Spatio-Temporal Modeling for Forecasting High-Risk Freshwater Cyanobacterial Harmful Algal Blooms in Florida

Mark H. Myer [1], Erin Urquhart [2], Blake A. Schaeffer [3] and John M. Johnston [4*]

[1] US Environmental Protection Agency, Oak Ridge Institute for Science and Education (ORISE), Athens, GA, United States, [2] US Environmental Protection Agency, Oak Ridge Institute for Science and Education (ORISE), Research Triangle Park, NC, United States, [3] US Environmental Protection Agency, Center for Exposure Measurement and Modeling, Research Triangle Park, NC, United States, [4] US Environmental Protection Agency, Center for Exposure Measurement and Modeling, Athens, GA, United States

Due to the occurrence of more frequent and widespread toxic cyanobacteria events, the ability to predict freshwater cyanobacteria harmful algal blooms (cyanoHAB) is of critical importance for the management of drinking and recreational waters. Lake system specific geographic variation of cyanoHABs has been reported, but regional and state level variation is infrequently examined. A spatio-temporal modeling approach can be applied, via the computationally efficient Integrated Nested Laplace Approximation (INLA), to high-risk cyanoHAB exceedance rates to explore spatio-temporal variations across statewide geographic scales. We explore the potential for using satellite-derived data and environmental determinants to develop a short-term forecasting tool for cyanobacteria presence at varying space-time domains for the state of Florida. Weekly cyanobacteria abundance data were obtained using Sentinel-3 Ocean Land Color Imagery (OLCI), for a period of May 2016–June 2019. Time and space varying covariates include surface water temperature, ambient temperature, precipitation, and lake geomorphology. The hierarchical Bayesian spatio-temporal modeling approach in R-INLA represents a potential forecasting tool useful for water managers and associated public health applications for predicting near future high-risk cyanoHAB occurrence given the spatio-temporal characteristics of these events in the recent past. This method is robust to missing data and unbalanced sampling between waterbodies, both common issues in water quality datasets.

Keywords: harmful algal blooms, cyanobacteria, hierarchical Bayes, integrated nested Laplace approximation, remote sensing, predictive modeling

INTRODUCTION

Harmful algal blooms are environmental events that occur when algal populations achieve sufficiently high density resulting in possible adverse ecological and public health effects (Smayda, 1997). Harmful cyanobacteria blooms (cyanoHABs) are made up of naturally occurring photosynthetic prokaryotes found in various aquatic systems and can produce toxins (cyanotoxins). Toxic cyanoHABs are common components of Florida's surface waters and have

been identified since the late 1980s (Carmichael, 1992; Chapman and Schelske, 1997; Burns, 2008) in Florida's freshwater and brackish environments, including those used for recreation, source waters used for drinking water supply, and finished drinking waters (Williams et al., 2001, 2007; Florida Fish Wildlife Conservation Commission, 2017). Florida freshwater cyanoHABs have consisted primarily of the genera *Microcystis*, *Anabaena,* and *Cylindrospermopsis* and their associated toxins: microcystins, anatoxin-a, and cylindrospermopsin, respectively (Burns et al., 2002; Phlips et al., 2002). Upon numerous sampling occasions, microcystin levels in Florida's recreational waters have exceeded the existing World Health Organization (WHO) (Chorus and Bartram, 1999; World Health Organization, 2003) and EPA (US EPA, 2019) guidelines for cyanobacterial toxins, underscoring the potential impact of cyanoHABs on public health (Williams et al., 2007).

Several of Florida's largest aquatic systems including Lake Okeechobee (Havens et al., 1998; Havens and Steinman, 2015); the Harris Chain of Lakes (Williams et al., 2001, 2007); and the St. Johns, St. Lucie and Caloosahatchee (Glibert et al., 2006; Boyer and FitzPatrick, 2016; United States Army Corps of Engineers, 2016) rivers have experienced the increasing adverse impacts of cyanoHABs. In 2005, the St. Johns County Department of Health released Florida's first official health alert for a toxigenic harmful algal bloom. In response, the St. Johns River Water Management District began routine daily sampling and issued weekly releases informing the public of high microcystin concentrations and risk (Williams et al., 2007). In July 2016, concentrations (4.5 mg L^{-1}) of microcystin were detected in the river-dominated Caloosahatchee and St. Lucie estuaries following very heavy rainstorms in Florida (Oehrle et al., 2017). Strong storms resulted in reservoir operators increasing the outflow from Lake Okeechobee causing the incursion of a toxic *M. aeruginosa* bloom into the St. Lucie Estuary (Boyer and FitzPatrick, 2016; United States Army Corps of Engineers, 2016).

Identification and quantification of the environmental factors that contribute to the proliferation of cyanoHABs in freshwater systems continues to be a topic of scientific research. It is generally understood that dense concentrations of cyanoHABs result from a combination of excess total anthropogenic nutrient loads, particularly phosphorus (Michalak et al., 2013). Other factors that can be important drivers of cyanoHAB abundance are positive associations with lake depth, water column stability, and water temperature (Paerl and Huisman, 2008; Taranu et al., 2012; Beaulieu et al., 2013) and a negative association with wind speed (Millie et al., 2014). Landscape alterations such as urbanization or agricultural practice can change sediment loading and further alter nutrient availability in watersheds (Lunetta et al., 2015).

Modeling efforts to identify and predict harmful algal blooms have used several approaches, including classical multivariate analysis with LOWESS smoothing (Downing et al., 2001), continuous artificial neural networks (Millie et al., 2014), linear, non-linear, and mixed-effect models (Beaulieu et al., 2014), and Bayesian modeling (Obenour et al., 2014).

These data-driven modeling efforts do not address spatial and temporal correlation in bloom occurrence. Failure to account for spatial and temporal autocorrelation violates the assumption of independent and identically distributed data and may lead to biased model estimates. Further, incorporating Bayesian inference into predictive models allows inclusion of prior knowledge, either through literature review and expert opinion or by using mathematical techniques to generate informative priors. We therefore aim to improve upon cyanoHAB models by addressing spatio-temporal correlations using Bayesian hierarchical models, leading to estimates of explanatory variable effects that are more reliable for scientific inference.

The integrated nested Laplace approximations (INLA) method (Rue et al., 2009) offers a simple way to compute complicated hierarchical models that include spatial and temporal structure. Large computational times remain a major drawback of modeling spatial correlation. However, a recent solution has been developed using stochastic partial differential equations (SPDE) to provide a faster and computationally cheaper solution to these modeling problems (Lindgren et al., 2011). R-INLA is a modeling framework developed and implemented for the statistical software R (R Core Team, 2015) that allows complex Bayesian spatial modeling with far fewer computational resources than previous approaches (Rue et al., 2017). One benefit of R-INLA is that it is applicable at any spatial and/or temporal scale. The SPDE approach allows the user to model spatial correlation and constructs flexible fields that are better able to handle complex spatial structures than alternative spatial correlation models such as kriging (Cameletti et al., 2013), making this approach appropriate for inland waterbody modeling because we see both regional clustering of blooms and vast areas of water and land with no cyanobacteria observations. See **Supplemental Material 2**, Statistical Appendix for definitions and discussion of these statistical terms.

In this study, we apply Bayesian spatio-temporal models using R-INLA for the purposes of mapping and predicting exceedance probabilities of World Health Organization (WHO) high-risk for recreational exposure blooms (>100,000 cells/mL) in lakes across Florida. Mapping state level cyanoHAB exceedance probabilities identifies areas where estimates are higher or lower than the state average and provides insight through visualization of patterns in both space and time in addition to inferences drawn from the effects of predictor variables.

METHODS

Study Area

This study was conducted for 103 lakes in the state of Florida, United States. Lakes and waterbodies within three Water Management Districts: St. Johns River, Southwest Florida, and South Florida, were included in the analysis (**Figure 1**). Lake water pixels were extracted using the USGS/EPA Hydrography Dataset Plus (NHDPlus) version 2 (http://www.horizon-systems.com/nhdplus/NHDplusV2_home.php). All NHDPlus features classified as lakes and reservoirs were selected using U.S. Environmental Protection Agency's 2012 National Lakes Assessment (NLA) site evaluation guidelines. Lakes in the NHDPlus shapefile with a minimum of three 300 m water pixels remaining after a land adjacency QA flag was applied were considered resolvable waterbodies. The number of pixels
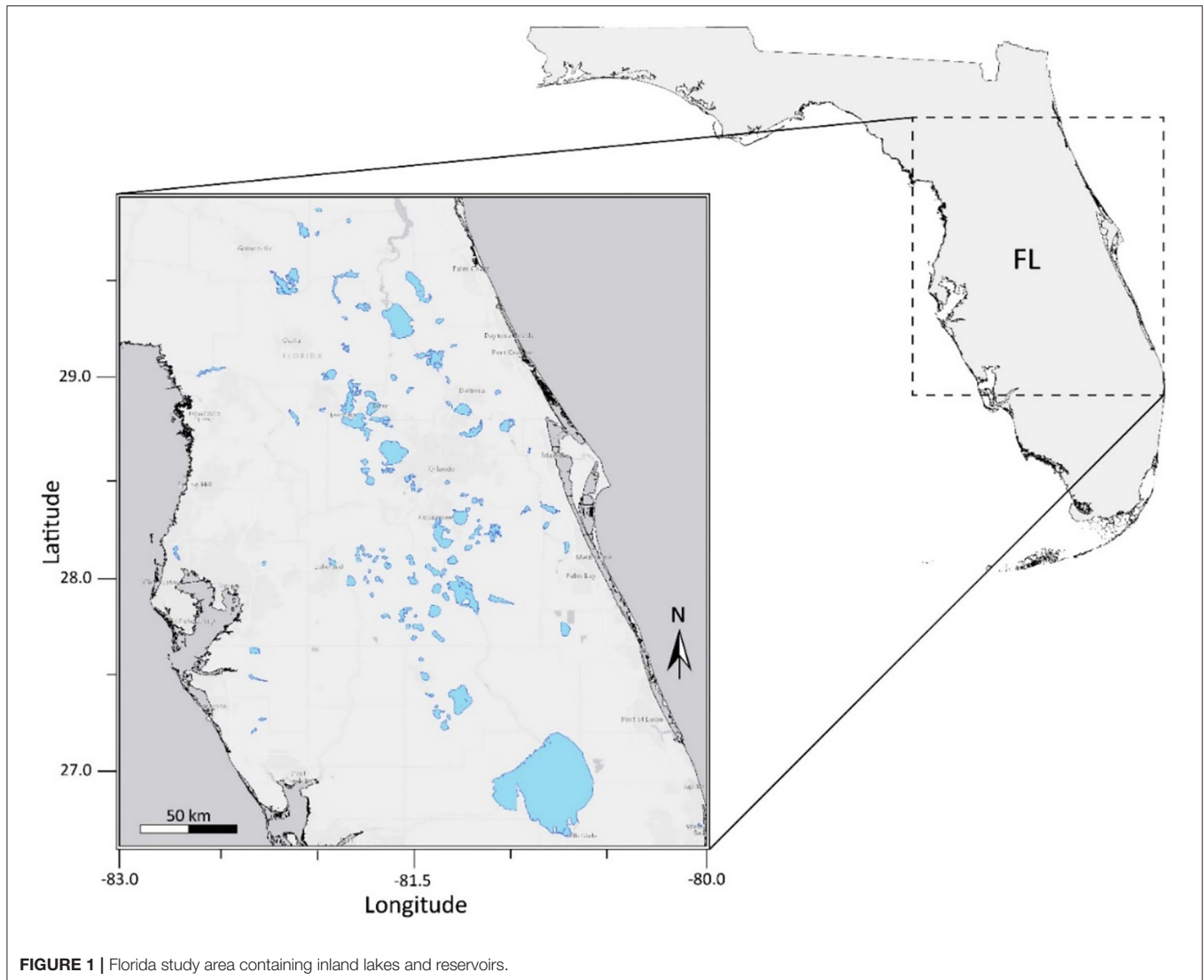
FIGURE 1 | Florida study area containing inland lakes and reservoirs.

included in each water body ranged from 3 to 14,499. The mean and median number of pixels in each water body were 228 and 23, respectively. The number of pixels in each water body considered in the analysis are provided in **Supplemental Material 2**. Waterbodies classified as intermittent, estuarine, rivers, streams, or waterbodies with a surface area <27 hectares are considered "unresolvable water" and thus QA flagged based on NLA criteria. Waterbodies with missing data were excluded from statistical analysis. The OLCI satellite scene extent for our study area is located within the NLA Coastal Plain (CPL) ecoregion. According to findings from the 2012 NLA report, 23% of lakes within the CPL ecoregion pose a high risk to exposure of cyanobacteria and cyanotoxins (US EPA, 2009).

Satellite-Derived Cyanobacteria Data Acquisition and Preparation

Satellite-derived cyanobacteria abundance data was obtained from Sentinel-3 imagery from May 2016 through May 2019 ($n = 217$ weeks). Standard OLCI Level-1B data were obtained

from the NASA Ocean Biology Processing Group (https://oceandata.sci.gsfc.nasa.gov). Sentinel-3 OLCI Florida granules were in the time window of 15:20–16:00 with a frame along track coordinate of 2,520. The cyanobacteria index (CI) was calculated using a spectral shape curvature method, as originally described in Wynne et al. (2008), updated in Lunetta et al. (2015), and algorithm progression fully detailed in Coffer et al. (2020). The CI-Multi value is then converted to cyanobacteria abundance (cells/mL) as described in Wynne et al. (2010). This algorithm was previously validated across Florida watersheds (Lunetta et al., 2015; Tomlinson et al., 2016). Composite images of the maximum cyanobacteria abundance at each resolvable satellite pixel were obtained from the individual scenes within sequential 7-day (1 week) periods. Weekly maximum cyanobacteria abundance was used to minimize the effect of clouds or wind that might otherwise reduce detection of the bloom. Many cyanoHAB genera such as *Microcystis*, *Aphanizomenon*, and *Dolichospermum* have buoyancy control and will typically float to the surface in the day in the absence

of strong winds (Visser et al., 2015). Wind can mix the blooms into the water column, diluting the surface concentration seen by the satellite (Wynne et al., 2010). Mishra et al. (2005) found satellite measures in the red region, similar to the CI-multi, typically only penetrate to a depth of 2 m or less. The OLCI sensor was selected because of public availability of data, spectral range to support deriving cyanobacteria concentrations, and 2–3 day repeat cycle (Urquhart and Schaeffer, 2020). For each water body, the satellite-derived cyanobacterial counts were averaged spatially each week to obtain one summary observation. This 1-week averaging and spatial aggregation was necessary due to cloud cover and satellite repeat cycles, though it results in a loss of information relative to using all available cyanobacterial abundance measurements without averaging and may cause short-term blooms to be missed.

Covariates

We limited covariates that were relevant to cyanobacterial growth and were readily available at the state and U.S. national scale on a weekly basis for operational forecasting. Land use and nutrients are not readily available on a weekly time frame for the continental US. In addition, nutrient transport and availability is complex and system specific, with data lacking for most systems. Wind speed and direction were not included as it was not available for the entire OLCI time period during our analysis.

We retrieved and processed raster and GIS datasets into a set of covariates for the state of Florida at 300 m × 300 m resolution. These data were identified in our review of the cyanobacterial modeling literature as environmental determinants likely to be associated with cyanobacteria bloom occurrence including ambient temperature, surface water temperature, precipitation, and static geomorphic lake conditions (**Table 1**). All data were obtained from public sources. We scaled all covariates to the same resolution as our satellite imagery in order to exclude measures along the land and water interface, because we masked our lake imagery with a 300 m buffer to exclude mixed pixel contamination. Landsat surface water temperature (WTEMP) was upscaled from 30 to 300 m, while PRISM precipitation (PRECIP) and ambient air temperature (ATEMP) were downscaled from 4 km to 300 m. Upscaling is a process that transfers information from local scale to large scale. Downscaling, conversely, transfers from large scale to local scale. Processes that are heterogenous at small scales become

homogenous at larger scales. All covariate data were temporally binned into weekly means by taking the arithmetic mean of 7 days within the numerical week (1 through 53) that bound each observation. Due to cloud cover and non-uniform scene acquisition, the monthly climatology of the Landsat Analysis Ready Data (ARD) surface water temperature product (Cook et al., 2014; Schaeffer et al., 2018) was used when the corresponding weekly data were not available. Spatial geographic coordinates were represented in kilometers using the Albers equal-area conic projection. Parameter-elevation Regressions on Independent Slopes Model (PRISM) mean, maximum, and minimum air temperatures and precipitation for North America was downloaded from the Oregon State PRISM Climate Group (PRISM Climate Group, 2004) through June 2019. Landsat ARD surface temperature was downloaded through USGS Earth Explorer through May 2019 in horizontal tiles 25, 26, and 27 and vertical tiles 16, 17, and 18.

All variables were averaged over all resolvable pixels within each water body each week, resulting in one summary observation per week for each water body. To aid in interpretation of relative variable influences on the response, predictor variables were centered and scaled by subtracting the mean and dividing by the standard deviation, transforming them into Z-scores. On this scale, a value of zero represents the mean, the units are standard deviations from the mean, and the value of the variable's coefficient in the model represents the effect in log-odds of a one standard deviation increase.

To identify the set of predictor variables most appropriate for modeling inland cyanobacterial bloom presence in Florida, we used non-spatial generalized linear models (**Supplementary Formula 4**), implemented in the R *glmulti* package (Calcagno and de Mazancourt, 2010), confirming the choice using stepwise-selection implemented in the R *stepAIC* package (Ripley, 2002). In this way, we constructed all possible non-redundant models with every combination of covariates. The best model was determined using Akaike's information criterion (AIC, **Supplementary Formula 6**), which is an estimator of out-of-sample prediction error and is used for model comparison (Akaike, 1998). See **Supplementary Material** for a discussion of the use of AIC. This step eliminated two variables, Maximum Lake Depth and Estimated Lake Volume. The best model, with the resulting fixed effect variables, was then used in the Bayesian spatio-temporal model.

TABLE 1 | Summary information for environmental datasets used for covariate selection.

Source	Variable name	Abbreviation	Units	Resolution
PRISM (Daly et al., 1997)	Air temperature	ATEMP	Degrees celsius (°C)	4 km; daily
	Precipitation	PRECIP	Millimeters rainfall (mm)	13.8 km; hourly
Landsat Analysis Ready Data (ARD) (Cook et al., 2014)	Surface water temperature	WTEMP	Degrees celsius (°C)	30 m; 16 and 8-day
lakeMorpho R-package (Hollister and Stachelek, 2017)	Surface area	Area	Square meters (m^2)	Static
	Mean lake depth	dMean	Meters (m)	Static

Hierarchical Bayesian Model Specification and Accuracy

We employed a hierarchical Bayesian spatio-temporal model to estimate exceedance probability over the WHO high-risk threshold with respect to environmental predictors. The response variable was the presence or absence of a high-risk cyanobacteria bloom, defined as a waterbody-wide average cell count above 100,000 cells/mL. Bayesian parameter estimates and prediction in the form of marginal posterior probability distributions were obtained via the R-INLA approach. For this study, weakly informative penalized-complexity priors were generated for all regression coefficients (fixed-effect parameters) and hyperparameters, allowing our large number of observations ($n = 11,096$) to inform the posterior distributions (Simpson et al., 2017). For the temporal component, we used a first order temporal autoregressive process (AR1, **Supplementary Formula 5**), which models the effect of time on bloom probability in each location as a function of the concentration in the previous week at that location plus an error term (Potzelberger, 1990). Spatial covariance was addressed using the INLA SPDE approach (Lindgren et al., 2011). The spatial effect represents residual error that can be attributed to location and may reflect the influence of an unmeasured or unmeasurable predictor that varies in space. We used a binomial logistic spatiotemporal model to predict harmful algal blooms (Formula 1).

$$logit\left(y_{st}\right) = \beta_1 X_1 + \ldots + \beta_n X_n + u_s + \mu_t$$

Formula 1. Generalized model structure for a binomial logistic model with spatial and temporal components.

y_{st} is the odds in favor of harmful bloom exceedance at location s and week t, $\beta_1, \ldots, \beta_n$ are the n regression coefficients, $X_1, \ldots, X_n$ are the n fixed independent variables, u_s is the value of the spatial random effect at location s, and μ_t is the value of the AR1 temporal random effect at time t. The *logit* is the logarithm of the odds in favor of an event and is also referred to as *log odds*. The β_n coefficients are in the logit or log-odds scale, which can be converted to probabilities as shown in Formula (2).

$$P\left(event\right) = \frac{e^{log-odds}}{1 + e^{log-odds}}$$

Formula 2. Converting log-odds to probability.

The odds were converted to probabilities of harmful algal bloom exceedance for model evaluation and predictive performance evaluation. A comprehensive mathematical overview of INLA SPDE can be found elsewhere (Rue et al., 2009; Lindgren and Rue, 2015). More detailed technical explanations of INLA SPDE applied to ecological and epidemiological modeling are available (Cosandey-Godin et al., 2014; Khana et al., 2018; Myer and Johnston, 2018). See **Supplementary Material** for a general discussion and explanation of the statistical methods used in this study along with definitions and further equations.

The utility of the spatial and temporal effects was assessed by fitting the model with and without random effects. The models were compared using the Deviance Information Criterion (DIC,

Supplementary Formula 7) to determine if the inclusion of the spatial and temporal effects improved model fit. Model performance was evaluated through holdout cross-validation in which the dataset was divided into three compartments: 80% of the data was randomly selected for training, 20% of the data was held out for validation, and the most recent week of data available at the time of the study was obtained for prediction. The model was created based on the training data only, and then model predictive power was assessed on the validation and prediction datasets. Model predictive power was determined by calculating the Area Under Curve (AUC), and by evaluating the sensitivity (true positive rate, **Supplementary Formula 8**), and specificity (true negative rate, **Supplementary Formula 9**) in predicting the holdout validation and prediction datasets after optimizing the logistic cutoff. Cutoff optimization, which chooses a value for the logistic predictor above which we consider a bloom prediction as positive, was performed by maximizing Youden's index on the validation dataset. Youden's index, sometimes referred to as the J-statistic, is defined in Formula (3).

$$J = sensitivity + specificity - 1$$

Formula 3. Youden's Index, or the J-statistic. Maximizing the value of this statistic provides the maximum overall accuracy for a binomial predictor.

Youden optimization attempts to find the cutoff at which sensitivity and specificity are balanced and at a maximum. All statistical analyses were conducted using R 3.3.4 on a compute cluster with 128 nodes and 4,096 cores.

RESULTS

The likelihood of high-risk cyanoHAB presence was estimated for 103 lakes and waterbodies in Florida. The number of high-risk blooms in individual lakes across Florida ranged from 0 to 146 weeks from 2016 to 2019 (**Figure 2**). During the study period, 22.5% of total bloom weeks were classified as high risk ($n = 3,149$ bloom weeks). Average waterbody-wide cyanobacterial concentration in lakes classified as high-risk was 376,504 cells/mL, more than three times the WHO "high" threshold of 100,000 cells/mL.

Model Selection

The model fit of our four candidate models was evaluated using the Deviance Information Criterion (DIC), with lower DIC values indicating better fit (**Table 2**). All four models included the same fixed effect predictors. The first model (M1), a non-hierarchical baseline model that did not incorporate random effects nor correlation features, exhibited a DIC of 14,195 and a ∼7 s computation time. The addition of a temporal effect (M2) improved the DIC value and only slightly increased the computation time. The addition of a spatial component (M3) dramatically improved the DIC value, but approximately quadrupled computation time. The addition of the spatial and temporal autocorrelation structure (M4) further improved the DIC value but increased the computation time to ∼53 s. We concluded that computational time was not an issue of
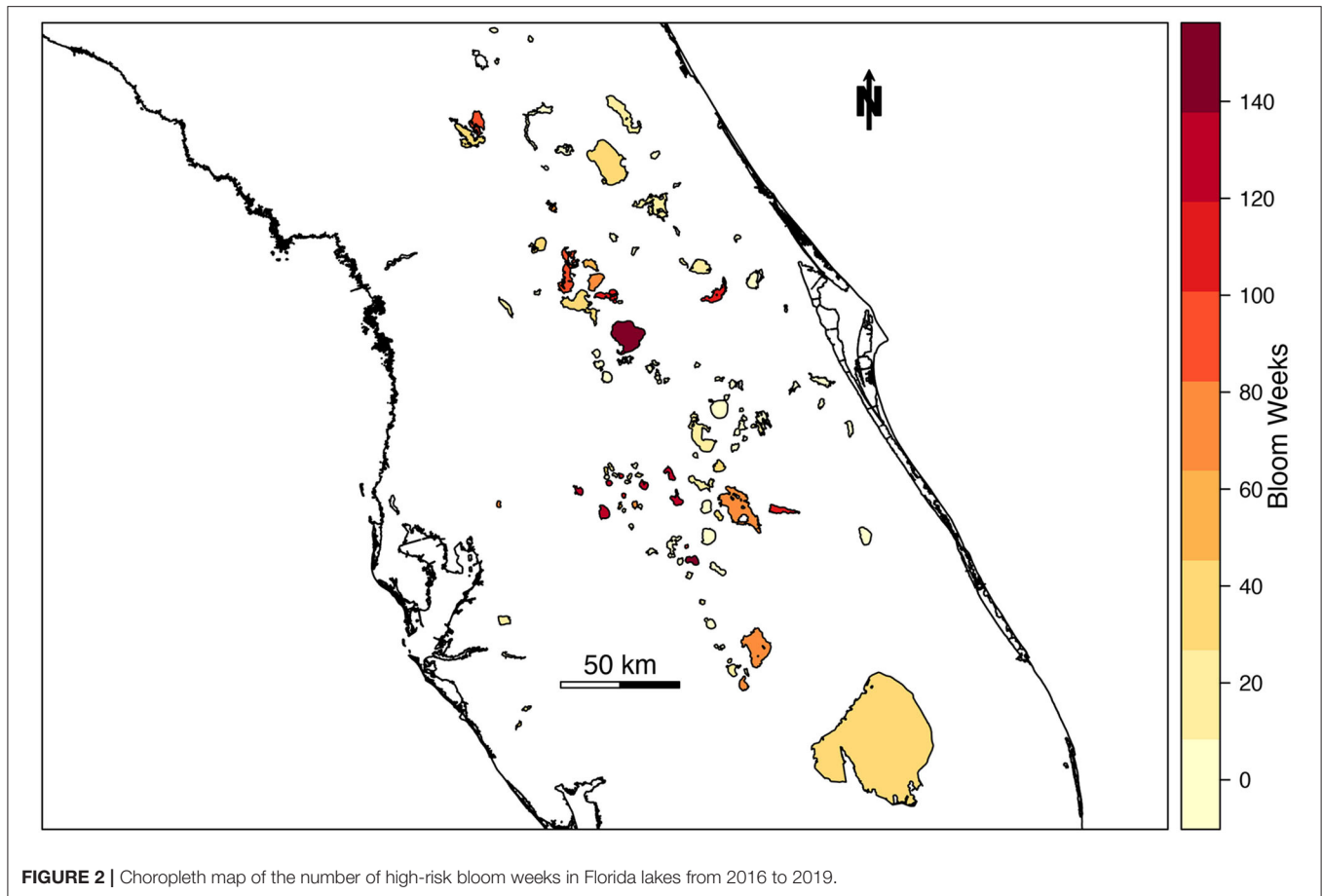
FIGURE 2 | Choropleth map of the number of high-risk bloom weeks in Florida lakes from 2016 to 2019.

TABLE 2 | Models considered with corresponding performance information.

Model	Model description	DIC	Computation time (s)
M1	Non-hierarchical model	21,381	~7
M2	Temporal component only	21,294	~10
M3	Spatial component only	7,758	~45
M4	Spatial + temporal model	7,601	~53

concern with this model, due to the favorable computational characteristics of the INLA SPDE method. Thus, the M4 hierarchical model with a full set of time varying, fixed effect covariates and spatio-temporal correlation effects provided the best model fit and was selected as the best model with subsequent results presented as follows.

Fixed Effect Covariates

The mean posterior coefficients of the fixed effect covariates are presented in log-odds and signify the estimated response to a one standard deviation change in the predictor variable when all other variables are held constant. For the purposes of determining statistical importance, we utilized an alpha level of 0.05 and considered a variable with a 95% credible

interval that did not encompass zero to have an important effect on the response. Using stepwise-selection and the resulting AIC values, ambient air temperature, surface water temperature, precipitation, lake area, and mean lake depth were selected as fixed effect predictors in the bloom estimation models (**Supplementary Table 1**). Fixed effect coefficients and 95% Bayesian credible intervals for the covariates included in the full spatio-temporal model (M4) are provided in **Table 3**. Posterior distributions of these fixed effect variables are presented as log-odds of the scaled covariate variables. As expected, a significant positive association is observed between surface water temperature (WTEMP) and high-risk bloom presence in Florida. For a one standard deviation increase in surface water temperature (6.23°C), the expected change in high-risk bloom log odds is 0.17 (or 1.18 times greater odds). Further, mean lake depth (DMEAN) exhibited a significant positive association with high-risk bloom presence, with an increase in one standard deviation (0.75 m) leading to an expected increase in log odds of 2.70 (or 14.88 times greater odds). Ambient air temperature (ATEMP) had a significant and negative effect on the likelihood of high-risk bloom presence in Florida, with a one standard deviation increase in ambient air temperature (4.79°C) resulting in change in high-risk bloom log odds of −0.23 (or 0.79 times lower odds).

TABLE 3 | Posterior estimates (mean, St. Dev., quantiles) for fixed effects.

Variable	Mean posterior	St. Dev.	0.025	0.5	0.975
ATEMP	−0.23	0.07	−0.37	−0.23	−0.08
WTEMP	0.17	0.05	0.08	0.17	0.26
PRECIP	−0.01	0.04	−0.09	−0.01	0.07
AREA	−0.11	0.66	−1.40	−0.11	1.19
DMEAN	2.70	0.52	1.68	2.70	3.72

TABLE 4 | Posterior estimates (mean, St. Dev., quantiles) for random effects.

Parameter (random effect)	Posterior mean	St. Dev.	0.025	0.5	0.975
Temporal variance	0.19	0.09	0.05	0.15	0.62
Spatial variance	38.51	7.85	25.63	37.61	56.35
Spatial correlation range ρ	16.76	3.44	10.72	16.54	24.14
AR (1) parameter α	0.90	0.09	0.68	0.92	0.99

Random Effects and Hyperparameters

The temporal model component (**Figure 3**) shows how the temporal effect on the odds of a high-risk bloom event vary throughout the year. This effect represents residual error that can be attributed to time and can be interpreted as the influence in the model of an unmeasured or unmeasurable predictor that varies in time. The log odds of high-risk bloom presence were higher from spring (week 15, approximately April 6) to late summer (week 35, approximately August 17) all other variables held constant. The posterior mean of the AR (1) parameter indicated an autocorrelation effect of 0.90 (95% CI = 0.68:0.99), which indicates that the temporal effect depends strongly on previous values and does not change quickly throughout the year (**Table 4**).

The spatial random effect (**Figure 4**) indicates that there is a significant amount of spatial variation in the mean concentration of cyanoHABs in lakes across Florida. The spatial random effect represents residual error that can be attributed to location and may reflect the influence of an unmeasured or unmeasurable predictor that varies in space. The posterior estimates (mean, standard deviation, 95% CI) for the random hyperparameters are collected in **Table 4**. The variance of the spatial effect showed a wide posterior distribution (95% CI = 25:56), indicating that the variability in bloom odds attributable to location is high. The posterior mean of the spatial correlation range (the distance at which spatial correlation declines to ∼0.1) was 16.8 km with a standard deviation of 3.4 km. This range indicates the approximate distance between lakes within which the odds of an algal bloom can be considered correlated.

Model Performance

In order to test the performance of the final spatio-temporal model (M4), we used holdout data to determine if M4 correctly estimates the observed data in space and time. Twenty percent (20%) of the data were randomly set aside as a holdout validation

TABLE 5 | Results of the statistical evaluation metrics obtained using a cutoff point of 0.365.

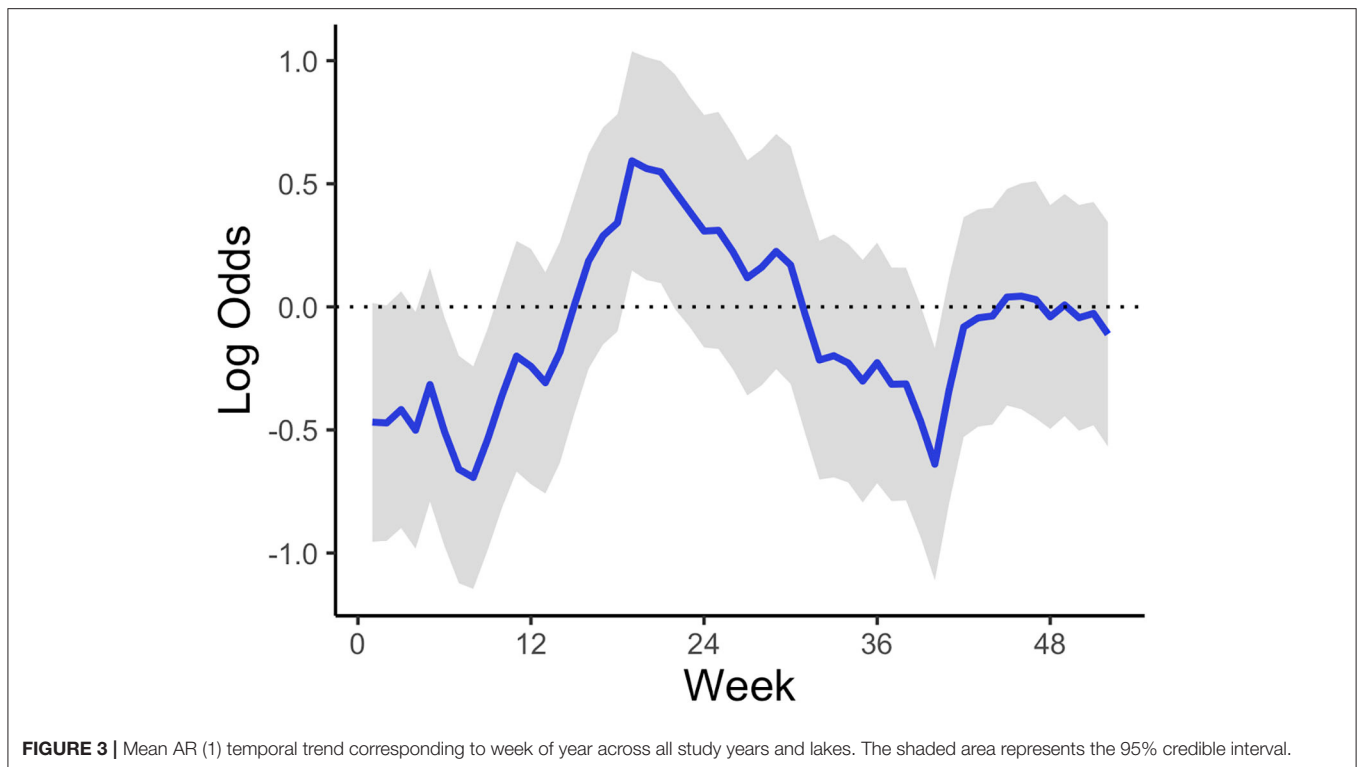| Metric	Validation dataset	Prediction dataset
AUC	0.95	0.89
Sensitivity	0.88	0.82
Specificity	0.88	0.82
Accuracy	0.88	0.82

dataset ($n = 2{,}775$ observations). Instead of using an arbitrary cutoff threshold to assign the predicted high-risk bloom value as positive or negative, we calculated a Youden-optimized cutoff point of 0.365. Traditional logistic model evaluation statistics including Area Under Curve (AUC), sensitivity (true positive rate), and specificity (true negative rate) provide insight into the model performance for both holdout dataset scenarios (**Table 5**). An AUC > 0.5 indicates that the model predicts better than chance alone. The resulting AUC between observations and predictions of the holdout validation dataset was 0.95, sensitivity was 0.88, specificity was 0.88, and accuracy was 0.88.

Prediction

For practical application reasons, we were also interested in assessing the model's capability to predict or forecast future bloom presence for a week in which the model was untrained. Predictive power of M4 was tested by reserving response and fixed effect data for the most recent week of available data at the time of analysis (May 27th through June 2nd, 2019) as a holdout prediction dataset ($n = 103$ observations). The resulting AUC between observations and predictions of the holdout prediction dataset is 0.89, sensitivity is 0.82, and specificity is 0.82 (**Table 5**). Our prediction objective was to get a probability of high-risk bloom exceedance, for a given week, using a threshold of 100,000 cyanobacteria cells/mL. As expected, higher exceedance probabilities are detected in lakes with >100 bloom weeks such as Lake Apopka located in central Florida (**Figure 5**).

DISCUSSION

Here we present a hierarchical Bayesian spatio-temporal modeling approach in R-INLA to estimate the likelihood of high-risk cyanoHABs in Florida inland waterbodies. Using DIC to evaluate model performance, the full spatio-temporal model (M4, **Table 2**) was selected as the best model and used to forecast the likelihood of bloom occurrence across Florida lakes for a week outside of the dataset, with AUC 0.93 (**Table 5**). The spatial random effect (**Figure 4**) identified residual error attributed to location that may reflect the influence of an unmeasured or unmeasurable predictor that varies across the landscape. The variance of the spatial effect also indicated the variability in cyanoHAB bloom odds for each lake is high. The distance at which spatial correlation declines below a meaningful threshold (0.1) indicates ∼17 km as the inter-lake distance within which the odds of cyanoHAB can be considered correlated. Because our model did not include any measures of nutrient input or

FIGURE 3 | Mean AR (1) temporal trend corresponding to week of year across all study years and lakes. The shaded area represents the 95% credible interval.
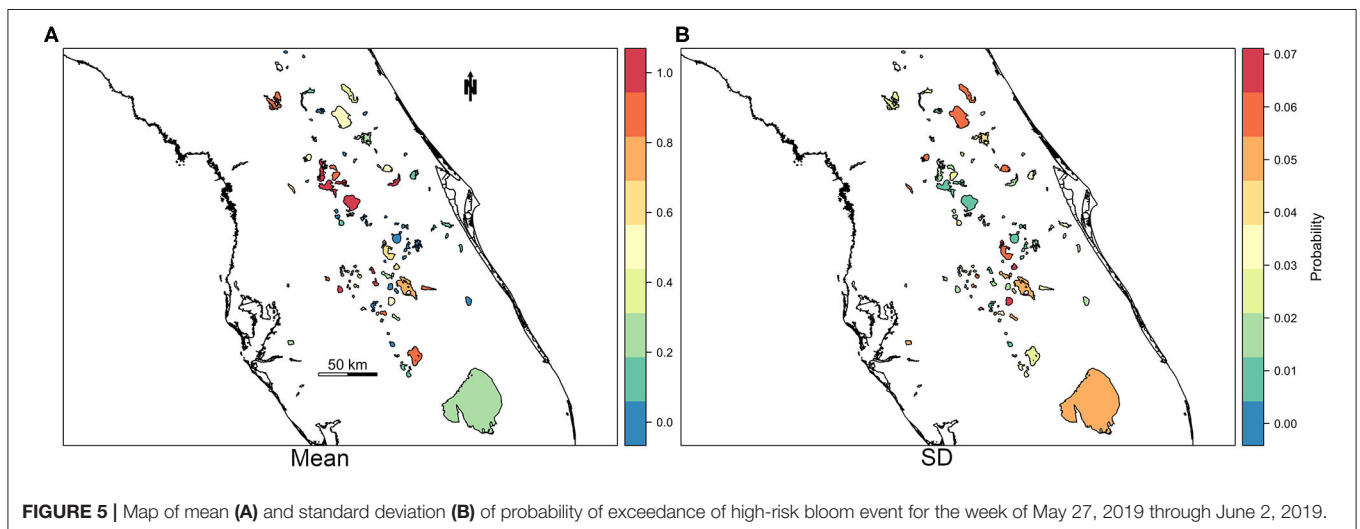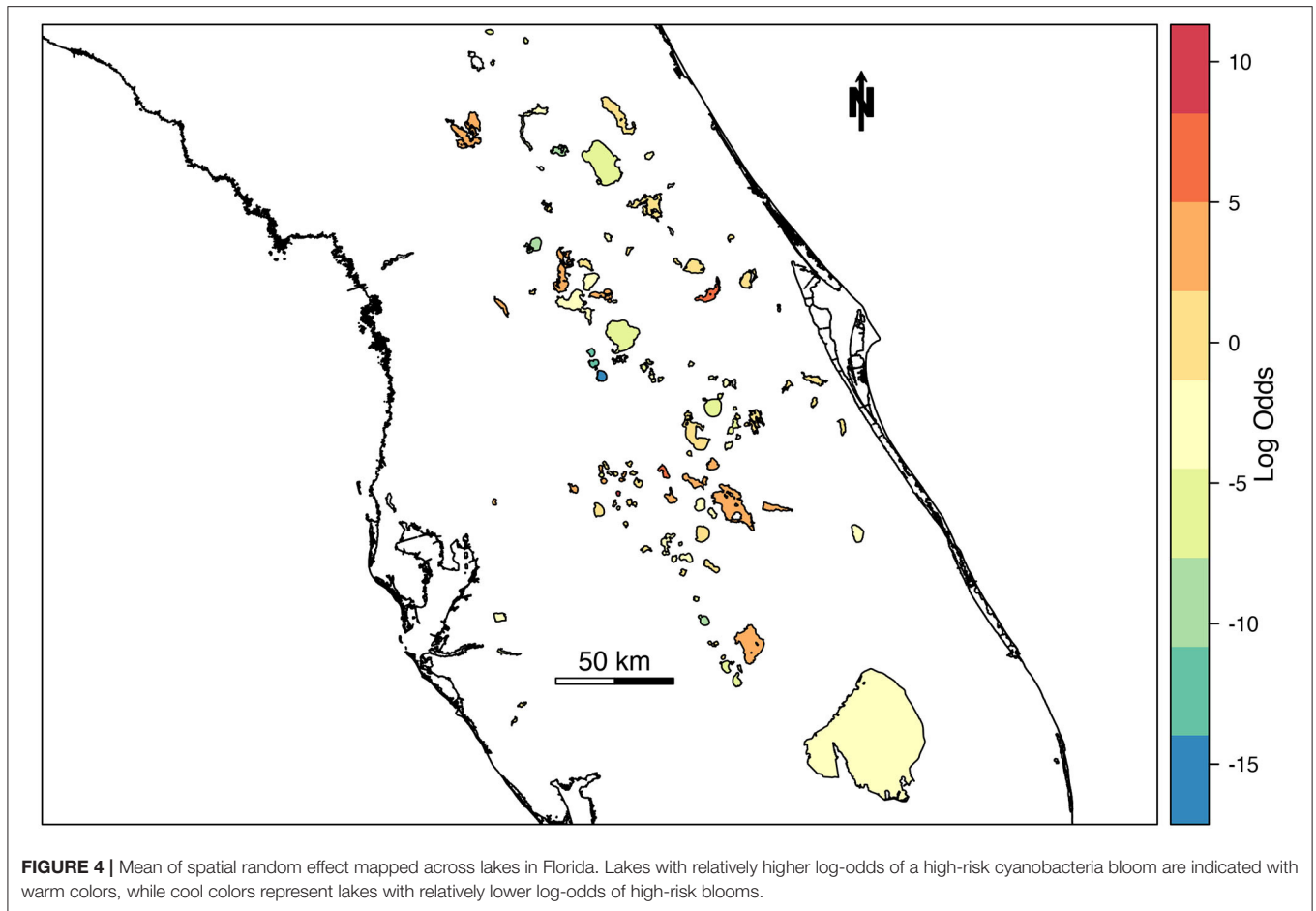
land use, we believe that it is likely the spatial component of the model is related to land cover and its resultant effect on eutrophication. This landscape effect varies at long spatial scales and is known to have a significant effect on the likelihood of cyanoHABs (Doubek et al., 2015; Marion et al., 2017). Another related possibility is that the spatial correlation represents the distance between watersheds experiencing similar environmental stressors such as nutrient input, making the lakes fed by those watersheds more similar in terms of cyanoHAB risk.

Water surface temperature had a positive effect on cyanoHAB risk, an association that has been observed in many studies to date (Paerl and Huisman, 2008; Wynne et al., 2010; Kosten et al., 2012; Taranu et al., 2012; Beaulieu et al., 2013; Cha et al., 2017). Several mechanisms operate concurrently to the advantage of cyanobacteria at higher water temperatures. The optimal temperature for cyanobacterial growth and photosynthesis is above 20°C, with some species experiencing optimal growth at 30°C or higher (Konopka and Brock, 1978; Lürling et al., 2013; Giannuzzi et al., 2016; Melina Celeste et al., 2017). At high temperatures cyanobacteria have a competitive advantage over green algae, diatoms, and other phytoplankton which favors cyanobacteria dominance. However, some common species of cyanoHAB in the genus *Microcystis* have been found to produce fewer toxins at high temperatures which may mitigate health risks (Runnegar et al., 1983; Rapala et al., 1997; Melina Celeste et al., 2017), although these findings are disputed (Lehman et al., 2008; Davis et al., 2009). Bloom-forming cyanobacteria contain many cylindrical gas vacuoles that impart buoyancy (Walsby, 1977). This buoyancy causes the cells to float to the surface

where they are exposed to more light and can outcompete sinking phytoplankton, and there is evidence that increased water temperature improves the ability of cyanobacteria to stay afloat (Kromkamp et al., 1988; Huisman et al., 2004).

Contrary to our expectations, mean lake depth (DMEAN) was a strongly positive correlate of cyanoHAB bloom odds. Our prior belief was that a shallower lake would have a diminished capacity to buffer changes in nutrient input, leading to a heightened risk for HAB blooms. However, in our data this does not seem to be the case. The lakes in our dataset, and Florida lakes in general, are shallow. The average lake in our study had a mean lake depth of only 0.75 m, and the deepest lake had a mean depth of 3.68 m. To contextualize the large positive coefficient of the lake depth effect, we note that an increase in mean depth of 0.75 m, the unit of increase in our model, represents a doubling in depth for the average lake. Satellite penetration in the red spectrum is 2 m or less in oligotrophic waters (Mishra et al., 2005). Given the focus of our model on cyanoHABs at the >100,000 cells/ml threshold, the penetration depth is likely only a few centimeters. It is possible that optically shallow water could cause bottom reflectance or benthic cyanobacteria could cause artifacts. However, Coffer et al. (2020) found the satellite derived phenology of cyanobacteria in Florida followed well-accepted ecological trends and a few lakes that had a peak of biomass in the winter were supported by independent field observations.

Classic models of lake eutrophication posit a strong effect of mixing between the epilimnion and hypolimnion (O'Melia, 1972; Imboden, 1974). Deep lakes are considered to have a greater inertia with respect to changing nutrient conditions;

FIGURE 4 | Mean of spatial random effect mapped across lakes in Florida. Lakes with relatively higher log-odds of a high-risk cyanobacteria bloom are indicated with warm colors, while cool colors represent lakes with relatively lower log-odds of high-risk blooms.

FIGURE 5 | Map of mean **(A)** and standard deviation **(B)** of probability of exceedance of high-risk bloom event for the week of May 27, 2019 through June 2, 2019.

Imboden called this effect a "memory" and found that deep lakes take longer to change from oligotrophy to eutrophy and vice-versa. Shallow lakes do not form the distinctive temperature-delineated layers of a stratified lake and are polymictic: experiencing constant mixing by wind and temperature that leaves the water at a generally homogenous state. This mixing leads to resuspension of nutrients from sediment a higher nutrient load in the water column (Taranu and Gregory-Eaves, 2008; Liu et al., 2012). The lakes in our study are not deep enough to stratify and should be considered polymictic.

Because all of our lakes are shallow, nutrients cannot sequester within the hypolimnion and sediment and the association between depth and oligotrophy found in stratified lakes does not apply.

In a study of cyanobacteria growth in shallow lakes (<6 m), Kosten et al. found similar results to our study, with water temperature predicting a higher proportion of cyanobacteria among the phytoplankton community (Kosten et al., 2012). A study of the National Lake Assessment dataset, containing 1,147 lakes and spanning the continental U.S., found that when lakes were divided into shallow (<6 m) or deep, their models were less able to predict cyanobacterial concentrations in the shallow lakes (Beaulieu et al., 2013). Associations of HAB volume with nitrogen content and water temperature were consistent across all lakes, but there was a decrease of two-thirds in variation explained when lakes were shallow. A national-scale study in the United Kingdom that investigated both shallow and deep lakes suggested that water residence time, a correlate of depth, was a positive predictor of HAB biovolume, especially in lakes with significant opacity (Carvalho et al., 2011). An in-depth study of a shallow (mean depth 1.2 m), warm, polymictic lake in the Mediterranean that was substantially similar to the lakes in our study found that water residence time was a strong predictor of harmful algae concentration (Romo et al., 2013). We propose that in our study, mean lake depth is acting as a correlate of water residence time, explaining its positive association with HAB blooms.

Air temperature had a negative effect on cyanoHAB risk, a result that seems counterintuitive when considered alongside the positive effect of water surface temperature. Our model suggests that cyanoHAB risk is heightened when air temperature is lower relative to surface water temperature, a condition that can occur in shallow lakes during the transition from summer to fall, or when lake temperature is artificially increased by impervious surface runoff or industrial input (Sabouri et al., 2013). In our dataset, the peak mean water temperature was significantly higher than the peak mean air temperature observed (36.5 and 30.3°C, respectively). The higher specific heat of water is responsible for this effect, with lakes serving as a heat sink that can persist after air temperature has cooled.

The predictive accuracy of our model on a holdout dataset was high, with ∼82% of held-out observations for the week of May 27, 2019 predicted correctly. Accuracy was also good in our validation dataset, with 88% of observations correctly classified. The accuracy of our simple model is encouraging but has a few caveats. This model is not expected to remain accurate for more than 2 weeks into the future, because wide spatial coverage meteorological estimates become less reliable beyond that time scale. Additionally, the autocorrelation parameter, or α, was 0.9, indicating that the change in the temporal component of risk of a cyanoHAB is strongly related to the past week's bloom conditions. While this temporal component was not as significant a contribution to model fit as the spatial component that considered lake location (**Table 2**), we caution that interpretation of our model's accuracy should consider that due to system inertia, a fairly accurate prediction of near-future bloom conditions can be made by simply extrapolating that current conditions will continue unchanged.

Although we have developed a reliable model for forecasting cyanoHAB odds for lakes in central Florida lakes for a week of interest outside of the modeled dataset, there are several limitations of this study. While remote sensing provides a continuous cyanoHAB data source, the relatively coarse 300 m sensor resolution, presence of cloud cover, and occasional missing data due to waterbody misclassification, limits our predictive ability in smaller inland waterbodies. Additional satellite limitations are reviewed in Coffer et al. (2020) and Clark et al. (2017). Future work could assess the applicability of higher resolution sensors such as the European Commission's Copernicus Sentinel-2. Our covariates were not all available at the same spatial scale and had to be resampled to the same 300 m resolution as our lake imagery. As a result, lake areas with highly irregular or elongated narrow reaches may be underrepresented vs. lake areas with broader widths. It is possible that CyanoHAB was present in a specific part of the lake, but satellite data didn't capture it because of land proximity. This study incorporated several environmental covariates known to be associated with cyanoHAB occurrence; however, our results showed a strong spatially varying effect, the cause of which is undetermined in the present analysis. Future studies could explore additional environmental determinants such as nutrient loadings. Nutrients certainly play a role related to cyanoHABs and nutrient eutrophication is a "wicked" problem as defined by Thornton et al. (2013), meaning that the issue is convoluted and precludes a simple solution. However, modeling nutrient transport and availability is complex and system specific, with model validation data lacking for most systems. This complexity, specificity and lack of data limits the ability to scale such models across spatial and temporal scales as described in Soranno et al. (2015). Therefore, we limited the model to data that would be readily available and most relevant to cyanoHAB biomass.

In the present study we applied R-INLA for the purpose of mapping/predicting exceedance probability of high-risk cyanoHABs at the level of a whole lake. Future work will apply the spatial-temporal modeling approach at the sub-lake level, particularly in large systems with greater geographic bloom variability such as Lake Okeechobee in South Florida.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: US EPA Environmental Dataset Gateway, https://edg.epa.gov/metadata/catalog/main/home.page. DOI: 10.23719/1518919.

AUTHOR CONTRIBUTIONS

EU and BS developed the project, gathered data, and conducted the initial round of analysis including determining algae levels from remote sensing data. MM and JJ provided editing and technical advice during the initial analysis round. MM conducted

the final analysis. All authors contributed equally to the writing of the final manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fenvs.2020.581091/full#supplementary-material

REFERENCES

Akaike, H. (1998). "A Bayesian analysis of the minimum AIC procedure," in *Selected Papers of Hirotugu Akaike*, eds E. Parzen, K. Tanabe, and G. Kitagawa (New York, NY: Springer), 275–280.

Beaulieu, M., Pick, F., and Gregory-Eaves, I. (2013). Nutrients and water temperature are significant predictors of cyanobacterial biomass in a 1147 lakes data set. *Limnol. Oceanogr.* 58, 1736–1746. doi: 10.4319/lo.2013.58.5.1736

Beaulieu, M., Pick, F., Palmer, M., Watson, S., Winter, J., Zurawell, R., et al. (2014). Comparing predictive cyanobacterial models from temperate regions. *Can. J. Fisher. Aquat. Sci.* 71, 1830–1839. doi: 10.1139/cjfas-2014-0168

Boyer, G., and FitzPatrick, K. (2016). *Water Quality Final Report*. Available online at: https://www.martin.fl.us/martin- county- services/water- quality- reports (accessed June 14, 2016).

Burns, J. (2008). "Toxic cyanobacteria in Florida waters," in *Cyanobacterial Harmful Algal Blooms: State of the Science and Research Needs*, ed H. K. Hudnell (New York, NY: Springer), 127–137.

Burns, J., Williams, C. D., and Chapman, A. D. (2002). Cyanobacteria and their toxins in Florida surface waters. *Paper presented at the Health Effects of Exposure to Cyanobacteria Toxins: State of the Science* (Sarasota, FL).

Calcagno, V., and de Mazancourt, C. (2010). glmulti: an R package for easy automated model selection with (generalized) linear models. *J. Stat. Softw.* 34, 1–29. doi: 10.18637/jss.v034.i12

Cameletti, M., Lindgren, F., Simpson, D., and Rue, H. (2013). Spatio-temporal modeling of particulate matter concentration through the SPDE approach. *AStA Adv. Stat. Anal.* 97, 109–131. doi: 10.1007/s10182-012-0196-3

Carmichael, W. (1992). *A Status Report of Planktonic Cyanobacteria (Blue- Green Algae) and Their Toxins*. Cincinnati, OH: United States Environmental Protection Agency.

Carvalho, L., Miller, C. A., Scott, E. M., Codd, G. A., Davies, P. S., and Tyler, A. N. (2011). Cyanobacterial blooms: statistical models describing risk factors for national-scale lake assessment and lake management. *Sci. Total Environ.* 409, 5353–5358. doi: 10.1016/j.scitotenv.2011.09.030

Cha, Y., Cho, K. H., Lee, H., Kang, T., and Kim, J. H. (2017). The relative importance of water temperature and residence time in predicting cyanobacteria abundance in regulated rivers. *Water Res.* 124, 11–19. doi: 10.1016/j.watres.2017.07.040

Chapman, A. D., and Schelske, C. L. (1997). Recent appearance of cylindrospemopsis (cyanobacteuia) in five hypereutropic florida lakes. *J. Phycol.* 33, 191–195. doi: 10.1111/j.0022-3646.1997.00191.x

Chorus, E. I., and Bartram, J. (1999). *Toxic Cyanobacteria in Water: A Guide to Their Public Health Consequences, Monitoring and Management*. Geneva: World Health Organization.

Clark, J. M., Schaeffer, B. A., Darling, J. A., Urquhart, E. A., Johnston, J. M., Ignatius, A. R., et al. (2017). Satellite monitoring of cyanobacterial harmful algal bloom frequency in recreational waters and drinking water sources. *Ecol. Indic.* 80, 84–95. doi: 10.1016/j.ecolind.2017.04.046

Coffer, M. M., Schaeffer, B. A., Darling, J. A., Urquhart, E. A., and Salls, W. B. (2020). Quantifying national and regional cyanobacterial occurrence in US lakes using satellite remote sensing. *Ecol. Indic.* 111:105976. doi: 10.1016/j.ecolind.2019.105976

Cook, M., Schott, J., Mandel, J., and Raqueno, N. (2014). Development of an operational calibration methodology for the Landsat thermal data archive and initial testing of the atmospheric compensation component of a Land Surface Temperature (LST) product from the archive. *Remote Sens.* 6, 11244–11266. doi: 10.3390/rs61111244

Cosandey-Godin, A., Krainski, E. T., Worm, B., and Flemming, J. M. (2014). Applying Bayesian spatiotemporal models to fisheries bycatch in the Canadian Arctic. *Can. J. Fisher. Aquat. Sci.* 72, 186–197. doi: 10.1139/cjfas-2014-0159

Daly, C., Taylor, G., and Gibson, W. (1997). "The PRISM approach to mapping precipitation and temperature," in *Proc 10th AMS Conf. on Applied Climatology*. Reno, NV: American Meteorological Society.

Davis, T. W., Berry, D. L., Boyer, G. L., and Gobler, C. J. (2009). The effects of temperature and nutrients on the growth and dynamics of toxic and non-toxic strains of microcystis during cyanobacteria blooms. *Harmful Algae* 8, 715–725. doi: 10.1016/j.hal.2009.02.004

Doubek, J. P., Carey, C. C., and Cardinale, B. J. (2015). Anthropogenic land use is associated with N-fixing cyanobacterial dominance in lakes across the continental United States. *Aquat. Sci.* 77, 681–694. doi: 10.1007/s00027-015-0411-x

Downing, J. A., Watson, S. B., and McCauley, E. (2001). Predicting cyanobacteria dominance in lakes. *Can. J. Fisher. Aquat. Sci.* 58, 1905–1908. doi: 10.1139/f01-143

Florida Fish and Wildlife Conservation Commission (2017). *Cyanobacteria in Florida Waters*. Available online at: http://myfwc.com/research/redtide/general/cyanobacteria

Giannuzzi, L., Krock, B., Minaglia, M. C. C., Rosso, L., Houghton, C., Sedan, D., et al. (2016). Growth, toxin production, active oxygen species and catalase activity of *Microcystis aeruginosa* (*Cyanophyceae*) exposed to temperature stress. *Comp. Biochem. Physiol. Part C Toxicol. Pharmacol.* 189, 22–30. doi: 10.1016/j.cbpc.2016.07.001

Glibert, P. M., Harrison, J., Heil, C., and Seitzinger, S. (2006). Escalating worldwide use of urea – a global change contributing to coastal eutrophication. *Biogeochemistry* 77, 441–463. doi: 10.1007/s10533-005-3070-5

Havens, K. E., Phlips, E. J., Cichra, M. F., and Li, B.-l. (1998). Light availability as a possible regulator of cyanobacteria species composition in a shallow subtropical lake. *Freshw. Biol.* 39, 547–556. doi: 10.1046/j.1365-2427.1998.00308.x

Havens, K. E., and Steinman, A. D. (2015). Ecological responses of a large shallow lake (Okeechobee, Florida) to climate change and potential future hydrologic regimes. *Environ. Manage.* 55, 763–775. doi: 10.1007/s00267-013-0189-3

Hollister, J., and Stachelek, J. (2017). lakemorpho: Calculating lake morphometry metrics in R. *F1000Res.* 6:1718. doi: 10.12688/f1000research.12512.1

Huisman, J., Sharples, J., Stroom, J. M., Visser, P. M., Kardinaal, W. E. A., Verspagen, J. M. H., et al. (2004). Changes in turbulent mixing shift competition for light between phytoplankton species. *Ecology* 85, 2960–2970. doi: 10.1890/03-0763

Imboden, D. M. (1974). Phosphorus model of lake eutrophication. *Limnol. Oceanogr.* 19, 297–304. doi: 10.4319/lo.1974.19.2.0297

Khana, D., Rossen, L. M., Hedegaard, H., and Warner, M. (2018). A Bayesian spatial and temporal modeling approach to mapping geographic variation in mortality rates for subnational areas with R-INLA. *J. Data Sci.* 16, 147–182. doi: 10.6339/JDS.201801_16(1).0009

Konopka, A., and Brock, T. D. (1978). Effect of temperature on blue-green algae (cyanobacteria) in lake mendota. *Appl. Environ. Microbiol.* 36, 572–576. doi: 10.1128/AEM.36.4.572-576.1978

Kosten, S., Huszar, V. L. M., Bécares, E., Costa, L. S., van Donk, E., Hansson, L.-A., et al. (2012). Warmer climates boost cyanobacterial dominance in shallow lakes. *Glob. Chang. Biol.* 18, 118–126. doi: 10.1111/j.1365-2486.2011.02488.x

Kromkamp, J., Botterweg, J., and Mur, L. R. (1988). Buoyancy regulation in *Microcystis aeruginosa* grown at different temperatures. *FEMS Microbiol. Ecol.* 4, 231–237. doi: 10.1111/j.1574-6968.1988.tb02669.x

Lehman, P., Boyer, G., Satchwell, M., and Waller, S. (2008). The influence of environmental conditions on coffseasonal variation of microcystis abundance and microcystins concentration in San Francisco Estuary. *Hydrobiologia* 600, 187–204. doi: 10.1007/s10750-007-9231-x

Lindgren, F., and Rue, H. (2015). Bayesian spatial modelling with R-INLA. *J. Stat. Softw.* 63, 1–25. doi: 10.18637/jss.v063.i19

Lindgren, F., Rue, H., and Lindstrom, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *J. R. Statist. Soc B.* 73, 763–775. doi: 10.1111/j.1467-9868.2011.00777.x

Liu, W., Li, S., Bu, H., Zhang, Q., and Liu, G. (2012). Eutrophication in the Yunnan Plateau lakes: the influence of lake morphology, watershed land use, and socioeconomic factors. *Environ. Sci. Pollut. Res.* 19, 858–870. doi: 10.1007/s11356-011-0616-z

Lunetta, R. S., Schaeffer, B. A., Stumpf, R. P., Keith, D., Jacobs, S. A., and Murphy, M. S. (2015). Evaluation of cyanobacteria cell count detection derived from MERIS imagery across the eastern USA. *Remote Sens. Environ.* 157, 24–34. doi: 10.1016/j.rse.2014.06.008

Lürling, M., Eshetu, F., Faasen, E. J., Kosten, S., and Huszar, V. L. M. (2013). Comparison of cyanobacterial and green algal growth rates at different temperatures. *Freshw. Biol.* 58, 552–559. doi: 10.1111/j.1365-2427.2012.02866.x

Marion, J. W., Zhang, F., Cutting, D., and Lee, J. (2017). Associations between county-level land cover classes and cyanobacteria blooms in the United States. *Ecol. Eng.* 108, 556–563. doi: 10.1016/j.ecoleng.2017.07.032

Melina Celeste, C. M., Lorena, R., Jorge Oswaldo, A., Sandro, G., Daniela, S., Dario, A., et al. (2017). Mathematical modeling of *Microcystis aeruginosa* growth and [D-Leu1] microcystin-LR production in culture media at different temperatures. *Harmful Algae* 67, 13–25. doi: 10.1016/j.hal.2017.05.006

Michalak, A. M., Anderson, E. J., Beletsky, D., Boland, S., Bosch, N. S., Bridgeman, T. B., et al. (2013). Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions. *Proc. Natl. Acad. Sci. U.S.A.* 110, 6448–6452. doi: 10.1073/pnas.1216006110

Millie, D. F., Weckman, G. R., Fahnenstiel, G. L., Carrick, H. J., Ardjmand, E., Young, W. A., et al. (2014). Using artificial intelligence for CyanoHAB niche modeling: discovery and visualization of Microcystis–environmental associations within western Lake Erie. *Can. J. Fisher. Aquat. Sci.* 71, 1642–1654. doi: 10.1139/cjfas-2013-0654

Mishra, D. R., Narumalani, S., Rundquist, D., and Lawson, M. (2005). Characterizing the vertical diffuse attenuation coefficient for downwelling irradiance in coastal waters: implications for water penetration by high resolution satellite data. *ISPRS J. Photogramm. Remote Sens.* 60, 48–64. doi: 10.1016/j.isprsjprs.2005.09.003

Myer, M. H., and Johnston, J. M. (2018). Spatiotemporal Bayesian modeling of West Nile virus: identifying risk of infection in mosquitoes with local-scale predictors. *Sci. Total Environ.* 650, 2818–2829. doi: 10.1016/j.scitotenv.2018.09.397

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

Oehrle, S., Rodriguez-Matos, M., Cartamil, M., Zavala, C., and Rein, K. S. (2017). Toxin composition of the 2016 *Microcystis aeruginosa* bloom in the St. Lucie Estuary, Florida. *Toxicon* 138, 169–172. doi: 10.1016/j.toxicon.2017.09.005

O'Melia, C. R. (1972). An approach to the modeling of lakes. *Schweizerische Zeitschrift Hydrol.* 34, 1–33. doi: 10.1007/BF02502798

Paerl, H. W., and Huisman, J. (2008). Blooms like it hot. *Science* 320, 57–58. doi: 10.1126/science.1155398

Phlips, E. J., Bledsoe, E., Cichra, M. F., Badylak, S., and Frost, J. (2002). The distribution of potentially toxic cyanobacteria in Florida. *Paper presented at the Health Effects of Exposure to Cyanobacteria Toxins: State of the Science* (Sarasota, FL).

Potzelberger, K. (1990). A characterization of random-coefficient AR(1) models. *Scotchastic Process. Appl.* 34, 171–180. doi: 10.1016/0304-4149(90)90062-W

PRISM Climate Group (2004). Oregon State University. Available online at: http://prism.oregonstate.edu (accessed Febuary 4, 2004).

R Core Team (2015). *R: A Language and Environment for Statistical Computing.* Vienna. Available online at: https://www.R-project.org

Rapala, J., Sivonen, K., Lyra, C., and Niemela, S. I. (1997). Variation of microcystins, cyanobacterial hepatotoxins, in *Anabaena* spp. as a function of growth stimuli. *Appl. Environ. Microbiol.* 63, 2206–2212. doi: 10.1128/AEM.63.6.2206-2212.1997

Ripley, B. (2002). *Modern Applied Statistics With S.* New York, NY: Springer.

Romo, S., Soria, J., Fernandez, F., Ouahid, Y., and Baron-Sola, Á. (2013). Water residence time and the dynamics of toxic cyanobacteria. *Freshw. Biol.* 58, 513–522. doi: 10.1111/j.1365-2427.2012.02734.x

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B* 71, 319–392. doi: 10.1111/j.1467-9868.2008.00700.x

Rue, H., Riebler, A., Sørbye, S. H., Illian, J. B., Simpson, D. P., and Lindgren, F. K. (2017). Bayesian computing with INLA: a review. *Annu. Rev. Stat. Appl.* 4, 395–421. doi: 10.1146/annurev-statistics-054045

Runnegar, M. T. C., Falconer, I. R., Jackson, A. R. B., and McInnes, A. (1983). Toxin production by *Microcystis aeruginosa* cultures. *Toxicon* 21, 377–380. doi: 10.1016/0041-0101(83)90233-7

Sabouri, F., Gharabaghi, B., Mahboubi, A. A., and McBean, E. A. (2013). Impervious surfaces and sewer pipe effects on stormwater runoff temperature. *J. Hydrol.* 502, 10–17. doi: 10.1016/j.jhydrol.2013.08.016

Schaeffer, B. A., Iiames, J., Dwyer, J., Urquhart, E., Salls, W., Rover, J., et al. (2018). An initial validation of Landsat 5 and 7 derived surface water temperature for U.S. lakes, reservoirs, and estuaries. *Int. J. Remote Sens.* 39, 7789–7805. doi: 10.1080/01431161.2018.1471545

Simpson, D., Rue, H., Riebler, A., Martins, T. G., and Sørbye, S. H. (2017). Penalising model component complexity: a principled, practical approach to constructing priors. *Stat. Sci.* 32, 1–28. doi: 10.1214/16-STS576

Smayda, T. J. (1997). What is a bloom? A commentary. *Limnol. Oceanogr.* 42, 1132–1136. doi: 10.4319/lo.1997.42.5_part_2.1132

Soranno, P. A., Cheruvelil, K. S., Wagner, T., Webster, K. E., and Bremigan, M. T. (2015). Effects of land use on lake nutrients: the importance of scale, hydrologic connectivity, and region. *PLoS ONE.* 10:e0135454. doi: 10.1371/journal.pone.0135454

Taranu, Z. E., and Gregory-Eaves, I. (2008). Quantifying relationships among phosphorus, agriculture, and lake depth at an inter-regional scale. *Ecosystems* 11, 715–725. doi: 10.1007/s10021-008-9153-0

Taranu, Z. E., Zurawell, R. W., Pick, F., and Gregory-Eaves, I. (2012). Predicting cyanobacterial dynamics in the face of global change: the importance of scale and environmental context. *Glob. Chang. Biol.* 18, 3477–3490. doi: 10.1111/gcb.12015

Thornton, J. A., Harding, W. R., Dent, M., Hart, R. C., Lin, H., Rast, C. L., et al. (2013). Eutrophication as a 'wicked' problem. *Lakes Reserv.* 18, 298–316. doi: 10.1111/lre.12044

Tomlinson, M. C., Stumpf, R. P., Wynne, T. T., Dupuy, D., Burks, R., Hendrickson, J., et al. (2016). Relating chlorophyll from cyanobacteria-dominated inland waters to a MERIS bloom index. *Remote Sens. Lett.* 7, 141–149. doi: 10.1080/2150704X.2015.1117155

United States Army Corps of Engineers (2016). *Corps to Reduce Flows From Lake Okeechobee*. Available online at: http://www.saj.usace.army.mil/Media/News-Releases/Article/821885/corps-to-reduce-flows-from-lake-okeechobee/

Urquhart, E. A., and Schaeffer, B. A. (2020). Envisat MERIS and sentinel-3 OLCI satellite lake biophysical water quality flag dataset for the contiguous United States. *Data Brief* 28:104826. doi: 10.1016/j.dib.2019.104826

US EPA (2009). *National Lakes Assessment: A Collaborative Survey of the Nation's Lakes*. Washington, DC.

US EPA (2019). *Draft Technical Support Document: Implementing the 2019 Recommended Human Health Recreational Ambient Water Quality Criteria or Swimming Advisories for Microcystins and Cylindrospermopsin*. US Environmental Protection Agency Office of Water.

Visser, P. M., Ibelings, B. W., Bormans, M., and Huisman, J. (2015). Artificial mixing to control cyanobacterial blooms: a review. *Aquat. Ecol.* 43, 12–19. doi: 10.1007/s10452-015-9537-0

Walsby, A. E. (1977). The gas vacuoles of blue-green algae. *Sci. Am.* 237, 90–97. doi: 10.1038/scientificamerican0877-90

Williams, C. D., Aubel, M. T., Chapman, A. D., and D'Aiuto, P. E. (2007). Identification of cyanobacterial toxins in Florida's freshwater systems. *Lake Reserv. Manag.* 23, 144–152. doi: 10.1080/07438140709353917

Williams, C. D., Burns, J., Chapman, A. D., Flewelling, L., Pawlowicz, M., and Carmichael, W. (2001). *Assessment of Cyanotoxins in Florida's Lakes, Reservoirs, and Rivers. 1999. Final Annual Report to the Florida Harmful Algal Bloom Task Force*. St. Petersburg, FL.

World Health Organization (2003). *Guidelines for Safe Recreational Water Environments: Coastal and Fresh Waters*. Vol. 1. Geneva: World Health Organization.

Wynne, T. T., Stumpf, R. P., Tomlinson, M. C., and Dyble, J. (2010). Characterizing a cyanobacterial bloom in western Lake Erie using satellite imagery and meteorological data. *Limnol. Oceanogr.* 55, 2025–2036. doi: 10.4319/lo.2010.55.5.2025

Wynne, T. T., Stumpf, R. P., Tomlinson, M. C., Warner, R. A., Tester, P. A., Dyble, J., et al. (2008). Relating spectral shape to cyanobacterial blooms in the Laurentian Great Lakes. *Int. J. Remote Sens.* 29, 3665–3672. doi: 10.1080/01431160802007640