., Obenour, D.R., 2019. Modeling biophysical controls on hypoxia in a shallow estuary using a Bayesian mechanistic approach. Environ. Model Softw. 120, 104491.

Katin, A., Del Giudice, D., Obenour, D.R., 2022. Temporally resolved coastal hypoxia forecasting and uncertainty assessment via Bayesian mechanistic modeling. Hydrol. Earth Syst. Sci. 26, 1131–1143. https://doi.org/10.5194/hess-26-1131-2022.

Lam, D., Schertzer, W., McCrimmon, R., Charlton, M., Millard, S., 2008. Modeling phosphorus and dissolved oxygen conditions pre- and post-Dreissena arrival in Lake Erie. In: Munawar, M., Heath, R. (Eds.), Checking the Pulse of Lake Erie. Aquatic Ecosystem Health and Management Society, Burlington, Ontario.

Liu, H., Zhang, W., Irwin, E., Kast, J., Aloysius, N., Martin, J., Kalcic, M., 2020b. Best management practices and nutrient reduction: an integrated economic-hydrologic model of the Western Lake Erie Basin. Land Economics 96 (November 2020), 510–530.

Liu, Q., Row, M.D., Anderson, E.J., Stow, C.A., Stump, R.P., Johengen, T.H., 2020a. Probabilistic forecast of microcystin toxin using satellite remote sensing, *in situ* observations and numerical modeling. Environ. Model Softw. 128, 104705.

Lunn, D., Jackson, C., Best, N., Thomas, A., Spiegelhalter, D., 2013. The BUGS Book. A Practical Introduction to Bayesian Analysis. Chapman Hall, London.

Lunn, D.J., Thomas, A., Best, N., Spiegelhalter, D., 2000. WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. Stat. Comput. 10, 325–337.

Luo, Y.Q., Ogle, K., Tucker, C., Fei, S.F., Gao, C., LaDeau, S., Clark, J.S., Schimel, D.S., 2011. Ecological forecasting and data assimilation in a data-rich era. Ecol. Appl. 21 (5), 1429–1442.

Maccoux, M.J., Dove, A., Backus, S.M., Dolan, D.M., 2016. Total and soluble reactive phosphorus loadings to Lake Erie: a detailed accounting by year, basin, country, and tributary. J. Great Lakes Res. 42, 1151–1165. https://doi.org/10.1016/j.jglr.2016.08.005.

Manning, N.F., Wang, Y.C., Long, C.M., Bertani, I., Sayers, M.J., Bosse, K.R., Shuchman, R.A., Scavia, D., 2019. Extending the forecast model: predicting harmful algal blooms at multiple spatial scales. J. Great Lakes Res. 45, 587–595.

Martin, J., Kalcic, M., Aloysius, N., Apostel, A., Brooker, M., Evenson, G., Kast, J., Kujawa, H., Murumkar, A., Becker, R., Boles, C., Confesor, R., Dagnew, A., Guo, T., Long, C., Muenich, R., Scavia, D., Redder, T., Robertson, D., Wang, Y.-C., 2021. Evaluating management options to reduce Lake Erie algal blooms using an ensemble of watershed models. J. Environ. Management. 28, 111710.

Matisoff, G., Kaltenberg, E.M., Steely, R.L., Hummel, S.K., Seo, J., Gibbons, K.J., Bridgeman, T.B., Seo, Y., Behbahani, M., James, W.F., Johnson, L.T., Doan, P., Dittrich, M., Evans, M.A., Chaffin, J.D., 2016. Internal loading of phosphorus in western Lake Erie. J. Great Lakes Res. 42, 775–788.

Matli, V.R.R., Fang, S., Guinness, J., Rabalais, N.N., Craig, J.K., Obenour, D.R., 2018. Space-time geostatistical assessment of hypoxia in the northern Gulf of Mexico. Environ. Sci. Technol. 52, 12484–12493.

Matli, V.R.R., Laurent, A., Fennel, K., Craig, K., Krause, J., Obenour, D.R., 2020. Fusion-based hypoxia estimates: combining geostatistical and mechanistic models of dissolved oxygen variability. Environ. Sci. Technol. 54, 13016–13025.

Michalak, A.M., Anderson, E.J., Beletsky, D., Boland, S., Bosch, N.S., Bridgeman, T.B., Chaffin, J.D., Cho, K., Confesor, R., Daloğlu, I., DePinto, J.V., Evans, M.A., Fahnenstiel, G.L., He, L., Ho, J.C., Jenkins, L., Johengen, T.H., Kuo, K.C., LaPorte, E., Liu, X., McWilliams, M.R., Moore, M.R., Posselt, D.J., Richards, R.P., Scavia, D., Steiner, A.L., Verhamme, E., Wright, D.M., Zagorski, M.A., 2013. Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions. Proc. Natl. Acad. Sci. 110, 6448–6452. https://doi.org/10.1073/pnas.1216006110.

NASA, 2020. Ecological Forecasting: Strengthening Ecosystems. National Aeronautics and Space Administration Webpage. Accessed November 2020 https://appliedsciences.nasa.gov/what-we-do/ecological-forecasting.

NOAA, 2020. Predicting human health and coastal economies with early warnings. NOAA Ecological Forecasting Webpage. Accessed November 2020 https://oceanservice.noaa.gov/ecoforecasting/.

NOAA, 2021. Samples of Lake Erie ensemble annual HAB forecasts. https://nccospublicstor.blob.core.windows.net/hab-data/bulletins/lake-erie/current/bulletin_current.pdf. https://www.noaa.gov/news-release/smaller-summer-harmful-algal-bloom-predicted-for-western-lake-erie.

NOAA"s Weather Prediction Center, 2021. Quantitative Precipitation Forecasts. https://origin.wpc.ncep.noaa.gov/qpf/day1-7.shtml.

North Carolina Sea Grant, 2020. Midsummer Neuse River Forecast Shows Greater Potential for Fish Kills. Webpage. Accessed November 2020 https://ncseagrant.ncsu.edu/currents/2020/06/midsummer-neuse-river-forecast-shows-greater-potential-for-fish-kills/.

Obenour, D.R., Gronewold, A.D., Stow, C.A., Scavia, D., 2014. Using a Bayesian hierarchical model to improve Lake Erie cyanobacteria bloom forecasts. Water Resour. Res. 50, 7847–7860. https://doi.org/10.1002/2014WR015616.

Palm-Forster, L.H., Swinton, S.M., Redder, T.M., DePinto, J.V., Boles, C.M.W., 2016. Using conservation auctions informed by environmental performance models to reduce agricultural nutrient flows into Lake Erie. J. Great Lakes Res. 42, 1357–1371.

Pappenberger, F., Beven, K.J., 2006. Ignorance is bliss: or seven reasons not to use uncertainty analysis. Water Resour. Res. 42, W05302.

Payne, M.R., Hobday, A.J., MacKenzie, B.R., Tommasi, D., Dempsey, D.P., Fassler, S., Haynie, A.C., Ji, R., Liu, G., Lynch, P.D., Matei, D., Miesner, A.K., Mills, K.E., Strand, K.O., Villarino, E., 2017. Lessons from the first generation of marine ecological forecast products. Front. Mar. Sci. 4, 289.

Phillips, G., Kelly, A., Pitt, J.A., Sanderson, R., Taylor, E., 2005. The recovery of a very shallow eutrophic lake, 20 years after the control of effluent derived phosphorus. Freshw. Biol. 50, 1628–1638. https://doi.org/10.1111/j.1365-2427.2005.01434.x.

Press Releases, 2021. https://news.umich.edu/changing-weather-patterns-mix-up-the-size-duration-of-annual-chesapeake-bay-dead-zone/. https://news.umich.edu/smaller-summer-harmful-algal-bloom-predicted-for-western-lake-erie/. https://news.umich.edu/average-sized-dead-zone-forecast-for-gulf-of-mexico/.

Qian, S.S., Cuffney, T.F., Alameddine, I., McMahon, G., Reckhow, K.H., 2010. On the application of multilevel modeling in environmental and ecological studies. Ecology 91 (2), 355–361.

R Core Team, 2015. R: A Language and Environment for Statistical Computing.

Raftery, A.E., 2016. Use and communication of probabilistic forecasts. Statistical Analysis and Data Mining: The ASA Data Science Journal 9 (6), 397–410.

Robertson, D.M., Diebel, M.W., 2020. Improtance of accurately quantifying internal loading in developing phosphorus reduction strategies for a chain of shallow lakes. Lake and Reservoir Management 36, 391–411.

Ross, Andrew C., Ross, A.C., Stock, C.A., Dixon, K.W., Friedrichs, M.A.M., Hood, R.R., Li, M., Pegion, K., Saba, V., Vecchi, G.A., 2020. Estuarine forecasts at daily weather to subseasonal time scales. Earth and Space Sci. 7, e2020EA001179.

Rowe, M.D., Anderson, E.J., Wynne, T.T., Stumpf, R.P., Fanslow, D.L., Kijanka, K., Vanderploeg, H.A., Strickler, J.R., Davis, T.W., 2016. Vertical distribution of buoyant Microcystis blooms in a Lagrangian particle tracking model for short-term forecasts in Lake Erie. J. Geophys. Res. Oceans 121, 5296–5314. https://doi.org/10.1002/2016jc011720.

Sabo, R.D., Sullivan, B., Wu, C., Trentacoste, E., Zhang, Q., Shenk, G., Bhatt, G., Linker, L.C., 2022. Major Point and Nonpoint Sources of Nutrient Pollution to Surface Water Have Declined Throughout the Chesapeake Bay Watershed. https://doi.org/10.1088/2515-7620/ac5db6.

Sayers, M., Fahnenstiel, G.L., Shuchman, R.A., Whitley, M., 2016. Cyanobacteria blooms in three eutrophic basins of the Great Lakes: a comparative analysis using satellite remote sensing. Int. J. Remote Sens. 37, 4148–4171.

Scavia, D., Donnelly, K.A., 2007. Reassessing hypoxia forecasts for the Gulf of Mexico. Env. Sci. Technol. 41, 8111–8117.

Scavia, D., Rabalais, N.N., Turner, R.E., Justic, D., Wiseman Jr., W., 2003. Predicting the response of Gulf of Mexico hypoxia to variations in Mississippi River nitrogen load. Limnol. Oceanogr. 48 (3), 951–956.

Scavia, D., Kelly, E.L.A., Hagy, J.D., 2006. A simple model for forecasting the effects of nitrogen loads on Chesapeake Bay hypoxia. Estuar. Coasts 29 (4), 674–684.

Scavia, D., David Allan, J., Arend, K.K., Bartell, S., Beletsky, D., Bosch, N.S., Brandt, S.B., Briland, R.D., Daloğlu, I., DePinto, J.V., et al., 2014. Assessing and addressing the re-eutrophication of Lake Erie: Central basin hypoxia. J. Gt. Lakes Res. 40 (2), 226–246.

Scavia, D., DePinto, J.V., Bertani, I., 2016. A multi-model approach to evaluating target phosphorus loads for Lake Erie. J. Gt. Lakes Res. 42 (6), 1139–1150.

Scavia, D., Bertani, I., Obenour, D.R., Turner, R.E., Forrest, D.R., Katin, A., 2017a. Ensemble modeling informs hypoxia management in the northern Gulf of Mexico. Proc. Natl. Acad. Sci. 114, 8823–8828.

Scavia, D., Kalcic, M., Logsdon Muenich, R., Aloysius, N., Bertani, I., Boles, C., Confesor, R., DePinto, J., Gildow, M., Martin, J., Read, J., Redder, T., Robertson, D., Sowa, S., Wang, Y., Yen, H., 2017b. Multiple models guide strategies for agricultural nutrient reductions. Front. Ecol. Environ. 15, 126–132.

Scavia, D., Justic, D., Obenour, D., Craig, K., Wang, L., 2019. Hypoxic volume is more responsive than hypoxic area to nutrient load reductions in the northern Gulf of Mexico – and it matters to fish and fisheries. Env. Res. Lett. 14, 024012.

Scavia, D., Wang, Y.-C., Obenour, D.R., Apostel, A., Basile, S.J., Kalcic, M.M., Kirchhoff, C.J., Miralha, L., Muenich, R.L., Steiner, A.L., 2021a. Quantifying uncertainty cascading from climate, watershed, and lake models in harmful algal bloom predictions. Science of the Total Environment 759 (10 March 2021), 143487. https://doi.org/10.1016/j.scitotenv.2020.143487.

Scavia, D., Bertani, I., Bever, A.J., Blomquist, J.D., Friedrichs, M.A.M., Linker, L.C., Michael, B.D., Murphy, R.R., Shenk, G.W., Testa, J.M., 2021b. Advancing estuarine ecological forecasts: seasonal hypoxia in Chesapeake Bay. Ecol. Appl. https://doi.org/10.1002/eap.2384.

Scavia, D., Wang, Y.-C., Bertani, I., Obenour, D., Long, C.M., 2021c. Western Lake Erie Harmful Algal Bloom Forecast. Chesapeake Bay Hypoxia Forecast, Gulf of Mexico Hypoxia Forecast http://scavia.seas.umich.edu/hypoxia-forecasts/.

Schindler, D.E., Hilborn, R., 2015. Prediction, precaution, and policy under global change. Science 347, 953–954.

Shamshirband, S., Nodoushan, E.J., Adolf, J.E., Manaf, A.A., Mosavi, A., Chau, K.-W., 2019. Ensemble models with uncertainty analysis for multi-day ahead forecasting of chlorophyll a concentration in coastal waters. Engineering Applications of Computational Fluid Mechanics 13 (1), 91–101 (2019).

Smith, D.R., King, K.W., Williams, M.R., 2015. What is causing the harmful algal blooms in Lake Erie? J. Soil Water Conserv. 70, 27A–29A.

Stackpoole, S., Sabo, R., Falcone, J., Sprague, L., 2021. Long-term Mississippi River trends expose shifts in the river load response to watershed nutrient balances between 1975 and 2017. Water Resour. Res. 57, e2021WR030318. https://doi.org/10.1029/2021WR030318.

Stow, C.A., Cha, Y., Johnson, L.T., Confesor, R., Richards, R.P., 2015. Long-term and seasonal trend decomposition of Maumee River nutrient inputs to western Lake Erie. Environ. Sci. Technol. 49, 3392–3400.

Stumpf, R.P., Wynne, T.T., Baker, D.B., Fahnenstiel, G.L., 2012. Interannual variability of cyanobacterial blooms in Lake Erie. PLoS One 7 (8), e42444. https://doi.org/10.1371/journal.pone.0042444.

Stumpf, R.P., Johnson, L.T., Wynne, T.T., Baker, D.B., 2016. Forecasting annual cyanobacterial bloom biomass to inform management decisions in Lake Erie. J. Gt. Lakes Res. 42, 1174–1183. https://doi.org/10.1016/j.jglr.2016.08.006.

Sturtz, S., Ligges, U., Gelman, A., 2005. R2WinBUGS: a package for running WinBUGS from R. J. Stat. Softw. 12, 1–16.

Testa, J.M., Clark, J.B., Dennison, W.C., Donovan, E.C., Fisher, A.W., Ni, W., Parker, M., Scavia, D., Spitzer, S.E., Waldrop, A.M., Vargas, V.M.D., Ziegler, G., 2017. Ecological forecasting and the science of hypoxia in Chesapeake Bay. Bioscience 67 (7), 614–626.

Treuer, G., Kirchhoff, C., Lemos, M.C.F., McGrath, 2021. Challenges of managing harmful algal blooms in US drinking water systems. Nat Sustain. https://doi.org/10.1038/s41893-021-00770-y.

Valette-Silver, N., Scavia, D., 2003. Ecological forecasting: new tools for coastal and marine ecosystem management. NOAA Technical Memorandum NOS NCCOS 1 116 pp. http://scavia.seas.umich.edu/wp-content/uploads/2009/11/noaa_ecological_forecasting_book1.pdf.

Verhamme, E., Redder, T., Schlea, D., Grush, J., Bratton, J., DePinto, J., 2016. Development of the Western Lake Erie Ecosystem Model (WLEEM): application to connect phosphorus loads to cyanobacteria biomass. J. Great Lakes Res. 42 (6), 1193–1205.

VIMS, 2020. Chesapeake Bay Hypoxia Forecast. Webpage. Accessed November 2020. https://www.vims.edu/research/topics/dead_zones/forecasts/cbay/index.php.

Walters, C.J., Holling, C.S., 1990. Large-scale management experiments and learning by doing. Ecology 71, 2060–2068.

Wang, Q., Boegman, L., 2021. Multi-year simulation of Western Lake Erie hydrodynamics and biogeochemistry to evaluate nutrient management scenarios. Sustainability 13, 7516. https://doi.org/10.3390/su13147516.

Watson, S.B., Miller, C., Arhonditsis, G., Boyer, G.L., Carmichael, W., Charlton, M.N., Confesor, R., Depew, D.C., Höök, T.O., Ludsin, S.A., Matisoff, G., McElmurry, S.P., Murray, M.W., Peter Richards, R., Rao, Y.R., Steffen, M.M., Wilhelm, S.W., 2016. The re-eutrophication of Lake Erie: harmful algal blooms and hypoxia. Harmful Algae 56, 44–66. https://doi.org/10.1016/j.hal.2016.04.010.

Westgate, M.J., Likens, G.E., Lindenmayer, D.B., 2013. Adaptive management of biological systems: A review. Biol. Conserv. 158, 128–139.

White, E.P., Yenni, G.M., Taylor, S.D., Christensen, E.M., Bledsoe, E.K., Simonis, J.L., Ernest, S.M., 2019. Developing an automated iterative near-term forecasting system for an ecological study. Methods Ecol. Evol. 10 (3), 332–344.

Wynne, T.T., Stumpf, R.P., Tomlinson, M.C., Schwab, D.J., Watabayashi, G.Y., Christensen, J.D., 2011. Estimating cyanobacterial bloom transport by coupling remotely sensed imagery and a hydrodynamic model. Ecol. Appl. 21, 2709–2721.

Wynne, T.T., Stumpf, R.P., Tomlinson, M.C., Fahnenstiel, G.L., Dyble, J., Schwab, D.J., Joshi, S.J., 2013. Evolution of a cyanobacterial bloom forecast system in western Lake Erie: development and initial evaluation. J. Great Lakes Res. 39, 90–99.

Zhang, H., Hu, W., Gu, K., Li, Q., Zheng, D., Zhai, S., 2013. An improved ecological model and software for short-term algal bloom forecasting. Environ. Model Softw. 43, 152–162.

Zhang, H., Boegman, L., Scavia, D., Culver, D.A., 2016. Spatial distributions of external and internal phosphorus loads in Lake Erie and their impacts on phytoplankton and water quality. J. Great Lakes Res. 42, 1212–1227.