

ARTICLE

Using near-term forecasts and uncertainty partitioning to inform prediction of oligotrophic lake cyanobacterial density

Mary E. Lofton¹  | Jennifer A. Brentrup²  | Whitney S. Beck³  |
 Jacob A. Zwart⁴  | Ruchi Bhattacharya⁵  | Ludmila S. Brighenti⁶  |
 Sarah H. Burnet⁷  | Ian M. McCullough⁸  | Bethel G. Steele⁹  |
 Cayelan C. Carey¹  | Kathryn L. Cottingham²  | Michael C. Dietze¹⁰  |
 Holly A. Ewing¹¹  | Kathleen C. Weathers⁹  | Shannon L. LaDeau⁹ 

¹Department of Biological Sciences, Virginia Tech, Blacksburg, Virginia, USA

²Department of Biological Sciences, Dartmouth College, Hanover, New Hampshire, USA

³Department of Biology and Graduate Degree Program in Ecology, Colorado State University, Fort Collins, Colorado, USA

⁴U.S. Geological Survey, Integrated Information Dissemination Division, Middleton, Wisconsin, USA

⁵Legacies of Agricultural Pollutants (LEAP), University of Waterloo, Waterloo, Ontario, Canada

⁶Universidade do Estado de Minas Gerais, Divinópolis, Brazil

⁷Department of Fish and Wildlife Resources, University of Idaho, Moscow, Idaho, USA

⁸Department of Fisheries and Wildlife, Michigan State University, East Lansing, Michigan, USA

⁹Cary Institute of Ecosystem Studies, Millbrook, New York, USA

¹⁰Department of Earth and Environment, Boston University, Boston, Massachusetts, USA

¹¹Environmental Studies, Bates College, Lewiston, Maine, USA

Correspondence

Mary E. Lofton

Email: melofton@vt.edu

Present address

Jennifer A. Brentrup, Biology and Environmental Studies Department, St. Olaf College, Northfield, Minnesota, USA

Whitney S. Beck, U.S. Environmental Protection Agency, Washington, District of Columbia, USA

Funding information

Auburn Water District/Lewiston Water Division; Bates College; Dartmouth College; Division of Biological Infrastructure, Grant/Award Numbers: 0434684, 1933016; Division of Computer

Abstract

Near-term ecological forecasts provide resource managers advance notice of changes in ecosystem services, such as fisheries stocks, timber yields, or water quality. Importantly, ecological forecasts can identify where there is uncertainty in the forecasting system, which is necessary to improve forecast skill and guide interpretation of forecast results. Uncertainty partitioning identifies the relative contributions to total forecast variance introduced by different sources, including specification of the model structure, errors in driver data, and estimation of current states (initial conditions). Uncertainty partitioning could be particularly useful in improving forecasts of highly variable cyanobacterial densities, which are difficult to predict and present a persistent challenge for lake managers. As cyanobacteria can produce toxic and unsightly surface scums, advance warning when cyanobacterial densities are increasing could help managers mitigate water quality issues. Here, we fit 13 Bayesian

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *Ecological Applications* published by Wiley Periodicals LLC on behalf of The Ecological Society of America.

and Network Systems, Grant/Award Number: 1737424; Division of Emerging Frontiers, Grant/Award Numbers: 0842112, 0842125, 0842267; Division of Environmental Biology, Grant/Award Numbers: 0749022, 1010862, 1137327, 1638577, 1702991, 1753639, MSB-1638575, MSB-1926050; Lake Sunapee Protective Association; Midge Eliassen Fellowship; NSF, United States National Science Foundation, Grant/Award Numbers: CNS-1737424, DBI-0434684, DEB-0749022, DEB-1010862, EF-0842112, EF-0842125, EF-0842267, EF-1137327, EF-1638577, EF-1702991, ICER-1517823, MSB EF-1638575; Virginia Polytechnic Institute and State University; Virginia Water Resources Research Center

Handling Editor: Eric J. Ward

state-space models to evaluate different hypotheses about cyanobacterial densities in a low nutrient lake that experiences sporadic surface scums of the toxin-producing cyanobacterium, *Gloeotrichia echinulata*. We used data from several summers of weekly cyanobacteria samples to identify dominant sources of uncertainty for near-term (1- to 4-week) forecasts of *G. echinulata* densities. Water temperature was an important predictor of cyanobacterial densities during model fitting and at the 4-week forecast horizon. However, no physical covariates improved model performance over a simple model including the previous week's densities in 1-week-ahead forecasts. Even the best fit models exhibited large variance in forecasted cyanobacterial densities and did not capture rare peak occurrences, indicating that significant explanatory variables when fitting models to historical data are not always effective for forecasting. Uncertainty partitioning revealed that model process specification and initial conditions dominated forecast uncertainty. These findings indicate that long-term studies of different cyanobacterial life stages and movement in the water column as well as measurements of drivers relevant to different life stages could improve model process representation of cyanobacteria abundance. In addition, improved observation protocols could better define initial conditions and reduce spatial misalignment of environmental data and cyanobacteria observations. Our results emphasize the importance of ecological forecasting principles and uncertainty partitioning to refine and understand predictive capacity across ecosystems.

KEYWORDS

algae, Bayesian model, blooms, ecological forecasting, hindcast, lake, oligotrophic, phytoplankton, scums, state-space model, uncertainty partitioning, variance partitioning

INTRODUCTION

Near-term ecological forecasts, defined as daily to decadal predictions of the state of ecosystems (Clark et al., 2001; Dietze et al., 2018), can be helpful to resource managers in systems ranging from fisheries stocks to disease outbreaks in protected species populations (Hobbs et al., 2015; Kuikka et al., 2014). For example, near-term forecasts have been used to provide projections for alternate management decisions in salmon fisheries and forests (Kuikka et al., 2014; Thomas et al., 2018, 2020), help managers allot fisheries take quotas (or avoid bycatch; Hobday et al., 2019), and provide advance notice of public safety hazards such as red tides (McGowan et al., 2017; Stumpf et al., 2009). Effective near-term forecasts include fully specified uncertainty by quantifying the total variance around a prediction and identifying the relative contributions of different sources of uncertainty (Dietze et al., 2018; Table 1).

Uncertainty in ecological forecasts may arise from several different sources, including initial conditions

uncertainty, parameter uncertainty, process uncertainty, observation uncertainty, driver or covariate data uncertainty, and random effects uncertainty (Dietze, 2017a; Table 1). Partitioning the variance associated with a forecast into these components allows for more targeted efforts to understand and improve forecasts to inform management of natural resources (Bauer et al., 2015; Page et al., 2018). For example, the dominant contributor to uncertainty in weather forecasts is from initial conditions because the atmosphere's internal instability amplifies even small errors in estimates of current states, and the physical processes controlling weather given a set of current conditions are relatively well defined (Dietze, 2017b). This has directed weather forecasters to prioritize efforts to better measure current (initial) atmospheric conditions (Bauer et al., 2015). In contrast, the dominance of process uncertainty in a forecast indicates that researchers need to consider alternative model structures and additional or different explanatory variables to describe the biological or ecological process of interest (e.g., Page et al., 2017; Thomas et al., 2018).

TABLE 1 Terms associated with partitioning uncertainty in ecological models and forecasts

Term	Definition	Example
Credible interval	Range of values within which a parameter or model prediction falls with a specified probability; does not include observation uncertainty	Range of chlorophyll <i>a</i> values within which 95% of possible latent values forecasted for tomorrow fall; incorporates initial conditions, process, parameter, and driver data uncertainty
Driver data uncertainty	Uncertainty in the estimate or measurement of driver data (environmental predictors of the forecasted state)	Uncertainty in observations of soil temperature needed to drive a soil respiration model; uncertainty in weather forecasts
Hindcast	Predictions with specified uncertainty of a past time period using data (withheld from model calibration) that are iteratively assimilated into the model (Jolliffe & Stephenson, 2003)	Making model predictions for tick abundances observed 2 years ago using a model calibrated to observations from 10 years prior
Initial conditions uncertainty	Uncertainty associated with the starting conditions of a forecasting model run	Uncertainty in initial focal states, such as fish abundance, chlorophyll <i>a</i> , or soil carbon stock
Observation uncertainty	Difference between the observed data and the true (latent) state that the model is designed to predict; does not propagate forward, so it does not affect the credible interval	Calibration uncertainty in a temperature sensor; sampling uncertainty when estimating species abundance
Parameter uncertainty	Variance around the model parameter estimates	Uncertainty in the growth rate parameter in a timber yield model
Predictive interval	Range of values within which predicted observations are expected to fall with a specified probability; includes observation uncertainty in addition to other uncertainties and so is larger than the credible interval; should be used when comparing models to observed data	Range of chlorophyll <i>a</i> values within which 95% of possible observation values forecasted for tomorrow fall
Process uncertainty	Uncertainty due to model specification (ecological processes that are simplified, absent, or incorrectly represented by the model) or inherent stochasticity in the system	Uncertainty arising from not including an important life history stage in a population growth model; uncertainty arising from demographic stochasticity in plankton communities
Random effects uncertainty	Uncertainty associated with estimation of random effects, which are used to describe shared variance across groups in space and time	Uncertainty in the value of a random site effect representing the degree to which beetle communities vary together across different sampling sites in a metacommunity model

Note: Definitions are adapted from Dietze (2017a) unless otherwise specified.

Estimating uncertainty has become more common in ecological analyses that generate forecasts (see studies in Appendix S1: Table S1 for examples). However, formal uncertainty partitioning that includes all the potential sources of forecast uncertainty is less common and methods are not standardized, making it difficult to compare how different components of uncertainty contribute across ecological systems or among focal state variables. For example, while studies by Gertner et al. (1996), Valle et al. (2009), Wang et al. (2009), and Thomas et al. (2018) (Appendix S1: Table S1) all forecast metrics of forest biomass and productivity, differences in how they estimate

and partition uncertainty limit synthetic understanding of how uncertainties in process structure or parameter estimation, for instance, limit confidence in forest productivity forecasts.

Forecasting freshwater cyanobacterial dynamics has been a persistent challenge for researchers and water quality managers (Janssen et al., 2019; Rouso et al., 2020), and to our knowledge uncertainty partitioning analysis has not been used to support forecasting capacity in this system. Cyanobacteria are increasing in many lakes and reservoirs worldwide due to climate and land-use change, posing substantial problems for

drinking water managers and other stakeholders (Carey, Ibelings, et al., 2012; O'Neil et al., 2012; Paerl et al., 2011). Many cyanobacterial taxa create toxic and unsightly scums that cause taste and odor problems and clog filters at drinking water treatment plants; consequently, knowing when cyanobacterial density is likely to increase could allow managers to take preemptive action to mitigate deleterious water quality effects (Ibelings et al., 2016; Stroom & Kardinaal, 2016).

Despite substantial research on drivers of cyanobacterial dominance (e.g., Carey, Ibelings, et al., 2012; Paerl & Otten, 2013) and technological developments permitting high-frequency observations of cyanobacterial density (e.g., Catherine et al., 2012; Le Vu et al., 2011), cyanobacterial abundance predictions often deviate substantially from observations (Hamilton et al., 2009; Rigosi et al., 2010; Rouso et al., 2020). Cyanobacterial densities can change rapidly on timescales of days to weeks (Carpenter et al., 2020; Dokulil & Teubner, 2000; Huisman & Hulot, 2005; Rolland et al., 2013), with densities in many lakes remaining relatively low for much of the year and then rapidly increasing from one sample period to the next (e.g., Bormans et al., 2005; Carey et al., 2014; Rolland et al., 2013), making prediction difficult. In a review of process-based phytoplankton succession models, Rigosi et al. (2010) observed that prediction errors of 40% or more were common for biomass of individual phytoplankton groups, such as cyanobacteria, and that the timing of maximum phytoplankton biomass was often mis-predicted by as much as one month. A systematic literature review of cyanobacteria bloom models by Rouso et al., 2020 compared the coefficient of determination (R^2) between predicted and observed values that were reported for data-driven models, such as artificial neural networks and decision trees, and found that at a 1-week forecast horizon, R^2 values ranged from ~ 0.1 to 0.9, with most models falling between 0.5 and 0.65. Hamilton et al. (2009) had good success in forecasting cyanobacteria bloom probability at a one-month timestep but did not attempt to forecast cyanobacterial density as a response variable. Moreover, few cyanobacterial forecasting studies have examined forecast uncertainty (Janssen et al., 2019; but see Hamilton et al., 2009; Huang et al., 2013; Massoud et al., 2018; Page et al., 2017), and fewer still have conducted uncertainty partitioning.

Cyanobacterial blooms are often associated with high nutrient levels (Dokulil & Teubner, 2000), and much of the effort to predict cyanobacterial densities has been focused on nutrient-rich lakes (reviewed by Rouso et al., 2020). As a result, prediction efforts in low-nutrient lakes have lagged behind and understanding why cyanobacterial densities change over the short term in such lakes is especially challenging. Consequently, teasing apart the different sources of uncertainty and their

relative importance to cyanobacterial forecast precision may help prioritize research efforts in economically important oligotrophic waterbodies. Increases in cyanobacteria densities have been documented in north temperate low-nutrient lakes throughout the United States (Carey, Ewing, et al., 2012; Sterner et al., 2020), Canada (Winter et al., 2011), and Europe (Freeman et al., 2020), and these increases are often associated with significant economic losses and public health concerns (Dodds et al., 2009; Mueller et al., 2016; Stoddard et al., 2016). High water quality in oligotrophic lakes provides substantial economic benefit through recreational use and high lakeside property values (Mueller et al., 2016; Wilson & Carpenter, 1999). Moreover, some oligotrophic systems are permitted as drinking water sources with reduced filtration requirements when their water quality meets U.S. Environmental Protection Agency (U.S. EPA) standards, thereby reducing water treatment costs (U.S. EPA, 1991; Kauffman, 2016).

Prior studies provide several hypotheses for what environmental drivers likely trigger cyanobacterial growth or accumulation of cyanobacterial surface scums apart from nutrient concentrations, including higher temperatures, which result in increased growth (Hamilton et al., 2009; Paerl & Huisman, 2008); light-induced triggering of cell germination and growth (Karlsson-Elfgren et al., 2004; Roelofs & Oglesby, 1970); water column mixing resulting in more recruitment of dormant cells from the sediment and/or dilution of surface water cyanobacterial density, which can occur due to temperature changes, precipitation events, or wind (Carey et al., 2014; de Eyto et al., 2016; Jennings et al., 2012; Kuha et al., 2016); strong thermal stratification resulting in greater incidence of surface scums (Carey, Ibelings, et al., 2012); and wind-driven aggregation of cells or colonies in nearshore zones (Cyr, 2017; Roelofs & Oglesby, 1970). The development of forecast models with uncertainty partitioning is needed to compare and evaluate these hypotheses in a predictive framework.

While there are a variety of techniques that can be used to develop forecast models with partitioned uncertainty, Bayesian state-space models (SSMs) are particularly suitable (Clark, 2007; Dietze, 2017a; Hobbs & Hooten, 2015). State-space models focus on estimating the true, latent state of the system by explicitly accounting for observation and process uncertainty. These dynamic models are structured so that each modeled latent state is a function of the previous latent state, independent of observations at other time points (Hobbs & Hooten, 2015; Dietze, 2017a; Figure 1). Importantly, Bayesian state-space models estimate latent states as distributions (Hobbs & Hooten, 2015). This feature permits forecasters to sample from the most recent SSM estimated latent state to quantify uncertainty in the current

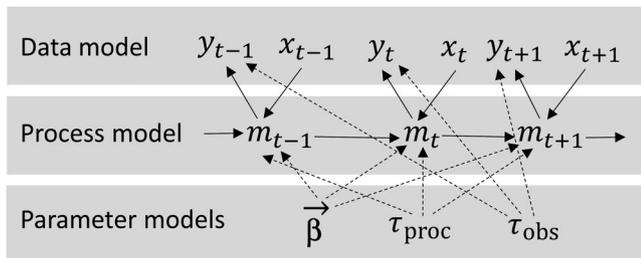


FIGURE 1 Directed acyclic graph (DAG) of a Bayesian state-space model, where y_t is the observed cyanobacterial density at time t , x_t are driver data (physical covariates) at time t , m_t is the estimated true, or latent, cyanobacterial density at time t , β is a vector of parameters in the process model (slope, intercept, etc.), and τ_{proc} and τ_{obs} are the precisions of normal distributions representing process error and observation error, respectively. Parameters are modeled as distributions in the parameter model. Parameters, along with driver data, determine the predicted latent states in the process model, which are fitted to observations using the data model

state conditions that will be used to initiate a subsequent forecast. Thus, initial conditions change each time a forecast is produced to reflect the updated estimate of “today’s” state of the system (Dietze, 2017b; Thomas et al., 2020). Furthermore, Bayesian SSMS use distributions rather than fixed values to represent all unknown values, including parameters and as-yet-unobserved future values for driver variables, allowing for quantification of uncertainty associated with each of these components and missing data.

We developed and evaluated a suite of Bayesian SSMS with different structures and tested different environmental variables hypothesized to be important in driving cyanobacterial density, including water temperature, thermal stability, wind, and precipitation. We fit each model to historical weekly cyanobacterial densities of *Gloeotrichia echinulata* measured from 2009 to 2014 in Lake Sunapee, New Hampshire, USA. We then generated “forecasts” of cyanobacterial density for 2015–2016, which we hereafter refer to as hindcasts because although they were generated as forecasts from the perspective of the model, the forecast period had already occurred (see “hindcast” definition in Table 1). We assessed and conducted uncertainty partitioning of our hindcasts to address the following questions: (1) Which model structures and environmental covariates best predict oligotrophic lake cyanobacterial density over 1–4 week forecast horizons? (2) What are the dominant sources of uncertainty in oligotrophic lake cyanobacterial forecasts? And (3) how do the relative contributions of different sources of uncertainty vary among models with differing complexity and environmental covariates? We

discuss how our results can inform future efforts to forecast oligotrophic lake cyanobacterial density and relate to patterns of predictive uncertainty observed in other ecosystems.

METHODS

Focal cyanobacterium

Gloeotrichia echinulata is a colonial, filamentous cyanobacterium commonly found in oligotrophic north temperate lakes in North America and Europe (Carey, Ewing, et al., 2012; Freeman et al., 2020; Karlsson-Elfgren et al., 2005; Winter et al., 2011). *G. echinulata* can form surface scums and produce toxins during the warm summer months (Carey, Ewing, et al., 2012; Karlsson-Elfgren et al., 2005). In the fall and winter, *G. echinulata* enters a dormant life stage on lake sediments in response to adverse environmental conditions, resulting in a seasonal cycle wherein vegetative *G. echinulata* is first observed in the water column in late spring, water column densities peak in the summer, and then densities decrease as *G. echinulata* enters dormancy in the fall (Roelofs & Oglesby, 1970).

Occurrence of *G. echinulata* surface scums in oligotrophic, north temperate lakes appears to have increased in recent decades (Carey et al., 2008; Carey, Ewing, et al., 2012; Winter et al., 2011), motivating researchers to improve understanding and prediction of *G. echinulata* density in these ecosystems. While nutrients are often a driver of cyanobacterial growth in eutrophic lakes (Dokulil & Teubner, 2000), current understanding of dynamics in oligotrophic lakes suggests that physical drivers (i.e., water temperature, light, wind) may be important for determining *G. echinulata* densities (Carey et al., 2014; Cyr, 2017; Karlsson-Elfgren et al., 2004; Roelofs & Oglesby, 1970), and we focus on these physical variables as predictors of *G. echinulata* density in this study.

Study site

We sampled *G. echinulata* surface abundance and collected environmental data weekly in May–October from 2009 to 2016 at two nearshore sites in Lake Sunapee, New Hampshire, USA, a recreational lake with high property values that also serves as a public drinking water supply (Figure 2). Lake Sunapee is a large, oligotrophic lake (43°24' N, 72°2' W, maximum depth = 33.7 m, surface area = 16.69 km², volume = 1.94 × 10³ m³, mean depth = 11.6 m; Ward et al., 2020). While high-nutrient

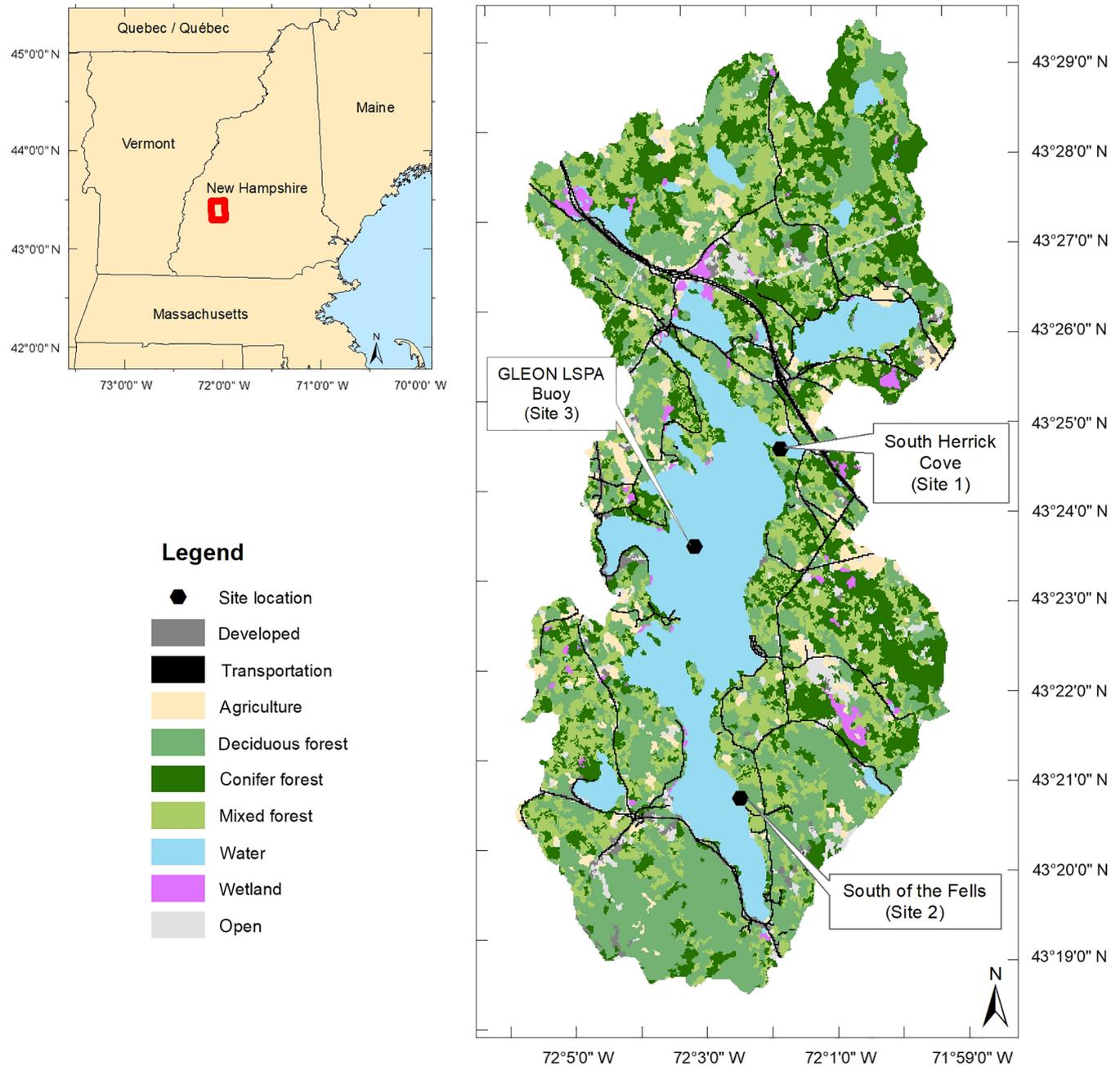


FIGURE 2 Map of Lake Sunapee, New Hampshire, USA with locator map (inset). Data from Site 1 were used for Bayesian state-space models, data from Site 2 were used to inform priors for Site 1 models, and data from Site 3 provided lake-level covariates for Site 1 models. Land use data was obtained from the New Hampshire Geographically Referenced ANalysis and Information transfer system (GRANIT) land cover assessment in 2001

(eutrophic) lakes can have total phosphorus (TP) concentrations $\geq 24 \mu\text{g/L}$ and total nitrogen (TN) concentrations ranging from $\sim 400\text{--}1600 \mu\text{g/L}$ (Carlson, 1977; Carlson & Simpson, 1996; Gibson et al., 2000), mean TP concentration in the surface waters of oligotrophic Lake Sunapee between 2009–2016 was $1.37 \pm 17.4 \mu\text{g/L}$ (mean \pm SD; Steele et al., 2021) and mean TN concentration from 2009 to 2012 was $172 \pm 25 \mu\text{g/L}$ (Cottingham, 2020). Lake Sunapee mean Secchi depth was $7.0 \pm 2.0 \text{ m}$ (Steele et al., 2021). Lake Sunapee typically thermally stratifies

from June–September with a mean seasonal thermocline depth of 7–9 m from 2009 to 2016 (Lake Sunapee Protective Association (LSPA), Weathers, et al., 2020). The watershed ($\sim 107 \text{ km}^2$ not including lake surface area) is 80% forested, but shoreline development has been increasing in recent decades (Cobourn et al., 2018).

Our research team began a weekly *G. echinulata* monitoring program at two sampling sites in collaboration with the LSPA in 2005 (Carey et al., 2008, 2014). Multiple sites were sampled to represent the high spatial

variability in *G. echinulata* densities among sites, with both high and low densities often observed concurrently at different sites (Cottingham et al., 2020a). The focal sampling site for this study (Site 1; Figure 2) was chosen because it frequently exhibits high local densities of *G. echinulata*, which are of concern to the LSPA. We used data from a second nearshore site (Site 2) only to generate informed priors for *G. echinulata* observation error and nearshore water temperature. Data from Site 2 were not used to inform estimates of *G. echinulata* density latent states or included in any hindcasting analyses due to the high variability of *G. echinulata* density among sites and our aim of estimating local site densities rather than a lake-wide mean density. We focused our analyses on 2009–2016 for this study because those years had at least 20 weeks of sampling data (Cottingham et al., 2020a). We used data from 2009 to 2014 for model fitting to provide as much data as possible for parameter estimation while still reserving enough data to conduct hindcasts over multiple years (2015–2016). During our 8-year study period there were six missing weekly *G. echinulata* observations, four of which occurred during the 2015–2016 hindcasting period. Notably, while *G. echinulata* is increasingly reported across a number of low nutrient lakes in the northeastern United States (Carey, Ewing, et al., 2012; Winter et al., 2011), Lake Sunapee has the longest duration of weekly *G. echinulata* monitoring data available for any oligotrophic lake, to the best of our knowledge. Thus, the Lake Sunapee dataset provides a unique opportunity to examine environmental drivers and sources of uncertainty in hindcasts of *G. echinulata* density over time.

***G. echinulata* data collection and sample processing**

Gloeotrichia echinulata surface abundance at both nearshore sites was sampled each week in the top 1 m of the water column by combining two vertical tows from 1 m to the surface using a 30 cm diameter, 80- μ m mesh plankton net (Wildlife Supply Co., Yulee, Florida, USA). *G. echinulata* were transferred from the net and preserved in opaque plastic bottles using Lugol's iodine (Carey et al., 2014). Total *G. echinulata* samples were counted using a Leica MZ12 dissecting microscope (Leica, Buffalo Grove, Illinois, USA). Density was quantified according to the number of colonies and filament bundles (immature, developing colonies) per liter rather than biovolume following protocols used in previous studies of *G. echinulata* (Barbiero & Welch, 1992; Karlsson-Elfgren et al., 2005; Roelofs & Oglesby, 1970). We then converted abundance to density by dividing the total number of colonies and

filament bundles in each sample by the volume of water sampled by the plankton net (Carey et al., 2014). All data are publicly available through the Environmental Data Initiative repository (Cottingham, 2020; Cottingham et al., 2020a, 2020b; Lake Sunapee Protective Association (LSPA), Steele, et al., 2020; Lake Sunapee Protective Association (LSPA), Weathers, et al., 2020; Lofton et al., 2022; Steele et al., 2021).

Physical driver data

To study the effect of temperature on *G. echinulata* growth, water temperature was monitored hourly using Onset loggers at our nearshore sampling sites (Sites 1 and 2; Figure 2; Cottingham et al., 2020b). Growing degree days (GDD), a measure of heat accumulation during the growing season, were calculated using these water temperatures for each day when *G. echinulata* was sampled. To investigate effects of thermal stratification on *G. echinulata* surface density, water temperature profiles from the Global Lake Ecological Observatory Network (GLEON) buoy, deployed in the lake by the LSPA since 2007 (Site 3; Figure 2), were used to calculate Schmidt stability, a measure of thermal stratification strength that indicates the amount of energy required to homogenize temperature across the water column (Idso, 1973; LSPA, Weathers, et al., 2020). To examine whether wind could drive nearshore aggregation of *G. echinulata* colonies, wind data from the LSPA/GLEON buoy (Site 3) were aggregated from minute and hourly scales to calculate daily summary statistics (LSPA, Steele, et al., 2020). Solar radiation data from the North American Land Data Assimilation System Phase 2 (NLDAS-2) forcing data set (<https://ldas.gsfc.nasa.gov/nldas>; Lofton et al., 2022) and photosynthetically active radiation (PAR) data from the LSPA/GLEON buoy (LSPA, Steele, et al., 2020) were similarly aggregated to determine whether surface water light availability was an important predictor of *G. echinulata* density. Finally, we calculated summary statistics of daily precipitation data from the Parameter-elevation Relationships on Independent Slopes Model (PRISM) model (<http://www.prism.oregonstate.edu>; Lofton et al., 2022) to examine the effect of storm events and subsequent water column mixing on *G. echinulata* pelagic populations (see Appendix S1: Section S1 and scripts in Data S1: 1_Get_data/) for further information on environmental data processing). To align potential physical driver data most closely with weekly *G. echinulata* density observations, we used daily summary statistics (e.g., daily mean water temperature; daily sum of precipitation) of each driver on the day of *G. echinulata* sampling each week in our state-space models, except when otherwise specified, such as when we examined lagged values or moving averages.

Selection of physical covariates for Bayesian models

In keeping with a Bayesian approach, where previous knowledge of a system is incorporated into the current analysis, we leveraged the expertise of our working group and published research to develop a list of 82 candidate physical variables hypothesized to influence *G. echinulata* densities (Carey et al., 2014; Carey, Ibelings, et al., 2012; Cyr, 2017; de Eyto et al., 2016; Hamilton et al., 2009; Jennings et al., 2012; Karlsson-Elfgren et al., 2004; Kuha et al., 2016; Paerl & Huisman, 2008; Roelofs & Oglesby, 1970).

We then performed a standardized variable selection process to determine which potential drivers to include in Bayesian state-space models (Appendix S1: Section S2 and Data S1: 2_Covariate_correlation_analysis/2_Covariate_correlation.R). We used Spearman correlations to prioritize inclusion in our Bayesian models and applied a Holm-Bonferroni correction to p values to account for multiple comparisons and prevent false positive identification of significant associations. The full list of covariate summary statistics is in Appendix S1: Table S2. This approach identified eight variables significantly associated with weekly *G. echinulata* densities (based on the model fit period of 2009–2014) for further evaluation: daily minimum water temperature on the sampling day (MinWaterTemp), daily minimum water temperature with a 1-week lag (MinWaterTempLag), seven-day moving average of water temperature (WaterTempMA), weekly difference in median Schmidt stability (Δ Schmidt), daily maximum Schmidt stability with a 1-week lag (SchmidtLag), daily mean of a wind direction indicator variable with a two-day lag (WindDir; see Appendix S1: Section S1 and Data S1: 1_Get_data/1_Get_buoy_wind.R for details on wind indicator variable calculation), growing degree days (GDD), and daily sum of precipitation (Precip).

Development of Bayesian state-space models

During model development, a total of 18 Bayesian SSMs were fit to data collected from Site 1, and these SSMs increased in complexity from a random walk with no covariates to models containing one or two of the eight prioritized physical driver variables (Table 2; see model code in DataS1: 4.1_JAGS_models/). Models were developed to reflect the ecology and observed density patterns of *G. echinulata*, which typically exists at low densities interspersed with rare, high-density events during the growing season and is dormant or at extremely low densities during winter (Carey, Ewing, et al., 2012; Cottingham et al., 2020a; Karlsson-Elfgren et al., 2005; Roelofs &

Oglesby, 1970). As a result, the expected density of *G. echinulata* during our study period was considered to be a low-density state, perhaps lower than the mean of our observed data, which includes rare, high-density events. For the purposes of model comparison, we considered four baseline models: first, a random walk process model (RW model; Table 2); second, a random walk model with an offset that allows the model to achieve a non-zero equilibrium (OffsetRW model); third, a model with an autocorrelation term (AC model); and fourth, a model including both an offset and an autocorrelation term (BaseLM model). Importantly, while the offset term (β_0 ; Table 2) permits models to achieve a non-zero equilibrium, our hindcasts are conducted over short timescales (1–4 weeks) such that we do not expect our models to achieve equilibrium over the course of the forecast horizon. All four baseline models included an informed prior distribution for observation error that was developed using data from Site 2 (Appendix S1: Section S3). We also assessed a random walk model with a random year effect as a possible baseline model, but because we were unable to estimate a non-zero year effect (Appendix S1: Table S3), we did not include a random year effect in subsequent models.

The structure of our BaseLM process within the full Bayesian SSM is provided in the following equations (Equations 1–3):

$$\text{Data model: } y_t \sim N(m_t, \tau_{\text{obs}}) \quad (1)$$

$$\text{Process model: } m_t \sim N((\beta_0 + \beta_1 \times m_{t-1}), \tau_{\text{proc}}) \quad (2)$$

$$\begin{aligned} \text{Parameter model: } \vec{\beta} &\sim N(\vec{0}, \vec{\tau}_{\beta}), \tau_{\text{proc}} \\ &\sim \text{Gamma}(\alpha_{\text{proc}}, \beta_{\text{proc}}), \\ \tau_{\text{obs}} &\sim \text{Gamma}(\alpha_{\text{obs}}, \beta_{\text{obs}}), \\ m_0 &\sim N(-5, 100) \end{aligned} \quad (3)$$

where y_t is the current observation of *G. echinulata* density, m_t is the current latent state of *G. echinulata* density, τ_{obs} is the precision of the normal distribution specifying the objective function or data model, $\vec{\beta}$ is the vector of coefficients of the linear model process where β_0 is a bias term and β_1 is an autocorrelation term, τ_{proc} is the precision of the normal distribution specifying process uncertainty, $\vec{0}$ and $\vec{\tau}_{\beta}$ are respectively the vectors of mean and precision for process coefficient normal priors, α_{proc} and β_{proc} are respectively the shape and rate for the gamma prior on precision in the process model, α_{obs} and β_{obs} are respectively the shape and rate for the gamma prior on precision in the data model (informed using data from Site 2), m_0 is the latent state of *G. echinulata* density at $t = 0$ (which is the beginning of the sampling season each

TABLE 2 List of Bayesian state-space models and covariates used in hindcast analysis

Model name	Model description	Process model	Covariates
RW	Random walk	$m_{t+1} \sim N(m_t, \tau_{\text{proc}})$	
OffsetRW	Random walk with an offset constant	$m_{t+1} \sim N((\beta_0 + m_t), \tau_{\text{proc}})$	
AC	State-space model with autocorrelation term	$m_{t+1} \sim N((\beta_1 \times m_t), \tau_{\text{proc}})$	
BaseLM	State-space model with both offset and autocorrelation	$m_{t+1} \sim N((\beta_0 + \beta_1 \times m_t), \tau_{\text{proc}})$	
MinWaterTemp	BaseLM with a single linear covariate	$m_{t+1} \sim N((\beta_0 + \beta_1 \times m_t + \beta_2 \times x_t), \tau_{\text{proc}})$	Minimum water temperature on sampling day
MinWaterTempLag	BaseLM with a single linear covariate	$m_{t+1} \sim N((\beta_0 + \beta_1 \times m_t + \beta_2 \times x_t), \tau_{\text{proc}})$	Minimum water temperature 1 week prior to the sampling day
WaterTempMA	BaseLM with a single linear covariate	$m_{t+1} \sim N((\beta_0 + \beta_1 \times m_t + \beta_2 \times x_t), \tau_{\text{proc}})$	7-day Moving average of water temperature including the sampling day
SchmidtLag	BaseLM with a single linear covariate	$m_{t+1} \sim N((\beta_0 + \beta_1 \times m_t + \beta_2 \times x_t), \tau_{\text{proc}})$	Maximum Schmidt stability 1 week prior to the sampling day
WindDir	BaseLM with a single linear covariate	$m_{t+1} \sim N((\beta_0 + \beta_1 \times m_t + \beta_2 \times x_t), \tau_{\text{proc}})$	Proportion of daily wind measurements blowing towards Site 1 with a 2-day lag
GDD	BaseLM with a single quadratic covariate	$m_{t+1} \sim N((\beta_0 + \beta_1 \times m_t + \beta_2 \times x_t + \beta_3 \times x_t^2), \tau_{\text{proc}})$	Growing degree days
Schmidt+Wind	BaseLM with two linear covariates	$m_{t+1} \sim N((\beta_0 + \beta_1 \times m_t + \beta_2 \times x1_t + \beta_3 \times x2_t), \tau_{\text{proc}})$	Maximum Schmidt stability 1 week prior to the sampling day and proportion of daily wind measurements blowing towards Site 1 with a 2-day lag
Temp+Wind	BaseLM with two linear covariates	$m_{t+1} \sim N((\beta_0 + \beta_1 \times m_t + \beta_2 \times x1_t + \beta_3 \times x2_t), \tau_{\text{proc}})$	Minimum water temperature on sampling day and proportion of daily wind measurements blowing towards Site 1 with a 2-day lag
Wind+GDD	BaseLM with one linear and one quadratic covariate	$m_{t+1} \sim N((\beta_0 + \beta_1 \times m_t + \beta_2 \times x1_t + \beta_3 \times x2_t + \beta_4 \times x2_t^2), \tau_{\text{proc}})$	Proportion of daily wind measurements blowing towards Site 1 with a 2-day lag and growing degree days

Notes: m_t is the latent state of *Gloeotrichia echinulata* density at time t , $N()$ represents a normal distribution with mean and precision (τ_{proc} ; precision is the inverse of the variance). X , $x1$, and $x2$ are environmental covariates in single-covariate and two-covariate models. $\vec{\beta}$ represents parameters for the process model equations. Models that were fit to 2009–2014 data but subsequently excluded from the hindcast analysis due to coefficient estimates close to 0 can be found in Appendix S1: Table S4.

year). The mean and precision for the normal distribution prior on m_0 were informed by expert opinion (author H. Ewing).

We developed a series of one-covariate models by extending our BaseLM to include a single physical covariate from those identified in our selection process (MinWaterTemp, MinWaterTempLag, WaterTempMA, Δ Schmidt, SchmidtLag, WindDir, Precip, and GDD;

Table 2; Appendix S1: Table S4). The influence of GDD on *G. echinulata* was visibly nonlinear in our preliminary analyses (Appendix S1: Figure S1), so a quadratic term was included in the model to evaluate the influence of GDD on *G. echinulata* growth. We subsequently developed two-covariate models based on which covariates were estimated to have non-zero coefficients during fitting of single-covariate models (Schmidt+Wind,

Temp+Wind, Schmidt+Temp, Schmidt+GDD, Wind+GDD; Table 2; Appendix S1: Table S4).

We fit each model over a 6-year period from 2009 to 2014, assessed model performance during a 2-year hindcasting period of 2015–2016, and then conducted uncertainty partitioning.

Model fitting using 2009–2014 data

We fit each SSM to weekly observations collected in 2009–2014 using the R packages *rjags* and *runjags* (*rjags* v.4-8, *runjags* v. 2.0.4-2; Denwood & Plummer, 2019; Plummer et al., 2019), which call the associated package for Just Another Gibbs Sampler (JAGS v. 4.3.0; <https://sourceforge.net/projects/mcmc-jags/>). All modeling and analyses were conducted in the R statistical environment (R version 4.0, R Core Team, 2020; Data S1: 4.2_Calibrate_Bayesian_models/4.2_Calibrate_Bayesian_models.R). We natural log-transformed *G. echinulata* densities to avoid prediction of negative concentrations and standardized all covariates using *Z* scores to facilitate model convergence. We ran three Markov chain Monte Carlo (MCMC) chains for each model, with an adaptation period of 5000 iterations, a burn-in of 10,000 iterations, and a sample size of 50,000 iterations, which we thinned to 7500 samples for hindcasting and model assessment. We evaluated convergence using the potential scale reduction factor of the Gelman-Rubin statistic, sometimes referred to as \hat{R} , where a value approaching 1 indicates that the model has converged well on a parameter estimate both within and among MCMC chains (Appendix S1: Tables S5 and S6). Missing data occurred for several of our candidate environmental drivers, so NA values were imputed using a missing data model with a Gaussian prior with mean and variance of observations from the same week across the model fitting period (2009–2014). Each model was structured around the sampling season, extending from the last week in May to the first week in October for 20 weeks per year, with a single, inter-season time step between October and May. To accommodate the gap in data collection between sampling seasons, the latent state of *G. echinulata* density at the start of each sampling season was drawn from a prior informed using expert opinion (author H. Ewing). We used an informed prior for start-of-season *G. echinulata* densities each year to ensure reasonable early season estimates and accommodate the possibility (and expert opinion) that springtime densities would not necessarily be similar to early fall densities the previous year (Carey et al., 2014; de Senerpont Domis et al., 2013; Hampton et al., 2016; Özkundakci et al., 2015; Phillips & Fawley, 2002; Wiedner & Nixdorf, 1998). Models with one or more covariates for which we were unable to estimate

non-zero coefficients during model fitting were excluded from the hindcasting analysis (Appendix S1: Table S4).

Hindcasting using 2015–2016 data

To assess the forecast skill of our Bayesian state-space models, we constructed an uncertainty partitioning algorithm in R (Figure 3) to produce 1-week-ahead through 4-week-ahead hindcasts of *G. echinulata* density in 2015–2016 with the propagated uncertainty estimates from each model (Data S1: 4.3_Hindcasting/4.3_Hindcasting.R). The primary purposes of this algorithm were to (1) facilitate calculation of the relative contributions of various sources of uncertainty to total hindcast uncertainty, and (2) to retain the important temporal autocorrelation of hindcasted covariates such as water temperature by using an ensemble approach (described below). We focus on results from the 1-week-ahead and 4-week-ahead hindcasts here for brevity; however, full model output is available in the Environmental Data Initiative repository (Lofton et al. 2022).

We assimilated data by iteratively adding 1 week of data to our model input dataset and re-running our SSMs to update parameter estimates and initial state conditions (“today’s” *G. echinulata* density) for hindcasts (Figure 3 Step 1). The posterior distribution for each parameter and initial condition was then fed into an uncertainty partitioning algorithm (Figure 3 Step 2 a, b), outside of JAGS. This algorithm ran the process model 4 weeks into the future while conducting Monte Carlo uncertainty propagation (Figure 3 Step 3). Importantly, the algorithm produced hindcasts using different combinations of uncertainty sources (e.g., just parameter uncertainty, or parameter and initial conditions uncertainty, etc.), which was needed for uncertainty partitioning.

We hindcasted “future” driver data for each physical covariate using data observations from 2009 to 2014 for the 2015 hindcasts and from 2009 to 2015 for the 2016 hindcasts. These historical driver time series were resampled with replacement for each of the 7500 hindcast model iterations. We chose to hindcast covariates using an ensemble of historical observations, rather than imputing them within our JAGS model, to account for temporal autocorrelation in driver data (Figure 3, Step 2c). As hindcasts were running, driver data from 2015 to 2016 were assimilated along with *G. echinulata* observations and thereby used to update posteriors throughout the hindcasting period. As a check on our methods for generating physical covariate hindcasts (e.g., water temperature hindcasts), we also generated *G. echinulata* density hindcasts using the observed values of drivers in 2015–2016, and compared the results to hindcasts generated using hindcasted drivers (Appendix S1: Table S7; Data S1:

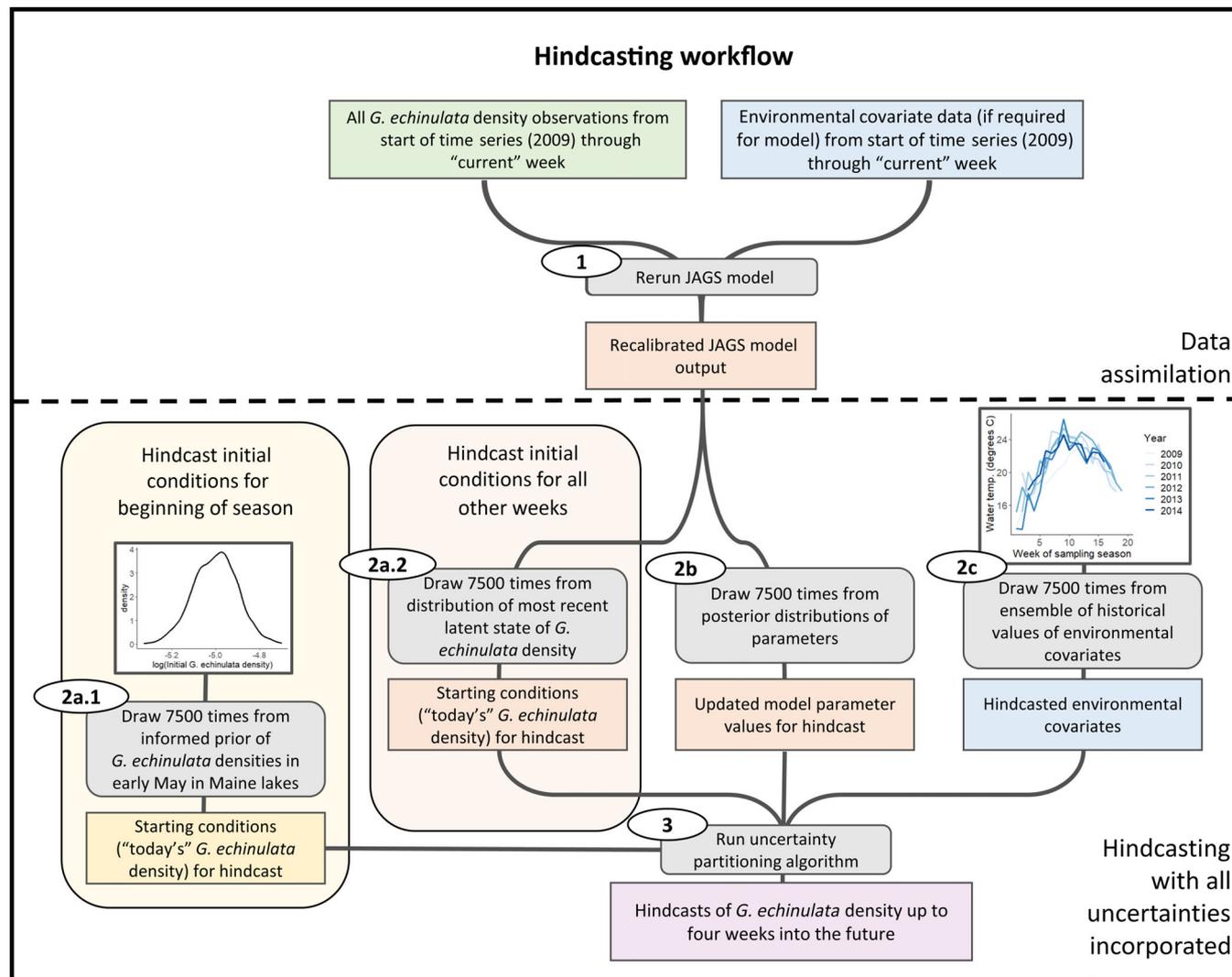


FIGURE 3 Hindcasting workflow diagram. The dotted line indicates the transition between the data assimilation (updating model fits) and hindcasting (including uncertainty partitioning) stages of the workflow. Step 1: Data is assimilated by iteratively adding 1 week of *Gloeotrichia echinulata* and physical driver data to the model input dataset and re-running SSMs in JAGS to update parameter estimates and initial state conditions (“today’s” *G. echinulata* density) for hindcasts. Step 2: Draws from the posterior distribution for each week’s updated initial condition (2a.1, 2a.2) and parameters (2b), as well as from the historical driver data time series in 2009–2014 (2c), are used to generate hindcast input data. During the first week of the sampling season initial conditions are drawn from an informed prior using data from Maine lakes (2a.1); during all other weeks of the season the initial conditions are drawn from the most recent estimated latent state of the recalibrated JAGS model output (2a.2). Step 3: Hindcast input data is fed into an uncertainty partitioning algorithm, outside of JAGS, which runs the process model 4 weeks into the future

4.3_Hindcasting/4.3_Hindcasting_known_covars.R). For this method-check exercise, any missing driver data from 2015 to 2016 (5.7% of observations) were gap-filled using the mean of observations during the same week across all previous years.

Following observations that model ensembles can provide more skilled predictions than a single model even when some ensemble members are low performing (Johansson et al., 2019), we also used the output of the uncertainty partitioning algorithm to generate a simple, unweighted model ensemble to determine if it could out-

perform our individual models (see Appendix S1: Section S4 and Data S1: 7_Model_ensemble/7_Model_ensemble.R).

Our primary criterion for hindcast model assessment was based on predictive loss, calculated using the root mean square error (RMSE) of predictions and the variance of the predictive interval (defined in Table 1) via the following equation:

$$\text{Predictive loss} = \sqrt{\text{RMSE}^2 + \text{predictive interval variance}}. \tag{4}$$

The model with the smallest predictive loss at a particular forecast horizon indicates the best-performing model at that horizon (Gelfand & Ghosh, 1998). Similar to Akaike Information Criterion (AIC), specific values of predictive loss are not assigned a particular interpretation; rather, there is a range of differences that are meaningful (Clark, 2007; Gelfand & Ghosh, 1998). To assess differences in predictive loss among models, we subtracted the predictive loss of the best-performing model from the predictive loss of all other models to calculate change in predictive loss (ΔPL), with smaller ΔPL indicating better-performing models and the best-fit model having a ΔPL of 0. We note that ΔPL should be interpreted within the context of the magnitude of predictive loss across models (i.e., if predictive loss is high across models, small differences in ΔPL may be less meaningful) as well as the distribution of the predicted variable (in our case, logged *G. echinulata* density; Appendix S1: Figure S2). For our study, we considered a difference in $\Delta\text{PL} > 0.1 \ln(\text{colonies/L})$ to be a meaningful difference between models.

To further assess model skill, we also calculated the RMSE, standard deviation of the predictive interval (predictive SD), the percent of observations falling within the 95% predictive interval (coverage), the mean difference between median predicted and observed values (bias), and the difference in model timesteps between when maximum *G. echinulata* density was observed during the hindcasting period and when each model predicted maximum *G. echinulata* density (peak timing, where -1 indicates the model predicted maximum *G. echinulata* density one timestep early and 1 indicates the model predicted maximum density one timestep late; Table 3; Data S1: 6_Output_analysis/6B_Hindcast_output_analysis.R).

Uncertainty partitioning of 2015–2016 hindcasts

We used the uncertainty partitioning algorithm to conduct uncertainty partitioning of our 2015–2016 cyanobacterial density hindcasts via a one-at-a-time-ahead approach, where all sources of uncertainty were initially held at fixed values and then sequentially added back into the hindcasts (Data S1: 4.4_Uncertainty_partitioning/4.4_Uncertainty_partitioning.R). For example, all model parameter values were initially set to the mean of the posterior distribution of the re-fitted model for all 7500 hindcasting iterations; then, when we wanted to add parameter uncertainty to our hindcasts, we allowed parameter values to be drawn from the full posterior distribution, resulting in a variety of possible parameter values and subsequent estimation of uncertainty in those parameters. We added sources of uncertainty

to our hindcasts in the following order: *G. echinulata* starting conditions (initial condition) uncertainty, parameter uncertainty, driver data uncertainty, and process uncertainty. The order of uncertainties is important because different sources of uncertainty can interact with each other (Dietze, 2017b). We were then able to calculate the relative contribution of each uncertainty source to total hindcast variance based on the incremental increase in variance as each source of uncertainty was added. Not all models included all the potential sources of uncertainty (e.g., the random walk model does not have driver data uncertainty because it does not include any physical covariates).

Observation uncertainty is not included in our partitioning results because it does not propagate and therefore does not affect uncertainty about the latent state of the system (Dietze, 2017a). However, to examine the relative importance of observation error in our study system, we assessed the estimated value of τ_{obs} , which is the precision ($1/\text{SD}^2$) of the normal distribution used to fit *G. echinulata* latent states to *G. echinulata* observations in the data model component of our SSMS (Figure 1; Equation 1).

All novel code associated with our analysis is published via Zenodo and included in the code supplement to this manuscript (10.5281/zenodo.5189281; Data S1).

RESULTS

Variability in *G. echinulata* density

Median *G. echinulata* density during the summertime study period from 2009 to 2016 was 0.25 ± 8.2 colonies/L (median \pm SD; Figure 4). During the model fitting period (2009–2014), *G. echinulata* density ranged from an annual maximum density of 1.2 colonies/L in 2012 to 81.6 colonies/L in 2013. Notably, while the model fitting years included two periods of high *G. echinulata* density with visible surface scums (44.3 colonies/L in August 2010 and 81.6 colonies/L in September 2013), maximum density during the 2015–2016 hindcasting period was only 14.1 colonies/L (Figure 4). Visualizations of physical drivers of *G. echinulata* density included in SSMS are provided in the Supplement (Appendix S1: Figures S3–S10).

Assessment of models fit to *G. echinulata* data

Gloeotrichia echinulata growth was dependent on *G. echinulata* density at the previous timestep, as indicated by a converged coefficient value ranging from 0.63 to 0.83 ± 0.04 to 0.10 (mean \pm SD) for the β_1 parameter

TABLE 3 Hindcasting results across models for the 2015–2016 hindcasting period

Model name	RMSE, ln (colonies/L)		Predictive SD, ln (colonies/L)		Predictive loss, ln (colonies/L)		Δ predictive loss (ΔPL), ln (colonies/L)		Coverage (%)		Peak timing (timesteps)		Bias (colonies/L)	
	1 week	4 weeks	1 week	4 weeks	1 week	4 weeks	1 week	4 weeks	1 week	4 weeks	1 week	4 weeks	1 week	4 weeks
RW	1.89	2.23	1.63	3	2.5	3.74	0.25	1.36	97.2	100	1	14	-0.43	-0.97
OffsetRW	1.88	2.1	1.62	2.99	2.49	3.65	0.24	1.27	97.2	100	1	14	-0.04	0.26
AC	1.68	1.5	1.5	2.21	2.26	2.67	0.01	0.29	100	100	1	14	-0.65	-1.18
BaseLM	1.67	1.61	1.51	2.09	2.25	2.64	0	0.26	94.4	100	1	14	-0.92	-1.51
MinWaterTemp	1.81	1.6	1.43	1.84	2.31	2.43	0.06	0.05	94.4	96.8	14	12	-0.93	-1.41
MinWaterTempLag	1.79	1.63	1.45	1.8	2.31	2.43	0.06	0.05	97.2	87.1	14	12	-1	-1.45
WaterTempMA	1.78	1.6	1.44	1.83	2.3	2.43	0.05	0.05	94.4	90.3	14	12	-0.95	-1.43
SchmidtLag	1.74	1.59	1.46	2.04	2.27	2.59	0.02	0.21	97.2	100	14	12	-0.9	-1.42
WindDir	1.77	1.55	1.5	1.99	2.32	2.52	0.07	0.14	94.4	100	1	14	-0.96	-1.51
GDD	1.83	1.59	1.43	1.84	2.32	2.43	0.07	0.05	94.4	96.8	14	14	-1.08	-1.4
Schmidt+Wind	1.82	1.55	1.46	1.98	2.33	2.51	0.08	0.13	94.4	100	14	13	-0.93	-1.43
Temp+Wind	1.88	1.56	1.43	1.8	2.37	2.38	0.12	0	91.7	93.5	14	14	-0.95	-1.42
Wind+GDD	1.88	1.55	1.43	1.82	2.36	2.39	0.11	0.01	91.7	93.5	14	13	-0.92	-1.35
Ensemble	1.80	1.56	1.46	1.89	2.32	2.45	0.07	0.07	97.2	96.8	14	14	-0.97	-1.43

Notes: RMSE, root mean square error; predictive variance, mean variance of the predictive interval; predictive loss, $\sqrt{\text{RMSE}^2 + \text{predictive variance}}$; Δ predictive loss, the difference between predictive loss for each model and the best-performing model for that forecast horizon; coverage, the percentage of observations falling within the 95% predictive interval; peak timing, the number of weeks between peak *G. echinulata* density during the hindcasting period and when the model predicted peak density; bias, mean difference between median predicted and observed values. Note that all assessment metrics are conducted on log-transformed data except for mean bias.

across models (Appendix S1: Table S5; see Table 2 for description of β_1 parameter). Parameter estimates from fitted models indicated that *G. echinulata* growth was positively associated with increases in water temperature, high Schmidt stability, and a higher daily proportion of wind blowing towards the focal nearshore site (see Appendix S1: Tables S5 and S6 for model coefficient values). The coefficient on the quadratic term for growing degree days based on water temperature (GDD) converged at -0.58 ± 0.17 (Appendix S1: Table S6), indicating that increases in GDD at high values (i.e., late in the sampling season) were associated with decreasing *G. echinulata* growth.

Some variables that seemed promising based on the a priori covariate selection protocol had estimated model coefficients close to 0 in SSMs fit to 2009–2014 data (Precip, ΔSchmidt; Appendix S1: Table S6), indicating a limited association with *G. echinulata* growth. As a result, these models were excluded from the hindcasting analysis.

Using covariates that had non-zero coefficients during fitting of single-covariate models (MinWaterTemp, SchmidtLag, WindDir, GDD), we subsequently fit 5 two-covariate models (Schmidt+Temp, Temp+Wind, Schmidt+Wind, Schmidt+GDD, Wind+GDD). We were unable to estimate non-zero coefficients for Schmidt

stability in the Schmidt+Temp and Schmidt+GDD models (Appendix S1: Table S6), and so these two models were excluded from the hindcasting analysis.

Physical drivers did not improve on BaseLM model for 1-week-ahead hindcasts

All single and two-covariate models as well as the BaseLM and AC models had improved performance over the baseline RW and OffsetRW models for 1-week-ahead hindcasts based on predictive loss and ΔPL (Table 3). The BaseLM model had the smallest predictive loss (indicating best performance and a ΔPL of 0 ln(colonies/L)) and the lowest RMSE (Table 3 and Figure 5; models not shown in Figure 5 can be found in Appendix S1: Figures S11 and S12). The baseline AC model also performed well at the 1-week horizon (ΔPL of 0.01 ln (colonies/L)) and had higher coverage and lower bias than the BaseLM model, demonstrating the importance of an autocorrelation term in predicting *G. echinulata* density. The magnitude of ΔPL between models with covariates and the best-performing BaseLM model were mostly small (ΔPL < 0.1 ln(colonies/L)), indicating that models with covariates performed similarly to the BaseLM model at the 1-week horizon (i.e., addition of explanatory

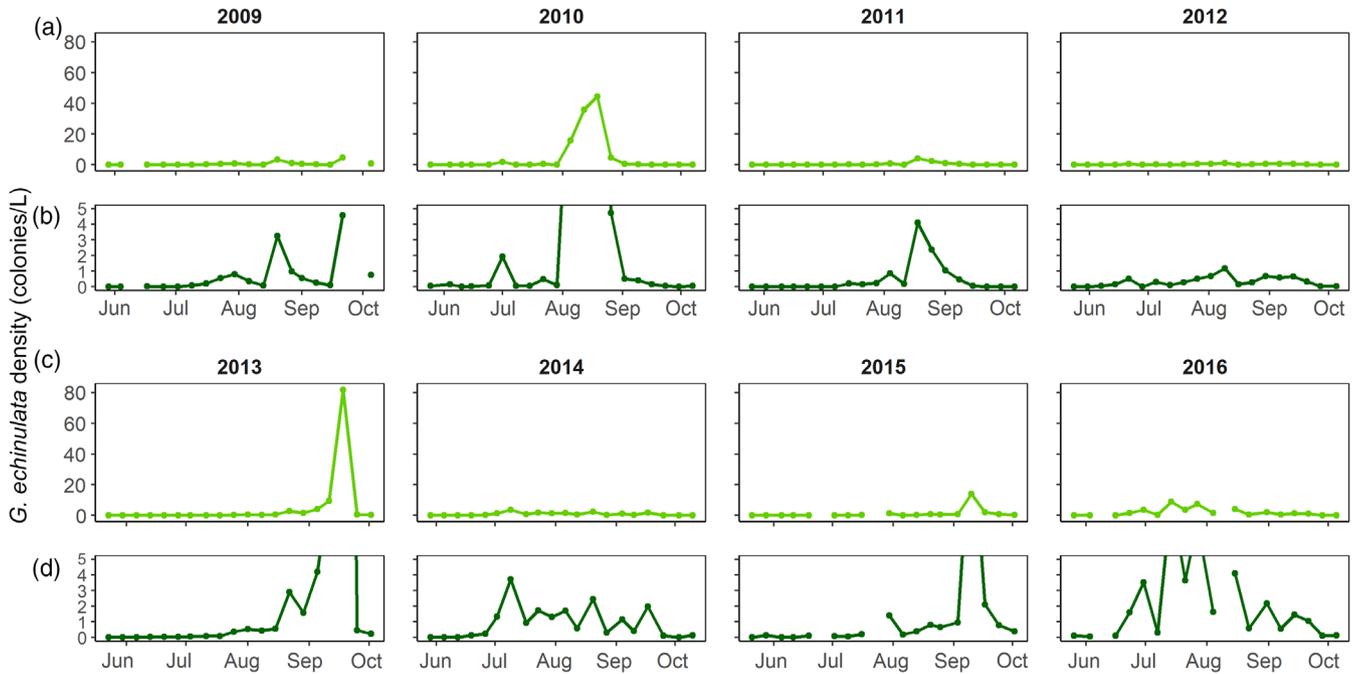


FIGURE 4 Time series of *G. echinulata* density at site 1 in Lake Sunapee from 2009 to 2016 (a, c); panels (b) and (d) show a reduced scale to better illustrate variability at low density

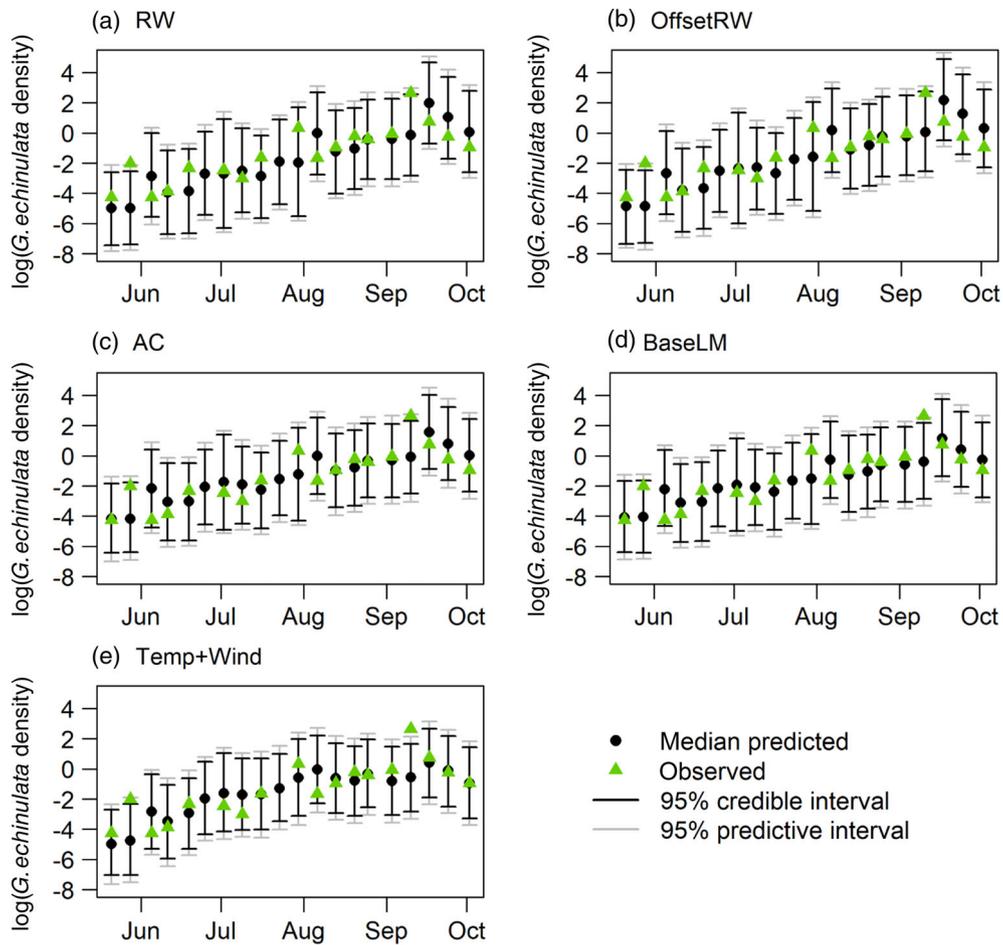


FIGURE 5 Time series of median predicted and observed *G. echinulata* density (colonies/L) for 1-week-ahead hindcasts in 2015 for null models (a–d) and best-performing models (d–e, as the null BaseLM model was also a best-performing model; Table 3). Similar figures for 2016 hindcasts and models not shown here may be in found in the supplemental material (Appendix S1: Figures S11, S12 and S15)

physical covariates did not improve hindcast skill). The exceptions were the Temp+Wind and Wind+GDD models, which had decreased hindcast skill compared to the BaseLM model (Δ PL of 0.11 and 0.12 ln(colonies/L), respectively; Table 3).

No model correctly predicted the week or magnitude of peak *G. echinulata* density during the hindcasting period, which was observed on 10 September 2015. The four baseline models (RW, OffsetRW, AC, and BaseLM) and one covariate model (WindDir) predicted peak density one timestep after it occurred (i.e., peak timing = 1, meaning the model predicted peak density 1 week late; Table 3), likely due to dependency of *G. echinulata* density predictions on density at the previous timestep.

Relative model performance at the 1-week horizon did not change when hindcasts were conducted using observed values of drivers from 2015 to 2016 rather than hindcasted drivers; the BaseLM and AC models were still the best-performing models (Appendix S1: Table S7).

Water temperature models more skilled than BaseLM at 4-week forecast horizon

Models containing water temperature covariates outperformed the BaseLM model at the 4-week horizon (Table 3; Figure 6; models not shown in Figure 6 may be found in Appendix S1: Figures S13 and S14). The best-performing model according to Δ PL at the 4-week horizon was Temp+Wind. While some models had a lower RMSE than Temp+Wind (AC, WindDir, Schmidt+Wind, Wind+GDD), Temp+Wind had the lowest predictive SD, indicating a narrower predictive interval around the median predicted value and therefore the smallest predictive loss (Figure 6 and Table 3). Other models containing water temperature covariates (MinWaterTemp, MinWaterTempLag, WaterTempMA, GDD, Wind+GDD) also performed well at the 4-week horizon, all with Δ PL ≤ 0.05 ln(colonies/L) (Table 3). Covariate models containing no water temperature covariates (WindDir, SchmidtLag, Schmidt+Wind) had

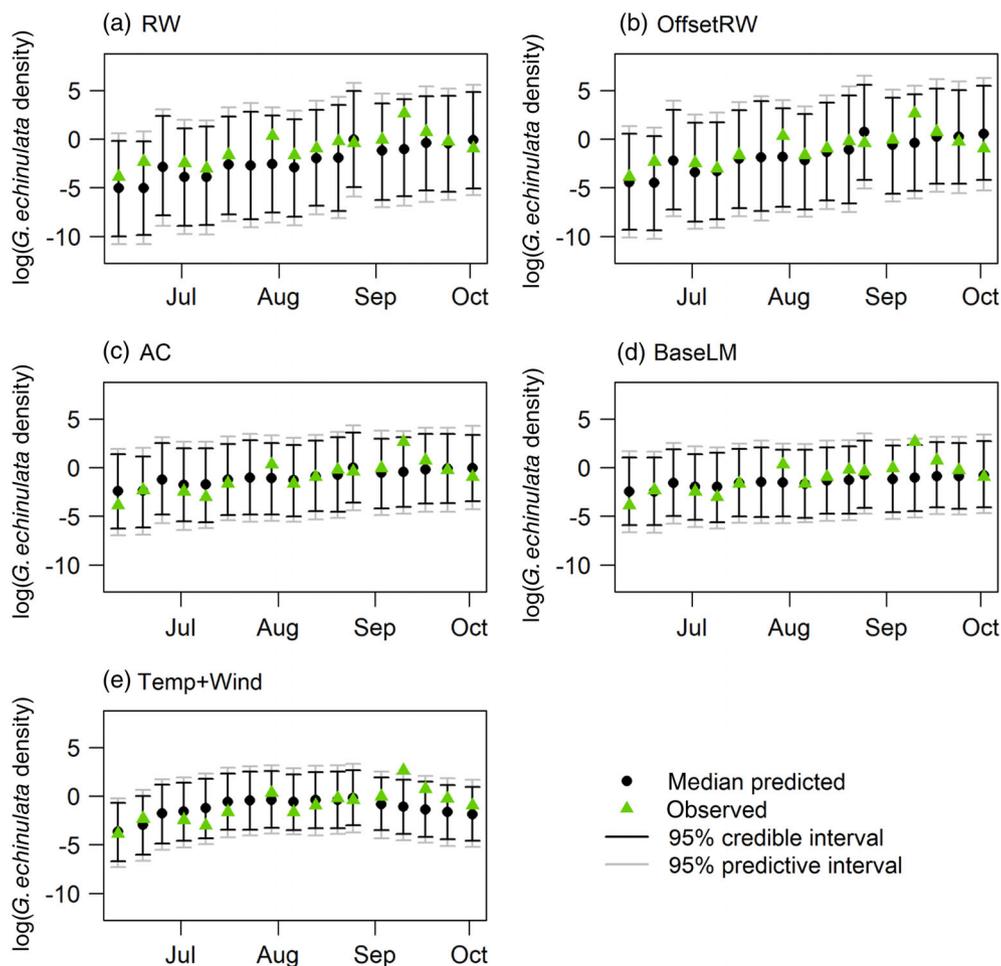


FIGURE 6 Time series of median predicted and observed *G. echinulata* density (colonies/L) for 4-week-ahead hindcasts in 2015 for null models (a–d) and best-performing models (d–e, as the null BaseLM model was also a best-performing model; Table 3). Similar figures for 2016 hindcasts and models not shown here may be in found in the supplemental material (Appendix S1: Figures S13, S14 and S15). Note the y-axis change between Figures 5 and 6 to accommodate larger credible and predictive intervals at the 4-week forecast horizon

higher predictive SD (i.e., wider predictive intervals) and therefore decreased performance (all $\Delta\text{PL} > 0.1 \ln(\text{colonies}/\text{L})$) compared to water temperature covariate models. All four baseline models (RW, OffsetRW, AC, BaseLM) had a $\Delta\text{PL} > 0.2 \ln(\text{colonies}/\text{L})$ at the 4-week horizon.

Despite the improvement of water temperature models over the BaseLM model, no model successfully predicted the week of peak *G. echinulata* density at the 4-week horizon (Figure 6). Patterns in predictive skill across models at the 4-week horizon did not change when hindcasts were generated using observed values for drivers, with one exception: models containing GDD as a covariate were relatively less skilled when hindcasts used known rather than hindcasted drivers as input (Appendix S1: Table S7). The unweighted model ensemble did not improve on the top-performing models at either the 1-week or 4-week forecast horizon, with a ΔPL of $0.07 \ln(\text{colonies}/\text{L})$ at both horizons (Table 3; Appendix S1: Section S4, Figure S15).

Process uncertainty dominated hindcast credible intervals

Process error represented the largest proportion of uncertainty for all models, which is consistent with our finding that models both with and without covariates had little predictive power (i.e., high predictive loss). The proportion of variance attributed to process uncertainty increased with hindcast horizon, coincident with a reduction in initial conditions uncertainty (Figure 7; models not shown in Figure 7 can be found in Appendix S1: Figure S16), as expected due to the decreasing influence of initial conditions as the forecast horizon increases (Dietze, 2017a). Neither increases in model structural complexity nor differences in model covariates substantially decreased the proportional contribution of process uncertainty (Figure 8). The mean contribution of process uncertainty across the hindcasting period ranged from 71% of hindcast uncertainty in the OffsetRW model to

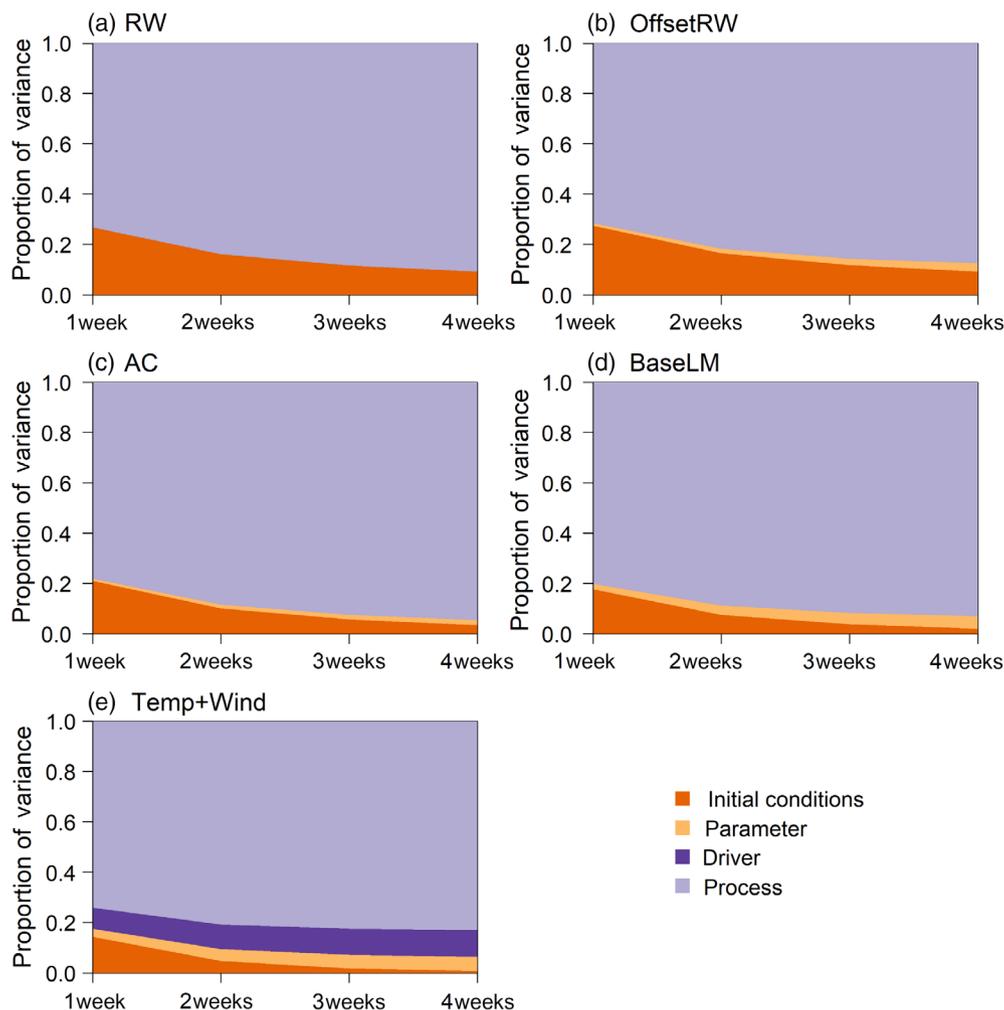


FIGURE 7 Uncertainty partitioning of the 1-week-ahead to 4-week-ahead hindcasts averaged across 2015–2016 for null models (a–d) and best-performing models (d–e; Table 3). Similar figures for other models may be found in the supplemental material (Appendix S1: Figure S16)

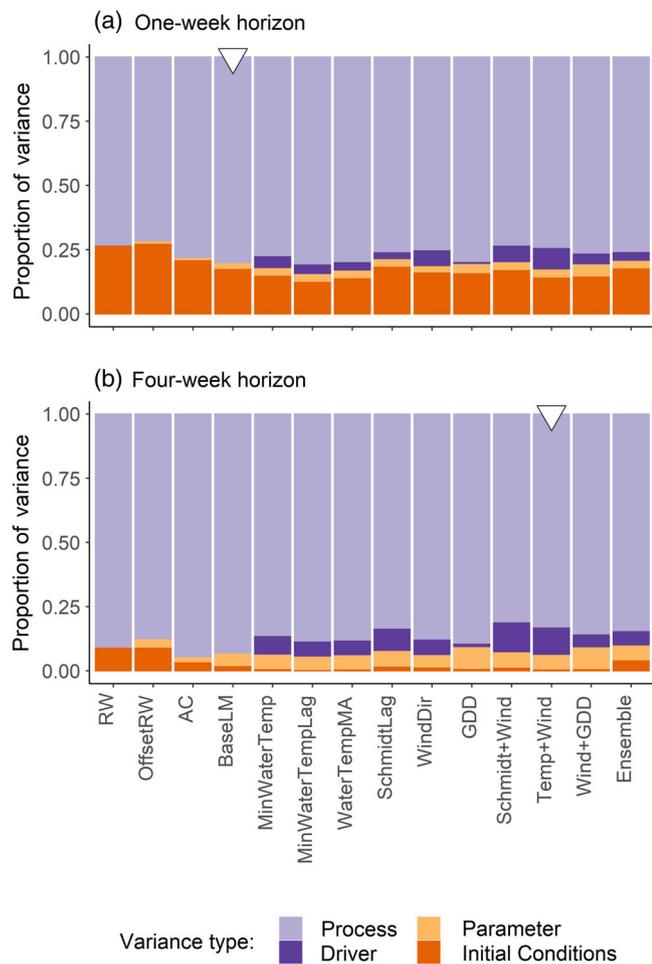


FIGURE 8 Uncertainty partitioning for (a) 1-week-ahead and (b) 4-week-ahead hindcasts averaged across the 2015–2016 hindcasting period across models. White triangles indicate a best-performing model at the respective forecast horizon as assessed by change in predictive loss (Δ PL; Table 3)

81% in the MinWaterTempLag model for 1-week-ahead hindcasts, and from 81% in the Schmidt+Wind model to 94% in the AC model for 4-week-ahead hindcasts. The relative contribution of process uncertainty to total hindcast uncertainty varied across the hindcasting period for individual models but was never <40% at any timestep for any model (summary statistics of the contributions of all uncertainty sources during 2015–2016 can be found in Appendix S1: Tables S8 and S9).

The second largest component of uncertainty in hindcasts was due to estimation of initial conditions (i.e., the log-transformed density of *G. echinulata* on the hindcast issue date), although this source of uncertainty quickly declined to negligible levels by the 4-week-ahead forecast horizon for all models (Figures 7 and 8). Averaged across the hindcasting period, initial conditions uncertainty contributions ranged from 13% (MinWaterTempLag) to 28% (OffsetRW) of the uncertainty for 1-week-ahead

hindcasts but comprised only 1% to 9% of total uncertainty for 4-week ahead hindcasts. As expected, initial conditions uncertainty was largest (31%–59% of total uncertainty) for 1-week-ahead hindcasts following a week with a missing *G. echinulata* observation, as no data were available to constrain the estimated latent state (e.g., Appendix S1: Figure S17). Parameter and driver error had negligible contributions to total hindcast uncertainty for both 1-week-ahead and 4-week-ahead hindcasts (Figures 7 and 8; Appendix S1: Tables S8 and S9).

Observation uncertainty

Observation uncertainty (τ_{obs}) was substantial for all models, ranging from 1.70 to 1.81 $\ln(\text{colonies/L})^{-2}$ across models (Appendix S1: Table S5), consistent with expectations as phytoplankton count data are notoriously variable (Rott et al., 2007; Vuorio et al., 2007). The estimated values of τ_{obs} correspond to a standard deviation of $\sim 0.75 \ln(\text{colonies/L})$, which can be interpreted as a factor of ~ 2.1 colonies/L difference between latent (predicted) and observed values.

DISCUSSION

Understanding ecological systems to better forecast future events is a critical challenge for managing resources and public health. Use of standardized ecological forecasting approaches such as fully specified uncertainty partitioning provides a much-needed framework for prioritizing research efforts to meet this challenge (Dietze et al., 2018). While there are numerous hypotheses and studies linking physical drivers to the *G. echinulata* surface scums (Carey et al., 2014; Hyenstrand et al., 2000; Istvánovics et al., 1993; Karlsson-Elfgren et al., 2005; Napiórkowska-Krzebietke & Hutorowicz, 2015; Roelofs & Oglesby, 1970) that make water quality management challenging in low-nutrient lakes, few have fully evaluated the predictive influence of these physical variables.

While models containing information about water temperature converged on non-zero covariate coefficients during model fitting and out-performed the null models at the 4-week hindcast horizon, even these models had low forecasting skill, particularly when *G. echinulata* densities were high. Our results reiterate that significant explanatory variables in model fitting are not necessarily effective predictive variables in near-term ecological forecasts (following Ward et al., 2014), and that models that adequately capture low densities may not successfully predict rare high-density events. The dominance of process uncertainty in our hindcasts emphasizes that

G. echinulata densities are probably a product of both growth and movement of colonies, and that data collection to facilitate addition of process model structure related to *G. echinulata* life history and biogeochemical processes might be valuable in future forecasting efforts. The importance of initial conditions uncertainty in our study indicates that spatial and temporal misalignment of driver data and density observations are ongoing challenges in this forecasting system, and that the imperfect observation of both *G. echinulata* density and physical covariates substantially affects forecast skill.

Of all the physical covariates we examined, water temperature was important in both fitted and hindcast models and remains a promising driver for predicting *G. echinulata* density. Water temperature was positively associated with *G. echinulata* density across models (see Appendix S1: Tables S5 and S6 for coefficient estimates). In addition, models containing water temperature covariates were more skilled than other models at the 4-week hindcast horizon (all water-temperature-containing models had a $\Delta\text{PL} \leq 0.05$; Table 3). These findings are consistent with studies demonstrating that cyanobacteria benefit from warmer temperatures (e.g., Carey, Ibelings, et al., 2012; Paerl & Huisman, 2008) and that water temperature is a good predictor of cyanobacterial density (Ho & Michalak, 2020; Rousoo et al., 2020). Our results indicate that a minimum water temperature predictor (e.g., MinWaterTemp, MinWaterTempLag) may be useful for forecasting *G. echinulata* density, consistent with findings from a previous study examining predictors of *Lyngbya majuscula* blooms in an Australian bay (Hamilton et al., 2009). However, we were unable to identify any physical covariates that improved *G. echinulata* density predictions over the BaseLM model at the 1-week horizon, indicating that water temperature is likely not adequate to forecast cyanobacterial densities at this time scale.

Process uncertainty dominated hindcast uncertainty across all models. Neither increases in model structural complexity nor differences in model covariates substantially decreased the proportional contribution of process uncertainty to forecast uncertainty. The predominance of process uncertainty, coupled with low parameter uncertainty (Figure 8), indicates a substantial need for both field research and modeling studies to better understand and represent *G. echinulata* biology. Some of the physical covariates we explored may sufficiently explain weekly differences in the most frequently observed low densities, but none of the models we fit had skill at forecasting peak abundances, which appeared and declined rapidly relative to observation frequency (i.e., high-density events seldom lasted more than 1 week). In theory, it is possible that *G. echinulata* dynamics are dominated by stochasticity (e.g., Carpenter et al., 2020), in which case improved understanding of the

process structure would not effectively reduce process uncertainty. However, our results suggest that additional process model structure related to the biology of the focal cyanobacterium could improve forecast skill.

Because cyanobacteria can quickly increase from a low-density to a high-density state, applying a switching state-space or stochastic volatility model (Achcar et al., 2011; Holan et al., 2009) that incorporates the transition probability from a low-density to a high-density state might improve *G. echinulata* density forecasts. This transition probability can vary through time (e.g., Pinto & Spezia, 2016) according to the seasonal probability of observing a high-density event, and such a model structure would permit relaxation of the assumption that process uncertainty is constant. Given the ecology of *G. echinulata* in north temperate lakes, where blooms are more likely in the summer months and improbable in winter, allowing for variability in state transition probability and process uncertainty would be a potentially valuable feature in future models.

Tradeoffs in the relative importance of driver and process uncertainty in the progression of our hindcasts suggest hypotheses about missing or misrepresented processes that could be investigated in future studies. For example, several of the top-performing models at the 4-week horizon (Temp+Wind, MinWaterTemp, MinWaterTempLag, WaterTempMA) exhibited low driver uncertainty but high process uncertainty during the last several weeks of the 2015 sampling season (Appendix S1: Figure S18). This is likely because while water temperatures decreased during this time (Appendix S1: Figures S3–S5), *G. echinulata* density increased (Figure 4), which is the opposite of the relation specified by our process model, where water temperature is positively correlated with density. Consequently, we can hypothesize that either (A) the relation between density and water temperature changes during the late summer and fall or (B) some other biological or ecological process (e.g., germination of dormant colonies; see Carey et al., 2014, or phenological senescence; see Winder & Cloern, 2010) is driving *G. echinulata* density during this part of the year. Our first hypothesis could be tested by allowing the parameters governing the relation between density and candidate environmental covariates (water temperature, wind) to vary through time (e.g., Nessler & Wilberg, 2019) or by specifying a nonlinear relation between water temperature and density. The uncertainty partitioning time series for the GDD and Wind+GDD models provide some support for including a non-linear water temperature relation to capture end-of-season densities, as process uncertainty for these models is relatively lower at the end of the 2015 sampling season (Appendix S1: Figure S18). In the future, our second hypothesis might be addressed by collecting more detailed

biological data to build a model incorporating additional life history stages beyond vegetative growth in the water column. For example, it is well-documented that recruitment from the sediments to the pelagic zone is an important life stage for *G. echinulata*, potentially contributing 4%–40% of the water column population each week (e.g., Barbiero & Welch, 1992; Carey et al., 2014). Including processes or covariates at multiple time lags representing recruitment of *G. echinulata* from the sediments may improve water column abundance forecasts. Finally, including biogeochemical processes in addition to physical processes in the model structure may help further explain variation in *G. echinulata* abundance. Unfortunately, we did not have enough biogeochemical data available to constrain these processes and covariates.

While the contribution of driver data uncertainty (accuracy of driver measurements and forecasts) to our hindcasts was small, spatial mismatches between driver data and response variable data may also contribute to process uncertainty. Thus, the inclusion of more variables collected close to our nearshore sampling sites, rather than variables collected in the deep-water pelagic zone, might reduce process uncertainty by better characterizing the effect of physical drivers on localized nearshore processes. In addition, local site variables have been found to be important in driving benthic recruitment (Carey et al., 2014), so inclusion of these variables could be a complementary approach to including benthic recruitment in models. Alternatively, the relatively low contribution of driver data uncertainty to hindcasts could be due, in part, to our method of hindcasting physical covariates using an ensemble of historical observations. While unavailable for our study system, near-term water temperature forecasts generated using numerical simulation models (e.g., Thomas et al., 2020) represent an alternative method for hindcasting physical covariates while retaining temporal autocorrelation in hindcasted driver data.

As with weather forecasting, initial conditions estimation was an important component of forecast uncertainty in this system. This suggests that increasing the spatial or temporal frequency of observations could improve forecast skill (e.g., Fox et al., 2018), as cyanobacterial densities can be spatially heterogeneous (Franks, 1997; Serizawa et al., 2008; Wynne & Stumpf, 2015) and change quickly on short timescales (Dokulil & Teubner, 2000; Huisman & Hulot, 2005; Rolland et al., 2013). While we used data from a second nearshore sampling site to constrain our observation error distribution, future research could incorporate data from other sampling sites into a hierarchical SSM to predict local *G. echinulata* densities at multiple sites. A hierarchical SSM would allow for spatial relationships between sites, thereby potentially reducing uncertainty due to missing observations at an individual site, as

information on states or parameters is “borrowed” from nearby sites (Clark, 2003). A multi-site model would also provide additional information to LSPA stakeholders regarding nearshore cyanobacterial high-density events in different coves across Lake Sunapee.

Because sampling and counting *G. echinulata* is labor-intensive, increasing observational frequency might necessitate assimilating other measures of cyanobacterial abundance into forecasts, such as fluorescence-based biomass measurements (e.g., Catherine et al., 2012) and spectrophotometric pigment analysis (e.g., Küpper et al., 2007; Thrane et al., 2015). Furthermore, as phytoplankton counts are notoriously variable (Rott et al., 2007; Vuorio et al., 2007), increased spatiotemporal sampling frequency and incorporation of measures of cyanobacterial abundance besides counts might constrain the high observation uncertainty in *G. echinulata* density data, thereby improving comparisons of models to data. However, before investing in costly increased in situ monitoring, the potential benefit of increased sampling effort could be determined through simulated data experiments exploring how different sampling techniques and frequencies affect forecast precision (following Dietze, 2017a).

Due to the myriad modeling, assessment, and uncertainty analysis approaches that are applied across the ecological literature, intercomparison among modeling and forecasting studies is difficult even among studies that focus on the same response variable (e.g., Rigosi et al., 2010; Rouso et al., 2020), let alone across ecosystems or response variables. However, preliminary commonalities between our oligotrophic lake cyanobacterial density hindcasts and other uncertainty partitioning analyses do emerge. For example, our hindcasts were dominated by process uncertainty and similar results have been reported for ecological forecasts ranging from forest biomass and productivity (Thomas et al., 2018; Valle et al., 2009) to vertebrate species distributions (Diniz-Filho et al., 2009; Watling et al., 2015). In addition, our finding that uncertainty in the initial condition of *G. echinulata* density at the start of a forecast is an important contributor to forecast uncertainty is consistent with terrestrial carbon forecasts at the annual scale (Fox et al., 2018) and lake chlorophyll *a* forecasts at the weekly scale (Huang et al., 2013). However, differences in modeling approach or methods of uncertainty quantification preclude definitive or quantitative comparisons across these studies.

Access to data and standardized expectations for uncertainty partitioning are critical to the iterative improvement of forecast skill. Our study was enabled both by collaborative sharing of long-term data through the Global Lake Ecological Observatory Network, which facilitated fitting and validation of hindcasting models

over many years (Cottingham et al., 2020a, 2020b; LSPA, Weathers, et al., 2020; LSPA, Steele, et al., 2020), and access to publicly available R code examples of how to conduct uncertainty partitioning (https://github.com/EcoForecast/EF_Activities). As such, our study illustrates the importance of open science and findable, accessible, interoperable, and reusable (FAIR) scientific practices with respect to data and code (Powers & Hampton, 2019; Wilkinson et al., 2016) to reduce barriers to adoption of techniques such as uncertainty partitioning and advance the field of ecological forecasting.

Developing forecasts for cyanobacterial density is especially challenging and uncertainty partitioning in these highly dynamic systems can help prioritize research to improve process understanding or sampling design. Overall, despite considering dozens of possible environmental covariates, our hindcasts were not skilled enough to predict the sudden, infrequent increases in cyanobacterial density that cause concern for water resource managers and other stakeholders in both oligotrophic and eutrophic lakes. While the difficulty in predicting near-term cyanobacterial densities is well-established, our work highlights the utility of formal uncertainty partitioning for providing insight on how to target data collection and modeling efforts to improve forecast skill, following Dietze et al. (2018). Our study determined that water temperature is a promising driver for forecasting cyanobacterial densities, and that inclusion of additional process model structure to better represent *G. echinulata* biology could improve forecast skill by reducing process uncertainty. In addition, our findings indicate that increased observational frequency, addition of alternate measures of *G. echinulata* density, or incorporation of multiple sampling sites into a single model could improve forecasts by reducing initial conditions uncertainty. Even if our initial forecasting efforts are not very skilled, the process of iteratively confronting our models with data and quantitatively examining forecast uncertainty provides a path for improvement (Bauer et al., 2015), equipping ecologists with the information they need to ultimately develop actionable forecasts to improve resource management.

AUTHOR CONTRIBUTIONS

Mary E. Lofton led development of Bayesian state-space models with Shannon L. LaDeau and conducted hindcasts and variance partitioning, with substantial contributions from Jennifer A. Brentrup, Whitney S. Beck, Jacob A. Zwart, and Michael C. Dietze. Kathryn L. Cottingham, Holly A. Ewing, Kathleen C. Weathers, and Cayelan C. Carey contributed data. Bethel G. Steele led publication of data sets. All authors provided feedback on the design and scope of the modeling effort, the interpretation of results, and edited and approved the final version of the manuscript.

ACKNOWLEDGMENTS

The study was initiated as part of a Global Lake Ecological Observatory Network (GLEON) Fellowship Program Bayesian statistics workshop. The long-term Sunapee data collection was funded by the National Science Foundation (NSF DBI-0434684, DEB-0749022, EF-0842267, EF-0842112, EF-0842125, DEB-1010862, CNS-1737424, ICER-1517823), the Lake Sunapee Protective Association, Dartmouth College, Auburn Water District/Lewiston Water Division, and Bates College. The forecasting project was funded by NSF MSB DEB-1638575, EF-1137327, DEB-1702991, and DEB-1638577. Mary E. Lofton and Cayelan C. Carey were supported by NSF CNS-1737424, DEB-1753639, DBI-1933016, and MSB DEB-1926050. Mary E. Lofton was supported by a Virginia Water Resources Research Center William R. Walker Graduate Research Fellow Award and a Virginia Tech College of Science Make-A-Difference Roundtable Scholarship. Jennifer A. Brentrup was supported by funding from the Ecology, Evolution, Environment, and Society program at Dartmouth College and LSPA's Midge Eliassen Fellowship. The Eliassen and Fichter families provided generous access to study sites. Special thanks to all of the participants in the December 2017 GLEON Bayesian statistics workshop for idea development, to John Foster and participants in the January 2019 Near-term Ecological Forecasting Initiative workshop for assistance with model development, to the Spring 2020 Ecological Forecasting seminar class at Virginia Tech for help with literature review, and to Hilary Dugan and Nicole Ward for help downloading and processing NLDAS data. GLEON and Ecological Forecasting Initiative (EFI) colleagues provided helpful comments, and two anonymous Reviewers provided thoughtful reviews that substantially improved both the manuscript and the accompanying code repository. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

CONFLICT OF INTEREST

None.

DATA AVAILABILITY STATEMENT

Data are already published, with those publications properly cited in this submission. Data (Cottingham, 2020; Cottingham et al., 2020a, 2020b; Lake Sunapee Protective Association (LSPA), Steele, et al., 2020; Lake Sunapee Protective Association (LSPA), Weathers, et al., 2020; Lofton et al., 2022; Steele et al., 2021) are available from the Environmental Data Initiative and Zenodo. All novel code associated with this manuscript is already published at the following DOI via Zenodo: 10.5281/zenodo.5189281 (<https://doi.org/10.5281/zenodo.5189281>) and is also available in a compressed zip file (Data S1) associated with this submission.

ORCID

Mary E. Lofton  <https://orcid.org/0000-0003-3270-1330>

Jennifer A. Brentrup  <https://orcid.org/0000-0002-4818-7762>

Whitney S. Beck  <https://orcid.org/0000-0002-0306-6086>

Jacob A. Zwart  <https://orcid.org/0000-0002-3870-405X>

Ruchi Bhattacharya  <https://orcid.org/0000-0001-5657-9603>

Ian M. McCullough  <https://orcid.org/0000-0002-6832-674X>

Bethel G. Steele  <https://orcid.org/0000-0003-4365-4103>

Cayelan C. Carey  <https://orcid.org/0000-0001-8835-4476>

Kathryn L. Cottingham  <https://orcid.org/0000-0003-3169-5047>

Michael C. Dietze  <https://orcid.org/0000-0002-2324-2518>

Holly A. Ewing  <https://orcid.org/0000-0001-7077-8473>

Kathleen C. Weathers  <https://orcid.org/0000-0002-3575-6508>

Shannon L. LaDeau  <https://orcid.org/0000-0003-4825-5435>

REFERENCES

- Achcar, J. A., E. R. Rodrigues, and G. Tzintzun. 2011. "Using Stochastic Volatility Models to Analyse Weekly Ozone Averages in Mexico City." *Environmental and Ecological Statistics* 18: 271–90.
- Barbiero, R. P., and E. B. Welch. 1992. "Contribution of Benthic Blue-Green Algal Recruitment to Lake Populations and Phosphorus Translocation." *Freshwater Biology* 27: 249–60.
- Bauer, P., A. Thorpe, and G. Brunet. 2015. "The Quiet Revolution of Numerical Weather Prediction." *Nature* 525: 47–55.
- Bormans, M., P. W. Ford, and L. Fabbro. 2005. "Spatial and Temporal Variability in Cyanobacterial Populations Controlled by Physical Processes." *Journal of Plankton Research* 27: 61–70.
- Carey, C. C., H. A. Ewing, K. L. Cottingham, K. C. Weathers, R. Q. Thomas, and J. F. Haney. 2012. "Occurrence and Toxicity of the Cyanobacterium *Gloeotrichia echinulata* in Low-Nutrient Lakes in the Northeastern United States." *Aquatic Ecology* 46: 395–409.
- Carey, C. C., B. W. Ibelings, E. P. Hoffmann, D. P. Hamilton, and J. D. Brookes. 2012. "Eco-Physiological Adaptations that Favour Freshwater Cyanobacteria in a Changing Climate." *Water Research* 46: 1394–407.
- Carey, C. C., K. C. Weathers, and K. L. Cottingham. 2008. "*Gloeotrichia echinulata* Blooms in an Oligotrophic Lake: Helpful Insights from Eutrophic Lakes." *Journal of Plankton Research* 30: 893–904.
- Carey, C. C., K. C. Weathers, H. A. Ewing, M. L. Greer, and K. L. Cottingham. 2014. "Spatial and Temporal Variability in Recruitment of the Cyanobacterium *Gloeotrichia echinulata* in an Oligotrophic Lake." *Freshwater Science* 33: 577–92.
- Carlson, R. E. 1977. "A Trophic State Index for Lakes." *Limnology and Oceanography* 22: 361–9.
- Carlson, R., and J. Simpson. 1996. *A Coordinator's Guide to Volunteer Lake Monitoring Methods*. Madidson, WI: North American Lake Management Society.
- Carpenter, S. R., B. M. S. Arani, P. C. Hanson, M. Scheffer, E. H. Stanley, and E. Van Nes. 2020. "Stochastic Dynamics of Cyanobacteria in Long-Term High-Frequency Observations of a Eutrophic Lake." *Limnology and Oceanography Letters* 5: 331–6.
- Catherine, A., N. Escoffier, A. Belhocine, A. B. Nasri, S. Hamlaoui, C. Yéprémian, C. Bernard, and M. Troussellier. 2012. "On the Use of the FluoroProbe[®], a Phytoplankton Quantification Method Based on Fluorescence Excitation Spectra for Large-Scale Surveys of Lakes and Reservoirs." *Water Research* 46: 1771–84.
- Clark, J. S. 2003. "Uncertainty and Variability in Demography and Population Growth: A Hierarchical Approach." *Ecology* 84: 1370–81.
- Clark, J. 2007. *Models for ecological data*. Princeton, NJ: Princeton University Press.
- Clark, J. S., S. R. Carpenter, M. Barber, S. Collins, A. Dobson, J. A. Foley, D. M. Lodge, et al. 2001. "Ecological Forecasts: An Emerging Imperative." *Science* 293: 657–60.
- Cobourn, K. M., C. C. Carey, K. J. Boyle, C. Duffy, H. A. Dugan, K. J. Farrell, L. Fitchett, et al. 2018. "From Concept to Practice to Policy: Modeling Coupled Natural and Human Systems in Lake Catchments." *Ecosphere* 9: 1–15.
- Cottingham, K. L. 2020. "Surface Pelagic Total Nitrogen Concentrations in Lake Sunapee, USA 2009 to 2012." Environmental Data Initiative. <https://doi.org/10.6073/pasta/57282aff9a884449864399012e36abf2>
- Cottingham, K. L., C. C. Carey, and K. C. Weathers. 2020a. "*Gloeotrichia echinulata* Density at Four Nearshore Sites in Lake Sunapee, NH, USA from 2005–2016 ver. 2." Environmental Data Initiative. <https://doi.org/10.6073/pasta/b6f418436088b14666a02467797ff1ad>
- Cottingham, K. L., C. C. Carey, and K. C. Weathers. 2020b. "High-Frequency Temperature Data from Four near-Shore Sites, Lake Sunapee, NH, USA, 2006–2018." Environmental Data Initiative. <https://doi.org/10.6073/pasta/3e325757f0e981d91cd297f257f05f55>
- Cyr, H. 2017. "Winds and the Distribution of Nearshore Phytoplankton in a Stratified Lake." *Water Research* 122: 114–27.
- de Eyto, E., E. Jennings, E. Ryder, K. Sparber, M. Dillane, C. Dalton, and R. Poole. 2016. "Response of a Humic Lake Ecosystem to an Extreme Precipitation Event: Physical, Chemical, and Biological Implications." *Inland Waters* 6: 483–98.
- de Senerpont Domis, L. N., J. J. Elser, A. S. Gsell, V. L. M. Huszar, B. W. Ibelings, E. Jeppesen, S. Kosten, et al. 2013. "Plankton Dynamics under Different Climatic Conditions in Space and Time." *Freshwater Biology* 58: 463–82.
- Denwood, M. J. 2016. "runjags: An R Package Providing Interface Utilities, Model Templates, Parallel Computing Methods and Additional Distributions for MCMC Models in JAGS." *Journal of Statistical Software*, 71(9): 1–25. doi:10.18637/jss.v071.i09
- Dietze, M. C. 2017a. *Ecological Forecasting*, First ed. Princeton, NJ: Princeton University Press.
- Dietze, M. C. 2017b. "Prediction in Ecology: A First-Principles Framework: A." *Ecological Applications* 27: 2048–60.
- Dietze, M. C., A. Fox, L. M. Beck-Johnson, J. L. Betancourt, M. B. Hooten, C. S. Jarnevich, T. H. Keitt, et al. 2018. "Iterative near-Term Ecological Forecasting: Needs, Opportunities, and Challenges." *Proceedings of the National Academy of Sciences of the United States of America* 115: 1424–32.

- Diniz-Filho, A. F., L. M. Bini, T. F. Rangel, R. D. Loyola, C. Hof, D. Nogue, and M. B. Arau. 2009. "Partitioning and Mapping Uncertainties in Ensembles of Forecasts of Species Turnover under Climate Change." *Ecography* 32: 897–906.
- Dodds, W., W. W. Bouska, J. L. Eitzmann, T. J. Pilger, K. L. Pitts, A. J. Riley, J. T. Schloesser, and D. J. Thornbrugh. 2009. "Eutrophication of U.S. Freshwaters: Analysis of Potential Economic Damages." *Environmental Science & Technology* 43: 12–9.
- Dokulil, M. T., and K. Teubner. 2000. "Cyanobacterial Dominance in Lakes." *Hydrobiologia* 438: 1–12.
- Fox, A. M., T. J. Hoar, J. L. Anderson, A. F. Arellano, W. K. Smith, M. E. Litvak, N. MacBean, D. S. Schimel, and D. J. P. Moore. 2018. "Evaluation of a Data Assimilation System for Land Surface Models Using CLM4.5." *Journal of Advances in Modeling Earth Systems* 10: 2471–94.
- Franks, P. J. S. 1997. "Spatial Patterns in Dense Algal Blooms." *Limnology and Oceanography* 42: 1297–305.
- Freeman, E. C., I. F. Creed, B. Jones, and A.-K. Bergstrom. 2020. "Global Changes May Be Promoting a Rise in Select Cyanobacteria in Nutrient-Poor Northern Lakes." *Global Change Biology* 26: 4966–87.
- Gelfand, A. E., and S. K. Ghosh. 1998. "Model Choice: A Minimum Posterior Predictive Loss Approach." *Biometrika Trust* 85: 1–11.
- Gertner, G., P. Parysow, and B. Guan. 1996. "Projection Variance Partitioning of a Conceptual Forest Growth Model with Orthogonal Polynomials." *Forest Science* 42: 474–86.
- Gibson, G., R. Carlson, J. Simpson, E. Smeltzer, J. Gerritson, S. Chapra, S. Heiskary, J. Jones, and R. Kennedy. 2000. *U.S. EPA Nutrient Criteria Technical Guidance Manual Lakes and Reservoirs*. Washington, D.C.: United States Environmental Protection Agency.
- Hamilton, G., R. McVinish, and K. Mengersen. 2009. "Bayesian Model Averaging for Harmful Algal Bloom Prediction." *Ecological Applications* 19: 1805–14.
- Hampton, S. E., A. W. E. Galloway, S. M. Powers, T. Ozersky, K. H. Woo, R. D. Batt, S. G. Labou, et al. 2016. "Ecology under Lake Ice." *Ecology Letters* 20: 98–111.
- Ho, J. C., and A. M. Michalak. 2020. "Exploring Temperature and Precipitation Impacts on Harmful Algal Blooms across Continental U.S. Lakes." *Limnology and Oceanography* 65: 992–1009.
- Hobbs, N. T., C. Geremia, J. Treanor, R. Wallen, P. J. White, M. B. Hooten, and J. C. Rhyhan. 2015. "State-Space Modeling to Support Management of Brucellosis in the Yellowstone Bison Population." *Ecological Monographs* 85: 525–56.
- Hobbs, N. T., and M. B. Hooten. 2015. *Bayesian Models: A Statistical Primer for Ecologists*, First ed. Princeton, NJ: Princeton University Press.
- Hobday, A. J., J. R. Hartog, J. P. Manderson, K. E. Mills, M. J. Oliver, A. J. Pershing, and S. Siedlecki. 2019. "Ethical Considerations and Unanticipated Consequences Associated with Ecological Forecasting for Marine Resources." *ICES Journal of Marine Science* 76: 1244–56.
- Holan, S. H., G. M. Davis, M. L. Wildhaber, A. J. DeLonay, and D. M. Papoulias. 2009. "Hierarchical Bayesian Markov Switching Models with Application to Predicting Spawning Success of Shovelnose Sturgeon." *Journal of the Royal Statistical Society Series C: Applied Statistics* 58: 47–64.
- Huang, J., J. Gao, J. Liu, and Y. Zhang. 2013. "State and Parameter Update of a Hydrodynamic-Phytoplankton Model Using Ensemble Kalman Filter." *Ecological Modelling* 263: 81–91.
- Huisman, J., and F. D. Hulot. 2005. "Population Dynamics of Harmful Cyanobacteria." In *Harmful Cyanobacteria*. Aquatic Ecology Series, Vol 3, edited by J. Huisman, H. C. P. Matthijs, and P. M. Visser, 143–76. Dordrecht: Springer.
- Hyenstrand, P., E. Rydin, and M. Gunnerhed. 2000. "Response of Pelagic Cyanobacteria to Iron Additions—Enclosure Experiments from Lake Erken." *Journal of Plankton Research* 22: 1113–26.
- Ibelings, B. W., M. Bormans, J. Fastner, and P. M. Visser. 2016. "CYANOCOST Special Issue on Cyanobacterial Blooms: Synopsis—a Critical Review of the Management Options for their Prevention, Control and Mitigation." *Aquatic Ecology* 50: 1–11.
- Idso, S. B. 1973. "On the Concept of Lake Stability." *Limnology and Oceanography* 18: 681–3.
- Istvánovics, V., K. Pettersson, M. A. Rodrigo, D. Pierson, J. Padišák, and W. Colom. 1993. "Gloeotrichia echinulata, a Colonial Cyanobacterium with a Unique Phosphorus Uptake and Life Strategy." *Journal of Plankton Research* 15: 531–52.
- Janssen, A. B., J. H. Janse, A. H. Beusen, M. Chang, J. A. Harrison, I. Huttunen, X. Kong, et al. 2019. "How to Model Algal Blooms in any Lake on Earth." *Current Opinion in Environmental Sustainability* 36: 1–10.
- Jennings, E., S. Jones, L. Arvola, P. A. Staehr, E. Gaiser, I. D. Jones, K. C. Weathers, G. A. Weyhenmeyer, C. Y. Chiu, and E. De Eyto. 2012. "Effects of Weather-Related Episodic Events in Lakes: An Analysis Based on High-Frequency Data." *Freshwater Biology* 57: 589–601.
- Johansson, M. A., K. M. Apfeldorf, S. Dobson, J. Devita, A. L. Buczak, B. Baugher, L. J. Moniz, et al. 2019. "An Open Challenge to Advance Probabilistic Forecasting for Dengue Epidemics." *Proceedings of the National Academy of Sciences USA* 116: 24268–74.
- Jolliffe, I., and D. Stephenson. 2003. *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, Second ed. Chichester: Wiley.
- Karlsson-Elfgren, I., P. Hyenstrand, and E. Rydin. 2005. "Pelagic Growth and Colony Division of *Gloeotrichia echinulata* in Lake Erken." *Journal of Plankton Research* 27: 145–51.
- Karlsson-Elfgren, I., K. Rengefors, and S. Gustafsson. 2004. "Factors Regulating Recruitment from the Sediment to the Water Column in the Bloom-Forming Cyanobacterium *Gloeotrichia echinulata*." *Freshwater Biology* 49: 265–73.
- Kauffman, G. J. 2016. "Economic Value of Nature and Ecosystems in the Delaware River Basin." *Journal of Contemporary Water Research & Education* 158: 98–119.
- Kuha, J., L. Arvola, P. Hanson, J. Huotari, T. Huttula, J. Juntunen, M. Jarvinen, et al. 2016. "Response of Boreal Lakes to Episodic Weather-Induced Events." *Inland Waters* 6: 523–34.
- Kuikka, S., J. Vanhatalo, H. Pulkkinen, S. Mäntyniemi, and J. Corander. 2014. "Experience in Bayesian Inference in Baltic Salmon Management." *Statistical Science* 29: 42–9.
- Küpper, H., S. Seibert, and A. Parameswaran. 2007. "Fast, Sensitive, and Inexpensive Alternative to Analytical Pigment HPLC: Quantification of Chlorophylls and Carotenoids in Crude Extracts by Fitting with Gauss Peak Spectra." *Analytical Chemistry* 79: 7611–27.
- Lake Sunapee Protective Association (LSPA), B. Steele, and K. C. Weathers. 2020. "Lake Sunapee High Frequency Weather Data measured on buoy - 2007–2017 ver. 3." Environmental Data Initiative. <https://doi.org/10.6073/pasta/175cfe22aaeee7349285d2c9fd298d9>

- Lake Sunapee Protective Association (LSPA), K. C. Weathers, and B. G. Steele. 2020. "Lake Sunapee Instrumented Buoy: High Frequency Water Temperature and Dissolved Oxygen Data – 2007–2019." Environmental Data Initiative. <https://doi.org/10.6073/pasta/70c41711d6199ac2758764ecfcb9815e>
- Le Vu, B., B. Vinçon-Leite, B. J. Lemaire, N. Bensoussan, M. Calzas, C. Drezen, J. F. Deroubaix, et al. 2011. "High-Frequency Monitoring of Phytoplankton Dynamics within the European Water Framework Directive: Application to Metalimnetic Cyanobacteria." *Biogeochemistry* 106: 229–42.
- Lofton, M. E., J. A. Brentrup, W. S. Beck, J. A. Zwart, R. Bhattacharya, L. S. Brighenti, S. H. Burnet, et al. 2022. "Lake Sunapee *Gloeotrichia echinulata* Density near-Term Hindcasts from 2015–2016 and Meteorological Model Driver Data, Including Shortwave Radiation and Precipitation from 2009–2016 ver 1." Environmental Data Initiative Staging Environment. <https://doi.org/10.6073/pasta/af14c8e8f2b0a793fab067536fb8abeb>
- Massoud, E. C., J. Huisman, E. Beninca, M. C. Dietze, W. Bouten, and J. A. Vrugt. 2018. "Probing the Limits of Predictability: Data Assimilation of Chaotic Dynamics in Complex Food Webs." *Ecology Letters* 21: 93–103.
- McGowan, J. A., E. R. Deyle, H. Ye, M. L. Carter, C. T. Perretti, M. Hilbern, K. Seger, A. de Verniel, and G. Sugihara. 2017. "Predicting Coastal Algal Blooms in Southern California." *Ecology* 98: 1419–33.
- Mueller, H., D. P. Hamilton, and G. J. Doole. 2016. "Evaluating Services and Damage Costs of Degradation of a Major Lake Ecosystem." *Ecosystem Services* 22: 370–80.
- Napiórkowska-Krzebietke, A., and A. Hutorowicz. 2015. "The Physicochemical Background for the Development of Potentially Harmful Cyanobacterium *Gloeotrichia echinulata* J. S. Smith ex Richt." *Journal of Elementology* 20: 363–76.
- Nesslage, G. M., and M. J. Wilberg. 2019. "A Performance Evaluation of Surplus Production Models with Time-Varying Intrinsic Growth in Dynamic Ecosystems." *Canadian Journal of Fisheries and Aquatic Sciences* 76: 2245–55.
- O'Neil, J. M., T. W. Davis, M. A. Burford, and C. J. Gobler. 2012. "The Rise of Harmful Cyanobacteria Blooms: The Potential Roles of Eutrophication and Climate Change." *Harmful Algae* 14: 313–34.
- Özkundakci, D., A. S. Gsell, T. Hintze, H. Täuscher, and R. Adrian. 2015. "Winter Severity Determines Functional Trait Composition of Phytoplankton in Seasonally Ice-Covered Lakes." *Global Change Biology* 22: 284–98.
- Paerl, H. W., N. S. Hall, and E. S. Calandrino. 2011. "Controlling Harmful Cyanobacterial Blooms in a World Experiencing Anthropogenic and Climatic-Induced Change." *Science of the Total Environment* 409: 1739–45.
- Paerl, H. W., and J. Huisman. 2008. "Blooms like it Hot." *Science* 320: 57–8.
- Paerl, H. W., and T. G. Otten. 2013. "Harmful Cyanobacterial Blooms: Causes, Consequences, and Controls." *Microbial Ecology* 65: 995–1010.
- Page, T., P. J. Smith, K. J. Beven, I. D. Jones, J. A. Elliott, S. C. Maberly, E. B. Mackay, M. De Ville, and H. Feuchtmayr. 2017. "Constraining Uncertainty and Process-Representation in an Algal Community Lake Model Using High Frequency in-Lake Observations." *Ecological Modelling* 357: 1–13.
- Page, T., P. J. Smith, K. J. Beven, I. D. Jones, J. A. Elliott, S. C. Maberly, E. B. Mackay, M. De Ville, and H. Feuchtmayr. 2018. "Adaptive Forecasting of Phytoplankton Communities." *Water Research* 134: 74–85.
- Phillips, K. A., and M. W. Fawley. 2002. "Winter Phytoplankton Community Structure in Three Shallow Temperate Lakes during Ice Cover." *Hydrobiologia* 470: 97–113.
- Pinto, C., and L. Spezia. 2016. "Markov Switching Autoregressive Models for Interpreting Vertical Movement Data with Application to an Endangered Marine Apex Predator." *Methods in Ecology and Evolution* 7: 407–17.
- Plummer, M., A. Stukalov, and M. Denwood. 2019. "Bayesian Graphical Models Using MCMC." <https://mcmc-jags.sourceforge.io>
- Powers, S. M., and S. E. Hampton. 2019. "Open Science, Reproducibility, and Transparency in Ecology." *Ecological Applications* 29: 1–8.
- R Core Team. 2020. "R: A language and environment for statistical computing." Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Rigosi, A., W. Fleenor, and F. Rueda. 2010. "State-of-the-Art and Recent Progress in Phytoplankton Succession Modelling." *Environmental Reviews* 18: 423–40.
- Roelofs, T. D., and R. T. Oglesby. 1970. "Ecological Observations on the Planktonic Cyanophyte *Gloeotrichia echinulata*." *Limnology and Oceanography* 15: 224–9.
- Rolland, D. C., S. Bourget, A. Warren, I. Laurion, and W. F. Vincent. 2013. "Extreme Variability of Cyanobacterial Blooms in an Urban Drinking Water Supply." *Journal of Plankton Research* 35: 744–58.
- Rott, E., N. Salmaso, and E. Hoehn. 2007. "Quality Control of Utermöhl-Based Phytoplankton Counting and Biovolume Estimates—an Easy Task or a Gordian Knot?" *Hydrobiologia* 578: 141–6.
- Rouso, B. Z., E. Bertone, R. Stewart, and D. P. Hamilton. 2020. "A Systematic Literature Review of Forecasting and Predictive Models for Cyanobacteria Blooms in Freshwater Lakes." *Water Research* 182: 115959.
- Serizawa, H., T. Amemiya, and K. Itoh. 2008. "Patchiness in a Minimal Nutrient—Phytoplankton Model." *Journal of Biosciences* 33: 391–403.
- Steele, B., K. C. Weathers, and Lake Sunapee Protective Association (LSPA). 2021. "Quality controlled in situ data from multiple locations in Lake Sunapee, NH, USA from the Lake Sunapee Protective Association's Long-term Monitoring Program, 1986–2020 (Version v2020.1)." Zenodo. <https://doi.org/10.5281/zenodo.4652076>.
- Sturner, R. W., K. L. Reint, B. M. Lafrancois, S. Brovold, and T. R. Miller. 2020. "A First Assessment of Cyanobacterial Blooms in Oligotrophic Lake Superior." *Limnology and Oceanography* 65: 2984–98.
- Stoddard, J. L., J. Van Sickle, A. T. Herlihy, J. Brahney, S. Paulsen, D. V. Peck, R. Mitchell, and A. I. Pollard. 2016. "Continental-Scale Increase in Lake and Stream Phosphorus: Are Oligotrophic Systems Disappearing in the United States?" *Environmental Science and Technology* 50: 3409–15.
- Stroom, J. M., and W. E. A. Kardinaal. 2016. "How to Combat Cyanobacterial Blooms: Strategy toward Preventive Lake Restoration and Reactive Control Measures." *Aquatic Ecology* 50: 1–36.

- Stumpf, R. P., C. Mi, J. A. Tomlinson, B. Calkins, K. Kirkpatrick, K. Fisher, R. C. Nierenberg, and T. T. Wynne. 2009. "Skill Assessment for an Operational Algal Bloom Forecast System." *Journal of Marine Systems* 76: 151–61.
- Thomas, R. Q., R. J. Figueiredo, V. Daneshmand, B. J. Bookout, L. K. Puckett, and C. C. Carey. 2020. "A Near-Term Iterative Forecasting System Successfully Predicts Reservoir Hydrodynamics and Partitions Uncertainty in Real Time." *Water Resources Research* 56: e2019WR026138.
- Thomas, R. Q., A. L. Jersild, E. B. Brooks, V. A. Thomas, and R. H. Wynne. 2018. "A Mid-Century Ecological Forecast with Partitioned Uncertainty Predicts Increases in Loblolly Pine Forest Productivity." *Ecological Applications* 28: 1503–19.
- Thrane, J. E., M. Kyle, M. Striebel, S. Haande, M. Grung, T. Rohrlack, and T. Andersen. 2015. "Spectrophotometric Analysis of Pigments: A Critical Assessment of a High-Throughput Method for Analysis of Algal Pigment Mixtures by Spectral Deconvolution." *PLoS ONE* 10: 1–24.
- U.S. EPA. 1991. *Guidance Manual for Compliance with the Filtration and Disinfection Requirements for Public Water Systems Using Surface Water Sources*. Washington, D.C.: United States Environmental Protection Agency.
- Valle, D., C. L. Staudhammer, W. P. Cropper, and P. R. Van Gardingen. 2009. "The Importance of Multimodel Projections to Assess Uncertainty in Projections from Simulation Models." *Ecological Applications* 19: 1680–92.
- Vuorio, K., L. Lepistö, and A. L. Holopainen. 2007. "Intercalibrations of Freshwater Phytoplankton Analyses." *Boreal Environment Research* 12: 561–9.
- Wang, G., T. Oyana, M. Zhang, S. Adu-prah, S. Zeng, H. Lin, and J. Se. 2009. "Mapping and Spatial Uncertainty Analysis of Forest Vegetation Carbon by Combining National Forest Inventory Data and Satellite Images." *Forest Ecology and Management* 258: 1275–83.
- Ward, E. J., E. E. Holmes, J. T. Thorson, and B. Collen. 2014. "Complexity Is Costly: A Meta-Analysis of Parametric and Non-parametric Methods for Short-Term Population Forecasting." *Oikos* 123: 652–61.
- Ward, N. K., B. G. Steele, K. C. Weathers, K. L. Cottingham, H. A. Ewing, P. C. Hanson, and C. C. Carey. 2020. "Differential Responses of Maximum Versus Median Chlorophyll-a to Air Temperature and Nutrient Loads in an Oligotrophic Lake over 31 years." *Water Resources Research* 56: e2020WR027296.
- Watling, J. I., L. A. Brandt, D. N. Bucklin, I. Fujisaki, F. J. Mazzotti, S. S. Roma, and C. Speroterra. 2015. "Performance Metrics and Variance Partitioning Reveal Sources of Uncertainty in Species Distribution Models." *Ecological Modelling* 309–310: 48–59.
- Wiedner, C., and B. Nixdorf. 1998. "Success of Chrysophytes, Cryptophytes and Dinoflagellates over Blue-Greens (Cyanobacteria) during an Extreme Winter (1995/96) in Eutrophic Shallow Lakes." *Hydrobiologia* 369–370: 229–35.
- Wilkinson, M. D., M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3: 160018.
- Wilson, M. A., and S. R. Carpenter. 1999. "Economic Valuation of Freshwater Ecosystem Services in the United States: 1971–1997." *Ecological Applications* 9: 772–83.
- Winder, M., and J. E. Cloern. 2010. "The Annual Cycles of Phytoplankton Biomass." *Philosophical Transactions of the Royal Society B: Biological Sciences* 365: 3215–26.
- Winter, J. G., A. M. Desellas, R. Fletcher, L. Heintsch, A. Morley, L. Nakamoto, and K. Utsumi. 2011. "Algal Blooms in Ontario, Canada: Increases in Reports since 1994." *Lake and Reservoir Management* 27: 107–14.
- Wynne, T. T., and R. P. Stumpf. 2015. "Spatial and Temporal Patterns in the Seasonal Distribution of Toxic Cyanobacteria in Western Lake Erie from 2002–2014." *Toxins* 7: 1649–63.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Lofton, Mary E., Jennifer A. Brentrup, Whitney S. Beck, Jacob A. Zwart, Ruchi Bhattacharya, Ludmila S. Brighenti, Sarah H. Burnet, et al. 2022. "Using Near-Term Forecasts and Uncertainty Partitioning to Inform Prediction of Oligotrophic Lake Cyanobacterial Density." *Ecological Applications* 32(5): e2590. <https://doi.org/10.1002/eap.2590>