

**Running head:** Chl-a forecasts at multiple time scales

**Title:** Near-term phytoplankton forecasts reveal the effects of model time step and forecast horizon on predictability

**Authors:** Whitney M. Woelmer<sup>11</sup>, R. Quinn Thomas<sup>22</sup>, Mary E. Lofton<sup>13</sup>, Ryan P. McClure<sup>14</sup>, Heather L. Wander<sup>15</sup>, Cayelan C. Carey<sup>16</sup>

<sup>1</sup>Department of Biological Sciences, Virginia Tech, Blacksburg, Virginia, USA

<sup>2</sup>Department of Forest Resources and Environmental Conservation, Virginia Tech, Blacksburg, Virginia, US

### **Open data statement**

Forecast data, metadata, and code are provided as private-for-peer review in a Zenodo repository (<https://doi.org/DOI:10.5281/zenodo.5963867>) and all data needed to run the analyses are published in the Environmental Data Initiative repository:

---

<sup>1</sup> Corresponding author. Address: Department of Biological Sciences, Derring Hall Room 2125, 926 West Campus Drive, Mail Code 0406, Blacksburg, VA 24061, USA. Phone: +1-540-231-0788. Fax: +1-540-231-9307. Email: [wwoelmer@vt.edu](mailto:wwoelmer@vt.edu), ORCID ID: 0000-0001-5147-3877

<sup>2</sup> Email: [rqthomas@vt.edu](mailto:rqthomas@vt.edu), ORCID ID: 0000-0003-1282-7825

<sup>3</sup> Email: [melofton@vt.edu](mailto:melofton@vt.edu), ORCID ID: 0000-0003-3270-1330

<sup>4</sup> Email: [ryan333@vt.edu](mailto:ryan333@vt.edu), ORCID ID: 0000-0001-6370-3852

<sup>5</sup> Email: [hwander@vt.edu](mailto:hwander@vt.edu), ORCID ID: 0000-0002-3762-6045

<sup>6</sup> Email: [cayelan@vt.edu](mailto:cayelan@vt.edu), ORCID ID: 0000-0001-8835-4476

**Meteorological data:** <https://doi.org/10.6073/pasta/890e4c11f4348b3ceda802732ffa48b4>

**Discharge data:** <https://doi.org/10.6073/pasta/8d22a432aac5560b0f45aa1b21ae4746>

**Observed chlorophyll-a data:**

<https://doi.org/10.6073/pasta/5448f9d415fd09e0090a46b9d4020ccc>

<https://doi.org/10.6073/pasta/88896f4a7208c9b7bddcf498258edf78>

Submitted as an Article to *Ecological Applications*

Declarations of interest: none

## ABSTRACT

As climate and land use increase the variability of many ecosystems, forecasts of ecological variables are needed to inform management and use of ecosystem services. In particular, forecasts of phytoplankton would be especially useful for drinking water management, as phytoplankton populations are exhibiting greater fluctuations due to human activities. While phytoplankton forecasts are increasing in number, many questions remain regarding the optimal model time step (the temporal frequency of the forecast model output), time horizon (the length of time into the future a prediction is made) for maximizing forecast performance, as well as what factors contribute to uncertainty in forecasts and their scalability among sites. To answer these questions, we developed near-term, iterative forecasts of phytoplankton 1 to 14 days into the future using forecast models with three different time steps (daily, weekly, fortnightly), that included a full uncertainty partitioning analysis at two drinking water reservoirs. We found that forecast accuracy varies with model time step and forecast horizon, and that forecast models can outperform null estimates under most conditions. Weekly and fortnightly forecasts consistently outperformed daily forecasts at 7-day and 14-day horizons, a trend which increased up to the 14-

day forecast horizon. Importantly, our work suggests that forecast accuracy can be increased by matching the forecast model time step to the forecast horizon for which predictions are needed. We found that model process uncertainty was the primary source of uncertainty in our phytoplankton forecasts over the forecast period, but parameter uncertainty increased during phytoplankton blooms and when scaling the forecast model to a new site. Overall, our scalability analysis shows promising results that simple models can be transferred to produce forecasts at additional sites. Altogether, our study advances our understanding of how forecast model time step and forecast horizon influence the forecastability of phytoplankton dynamics in aquatic systems, and adds to the growing body of work regarding the predictability of ecological systems broadly.

**KEYWORDS:** autoregressive model, Bayesian model, blooms, chlorophyll-a, ecological forecasting, hindcast, historical monitoring, iterative, management, phytoplankton, time series analysis, water quality

## INTRODUCTION

Globally, ecosystems are experiencing unprecedented climate and land use change (Vitousek 1994, Solomon et al. 2007), resulting in increased variability in their functioning (e.g., Smith et al. 2009, Komatsu et al. 2019, Rillig et al. 2019). As a result of this greater variability, management of the ecosystem services upon which society depends is increasingly challenging (Manning et al. 2019b, Hines et al. 2019). Thus, new predictive approaches to anticipate future ecosystem responses to global change are needed to assist managers and the public as they respond to changing ecosystem states (Clark et al. 2001, Dietze et al. 2018).

Freshwater ecosystems, which provide integral services to society (e.g., drinking water, fisheries, irrigation, hydropower), are particularly threatened by human activities (Millenium Ecosystem Assessment 2005, Jimenez Cisneros et al. 2014). Freshwater lakes and reservoirs are experiencing rapid changes in water quality, including increased phytoplankton biomass in many waterbodies, due to climate and land use change (Carey et al. 2012, Paerl and Paul 2012, Sinha et al. 2012, Ho et al. 2019). Increased phytoplankton biomass poses a potent risk to drinking water (Carpenter et al. 1998). Because of their toxins, odors, and scums, high phytoplankton concentrations increase the need for drinking water treatment, which is estimated to cost >\$2 billion annually in the U.S. (Dodds et al. 2009).

Anticipating future lake and reservoir phytoplankton concentrations is critical to drinking water management. There is a great need to predict both phytoplankton blooms, which are large, ephemeral aggregations of phytoplankton biomass that can have substantial negative effects on drinking water quality (Tarczyńska et al. 2001, Qin et al. 2010, Cheung et al. 2013, Ewerts et al. 2013), as well as “baseline,” or non-bloom, phytoplankton concentrations in drinking water lakes

and reservoirs. Having forecasts of both bloom and non-bloom phytoplankton conditions provide complementary information for drinking water managers. For example, knowing preemptively when a high-impact yet rare bloom will occur would allow managers to order additional treatment supplies in advance or provide swimming closure notices before the bloom occurs (Carey et al. 2021a). Similarly, daily water treatment operations could be improved by forecasts of non-bloom conditions, which would provide important information on "typical" water quality conditions that occur over the majority of the year, allowing managers to choose depths from which to draw water. Here, we define predictions of blooms as the prediction of maximum or peak phytoplankton concentrations in a year, whereas predictions of non-bloom conditions are represented by mean or median phytoplankton concentrations. Despite the great need for scalable phytoplankton forecasts across many waterbodies, it remains unclear how well they can be predicted under bloom and non-bloom conditions.

Providing near-term, iterative forecasts, or future estimates of ecological variable(s) with quantified uncertainty, may be especially useful for management as they provide information on a timescale that is relevant for decision-making (Carey et al. 2021a). Additionally, a full analysis of the relative contribution of different sources of uncertainty (Table 1) in forecasts is crucial to both properly inform management decisions (Morss et al. 2008, Berthet et al. 2016, Dietze 2017a) and iteratively improve forecast performance by constraining large sources of uncertainty (Luo et al. 2011, Dietze 2017a). Finally, building a forecasting framework which readily scales from one location to another is critical to furthering the field of ecological forecasting.

Current phytoplankton forecasts are often made at multiple forecast horizons to provide predictions at several timepoints in the future. Forecast horizon, or the length of time into the future at which a forecast is made, is generally expected to decrease the forecast's predictive

accuracy (Petchey et al. 2015, Dietze 2017). However, forecasts which extend to longer time horizons (~1-2 weeks) may be more useful decision support tools than 1-2 day forecasts by providing managers longer lead times for implementing interventions (e.g., Carey et al. 2021a, Thomas et al. 2020). Forecasts are often made at longer time horizons by propagating forecasts forward over multiple time steps. Importantly, the time step of a forecast, or the temporal frequency of the forecast model output, may also influence prediction accuracy by representing different processes which occur over different time periods. However, the forecast time step is not generally taken into account when choosing a forecast model. As a result, finding a balance between the time horizon at which forecast accuracy deteriorates and the time horizon which provides the most useful decision support tool for managers is a critical step in advancing ecological forecast applications. For predictions made further into the future, it remains unknown whether models which are developed for that time step perform better than models which simply propagate daily forecasts out to multiple time steps, a focus of this study.

While the number of phytoplankton forecasts have increased in recent years, there is still a substantial range in the time step at which phytoplankton are forecasted (reviewed by Rousso et al. 2020). A majority of phytoplankton forecasts are made using a daily or sub-daily time step, which is then used to forecast time horizons ranging from 1 day to 14 days into the future (e.g., Chen et al. 2015, Xiao et al. 2017, Page et al. 2018, Kehoe et al. 2019, Loos et al. 2020). Daily or sub-daily forecasts have variable but reasonable success forecasting non-bloom dynamics and timing of bloom events, but are often unable to capture the magnitude of phytoplankton blooms, resulting in under-prediction (Huang et al. 2013, Chen et al. 2015, Rajaei and Boroumand 2015, Recknagel et al. 2016, Massoud et al. 2018) or over-prediction (Mao et al. 2009, Page et al. 2018, Loos et al. 2020) of bloom concentrations. Moreover, these forecasts may be less accurate

at forecasting multiple days or a week into the future, a lead time which is needed for drinking water managers to change treatment operations (e.g., Thomas et al. 2020). In contrast, phytoplankton forecasts which use much longer time steps (i.e., monthly or seasonal) show relatively high performance at predicting both bloom and non-bloom conditions (Rajaei and Boroumand 2015, Manning et al. 2019a, Park et al. 2019), but lose the ability to capture short-term variability due to the coarseness of their model time step.

Forecasts on a weekly or fortnightly time step are comparatively rare (but see Recknagel et al. 2016), and none of which we are aware compare the relative success of prediction at multiple time scales within the same ecosystem. However, given that most lake and reservoir monitoring programs typically collect weekly or fortnightly manual phytoplankton samples (Arhonditsis et al. 2004, Wang et al. 2011, Read et al. 2015, Gerling et al. 2016, McGowan et al. 2017), developing forecasting workflows on a weekly or fortnightly time step may be readily available for many waterbodies. In the meantime, the few but increasing number of lakes and reservoirs that have high-frequency phytoplankton sensors (e.g., Marcé et al. 2016) provide an ideal testbed for assessing the performance of phytoplankton forecasting models which operate on different time steps (*sensu* Hamilton et al. 2015).

To examine the effect of model time step on phytoplankton forecast skill, we developed ~600 days of near-term, iterative phytoplankton forecasts with quantified uncertainty for a drinking water reservoir. We made forecasts during bloom and non-bloom conditions using three time series models built on different time steps: daily, weekly, and fortnightly. To assess the scalability of our forecasting framework, we adapted the framework and applied it to an additional drinking water reservoir to produce one year of forecasts, and compared the skill of the forecasts at both sites. We addressed four questions in this study: 1) How well can we

forecast phytoplankton over the entire forecast period, as well as under bloom and non-bloom conditions? 2) What is the effect of forecast model time step on forecast performance?, 3) How do the major sources of phytoplankton forecast uncertainty vary with forecast horizon and over time?, and 4) How do forecast skill and uncertainty contributions change when scaled to an additional study site?

## METHODS

### *Study site*

Forecasts were produced for near-surface chlorophyll-a (henceforth ‘chl-a’, a common metric of phytoplankton biomass) at Falling Creek Reservoir (FCR). FCR is a small (0.119 km<sup>2</sup>), shallow (maximum depth = 9.3 m), eutrophic reservoir located in Vinton, Virginia, USA (37.30°N, 79.84°W, Figure 1). FCR is dimictic, with thermal stratification occurring annually from approximately May-October (Carey et al. 2019). FCR has a history of phytoplankton blooms and is primarily fed by one major upstream tributary that has been monitored with a weir since 2013 (Gerling et al. 2014, 2016). The reservoir is owned and operated as a drinking water source by the Western Virginia Water Authority (WVWA), who partnered with our research team throughout forecast development (described by Carey et al. 2021a). This partnership informed several aspects of our forecasting framework, including the location of the chl-a sensor used for updating forecasts, the time steps of interest for management decision-making, and forecast delivery mode. While we note that this study produced hindcasts (following the definition of Jolliffe and Stephenson 2003), we refer to them as forecasts for consistency.



### *Forecasting overview*

Forecasts of near-surface (~1.0 m depth) chl-a at FCR were produced over the course of ~600 days (from 2 January 2019 to 15 August 2020) using daily, weekly, and fortnightly autoregressive (AR) linear models developed using observational sensor data from FCR. For each forecast, model driver and validation data were automatically collected via automated sensors at the study site and wirelessly uploaded to an online data repository using secure sensor gateways (Figure 2: Step 1). At each model time step, new data up to the day being forecasted were appended to the historical training dataset and used to re-parameterize the AR models (Figure 2: Steps 2-3). Forecasted model driver data included shortwave radiation and forecasted discharge to the reservoir from the major inflow. Shortwave forecasts were downloaded from the National Atmospheric and Oceanic Administration Global Ensemble Forecasts System (NOAA GEFS) repository. Forecasts of future discharge were modeled using observed sensor discharge measurements on the inflow at FCR and NOAA forecasted precipitation as inputs (see Thomas et al. 2020 for a detailed description) (Figure 2: Step 4). Uncertainty was propagated for four different uncertainty types (process, initial condition, parameter, and driver data, see Table 1 for definitions) (Figure 2: Step 5).

We generated probabilistic daily forecasts which had a 1-day, 2-day, 3-day... up to 14-day time horizon, weekly forecasts which had a 1-week and 2-week (i.e., 7-day and 14-day) time horizon, and fortnightly forecasts which had a 2-week time horizon (i.e., 14-day; Figure 2: Step 6). Thus, to summarize, there were 14 forecast horizons for the daily model, two horizons for the weekly model, and one horizon for the fortnightly model. To develop and run our forecast models, we used a combination of linear parametric and Bayesian statistical methods. We used

parametric linear model selection on historical data to select model covariates and initial parameter values (see *Model development and training*, below). To produce forecasts, we applied our model in a Bayesian framework (see *Bayesian forecasting framework*, below).

All driver data is published in the Environmental Data Initiative (EDI) repository (Woelmer et al. 2021, staging link) and all forecast output and code is available in Zenodo (<https://doi.org/DOI:10.5281/zenodo.5963867>).

### *Reservoir monitoring dataset overview*

Our research team has monitored physical, chemical, and biological conditions at FCR since 2013. This long-term dataset was used to develop and train the AR models prior to the beginning of the forecast period and includes manual and automated sensor measurements of chl-a at the deepest site of the reservoir, meteorological variables from a meteorological station on the dam of the reservoir, and sensor measurements of discharge at the major inflow stream to the reservoir (Figure 1; see Appendix S1: Section S1 for more details on each of these datasets). Because sampling primarily occurred from May to October in 2013-2016, we began running our models during the second week of May so that observations in the first week of May served as the AR lag, not the October observations.

### *Model development and training*

We used standard linear model selection techniques (Appendix S1: Section S2) to select covariates for the AR models (following Box and Jenkins, 1970; Quinn and Keough 2002) for forecasting surface chl-a (1.0 m) using the monitoring dataset described above. We chose an autoregressive linear model to capture the high temporal autocorrelation common in

phytoplankton, as was demonstrated in our dataset (Appendix S1: Figure S4). While this model is relatively simple in its representation of phytoplankton dynamics, we wanted to test the ability of a data-driven model that was easy to implement for examining the effect of different model time steps and horizons at multiple sites, following numerous previous phytoplankton predictive studies (Rousso et al. 2020).

We used a weekly dataset of surface chl-a from 2013-2016 to select covariates for our AR model. We chose the weekly dataset, as opposed to the daily dataset, because it had the longest historical data coverage, dating back to 2013, whereas the daily dataset only began in summer 2018. When developing our linear model, we aimed to meet the assumptions that variables are normally distributed, and found that square-root transformation of our chlorophyll data led to a better fit of normality than log-transformation. Meteorological and discharge data were aggregated to daily summaries (mean, median, maximum, and minimum) on the same day as the chl-a observations. Any weeks with missing data from May-October over the four years (n=18 out of 107 weeks total) were linearly interpolated using the `na.approx()` function in the ‘zoo’ package (Zeileis and Grothendieck, 2005) to allow for a consistent time interval in the training data. To determine the influence of these linearly interpolated training data points on model performance, we conducted an identical model analysis using only observed data without interpolated data points, which showed no substantial change in model accuracy (Appendix S1: Figure S1). Variables that did not follow a normal distribution were transformed to meet the assumptions of a linear model (Appendix S1: Table S2). Our best-performing weekly model (which also serves as the process model for our Bayesian analysis) was:

$$C_{t+n} = \beta_1 + \beta_2 C_t + \beta_3 S_{t+n} + \beta_4 D_{t+n} + \varepsilon \quad (\text{Eq. 1})$$

Where  $C_t$  is the AR term or mean daily chl-a at the current time,  $t$ . The response,  $C_{t+n}$ , is the mean daily chl-a concentration at the forecasted time horizon,  $n$  days in the future, where  $n$  is the time step at each time horizon of the model (e.g.,  $n = 1-14$  days for daily forecasts, 7 and 14 days for weekly forecasts, and 14 days for fortnightly forecasts).  $S_{t+n}$  is the mean daily shortwave on the next time horizon, and  $D_{t+n}$  is the mean daily discharge on the next time horizon.  $\epsilon$  is normally distributed random noise with mean = 0 and standard deviation of  $\sigma$ . Parameter ( $\beta_1$ ,  $\beta_2$ ,  $\beta_3$ , and  $\beta_4$ ), and process error standard deviation ( $\sigma$ ) values from the linear model fitting were used as initial parameter values in a Markov Chain Monte Carlo simulation when forecasting (see ‘*Bayesian forecasting framework*’ for more details).

Given that the model drivers likely have varying effects on phytoplankton at different time scales, we performed the model selection analysis separately for the daily and fortnightly datasets to determine if the same covariates would be significant for the different time scales. We found that for both the fortnightly and daily datasets, the same model (covariates of mean shortwave radiation and mean discharge) was within 2 AICc units of the top model (Tables S6-S8, indicating a similar model performance (Burnham and Anderson 2002)). Therefore, we used the same weekly model covariates (an AR term, shortwave radiation, and discharge) for all model time steps (daily, weekly, fortnightly) in our forecast analysis.

To examine the influence of drivers averaged over daily to weekly scales, we developed models using drivers which were cumulatively averaged over a week compared to drivers which were point estimates on the day of the forecast. We found little difference in the effect of model prediction accuracy (Appendix S1: Figure S2). Thus, the results presented here use driver data which are forecasted or observed on the day being predicted.

Although the datasets used to develop AR forecasts spanned different durations and temporal resolutions (Appendix S1: Table S2), we used a similar number of observations to train all models with differing time steps to enable comparability of model forecast performance when logistically possible. Due to data availability, the training dataset for the daily chl-a models ranged from 15 August 2018 to 15 December 2018 ( $n = 110$  observations), while the weekly and fortnightly dataset ranged from May-October each year in 2013-2016 ( $n = 107$  weekly observations and  $n = 53$  fortnightly observations).

### *Bayesian forecasting framework*

Using the model covariates selected from the weekly training dataset (mean shortwave radiation and mean discharge) and initial parameter values, we used a Bayesian framework to fit distributions for model parameters on each forecast day for the daily, weekly, and fortnightly models following the workflow outlined in Figure 2, Step 2a. This fitting occurred iteratively on every day in which a forecast was produced. Forecasts were produced and evaluated from 2 January 2019 to 15 August 2020 (hereafter, the full forecast period) for all forecast models.

Our state-space Bayesian model followed the same form as Eq. 1, except with the process model and data model separated to allow the partitioning of model process and observation uncertainty. The process model was:

$$C_{t+n}^{latent} \sim normal(\beta_1 + \beta_2 C_t^{latent} + \beta_3 S_{t+n} + \beta_3 D_{t+n}, \sigma_{add}) \quad (\text{Eq. 2})$$

with the following data model:

$$y_{t+n} \sim normal(C_{t+n}^{latent}, \sigma_{obs}) \quad (\text{Eq. 3})$$

where  $\sigma_{add}$  and  $\sigma_{obs}$  are the standard deviation parameters describing normally distributed additive process uncertainty and normally distributed observation uncertainty, respectively.  $y$  are

the observations and  $C_t^{latent}$  are the latent states that represent the modeled distribution before observation noise is added. All other parameters are the same as described above in Eq. 1. We used uninformative priors for the  $\beta_1, \beta_2, \beta_3$ , and  $\beta_4$  parameters by assuming normal distributions with large standard deviations:

$$\beta_i \sim normal(0, 1000) \text{ (Eq. 4)}$$

Similarly,  $\sigma_{add}$  had an uninformative prior with a uniform distribution:

$$\sigma_{add} \sim uniform(0.001, 100) \text{ (Eq. 5)}$$

$\sigma_{obs}$  was a constant that was estimated using the standard deviation of a linear regression between our two sensor datasets of chl-a derived from the CTD chl-a and EXO chl-a (Appendix S1: Section S3).

To estimate posterior distributions, we ran four MCMC chains, with an adaptation period of 1,000 iterations, a burn-in period of 1,000 iterations, and a sample size of 10,000 iterations. Initial starting values for each parameter were taken from the linear model fit for each parameter value (see *Model development and training*). The latent state was initialized using the first observation in the training dataset, and the MCMC was run using all data up to the day being forecasted. The posterior output from the MCMC distribution for all parameters and latent states were saved as inputs to the forecast model to quantify uncertainty (Figure 2, Step 5). MCMC chains were assessed for convergence using the potential scale reduction factor of the Gelman-Rubin statistic ( $R^\wedge$ ). All  $R^\wedge$  values for all parameters in all forecast models were 1 or 1.01, indicating that the model had converged on a parameter estimate both within and among MCMC chains.

Observed model driver and validation data were automatically uploaded to a GitHub repository in real-time (Figure 2, Step 1) (<https://github.com/FLARE-forecast/FCRE-data>) and

added (or assimilated) into the training dataset for fitting model parameter distributions in the Bayesian framework (Figure 2, Step 2). New data were assimilated according to the time step of the model (i.e., daily, weekly or fortnightly). All models (daily, weekly, and fortnightly) were re-fit using the entire dataset (including training data) up to the day being forecast using MCMC (Figure 2, Step 3) methods for each forecast day to allow the parameter values to evolve over time. MCMC analyses were carried out using the 'rjags' package (Plummer, 2019) within the R statistical environment. For each forecast model and for every forecast run, we estimated the posterior distributions of parameters using MCMC, then sampled the posterior distributions of parameters, process error, and latent states to combine with forecasted driver variables to produce forecasts of chl-a at each forecast horizon.

### *Forecast evaluation*

To address Q1, we evaluated our forecasts' performance post-hoc for three time periods: 1) the entire ~600 day period during which multiple forecasts were produced at FCR, 2) non-bloom conditions, and 3) bloom conditions. Bloom conditions were determined as any time when observed chl-a concentrations were above a bloom threshold. The threshold for bloom conditions was determined as four times the standard deviation of historical chl-a at FCR from CTD chl-a dataset (17.1  $\mu\text{g/L}$ ; Carey et al. 2019). We chose four standard deviations as our threshold, rather than three standard deviations (e.g., Healy 1979), because of the high-frequency nature of our sensor data, which may increase the likelihood of outliers due to sensor fouling and phytoplankton quenching (Hamilton et al. 2010, McBride and Rose 2018, Roussio et al. 2021). Forecasts were evaluated as occurring under bloom conditions any time the observed chl-a was above this threshold on the day the forecast was initiated or any days when chl-a was above this

threshold within the 14-day forecast horizon. We included this second set of days in order to evaluate anticipatory predictions of bloom conditions before they occurred. Forecast performance over time was calculated for all three of these periods using root mean squared error (RMSE).

We also quantified the performance of our forecasts relative to a null persistence forecast (following Harris et al. 2018). A null persistence model assumes that the chl-a concentration at the next time step is the same as the current time step while accounting for process error and observation uncertainty. The null forecast was fit for each forecast day in *rjags* using a random walk MCMC simulation following the same methods as described above in *Methods: Bayesian Forecasting Framework* in which process error was estimated and observation uncertainty was calculated by sampling from a normal distribution around the observed chl-a concentration (s.d. = 0.21, Appendix S1: Section S3). To create an ensemble null forecast, we randomly sampled 441 times from the distribution of predicted chl-a for each forecast day. For RMSE statistics, the mean of all ensembles was used.

Lastly, to address Q2, we performed the same forecast evaluation as above and compared the relative skill of our forecast models across matching forecast horizons (e.g., we compared daily forecasts to weekly forecasts at the 7-day horizons, and daily, weekly, and fortnightly forecasts at the 14-day horizon).

#### *Uncertainty quantification and partitioning (Step 5 in Figure 2)*

To address Q3, we quantified total uncertainty in all forecasts and partitioned uncertainty among the individual sources (process, initial condition, parameter, and driver data, Table 1) using a separate post-hoc analysis (Figure 2, Step 5). All sources of uncertainty were included in



each forecast (Figure 2, Step 6). We chose 441 ensembles by multiplying the 21 NOAA GEFS ensembles by 21 discharge forecast ensembles. The resulting number of 441 ensembles was both within the constraints of computational abilities and large enough to allow for a reasonable spread of uncertainty. Further, 441 ensembles was below the effective sample size of all parameters within the MCMC chain, and therefore an appropriate number of samples to pull from the posterior distributions (Appendix S1: Table S9).

To quantify initial condition and parameter uncertainty, we sampled a randomly selected index from the joint posterior distributions of the latent state and parameter distributions, respectively, 441 times. Similarly, process error was estimated using a normal distribution with a mean of zero and a standard deviation sampled from the MCMC posterior distribution of  $\sigma_{add}$  (Eq. 5) for each of the 441 ensemble members. Meteorological driver data uncertainty was estimated by running the forecasts with the 21 unique NOAA GEFS meteorological forecasts, looping through the 21 ensembles for all 441 forecast ensemble members. Because the NOAA GEFS forecasts were statistically downscaled from regional forecasts to our study location (see Thomas et al. 2020 for methods), there is inherent uncertainty in downscaling, although we did not quantify its contribution in this study. Discharge driver data uncertainty was estimated by randomly sampling from a normal distribution around the discharge forecast for each ensemble member, where the normal distribution was defined using the discharge forecast estimate as the mean and the standard deviation of the residuals of the linear model between observed and forecasted discharge as the standard deviation in the normal distribution (following Thomas et al. 2020).

To partition the relative contributions of uncertainty sources through time, we performed a post-hoc uncertainty analysis conducted for all forecast models. We quantified each individual

source of uncertainty by propagating only that source, using the methods described above, while removing the contribution of the other sources for each week forecast (daily, weekly, and fortnightly). In the analysis, initial condition uncertainty was removed by initializing the forecast using the mean value of the posterior distribution of the latent state of chl-a, parameter uncertainty was removed by using the mean values of the posterior parameter distribution, process uncertainty was removed by not adding any process uncertainty (i.e., not sampling from the normal distribution describing model error), meteorological uncertainty was removed by using only one of the members from the weather forecast ensemble, and discharge uncertainty was removed by not sampling from the normal distribution describing error in the discharge model. The proportion of variance in the forecast output for each uncertainty source relative to total variance summed across all sources was then calculated over the forecast period and for each forecast horizon.

#### *Forecast scalability to new locations*

To address Q4 and test the scalability of our forecasting framework to other locations, we produced forecasts at an additional drinking water reservoir, Beaverdam Reservoir (BVR). BVR is also located in Vinton, Virginia (37.31, -79.82) and is owned and operated by the WVWA (Appendix S1: Figure S3). BVR is slightly larger (0.39 km<sup>2</sup>) and deeper (13.0 m maximum depth) than FCR and is fed by numerous small inflow streams. Discharge in two of the largest contributing inflows has been sampled occasionally, allowing us to produce modelled estimates of inflow discharge to the reservoir. Like FCR, BVR is dimictic with thermal stratification occurring annually from approximately May-October (Carey et al. 2019). BVR has a history of deep-water phytoplankton blooms (Hamre et al. 2018), but has overall lower phytoplankton

concentrations at the surface than FCR, allowing for a novel comparison between forecasts of surface phytoplankton in a system with lower surface phytoplankton concentrations (FCR historical median chl-a: 5.4  $\mu\text{g/L}$ ; BVR historical median chl-a: 3.4  $\mu\text{g/L}$ ). Because the aim of this analysis was to assess the scalability of our forecasting framework to a new site, we chose to use only the weekly model of our three (daily, weekly, fortnightly) forecast model timesteps at BVR, as it was the intermediate option of the three and provided training data dating back to 2013.

Moreover, BVR is an ideal test case for our scalability analysis due to its availability of data streams for applying our forecasting model. In August 2020, an EXO sonde collecting chl-a fluorescence data was installed at the surface ( $\sim 1.5$  m) of BVR (Carey et al. 2021b). Because of its proximity to FCR, we used the same observational meteorological data (daily mean shortwave radiation) for model training, and NOAA GEFS forecasts of shortwave radiation to drive the forecasts.

Producing estimates of inflow to BVR was slightly more challenging than in FCR because the reservoir's inflows have not been routinely monitored, nor is there a high-frequency sensor measuring discharge into the reservoir. However, we used a publicly available watershed model, the Thornthwaite-Mather Water Balance (TWMB) model (see Appendix S1: Section S4 for a full description), coupled with observed soil characteristic data (Soil Survey Staff 2021), and observed precipitation and air temperature from our meteorological station to model daily inflow rates to BVR over the training period (2013-2016). When producing forecasts of discharge at BVR, we used NOAA GEFS forecasts of daily precipitation and air temperature to drive the TWMB model.

In summary, to produce weekly forecasts of chl-a at BVR, we followed the same framework as at FCR (see Methods: Forecasting Framework above), resulting in 1-week and 2-week forecasts every day from 1 August 2020 – 31 August 2021. On each day a forecast was made, all available observational data up to the day being forecasted was used to fit the model in a Bayesian framework. We then used forecasts of shortwave radiation from NOAA GEFS, as well as NOAA GEFS forecasts of precipitation to drive the TWMB inflow model, to drive our chl-a forecasts. Additionally, we produced 1-week and 2-week forecasts at FCR over this same period and conducted a full uncertainty analysis for both FCR and BVR to compare the relative performance of the forecast model at both sites.

## RESULTS

### *Observed chl-a patterns during study period*

Chl-a exhibited substantial variability over the forecast period, which included two blooms in July 2019 and March-April 2020 and rapid declines due to two copper sulfate dosing events in February and March 2019 (Appendix S1: Figure S4). Observed chl-a over the forecast period exhibited a median chl-a concentration of  $7.3 \pm 7.2$   $\mu\text{g/L}$  (1 s.d.), and a maximum concentration of 55.7  $\mu\text{g/L}$ , which occurred during the July 2019 bloom. The magnitude of this large bloom greatly exceeded any previous surface bloom recorded by chl-a sensors in the reservoir (Appendix S1: Figure S4). For comparison, the training dataset from May to October of 2013-2016 had a slightly lower median concentration of  $4.9 \pm 4.3$   $\mu\text{g/L}$ , and the highest recorded bloom during this period was 24.2  $\mu\text{g/L}$  in October 2014 (Appendix S1: Figure S4). Further, the two blooms which occurred during the forecast period exhibited different patterns. The 2019 bloom lasted from 17 July to 5 August 2019, whereas the 2020 bloom lasted intermittently from

16 March to 23 April 2020 and was much more variable in concentration (Appendix S1: Figure S4).

The reservoir was actively managed as a drinking water supply during the forecast period and was treated with copper sulfate twice in 2019 to reduce phytoplankton concentrations. Managers added 200 lbs (90.7 kg) of copper sulfate on 28 February 2019 and 100 lbs (45.4 kg) on 20 March 2019, effectively decreasing chl-a concentrations to  $\sim 0$   $\mu\text{g/L}$  on both days. These time periods were excluded from subsequent analyses because they are not an ecological phenomenon instantiated in the model and thus could not have been predicted.

*Q1: Forecasts over the full ~600 day forecast period*

Aggregated over the entire forecast period, daily, weekly, and fortnightly forecasts predicted non-bloom chl-a dynamics with consistent accuracy (Figure 3), although bloom events were predicted with variable accuracy. Daily forecasts never outperformed a daily null model, but weekly and fortnightly forecasts performed better or the same as their respective null models (Figure 4a).

Daily forecasts at a 1-day horizon over the entire forecast period recreated observed dynamics with an overall RMSE of 3.9  $\mu\text{g/L}$  (Figure 3, Appendix S1: Table S10), less than the observed historical standard deviation in chlorophyll-a in this system (4.3  $\mu\text{g/L}$ ), a range of error which should still allow managers to make decisions with confidence. However, daily forecasts at the 1-day horizon did not perform better than the null model, which also had a lower RMSE (2.8  $\mu\text{g/L}$ ) than the historical standard deviation (Figure 4a, Appendix S1: Table S10). Weekly forecasts predicted chl-a with an RMSE of 6.1  $\mu\text{g/L}$  seven days ahead and an RMSE of 6.8  $\mu\text{g/L}$  at 14 days ahead (Figure 3, Appendix S1: Table S10), outperforming the null model at both

forecast horizons (Figure 4a). Similarly, fortnightly forecasts also outperformed the null model at the 14-day horizon, with an RMSE of 7.0  $\mu\text{g/L}$ , compared to 7.9  $\mu\text{g/L}$  for the null model (Figure 4a, Appendix S1: Table S10).

Based on AR model selection, our forecasts were driven by covariates of shortwave radiation and discharge into the reservoir. Both discharge and shortwave showed a negative relationship with chl-a in all forecast models, although the coefficients showed some variation in magnitude over time. In contrast, the AR chl-a coefficient, intercept, and error term all remained positive throughout the forecast period, although their magnitude varied (Appendix S1: Figure S6-S8). The parameter values for the environmental predictors (solar radiation, discharge) were much higher in the weekly forecasts than daily, and slightly higher in the fortnightly forecasts than the weekly forecasts, indicating that solar radiation and discharge were stronger predictors of chl-a at longer time horizons. In contrast, the AR chl-a coefficient showed the opposite trend (higher in daily forecasts, and lower in weekly and fortnightly). This pattern indicates that daily forecasts are much more sensitive to the AR chl-a term than to environmental predictors, while weekly and fortnightly forecasts have more weight on the environmental predictors relative to daily forecasts.

#### *Q1: Forecasts during non-bloom vs. bloom conditions*

Forecast performance was consistently and substantially higher during non-bloom than bloom conditions (Figure 4b vs. 4c). During non-bloom conditions, daily forecasts still did not outperform the null model but at least had similar performance as the null until the 9-day forecast horizon (Figure 4b). In contrast, weekly and fortnightly forecasts were quite accurate under non-bloom conditions and outperformed the null persistence models at all forecast horizons (Figure

4b), with an overall RMSE of 3.5  $\mu\text{g/L}$  at the 7-day horizon and 3.1  $\mu\text{g/L}$  at the 14-day horizon for weekly forecasts (Figure 4b) and 2.8  $\mu\text{g/L}$  at the 14-day horizon for fortnightly forecasts (Appendix S1: Table S10).

In contrast, forecasts of blooms did not consistently outperform null forecasts. Bloom conditions were rare and occurred over 19 days during July and August 2019 and intermittently over a period of 25 days in March and April 2020, for a total of 44 out of 590 days, or 7.5% of the forecast period. During these bloom periods, daily forecasts always performed worse than the null model (Figure 4c, Appendix S1: Table S10), and weekly and fortnightly forecasts performed the same or worse than the null model. Further, forecast accuracy appreciably decreased during blooms, with RMSE values ranging from 12  $\mu\text{g/L}$  at the 1-day forecast horizon to 23.8  $\mu\text{g/L}$  for fortnightly forecasts at the 14-day horizon (Appendix S1: Table S10).

Forecast performance was markedly different between the July 2019 and March/April 2020 blooms (Figure 3). For the July 2019 bloom, daily forecasts were able to predict the magnitude of the bloom at all forecast horizons, but only accurately predicted the timing of the bloom at a 1-day horizon (Figure 3a-c). At 7-day and 14-day horizons, peak chl-a predictions highly exceeded the observed magnitude of the bloom, and were predicted only after the observed event occurred, with very large uncertainty. In contrast, during the smaller, less persistent bloom which occurred in March-April of 2020, daily forecasts were able to predict both bloom magnitude and timing.

In contrast to daily forecasts, weekly forecasts did not successfully predict the magnitude of observed chl-a concentration for the July 2019 bloom, with peak values being much lower than observed chl-a concentrations (Figure 3d-e). Like the daily forecasts, weekly forecasts much more accurately predicted the timing and magnitude of the March-April 2020 bloom, with

observed concentrations falling within the confidence intervals of the forecast ensemble for almost all 7-day forecasts and 14-day forecasts.

Fortnightly forecasts underpredicted the magnitude and did not predict the timing of the July 2019 bloom (Figure 3e). However, the forecast performed much better during the March-April 2020 bloom, with only a few days of observed concentrations falling outside of the confidence intervals of the forecast.

### *Q2: Effect of model time step on forecast performance*

There was a consistent trend in the effect of model time step on forecast performance, as determined by comparing the forecasts generated by multiple forecast models (daily, weekly, fortnightly) at the 7-day and 14-day time horizons. Weekly and fortnightly forecasts consistently outperformed daily forecasts, a trend which increased with forecast horizon except during bloom conditions (Figure 4). Over the full forecast period, the weekly and fortnightly forecasts had a  $>5$   $\mu\text{g/L}$  improvement in RMSE over the daily model at the 14-day forecast horizon (Figure 4a, Appendix S1: Table S10). However, the effect of model time step was not as pronounced between the weekly and fortnightly models, with weekly forecasts sometimes only slightly outperforming fortnightly forecasts (e.g., Figure 4a vs. 4c).

### *Q3: Uncertainty partitioning analysis*

Process error was the dominant source of uncertainty at all forecast horizons for all forecast models, except for the 1-day forecast horizon, which had initial condition uncertainty as the dominant source (Figure 5, S9). The relative contributions of the different uncertainty sources varied over time for all forecast models (Figure 5). Process error remained the dominant



source of uncertainty for most of the forecast period at all horizons greater than 1 day. However, parameter uncertainty increased dramatically during the March 2019 copper sulfate dosing and the July 2019 and March/April 2020 blooms. During these events, parameter uncertainty became the dominant source of uncertainty for daily forecasts at 7-day and 14-day horizons (Figure 5a, c) and increased sharply in importance for weekly and fortnightly forecasts (Figure 5b, d).

Meteorological driver data uncertainty was a small contributor to the total uncertainty for all forecast models but increased in contribution in the late winter and early spring of 2019 (Figure 5b, c, d). Discharge uncertainty made up a very small yet consistent proportion of overall uncertainty throughout the year for all forecast models and was a larger contribution of uncertainty at the 14-day horizon for all forecasts (Figure 5b, d).

Interestingly, the July 2019 bloom resulted in a greater increase in parameter uncertainty and total forecast uncertainty than the March/April 2020 bloom (Figure 5). The increase in both parameter and total uncertainty in July 2019 was much larger for daily forecasts than for weekly or fortnightly forecasts. In particular, total variance increased by over seven-fold for the daily model at the 14-day forecast horizon during the July 2019 bloom. Total variance also increased during copper sulfate dosing events (late February and March) although this increase was much smaller for weekly forecasts than for daily forecasts (Figure 5a, c).

#### *Q4: Forecast scalability to new locations*

Our forecast model framework was successfully scaled from FCR to BVR, producing weekly forecasts of surface chl-a at both sites from August 2020-August 2021. Forecasts at BVR matched well with observed chl-a at both the 1-week and 2-week horizon (RMSE = 2.44  $\mu\text{g/L}$ ; 3.07  $\mu\text{g/L}$ , respectively), with a small decline in performance up to the 2-week forecast horizon

(Figure 6a, b). All forecasts at BVR showed very high uncertainty around the forecast mean throughout the forecast time period. Forecasts of chl-a at FCR over the same time period (RMSE = 5.36  $\mu\text{g/L}$ ; 5.16  $\mu\text{g/L}$ ) were less accurate but had lower overall uncertainty (Figure 6c, d).

While process uncertainty is still by far the largest contributor of overall uncertainty (Figure 7) in forecasts for FCR and BVR, parameter uncertainty had a higher contribution in BVR than in FCR. Additionally, while the overall contribution is much lower, both BVR and FCR show a slight increase in weather driver uncertainty over the late spring to mid-summer time period (~February to August). Discharge driver uncertainty remained a very small contribution to overall uncertainty in FCR but was negligible in BVR. Total variance between the two sites was similar at both 1-week and 2-week forecast horizons.

## DISCUSSION

While the number of ecological forecasts is increasing (Luo et al. 2011, Rousso et al. 2020, Lewis et al. 2021), many questions remain regarding the appropriate time scale at which to develop ecological models for forecasting and management applications, the time horizon and conditions under which ecological variables are predictable, the major sources of uncertainty in forecasts, and their scalability across waterbodies (Clark et al. 2001, Petchey et al. 2015, Dietze et al. 2018). Our work indicates that forecast accuracy varies with model time step and forecast horizon, and that weekly and fortnightly chl-a forecast models can outperform null estimates under most conditions. Importantly, our work also shows that sources of uncertainty, as well as forecast accuracy, do not remain constant through time, and that examining changes in uncertainty can elucidate mechanisms behind changes in forecast performance (e.g. during bloom conditions when parameter uncertainty increases and forecast skill decreases). Further, we

show that our forecasting framework can produce accurate forecasts when scaled to another site. Altogether, our study advances understanding of forecasting phytoplankton dynamics in aquatic systems, and adds to the growing body of work regarding the forecastability of rare, short-lived events (such as blooms) in ecological systems broadly, which we explore below.

### *Importance of the forecast model time step*

Forecast performance differed with model time step and horizon (Figure 4), demonstrating that both are important considerations when choosing a forecast model. For the entire ~600 day forecast period (i.e., primarily non-bloom conditions), both weekly and fortnightly forecasts outperformed daily forecasts at all horizons, indicating that forecast accuracy depends on the time step of your model. Specifically, our study shows that tuning a forecast model to a longer time step may produce better forecasts than propagating a daily model to multiple horizons. In practical terms, this indicates that if you are aiming to forecast conditions in one week, rather than in one day, forecast performance may be improved by creating a model at the weekly timestep. Interestingly, fortnightly forecasts did not outperform weekly forecasts, which may be due to a saturating effect of model time step on forecast accuracy in predicting phytoplankton ~one week ahead. Although phytoplankton are known to quickly respond to changes on hourly to daily time scales (Reynolds 2006), our study demonstrates the benefit of exploring the relevance of the weekly time scale, a time period which may be better able to capture slower moving processes, such as nutrient delivery from inflow streams (Liu et al. 2019). Further, weekly to fortnightly lead times may actually be more useful for management decisions than daily forecasts, providing more time to implement a water quality intervention, such as ordering chemicals or filters for water treatment or choosing the depth of

extraction within the reservoir (Carey et al. 2021a).

### *Forecast skill compared to a null model*

All the forecasts performed the same or worse than a null persistence model during bloom conditions, and only weekly and fortnightly forecasts outperformed the null during non-bloom conditions (Figure 4a). Forecasts likely underperformed compared to a null model during bloom conditions due to an inability of our forecast models to accurately replicate the processes occurring during blooms. While weekly and fortnightly forecast models were more skilled at predicting changes in phytoplankton biomass during non-bloom conditions than a null model, the information added by our forecasted covariates was not sufficient to increase forecast skill over the null model during bloom conditions.

It is not unusual for phytoplankton forecasts to perform similarly or worse than null models, especially at the daily scale. While many forecasts of phytoplankton variables do not currently compare forecasts with a null model (Lewis et al. 2021), those that do report variable performance, with many showing that forecast skill is only greater than a null at monthly to yearly forecast horizons. For example, Park et al. (2019) found that their forecasts of marine chl-*a* concentrations outperformed a null model at horizons ranging from one month to one year. In contrast, Page et al. (2018) and Kehoe et al. (2019) both found that their daily phytoplankton forecasts performed similarly or worse than a null persistence model, in some cases with no improvement as forecast horizon increased. These two phytoplankton studies support our findings that daily forecasts did not always perform better than a null model. However, the variability in null performance across different types of ecological forecasts (e.g., Harris et al. 2018, Massoud et al. 2018, Yeager et al. 2018, Lovenduski et al. 2019) underscores the need for

more studies to integrate null models into forecast evaluation metrics (following Harris et al. 2018, Dietze 2017a, Lewis et al. 2021).

### *Challenges in forecasting phytoplankton blooms*

There are multiple potential reasons why forecast accuracy was much worse during bloom conditions relative to non-bloom conditions (Figure 4). First, the decrease in forecast performance may be due to a decrease in autocorrelation of phytoplankton biomass during blooms, which result in rapid and steep increases in biomass, as opposed to non-bloom conditions, when biomass is relatively stable from day to day. Second, our 2013-2016 training data included very few blooms, and even though blooms did occur in our forecast period, a majority of the time series remains dominated by non-bloom conditions (92.5%). This means that the training dataset was inherently biased to predicting non-bloom, as opposed to bloom conditions. Third, the relative importance of processes governing phytoplankton populations may change between bloom and non-bloom conditions (Gray et al. 2019, Ho et al. 2019, Ho and Michalak 2019). Altogether, our work follows many other studies that have also observed that phytoplankton blooms are notoriously difficult to predict (e.g., Rigosi et al. 2011, Huang et al. 2013, Chen et al. 2015, Rigosi et al. 2015, Recknagel et al. 2016, Massoud et al. 2018, Page et al. 2018, Loos et al. 2020, McGowan et al. 2017), and underscore the need to expand our understanding of phytoplankton bloom dynamics.

Forecast accuracy differed between the July 2019 and March/April 2020 bloom events, indicating that phytoplankton dynamics during some blooms may be more predictable than

others. It is possible that the differences in predictive ability between blooms may be because of different phytoplankton taxa underlying the two blooms, but unfortunately we do not have the data to test this hypothesis (Appendix S1: Section S5). However, all forecast models better predicted the March/April 2020 bloom than the July 2019 bloom (maximum concentration = 55.7  $\mu\text{g/L}$ ), likely because the March/April 2020 bloom was much lower in peak concentrations (maximum concentration = 24.2  $\mu\text{g/L}$ ), and lasted a longer period of time, allowing the forecasts to adjust to elevated chl-a concentrations (see Appendix S1: Figure S6-S8 show dramatic changes in parameter values in response to bloom events).

Capturing both the timing of onset and magnitude of phytoplankton blooms is a critical benchmark of phytoplankton forecasts. While daily forecasts at a 1-day horizon recreated both the timing and magnitude of the two blooms, daily forecasts at time horizons greater than one day overpredicted the magnitude and missed the timing of the July 2019 bloom (Figure 3a, b, c). The autoregressive component of our forecast models likely contributed to the mismatch in the timing of bloom prediction at longer time horizons (e.g., Figure 3c). As the model was updated with the newest observed chl-a concentration, the daily model was able to predict higher concentrations, but only after concentrations were already elevated. As a result, the model did not accurately predict the bloom ahead of time, but only responded to already elevated initial conditions. In contrast, weekly and fortnightly forecasts did not accurately predict the magnitude or timing of the July 2019 bloom (Figure 3e-f). Accurately predicting both the timing and magnitude of blooms has long evaded phytoplankton research (Rigosi et al. 2011, Huang et al. 2013, Chen et al. 2015b, Recknagel et al. 2016, McGowan et al. 2017, Massoud et al. 2018, Page et al. 2018, Loos et al. 2020), with many of these studies successfully predicting non-bloom

conditions with decreased accuracy during bloom conditions, highlighting the importance of further understanding the factors contributing to the onset of phytoplankton blooms.

### *Forecast model simplicity*

Forecasting phytoplankton using a simple empirical modeling approach has limitations and benefits. Process-based approaches (e.g., Mao et al. 2009, Xiao et al. 2017, Page et al. 2018, Loos et al. 2020) are more likely to simulate numerous interacting processes, leaning less on historical patterns than empirical approaches (*sensu* Poff 2018). In our case especially, the lack of large blooms in our training dataset (92.5% non-bloom) inherently limited the ability of the empirical model to predict bloom dynamics. However, even state-of-the-art process-based models often fail to reproduce phytoplankton dynamics (e.g., Page et al. 2018, Loos et al. 2020), and numerous studies document that simpler models consistently out-perform more complicated models, especially when used in predictive applications (Ward et al. 2014, Chevalier and Knape 2020, Rouso et al. 2020, Wood et al. 2020). Further, using covariates which are readily predictable is crucial for forecasting applications, making simpler models with fewer covariates easier to convert from explanatory to predictive applications. Similarly, process-based forecasts such as Loos et al. (2020) and Page et al. (2018) require repeated updating of model states such as nutrient concentrations and phytoplankton functional groups. In lieu of frequent observations, these studies have to rely on model simulations to estimate these states.

Lastly, while our forecast model does not provide a direct explanation of mechanisms driving phytoplankton dynamics, our model selection procedure identifies covariates which are both readily forecastable and improve forecasts of phytoplankton over other simple models such as a null persistence model. Ultimately, a more balanced approach which leverages both empirical and process-based approaches may help to expand our predictive ability in phytoplankton ecology (e.g., Buckley et al. 2018, Briscoe et al. 2019, Read et al. 2019, Geary et al. 2020).

### *Forecast uncertainty*

Uncertainty partitioning analyses can increase understanding of the mechanisms which drive overall forecast uncertainty, ultimately leading to improvement in forecast skill (Dietze 2017a, Harris et al. 2018). At all forecast horizons >1 day, forecast uncertainty was dominated by process uncertainty, indicating that a better understanding of the mechanisms driving phytoplankton dynamics is critical to improving future forecasts. However, under bloom conditions (e.g., July 2019), parameter uncertainty increased substantially, likely an indication that parameter distributions were not inclusive of values which could accurately recreate bloom dynamics (Appendix S1: Figure S6-S8). Importantly, the increase in parameter uncertainty may be inherently linked to process uncertainty, given that the parameter values in our linear model were not able to reflect the abrupt changes in phytoplankton biomass, which may be better addressed by employing a different model structure entirely (e.g., exponential growth, see Rouso et al. 2020). Forecasts at BVR also exhibit high parameter uncertainty. To the best of our knowledge, it is not common for parameter uncertainty to be a dominant source of uncertainty in phytoplankton forecasts, although many phytoplankton forecasts do not specifically quantify this uncertainty source at all. Two other studies found parameter uncertainty dominated in forecasts



of red pine growth and snow geese populations (Gertner et al. 1996, Gauthier et al. 2016, respectively), although both studies used process-based models.

Despite many forecasting studies citing driver uncertainty as the dominant source of uncertainty (e.g., Mbogga et al. 2010, Dietze 2017, Jiang et al. 2018, Thomas et al. 2020), we found that driver uncertainty (meteorological and discharge) contributed only a very small portion of forecast uncertainty. While the contribution of meteorological driver uncertainty varied throughout the year (Figure 5), it only contributed a very small fraction of total uncertainty, despite that we used the same weather data forecast product (NOAA GEFS) as other studies.

Overall, our uncertainty analysis demonstrates that the dominant sources of uncertainty in phytoplankton forecasts may be different under bloom vs. non-bloom conditions, but that until process uncertainty is adequately constrained, other sources of uncertainty may contribute a minimal proportion of overall uncertainty. As a result, focusing efforts on improving process representation of bloom and non-bloom dynamics must be a priority before improvements to driver or initial condition uncertainty can be expected to substantially improve forecasts.

### *Forecast scalability*

Our scalability analysis showed that forecast performance remained high when applying the same forecast model and uncertainty structure in a new waterbody. Over a full year, forecasts at BVR accurately predicted chl-a concentrations within 2.44  $\mu\text{g/L}$  (RMSE) one week ahead, a promising result suggesting that autoregressive models can be readily applied in new systems. Despite the high performance of the forecast mean when compared to observational chl-a,

overall forecast uncertainty bounds were higher than in forecasts produced for FCR over this same time period. This is due to increased process and parameter uncertainty, likely an indication that covariates in the autoregressive model selected for FCR may not be appropriate for BVR, which has distinct phytoplankton and hydrological dynamics (e.g., Hamre et al. 2017, Hamre et al. 2018). Over the forecasted time period in BVR, we did not see any major bloom events, providing limited support for examining how our forecast model performs under bloom conditions specifically. Additionally, BVR has numerous inflow streams, as opposed to a single primary inflow at FCR, which may lessen the importance of discharge as a driver for BVR's phytoplankton dynamics. Indeed, our uncertainty analysis between BVR and FCR showed that parameter uncertainty was higher in BVR than FCR, and that process uncertainty was dominant for both reservoirs, providing support for our hypothesis that different predictors may be important in BVR than in FCR.

Our scalability analysis indicates that our forecasting framework can be successfully transferred to other waterbodies. When scaling this framework to other sites, we recommend developing unique autoregressive models (with full model selection) for individual study sites to determine which covariates are most informative to driving dynamics in individual ecosystems. In lieu of long-term, high-frequency data streams for drivers such as inflow discharge, we demonstrate an application which uses freely available model structures (e.g., the TWMB model) which can use available data to estimate the necessary data streams. Other examples include using nationally available meteorological data products such as NLDAS or USGS stream gauge data. Overall, our scalability analysis is promising for expanding forecasting frameworks beyond a single system.

## *Conclusions*

Our forecasting system successfully predicted phytoplankton biomass at multiple time scales over the course of ~600 days, requiring only two driver data streams and as little as 5 months training data. Our forecasting system highlights the feasibility of producing ecological forecasts which leverage historical monitoring datasets and forecastable model covariates at multiple sites, as well as the value of performing a formal uncertainty analysis on forecasts. Additionally, we emphasize the importance of considering the time step of a forecast model when producing forecasts at longer time horizons, which is critical to improving the use of forecasts as decision support tools for managers and the public. Further, by applying our forecasting system to an additional study site, we show that simple forecast models can be adapted for new locations using limited input data. Ultimately, this study provides insight on the predictability of freshwater phytoplankton dynamics as well as helpful considerations for developing ecological forecasts in a diverse set of ecosystems.

## ACKNOWLEDGEMENTS

Author contribution statement: CCC and WMW conceptualized the study, with substantial help from RQT. The forecasting framework was developed by RQT and adapted for use in this study by WMW. All models and forecasts were run and analyzed by WMW. MEL and RPM led the extensive data collection in this study. HLW built the code-base for producing inflow discharge estimates at Beaverdam Reservoir. The manuscript was written by WMW, with substantial editing and reviewing from CCC; all authors provided feedback and approved the final version. We thank the Western Virginia Water Authority (WVWA) for their long-term access to field sites and continuing support and engagement in this research. The Smart Reservoir and FLARE

team - in particular, R.J. Figueiredo, V. Daneshmand, B.J. Bookout, and A. Breef-Pilz - were critical in developing the forecasting workflows that enabled this project. We thank many field and lab assistants, especially A. Gerling, K. Hamre, A. Hounshell, D. Howard, J. Doubek, K. Krueger, A. Lewis, N. Hammond, J. Maze, Z. Munger, N. Ward, and J. Wynne, as well as B. Niederlehner for her tireless efforts in sample analysis. We thank the Virginia Tech Stream Team, Carey Lab, Global Lake Ecological Observatory Network, and Ecological Forecasting Initiative (EFI), especially the EFI Student Association for their support, feedback, and thoughtful discussions. This work was supported by the National Science Foundation (CNS-1737424, DEB-1753639, DBI-1933016, DBI-1933102, DEB-1926050, a Graduate Research Fellowship to WMW), and the WVWA.

## REFERENCES

- Arhonditsis, G. B., M. Winder, M. T. Brett, and D. E. Schindler. 2004. Patterns and mechanisms of phytoplankton variability in Lake Washington (USA). *Water Research* 38:4013–4027.
- Berthet, L., O. Piotte, É. Gaume, R. Marty, and C. Ardilouze. 2016. Operational forecast uncertainty assessment for better information to stakeholders and crisis managers. *E3S Web of Conferences* 7.
- Box, G. E. P., and Jenkins, G. M. 1970. *Time Series Analysis Forecasting and Control*. San Francisco: Holden-Day.
- Burnham, K. P. and Anderson, D. R. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd edn. Springer, Berlin.
- Briscoe, N. J., J. Elith, R. Salguero-Gómez, J.J. Lahoz-Monfort, J.S. Camac, K.M. Giljohann, M.H. Holden, B.A. Hradsky, M.R. Kearney, S.M. McMahon, B.L. Phillips, T.J. Regan, J.R. Rhodes, P.A. Vesk, B.A. Wintle, J.D.L. Yen, and G. Guillera-Arroita. 2019. Forecasting species range dynamics with process-explicit models: matching methods to applications. *Ecology Letters*, 22(11), 1940–1956.
- Buckley, L.B. A.F. Cannistra, and A. John. 2018. Leveraging Organismal Biology to Forecast the Effects of Climate Change. *Integrative & Comparative Biology* 58(1):38-51.

Carey, C. C., W. M. Woelmer, M. E. Lofton, R. J. Figueiredo, B. J. Bookout, R. S. Corrigan, V.

Daneshmand, A. G. Hounshell, D. W. Howard, A. S. L. Lewis, R. P. McClure, H. L.

Wander, N. K. Ward, R. Q. Thomas. 2021a. Advancing lake and reservoir water quality management with near-term, iterative ecological forecasting. *Inland Waters*, DOI:

10.1080/20442041.2020.1816421

Carey, C.C., A. Breef-Pilz, and B.J. Bookout. 2021b. Time series of high-frequency sensor data

measuring water temperature, dissolved oxygen, pressure, conductivity, specific

conductance, total dissolved solids, chlorophyll a, phycocyanin, and fluorescent dissolved organic matter at discrete depths in Beaverdam Reservoir, Virginia, USA in 2020 ver 1.

Environmental Data Initiative.

Carey, C. C., A. S. Lewis, R. P. McClure, A. B. Gerling, J. P. Doubek, S. Chen, M. E. Lofton,

and K. D. Hamre. 2019. Time series of high-frequency profiles of depth, temperature,

dissolved oxygen, conductivity, specific conductivity, chlorophyll a, turbidity, pH,

oxidation-reduction potential, photosynthetic active radiation, and descent rate for

Beaverdam Reservoir, Carvins Cove Reservoir, Falling Creek Reservoir, Gatewood

Reservoir, and Spring Hollow Reservoir in Southwestern Virginia, USA 2013-2019 ver

10. Environmental Data Initiative. DOI:

[10.6073/pasta/1fc7d2a5c69c6a651793dba06d375ae2](https://doi.org/10.6073/pasta/1fc7d2a5c69c6a651793dba06d375ae2)

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* 46:1394–1407.

- Carpenter, S. R., Bolgrien, D., Lathrop, R. C., Stow, C. A., Reed, T., and Wilson, M. A. 1998. Ecological and economic analysis of lake eutrophication by nonpoint pollution. *Austral Ecology*, 23:68–79.
- Chen, Q., T. Guan, L. Yun, R. Li, and F. Recknagel. 2015. Online forecasting chlorophyll a concentrations by an autoregressive integrated moving average model: Feasibilities and potentials. *Harmful Algae* 43:58–65.
- Cheung, M. Y., Liang, S., and Lee, J. 2013. Toxin-producing cyanobacteria in freshwater: A review of the problems, impact on drinking water safety, and efforts for protecting public health. *Journal of Microbiology* 51:1–10.
- Chevalier, M. and J. Knappe. 2020. The cost of complexity in forecasts of populations abundances is reduced but not eliminated by borrowing information across space using a hierarchical approach. *Oikos*. 129:249-260.
- Clark, J. S., S. R. Carpenter, M. Barber, S. Collins, A. Dobson, J. A. Foley, D. M. Lodge, M. Pascual, R. P. Jr, W. Pizer, C. Pringle, W. V. Reid, K. A. Rose, O. Sala, W. H. Schlesinger, D. H. Wall, and D. Wear. 2001. Ecological forecasts: An emerging imperative. *Science*. 293:657–661.
- Dietze, M. C. 2017a. Ecological forecasting. Princeton University Press, Princeton.
- Dietze, M. C. 2017b. Prediction in ecology: A first-principles framework. *Ecological Applications* 27:2048–2060.
- Dietze, M. C., A. Fox, L. M. Beck-Johnson, J. L. Betancourt, M. B. Hooten, C. S. Jarnevich, T. H. Keitt, M. A. Kenney, C. M. Laney, L. G. Larsen, H. W. Loescher, C. K. Lunch, B. C. Pijanowski, J. T. Randerson, E. K. Read, A. T. Tredennick, R. Vargas, K. C. Weathers,

- and E. P. White. 2018. Iterative near-term ecological forecasting: Needs, opportunities, and challenges. *Proceedings of the National Academy of Sciences* 115:1424–1432.
- Dodds, W. K., 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* 43:12–19.
- Ewerts, H., A. Swanepoel, and H. H. du Preez. 2013. Efficacy of conventional drinking water treatment processes in removing problem-causing phytoplankton and associated organic compounds. *Water SA*. 39:739–749.
- Gauthier, G., G. Péron, J. D. Lebreton, P. Grenier, and L. Van Oudenhove. 2016. Partitioning prediction uncertainty in climate-dependent population models. *Proceedings of the Royal Society B: Biological Sciences* 283: 20162353
- Geary, W.L., M. Bode, T.S. Doherty, E.A. Fulton, D.G. Nimmo, A.I.T. Tulloch, V.J.D. Tulloch, and E.G. Ritchie. 2020. A guide to ecosystem models and their environmental applications. *Nature Ecology & Evolution* 4:1459–147.
- Gerling, A. B., R. G. Browne, P. A. Gantzer, M. H. Mobley, J. C. Little, and C. C. Carey. 2014. First report of the successful operation of a side stream supersaturation hypolimnetic oxygenation system in a eutrophic, shallow reservoir. *Water Research* 67:129–143.
- Gerling, A. B., Z. W. Munger, J. P. Doubek, K. D. Hamre, P. A. Gantzer, J. C. Little, and C. C. Carey. 2016. Whole-catchment manipulations of internal and external loading reveal the sensitivity of a century-old reservoir to hypoxia. *Ecosystems* 19:555–571.
- Gertner, G., P. Parysow, and B. Guan. 1996. Projection variance partitioning of a conceptual forest growth model with orthogonal polynomials. *Forest Science* 42:474–486.



- Hamilton, D. P., C. C. Carey, L. Arvola, P. Arzberger, C. Brewer, J. J. Cole, E. Gaiser, P. C. Hanson, B. W. Ibelings, E. Jennings, T. K. Kratz, F. P. Lin, C. G. McBride, D. de M. Marques, K. Muraoka, A. Nishri, B. Qin, J. S. Read, K. C. Rose, E. Ryder, K. C. Weathers, G. Zhu, D. Trolle, and J. D. Brookes. 2015. A Global Lake Ecological Observatory Network (GLEON) for synthesising high-frequency sensor data for validation of deterministic ecological models. *Inland Waters* 5:49–56.
- Hamilton, D. P., K. R. O’Brien, M. A. Burford, J. D. Brookes, and C. G. McBride. 2010. Vertical distributions of chlorophyll in deep, warm monomictic lakes. *Aquatic Sciences* 72:295–307.
- Hamre, K. D., Gerling, A. B., Munger, Z. W., Doubek, J. P., McClure, R. P., Cottingham, K. L., & Carey, C. C. 2017. Spatial variation in dinoflagellate recruitment along a reservoir ecosystem continuum. *Journal of Plankton Research*, 39:715–728.
- Hamre, K. D., Lofton, M. E., McClure, R. P., Munger, Z. W., Doubek, J. P., Gerling, A. B., ... Carey, C. C. 2018. In situ fluorometry reveals a persistent, perennial hypolimnetic cyanobacterial bloom in a seasonally anoxic reservoir. *Freshwater Science*, 37(3), 483–495.
- Hansson, L. A. 2000. Synergistic effects of food chain dynamics and induced behavioral responses in aquatic ecosystems. *Ecology*, 81:842–851.
- Harris, D. J., S. D. Taylor, and E. P. White. 2018. Forecasting biodiversity in breeding birds using best practices. *PeerJ* 6:e4278.
- Healy, M. J. 1979. Outliers in clinical chemical quality-control schemes. *Clinical Chemistry*, 25: 675–677

- Hines, J., A. Ebeling, A. D. Barnes, U. Brose, C. Scherber, S. Scheu, T. Tschardtke, W. W. Weisser, D. P. Gilling, A. M. Klein, and N. Eisenhauer. 2019. Chapter Nine - Mapping change in biodiversity and ecosystem function research: food webs foster integration of experiments and science policy. Pages 297–322 in N. Eisenhauer, D. A. Bohan, and A. J. Dumbrell, editors. *Advances in Ecological Research*. London, UK: Academic Press.
- Ho, J. C., A. M. Michalak, and N. Pahlevan. 2019. Widespread global increase in intense lake phytoplankton blooms since the 1980s. *Nature* 574:667–670.
- 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.
- Jiang, J., Y. Huang, S. Ma, M. Stacy, Z. Shi, D. M. Ricciuto, P. J. Hanson, and Y. Luo. 2018. Forecasting responses of a northern peatland carbon cycle to elevated CO<sub>2</sub> and a gradient of experimental warming. *Journal of Geophysical Research: Biogeosciences* 123:1057–1071.
- Jiménez Cisneros, B. E., T. Oki, N. W. Arnell, G. Benito, J. G. Cogley, P. Döll, T. Jiang, S. S. Mwakalila, T. Fischer, D. Gerten, R. Hock, S. Kanae, X. Lu, L. J. Mata, C. Pahl-Wostl, K. M. Strzepek, B. Su, B. van den Hurk. 2014. Freshwater resources. *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* : 229-269.
- Jolliffe, I. T., and Stephenson, D. B. (Eds.). 2003. *Forecast verification: A practitioner's guide in atmospheric science*. Chichester, UK: Wiley.

Kehoe, M. J., B. P. Ingalls, J. J. Venkiteswaran, and H. M. Baulch. 2019. Successful forecasting of harmful cyanobacteria blooms with high frequency lake data. *BioRxiv Preprint*.

<https://doi.org/10.1101/674325>

Komatsu, K. J., M. L. Avolio, N. P. Lemoine, F. Isbell, E. Grman, G. R. Houseman, S. E.

Koerner, D. S. Johnson, K. R. Wilcox, J. M. Alatalo, J. P. Anderson, R. Aerts, S. G. Baer,

A. H. Baldwin, J. Bates, C. Beierkuhnlein, R. T. Belote, J. Blair, J. M. G. Bloor, P. J.

Bohlen, E. W. Bork, E. H. Boughton, W. D. Bowman, A. J. Britton, J. F. Cahill, E.

Chaneton, N. R. Chiariello, J. Cheng, S. L. Collins, J. H. C. Cornelissen, G. Du, A.

Eskelinen, J. Firn, B. Foster, L. Gough, K. Gross, L. M. Hallett, X. Han, H. Harmens, M.

J. Hovenden, A. Jagerbrand, A. Jentsch, C. Kern, K. Klanderud, A. K. Knapp, J.

Kreyling, W. Li, Y. Luo, R. L. McCulley, J. R. McLaren, J. P. Megonigal, J. W. Morgan,

V. Onipchenko, S. C. Pennings, J. S. Prevéy, J. N. Price, P. B. Reich, C. H. Robinson, F.

L. Russell, O. E. Sala, E. W. Seabloom, M. D. Smith, N. A. Soudzilovskaia, L. Souza, K.

Suding, K. B. Suttle, T. Svejcar, D. Tilman, P. Tognetti, R. Turkington, S. White, Z. Xu,

L. Yahdjian, Q. Yu, P. Zhang, and Y. Zhang. 2019. Global change effects on plant

communities are magnified by time and the number of global change factors imposed.

*Proceedings of the National Academy of Sciences* 116:17867–17873.

Lewis, A. S. L., W. M. Woelmer, H. L. Wander, D. W. Howard, J. W. Smith, R. P. McClure, M.

E. Lofton, N. W. Hammond, R. S. Corrigan, R. Q. Thomas, and C. C. Carey. 2021.

Increased adoption of best practices in ecological forecasting enables comparisons of

forecastability. *Ecological Applications* 00(00):e02500. 10.1002/eap.2500

- Liu, X., J. Feng, and Y. Wang. 2019. Chlorophyll a predictability and relative importance of factors governing lake phytoplankton at different timescales. *Science of the Total Environment* 648:472–480.
- Loos, S., C. M. Shin, J. Sumihar, K. Kim, J. Cho, and A. H. Weerts. 2020. Ensemble data assimilation methods for improving river water quality forecasting accuracy. *Water Research* 171:115343.
- Lovenduski, N. S., G. B. Bonan, S. G. Yeager, K. Lindsay, and D. L. Lombardozzi. 2019. High predictability of terrestrial carbon fluxes from an initialized decadal prediction system. *Environmental Research Letters* 14:124074.
- Luo, Y., K. Ogle, C. Tucker, S. Fei, C. Gao, S. LaDeau, J. S. Clark, and D. S. Schimel. 2011. Ecological forecasting and data assimilation in a data-rich era. *Ecological Applications* 21:1429–1442.
- Manning, N. F., Y. C. Wang, C. M. Long, I. Bertani, M. J. Sayers, K. R. Bosse, R. A. Shuchman, and D. Scavia. 2019a. Extending the forecast model: Predicting Western Lake Erie harmful algal blooms at multiple spatial scales. *Journal of Great Lakes Research* 45:587–595.
- Manning, P., J. Loos, A. D. Barnes, P. Batáry, F. J. J. A. Bianchi, N. Buchmann, G. B. De Deyn, A. Ebeling, N. Eisenhauer, M. Fischer, J. Fründ, I. Grass, J. Isselstein, M. Jochum, A. M. Klein, E. O. F. Klingenberg, D. A. Landis, J. Lepš, R. Lindborg, S. T. Meyer, V. M. Temperton, C. Westphal, and T. Tschardtke. 2019b. Transferring biodiversity-ecosystem function research to the management of ‘real-world’ ecosystems. *Advances in Ecological Research* 61:323–356.

- Mao, J. Q., J. H. W. Lee, and K. W. Choi. 2009. The extended Kalman filter for forecast of algal bloom dynamics. *Water Research* 43:4214–4224.
- Marcé, R., G. George, P. Buscarinu, M. Deidda, J. Dunalska, E. De Eyto, G. Flaim, H. P. Grossart, V. Istvanovics, M. Lenhardt, E. Moreno-Ostos, B. Obrador, I. Ostrovsky, D. C. Pierson, J. Potužák, S. Poikane, K. Rinke, S. Rodríguez-Mozaz, P. A. Staehr, K. Šumberová, G. Waajen, G. A. Weyhenmeyer, K. C. Weathers, M. Zion, B. W. Ibelings, and E. Jennings. 2016. Automatic High Frequency Monitoring for Improved Lake and Reservoir Management. *Environmental Science and Technology* 50:10780–10794.
- Massoud, E. C., J. Huisman, E. Benincà, 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.
- Mbogga, M. S., X. Wang, and A. Hamann. 2010. Bioclimate envelope model predictions for natural resource management: dealing with uncertainty. *Journal of Applied Ecology* 47:731–740.
- McBride C. G., Rose K. C. 2018. Automated High-frequency Monitoring and Research. In: Hamilton D., Collier K., Quinn J., Howard-Williams C. (eds) *Lake Restoration Handbook*. Basel, Switzerland: Springer, Cham.
- McGowan, J.A., Deyle, E.R., Ye, H., Carter, M.L., Perretti, C.T., Seger, K.D., de Verneil, A. and Sugihara, G., 2017. Predicting coastal algal blooms in southern California. *Ecology*, 98(5), pp.1419-1433
- Millenium Ecosystem Assessment. 2005 *Ecosystems and human well-being: biodiversity synthesis*. World Resources Institute, Washington, D.C., USA.

- Morss, R. E., J. L. Demuth, and J. K. Lazo. 2008. Communicating uncertainty in weather forecasts: A survey of the U.S. public. *Weather and Forecasting* 23:974–991.
- Paerl, H. W., and V. J. Paul. 2012. Climate change: Links to global expansion of harmful cyanobacteria. *Water Research* 46:1349–1363.
- 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.
- Park, J., C. A. Stock, J. P. Dunne, X. Yang, and A. Rosati. 2019. Seasonal to multiannual marine ecosystem prediction with a global Earth system model. *Science*. 365:284-288.
- Petchey, O. L., M. Pontarp, T. M. Massie, S. Kéfi, A. Ozgul, M. Weilenmann, G. M. Palamara, F. Altermatt, B. Matthews, J. M. Levine, D. Z. Childs, B. J. McGill, M. E. Schaepman, B. Schmid, P. Spaak, A. P. Beckerman, F. Pennekamp, and I. S. Pearse. 2015. The ecological forecast horizon, and examples of its uses and determinants. *Ecology Letters* 18:597–611.
- Plummer, M. 2019. rjags: Bayesian Graphical Models using MCMC. R package version 4-9. <https://CRAN.R-project.org/package=rjags>
- Poff, N.L. 2018. Beyond the natural flow regime? Broadening the hydro-ecological foundation to meet environmental flows challenges in a non-stationary world. *Freshwater Biology*, 63(8), pp.1011-1021
- Qin, B., Zhu, G., Gao, G., Zhang, Y., Li, W., Paerl, H.W., and W.W. Carmichael. 2010. A drinking water Crisis in Lake Taihu, China: linkage to climatic variability and Lake management. *Environmental Management* 45:105-112.

- Quinn G. P., Keough M. J. 2002 Experimental design and data analysis for biologists. Cambridge University Press, Melbourne.
- Rajae, T., and A. Boroumand. 2015. Forecasting of chlorophyll-a concentrations in South San Francisco Bay using five different models. *Applied Ocean Research* 53:208–217.
- Read, E. K., V. P. Patil, S. K. Oliver, A. L. Hetherington, J. A. Brentrup, J. A. Zwart, K. M. Winters, J. R. Corman, E. R. Nodine, R. I. Woolway, H. A. Dugan, A. Jaimes, A. B. Santoso, G. S. Hong, L. A. Winslow, P. C. Hanson, and K. C. Weathers. 2015. The importance of lake-specific characteristics for water quality across the continental United States. *Ecological Applications* 25:943–955.
- Read, J. S., Jia, X., Willard, J., Appling, A. P., Zwart, J. A., Oliver, S. K., ... Kumar, V. (2019). Process-Guided Deep Learning Predictions of Lake Water Temperature. *Water Resources Research*, 55(11), 9173–9190.
- Recknagel, F., R. Adrian, J. Köhler, and H. Cao. 2016. Threshold quantification and short-term forecasting of *Anabaena*, *Aphanizomenon* and *Microcystis* in the polymictic eutrophic Lake Müggelsee (Germany) by inferential modelling using the hybrid evolutionary algorithm HEA. *Hydrobiologia* 778:61–74.
- Reynolds, C. S. 2006. The Ecology of Phytoplankton. Cambridge University Press.
- Rigosi, A., P. Hanson, D. P. Hamilton, M. Hipsey, J. A. Rusak, J. Bois, K. Sparber, I. Chorus, A. J. Watkinson, B. Qin, A. Rigosi, P. Hanson, D. P. Hamilton, M. Hipsey, J. A. Rusak, J. Bois, I. Chorus, A. J. Watkinson, B. Qin, B. Kim, and J. D. Brookes. 2015. Determining the probability of cyanobacterial blooms : the application of Bayesian networks in multiple lake systems. *Ecological Applications*. 25:186-199.

- Rigosi, A., R. Marcé, C. Escot, and F. J. Rueda. 2011. A calibration strategy for dynamic succession models including several phytoplankton groups. *Environmental Modelling and Software* 26:697–710.
- Rillig, M. C., M. Ryo, A. Lehmann, C. A. Aguilar-Trigueros, S. Buchert, A. Wulf, A. Iwasaki, J. Roy, and G. Yang. 2019. The role of multiple global change factors in driving soil functions and microbial biodiversity. *Science* 366:886–890.
- Rousso, 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.
- Rousso, B. Z., E. Bertone, R. A. Stewart, K. Rinke, D. P. Hamilton. 2021. Light-induced fluorescence quenching leads to errors in sensor measurements of phytoplankton chlorophyll and phycocyanin. *Water Research*, In Press.  
<https://doi.org/10.1016/j.watres.2021.117133>
- Sinha, R., Pearson, L. A., Davis, T. W., Burford, M. A., Orr, P. T., and Neilan, B. A. (2012). Increased incidence of *Cylindrospermopsis raciborskii* in temperate zones - Is climate change responsible? *Water Research* 46:1408–1419.
- Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. Web Soil Survey. 2021. Available online at the following link:  
<http://websoilsurvey.sc.egov.usda.gov/>.
- Smith M. D., Knapp A. K., Collins S. L. 2009. A framework for assessing ecosystem dynamics in response to chronic resource alterations induced by global change. *Ecology* 90:3279–3289.



- Solomon, S., Qin D., Manning M., Chen Z., Marquis M., et al. 2007. Technical summary. In Climate change 2007: the physical science basis. Contribution of Working Group I to the fourth assessment report of the Intergovernmental Panel on Climate Change. S Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller, editors. Cambridge University Press, Cambridge, UK.
- Tarczyńska, M., Z. Romanowska-Duda, T. Jurczak, and M. Zalewski. 2001. Toxic cyanobacterial blooms in a drinking water reservoir-causes, consequences and management strategy. *Water Science and Technology: Water Supply* 1:237–246.
- Thomas, R. Q, R. J. Figueiredo, V. Daneshmand, B. J. Bookout, L. Puckett, and C. C. Carey. 2020. A near-term iterative forecasting system successfully predicts reservoir hydrodynamics and partitions uncertainty. *Water Resources Research*. 56, e2019WR026138
- Vitousek, P. M. 1994. Beyond global warming: Ecology and global change. *Ecology* 75:1861–1876.
- Vollenweider R. A. 1975. Input-output models: with special reference to the phosphorus loading concept in limnology. *Aquatic Sciences*. 37:53–84.
- 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* 123:652–661.
- Wang, L., Q. Cai, Y. Xu, L. Kong, L. Tan, and M. Zhang. 2011. Weekly dynamics of phytoplankton functional groups under high water level fluctuations in a subtropical reservoir-bay. *Aquatic Ecology* 45:197–212.

- Wood, C. M., Loman, Z. G., McKinney, S. T., & Loftin, C. S. 2020. Testing prediction accuracy in short-term ecological studies. *Basic and Applied Ecology* 43:77–85.
- Xiao, X., J. He, H. Huang, T. R. Miller, G. Christakos, E. S. Reichwaldt, A. Ghadouani, S. Lin, X. Xu, and J. Shi. 2017. A novel single-parameter approach for forecasting algal blooms. *Water Research* 108:222–231.
- Yeager, S. G., G. Danabasoglu, N. A. Rosenbloom, W. Strand, S. C. Bates, G. A. Meehl, A. R. Karspeck, K. Lindsay, M. C. Long, H. Teng, and N. S. Lovenduski. 2018. Predicting near-term changes in the earth system: A large ensemble of initialized decadal prediction simulations using the community earth system model. *Bulletin of the American Meteorological Society* 99:1867–1886.
- Zeileis, A. and G. Grothendieck. 2005. zoo: S3 Infrastructure for Regular and Irregular Time Series. *Journal of Statistical Software*, 14: 1-27.

## TABLES

Table 1. Definitions of uncertainty sources that can contribute to total forecast uncertainty (derived from Dietze 2017a).

Uncertainty source	Definition
Driver data	Uncertainty in the forecasted estimates of model covariates (e.g., meteorology)
Initial condition	Uncertainty in the observed conditions when a forecast is created
Parameter	Uncertainty in model parameter values
Process	Uncertainty due to the inability of a model to reproduce observed conditions

## FIGURE LEGENDS

Figure 1. Map of Falling Creek Reservoir, Vinton, VA, USA (37.30°N, 79.84°W) study site showing locations of the weir, meteorological station, and chl-a sensor. Forecasts were generated for the location where chl-a measurements were collected.

Figure 2. Summarized workflow for development of forecasts, from data acquisition and quality assurance/quality control (QA/QC) (Step 1), assimilation of new data (Step 2), re-fitting of the model with new observations (Step 3), acquisition and processing of model driver data (Step 4), quantification of uncertainty sources (Step 5), and production of forecasts (Step 6). This process occurs once a day for each forecast model (daily, weekly, fortnightly). NOAA GEFS refers to the National Atmospheric and Oceanic Administration Global Ensemble Forecasts System.

Figure 3. Forecasted (black line) and observed (points) chl-a concentrations over ~600 days from daily (blue, panels a-c), weekly (green, panel d-e), and fortnightly (purple, panel f) forecasts; the columns are grouped by forecast horizon, at 1 day (left column), 7 days (middle column), and 14 days (right column). The black lines show the mean of 441 forecast ensembles, and the grey shaded area gives the 95% confidence intervals of the forecast ensembles. Please note the differences in axes among panels.

Figure 4. RMSE (root mean squared error) across 1-14 day forecast horizons for daily, weekly, and fortnightly forecasts and respective null models aggregated over a) the full ~600 day forecast period, b) non-bloom conditions only, and c) bloom conditions only. Note the different y-axis scale in panel c.

Figure 5. Relative proportion of forecast variance (uncertainty) over the forecast period (left y-axis) for a) daily forecasts at 1 day, b) daily forecasts at 7 days, c) daily forecasts at 14 days, d) weekly forecasts at 7 days, and e) weekly forecasts at 14 days, and f) fortnightly forecasts at 14 days. The black line represents the total variance over the forecasting time period (right y-axis).

Figure 6. Forecasted (black line) and observed (points) chl-a concentrations over 1 year at Beaverdam Reservoir (panels a, b) and Falling Creek Reservoir (panels c, d) from weekly (green, panel a, c), and fortnightly (purple, panel b, d) forecasts; the columns are grouped by forecast horizon, at 7 days (left column) and 14 days (right column). The black lines show the mean of 441 forecast ensembles, and the grey shaded area gives the 95% confidence intervals of the forecast ensembles. Please note the differences in axes among panels.

Figure 7. Relative proportion of forecast variance (uncertainty) for weekly forecasts from August 2020-August 2021. The top row shows the uncertainty analysis for Beaverdam Reservoir at 7 days (panel a) and 14 days (panel b), while the bottom row shows weekly forecasts for Falling Creek Reservoirs over the same time period at 7 days (panel c) and 14 days (panel d). The colors represent the relative contribution of each uncertainty source (left y-axis), while the black line represents the total variance over the forecast period (right y-axis).

## FIGURES

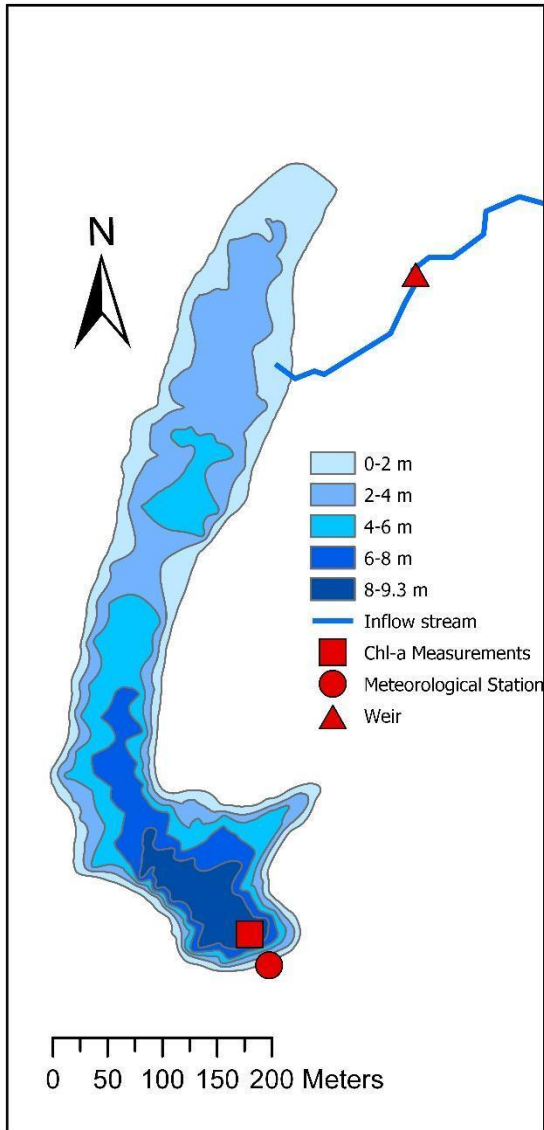


Figure 1. Map of Falling Creek Reservoir, Vinton, VA, USA (37.30°N, 79.84°W) study site showing locations of the weir, meteorological station, and chl-a sensor. Forecasts were generated for the location where chl-a measurements were collected.

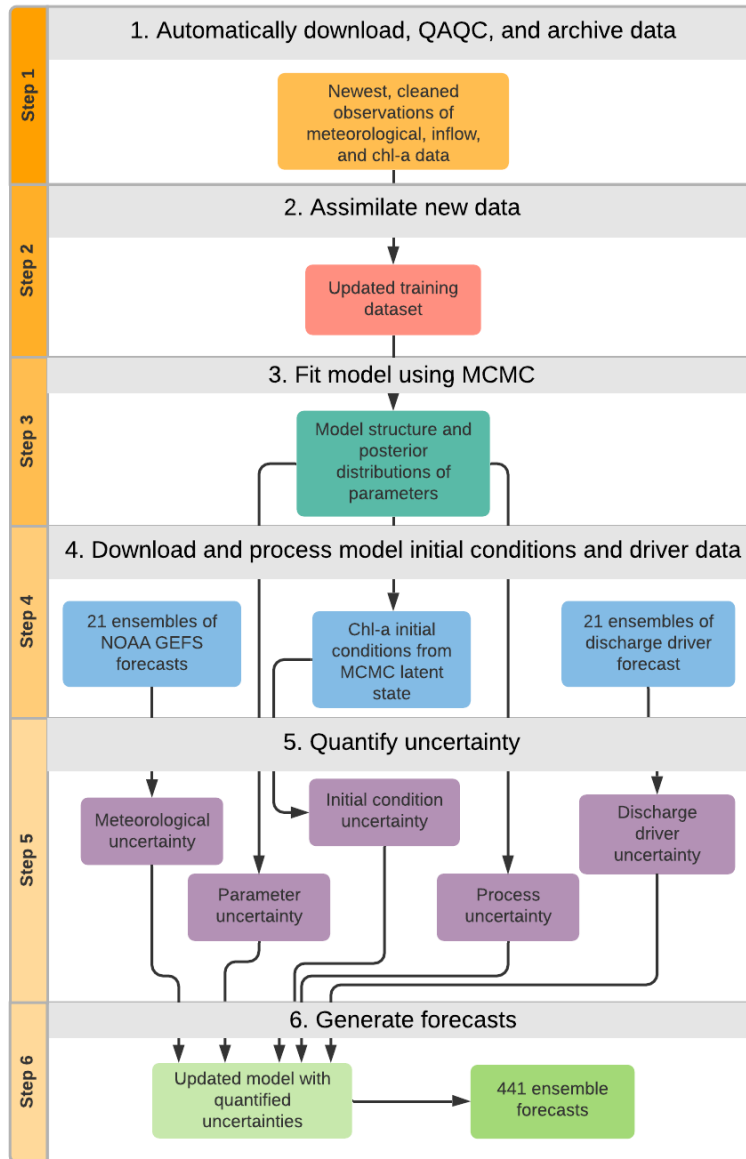


Figure 2. Summarized workflow for development of forecasts, from data acquisition and quality assurance/quality control (QA/QC) (Step 1), assimilation of new data (Step 2), re-fitting of the model with new observations (Step 3), acquisition and processing of model driver data (Step 4), quantification of uncertainty sources (Step 5), and production of forecasts (Step 6). This process occurs once a day for each forecast model (daily, weekly, fortnightly). NOAA GEFS refers to the National Atmospheric and Oceanic Administration Global Ensemble Forecasts System.



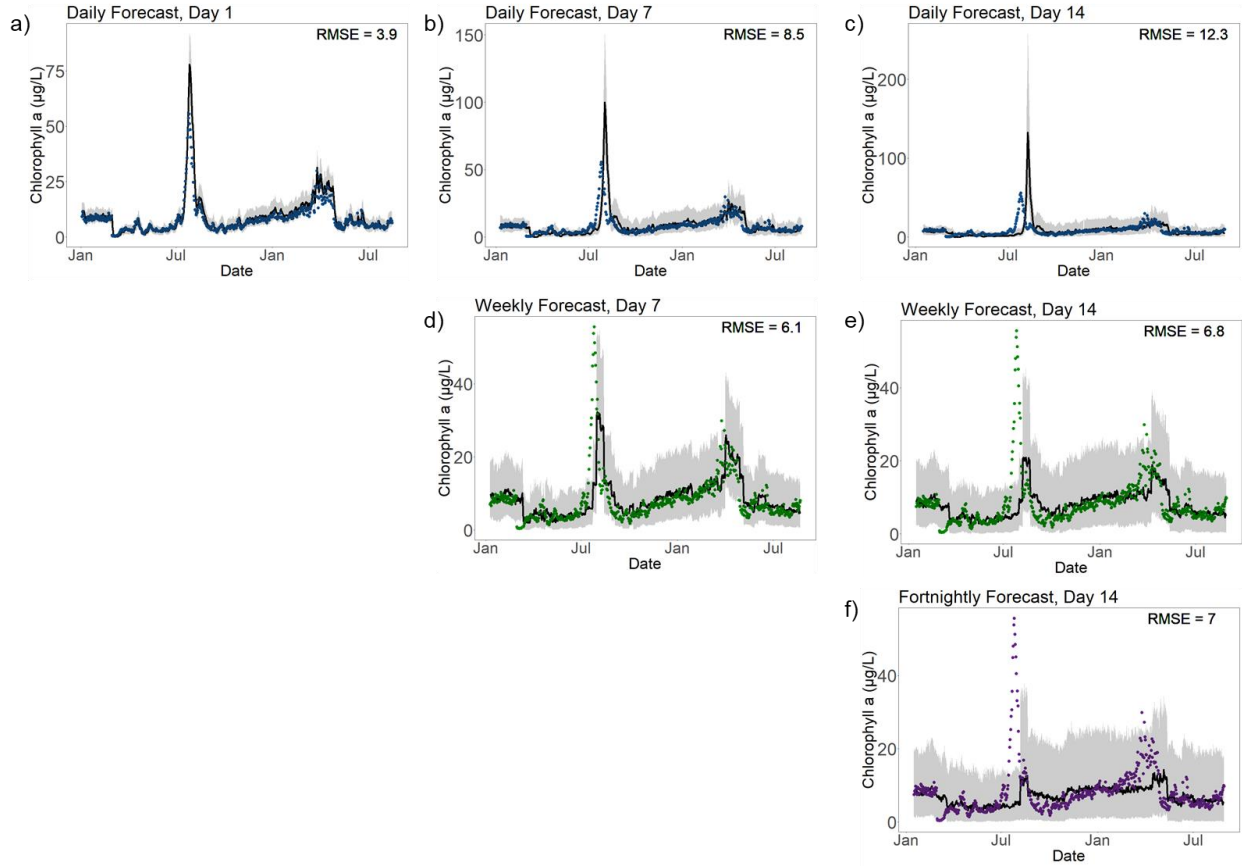


Figure 3. Forecasted (black line) and observed (points) chl-a concentrations over ~600 days from daily (blue, panels a-c), weekly (green, panel d-e), and fortnightly (purple, panel f) forecasts; the columns are grouped by forecast horizon, at 1 day (left column), 7 days (middle column), and 14 days (right column). The black lines show the mean of 441 forecast ensembles, and the grey shaded area gives the 95% confidence intervals of the forecast ensembles. Please note the differences in axes among panels.

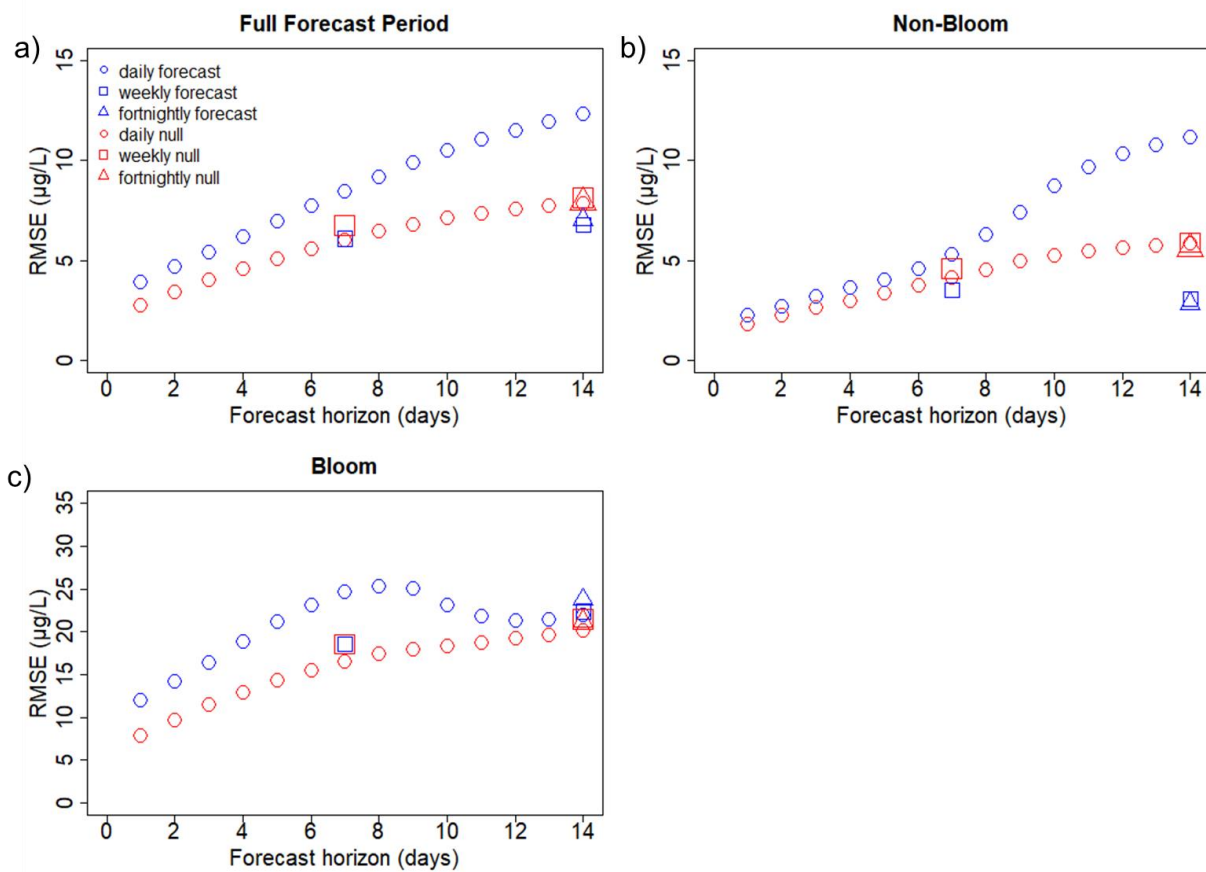


Figure 4. RMSE (root mean squared error) across 1-14 day forecast horizons for daily, weekly, and fortnightly forecasts and respective null models aggregated over a) the full ~600 day forecast period, b) non-bloom conditions only, and c) bloom conditions only. Note the different y-axis scale in panel c.

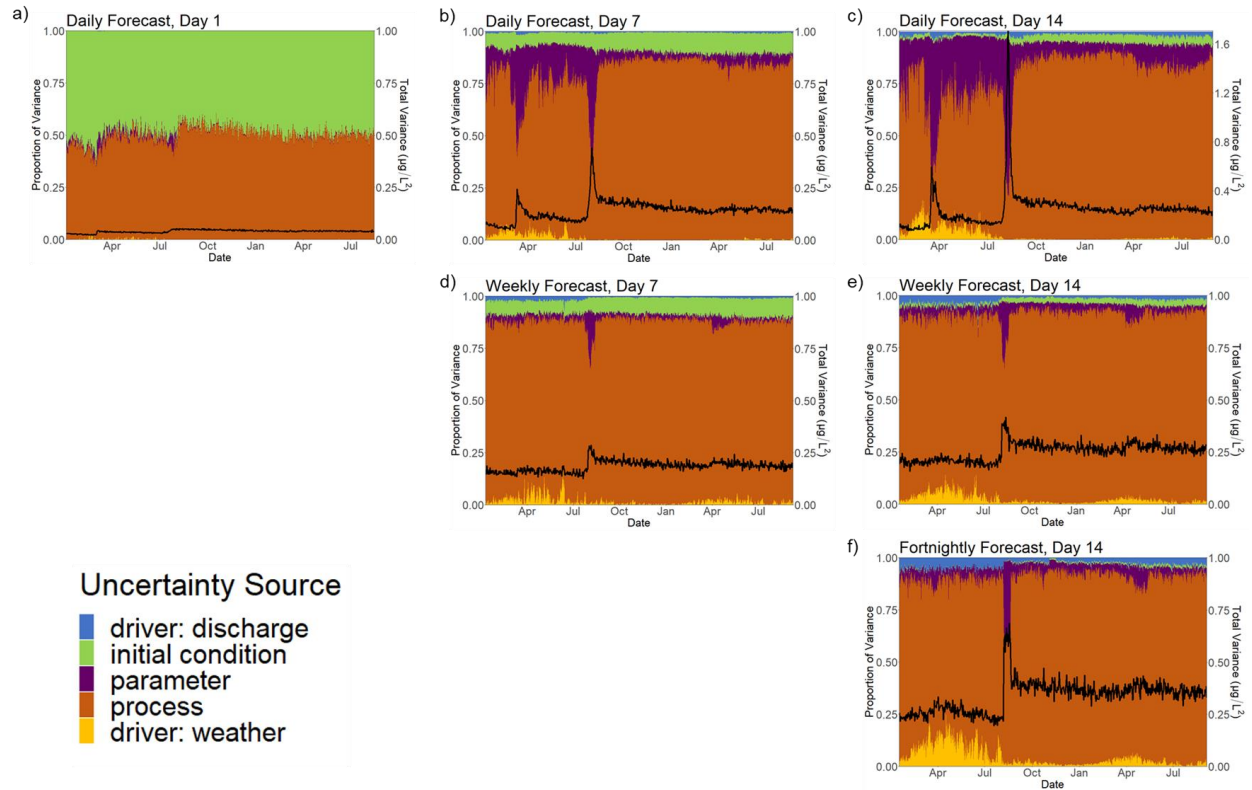


Figure 5. Relative proportion of forecast variance (uncertainty) over the forecast period (left y-axis) for a) daily forecasts at 1 day, b) daily forecasts at 7 days, c) daily forecasts at 14 days, d) weekly forecasts at 7 days, and e) weekly forecasts at 14 days, and f) fortnightly forecasts at 14 days. The black line represents the total variance over the forecasting time period (right y-axis).

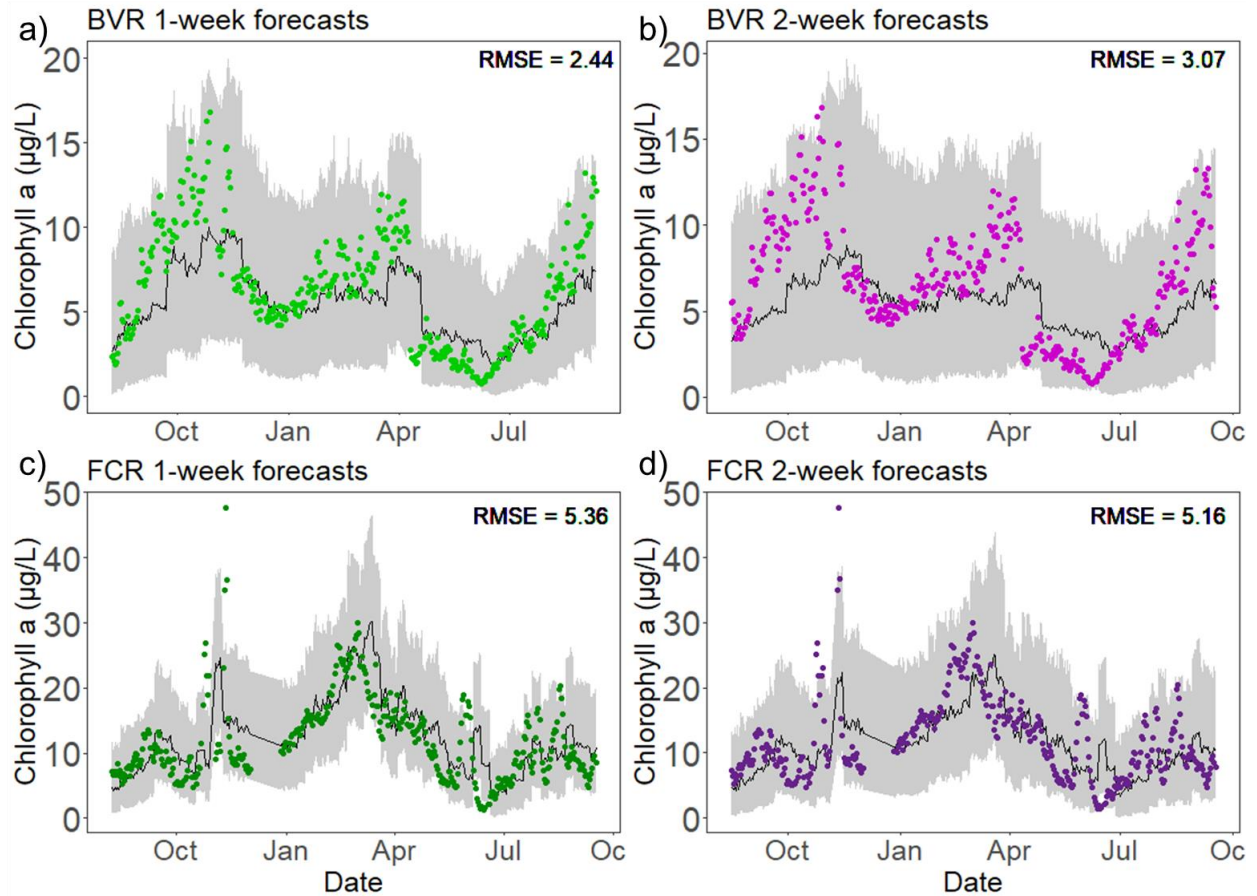


Figure 6. Forecasted (black line) and observed (points) chl-a concentrations over 1 year at Beaverdam Reservoir (panels a, b) and Falling Creek Reservoir (panels c, d) from weekly (green, panel a, c), and fortnightly (purple, panel b, d) forecasts; the columns are grouped by forecast horizon, at 7 days (left column) and 14 days (right column). The black lines show the mean of 441 forecast ensembles, and the grey shaded area gives the 95% confidence intervals of the forecast ensembles. Please note the differences in axes among panels.

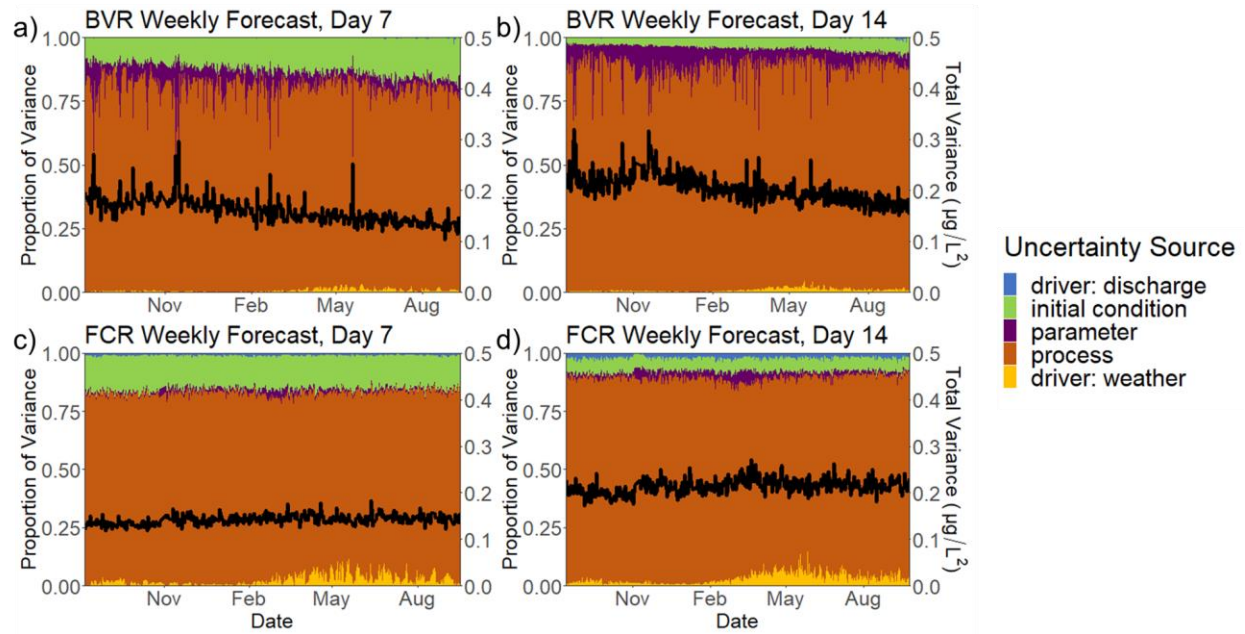


Figure 7. Relative proportion of forecast variance (uncertainty) for weekly forecasts from August 2020-August 2021. The top row shows the uncertainty analysis for Beaverdam Reservoir at 7 days (panel a) and 14 days (panel b), while the bottom row shows weekly forecasts for Falling Creek Reservoirs over the same time period at 7 days (panel c) and 14 days (panel d). The colors represent the relative contribution of each uncertainty source (left y-axis) , while the black line represents the total variance over the forecasting time period (right y-axis).

## **Appendix S1**

**Title:** Near-term phytoplankton forecasts reveal the effects of model time step and forecast horizon on predictability

**Authors:** Whitney M. Woelmer, R. Quinn Thomas, Mary E. Lofton, Ryan P. McClure, Heather L. Wander, Cayelan C. Carey

Accepted as an Article to *Ecological Applications*

**Pages:** 28

**Section:** 5

**Tables:** 10

**Figures:** 9

## Section S1. Reservoir monitoring dataset overview

*Surface chl-a:* The reservoir monitoring dataset includes chl-a data collected at multiple time scales over several years, which were used to develop AR forecasts at daily, weekly, and fortnightly time steps (Table S1; ‘datasets used in forecasting’). These data included weekly manual profiles of chl-a during 2013-2019 collected at the deepest site of the reservoir using a WET Labs ECO-AFL/FL chlorophyll fluorescence sensor deployed on an SBE 19plus V2 Conductivity, Temperature, and Depth (CTD) profiler (Seabird Electronics, Bellevue, WA) (henceforth referred to as ‘CTD chl-a’) (Carey et al. 2019). High-frequency measurements (10-minute intervals) of chl-a fluorescence were also measured from August 2018 to present using an EXO-sonde (YSI, Yellow Springs, OH) moored at ~1.0 m depth at the deepest site of FCR (henceforth referred to as ‘EXO chl-a’; Carey et al. 2021d).

*Meteorology:* A meteorological station was installed on the dam of FCR in July 2015 that measured ambient air temperature, relative humidity, rainfall, wind speed and direction, and shortwave and longwave radiation. Details regarding the meteorological sensors are outlined in Table S1, with comprehensive metadata provided by Carey et al. (2021b). Meteorological data from 2013-2015, before the meteorological station was installed on-site, and from two dates when local meteorological data were missing in late 2016 were obtained from the National Land and Data and Assimilation System (NLDAS-2; Xia et al. 2012).

*Inflow discharge:* Discharge at the major inflow to FCR (Figure 1) was measured from 2013 to present using an INW Aquistar PT2X pressure sensor (INW, Kirkland, WA) installed at a weir (Carey et al. 2021c). In spring 2019, a second pressure transducer, a Campbell Scientific CS451 pressure sensor, was installed at the weir. For both data streams, pressure was converted into discharge using a rating curve and weir-specific dimensions, following the method outlined by Carey et al. (2021c). All training data for discharge estimates were taken from the INW Aquistar pressure sensor. However, after the Campbell Scientific pressure sensor was installed, we transitioned to using that datastream for forecasting because it was automatically available for data download via a wirelessly streaming sensor gateway. A linear regression was developed to convert between the INW Aquistar PT2X and the Campbell Scientific CS451 pressure transducers (Section S3).

## References

- Carey, C. C., A. Breef-Pilz, B. J. Bookout, M. E. Lofton, and R. P. McClure. 2021b. Time series of high-frequency meteorological data at Falling Creek Reservoir, Virginia, USA 2015-2020 ver 5. Environmental Data Initiative. DOI: [10.6073/pasta/890e4c11f4348b3ceda802732ffa48b4](https://doi.org/10.6073/pasta/890e4c11f4348b3ceda802732ffa48b4)
- Carey, C. C., A. G. Hounshell, M. E. Lofton, F. Birgand, B. J. Bookout, R. S. Corrigan, A. B. Gerling, R. P. McClure, and W. M. Woelmer. 2021c. Discharge time series for the primary inflow tributary entering Falling Creek Reservoir, Vinton, Virginia, USA 2013-2021 ver 7. Environmental Data Initiative. DOI: [10.6073/pasta/8d22a432aac5560b0f45aa1b21ae4746](https://doi.org/10.6073/pasta/8d22a432aac5560b0f45aa1b21ae4746)
- Carey, C. C., W. M. Woelmer, A. S. Lewis, A. Breef-Pilz, D. W. Howard, and B. J. Bookout. 2021d. Time series of high-frequency sensor data measuring water temperature, dissolved oxygen, pressure, conductivity, specific conductance, total dissolved solids, chlorophyll a, phycocyanin, and fluorescent dissolved organic matter at discrete depths in

Falling Creek Reservoir, Virginia, USA in 2018-2020 ver 5. Environmental Data Initiative. DOI: [10.6073/pasta/88896f4a7208c9b7bddcf498258edf78](https://doi.org/10.6073/pasta/88896f4a7208c9b7bddcf498258edf78)  
Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., Luo, L., Alonge, C., Wei, H., Meng, J., Livneh, B., Lettenmaier, D., Koren, V., Duan, Q., Mo, K., Fan, Y., & Mocko, D. 2012. NCEP/EMC. 2009. NLDAS Secondary Forcing Data L4 Hourly 0.125 x 0.125 degree V002, Edited by David Mocko, NASA/GSFC/HSL, Greenbelt, Maryland, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC).



## Section S2. Linear parametric model selection methods

First, to remove correlated predictor variables, we examined all pairwise relationships among potential predictor variables using Pearson correlations. For variables which showed moderate correlation with one another ( $|\text{Pearson's } r| > 0.5$ ), the variable with the highest correlation with the response (chl-a) was chosen and the other removed from the analysis (Table S3, S4, S5 for daily, weekly, and fortnightly respectively). An AR lag of one time step was chosen (i.e., one day for daily models, one week for weekly models, two weeks for fortnightly models), based on the strength of autocorrelation between time steps within the weekly dataset using the `acf()` function in the R package ‘`astsa`’ (Figure S5; Stoffer, 2019). All analyses for this study were conducted in the R statistical environment (R version 4.0.3, R Core Development Team, 2020).

Second, using the training dataset described above, we fit all possible combinations of AR linear models using the function ‘`dredge`’ in the package ‘`MuMIn`’ (Barton, 2019), and selected the most parsimonious model using the corrected Akaike Information Criterion (AICc) (Table S6, S7, and S8).

### References

Barton, K. 2019. MuMIn: Multi-Model Inference. R package version 1.43.6.

<https://CRAN.R-project.org/package=MuMIn>

Stoffer, D. 2019. astsa: Applied Statistical Time Series Analysis. R package version 1.9.

<https://CRAN.R-project.org/package=astsa>

### Section S3. *Harmonizing multiple data sources*

Both the chl-a and discharge datasets have been collected using multiple sensors that cover different time periods and have different latencies, or the amount of time between data collection and data being available for integration into the forecasting workflow (Table S1t). To seamlessly compare measurements of the same variable collected by multiple sensors, we developed linear regressions. First, to convert from EXO chl-a measurements to weekly CTD chl-a measurements at the same depth, we aggregated the high-frequency EXO data to a daily mean that coincided with the time of the CTD chl-a measurement. A regression was developed between the two sensors, with the EXO data as the independent variable and CTD as the dependent variable. Second, a regression was also developed to harmonize discharge data collected by the INW Aquistar PT2X pressure sensor (collecting data from 2013-present; Table S1) and the Campbell Scientific CS451 pressure sensor (collecting data from summer 2019-present; Table S1). A regression was developed between the two sensors, with the Campbell discharge as the independent variable and the INW Aquistar discharge as the dependent variable. The slope, intercept,  $R^2$ , and standard deviation of the residuals of these two regressions is reported below.

Variable and Instrument Names	Slope	Intercept	$R^2$	Standard Deviation of residuals (used as observation error)
CTD to EXO (EXO on x-axis) chlorophyll-a ( $\mu\text{g/L}$ ) for FCR	0.605	0.255	0.68	0.21
INW to Campbell Scientific discharge ( $\text{m}^3/\text{s}$ )	-0.0047	-0.713	0.98	0.0039

#### Section S4. Thornthwaite-Mather Water Balance methods

We adapted the Thornthwaite-Mather Water Balance (TMWB) to model daily inflow in BVR from 2013-2016 using daily precipitation and air temperature data from the meteorological station located at Falling Creek Reservoir, as well as soil characteristics for the BVR watershed from the NRCS-USDA soil survey dataset (Soil Survey Staff 2021). After we calculated daily inflow, we then used forecasted NOAA GEFS precipitation and air temperature to drive the TMWB model and produce forecasts of daily inflow from 2 January 2019 - 15 August 2020.

To apply the TMWB model to calculate daily inflow in BVR, we first associated the soil characteristics and the map units for the BVR watershed to merge the spatial and soil characteristics data. Then, we calculated evapotranspiration and net precipitation to estimate the available soil water and total runoff for each timestep based on soil moisture conditions and precipitation from the previous day. We modified the TMWB model by changing the timestep from monthly to daily and adding a recharge equation to better estimate baseflow. Recharge (R) was calculated during precipitation events as

$$P - ExcessOut - Et$$

where P is the total precipitation, ExcessOut is runoff, and Et is actual evapotranspiration (Shuler and Mariner 2020). Inflow was then calculated as the sum of runoff and recharge.

#### References

- Shuler, C. K., & Mariner, K. E. (2020). Collaborative groundwater modeling: Open-source, cloud-based, applied science at a small-island water utility scale. *Environmental Modelling & Software*, 127, 104693.
- Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. Web Soil Survey. 2021. Available online at the following link: <http://websoilsurvey.sc.egov.usda.gov/>.

## **Section S5. Microscope identification of phytoplankton during 2019 bloom event**

Weekly phytoplankton samples taken at various depths in the surface waters of Falling Creek Reservoir (0.3 – 5 m) from 15 July to 05 August 2019 indicated that the phytoplankton community in the epilimnion was dominated by cyanobacterial and green algae taxa. *Dolichospermum* and *Pseudanabaena* were the dominant cyanobacterial taxa; the dominant green alga was a non-flagellated, unicellular nanoplankter with a maximum linear dimension averaging 5 µm. *Dolichospermum* was the dominant genus from 15 July to 29 July, comprising between 41% and 73% of total phytoplankton biovolume during this time period.

Samples were collected using a 4-L van Dorn depth sampler (Wildco, Yulee, FL, USA) and immediately preserved in 250 mL amber high-density polyethylene bottles by adding ~1% Lugol's iodine by volume to each sample. Before counting, three sub-samples of 20 mL from each sample were filtered and fixed onto permanent slides using 2-hydroxypropyl methacrylate (HPMA) following Crumpton 1987. Samples were subsequently enumerated on a Nikon Eclipse Ci microscope (Nikon, Minato City, Tokyo, Japan) at 400 until at least 300 natural units (either single cells or colonies) had been counted, with counting effort evenly divided amongst the three sub-samples. Phytoplankton were identified to genus when possible and the first ten natural units of each genus were measured and used to calculate biovolume via approximation to known geometric shapes following Hillebrand et al. 1999. All counts were conducted by M.E.L.

Unfortunately, we are unable to confirm the identity of the March/April 2020 bloom due to inability to sample during the COVID-19 pandemic which occurred during this time period.

## **References**

- Crumpton, W. G. 1987. A simple and reliable method for making permanent mounts of phytoplankton for light and fluorescence microscopy. *Limnology and Oceanography* 32:1154–1159.
- Hillebrand, H., C.-D. Dürselen, D. Kirschtel, U. Pollinger, and T. Zohary. 1999. Biovolume calculation for pelagic and benthic microalgae. *Journal of Phycology* 35:403–424.

Table S1. Sensor datasets used in daily, weekly, and fortnightly forecasts. Each of the steps in the first column map correspond to the steps in Figure 2. Data latency here refers to the time between data collection and data availability for use in forecasting models. NLDAS = North American Land Data Assimilation System; NOAA GEFS = National Oceanic and Atmospheric Administration, Global Ensemble Forecasts System.

Use in forecasting workflow	Variable	Units	Instrument used for measurement	Duration of data stream	Temporal resolution of measurements	Data latency
Training (Fig. 2, Step 2)	Discharge	m <sup>3</sup> /s	INW Aquistar PT2X pressure sensor	2013-present	10 minutes	Weekly to monthly
Forecasts (Fig. 2, Steps 2 and 4)	Discharge	m <sup>3</sup> /s	Campbell Scientific CS451 pressure sensor	2019-present	10 minutes	Daily
Training (Fig. 2, Step 2)	Surface chl-a	µg/L	SBE19plus CTD profiler	2013-present	1 week	Weekly
Forecasts (Fig. 2, Steps 2 and 4)	Surface chl-a	µg/L	EXO-Sonde	August 2018-present	15 minutes	Daily
Training (Fig. 2, Step 2)	Shortwave radiation	W/m <sup>2</sup>	NLDAS-2	2013-2015	1 hour	N/A
Training (Fig. 2, Step 2)	Shortwave radiation	W/m <sup>2</sup>	Hukseflux NR01 4-Component Net Radiometer	2015-present	5 minutes	Daily
Forecasts (Fig. 2, Step 4)	NOAA GEFS Shortwave Forecasts	W/m <sup>2</sup>	N/A	N/A	3 hours	Daily

Table S2. Candidate variables included in model fitting. Transformations were applied to improve fit with assumptions of a linear model.		
Variable	Forecasting Model	Transformation
Chl-a (ug/L)	Daily, weekly	Square root
Mean flow (m <sup>3</sup> /s)	Daily, weekly, fortnightly	--
Mean air temperature (°C)	Weekly	log
Mean wind speed (m/s)	Weekly	log
Median wind speed (m/s)	Fortnightly	log
Median relative humidity (%)	Weekly	--
Mean shortwave (W/m <sup>2</sup> )	Weekly, fortnightly	--
Mean relative humidity (%)	Daily	--
Median shortwave (W/m <sup>2</sup> )	Daily	--
Max wind speed (m/s)	Daily	log
Rain (mm)	Fortnightly	log



**Table S3.** Table of Spearman correlations for daily chlorophyll-a and drivers (variable units given in the first column). Variables that were removed from model selection due to correlations with other driver variables greater than 0.5 or less than -0.5 are shaded.

	Chl-a	Mean flow	Max flow	Min flow	Med flow	Mean RH	Med RH	Max SW	Mean SW	Med SW	Max WS	Mean WS	Med WS	Rain	Max temp	Mean temp	Med temp	Max RH
Chl-a (µg/L)	1																	
Mean flow (m³/s)	-0.11	1																
Max flow (m³/s)	-0.09	0.97	1															
Min flow (m³/s)	-0.11	0.97	0.92	1														
Median flow (m³/s)	-0.12	0.99	0.95	0.98	1													
Mean relative humidity (%)	0.29	-0.3	-0.22	-0.34	-0.3	1												
Median relative humidity (%)	0.28	-0.31	-0.23	-0.35	-0.3	0.96	1											
Max shortwave (W/m²)	-0.13	-0.17	-0.19	-0.13	-0.1	-0.39	-0.33	1										
Mean shortwave (W/m²)	-0.15	-0.18	-0.21	-0.15	-0.2	-0.39	-0.33	0.97	1									
Median shortwave (W/m²)	-0.03	-0.28	-0.26	-0.28	-0.3	0.12	0.14	0.67	0.7	1								
Max windspeed (m/s)	-0.11	0.3	0.3	0.27	0.29	-0.44	-0.43	0.12	0.09	-0.14	1							
Mean windspeed (m/s)	-0.09	0.29	0.29	0.26	0.29	-0.51	-0.5	0.11	0.1	-0.19	0.89	1						
Median windspeed (m/s)	-0.05	0.27	0.26	0.24	0.26	-0.5	-0.48	0.09	0.08	-0.21	0.78	0.96	1					
Rain (mm)	0.16	0.1	0.22	0.01	0.05	0.57	0.54	-0.4	-0.39	-0.01	-0.02	-0.09	-0.1	1				
Max temp (°C)	0.05	-0.58	-0.56	-0.58	-0.6	0.22	0.26	0.58	0.6	0.75	-0.17	-0.18	-0.19	-0.1	1			
Mean temp (°C)	0.08	-0.61	-0.57	-0.61	-0.6	0.3	0.31	0.51	0.54	0.77	-0.18	-0.21	-0.23	-0	0.96	1		
Median temp (°C)	0.07	-0.59	-0.56	-0.6	-0.6	0.3	0.31	0.49	0.52	0.77	-0.18	-0.21	-0.23	-0	0.94	0.99	1	
Max relative humidity (%)	0.2	-0.3	-0.25	-0.32	-0.3	0.72	0.69	-0.1	-0.13	0.15	-0.3	-0.36	-0.36	0.36	0.37	0.37	0.36	1



**Table S4.** Table of Spearman correlations for weekly chlorophyll-a and drivers (variable units given in the first column). Variables that were removed from model selection due to correlations with other driver variables greater than 0.5 or less than -0.5 are shaded.

	Chl-a	Water temp	Mean flow	Max flow	Min flow	Med flow	Max RH	Max temp	Mean temp	Med temp	Rain	Max WS	Mean WS	Med WS	Mean RH	Med RH	Max SW	Mean SW
Chl-a (µg/L)	1																	
Water temp (°C)	-0.04	1.00																
Mean flow (m³/s)	-0.29	-0.23	1.00															
Max flow (m³/s)	-0.27	-0.15	0.91	1.00														
Min flow (m³/s)	-0.31	-0.24	0.99	0.87	1.00													
Median flow (m³/s)	-0.29	-0.24	1.00	0.89	0.99	1.00												
Max relative humidity (%)	0.29	-0.11	0.09	0.16	0.03	0.07	1.00											
Max temp (°C)	-0.22	0.73	-0.05	0.00	-0.03	-0.05	-0.25	1.00										
Mean temp (°C)	-0.18	0.79	-0.03	0.03	-0.02	-0.04	-0.16	0.96	1.00									
Med temp (°C)	-0.20	0.79	-0.04	0.02	-0.03	-0.04	-0.17	0.95	0.99	1.00								
Rain (mm)	0.11	0.21	0.18	0.37	0.11	0.14	0.53	0.09	0.24	0.23	1.00							
Max windpseed (m/s)	-0.13	-0.20	0.18	0.17	0.19	0.19	0.07	-0.08	-0.14	-0.16	-0.04	1.00						
Mean windspeed (m/s)	-0.09	-0.21	0.10	0.12	0.10	0.10	0.11	-0.11	-0.16	-0.18	-0.06	0.90	1.00					
Median windspeed (m/s)	-0.06	-0.20	0.08	0.10	0.08	0.08	0.06	-0.11	-0.17	-0.19	-0.06	0.87	0.97	1.00				
Mean relative humidity (%)	0.18	-0.03	0.06	0.17	0.01	0.04	0.86	-0.22	-0.09	-0.10	0.57	-0.10	-0.02	-0.05	1.00			
Med relative humidity (%)	0.15	0.00	0.07	0.21	0.02	0.05	0.79	-0.19	-0.07	-0.08	0.55	-0.10	-0.01	-0.04	0.97	1.00		
Max shortwave (W/m²)	-0.42	0.28	0.06	0.01	0.09	0.06	-0.55	0.55	0.41	0.42	-0.32	0.07	0.03	0.03	-0.53	-0.49	1.00	
Mean shortwave (W/m²)	-0.43	0.37	0.06	0.01	0.09	0.07	-0.55	0.62	0.50	0.52	-0.31	0.07	0.02	0.01	-0.52	-0.48	0.97	1.00

**Table S5.** Table of Spearman correlations for fortnightly chlorophyll-a and drivers (variable units given in the first column). Variables that were removed from model selection due to correlations with other driver variables greater than 0.5 or less than -0.5 are shaded.

	Chl-a	Temp	Mean flow	Max flow	Min flow	Med flow	Max RH	Max temp	Mean temp	Med temp	Rain	Max WS	Mean WS	Med WS	Mean RH	Med RH	Max SW	Mean SW
Chl-a (µg/L)	1																	
Water temp (°C)	-0.05	1.00																
Mean flow (m³/s)	-0.16	-0.25	1.00															
Max flow (m³/s)	-0.14	-0.11	0.85	1.00														
Min flow (m³/s)	-0.19	-0.26	0.99	0.79	1.00													
Median flow (m³/s)	-0.15	-0.27	1.00	0.82	0.99	1.00												
Max relative humidity (%)	0.34	0.01	0.05	0.16	-0.01	0.04	1.00											
Max temp (°C)	-0.23	0.74	-0.15	-0.03	-0.14	-0.16	-0.21	1.00										
Mean temp (°C)	-0.22	0.78	-0.08	0.05	-0.07	-0.09	-0.13	0.97	1.00									
Median temp (°C)	-0.23	0.78	-0.08	0.04	-0.07	-0.09	-0.14	0.96	1.00	1.00								
Rain (mm)	0.12	0.31	0.14	0.42	0.06	0.12	0.53	0.19	0.31	0.31	1.00							
Max windspeed (m/s)	-0.01	-0.24	0.11	0.05	0.10	0.13	0.13	-0.20	-0.21	-0.24	0.00	1.00						
Mean windspeed (m/s)	0.05	-0.23	-0.02	-0.02	-0.02	0.00	0.12	-0.17	-0.19	-0.23	-0.03	0.90	1.00					
Median windspeed (m/s)	0.11	-0.26	-0.03	-0.02	-0.03	-0.02	0.07	-0.19	-0.24	-0.27	-0.04	0.86	0.97	1.00				
Mean relative humidity (%)	0.22	0.03	0.06	0.20	0.00	0.06	0.85	-0.16	-0.06	-0.06	0.56	-0.07	-0.02	-0.06	1.00			

Median relative humidity (%)	0.18	0.05	0.08	0.24	0.02	0.07	0.79	-0.13	-0.04	-0.04	0.53	-0.11	-0.04	-0.07	0.98	1.00		
Max shortwave (W/m <sup>2</sup> )	-0.35	0.25	0.01	-0.10	0.05	0.00	-0.60	0.56	0.45	0.46	-0.32	-0.13	-0.16	-0.15	-0.57	-0.53	1.00	
Mean shortwave (W/m <sup>2</sup> )	-0.37	0.33	0.04	-0.07	0.08	0.04	-0.60	0.64	0.55	0.56	-0.31	-0.12	-0.14	-0.14	-0.55	-0.51	0.96	1.00

Table S6. All potential models predicting chlorophyll a (chl-a) from daily training dataset, ranked by corrected Akaike's Information Criterion (AICc). Chl-a lag refers to the one-day autoregressive chlorophyll lag; median shortwave refers to median shortwave radiation; and df indicates degrees of freedom. The value in the variable columns refers to that variable's parameter in the model, no value indicates that the variable was not included in the model. We used multiple metrics to evaluate model fit, including log likelihood (logLik) and AICc; the delta indicates the number of AICc units the model is away from the best-fitting model; the weight represents the relative likelihood of the model compared to the other models.

Intercept	Chl-a lag	Mean flow	Mean Relative humidity	Median shortwave	Max. wind speed	df	logLik	AICc	delta	weight
0.23	0.77		0.00			4.00	26.85	-45.31	0.00	0.13
0.47	0.77				-0.07	4.00	26.61	-44.82	0.48	0.10
0.37	0.79					3.00	25.42	-44.61	0.69	0.09
0.19	0.78		0.00	0.00		5.00	27.60	-44.60	0.71	0.09
0.28	0.75	-1.13	0.00	0.00		6.00	28.57	-44.29	1.02	0.08
0.48	0.78			0.00	-0.08	5.00	27.26	-43.91	1.39	0.07
0.33	0.77		0.00		-0.04	5.00	27.20	-43.81	1.50	0.06
0.28	0.75	-0.63	0.00			5.00	27.20	-43.79	1.51	0.06
0.43	0.76	-0.84				4.00	26.03	-43.65	1.65	0.06
0.31	0.77		0.00	0.00	-0.05	6.00	28.10	-43.36	1.95	0.05
0.46	0.76	-1.26		0.00		5.00	26.87	-43.15	2.16	0.04
0.37	0.79			0.00		4.00	25.71	-43.02	2.28	0.04
0.49	0.76	-0.42			-0.06	5.00	26.73	-42.87	2.44	0.04
0.52	0.77	-0.84		0.00	-0.06	6.00	27.71	-42.57	2.74	0.03
0.33	0.75	-0.96	0.00	0.00	-0.03	7.00	28.70	-42.27	3.04	0.03

0.34	0.76	-0.45	0.00		-0.03	6.00	27.35	-41.86	3.45	0.02
------	------	-------	------	--	-------	------	-------	--------	------	------

Table S7. All potential model combinations from weekly training dataset, ranked by corrected Akaike's Information Criterion (AICc). Chl-a lag refers to the one-week autoregressive chlorophyll lag; median shortwave refers to median shortwave radiation; and df indicates degrees of freedom. The value in the variable columns refers to that variable's parameter in the model, no value indicates that the variable was not included in the model. We used multiple metrics to evaluate model fit, including log likelihood (logLik) and AICc; the delta indicates the number of AICc units the model is away from the best-fitting model; the weight represents the relative likelihood of the model compared to the other models.

Intercept	Air Temp	Chla lag	Mean Flow	Median Relative Humidity	Mean Shortwave	Mean Windspeed	df	logLik	AICc	delta	weight
1.65		0.46	-3.05		0.00		5.00	-64.25	139.10	0.00	0.26
1.13	0.19	0.47	-2.86		0.00		6.00	-63.71	140.27	1.16	0.14
1.73		0.46	-2.93		0.00	-0.08	6.00	-64.02	140.87	1.77	0.11
1.64		0.46	-3.05	0.00	0.00		6.00	-64.25	141.35	2.25	0.08
1.43		0.50			0.00		4.00	-66.69	141.76	2.66	0.07
0.79	0.24	0.50			0.00		5.00	-65.84	142.28	3.17	0.05
1.24	0.17	0.47	-2.79		0.00	-0.06	7.00	-63.59	142.32	3.21	0.05
1.25	0.22	0.46	-2.77	0.00	0.00		7.00	-63.64	142.41	3.31	0.05

1.56		0.50			0.00	-0.12	5.00	-66.24	143.06	3.96	0.04
1.74		0.46	-2.93	0.00	0.00	-0.08	7.00	-64.02	143.17	4.06	0.03
1.54		0.49		0.00	0.00		5.00	-66.66	143.91	4.81	0.02
1.04	0.29	0.50		0.00	0.00		6.00	-65.58	144.00	4.90	0.02
0.96	0.21	0.51			0.00	-0.09	6.00	-65.60	144.05	4.95	0.02
1.36	0.20	0.47	-2.70	0.00	0.00	-0.06	8.00	-63.53	144.52	5.42	0.02
1.69		0.50		0.00	0.00	-0.12	6.00	-66.20	145.24	6.14	0.01
1.19	0.26	0.50		0.00	0.00	-0.08	7.00	-65.36	145.86	6.76	0.01
0.25		0.55	-3.19	0.01			5.00	-69.43	149.45	10.34	0.00
0.76	-0.16	0.53	-3.38	0.01			6.00	-69.00	150.84	11.73	0.00
0.32		0.55	-3.08	0.01		-0.07	6.00	-69.27	151.38	12.27	0.00
0.89		0.55	-2.84				4.00	-71.53	151.45	12.34	0.00
0.11		0.58		0.01			4.00	-71.81	152.02	12.92	0.00
0.93	-0.18	0.53	-3.26	0.01		-0.10	7.00	-68.72	152.57	13.47	0.00
1.39	-0.15	0.54	-3.01				5.00	-71.14	152.88	13.77	0.00
0.70		0.59					3.00	-73.37	152.98	13.88	0.00
0.97		0.56	-2.72			-0.08	5.00	-71.33	153.26	14.15	0.00
0.24		0.59		0.01		-0.11	5.00	-71.47	153.53	14.43	0.00
0.45	-0.10	0.57		0.01			5.00	-71.63	153.86	14.76	0.00
0.82		0.59				-0.11	4.00	-73.01	154.41	15.30	0.00
1.58	-0.18	0.54	-2.89			-0.10	6.00	-70.82	154.48	15.37	0.00

1.04	-0.10	0.58					4.00	-73.20	154.79	15.68	0.00
0.70	-0.14	0.57		0.01		-0.13	6.00	-71.16	155.17	16.06	0.00
1.29	-0.14	0.58				-0.13	5.00	-72.70	156.00	16.89	0.00

Table S8. All potential model combinations from fortnightly training dataset, ranked by corrected Akaike's Information Criterion (AICc). Chl-a lag refers to the two-week autoregressive chlorophyll lag; median shortwave refers to median shortwave radiation; and df indicates degrees of freedom. The value in the variable columns refers to that variable's parameter in the model, no value indicates that the variable was not included in the model. We used multiple metrics to evaluate model fit, including log likelihood (logLik) and AICc; the delta indicates the number of AICc units the model is away from the best-fitting model; the weight represents the relative likelihood of the model compared to the other models.

Intercept	Chla lag	Mean Flow	Rain Sum	Mean Shortwave	Median Windspeed	df	logLik	AICc	delta	weight
1.08	0.33	26.61		0.00		5.00	0.51	17.56	0.00	0.49
0.58	0.35	26.11				4.00	-2.59	18.18	0.62	0.36
0.81	0.38	29.02	-0.04	0.00		6.00	1.87	22.26	4.70	0.05
0.81	0.35	28.91		0.00	0.22	6.00	1.64	22.72	5.17	0.04
0.36	0.38	27.78	-0.03			5.00	-2.19	22.95	5.39	0.03
0.41	0.36	27.28			0.11	5.00	-2.40	23.37	5.82	0.03
1.36	0.21					3.00	-9.48	27.62	10.06	0.00
0.49	0.40	31.73	-0.04	0.00	0.24	7.00	3.60	29.19	11.64	0.00
0.16	0.39	29.08	-0.03		0.12	6.00	-1.96	29.92	12.37	0.00
1.82	0.19			0.00		4.00	-8.74	30.47	12.91	0.00
1.60	0.21				-0.20	4.00	-9.25	31.50	13.95	0.00
1.51	0.19		0.02			4.00	-9.35	31.70	14.15	0.00
1.95	0.19			0.00	-0.13	5.00	-8.63	35.82	18.27	0.00
1.92	0.18		0.02	0.00		5.00	-8.67	35.91	18.35	0.00
1.73	0.19		0.02		-0.19	5.00	-9.15	36.87	19.31	0.00
2.03	0.18		0.01	0.00	-0.13	6.00	-8.57	43.13	25.58	0.00



Table S9. Effective sample size for each parameter in the daily, weekly, and fortnightly forecast models. Effective sample size was calculated on the first day of the forecasting period (1 January 2019), when all training and spin-up data were included in the dataset. IC Forecast = initial conditions for forecast.

	IC Forecast	Intercept	Chlorophyll-a	Discharge	Shortwave	Error
Daily	6619	243	301	2909	1582	1204
Weekly	37128	653	869	4241	1444	6629
Fortnightly	36642	757	941	6939	1721	11530

25

26

27

28

Table S10. Root mean square error (RMSE) of daily and weekly forecasts over the full year of forecasts, non-bloom conditions, and bloom conditions.							
Model time step	Day in future	Forecast, ~600 days	Null, ~600 days	Forecast non-bloom	Null non-bloom	Forecast bloom	Null bloom
Daily	1	3.9	2.8	2.1	1.7	11.3	7.4
Daily	2	4.7	3.4	2.5	2.1	13.1	9.0
Daily	3	5.4	4.0	3.0	2.5	14.7	10.3
Daily	4	6.2	4.6	3.5	2.9	16.5	11.4
Daily	5	7.0	5.1	4.0	3.4	18.5	12.6
Daily	6	7.7	5.6	4.6	3.8	19.8	13.4
Daily	7	8.5	6.0	5.4	4.2	21.3	14.3
Weekly	7*	6.1	6.7	3.5	4.6	16.1	16.2
Daily	8	9.2	6.5	6.4	4.6	21.3	14.7
Daily	9	9.9	6.8	7.5	5.0	20.8	14.9
Daily	10	10.5	7.1	8.9	5.3	18.9	15.0
Daily	11	11.1	7.4	9.9	5.6	17.6	15.0
Daily	12	11.5	7.6	10.5	5.7	17.2	15.4
Daily	13	11.9	7.7	11.0	5.9	17.1	15.5
Daily	14	12.3	7.9	11.4	5.9	17.5	15.9
Weekly	14*	6.8	8.1	3.0	5.9	17.7	17.0
Fortnightly	14'	7.0	7.9	2.7	5.6	18.8	16.9

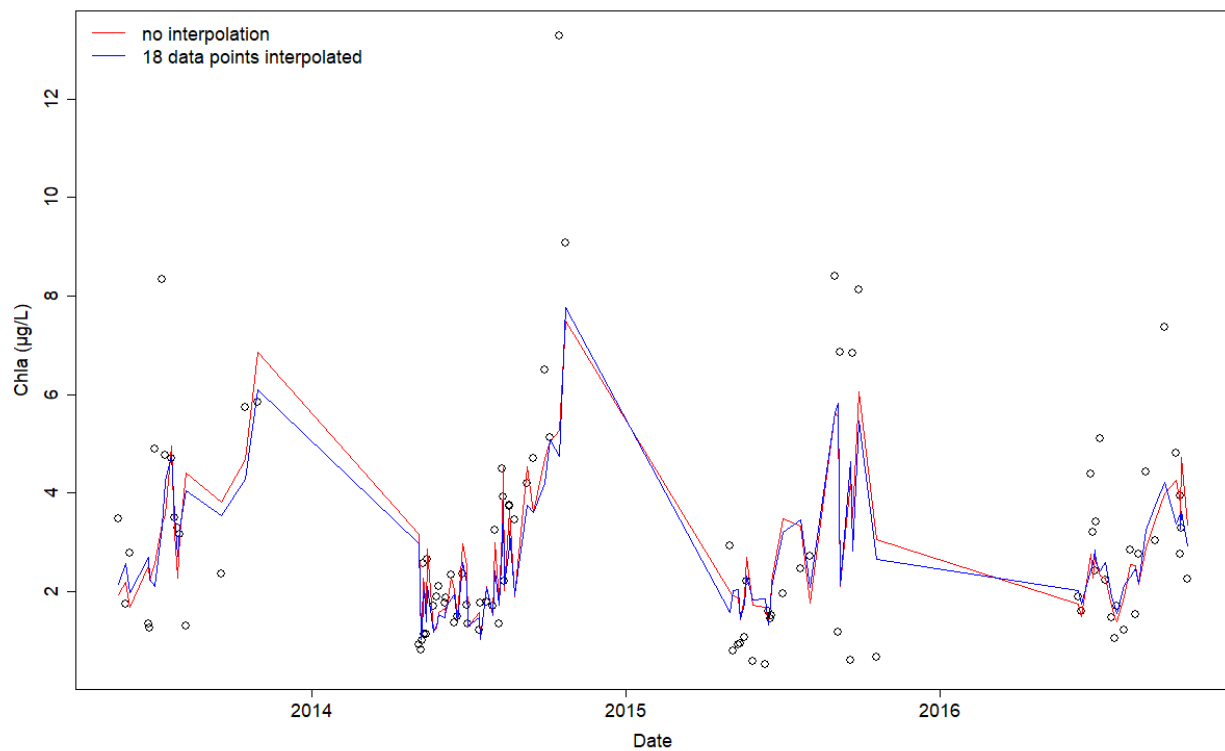
29

30

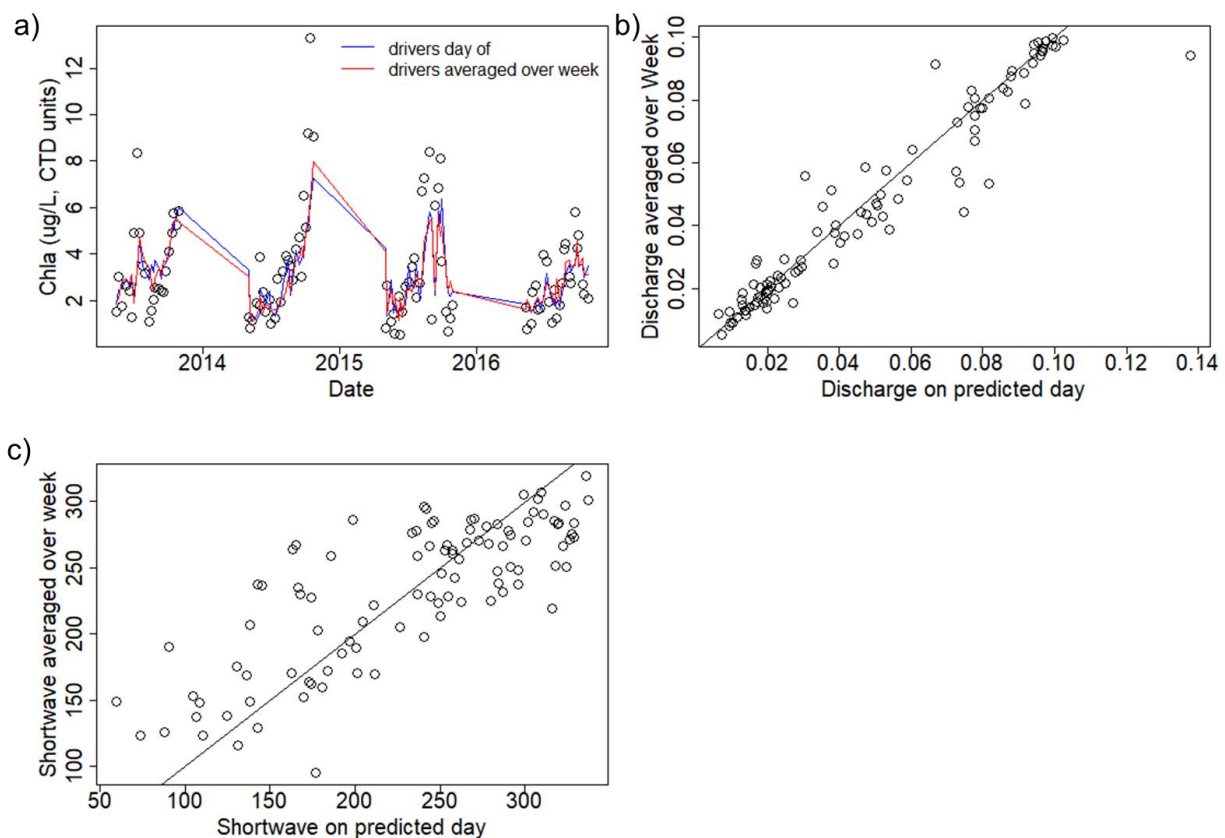
31

32

33

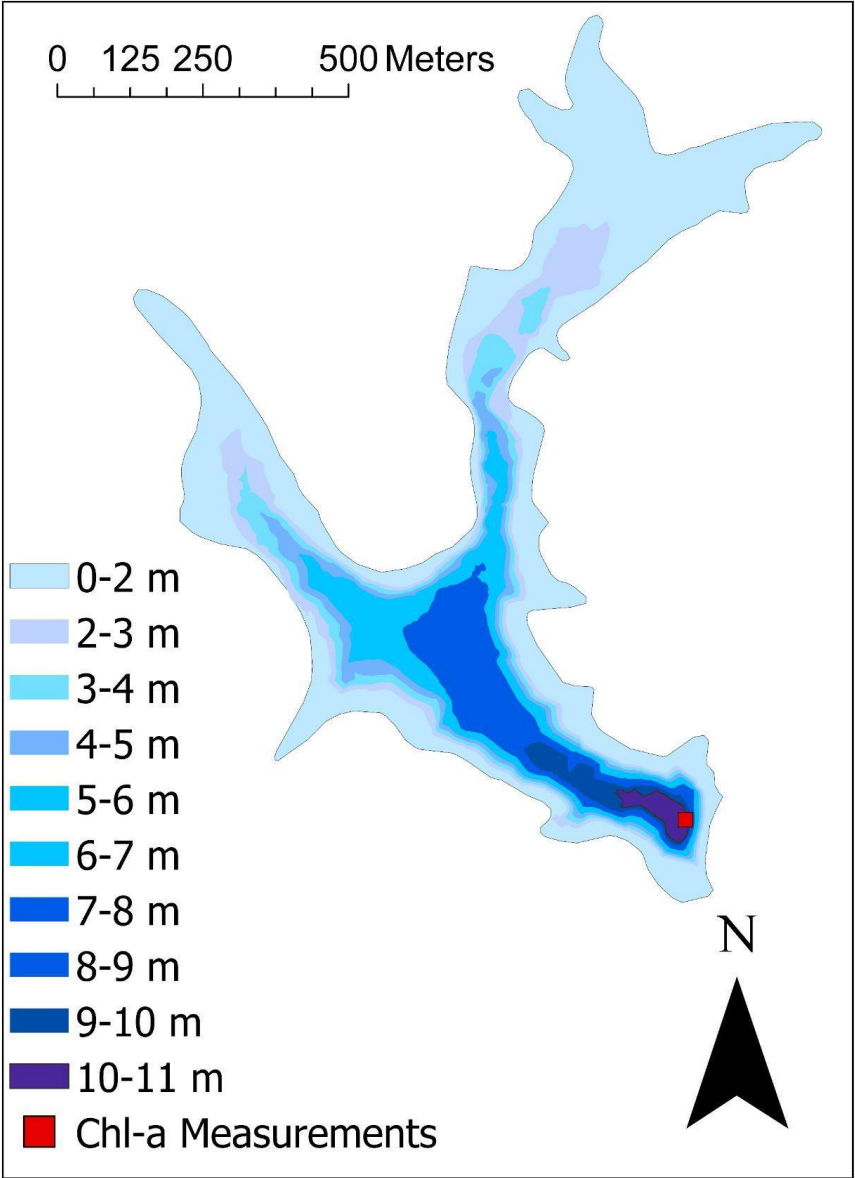


**Figure S1.** Observed chl-a (points) and predicted weekly chl-a using data from 2013-2016 (red line,  $R^2 = 0.38$ ), and data from 2013-2016 including 18 interpolated datapoints (blue line,  $R^2 = 0.44$ ).



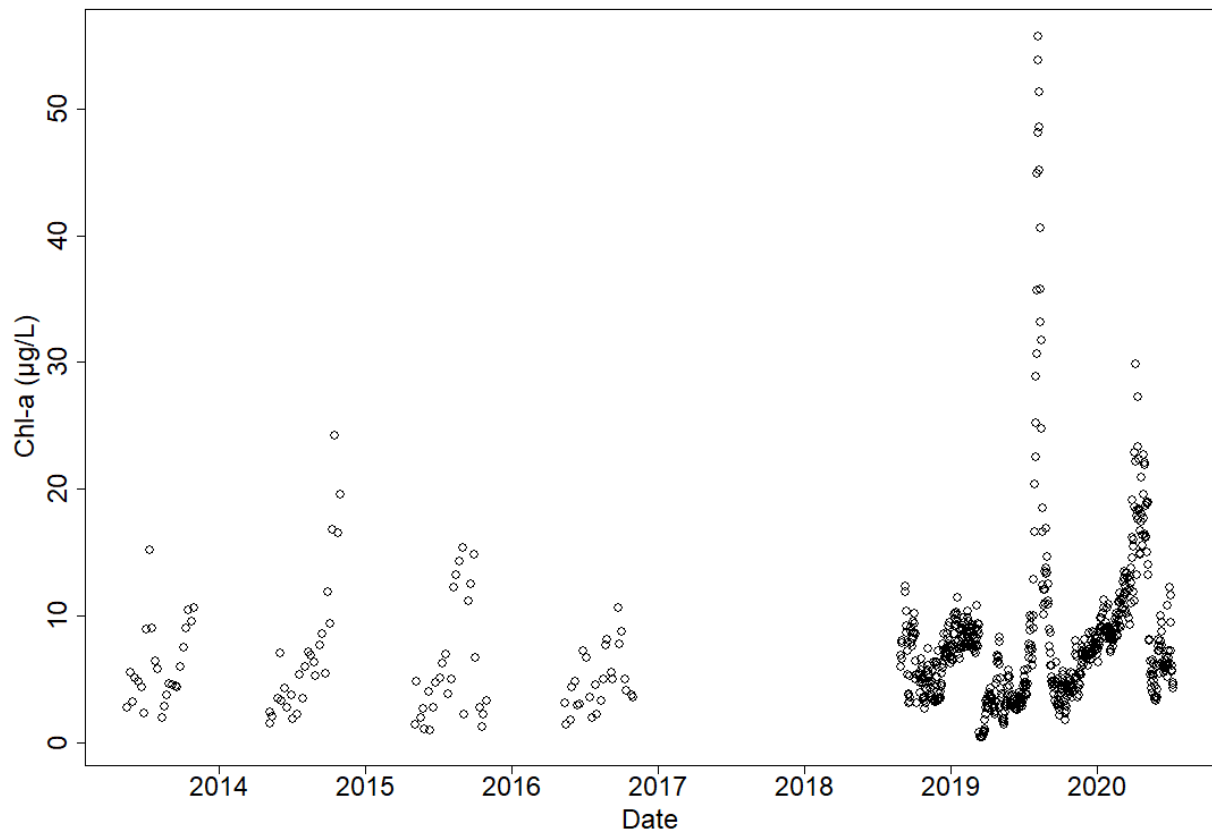
**Figure S2.** a) Relative performance of the same weekly linear model with mean shortwave and mean discharge covariates developed using driver variables averaged over a week period (red line) vs. driver variables derived from point estimates on the day being predicted (blue line). Observations are shown as the points for reference. b) 1:1 line of discharge driver data averaged over a week (y-axis) vs. point estimates of discharge on the predicted day (x-axis). c) 1:1 line of shortwave radiation driver data averaged over a week (y-axis) vs. point estimates of shortwave radiation on the predicted day (x-axis).

68  
69  
70  
71

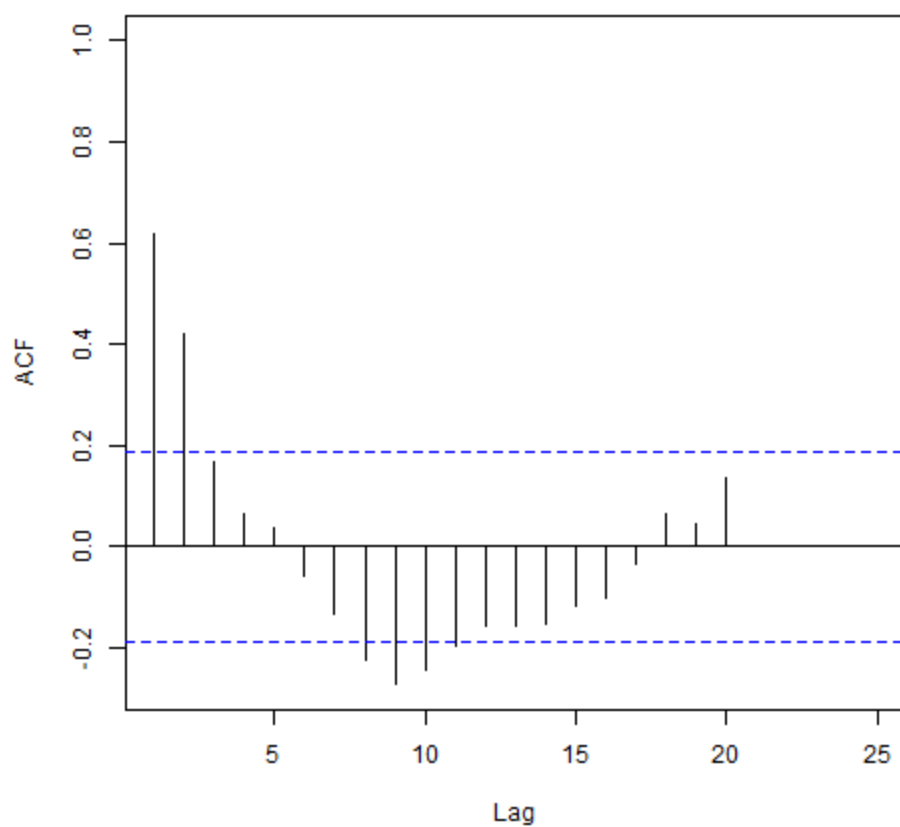


72  
73

**Figure S3.** Map of Beaverdam Reservoir, Vinton, VA, USA (37.31°N, 79.82°W) study site showing where chl-a measurements were taken at the dam. Forecasts were generated for the location where chl-a measurements were collected.

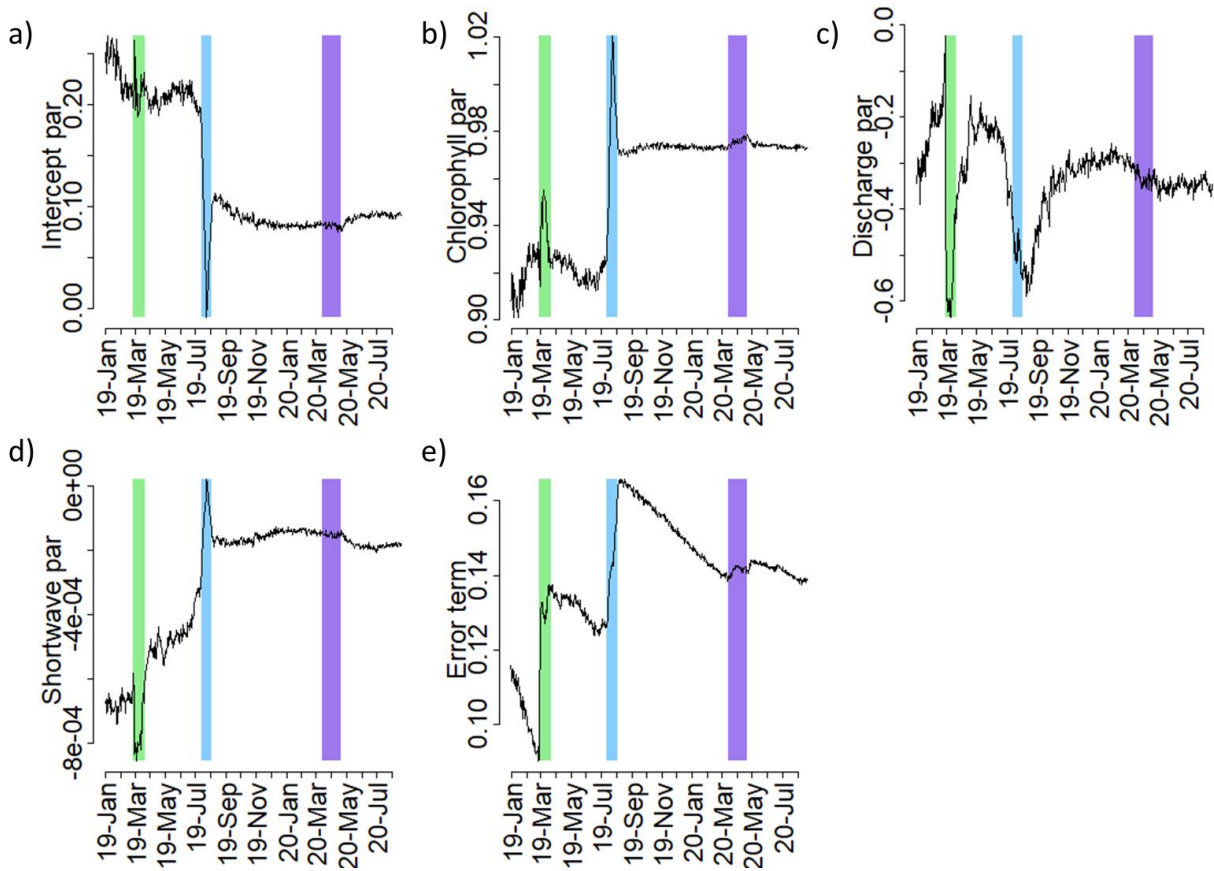


**Figure S4.** Observed time series of chl-a from the weekly training dataset (2013-2016) and spin-up and forecast time series for weekly and daily forecasts (Aug 2018-Dec 2019), using the harmonized CTD and EXO data, converted to EXO units (see Section S1).



**Figure S5.** Auto-correlation function (ACF) for chlorophyll-a in the CTD weekly training dataset, 2013-2016.

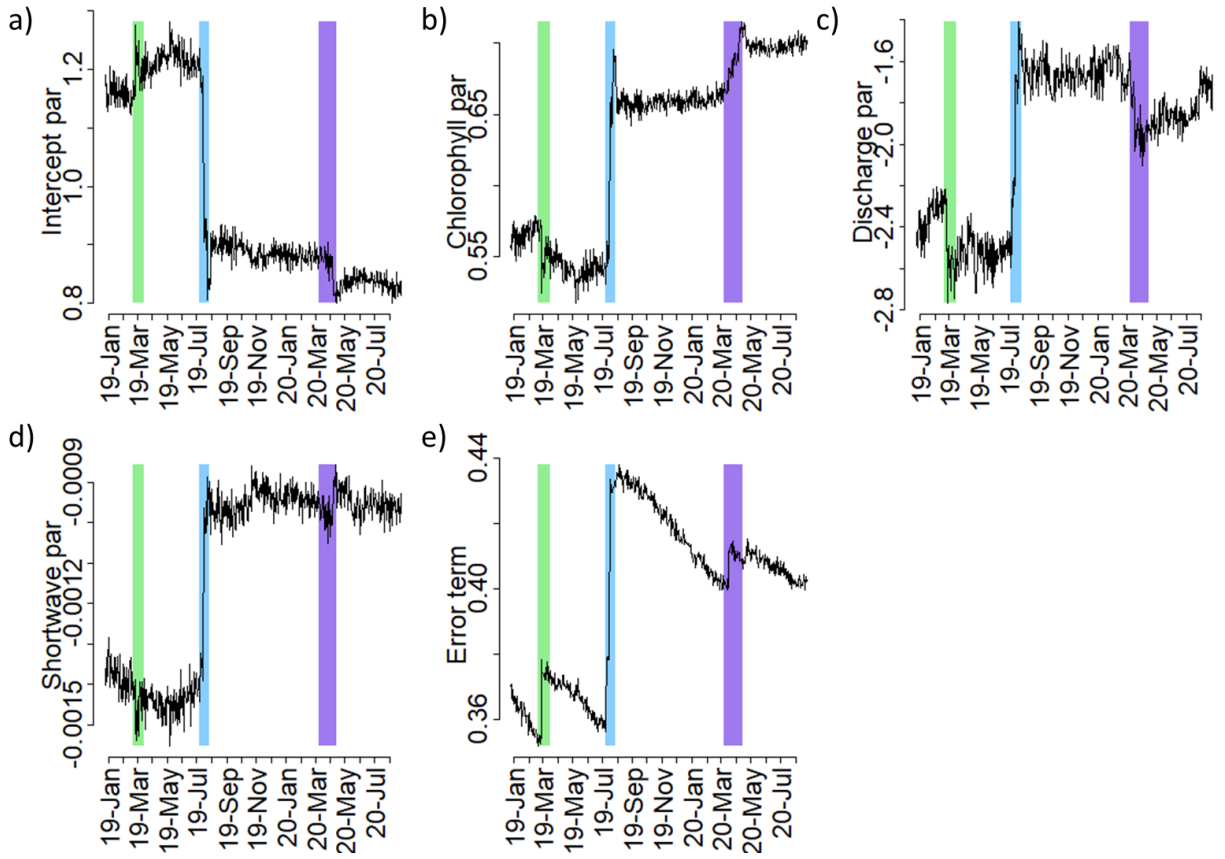
106  
107



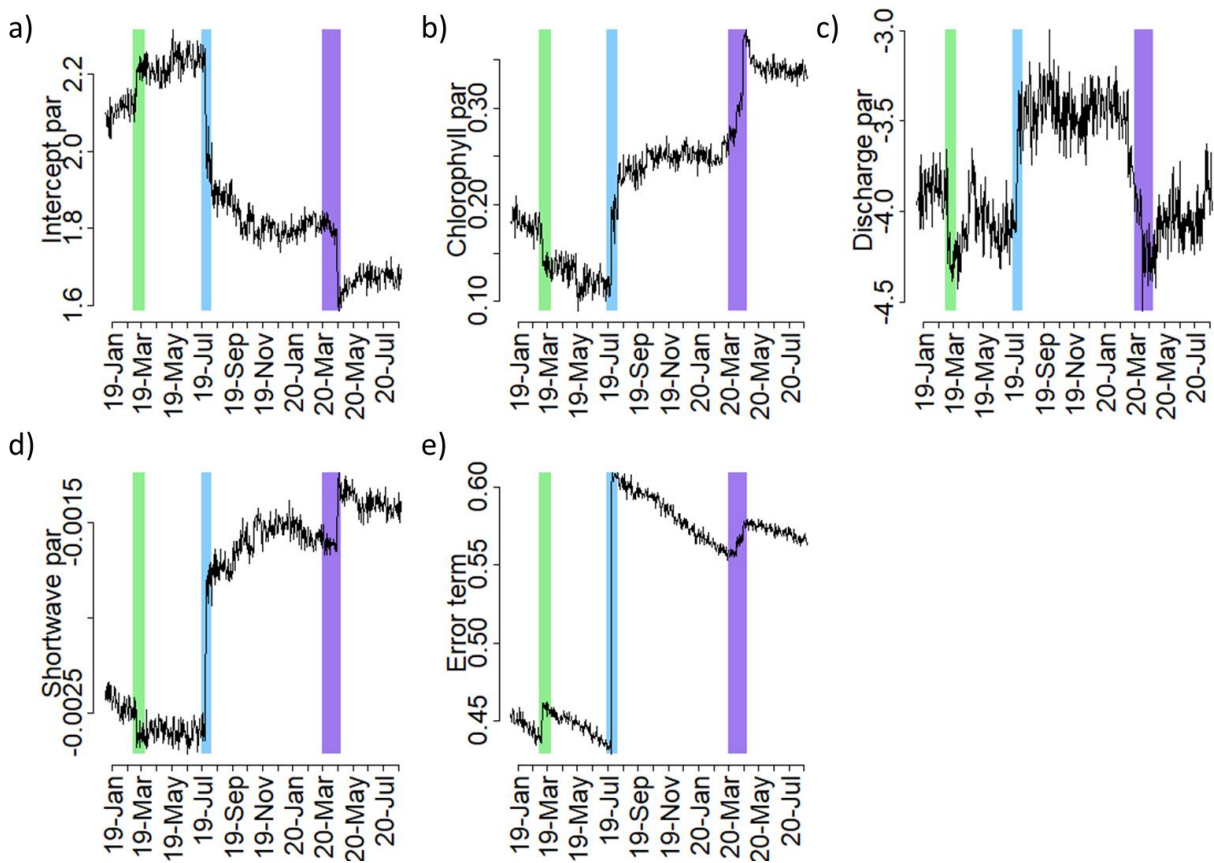
108

109 **Figure S6.** Parameter (par) evolution for daily forecasts throughout the forecasting period for a) the  
110 intercept parameter, b) chlorophyll lag term, c) discharge parameter, d) shortwave radiation parameter,  
111 and e) error term. See Equation 1 in the main text. The green shaded area denotes the time when the  
112 reservoir was dosed with copper sulfate, blue shaded area denotes the large bloom event in July 2019, and  
113 the purple shaded area denotes the March/April 2020 bloom.

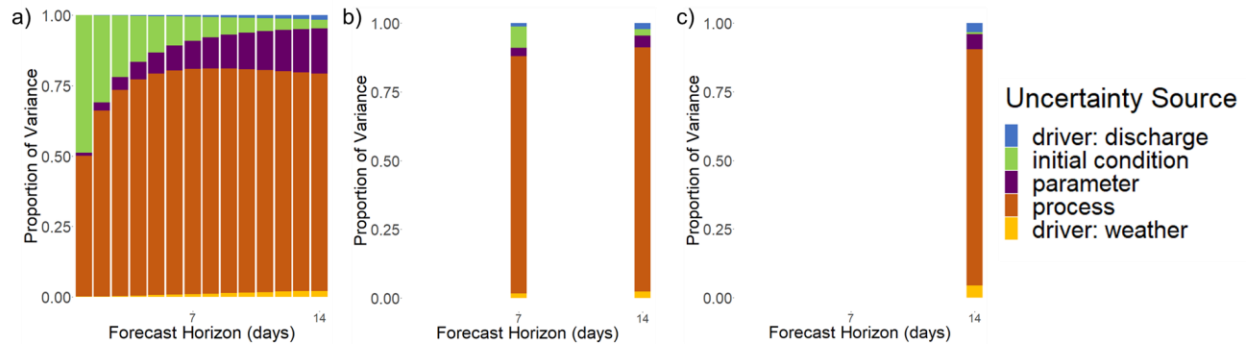




**Figure S7.** Parameter (par) evolution for weekly forecasts throughout the forecasting period for a) the intercept parameter, b) chlorophyll lag term, c) discharge parameter, d) shortwave radiation parameter, and e) error term. See Equation 1 in the main text. The green shaded area denotes the time when the reservoir was dosed with copper sulfate, blue shaded area denotes the large bloom event in July 2019, and the purple shaded area denotes the March/April 2020 bloom.



**Figure S8.** Parameter (par) evolution for fortnightly forecasts throughout the forecasting period for a) the intercept parameter, b) chlorophyll lag term, c) discharge parameter, d) shortwave radiation parameter, and e) error term. See Equation 1 in the main text. The green shaded area denotes the time when the reservoir was dosed with copper sulfate, blue shaded area denotes the large bloom event in July 2019, and the purple shaded area denotes the March/April 2020 bloom.



**Figure S9.** Proportion of variance contributed by different uncertainty sources to total forecast variance over the forecast horizon aggregated over the entire forecast period for a) daily, b) weekly, and c) fortnightly forecasts.

Meteorological data, parameter, initial condition, and discharge data uncertainty contributed a small proportion of the total uncertainty in all forecasts, with the relative contribution increasing slightly with forecast horizon (Figure S9a). Weekly forecasts showed little change in the contribution of uncertainty sources between 7-day and 14-day horizons, due to the dominance of process error as the major source of uncertainty. However, there was a small decrease in initial condition uncertainty and a small increase in parameter uncertainty with increasing forecast horizon (Figure S9b). Similar to daily forecasts, meteorological, discharge driver data, and parameter uncertainty were a small contribution to overall uncertainty for both weekly and fortnightly forecasts.