

**Building, applying, and communicating ecosystem understanding via freshwater
forecasts over time and space**

Whitney Marie Woelmer

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State
University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In

Biological Sciences

Cayelan C. Carey, Chair

Paul C. Hanson

Erin R. Hotchkiss

R. Quinn Thomas

August 18th, 2023

Blacksburg, Virginia

Keywords: decision-making, ecological forecasting, forecasting education, lake and
reservoir ecosystems, spatial and temporal variability, uncertainty, visualization science,
water quality

Copyright Whitney Marie Woelmer

Building, applying, and communicating ecosystem understanding via freshwater forecasts
over time and space

Whitney Marie Woelmer

ABSTRACT

Accelerating rates of change in ecosystems globally heighten the need for improved predictions of future ecological conditions. Freshwater lakes and reservoirs, which provide numerous ecosystem services, are particularly threatened by global change stressors and have already exhibited substantial changes to their physical, chemical, and biological functioning. Thus, to provide useful predictive tools for managing freshwater resources in the face of global change, we must improve our ability to build, apply, and communicate understanding of lake and reservoir ecosystem dynamics. To address this, I first *built* ecosystem understanding by conducting multiple whole-ecosystem surveys to quantify the spatial and temporal variability of biogeochemistry in two reservoirs over a year. We found that temporal heterogeneity was higher than spatial heterogeneity for most biogeochemical variables, with the stream-reservoir interface as a consistent hotspot of biogeochemical processing. Second, I *applied* ecosystem understanding by producing ecological forecasts of physical (water temperature), chemical (dissolved oxygen), and biological (chlorophyll-a) variables across three waterbodies using diverse modeling methods. I developed daily, weekly, and fortnightly forecasts of chlorophyll-a at two drinking water reservoirs using a Bayesian linear model, and found process uncertainty

dominated total forecast uncertainty. Additionally, I produced forecasts of water temperature and dissolved oxygen in an oligotrophic lake using a hydrodynamic-ecosystem model and found that water temperature was more predictable than oxygen despite variable performance over depth and between years. Across these two forecasting studies, forecast skill relative to a null model varied among water quality metrics: water temperature forecasts outperformed the null model up to 11 days ahead, oxygen forecasts outperformed the null model up to 2 days ahead, and chlorophyll-a forecasts outperformed the null model up to 14 days ahead. Third, to *communicate* forecasts for decision-making, I developed an educational module for undergraduate ecology students which taught important concepts on visualization and decision science. Following completion of the module, students' ability to identify methods for uncertainty communication increased significantly, as well as their understanding of the benefits of ecological forecasting. Overall, my dissertation provides insight into how reservoirs function in global biogeochemical cycles, the predictability of multiple water quality variables, and deepens our understanding of how to communicate ecosystem science for improved management and protection of ecosystems.

Building, applying, and communicating ecosystem understanding via freshwater forecasts
over time and space

Whitney Marie Woelmer

GENERAL AUDIENCE ABSTRACT

Human activities such as changing land use and climate have altered ecosystems around the world. As a result, many ecosystems no longer exhibit the same patterns from year to year as they have in the past, motivating the need for new tools to predict their future conditions. Among all ecosystems, freshwater lakes and reservoirs have been especially vulnerable due to human activities, threatening the critical services they provide to society, including drinking water and food, opportunities for recreational and cultural activities, and flood protection. Thus, because freshwater ecosystems are rapidly changing, forecasts of important water quality variables are needed. However, in order to use these forecasts effectively, a better understanding of how lakes and reservoirs change over space and time is needed, in addition to thoughtful communication of scientific findings for users. In my dissertation, I first addressed these needs by monitoring two reservoirs to examine how water quality varies over time and space. I found that water quality was more different over time (i.e., seasons) than over space (i.e., from one location to another within the reservoir). Second, I made ecological forecasts of water quality variables over space and time at different lakes. I forecasted phytoplankton levels in two drinking water reservoirs using a relatively simple model and water temperature

and dissolved oxygen in a large clear-water lake using a more complex model.

Comparing across forecasts in the three waterbodies, I found that my temperature forecasts provided valuable information above a baseline model up to 11 days ahead, oxygen forecasts up to 2 days ahead, and phytoplankton forecasts up to 14 days ahead. Third, because forecasts must be communicated effectively to be used as decision-making tools, I designed an educational module to teach forecast visualization and decision-making concepts. I found that students who completed the module better understood the benefits of forecasting for decision-making, and identified more ways to communicate forecasts after they completed the module. Overall, my dissertation shows the importance of measuring water quality over time and space to develop lake and reservoir forecasts, and demonstrates how forecast communication can help build students' understanding of the importance of forecasts for protecting and managing ecosystems.

[Type here]

Dedication

I dedicate this work to the Chanda family, especially, Malama, Bwalya, Nkatya, Chota, Annie, Gillie, and Bana Kabuswe, for showing me how to balance the important things in life wherever you are and when to laugh.

Acknowledgements

First, to Cayelan, who has guided me along so many turns in my PhD journey. You have taught me to advocate for my work and myself; to keep myself organized and accountable; how to lead a project to fruition; and how to motivate and maintain a team of people, along with so many other invaluable skills. I remember in one of my first lab meetings I said I hoped to be a contributing member soon and you told me I already was—you never fail to encourage and help your students find their potential. Thank you to my committee for your multifaceted advice. To Quinn for always making me think about my assumptions and encouraging me to adopt the latest coding or forecasting technology; to Erin for pushing me to identify the question and to think about the bigger picture; and to Paul for making me think conceptually and making me think about how the anything could be represented in an equation. I am a better scientist because of each of your individual contributions, and I hope to continue learning from you for years to come!

Thank you to several funding sources, including the National Science Foundation Graduate Research Fellowship, National Science Foundation grants DEB-1753639, CNS-1737424, DBI-1933016, DEB-1926050, Mary and George Schaeffer, Virginia Tech Biological Sciences Department, and to the Calhoun LPSA-VT Fellowship for supporting my research in so many ways.

I owe deep gratitude to several groups that were pivotal to my development as a scientist. You have all helped me to better understand my research in the broader context, and provided invaluable support and companionship during my time at Virginia Tech. I

especially want to thank the Carey Lab and Virginia Tech Reservoir Group, who are a powerhouse of a freshwater team that I am proud to have been part of whose willingness to provide feedback and always in the nicest way possible has made my work exponentially better. In particular Nicole Ward, Ryan McClure, Mary Lofton, Heather Wander, Abby Lewis, Dexter Howard, and Adrienne Breef-Pilz: all of whom have been invaluable colleagues, collaborators, and importantly, friends who have shown me the value of community and how to be a better team member. To Stream Team for being an academic family with a shared love for environmentalism and whose presence at Virginia Tech made me feel like I belonged, and especially to Elizabeth Shadle, Kristen Bretz, Stephen Plont, and Carla López Lloreda. To the Ecological Forecasting Initiative and the Ecological Forecasting Initiative Student Association for fostering many conversations along my journey to learning about forecasting and broadening my understanding of predictive ecology. To the Global Lake Ecological Observatory Network for being such a kind, welcoming group of people with an inspiring passion for lakes.

To many folks who made my time in Sunapee productive and meaningful, including Kak Weathers for showing me the meaning of building a scientific community dedicated to a beautiful place, Geoff Lizotte, June Fichter, Teriko MacConnell, Sue Godin, Kathleen Stowell, Nancy Brook Heckel, Susie Burbidge, and Elizabeth Harper for helping me find my way in Sunapee, and especially Jacob Wynne for being a great buddy in the field and in life—you always surprise me and I’ve learned so much from you since we first started working together.

Lastly to my family and friends who provide context for my life outside of my PhD: to Maggie, Madison, and Sam for keeping me grounded; to Tadhg for your selflessness, for

keeping me going and always stopping to help; and to my family for your unlimited support and earnest interest in what I have done for the past five years.

Table of Contents

Dedication	vii
Acknowledgements	viii
Table of Contents	xi
List of Figures	xiii
List of Tables	xxx
Chapter I: Introduction	1
References.....	14
Chapter II: The importance of time and space in biogeochemical heterogeneity and processing along the reservoir ecosystem continuum	33
Abstract	34
Introduction.....	34
Methods	40
Results	48
Discussion.....	55
Acknowledgements.....	67
References.....	69
Supporting Information	91
Chapter III: Near-term phytoplankton forecasts reveal the effects of model time step and forecast horizon on predictability	111
Abstract	112
Introduction.....	114
Methods	118
Results	130
Discussion.....	136
Acknowledgements.....	146
References.....	148
Supporting Information	171
Chapter IV: Performance of near-term, iterative water temperature and dissolved oxygen forecasts varies across time and depth	210

Abstract	211
Introduction.....	212
Methods	217
Results	226
Discussion.....	231
Acknowledgements.....	243
References.....	244
Supporting Information	265
Chapter V: Embedding communication concepts in forecasting training increases students' understanding of ecological uncertainty.....	286
Abstract	287
Introduction.....	289
Methods	293
Results	303
Discussion.....	308
Acknowledgements.....	318
References.....	320
Supporting Information	348
Chapter VI: Synthesis.....	370
References.....	386

List of Figures

Chapter I: Introduction

Figure 1 Iterative conceptual framework that integrates my four dissertation chapters (Ch.) to build, apply, and communicate ecological understanding via forecasting..... 32

Chapter II: The importance of time and space in biogeochemical heterogeneity and processing along the reservoir ecosystem continuum

Figure 1. Conceptual diagram showing a) Thornton et al.'s (1990) model of the spatial distribution of chlorophyll-a and nutrients (N and P) along the reservoir continuum, as well as selected field studies which have examined the distribution of b) nitrogen (N), c) chlorophyll-a (chl-a), and d) phosphorus (P) in reservoirs across the globe. While Thornton et al. (1990) did not distinguish between N and P, or total and dissolved nutrient fractions, we delineate these pools here when possible (total nutrients in solid lines, dissolved nutrients in dashed lines in panels b and d). We note that patterns of reservoir longitudinal distributions are meant to be qualitative and only represent relative minimums and maximums. Other than Marcé et al. (2021), we were unable to find analyses of the distributions of allochthonous and autochthonous carbon across the riverine, transitional, and lacustrine reservoir zones in the literature to match Thornton et al. (1990)'s model 84

Figure 2. Map of Falling Creek Reservoir (FCR) and Beaverdam Reservoir (BVR) with a) watershed area, and locations of stream and reservoir sampling sites and b) names of sampling sites on both reservoirs' bathymetry. Arrows in panel B indicate the approximate flow of water through the stream-reservoir continuum 85

Figure 3. Concentrations of biogeochemical variables along the reservoir continuum starting from the most upstream reservoir site of Beaverdam Reservoir (left panels) and Falling Creek Reservoir (right panels). Each panel represents a different biogeochemical variable, and colors represent the month of sampling, from 4 (April) to 10 (October). All units are denoted in each panel header. TN = total nitrogen, TP = total phosphorus, chl-a = chlorophyll-a, NH₄ = ammonium, NO₃ = nitrate, SRP = soluble reactive phosphorus, DOC = dissolved organic carbon, T-autoch = Peak T autochthonous carbon quality, A-alloch = Peak A allochthonous carbon quality. Because BVR has two upstream sites (one within each arm), these sites are plotted according to their distance from the corresponding stream site, meaning the first point on the x-axis is B1a and the second point on the x-axis is B1b. 86

Figure 4. a) Discharge at all stream (inflow and outflow) sites on all sampling dates, with Falling Creek Reservoir (FCR) sites represented in blue and Beaverdam Reservoir (BVR) sites in green. Site S1 refers to sites FS1 or BS1, while Site S2 refers to sites FS2 or BS2, with reservoir designated by color and panel. Sites designated as ‘Out’ represent the outflow dynamics for each reservoir and were estimated at FS3 at FCR and B4 in BVR. b) Specific conductance at all sites and for all sampling dates, with colors representing the month of sampling, from 4 (April) to 10 (October). Stream sites are represented by triangles or squares, while sites within the reservoir are represented by circles. Specific conductance was not collected at BVR sites in September 87

Figure 5. Heterogeneity, as measured by coefficient of variation, over space (CV_{space}) and time (CV_{time}) for BVR and FCR for total nutrients (TN and TP), chl-a, dissolved nutrients (NH₄, NO₃, SRP), dissolved organic carbon quantity (DOC), and dissolved

organic carbon quality (Peak T, autochthonous, and Peak A, allochthonous). Bars above pairing represent the pairing of groups for significance testing. ‘ns’ denotes a non-significant relationship ($p > 0.05$), a single asterisk ‘*’ represents $p < 0.05$, three asterisks ‘***’ represents $p < 0.0001$, and four asterisks ‘****’ represents $p < 0.00001$ 88

Figure 6. The change in biogeochemical variables, Δvar , along the reservoir continuum corrected by specific conductance and standardized by the area of the reservoir between sampling sites with each variable represented in individual panels. Δvar_{Site} , which represents the change between sites, is shown at each sampling site along the reservoir continuum and Δvar_{WR} , which represents the change from inflow stream sites and the furthest downstream site, is shown in diamond-shaped points at the furthest downstream site on the x-axis to the right of the vertical line. Note that the secondary (right) y-axis corresponds to Δvar_{WR} (large diamond points at far right of each panel) and presents the same units as Δvar_{Site} but on a smaller scale to show the range of variability in these values (which are of a smaller magnitude than Δvar_{Site} due to being standardized by the area of the entire reservoir). See *Methods: Spatially explicit processing of biogeochemical variables* for a full description of Δvar_{Site} and Δvar_{WR} (Eq. 1). Because BVR has two upstream sites (one within each arm), these are plotted according to their distance from the corresponding stream site, meaning the first point on the x-axis is B1a and the second point on the x-axis is B1b. Color represents the month in which the sample was collected, from 4 (April) to 10 (October). ‘BGV UNIT’ stands for the individual units of each biogeochemical variable, which are listed in each panel label. Samples were not collected at stream sites in September and are excluded from this analysis..... 89

Figure S1. Residence time (days) for Beaverdam Reservoir (BVR) and Falling Creek Reservoir (FCR) calculated for each sampling day (shown as points). Residence times were calculated by dividing the reservoir volume by the incoming flow to the reservoir for a given	102
Figure S2. Daily average air temperature and cumulative daily rainfall over the entire study period, measured at the dam of Falling Creek Reservoir. Sampling dates are indicated by the vertical blue lines.....	103
Figure S3. Surface water temperature and surface dissolved oxygen (DO) percent saturation (%) at all study sites on all sampling dates. In September, we did not visit all sites at Beaverdam Reservoir so data from that reservoir’s inflow sites (BS1 and BS2), as well as B4, are absent.....	104
Figure S4. Time series of thermocline depth (left) and Schmidt stability (right) in Beaverdam Reservoir (red) and Falling Creek Reservoir (blue). Note that thermocline depth is plotted with the y-axis reversed, to represent the depth from the surface of the reservoir. Calculations were performed using the rLakeAnalyzer package (Winslow et al. 2019).	105
Figure S5. Nutrient stoichiometry ratios along the reservoir continuum in Beaverdam Reservoir and Falling Creek Reservoir, starting from the most upstream reservoir site in each reservoir. Colors represent the month of sampling. TN = total nitrogen; TP = total phosphorus; DIN = dissolved inorganic nitrogen consisting of the sum of NH ₄ and NO ₃ ; SRP = soluble reactive phosphorus.	106
Figure S6. Heterogeneity, as measured by coefficient of variation, over space (CV _{space}) and time (CV _{time}) for nutrient stoichiometry ratios in BVR and FCR. ‘ns’ denotes no	

significant difference between groups ($p > 0.05$) and asterisks ‘*’ denotes significant differences between groups ($p < 0.05$). TN = total nitrogen; TP = total phosphorus; DIN = dissolved inorganic nitrogen consisting of the sum of NH_4 and NO_3 ; and SRP = soluble reactive phosphorus. 107

Figure S7. Δvar , or the change in biogeochemical variables calculated using Eq. S1, along the reservoir continuum with each variable represented in individual panels, but with specific conductance at the stream sites standardized to the specific conductance observed at the most upstream reservoir site to examine the effect of specific conductance on the calculations. Δvar_{Site} , which represents the change between sites, is plotted at the location of the downstream sampling site along the reservoir continuum and Δvar_{WR} , which represents the change from inflow stream sites and the furthest downstream site, is shown in diamond-shaped points to the right of the vertical line. *See Methods: Spatially explicit processing of biogeochemical variables* for a full description of Δvar_{Site} and Δvar_{WR} (Eq. 1). Color represents the month in which the sample was collected, from 4 (April) to 10 (October). ‘BGV UNIT’ stands for the individual units of each biogeochemical variable, which are listed in each panel label. Samples were not collected at stream sites in September and are excluded from this analysis. 108

Figure S8. Δvar , or the change in biogeochemical variables calculated using Eq. S1 (Text S4) not standardized by area, along the reservoir continuum (in comparison to the standardized metric shown in Figure 6). Δvar_{Site} which represents the change between sites, is shown at each sampling site along the reservoir continuum. Δvar_{WR} , which represents the change from inflow stream sites and the furthest downstream site, is shown in diamond-shaped points to the right of the vertical line. Because BVR has two upstream

sites (one within each arm), these are plotted according to their distance from the corresponding stream site, meaning the first point on the x-axis is B1a and the second point on the x-axis is B1b. Color represents the month in which the sample was collected, from 4 (April) to 10 (October). ‘BGV UNIT’ stands for the individual units of each biogeochemical variable, which are listed in each panel label. Samples were not collected at stream sites in September and are excluded from this analysis 109

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

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. 162

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..... 163

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. 165

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. 166

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) 167

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. 168

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). 169

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, $R^2 = 0.44$). 199

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). 200

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. 201

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) .. 202

Figure S5. Auto-correlation function (ACF) for chlorophyll-a in the CTD weekly training dataset, 2013-2016..... 203

Figure S6. Parameter (par) evolution for daily 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. 204

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. 205

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. 206

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 207

Chapter IV: Performance of near-term, iterative water temperature and dissolved oxygen forecasts varies across time and depth

Figure 1. Conceptual representation of the FLARE (Forecasting Lakes And Reservoir Ecosystems) framework which shows the daily steps of 1) downloading lake observational data and appending the new data to the long-term observational dataset, 2) downloading meteorological forecasts from NOAA needed as driver data for the lake model, 3) updating model initial conditions (IC) and tuning parameters via an ensemble Kalman filter (EnKF), 4) producing a forecast of 1 to 35-days-ahead with uncertainty, and 5) evaluating the forecast. CI = confidence interval. Awaiting permission from NOAA to use logo for publication. Use of the emblem or logo does not imply an endorsement by NOAA/NWS..... 259

Figure 2. Bathymetric map of Lake Sunapee, NH, USA (43.37°, -72.05°; (LSPA 2023), showing the location of the manually-collected data sampling site (circle) and the sensed buoy site (triangle). 260

Figure 3. High-frequency observations of temperature and oxygen from the buoy in Lake Sunapee over the study period (August - October) in 2021 and 2022 showing the time series of observations for a) temperature at 1.0 m, b) oxygen at 1.0 m, violin plots of observations for c) temperature at 1.0 m, d) oxygen at 1.0 m, and e) permutation entropy

of temperature at 1.0 m and 10.0 m in both years. Observations at 10.0 m are shown in the bottom row with the time series of e) temperature at 10.0 m, and f) oxygen at 10.0 m, and violin plots of g) temperature at 10.0 m, and h) oxygen at 10.0 m, as well as j) permutation entropy of oxygen at 1.0 m and 10.0 m in both years. 261

Figure 4. Forecast accuracy of a) water temperature ($^{\circ}\text{C}$), and b) dissolved oxygen (mg/L) and c) forecast skill (unitless; calculated using Eq. 1) of both temperature and oxygen over the full study period and across all depths at forecast horizons from 1-35 days into the future (note the difference in y-axis). In panels a and b, more accurate forecasts are represented by lower CRPS values. In Panel c, and forecasts which are less skillful than the null are shown below zero. 262

Figure 5. Forecast accuracy of a) temperature at 1.0 m, b) oxygen at 1.0 m, c) temperature at 10.0 m, and d) oxygen at 10.0 m as measured by the Continuous Ranked Probability Score (CRPS). The mean across all days of year within a given year is shown by the central solid line, with the shaded area representing the standard deviation across all days within a given year. Higher CRPS values indicate lower forecast accuracy. Note the difference in y-axis..... 263

Figure 6. Forecast skill across the two study years for a) temperature at 1.0 m, b) oxygen at 1.0 m, as well as the percent of forecasts which beat the null for c) temperature and 1.0 m, and d) oxygen and 1.0 m. Forecasts from 10.0 m are shown in the bottom row for e) temperature at 10.0 m, f) oxygen at 10.0 m, and percent of forecasts better than the null for g) temperature at 10.0 m, and h) oxygen at 10.0 m..... 264

Figure S1. Lake Sunapee high-frequency buoy observations (lines) and monthly profiles from depths > 10 m (points) in 2021 and 2022 of a) water temperature profiles from the surface to 30.0m and b) calculated thermocline depth using these profiles	273
Figure S2. Evolution of parameter values of the spin-up period in the two focal years of this study (2021 and 2022). Note that the spin-up time period started on different dates in each year but covered 35 days total. See Table S2 for a description of each parameter.	274
Figure S3. Climatology null calculated as the mean and standard deviation across all available data from 2007-2022. Only days of year with three or more observations on that date were included. Sensors were not all deployed over the same time frame since the buoy was first deployed in 2007, resulting in differences in data availability across days of year for each variable and depth.....	275
Figure S4. Absolute Error (AE) of temperature and oxygen forecasts in 2021 and 2022 at Lake Sunapee, NH at 1.0 m and 10.0 m depth across a 1-35 day forecast horizon.....	276
Figure S5. Weighted permutation entropy across years (2021 and 2022), depths (1.0 m and 10.0 m) and variables (water temperature and dissolved oxygen) in this study.	277
Figure S6. Comparison of dissolved oxygen in percent saturation and mg/L over the focal years in this study (2021 and 2022)	278
Figure S7. Forecast performance as Continuous Ranked Probability Score (CRPS) of forecasts at select (1-day, 7-day, and 21-day) horizons across the full forecast time period (August 04 to October 17). Dates are shown relative to the day of turnover in each year (2021-10-04 and 2022-09-23). CRPS is shown across years, depth, and variable for each horizon.	279

Figure S8. Forecast accuracy of climatology model across years, depths, and variables. The climatology null was calculated across all historical years with available observations (Fig. S1) and was evaluated against observations in 2021 and 2022, resulting in different skill scores for each year.	280
Figure S9. Meteorological variables over the two years in our study at Lake Sunapee, NH, USA. AirTemp = air temperature in °C; LongWave = longwave radiation in W m ⁻² ; Rain = precipitation in mm/day; RelHum = relative humidity in decimal percentage; ShortWave = shortwave radiation in W m ⁻² ; WindSpeed = wind speed in m s ⁻¹	281
Figure S10. Time series of meteorological variables which were used to drive FLARE forecasts at Lake Sunapee, NH in 2021 and 2022. The date of turnover is indicated by the vertical line. AirTemp = air temperature in °C; LongWave = longwave radiation in W m ⁻² ; Rain = precipitation in mm/day; RelHum = relative humidity in decimal percentage; ShortWave = shortwave radiation in W m ⁻² ; WindSpeed = wind speed in m s ⁻¹	282
Figure S11. Comparison of deepwater observations at 10.0 m, 15.0 m, 20.0 m, and 30.0 m from the deep site at Lake Sunapee, NH	283
Figure S12. Comparison of historical observations of dissolved oxygen and focal years in this study (2021 and 2022). Years with incomplete observations were removed (2007, 2008, 2013, 2014, 2017)	284
Figure S13. Comparison of historical observations of water temperature and focal years in this study (2021 and 2022). Years with incomplete observations were removed (2007, 2009, 2013, 2017)	285

Chapter V: Embedding communication concepts in forecasting training increases students' understanding of ecological uncertainty

Figure 1. Conceptual diagram of the two learning objectives (LOs) associated with uncertainty communication and foundational ecological forecasting concepts taught in Macrosystems EDDIE Module 8: “Using Ecological Forecasts to Guide Decision-Making.” Within the larger circles, each bubble shows the corresponding assessment questions which tested the effectiveness of the module in meeting these LOs.

Descriptions of the assessment questions are in Table 3 and the full assessment questions can be found in Appendix S1: Tables S1-2. Ecological forecast is abbreviated as EF. . 341

Figure 2. Screenshots of activities from the module R Shiny application showing A) Activity B, in which students (1) make decisions about how to manage a drinking water reservoir, using two different forecast visualizations (2a and 2b, which were shown separately to students within the application), while weighing the consequences on multiple objectives, such as maintaining good drinking water quality, preserving ecological health, maximizing economic benefit, and ensuring swimmer safety (3). In Activity C, students chose a forecast user and customized a visualization for that particular users' decision needs; for example, (B) shows a visualization and decision chosen for a swimmer, while (C) shows a visualization and decision chosen for a local policymaker..... 342

Figure 3. Percentage of students who answered multiple-choice questions correctly in the pre- and post-module surveys. Asterisks (***) indicate a statistically significant difference between the pre- and post-module survey according to a Wilcoxon signed-rank test ($p < 0.001$). Colors of the bars correspond to LO1: Foundational ecological

forecasting (EF) concepts (green) and LO2: Uncertainty Communication (orange; Figure 1, Table 3). A description of questions can be found in Table 3 and Appendix S1: Table S3. 344

Figure 4. Percentage of students who identified a) different benefits of ecological forecasting (Q3), and b) different ways to visualize uncertainty (Q4) in the pre- and post-module responses. These responses correspond to the c) total number of benefits identified by individual students in Q3, and d) the total number of ways to visualize uncertainty identified by individual students in Q4. Students were also given the option to state “I don’t know,” represented as “IDK” in panels a and b. The categories listed on the x-axis in panels a and b were determined through the methods outlined in Appendix S1: Text S1 and are listed in Table 4, respectively. Asterisks (***) indicate a significant difference ($p < 0.001$) between the pre- and post-survey responses according to paired Wilcoxon signed-rank tests. A full description of questions can be found in Appendix S1: Table S4. 345

Figure 5. Pre- and post-survey results across the two LOs showing a) the total percent correct for all students across each category, and b) the change in the number of correct answers for each student after module completion relative to each student’s percent correct before taking the module. Color in (b) corresponds to the number correct on the pre-survey only and points are jittered to improve legibility of individual points. Note that the number of questions corresponding to each LO varied, with three questions assessing foundational ecological forecasting concepts and five questions assessing uncertainty communication concepts. The percentage of correct answers are standardized to the total number of questions per LO in (a), while the number of questions answered correctly is

shown in (b). “Foundational” corresponds to LO1: Ecological Forecasting Concepts, and
“Communication” corresponds to LO2: Uncertainty Communication..... 347

Figure S1. The figure accompanying Questions 5 and 6 on the pre- and post-module
assessment survey. Forecast visualizations of: A) water temperature forecasts in Blue
Lake; and B) grassland biomass in the Great Plains region of the United States. Panel A
shows a forecast generated for surface (red) and bottom (blue) temperatures in Blue Lake
for 14 days into the future. Points represent measurements from the last 7 days, the black
line represents the mean forecast estimate, and the shaded polygon represents the
probabilistic forecasts of future temperatures. Panel B shows a forecast for the upcoming
year’s growing season. It is a forecast of the percent change in future grassland biomass
in the U.S. Great Plains region compared to the historical (1982-2019) average. 364

Figure S2. Screenshot from introductory PowerPoint lecture at beginning of module
explaining A) the concept of a forecast output and a forecast index, and B) different ways
to communicate uncertainty using numbers, words, icons, and figures as either forecast
output or forecast index. 365

Figure S3. Screenshots of activities from the module R Shiny application showing A)
Activity B, where students (1) make decisions about how to manage a drinking water
reservoir, using two different forecast visualizations (2a and 2b), while weighing the
consequences on multiple objectives (3). The transition from 2.1 to 2.2 shows how
forecasts update over time and following students’ management decisions. In Activity C,
students choose a forecast user and customize a visualization for that particular users’
decision needs showing B) a visualization and decision chosen for a swimmer, as
opposed to C) a visualization and decision chosen for a local policymaker. 366

Figure S4. Results from pre- and post-assessment Question 9 which asked students (n = 240) to identify the components of PrOACT, a structured decision-making technique ($p < 0.001$). Based on Wilcoxon signed-rank test, this question had a test statistic of 845.5, and effect size of 0.34, with 54% of student responses correct on the pre-module and 75% of student responses correct on the post-module. 368

Chapter VI: Synthesis

Figure 1. Iterative conceptual framework that integrates my four dissertation chapters (Ch.) to build, apply, and communicate ecological understanding via forecasting. 391

List of Tables

Chapter II: The importance of time and space in biogeochemical heterogeneity and processing along the reservoir ecosystem continuum

Table S1. Table of GPS coordinates, sub-watershed names, and corresponding area and percent contribution for each sub-watershed within Beaverdam Reservoir's overall watershed. The shaded rows denote the sub-watersheds included in this study..... 96

Table S2. GPS coordinates for all sites in this study. Refer to Figure 2 in the main text for a map of sites..... 97

Table S3. Analytical chemistry limits of quantitation for total and dissolved nitrogen, phosphorus, and carbon. 98

Table S4. Summary statistics (minimum, maximum, mean) of coefficients of variation (CV) for each reservoir and variable across space and time. (chl-a = chlorophyll-a, SRP = soluble reactive phosphorus, NO₃ = nitrate, NH₄ = ammonium, TP = total phosphorus, TN = total nitrogen, A = allochthonous carbon, T = autochthonous carbon, DOC = dissolved organic carbon) 99

Table S5. Areas of each reservoir section used to calculate the spatially explicit processing metric, $\Delta_{\text{var Site}}$ using Eq. 1. Area between sampling points was calculated using the Calculate Geometry tool in ArcGIS Pro (2021)..... 101

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

Table 1. Definitions of uncertainty sources that can contribute to total forecast uncertainty (derived from Dietze 2017a)..... 170

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.	179
Table S2. Candidate variables included in model fitting. Transformations were applied to improve fit with assumptions of a linear model.....	181
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....	182
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....	184
Table S5. Table of Spearman correlations for fortnightly chlorophyll-a and drivers (variable units given in the first column).....	186
Table S6. All potential models predicting chlorophyll a (chl-a) from daily training dataset, ranked by corrected Aikaike'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. 189

Table S7. All potential model combinations from weekly training dataset, ranked by corrected Aikake'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. 191

Table S8. All potential model combinations from fortnightly training dataset, ranked by corrected Aikake'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. 194

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. 196

Table S10. Root mean square error (RMSE) of daily and weekly forecasts over the full year of forecasts, non-bloom conditions, and bloom conditions. 197

Chapter IV: Performance of near-term, iterative water temperature and dissolved oxygen forecasts varies across time and depth

Table S1. Details of data availability, spin-up length, and forecast date ranges for all years forecasted in this study 269

Table S2. List of parameter values included in the Ensemble Kalman Filter (EnKF) parameter calibration in FLARE. The zones represent zones within the water column with increasing numbers corresponding to depth above the lake bottom (e.g., Zone 1 is the first zone on the bottom of the lake). 270

Table S3. Definitions and descriptions of uncertainty sources included in this study. The method of uncertainty quantification is described. For sources whose initial configuration is different from the quantification throughout the forecast run, that is described separately below..... 272

Chapter V: Embedding communication concepts in forecasting training increases students' understanding of ecological uncertainty

Table 1. Glossary of uncertainty communication and ecological forecasting terms taught in Macrosystems EDDIE Module 8: “Using Ecological Forecasts to Guide Decision-Making” as well as examples of how each term is applied to real-world, near-term forecasting 336

Table 2. Summary of courses which participated in assessment of Macrosystems EDDIE Module 8 “Using Ecological Forecasts to Guide Decision-Making.”	337
Table 3. Selected assessment questions and their corresponding learning objectives for Macrosystems EDDIE Module 8: “Using Ecological Forecasts to Guide Decision-Making.”	338
Table 4. Summary statistics of pre- and post-module assessment questions for assessment questions in this study (Q1-8).	339
Table S1. All module learning objectives (LOs) and associated assessment questions and module activities.	353
Table S2. List of ecological forecasts used in Activity A of Module 8 to introduce students to existing examples of ecological forecasts and their visualizations.	355
Table S3. Multiple-choice assessment questions for pre- and post-module assessment questions, with correct answers are indicated in bold.....	357
Table S4. Qualitative questions from pre- and post-module assessments which were graded using the method outlined in Text S1.	359
Table S5. Coding criteria used for Q3 (List some benefits of ecological forecasting) thematic bins. Bins were developed using the method described in Text S1.	360
Table S6. Coding criteria used for Q4 (describe two different ways uncertainty can be visualized) thematic bins. Bins were developed using the method described in Text S1.	

Chapter I: Introduction

Ecosystems are experiencing unprecedented and accelerating rates of change (Bruggemann et al. 2012; Sydeman et al. 2013; Mariani et al. 2022), motivating new approaches for understanding and predicting ecological dynamics (Clark et al. 2001; Luo et al. 2011; Dietze et al. 2018). Over the past several decades, the globe has experienced major shifts in land use, including conversion of millions of hectares of forest into cropland (Estoque et al. 2022), degradation of many ecosystem services such as flood protection and water provisioning (Hou et al. 2023), and a resulting decline in biodiversity (Habibullah et al. 2022; Faria et al. 2023). These land use changes are occurring simultaneously with anthropogenically-induced climate change, with shifts in air temperature and precipitation outside of historical ranges, and an increased frequency of severe weather (Tabari 2020; van der Wiel and Bintanja 2021; Todaro et al. 2022; IPCC 2023). Altogether, these pressures have resulted in broad changes to ecosystem health and productivity (Smith et al. 2009; Komatsu et al. 2019; Rillig et al. 2019), with cascading effects on human and economic health (Feng et al. 2019; Penuelas et al. 2020; de Vries 2021; Abbass et al. 2022).

To help managers understand and anticipate ecosystem changes in advance, a quantitative understanding of how ecosystems will continue to respond to climate and land use change into the future is critical for decision support (Clark et al. 2001; Dietze et al. 2018; Carey et al. 2022b). Ecological forecasts, or predictions of future ecosystem states with quantified uncertainty (Lewis et al. 2022), provide vast potential for predicting ecosystem change and subsequently informing ecosystem management and conservation (Dietze et al. 2018). **However, for decision support tools based on ecological forecasts to be useful for managers, we need**

both improved quantification of ecosystem variability for forecast models, as well as improved communication of ecosystem understanding for decision-making.

First, understanding spatial and temporal heterogeneity in ecosystem functioning is critical for informing the design of monitoring programs and predictive models needed for ecological forecasting. Specifically, an understanding of spatial and temporal heterogeneity can aid in identifying both sites within an ecosystem and time periods which dominate overall ecosystem functioning. Quantifying spatial and temporal heterogeneity is also necessary for developing realistic ecosystem models simulating these ecosystem dynamics. Moreover, quantifying multiple scales of temporal and spatial variability is important for understanding the uncertainty associated with ecosystem predictions and management decision-making.

Second, after building an understanding of ecosystem variability, forecasts can be used to apply that understanding. Forecasts of ecological variables can be made for a range of uses and aid in decision-making broadly. For example, forecasts of the distribution of certain species, such as the Atlantic Sturgeon in Chesapeake Bay, can be used to help fishers decide when and where to fish in order to avoid by-catch of this endangered species (Breece et al. 2018). There are many other examples how forecasts can guide natural resource decision-making, with broad utility across ecosystems (Bodner et al. 2021). However, simply producing forecasts does not guarantee their use in decision-making workflows.

Third, thoughtful consideration of how to communicate and embed these tools in decision-making frameworks is critical for realizing their potential in ecosystem management (von Winterfeldt 2013; Copp et al. 2021). Here, I refer to communication of a decision support tool as the delivery of forecast information to an end user, which may range from a simple word or phrase (e.g., “High risk of encountering endangered species”) to a detailed figure showing

forecasted probability distributions. Similarly, I define decision-making frameworks as the workflow that a user goes through to make a decision, including the delivery and interpretation of the communicated information and the implementation of a decision. Weather forecasts provide a good example of communication of scientific information within a decision-making framework which is ubiquitous to many people's daily lives in the United States (Demuth et al. 2011; Gerst et al. 2017; Fundel et al. 2019). Communication of weather forecasts has been thoughtfully designed for ease of integration within everyday use (i.e., delivery of information via mobile phone applications phones and regular television slots) and ease of perception (i.e., using symbols to communicate a sunny vs. rainy day). Forecasts which are communicated so that users can readily understand the information and subsequently apply it to make a decision have great potential for improving natural resource management and public understanding of ecosystems in the face of climate and land use change.

Understanding global change effects on freshwaters to manage their ecosystem services

Freshwater lakes and reservoirs provide critical ecosystem services and are disproportionately affected by global climate and land use change (Williamson et al. 2009). Situated at the lowest sites within their watershed, freshwaters are natural integrators of the landscape and respond rapidly to changes in surrounding terrestrial ecosystems (Adrian et al. 2009; Williamson et al. 2009). In addition, freshwater ecosystems also provide numerous services including drinking water and food production, flood protection, recreation, and tourism (Aylward et al. 2005, Janssen et al. 2021). Despite the many ecosystem services provided by freshwaters, they are disproportionately threatened globally by human activities (Smith 2003;

Dudgeon et al. 2006; Ormerod et al. 2010; Darwall et al. 2018), underscoring the importance of understanding and conserving these ecosystems.

There are many examples of how global change is substantially altering the physical, chemical, and biological water quality of lakes and reservoirs. In response to shifts in global air temperatures, wind speeds, and rainfall duration and intensity (Tabari 2020; van der Wiel and Bintanja 2021; Todaro et al. 2022), surface water temperatures are increasing in many lakes and reservoirs (O' Reilly et al. 2015; Piccolroaz et al. 2020; Woolway et al. 2020). Changes in lake water temperatures influence annual patterns of thermal stratification in lakes and reservoirs, disrupting mixing regimes and altering the duration of ice cover (Woolway et al. 2020). Further, lake and reservoir bottom-water oxygen concentrations have decreased globally in response to warmer and more stratified waters (Jane et al. 2021). Decreases in oxygen, especially in the bottom waters of lakes, can lead to internal nutrient loading (i.e., release of sediment-bound nitrogen, phosphorus, and/or metals under hypoxic conditions; Sondergaard et al. 2001). In conjunction with increased internal nutrient loading, external nutrient loading (i.e., delivery of nitrogen and phosphorus from point and non-point sources) has also been influenced globally by human activities, with implications for nutrient cycling within lakes and global biogeochemical cycles (Peñuelas and Sardans 2022; Wu et al. 2022). Altogether, land use change has substantially altered the nutrient chemistry of lake water (Mooij et al. 2005; Trolle et al. 2011; Creed et al. 2018) and phytoplankton dynamics (Winder and Sommer 2012). Some lakes are exhibiting increases in phytoplankton blooms (Hou et al. 2022; Liu et al. 2023), including those of toxic cyanobacteria (Paerl and Huisman 2008; Griffith and Gobler 2020). Overall, these shifts in physical, chemical, and biological variables have resulted in ecosystem-level changes in lake and reservoir water quality.

These global changes are altering the spatial and temporal heterogeneity in freshwater ecosystems, necessitating better quantification of ecosystem variability to inform monitoring programs and predictive models used for ecological forecasting. Monitoring programs need to be able to identify sampling locations and time periods in order to best capture overall ecosystem variability. For example, certain locations within a lake or reservoir may exhibit disproportionately high rates of biogeochemical processing (i.e., hotspots) which affect the overall ecosystem carbon, nitrogen, and phosphorus balance (Mcclain et al. 2003; Wik et al. 2016; Bernhardt et al. 2017; Carey et al. 2022a). Similarly, identifying time periods with high processing rates (i.e., hot moments) is also crucial for informing adaptive sampling frequencies, which is needed to constrain the contribution of a lake or reservoir to global biogeochemical budgets (Mcclain et al. 2003; Archer et al. 2015; Bernhardt et al. 2017). Finally, in addition to identifying ecosystem hotspots and hot moments, quantifying spatial and temporal heterogeneity is critical for developing realistic ecosystem models simulating these freshwater dynamics.

Forecasts of important water quality variables can inform decision-making

Forecasts of how multiple water quality variables (including water temperature, dissolved oxygen, and phytoplankton) will change in the future can inform a variety of decisions. First, water temperature determines the dynamics of all temperature-dependent physical processes, including lake stratification, ice cover, and mixing dynamics; atmospheric gas exchange and solubility of gasses throughout the water column; and habitat availability for temperature-sensitive species (Kirillin 2010; Richardson et al. 2017; Pilla and Williamson 2023). Knowing the timing of when stratification will occur and the water temperature at multiple depths can help inform decisions about selective water withdrawal for minimizing the effect of warm reservoir

water on downstream ecosystems (e.g., salmon spawning; Angilletta Jr et al. 2008; Calamita et al. 2021). Similarly, lake and reservoir water temperature and ice forecasts can guide winter transportation and recreation (Fujisaki-Manome et al. 2019).

Second, dissolved oxygen dynamics, which are in part dictated by water temperature dynamics, both drive and respond to a variety of ecosystem processes which are important for decision-making. For example, through consumption and production of oxygen via organismal respiration and primary productivity, respectively, dissolved oxygen dynamics are indicators of ecosystem metabolism and trophic state (Odum 1956). In addition, through controlling the release of reduced nutrients and metals from sediments under low oxygen, or hypoxic, conditions, oxygen is also an important determinant of water quality (Sondergaard et al. 2001). Through both of these mechanisms, predictions of the onset and magnitude of hypoxia are considered indicators of ecosystem function (Jane et al. 2021) and could serve as an early warning indicator of changing trophic state, informing decision-making about whether to install or operate oxygenation systems (Carey et al. 2022b) or enact watershed management efforts to mediate changes in trophic state (Mateus et al. 2014). Additionally, dissolved oxygen concentrations also regulate habitat availability for oxygen-sensitive species (Davis 1975; Kramer 1987; Pilla and Williamson 2023). As a result, predictions of dissolved oxygen concentrations can help inform decisions on whether to stock economically-important fish species based on habitat suitability.

Third, predictions of phytoplankton dynamics can be used to guide many decisions. Increased frequency and intensity of phytoplankton blooms is one of the most salient shifts in water quality in response to global change, with bloom events directly affecting many people's daily lives through beach closures, drinking water advisories, and over \$2 billion dollars

annually in water treatment costs across the United States (Anderson et al. 2000; Hoagland and Scatasta 2006; Dodds et al. 2009, City of Toledo, OH 2014). Advance notice of bloom events could be used to provide public notices of beach closures and drinking water advisories, allow treatment utilities to implement preemptive treatments or undertake management efforts to prevent blooms (e.g., copper sulfate dosing or surface water mixing; (Chen et al. 2018; Liu et al. 2023), and allow citizens to make informed decisions about personal safety risks when swimming (Lofton et al. 2023).

Altogether, continued alteration to historical patterns in lake water temperatures, dissolved oxygen dynamics, and phytoplankton dynamics are expected through the next century (Janse et al. 2015; Chapra et al. 2017; Woolway and Merchant 2019; Woolway et al. 2022; Desgué-Itier et al. 2023). As variability in these important water quality variables continues to increase outside of historical ranges, an improved mechanistic understanding and development of quantifiable predictions to support these decision-making needs is increasingly necessary.

Forecasts as a tool for informing decision-making in freshwater ecosystems

Near-term, iterative ecological forecasting provides an opportunity to both apply ecological understanding and develop quantitative decision-making tools with uncertainty to improve management of ecosystems (Dietze et al. 2018; Carey et al. 2022b; Lewis et al. 2023). Forecasts are often made within an iterative, automated forecasting cycle, in which iterative refers to repeated generation of forecasts, and automated refers to forecasts being made without human intervention on a regular interval (e.g., daily) or, in some cases, triggered by the collection of new observations (White et al. 2019). Forecasts are made on the scale of days to a decade into the future (i.e., the forecast horizon), evaluated with observations when they become

available, and updated to improve forecast predictions over time (Dietze 2017; Dietze et al. 2018). The number of ecological forecasts being generated has increased exponentially over the past 40 years (Lewis et al. 2022), enabling fundamental and applied questions about ecological predictability to be addressed (Lewis et al. 2023).

Many ecological forecasts of water quality have been developed and directly applied within decision-making frameworks. For example, forecasts of short-term ice dynamics in the Great Lakes region were developed and iteratively improved through participation in a targeted working group of social scientists, operational forecast providers, and forecast users to help identify decision needs for the forecast (Fujisake-Manome et al. 2022). Similarly, forecasts of the risk of phytoplankton blooms were co-developed with water quality managers at Lake Vansjø, Norway to address issues with toxic cyanobacteria causing swimming closures in this recreationally-important lake (Jackson-Blake et al. 2022). In addition, forecasts of the timing of peak landings of economically important marine and freshwater fish species have been directly integrated into decision frameworks for market operational planning (Mills et al. 2017).

Across freshwater forecasts more generally, forecasts of water quality variables (e.g., water temperature, dissolved oxygen concentrations, phytoplankton blooms) are relatively rare compared to water quantity forecasts (Lofton et al. 2023), despite the need for water quality forecasts for decision-making. Given that many water quality variables provide critical ecosystem services and can be used as indicators of changing water quality (Sánchez et al. 2007; Simões et al. 2008; Bhateria and Jain 2016), further development of water quality forecasts could provide valuable tools for decision-making.

Several attributes of forecasts are helpful for maximizing forecast utility for decision-making. Specifically, forecasts which are near-term (i.e., have days to decades-ahead forecast

horizons; Tulloch et al. 2020), have quantified accuracy (i.e., the forecasts are assessed relative to observations; Lewis et al. 2022), predict variables which directly inform specific decision applications (e.g., are targeted to address a question such as “will there be a phytoplankton bloom?” rather than “what is the phytoplankton concentration?”; Tulloch et al. 2020), are iteratively produced and accessible for users (Lewis et al. 2022), and incorporate uncertainty within both their quantification and communication are likely most useful for decision-makers.

Uncertainty inherent in forecasts poses a unique challenge for interpreting and using forecasts for decision-making. Despite that uncertainty is an inherent component of an ecological forecast, the number of ecological forecasts which incorporate uncertainty continues to remain low (Lewis et al. 2022, Lofton et al. 2023). This is likely because methods for both quantifying and communicating uncertainty are inconsistent across studies (Spiegelhalter et al. 2011; Wynne et al. 2023), despite that accurate quantification of uncertainty is critical for informing decision-making (Fundel et al. 2019). However, translating uncertainty thresholds for forecast communication tools is notoriously difficult (Spiegelhalter et al. 2011; Potter et al. 2012; Bonneau et al. 2015), with different representations of uncertainty resulting in differential interpretation and decision-making (Ramos et al. 2013; Cheong et al. 2016; McKenzie et al. 2016; Kinkeldey et al. 2017). Ultimately, educational resources which teach necessary quantitative skills, but also engage the next generation of scientists, are critical for developing impactful science (Enquist et al. 2017; Bakermans and Pfeifer 2018; Willson et al. 2023).

Building, applying, and communicating accurate, useful forecasts for decision-making

Producing forecasts for a range of water quality variables requires several important steps (Figure 1). First, *building* a basic understanding of ecosystem functioning and variability (Figure

1a) is critical for developing monitoring programs and models of water quality variables (Figure 1b). Once a foundational understanding of ecosystem variability is developed and implemented in models, *applying* ecosystem understanding through iterative forecast development is necessary for testing and quantifying our forecast's representation of observed conditions (Figure 1c). Next, the decision support tools developed from forecast outputs (Figure 1d) must be *communicated* effectively to forecast users (Figure 1e). If forecasts are effectively communicated, user engagement and interest in understanding and conserving ecosystems is likely to increase (Figure 1f), motivating further investigation to improve ecosystem understanding (Figure 1a). I describe each of these components in further detail below.

First, accurate forecasts rely on *building* an understanding of ecosystem variability and processes over space and time to inform development of ecosystem monitoring programs (Figure 1a, b). Many current water quality monitoring regimes collect lake and reservoir data on annual, monthly, or weekly frequencies, and are typically based on data collection at a single location near the deepest site of the waterbody. However, in some cases, additional sampling may be required to produce accurate forecasts. For example, to better understand spatial and temporal variability of ephemeral phytoplankton blooms, adaptive sampling of phytoplankton communities at multiple sites with high-frequency sensors may help improve phytoplankton forecasts (Den Uyl et al. 2022). By contrast, variables which change more slowly over time (e.g., hypolimnetic water temperature during stratified periods) may not need to be sampled as frequently and at as many sites to produce accurate forecasts. Ecosystem studies to elucidate where and when freshwater quality variables exhibit high rates of change are important for informing the design of monitoring programs to support forecasting.

Once monitoring data are available, they can be used to instantiate ecological understanding and *build* forecast models (Figure 1b). Researchers to date have used many different types of water quality models, ranging from statistical models (e.g., Ward et al. 2014; Lee et al. 2023), to machine learning models (e.g., Zhu et al. 2022), to process-based models (e.g., Soares and Calijuri 2021), which have a broad range of ability to predict water quality (Flynn 2005; Ejigu 2021; Lee et al. 2023). These models span a range of input requirements, include different levels of process information and driver datasets, and have differing levels of difficulty for parameterization, each of which lend themselves to varying suitability across many different applications (Ejigu 2021; Tredennick et al. 2021). For example, while statistical models are generally simpler to calibrate, process-based models typically include a better representation of underlying mechanisms and as such are expected to be better at predicting dynamics outside of historical baselines (Scibek and Allen 2006; Winslow et al. 2017; Me et al. 2018). In contrast, machine learning models tend to be highly accurate (Zhu et al. 2022) but require large amounts of input data and provide little explanation of underlying mechanisms in the processes being predicted (Peters et al. 2014; Lazer et al. 2014; Jia et al. 2018). Consequently, determining the most appropriate modeling approach for a forecasting application is a critical part of forecasting system development.

Second, iterative *application* of ecological understanding through the deployment of forecasts provides a testbed for ecological understanding (Figure 1c). Producing forecasts across time (e.g., inter- and intra-annually), space (e.g., across ecosystems or within an ecosystem), and water quality variables (e.g., water temperature, dissolved oxygen, and chlorophyll-a) can help broaden our understanding of when, where, and why predictability varies (Thomas et al. 2023). In addition, iterative and automated forecast production can also be an important test of the

robustness of cyberinfrastructure, which is critical for consistent delivery of forecasts for forecast users (Carey et al. 2022b).

Third, following production of a diverse catalog of forecasts, decision support tools can be developed for *communicating* forecasts to meet specific user needs (Figure 1d). Importantly, proper communication of forecast uncertainty within decision support tools for forecast users is critical to realizing the potential of forecasts for improved decision-making and increased investment in the study of ecosystems (Figure 1e, f). When uncertainty is not properly communicated, forecasts can be misinterpreted, in which users either assign their own level of uncertainty or interpret the forecast as having no uncertainty at all, leading to overconfidence (Berthet et al. 2016). This latter situation is exemplified in the Red River flood in the United States in 1997, when forecasters reported a flood crest with no uncertainty, leading to misinterpretation by decision-makers, and resulting in the failure of the city's dikes and major flood damage (Pielke 1999). In contrast, communicating well-constrained uncertainty estimates in forecasts can help users properly evaluate risks associated with decision-making. Despite its importance, training opportunities on how to effectively communicate forecast uncertainty for diverse audiences are limited (Brownell et al. 2013; Willson et al. 2023), necessitating the development of educational resources on this topic. Through proper communication of forecast information, there is potential to more effectively engage the general public in ecological conservation and management.

Taken together, these three components (building, applying, and communicating ecological understanding) represent critical aspects for advancing forecasting for freshwater science, motivating the central question of my dissertation: **How can we build, apply, and communicate ecosystem understanding in freshwater ecological forecasts?** My dissertation

addressed this question by conducting multiple whole-ecosystem field surveys, coupled with long-term data analyses and diverse modeling approaches to produce forecasts of multiple water quality variables, and curriculum development to teach ecological forecasting. Altogether, my overarching aim was to improve our understanding and management of freshwater ecosystems.

Scope of dissertation chapters

Overall, my dissertation aimed to *build* ecosystem understanding through field-based research (Figure 1a, b); *apply* ecological understanding through the iterative ecological forecasting cycle (Figure 1c) to develop decision support tools for management and understanding (Figure 1d); and improve *communication* of forecasts for community engagement (Figure 1e, f). In Chapter 2, I quantified the temporal and spatial heterogeneity of biogeochemistry in two reservoirs and identified hotspots of biogeochemical processing in water supply reservoirs. In Chapters 3 and 4, I used a suite of models to produce forecasts of water temperature and dissolved oxygen (Chapter 3) and phytoplankton (Chapter 4). Together, these two chapters build on one another to compare predictability across water quality variables, as well as examine forecast skill across temporal (Chapters 3 and 4) and spatial (Chapter 4) scales. In Chapter 5, I evaluated the effectiveness of a teaching module developed for undergraduates in teaching ecological forecasting and uncertainty communication concepts. This module examines the potential for incorporating these important concepts into undergraduate ecology curricula to increase engagement in ecological forecasting and ecology more broadly.

References

- Abbass K, Qasim MZ, Song H, et al (2022) A review of the global climate change impacts, adaptation, and sustainable mitigation measures. *Environmental Science and Pollution Research* 29:42539–42559. <https://doi.org/10.1007/s11356-022-19718-6>
- Adrian R, Reilly CMO, Zagarese H, et al (2009) Lakes as sentinels of climate change. *Limnology and Oceanography*. 54(6): 2283–2297. DOI: 10.4319/lo.2009.54.6_part_2.2283
- Anderson DM, Hole W, Kaoru Y, White AW (2000) Economic Impacts from Harmful Algal Blooms (HABs) in the United States. NOAA SeaGrant. URL: <https://repository.library.noaa.gov/view/noaa/34913>
- Angilletta Jr MJ, Ashley Steel E, Bartz KK, et al (2008) Big dams and salmon evolution: changes in thermal regimes and their potential evolutionary consequences. *Evolutionary Applications* 1:286–299. <https://doi.org/10.1111/j.1752-4571.2008.00032.x>
- Archer SK, Allgeier JE, Semmens BX, et al (2015) Hot moments in spawning aggregations: implications for ecosystem-scale nutrient cycling. *Coral Reefs* 34:19–23. <https://doi.org/10.1007/s00338-014-1208-4>
- Aylward B, Bandyopadhyay J, Belausteguigotia JC, Borkey P, Cassar AZ, Meadors L, et al (2005) Freshwater ecosystem services. *Ecosystems and human well-being: policy responses*, 3, 213-256.
- Bakermans MH, Pfeifer G (2018) A model for translational science in undergraduate classrooms. *Frontiers in Ecology and the Environment* 16:319–321. <https://doi.org/10.1002/fee.1920>

- Baron JS, Poff NL (2004) Sustaining healthy freshwater ecosystems. *Journal of Contemporary Water Research and Education*, 127:52-59. <https://pubs.er.usgs.gov/publication/1015156>
- Bernhardt ES, Blaszcak JR, Ficken CD, et al (2017) Control Points in Ecosystems: Moving Beyond the Hot Spot Hot Moment Concept. *Ecosystems* 20:665–682.
<https://doi.org/10.1007/s10021-016-0103-y>
- Berthet L, Piotte O, Gaume É, et al (2016) Operational forecast uncertainty assessment for better information to stakeholders and crisis managers. *E3S Web of Conferences* 7:.
<https://doi.org/10.1051/e3sconf/20160718005>
- Bhateria R, Jain D (2016) Water quality assessment of lake water: a review. *Sustainable Water Resources Management* 2:161–173. <https://doi.org/10.1007/s40899-015-0014-7>
- Bodner K, Rauen Firkowski C, Bennett JR, et al (2021) Bridging the divide between ecological forecasts and environmental decision making. *Ecosphere* 12:.
<https://doi.org/10.1002/ecs2.3869>
- Bonneau G, Hege H, Johnson CR, et al (2014) Chapter 1 Overview and State-of-the-Art of Uncertainty Visualization. *Scientific visualization: Uncertainty, multifield, biomedical, and scalable visualization*, 3-27
- Breece MW, Fox DA, Haulsee DE, et al (2018) Satellite driven distribution models of endangered Atlantic sturgeon occurrence in the mid-Atlantic Bight. *ICES Journal of Marine Science* 75:562–571. <https://doi.org/10.1093/icesjms/fsx187>
- Brownell SE, Price JV, Steinman L (2013) Science communication to the general public: Why we need to teach undergraduate and graduate students this skill as part of their formal

scientific training. *Journal of Undergraduate Neuroscience Education* 12:6–10. PMID: 24319399.

Calamita E, Vanzo D, Wehrli B, Schmid M (2021) Lake Modeling Reveals Management Opportunities for Improving Water Quality Downstream of Transboundary Tropical Dams. *Water Resources Research* 57:e2020WR027465.
<https://doi.org/10.1029/2020WR027465>

Carey CC, Hanson PC, Thomas RQ, et al (2022a) Anoxia decreases the magnitude of the carbon, nitrogen, and phosphorus sink in freshwaters. *Global Change Biology* 28:4861–4881.
<https://doi.org/10.1111/gcb.16228>

Carey CC, Woelmer WM, Lofton ME, et al (2022b) Advancing lake and reservoir water quality management with near-term, iterative ecological forecasting. *Inland Waters* 12:107–120.
<https://doi.org/10.1080/20442041.2020.1816421>

Chapra SC, Boehlert B, Fant C, et al (2017) Climate Change Impacts on Harmful Algal Blooms in U.S. Freshwaters: A Screening-Level Assessment. *Environmental Science and Technology* 51:8933–8943. <https://doi.org/10.1021/acs.est.7b01498>

Chen S, Carey CC, Little JC, et al (2018) Effectiveness of a bubble-plume mixing system for managing phytoplankton in lakes and reservoirs. *Ecological Engineering* 113:43–51.
<https://doi.org/10.1016/j.ecoleng.2018.01.002>

Cheong L, Bleisch S, Kealy A, et al (2016) Evaluating the impact of visualization of wildfire hazard upon decision-making under uncertainty. *International Journal of Geographical Information Science* 30:1377–1404. <https://doi.org/10.1080/13658816.2015.1131829>

- City of Toledo, OH (2014) Microcystin Event Preliminary Summary; City of Toledo Department of Public Utilities: Toledo, OH, 2014; <http://toledo.oh.gov/media/132055/Microcystin-Test-Results.pdf> cited in <https://pubs.acs.org/doi/10.1021/acs.est.7b01498>
- Clark JS, Carpenter SR, Barber M, et al (2001) Ecological Forecasts : An Emerging Imperative. *Science* 293:657–661. <https://doi.org/10.1126/science.293.5530.657>
- Copp GH, Vilizzi L, Wei H, et al (2021) Speaking their language – Development of a multilingual decision-support tool for communicating invasive species risks to decision makers and stakeholders. *Environmental Modelling and Software* 135:. <https://doi.org/10.1016/j.envsoft.2020.104900>
- Creed IF, Grimm NB, Trick CG, et al (2018) Global change-driven effects on dissolved organic matter composition : Implications for food webs of northern lakes. *Global Change Biology* 24(8): 3692–3714. <https://doi.org/10.1111/gcb.14129>
- Darwall W, Bremerich V, De Wever A, et al (2018) The *Alliance for Freshwater Life* : A global call to unite efforts for freshwater biodiversity science and conservation: The *Alliance for Freshwater Life* : A global call to unite efforts for freshwater biodiversity science and conservation. *Aquatic Conservation: Marine and Freshwater Ecosystems* 28:1015–1022. <https://doi.org/10.1002/aqc.2958>
- Davis JC (1975) Minimal Dissolved Oxygen Requirements of Aquatic Life with Emphasis on Canadian Species: a Review. *J Fish Res Bd Can* 32:2295–2332. <https://doi.org/10.1139/f75-268>

- de Vries W (2021) Impacts of nitrogen emissions on ecosystems and human health: A mini review. *Current Opinion in Environmental Science & Health* 21:100249.
<https://doi.org/10.1016/j.coesh.2021.100249>
- Demuth JL, Lazo JK, Morss RE (2011) Exploring Variations in People’s Sources, Uses, and Perceptions of Weather Forecasts. *Weather, Climate, and Society* 3:177–192.
<https://doi.org/10.1175/2011WCAS1061.1>
- Den Uyl PA, Thompson LR, Errera RM, et al (2022) Lake Erie field trials to advance autonomous monitoring of cyanobacterial harmful algal blooms. *Frontiers in Marine Science* 9. <https://doi.org/10.3389/fmars.2022.1021952>
- Desgué-Itier O, Melo L, Soares V, et al (2023) Past and future climate change effects on thermal regime and oxygen solubility of four peri-alpine lakes. *Hydrology and Earth System Sciences* 1:1–34. <https://doi.org/10.5194/hess-27-837-2023>
- Dietze MC (2017) Prediction in ecology: A first-principles framework: A. *Ecological Applications* 27:2048–2060. <https://doi.org/10.1002/eap.1589>
- Dietze MC, Fox A, Beck-Johnson LM, et al (2018) Iterative near-term ecological forecasting: Needs, opportunities, and challenges. *Proceedings of the National Academy of Sciences* 115:1424–1432. <https://doi.org/10.1073/pnas.1710231115>
- Dodds WK, Bouska WW, Eitzmann JL, et al (2009) Eutrophication of U.S. Freshwaters: Analysis of Potential Economic Damages. *Environ Sci Technol* 43:12–19.
<https://doi.org/10.1021/es801217q>

- Dudgeon D, Arthington AH, Gessner MO, et al (2006) Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological Reviews* 81:163–182.
<https://doi.org/10.1017/S1464793105006950>
- Ejigu MT (2021) Overview of water quality modeling. *Cogent Engineering* 8.
<https://doi.org/10.1080/23311916.2021.1891711>
- Enquist CAF, Jackson ST, Garfin GM, et al (2017) Foundations of translational ecology. *Frontiers in Ecology and the Environment* 15:541–550. <https://doi.org/10.1002/fee.1733>
- Estoque RC, Dasgupta R, Winkler K, et al (2022) Spatiotemporal pattern of global forest change over the past 60 years and the forest transition theory. *Environmental Research Letters* 17: <https://doi.org/10.1088/1748-9326/ac7df5>
- Faria D, Morante-Filho JC, Baumgarten J, et al (2023) The breakdown of ecosystem functionality driven by deforestation in a global biodiversity hotspot. *Biological Conservation* 283: <https://doi.org/10.1016/j.biocon.2023.110126>
- Feng Z, De Marco A, Anav A, et al (2019) Economic losses due to ozone impacts on human health, forest productivity and crop yield across China. *Environment International* 131:104966. <https://doi.org/10.1016/j.envint.2019.104966>
- Flynn KJ (2005) Castles built on sand: Dysfunctionality in plankton models and the inadequacy of dialogue between biologists and modellers. *Journal of Plankton Research* 27:1205–1210. <https://doi.org/10.1093/plankt/fbi099>

Fujisaki-Manome A, Gill D, Guo T, et al (2019) Knowledge Co-production in a Research-to-Operation (R2O) Process for Development of a Great Lakes Ice Forecast: Reflection from a Stakeholder Engagement. Authorea Preprints.

<https://doi.org/10.1002/essoar.10501135.1>

Fundel VJ, Fleischhut N, Herzog SM, et al (2019) Promoting the use of probabilistic weather forecasts through a dialogue between scientists, developers and end-users. Quarterly Journal of the Royal Meteorological Society 145:210–231.

<https://doi.org/10.1002/qj.3482>

Gerst MD, Kenney MA, Baer AE, et al (2019) Using Visualization Science to Improve Expert and Public Understanding of Probabilistic Temperature and Precipitation Outlooks.

Weather, Climate, and Society 12(1): 117-133. <https://doi.org/10.1175/WCAS-D-18-0094.1>

Griffith AW, Gobler CJ (2020) Harmful algal blooms: A climate change co-stressor in marine and freshwater ecosystems. Harmful Algae 91:101590.

<https://doi.org/10.1016/j.hal.2019.03.008>

Habibullah MS, Din BH, Tan SH, Zahid H (2022) Impact of climate change on biodiversity loss: global evidence. Environmental Science and Pollution Research 29:1073–1086.

<https://doi.org/10.1007/s11356-021-15702-8>

Henrich Bruggemann J, Rodier M, Guillaume MMM, et al (2012) Wicked social-ecological problems forcing unprecedented change on the latitudinal margins of coral reefs: The

- Case of southwest Madagascar. *Ecology and Society* 17:. <https://doi.org/10.5751/ES-05300-170447>
- Hoagland P, Scatasta S (2006) The Economic Effects of Harmful Algal Blooms. In: Granéli E, Turner JT (eds) *Ecology of Harmful Algae*. Springer, Berlin, Heidelberg, pp 391–402
- Hou X, Feng L, Dai Y, et al (2022) Global mapping reveals increase in lacustrine algal blooms over the past decade. *Nature Geoscience* 15:130–134. <https://doi.org/10.1038/s41561-021-00887-x>
- Hou Y, Wei X, Zhang M, et al (2023) A global synthesis of hydrological sensitivities to deforestation and forestation. *Forest Ecology and Management* 529:120718. <https://doi.org/10.1016/j.foreco.2022.120718>
- IPCC (2023) *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, 1st edn. Cambridge University Press
- Jackson-Blake LA, Clayer F, De Eyto E, et al (2022) Opportunities for seasonal forecasting to support water management outside the tropics. *Hydrology and Earth System Sciences* 26:1389–1406. <https://doi.org/10.5194/hess-26-1389-2022>
- Jane SF, Hansen GJA, Kraemer BM, et al (2021) Widespread deoxygenation of temperate lakes. *Nature* 594:66–70. <https://doi.org/10.1038/s41586-021-03550-y>

- Janse JH, Kuiper JJ, Weijters MJ, et al (2015) GLOBIO-Aquatic, a global model of human impact on the biodiversity of inland aquatic ecosystems. *Environmental Science and Policy* 48:99–114. <https://doi.org/10.1016/j.envsci.2014.12.007>
- Janssen ABG, Hilt S, Kosten S, et al (2021) Shifting states, shifting services: Linking regime shifts to changes in ecosystem services of shallow lakes. *Freshwater Biology* 66:1–12. <https://doi.org/10.1111/fwb.13582>
- Jia X, Karpatne A, Willard J, et al (2018) Physics Guided Recurrent Neural Networks For Modeling Dynamical Systems: Application to Monitoring Water Temperature And Quality In Lakes. *arXiv Preprint*. <https://doi.org/10.48550/arXiv.1810.02880>
- Kinkeldey C, MacEachren AM, Riveiro M, Schiewe J (2017) Evaluating the effect of visually represented geodata uncertainty on decision-making: systematic review, lessons learned, and recommendations. *Cartography and Geographic Information Science* 44:1–21. <https://doi.org/10.1080/15230406.2015.1089792>
- Kirillin G (2010) Modeling the impact of global warming on water temperature and seasonal mixing regimes in small temperate lakes. *Boreal Environment Research* 15:279–293. ISSN 1797-2469
- Komatsu KJ, Avolio ML, Lemoine NP, et al (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. <https://doi.org/10.1073/pnas.1819027116>

- Kramer DL (1987) Dissolved oxygen and fish behavior. *Environmental Biology of Fishes* 18:81–92. <https://doi.org/10.1007/BF00002597>
- Lazer D, Kennedy R, King G, Vespignani A (2014) The Parable of Google Flu: Traps in Big Data Analysis. *Science* 343:1203–1205. <https://doi.org/10.1126/science.1248506>
- Lee D-Y, Lee D-S, Cha Y, et al (2023) Data-driven models for predicting community changes in freshwater ecosystems: A review. *Ecological Informatics* 77:102163. <https://doi.org/10.1016/j.ecoinf.2023.102163>
- Lewis ASL, Rollinson CR, Allyn AJ, et al (2023) The power of forecasts to advance ecological theory. *Methods in Ecology and Evolution* 14:746–756. <https://doi.org/10.1111/2041-210X.13955>
- Lewis ASL, Woelmer WM, Wander HL, et al (2022) Increased adoption of best practices in ecological forecasting enables comparisons of forecastability. *Ecological Applications* 32:e2500. <https://doi.org/10.1002/eap.2500>
- Liu H, Chen S, Zhang H, et al (2023) Effects of copper sulfate algicide on the cell growth, physiological characteristics, the metabolic activity of *Microcystis aeruginosa* and raw water application. *Journal of Hazardous Materials* 445:130604. <https://doi.org/10.1016/j.jhazmat.2022.130604>
- Lofton ME, Howard DW, Thomas RQ, Carey CC (2023) Progress and opportunities in advancing near-term forecasting of freshwater quality. *Global Change Biology* 29:1691–1714. <https://doi.org/10.1111/gcb.16590>

- Luo Y, Ogle K, Tucker C, et al (2011) Ecological forecasting and data assimilation in a data-rich era. *Ecological Applications* 21:1429–1442. <https://doi.org/10.1890/09-1275.1>
- Mariani M, Connor SE, Theuerkauf M, et al (2022) Disruption of cultural burning promotes shrub encroachment and unprecedented wildfires. *Frontiers in Ecology and the Environment* 20:292–300. <https://doi.org/10.1002/fee.2395>
- Mateus M, Almeida C, Brito D, Neves R (2014) From Eutrophic to Mesotrophic: Modelling Watershed Management Scenarios to Change the Trophic Status of a Reservoir. *International Journal of Environmental Research and Public Health* 11:3015–3031. <https://doi.org/10.3390/ijerph110303015>
- Mcclain ME, Boyer EW, Dent CL, et al (2003) Biogeochemical Hot Spots and Hot Moments at the Interface of Terrestrial and Aquatic Ecosystems. *Ecosystems* 301–312. <https://doi.org/10.1007/s10021-003-0161-9>
- McKenzie G, Hegarty M, Barrett T, Goodchild M (2016) Assessing the effectiveness of different visualizations for judgments of positional uncertainty. *International Journal of Geographical Information Science* 30:221–239. <https://doi.org/10.1080/13658816.2015.1082566>
- Me W, Hamilton DP, McBride CG, et al (2018) Modelling hydrology and water quality in a mixed land use catchment and eutrophic lake: Effects of nutrient load reductions and climate change. *Environmental Modelling & Software* 109:114–133. <https://doi.org/10.1016/j.envsoft.2018.08.001>

- Mills KE, Pershing AJ, Hernández CM (2017) Forecasting the seasonal timing of Maine's lobster fishery. *Frontiers in Marine Science* 4:1–10.
<https://doi.org/10.3389/fmars.2017.00337>
- Mooij WM, Hülsmann S, De Senerpont Domis LN, et al (2005) The impact of climate change on lakes in the Netherlands: a review. *Aquatic Ecology* 39:381–400.
<https://doi.org/10.1007/s10452-005-9008-0>
- Odum HT (1956) Primary production in flowing waters. *Limnology and Oceanography* 1:102–117
- O' Reilly CM, Rowley RJ, Schneider P, et al (2015) Rapid and highly variable warming of lake surface waters around the globe. *Geophysical Research Letters* 42(24):10,773–10,781.
<https://doi.org/10.1002/2015GL066235>.Received
- Ormerod SJ, Dobson M, Hildrew AG, Townsend CR (2010) Multiple stressors in freshwater ecosystems. *Freshwater Biology* 55:1–4. <https://doi.org/10.1111/j.1365-2427.2009.02395.x>
- Paerl HW, Huisman J (2008) Blooms Like It Hot. *Science* 320:57–58.
<https://doi.org/10.1126/science.1155398>
- Penuelas J, Janssens IA, Ciais P, et al (2020) Anthropogenic global shifts in biospheric N and P concentrations and ratios and their impacts on biodiversity, ecosystem productivity, food security, and human health. *Global Change Biology* 26:1962–1985.
<https://doi.org/10.1111/gcb.14981>

Peñuelas J, Sardans J (2022) The global nitrogen-phosphorus imbalance. *Science* 375:266–267.

<https://doi.org/10.1126/science.abl4827>

Peters DPC, Havstad KM, Cushing J, et al (2014) Harnessing the power of big data: infusing the scientific method with machine learning to transform ecology. *Ecosphere* 5(6): 1-15.

<https://doi.org/10.1890/ES13-00359.1>

Piccolroaz S, Woolway RI, Merchant CJ (2020) Global reconstruction of twentieth century lake surface water temperature reveals different warming trends depending on the climatic zone. *Climatic Change* 160:427–442. <https://doi.org/10.1007/s10584-020-02663-z>

Pielke RA (1999) Who decides? Forecasts and responsibilities in the 1997 Red River flood.

Applied Behavioral Science Review 7:83–101. [https://doi.org/10.1016/S1068-8595\(00\)80012-4](https://doi.org/10.1016/S1068-8595(00)80012-4)

Pilla RM, Williamson CE (2023) Multidecadal trends in ultraviolet radiation, temperature, and dissolved oxygen have altered vertical habitat availability for *Daphnia* in temperate Lake Giles, USA. *Freshwater Biology* 68:523–533. <https://doi.org/10.1111/fwb.14044>

Potter K, Rosen P, Johnson CR (2012) From quantification to visualization: A taxonomy of uncertainty visualization approaches. *Uncertainty Quantification in Scientific Computing: 10th IFIP Advances in Information and Communication Technology* 377 AICT:226–247. https://doi.org/10.1007/978-3-642-32677-6_15

Ramos MH, Van Andel SJ, Pappenberger F (2013) Do probabilistic forecasts lead to better decisions? *Hydrology and Earth System Sciences* 17:2219–2232.

<https://doi.org/10.5194/hess-17-2219-2013>

- Richardson DC, Carey CC, Bruesewitz DA, Weathers KC (2017) Intra- and inter-annual variability in metabolism in an oligotrophic lake. *Aquatic Sciences* 79:319–333.
<https://doi.org/10.1007/s00027-016-0499-7>
- Rillig MC, Ryo M, Lehmann A, et al (2019) The role of multiple global change factors in driving soil functions and microbial biodiversity. *Science* 366:886–890.
<https://doi.org/10.1126/science.aay2832>
- Sánchez E, Colmenarejo MF, Vicente J, et al (2007) Use of the water quality index and dissolved oxygen deficit as simple indicators of watersheds pollution. *Ecological Indicators* 7:315–328. <https://doi.org/10.1016/j.ecolind.2006.02.005>
- Scibek J, Allen DM (2006) Comparing modelled responses of two high-permeability, unconfined aquifers to predicted climate change. *Global and Planetary Change* 50:50–62.
<https://doi.org/10.1016/j.gloplacha.2005.10.002>
- Simões F dos S, Moreira AB, Bisinoti MC, et al (2008) Water quality index as a simple indicator of aquaculture effects on aquatic bodies. *Ecological Indicators* 8:476–484.
<https://doi.org/10.1016/j.ecolind.2007.05.002>
- Smith MD, Knapp AK, Collins SL (2009) A framework for assessing ecosystem dynamics in response to chronic resource alterations induced by global change. *Ecology* 90:3279–3289. <https://doi.org/10.1890/08-1815.1>
- Smith V (2003) Eutrophication of freshwater and coastal marine ecosystems a global problem. *Environmental Science and Pollution Research* 10:126–139.
<https://doi.org/10.1065/espr2002.12.142>

- Soares LMV, Calijuri M do C (2021) Deterministic modelling of freshwater lakes and reservoirs: Current trends and recent progress. *Environmental Modelling and Software* 144:105143. <https://doi.org/10.1016/j.envsoft.2021.105143>
- Sondergaard M, Jensen PJ, Jeppesen E (2001) Retention and internal loading of phosphorus in shallow, eutrophic lakes. *The Scientific World Journal* 1:427–442. <https://doi.org/10.1100/tsw.2001.72>
- Spiegelhalter D, Pearson M, Short I (2011) Visualizing uncertainty about the future. *Science* 333:1393–1400. <https://doi.org/10.1126/science.1191181>
- Sydeman WJ, Santora JA, Thompson SA, et al (2013) Increasing variance in North Pacific climate relates to unprecedented ecosystem variability off California. *Global Change Biology* 19:1662–1675. <https://doi.org/10.1111/gcb.12165>
- Tabari H (2020) Climate change impact on flood and extreme precipitation increases with water availability. *Scientific Reports* 10:13768. <https://doi.org/10.1038/s41598-020-70816-2>
- Thomas RQ, McClure RP, Moore TN, et al (2023) Near-term forecasts of NEON lakes reveal gradients of environmental predictability across the US. *Frontiers in Ecology and the Environment* 21:220–226. <https://doi.org/10.1002/fee.2623>
- Todaro V, D’Oria M, Secci D, et al (2022) Climate Change over the Mediterranean Region: Local Temperature and Precipitation Variations at Five Pilot Sites. *Water (Switzerland)* 14(16). <https://doi.org/10.3390/w14162499>

- Tredennick AT, Hooker G, Ellner SP, Adler PB (2021) A practical guide to selecting models for exploration, inference, and prediction in ecology. *Ecology*. 102(6):e03336
<https://doi.org/10.1002/ecy.3336>
- Trolle D, Hamilton DP, Pilditch CA, et al (2011) Predicting the effects of climate change on trophic status of three morphologically varying lakes: Implications for lake restoration and management. *Environmental Modelling & Software* 26:354–370.
<https://doi.org/10.1016/j.envsoft.2010.08.009>
- Tulloch AIT, Hagger V, Greenville AC (2020) Ecological forecasts to inform near-term management of threats to biodiversity. *Global Change Biology* 26(10): 5816-5828.
<https://doi.org/10.1111/gcb.15272>
- van der Wiel K, Bintanja R (2021) Contribution of climatic changes in mean and variability to monthly temperature and precipitation extremes. *Communications Earth and Environment* 2:1–11. <https://doi.org/10.1038/s43247-020-00077-4>
- von Winterfeldt D (2013) Bridging the gap between science and decision making. *Proceedings of the National Academy of Sciences* 110:14055–14061.
<https://doi.org/10.1073/pnas.1213532110>
- Ward EJ, Holmes EE, Thorson JT, Collen B (2014) Complexity is costly: A meta-analysis of parametric and non-parametric methods for short-term population forecasting. *Oikos* 123:652–661. <https://doi.org/10.1111/j.1600-0706.2014.00916.x>

- White EP, Yenni GM, Taylor SD, et al (2019) Developing an automated iterative near- - term forecasting system for an ecological study. *Methods in Ecology and Evolution* 10(3):332–344. <https://doi.org/10.1111/2041-210X.13104>
- Wik M, Varner RK, Anthony KW, et al (2016) Climate-sensitive northern lakes and ponds are critical components of methane release. *Nature Geosciences* 9:99–105. <https://doi.org/10.1038/ngeo2578>
- Williamson CE, Saros JE, Vincent WF, Smol JP (2009) Lakes and reservoirs as sentinels, integrators, and regulators of climate change. *Limnology and Oceanography* 54:2273–2282. https://doi.org/10.4319/lo.2009.54.6_part_2.2273
- Willson AM, Gallo H, Peters JA, et al (2023) Assessing opportunities and inequities in undergraduate ecological forecasting education. *Ecology and Evolution* 13(5):e10001. <https://doi.org/10.1002/ece3.10001>
- Winder M, Sommer U (2012) Phytoplankton response to a changing climate. *Hydrobiologia* 698:5–16. <https://doi.org/10.1007/s10750-012-1149-2>
- Winslow LA, Hansen GJA, Read JS, Notaro M (2017) Large-scale modeled contemporary and future water temperature estimates for 10774 Midwestern U.S. Lakes. *Scientific Data* 4:170053. <https://doi.org/10.1038/sdata.2017.53>
- Woolway RI, Kraemer BM, Lenters JD, et al (2020) Global lake responses to climate change. *Nature Reviews Earth & Environment* 1:388–403. <https://doi.org/10.1038/s43017-020-0067-5>

Woolway RI, Merchant CJ (2019) Worldwide alteration of lake mixing regimes in response to climate change. *Nature Geoscience* 12:271–276. <https://doi.org/10.1038/s41561-019-0322-x>

Woolway RI, Sharma S, Smol JP (2022) Lakes in Hot Water: The Impacts of a Changing Climate on Aquatic Ecosystems. *BioScience* 72:1050–1061. <https://doi.org/10.1093/biosci/biac052>

Wu Z, Li J, Sun Y, et al (2022) Imbalance of global nutrient cycles exacerbated by the greater retention of phosphorus over nitrogen in lakes. *Nat Geosci* 15:464–468. <https://doi.org/10.1038/s41561-022-00958-7>

Wynne JH, Woelmer W, Moore TN, et al (2023) Uncertainty in projections of future lake thermal dynamics is differentially driven by lake and global climate models. *PeerJ* 11:e15445. <https://doi.org/10.7717/peerj.15445>

Zhu M, Wang J, Yang X, et al (2022) Eco-Environment & Health A review of the application of machine learning in water quality evaluation. *Eco-Environment & Health* 1:107–116. <https://doi.org/10.1016/j.eehl.2022.06.001>

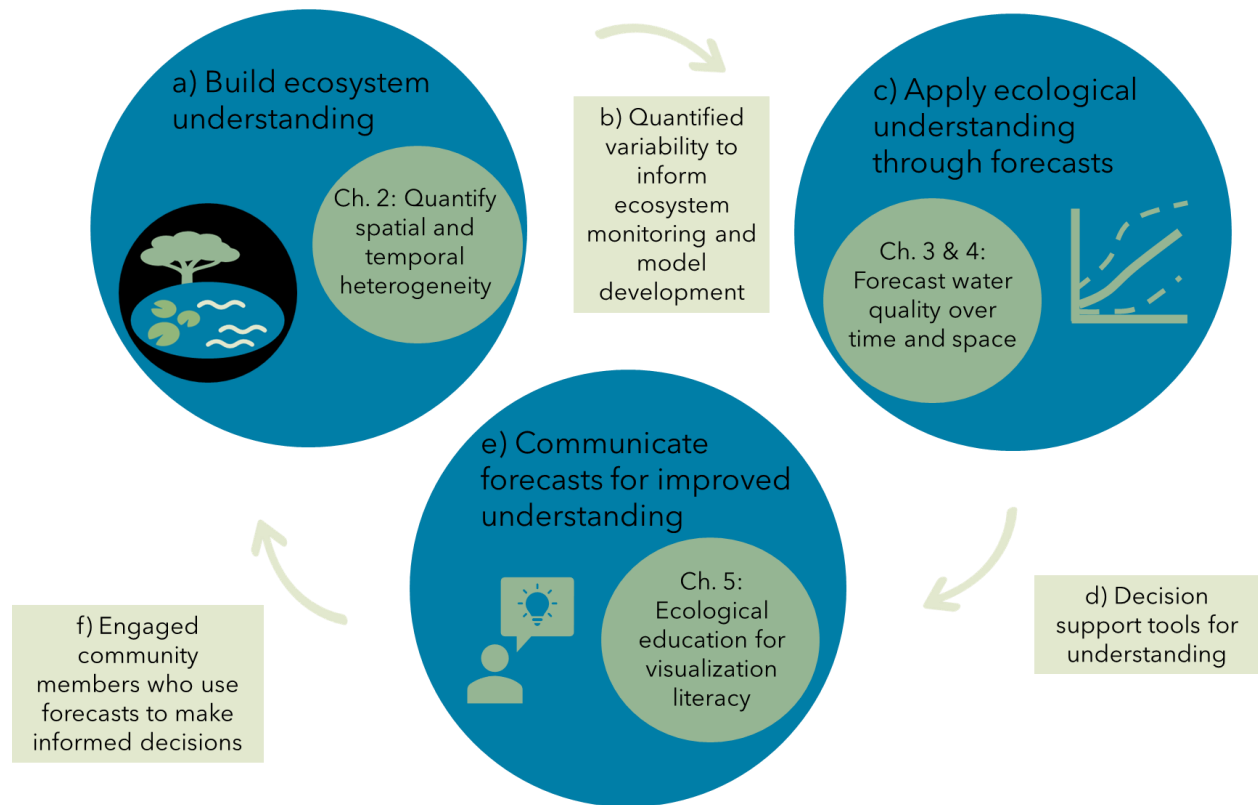


Figure 1. Iterative conceptual framework that integrates my four dissertation chapters (abbreviated as Ch.) to build, apply, and communicate ecological understanding via forecasting.

Chapter II: The importance of time and space in biogeochemical heterogeneity and processing along the reservoir ecosystem continuum*

Whitney M. Woelmer¹, Alexandria G. Hounshell^{1,2}, Mary E. Lofton¹, Heather L. Wander¹, Abigail S.L. Lewis¹, Durelle Scott³, Cayelan C. Carey¹

¹Department of Biological Sciences, Virginia Tech, 926 West Campus Drive, Blacksburg, Virginia 24061, USA.

² Current Address: National Center for Coastal Ocean Science, National Oceanic and Atmospheric Administration, Beaufort, NC, 28516 USA

³Department of Biological Systems Engineering, Virginia Tech, Blacksburg, VA, 26060 USA

*Published in *Aquatic Sciences* with the citation:

Woelmer, W. M., Hounshell, A. G., Lofton, M. E., Wander, H. L., Lewis, A. S., Scott, D., & Carey, C. C. (2023). The importance of time and space in biogeochemical heterogeneity and processing along the reservoir ecosystem continuum. *Aquatic Sciences*, 85(2), 66.

This article is published in *Aquatic Sciences* as Open Access with a CC-BY-ND copyright held by the authors, giving the authors permission to republish in this dissertation.

Abstract

Globally-significant quantities of carbon (C), nitrogen (N), and phosphorus (P) enter freshwater reservoirs each year. These inputs can be buried in sediments, respired, taken up by organisms, emitted to the atmosphere, or exported downstream. While much is known about reservoir-scale biogeochemical processing, less is known about spatial and temporal variability of biogeochemistry within a reservoir along the continuum from inflowing streams to the dam. To address this gap, we examined longitudinal variability in surface water biogeochemistry (C, N, and P) in two small reservoirs throughout a thermally-stratified season. We sampled total and dissolved fractions of C, N, and P, and chlorophyll-a from each reservoir's major inflows to the dam. We found that heterogeneity in biogeochemical concentrations was greater over time than space. However, dissolved nutrient and organic carbon concentrations had high site-to-site variability within both reservoirs, potentially as a result of shifting biological activity or environmental conditions. When considering spatially explicit processing, we found that certain locations within the reservoir, most often the stream-reservoir interface, acted as 'hotspots' of change in biogeochemical concentrations. Our study suggests that spatially explicit metrics of biogeochemical processing could help constrain the role of reservoirs in C, N, and P cycles in the landscape. Ultimately, our results highlight that biogeochemical heterogeneity in small reservoirs may be more variable over time than space, and that some sites within reservoirs play critically important roles in whole-ecosystem biogeochemical processing.

Introduction

Of freshwater ecosystems, reservoirs in particular play a disproportionately large role in global biogeochemical cycles, transforming and burying large amounts of carbon (C), nitrogen

(N), and phosphorus (P; Harrison et al. 2009; Maavara et al. 2015, 2020; Powers et al. 2015; Shaughnessy et al. 2019, Stratton and Grant, 2019). Reservoirs receive a much higher quantity of nutrients (N, P) and organic C than many naturally-formed lakes due to their high watershed-to-surface area ratio and corresponding large terrestrial drainage area, making them important for biogeochemical processing within the landscape (Thornton et al. 1990; Harrison et al. 2009; Hayes et al. 2017).

Within a reservoir, multiple processes interact to control the biogeochemical cycling and fate of C, N, and P. Due to longer residence times in reservoir basins than incoming streams (Nilsson et al. 2005), organic matter and nutrients which enter the reservoir are altered via biotic transformation (e.g., uptake by organisms, decomposition; Harrison et al. 2009; Shaughnessy et al. 2019) or deposited within reservoir sediments as they travel downstream (Syvitski et al. 2005; Tranvik et al. 2009; Clow et al. 2015). Altogether, the balance of these physical and biogeochemical processes results in major changes in the quantity and quality of material leaving the reservoir and directly drives the stoichiometry of C, N, and P exported to downstream ecosystems (Maranger et al. 2018; Wang et al. 2018, Carey et al. 2022a). Consequently, reservoirs can either act as nutrient and C sources (i.e., production within the waterbody) or sinks (i.e., burial in the waterbody or emission to the atmosphere; Harrison et al. 2009; Powers et al. 2015; Maranger et al. 2018). However, because most studies focus on biogeochemical processing at a single location within a reservoir, typically the deepest site (e.g., the U.S. Environmental Protection Agency National Lake Assessment), less is known about the spatial variation in biogeochemistry which may occur between a reservoir's inflow streams and its dam.

Reservoirs are physically heterogeneous ecosystems, which likely influences biogeochemical heterogeneity along the reservoir continuum. Here, we define “reservoir

continuum” as the spatial gradient from the inflowing stream-reservoir interface to the reservoir dam. Thornton et al. (1990) proposed a heuristic model that described the reservoir continuum as a longitudinal gradient within a reservoir consisting of riverine, transitional, and lacustrine zones, each with distinct physical, chemical, and biological characteristics. Broadly, Thornton et al. (1990) posited that nutrient availability should decrease from the riverine to transitional to lacustrine zone because of decreasing water velocity, lower allochthonous inputs of sediment and organic matter, and increasing sedimentation of particulate fractions (Figure 1a). According to this heuristic model, photosynthetic biological activity, or by proxy, chlorophyll-a (chl-a), should peak in the transitional zone (Figure 1a) due to relatively high nutrient and light availability for primary production (Figure 1a).

Empirical studies examining the spatial heterogeneity of biogeochemical variables in reservoirs have found inconsistent support for Thornton et al. (1990)’s heuristic model (Figure 1b-d). Some studies have observed peak concentrations of nutrients and chl-a occasionally following Thornton et al. (1990)’s model (Figure 1b-d; Borges et al. 2008; Rychtecky and Znachor 2011; Soares et al. 2012, Berberich et al. 2020), but often showing unclear or inconsistent patterns (Figure 1b-d; Gloss et al. 1980; Borges et al. 2008; Varol et al. 2012; Woldeab et al. 2018; Carneiro and Bini 2020) or patterns contrary to expectations (Figure 1b-d; Scott et al. 2009; Soares et al. 2012; Woldeab et al. 2018, Berberich et al. 2020). Observed discrepancies in spatial patterns of chl-a and nutrient concentrations among reservoirs may result from reservoir specific characteristics (e.g., morphology, land-use, management, age, etc.), as well as differences in the behavior of varying solutes (e.g., NO_3 vs. NH_4 , as well as total vs. dissolved fractions), and other biogeochemical processes which were not accounted for in Thornton et al. (1990)’s original model. Ultimately, differences in observed patterns of water

chemistry and chl-a among reservoirs suggest that more research is needed to characterize reservoir biogeochemical cycling across space and through time.

While Thornton et al. (1990)'s heuristic model does not directly predict organic C concentrations or types along the reservoir continuum, it does propose patterns for the sedimentation of different types of organic C. Specifically, Thornton et al. (1990)'s heuristic model predicts sedimentation of allochthonous C to be highest in the riverine region of the reservoir due to terrestrial material transported from the reservoir inflow. In contrast, autochthonous C sedimentation would be highest in the transitional zone, where chl-a is expected to peak, due to the contribution of algal biomass to autochthonous C production and sedimentation. These expectations regarding C sedimentation imply changes in organic C quantity and quality along the reservoir continuum are likely also occurring, potentially affecting multiple ecosystem processes including light penetration (Schindler et al. 1996; Solomon et al. 2015), and C lability (Guillemette et al. 2013), which is a critical control of decomposition rates (Strauss and Lamberti 2002; Jane and Rose 2018). Berberich et al. (2020) found that organic matter in sediments and porewater of the upstream riverine zone was dominated by terrestrially-derived organic matter, and observed an increase in autochthonous organic matter within sediments and further downstream. However, to the best of our knowledge, only one study has examined longitudinal trends in surface water reservoir organic C quality. Marcé et al. (2021) identified a transition from terrestrially-derived organic C to autochthonous organic C in surface water along the continuum of Sau Reservoir in Catalonia, Spain. Their findings suggest high rates of organic C processing within the reservoir from the one sampling day of the study, which motivates the need for more data over a longer period of time to examine the consistency of these spatial patterns, for both organic C and other biogeochemical variables.

Drivers which change temporally may play an equal or even more important role than drivers which change spatially in influencing biogeochemical heterogeneity within reservoirs. For example, seasonal changes in hydrology may be an important driver of biogeochemical heterogeneity (Williamson et al. 2021), both directly via loading or dilution of biogeochemical variables (Woldeab et al. 2018), as well as indirectly via reservoir residence time, which will vary throughout the year (Soares et al. 2012). Seasonality in air temperature and solar radiation alters biogeochemical processes such as mineralization, nutrient uptake, and biological growth rates (Wetzel 2001; Weathers et al. 2021) and can increase thermal stratification, resulting in decreased oxygen concentrations in reservoir bottom waters and increased release of certain nutrients from sediment in the summer (Nürnberg 1988; Rydin 2000; Boström et al. 1988). Temporally variable storm events can alter rates of primary production and respiration (Jennings et al. 2012; de Eyto et al. 2016; Kasprzak et al. 2017), as well as disrupt thermal stratification, mixing nutrient-rich hypolimnetic water into the surface. Changes in both allochthonous and autochthonous C are driven by seasonal dynamics, often with an influx of allochthonous material during snowmelt or autumn leaf senescence, and autochthonous C (e.g., phytoplankton and macrophytes) typically peaking in spring or summer (Sommer et al. 1986). Indeed, numerous studies examining seasonal variation in reservoirs have found significant trends in biogeochemical variables over the course of a year (Torres et al. 2007; Varol et al. 2012; Woldeab et al. 2018; Klippel et al. 2020; Waldo et al. 2021). Despite the many impacts of seasonality on biogeochemical variables, the relative magnitude of spatial and temporal heterogeneity in reservoir biogeochemistry remains unknown. In addition to naturally occurring changes over seasons, some reservoirs also experience substantial management activity which can lead to temporal variability in biogeochemistry. For example, hypolimnetic oxygenation (and

reservoir drawdown (can both disrupt typical seasonal limnological and biogeochemical trends, including thermal stratification, nutrient dynamics, and phytoplankton within the water column (e.g., Furey et al. 2004, Preece et al. 2019, Keller et al. 2021, Matsuzaki et al. 2022).

Analyzing trends in spatial variability of biogeochemical variables may help identify important sites where reservoir processing disproportionately occurs (i.e., biogeochemical hotspots, McClain et al. 2003 or control points, Bernhardt et al. 2017). Hotspots of biogeochemical processing often occur at ecosystem interfaces due to the occurrence of gradients, which either deliver limiting nutrients or provide substrate for chemical reactions (McClain et al. 2003; Sadro et al. 2011; Hotchkiss et al. 2018). Within a reservoir, the inflowing stream-reservoir interface may be a critical point for processing incoming biogeochemical variables (DelSontro et al. 2011, Berberich et al. 2020, Linkhorst et al. 2021, Marcé et al. 2021). However, it is likely that the importance of this ecosystem interface may vary over time. Studies documenting spatial hotspots of biogeochemical processing in reservoirs are rare (although see Marcé et al. 2021 for estimates of organic C hotspots, and DelSontro et al. 2011, Berberich et al. 2020, McClure et al. 2020, Linkhorst et al. 2021 for estimates of inorganic C hotspots), and to our knowledge none exist which examine hotspots across seasons or compare the relative importance of spatial heterogeneity to seasonal (or temporal) heterogeneity of multiple biogeochemical variables in reservoirs.

In this study, we examined the spatial and temporal heterogeneity of chl-a, C, N, and P along a reservoir continuum in two small, eutrophic reservoirs. Both reservoirs are located within ~3 km of each other, with similar age, geology, weather, and land-use, allowing us to constrain external driving factors in biogeochemical heterogeneity. We sampled surface water along a longitudinal gradient from stream inflow to the reservoir dam and outflow in both reservoirs

every month from April to October. Our main questions were: 1) How do concentrations of biogeochemical variables change along the reservoir continuum? 2) What is the relative magnitude of spatial and temporal heterogeneity in biogeochemistry along the reservoir continuum? and 3) Are there “hotspots” of biogeochemical processing along the reservoir continuum?

Methods

Study sites

Falling Creek Reservoir and Beaverdam Reservoir are both small (FCR: 0.119 km²; BVR: 0.39 km²), shallow (maximum depth FCR: 9.3 m; BVR: 11 m at the time of this study) dimictic reservoirs owned and operated by the Western Virginia Water Authority as drinking water supplies (Figure 2). Both reservoirs are located in Vinton, Virginia, USA (FCR at 37.30325, -79.8373; BVR at 37.31288, -79.8159) and are primarily surface water effluent reservoirs with spillways as their primary outflows.

BVR is located in an adjacent watershed to FCR and the two reservoirs are connected seasonally (Figure 2). BVR has an outflow pipe installed at an upstream site (B2) which delivers a small volume of water via an underground tunnel to the FCR watershed, feeding the inflow stream which enters FCR on the northeastern side of the reservoir (site FS2, Figure 2b). The outflow pipe is installed at ~2.4 m below full pond (maximum reservoir depth = 13.4 m). However, during the time of this study, water levels in BVR were 2.0-2.9 m below full pond, meaning that the pipe was either just under the surface (April-June 2019) or above the surface of the water (July-October 2019) throughout the study. As a result, delivery of water from BVR to FCR occurred only during the first three sampling events of this study (April-June 2019).

Moreover, in our two focal reservoirs, there were no management drawdowns during the study period. While there is a hypolimnetic oxygenation system in Falling Creek Reservoir, seasonal thermal stratification was not disrupted throughout the study period (Figure S4), so epilimnetic sampling was unlikely to be unaffected (Gerling et al. 2014).

We sampled the entire reservoir continua of BVR and FCR in this study. There are two major inflows to FCR, which we sampled immediately before they entered the reservoir at site FS1 located on a forested stream, and site FS2 located on a wetland-stream complex (Figure 2b). BVR has multiple inflow streams, but we chose to focus on two major forested inflows in the western (site BS1) and eastern (site BS2) arms, which represent the sub-watersheds with the largest contributing area within the major western and eastern arms, respectively (Figure 2b, Text S1). In addition, we monitored discharge, but not water chemistry, at the spillway of FCR (FS3) to estimate the volume of water leaving the reservoir. Drinking water was not regularly extracted from either reservoir during the time of the study, so the primary export of water from FCR was via the surface spillway (site FS3), whereas BVR's primary export was via a stream draining below site B4. Consequently, the primary flow of water in both reservoirs was from the inflowing streams to the dam, as is typically found in run-of-river reservoirs.

Within each reservoir, we chose sites along a gradient from the stream inflows to the reservoir dam, resulting in 4 reservoir sites in FCR (F1, F2, F3, and F4 in order of increasing distance from the inflow streams) and 5 reservoir sites in BVR (B1a in the western arm, B1b in the eastern arm, B2, B3, and B4 in order of increasing distance from the inflow streams; Table S2). We also monitored one site within each inflowing stream to FCR ($n = 2$) and BVR ($n = 2$) just upstream of the reservoir-stream interface, resulting in a total of 14 monitoring sites, 9 sites

within each reservoir, 4 inflowing stream sites, and one stream site downstream of FCR (Figure 2b).

Field sampling

We sampled all sites in both reservoirs (Figure 2) on the same day once every month from April to October 2019 ($n = 7$ sampling days), which allowed us to capture broad spatial and temporal changes in reservoir biogeochemistry. On each sampling day, we sampled surface water from downstream to upstream (e.g., from F4 to F1 and then onto the stream sites).

We measured a suite of surface physico-chemical variables (water temperature, dissolved oxygen, and specific conductance) using a handheld YSI and conductivity probe (Carey et al. 2022b). Additionally, water samples were collected from the surface at all sites and analyzed for chl-a, dissolved organic C (DOC), total N and P (TN, TP), and dissolved N and P (ammonium: NH_4 ; nitrate: NO_3 ; soluble reactive phosphorus: SRP) and DOC quality via excitation emission matrices (EEMs; Fellman et al. 2010). EEMs measure the natural fluorescence of dissolved organic matter to characterize DOC quality (Fellman et al. 2010); here, we specifically focused on metrics correlated with allochthonous vs. autochthonous origin of DOC. We refer readers to Carey et al. (2021), Hounshell et al. (2021), and Carey et al. (2022b) for detailed methods on sample collection, storage, and analysis of total and dissolved N, P, and DOC; EEMS; and chl-a; respectively.

We measured discharge at all stream sites (FS1, FS2, FS3, and BS1, BS2) using either a salt slug or flowmeter method, depending on the stream site (see Carey et al. 2022c for a detailed description of discharge measurement methods). Discharge data were used to calculate the transport of solutes from upstream to downstream and to calculate mass balance estimates.

Because we did not have a sampling location at the spillway of Beaverdam Reservoir, we calculated outflow using a mass balance approach based on modeled inflow and reservoir volume (Text S3).

In addition, we collected data for other environmental variables which might be associated with spatial and temporal heterogeneity, including whole-reservoir water residence time (Figure S1), meteorological variables (air temperature, precipitation, Figure S2), surface water temperature and dissolved oxygen (Figure S3), and thermal stratification metrics (thermocline depth and Schmidt stability, Figure S4). Thermal stratification metrics were calculated at the deepest site of each reservoir (F4 and B4) using full water column temperature profiles which were collected as part of both reservoirs' routine monitoring programs (Carey et al. 2022d, 2022e). Water residence times were calculated by dividing the total reservoir volume by the summed inflow to the reservoir for a given day (following Gerling et al. 2016). Thermocline depth and Schmidt stability were calculated at the deepest site using the R package rLakeAnalyzer (Winslow et al. 2019). Distance between sites within each reservoir were measured using GPS points and the Measure Tool in ArcGIS Pro (2021).

Chemical analyses

All chemical analyses were conducted within six months of collection following the standard methods outlined by Carey et al. (2020), Carey et al. (2021), and Hounshell et al. (2021). We used traditional peak-picking methods to identify broad types of autochthonous and allochthonous DOM within the EEMs (i.e., Fellman et al. 2010; Gabor et al. 2014), specifically focusing on Peaks T (Excitation: 275, Emission: 340) and A (Excitation: 260, Emission: 400-460). These two metrics represent microbial-like and humic-like fractions of dissolved organic

matter, which have been correlated with autochthonous-like (e.g., algal-derived) or allochthonous-like (e.g., terrestrial-derived) DOC, respectively (Fellman et al. 2010; Gabor et al. 2014). We used analytical limits of quantitation for nutrient and carbon analyses to interpret ecologically meaningful differences in concentrations (Table S3).

Examining biogeochemical distributions along the reservoir continuum

We first compared concentrations of nutrients and chl-a at each sampling site to expected spatial patterns of nutrients and chl-a from the riverine to the lacustrine zone according to Thornton et al. (1990)'s model. We limited this visual analysis to variables which were included in Thornton's model (i.e., chl-a and nutrients: TN, TP, NH₄, NO₃, SRP). We also compared temporal variability in concentrations of all biogeochemical variables (i.e., chl-a and nutrients: TN, TP, NH₄, NO₃, SRP, as well as DOC and carbon quality: Peaks T and A) across sampling days. Lastly, we examined nutrient stoichiometry to estimate the relative availability of N, and P, which influences phytoplankton nutrient limitation, among other ecosystem processes (Klausmeier et al. 2004; Hessen et al. 2013, Sterner and Elser 2002). Specifically, we calculated molar ratios of TN:TP, DIN:SRP (dissolved inorganic nitrogen to SRP), SRP:TP, and DIN:TN. We calculated DIN as the sum of NH₄ and NO₃.

Drivers of heterogeneity

To determine the relative magnitude of spatial and temporal heterogeneity of biogeochemistry within the reservoir continuum, we used the coefficient of variation (CV) as a metric of heterogeneity (following Sadro et al. 2012; Soares et al. 2012). We calculated CV across both space (CV_{space}) and time (CV_{time}). CV_{space} was calculated across all sites within each

individual reservoir (e.g., F1-4, B1-4, dark brown circles on Figure 2) on a given sampling day separately for each of the n=7 sampling days. In contrast, CV_{time} was calculated individually for each reservoir site (n = 4 sites at FCR, n = 5 at BVR; i.e., dark brown circles on Figure 2) across all 7 sampling dates. As a result, we calculated 7 values of CV_{space} for FCR and 7 values of CV_{space} for BVR, and 4 values of CV_{time} for FCR and 5 values of CV_{time} for BVR. We report the mean, minimum, and maximum of these distributions in Table S4.

We compared CV_{time} to CV_{space} across both reservoirs to estimate the relative magnitude of spatial and temporal heterogeneity. Additionally, we compared CV_{time} and CV_{space} between FCR and BVR to estimate differences between reservoirs. All comparisons were made using unpaired Wilcoxon tests (Alboukadel 2020).

Spatially explicit processing of biogeochemical variables

When analyzing reservoir biogeochemical processing, many studies calculate the mass balance of solutes coming into and out of a waterbody, thereby estimating whether the waterbody as a whole was a sink (retention of nutrients/carbon) or source (release of nutrients/carbon) for a given analyte (e.g., Vollenweider 1975; Torres et al. 2007; Brett and Benjamin 2008; Keys et al. 2019). However, because processing may not occur uniformly along a longitudinal gradient, we extended the mass balance approach to calculate a spatially explicit metric of change in biogeochemical mass standardized by the area between sampling sites along the stream-reservoir continuum in the surface water, Δ_{var} , using Equation 1:

$$\Delta_{var} = \frac{\frac{[var]_d * Q_d}{sp\ cond_d} - \frac{[var]_u * Q_u}{sp\ cond_u}}{A_R} \quad \text{Eq. 1}$$

where $[var]_d$ is the concentration of a given variable at the downstream reservoir site in units of $\mu\text{g/L}$, mg/L , or RFU (relative fluorescence units) depending on the analyte (see Table S3); Q_d is

the discharge at the downstream site (m^3/s), $sp\ cond_d$ is the specific conductance at the downstream reservoir site ($\mu\text{S}/\text{Cm}$), included here to account for unobserved intrusions of solutes into the reservoir which may occur between sites (more detail below), and A_R is the area of the reservoir section between two sampling sites (Table S5); $[var]_u$, Q_u , and $sp\ cond_u$ are similarly defined but at the upstream site. Q_u was calculated for each stream site on each sampling day. We recognize that standardizing processing rates by the area between sampling sites assumes one-directional flow through the reservoir. As a check on this assumption, and following the precedent of other studies (e.g., Vollenweider 1975; Torres et al. 2007; Keys et al. 2019), we also calculated Δ_{var} without standardizing by area (Text S4, Eq. S1).

Using Eq. 1, we calculated Δ_{var} both for the whole reservoir ($\Delta_{var}WR$) and along the reservoir continuum from one site to the next ($\Delta_{var}Site$) in BVR and FCR. As a result, $\Delta_{var}WR$ represents the traditional mass balance approach of estimating solutes coming into and leaving the whole reservoir corrected for the area of the whole reservoir. In comparison, $\Delta_{var}Site$ represents a localized version of the same metric, i.e., estimating solutes entering and leaving the surface water within a given location, calculated at multiple locations along the reservoir continuum and corrected for the area of the reservoir between sampling sites (Table S5). In the case of $\Delta_{var}WR$, the upstream site is the reservoir inflow(s), and the downstream site is the reservoir outflow; whereas for $\Delta_{var}Site$, the upstream and downstream sites are determined along the flow path of the reservoir continuum (e.g., when calculating $\Delta_{var}Site$ between F1 and F2, F1 is the upstream site and F2 is the downstream site). As a result, positive values indicate that the reservoir or site acted as a source of a given biogeochemical variable downstream, whereas negative values indicate that the reservoir or site acted as a sink. By comparing $\Delta_{var}WR$ for a given day to $\Delta_{var}Site$ throughout the reservoir, we can identify locations within the

reservoir which disproportionately contribute to whole-reservoir biogeochemical processing, acting as biogeochemical hotspots.

For most reservoir sites, $\Delta_{var}Site$ was calculated as the difference between the single upstream and single downstream sites. However, when there were two upstream stream sites (e.g., at F1 when both FS1 and FS2 flow into FCR; Figure 2b), the solute mass at both sites was summed to calculate the total input, divided by the reservoir area between sampling sites. At site B3, where the western and eastern arms of BVR converge, we calculated a weighted average of the mass of the two incoming sites, B2 and B1b, based on a weighted average of the incoming discharge rates.

Given the availability of data for this study, our spatially explicit mass balance approach provided a robust metric of spatial variability in reservoir biogeochemical processing with some caveats. First, given limited resources for monitoring discharge at within-reservoir sites, we used the same Q_d for all sites within a given reservoir on each day, measured at the outflow of the reservoir on each sampling date. Second, we strategically chose to monitor inflow streams to the reservoirs which contributed the most watershed area (Text S1), given limited resources for stream monitoring. This method assumed that the majority of incoming water was delivered through the streams we monitored, and additional sources of water and associated solutes were negligible. We accounted for any unmonitored intrusions of solutes (e.g., groundwater intrusion or smaller tributaries which were not monitored) through the use of specific conductance, which can be used as a conservative tracer due to limited biological uptake relative to C, N, or P at each site (Baker and Webster 2017; e.g. Triska et al. 1989; Casas-Ruiz et al. 2017). We assume that unknown, and thus unmeasured, sources to the reservoirs similarly follow this conservative tracer approach. Third, while this method does not partition out the specific abiotic or biotic

processes which may have occurred, it allowed us to estimate ecologically meaningful changes in biogeochemical variables along the reservoir continuum.

All analyses were conducted in R 4.0.3 (R Core Team, 2020), all code to recreate the analyses is available in the Zenodo repository (Woelmer, 2023), and all field and laboratory data are published with extensive metadata in the Environmental Data Initiative repository (Hounshell et al. 2021, Carey et al. 2020, Carey et al. 2021, Carey et al. 2022b, Carey et al. 2022c, Carey et al. 2022d, Carey et al. 2022e, Carey et al. 2022f).

Results

Distribution of biogeochemical variables along the reservoir continuum

In general, our observations of nutrients and chl-a (Figure 3) did not follow the predicted patterns of Thornton et al. (1990)'s model of the reservoir continuum (Figure 1a). Total nutrients (TN, TP) did not show a predictable decrease along the reservoir continuum in either reservoir, as was expected by Thornton et al. (1990)'s model (Figure 3a-b, d-e). Similarly, dissolved nutrients (NH_4 , NO_3 , SRP) did not show a consistent pattern of decreased concentrations along the continuum in either reservoir (Figure 3g-l). In FCR, we observed a marked increase in both NH_4 and NO_3 along the continuum (Figure 3j-k), contrary to expectations, whereas BVR's patterns were more inconsistent (Figure 3g-h). Lastly, we did not observe the expected peak in chl-a in the transitional zone within the reservoir, and concentrations remained relatively unchanged through the reservoir (Figure 3c, f), except in BVR in later months when concentrations increased at upstream reservoir sites (Figure 3c). Overall, FCR exhibited concentrations that were similar or higher than BVR for all biogeochemical variables. As a

result, we focused our comparisons of spatial and temporal variability among sites within a single reservoir, rather than comparing between FCR and BVR.

For TN and TP, both BVR and FCR showed inconsistent spatial variability over time, with peak concentrations occurring in different regions of the reservoir in different months (Figure 3a-b, d-e). Spatial variability, as estimated by differences in concentrations across sites, increased throughout the season with the largest range in concentrations across sites in BVR occurring in October for TN (258.5 $\mu\text{g/L}$) and in September for TP (11.6 $\mu\text{g/L}$). Similarly, the largest range in concentrations across sites in FCR occurred in September for TN (435.5 $\mu\text{g/L}$) and October for TP (14.8 $\mu\text{g/L}$). In both BVR and FCR, sampling days later in the year (August – October) tended to have higher ($>2\times$ in FCR) concentrations than earlier sampling days (April – July, Figure 3a-b, d-e).

Chl-a was relatively constant across space on all sampling days in FCR (Figure 3f; max. range = 5.7 $\mu\text{g/L}$ in August, min. range = 0.2 $\mu\text{g/L}$ in April), with relatively similar concentrations throughout the reservoir. Chl-a in BVR was similarly homogenous over space (Figure 3c; max. range = 7.6 $\mu\text{g/L}$ in October, min. range = 0.5 $\mu\text{g/L}$ in June), with the exception of September and October, when chl-a was higher in the riverine areas of the reservoir. Chl-a increased over the study period in both reservoirs, with the highest concentrations observed in September and October.

Dissolved N, P, and organic C were generally more variable over both space and time than total nutrients (Fig 3). Both NH_4 and NO_3 were variable across space in FCR (Figure 3j-k), showing peak concentrations of both variables in the most downstream location near the dam. NO_3 and NH_4 in BVR and SRP in both reservoirs varied both across space and time, but most observed values were below the limits of quantitation and therefore should be interpreted with

caution (Table S3). Finally, DOC was highly variable over both space and time in BVR (Figure 3m), with the largest range across sites occurring in July (2.4 mg/L), and the smallest range occurring in October (0.7 mg/L). While FCR did exhibit some spatial variability in DOC (Figure 3p), the trend was inconsistent over time.

Over our sampling sites and dates, autochthonous (Peak T) and allochthonous (Peak A) C varied more over time than space (Figure 3n-o, q-r), with relative fluorescent intensities increasing throughout the year. The one exception to this pattern was in August in BVR, when Peak T increased at the most downstream site B4 (range across sites = 0.10 RFU). Given that both autochthonous (Peak T) and allochthonous (Peak A) metrics showed very little change along the reservoir continuum, our data do not support a clear shift between autochthonous C and allochthonous C sources among sites in our reservoirs. Overall, there was greater fluorescent intensity for both Peaks A and T in FCR as compared to BVR across both time and space, reflective of higher DOC concentrations observed in FCR.

Nutrient stoichiometry was also highly variable over space and time (Figure S5). Molar ratios of TN:TP at reservoir sites in FCR and BVR were typically above Redfield ratios of 16:1 (Redfield 1934), indicating potential P limitation in both systems (mean ratio of 28.9 ± 14.6 , 1 SD). However, these ratios were variable over space, with no consistent increase or decrease observed along the continuum. Molar DIN:SRP was similarly variable over space and time, but showed a consistent increase along the reservoir continuum in FCR, on average from 1.3 at the most upstream site to 3.8 at the dam (Fig S5). This pattern did not occur in BVR, which ranged from a ratio of DIN:SRP of 2.0 and 3.7 at the most upstream sites (B1a and B1b, respectively) to 2.1 at the dam. Across FCR and BVR, mean molar DIN:TN ratios were 0.04 (± 0.02), indicating

that the particulate fraction of N dominated the total N pool. In contrast, SRP:TP ratios were 0.6 (± 0.45) in both reservoirs.

Ancillary environmental variables showed expected seasonal trends. Discharge at all four stream sites showed a clear decreasing trend over the course of the study period from April to October (Figure 4a). Specific conductance, which we used as a conservative tracer in our study, increased over time throughout the year at all study sites. In addition, specific conductance showed very little variation over space within the reservoir (Figure 4b), except for lower specific conductance (on average by 8.4 $\mu\text{S}/\text{cm}$) in reservoir sites than stream sites. Following Eq. 1, which used specific conductance as a conservative tracer throughout the reservoir, this difference in specific conductance had a negligible effect on $\Delta_{var}Site$, and may have resulted in under-estimation of biogeochemical processing at the stream-reservoir interface (Figure S7). Over our study period, water residence times were approximately 3 $\times$ longer in BVR than FCR, with a mean water residence time of 475 ± 301 (1 SE.) days for FCR and 1300 ± 335 days for BVR. The longest residence times occurred during October (Figure S1), which were 8 $\times$ longer in BVR and 73 $\times$ longer in FCR than the shortest residence times in April. Text S2 includes a description of meteorological variables (air temperature and precipitation, Figure S2), surface water temperature and dissolved oxygen (Figure S3), and thermal stratification metrics (thermocline depth and Schmidt stability, Figure S4) over the study period.

Heterogeneity across space and time

Across both reservoirs, heterogeneity over time (CV_{time}) was significantly higher than heterogeneity across space (CV_{space}) for TN, TP, Chl-a, SRP, Peak T, and Peak A (Figure 5a-c, f, h-i; SRP $p = 0.046$; all others $p < 0.0001$). For the remaining variables (NH_4 , NO_3 , and DOC;

Figure 5d-e, g), there was no significant difference in heterogeneity across space and time. FCR and BVR were significantly different from each other in heterogeneity of both space and time for most biogeochemical variables (Figure 5). FCR was significantly more temporally heterogeneous than BVR for several variables: TP, NH₄, NO₃, DOC, Peak T, and Peak A (Figure 5b, d-e, g-i). In contrast, BVR was significantly more temporally heterogeneous than FCR only for chl-a (mean CV_{time} BVR = 0.84; mean CV_{time} FCR = 0.75; Table S4). Over space, FCR was not significantly more heterogeneous than BVR for any variable, and BVR was more heterogeneous over space than FCR for chl-a only (mean CV_{space} BVR = 0.21; mean CV_{space} FCR = 0.1; Table S4, Figure 5c).

CV_{space} and CV_{time} were not significantly different from each other for the nutrient ratios, with the exception of SRP:TP, which was more heterogeneous over time than space (Figure S6, $p < 0.05$). Similarly, BVR and FCR were not significantly different in the magnitude of heterogeneity for any variables, except for CV_{space} for SRP:TP.

Spatially explicit processing of biogeochemical variables

When calculating spatially explicit processing between sites ($\Delta_{var}Site$), we found that the highest rates of processing typically occurred at the inflowing stream-reservoir interface, regardless of whether we used a processing metric standardized by area (Figure 6) or not (Figure S8). This pattern is exemplified by peak values of $\Delta_{var}Site$ in the regions between the stream sites and the most upstream reservoir sites in both reservoirs (i.e., between FS1/FS2 and F1; BS1 and B1a; BS2 and B1b). We found that the stream-reservoir interface disproportionately contributed to whole-reservoir dynamics, in which the magnitude and direction of the whole-ecosystem processing (as indicated by $\Delta_{var}WR$) matched that of $\Delta_{var}Site$ at these locations. For

example, the $\Delta_{var}Site$ data for DOC in BVR highlight that a majority of changes in DOC occurred between the inflowing stream sites and upstream areas of the reservoir, with negligible change occurring in the downstream areas of the reservoir. Consequently, the upstream areas of the reservoir were primarily responsible for the magnitude and directionality of $\Delta_{var}WR$, which shows that BVR was a source of DOC on most days (Figure 6m; e.g., between BS1 and B1a and BS2 and B1b, Figure 2b), despite some sites in the intermediate area of the reservoir (e.g., sites B2 and B3) functioning as a small sink of DOC on some days, especially in April. Within BVR, we found that while both arms of the reservoir typically had higher $\Delta_{var}Site$ values than sites further downstream, $\Delta_{var}Site$ between BS1 and B1a showed greater values than BS2 and B1b (e.g., Figure 6a), indicating that this metric was sensitive to the distance between the incoming stream site and the reservoir site (Figure 2, Table S5).

Large values of $\Delta_{var}Site$ at the stream-reservoir interface were often coincident with large values of $\Delta_{var}WR$, indicating that processing at the stream-reservoir interface disproportionately impacted whole-reservoir ($\Delta_{var}WR$) dynamics. This pattern was consistent for TN, TP, Chl-a, NH_4 , DOC, and Peaks T and A in BVR, and almost all biogeochemical variables in FCR, with the exception of DOC. The changes along the reservoir continuum in $\Delta_{var}Site$ for Peak T (a metric of autochthonous organic C) indicates an increase in autochthonous sources of C between the incoming stream and upstream reservoir sites for both reservoirs, with the exception of some months (July in BVR and June in FCR, Figure 6n, q). Interestingly, $\Delta_{var}Site$ for peak A (a metric of allochthonous C) followed a similar pattern, with an increase in allochthonous sources of C (peak A) between incoming stream sites and upstream reservoir sites, except for July in BVR, and June and August in FCR. Overall, patterns of DOC, Peak A, and Peak T were similar across days and sites in BVR.

Despite that the stream-reservoir interface often dominated whole-reservoir dynamics, there were some biogeochemical variables and sampling days for which sink-source dynamics changed throughout the reservoir, indicating that the reservoir functioned as both a sink and a source at different locations. These variables included TN in both reservoirs (Figure 6a, d), NO₃ and SRP in BVR (Figure 6h-i), and Peaks T and A in FCR (Figure 6q-r). For example, changes in $\Delta_{var}Site$ of Peak T were variable along the reservoir continuum in FCR, with some sampling days showing decreases from site to site in the intermediate region of the reservoir (e.g., site F2). Values of $\Delta_{var}Site$ for SRP in BVR were also spatially variable, in which the reservoir exhibited source dynamics at upstream sites, sink dynamics in the intermediate sites of the reservoir, and source dynamics again near the dam on a single sampling day (e.g., April, August, October). This variability over both space and time highlights that likely multiple interacting mechanisms were driving the processing of these biogeochemical variables along the reservoir continuum.

Further, the presence of large values of $\Delta_{var}Site$ at sites throughout the reservoir (e.g., NO₃ in BVR, which exhibited numerous fluctuations between sink and source dynamics, Figure 6h) were more common in BVR than in FCR, where large $\Delta_{var}Site$ values were primarily found in upstream areas of the reservoir. Variability in the location of peak values of $\Delta_{var}Site$ as well as increases in the magnitude of $\Delta_{var}WR$ were more common in FCR during April, when the discharge into the reservoir was >4× higher than any other sampling days. This result indicates that our metrics of processing were sensitive to large changes in hydrology and emphasizes the importance of time-discrete changes in the location of processing ‘hotspots.’

According to the whole-reservoir metric of processing, $\Delta_{var}WR$, BVR and FCR acted similarly as a sink of both NO₃ and TP at the whole-reservoir scale, and both reservoirs acted as a source of TN, DOC, and chl-a on most or all days (Figure 6). On all days, BVR was a sink for

NH₄, and on most days FCR was a sink or passive transporter (meaning no substantial change between the inflow and outflow) for NH₄, but acted as a source of NH₄ in April. Sink-source dynamics of SRP differed between FCR and BVR, with FCR primarily serving as a sink of SRP, whereas BVR switched from a sink to a source through time.

Discussion

Our study demonstrates that in two small reservoirs, multiple biogeochemical variables were more heterogeneous over time than space. While the two reservoirs did show substantial spatial variability in surface water biogeochemistry, these patterns were inconsistent over time, contrary to Thornton et al. (1990)'s heuristic model which describes distinct and consistent longitudinal zones along a reservoir continuum. Moreover, we observed “hotspots” of processing along the reservoir continuum where biogeochemical variables consistently and substantially changed, indicating that some locations disproportionately contributed to changes in biogeochemistry within a reservoir. For multiple variables we found that this hotspot location was frequently the interface between the stream and the reservoir, supporting other studies which posit that ecosystem boundaries or transitions often result in “hotspots” of biogeochemical processing (McClain et al. 2003, Bernhardt et al. 2017, Marcé et al. 2021). Ultimately, our analysis emphasizes the importance of examining spatial patterns and biogeochemical processing along the reservoir continuum.

Spatial and temporal variability in biogeochemistry along the reservoir continuum

BVR and FCR exhibited substantial variability in biogeochemistry across the reservoir continuum (space) and throughout the study period (time) (Figure 3). However, spatially, most

variables either showed inconsistent longitudinal variability over different sampling dates (e.g., NH_4 and DOC in BVR, SRP in FCR) or patterns contrary to previous expectations (e.g., NH_4 and NO_3 in FCR, chl-a in BVR). Ultimately, our results provide limited support for Thornton et al. (1990)'s model that biogeochemical variables change predictably along the reservoir continuum (following Figure 1). However, a major caveat of Thornton et al.'s (1990) model is the absence of an explicit time-dynamic component, which may be a key factor constraining the generality of their proposed trends. Our findings follow other studies that found inconsistent or divergent trends in reservoir spatial dynamics (Carneiro and Bini 2020, Borges et al. 2008, Varol et al. 2001, Woldeab et al. 2017, Scott et al. 2009, Soares et al. 2012), adding to a body of work examining longitudinal biogeochemistry in reservoirs.

A number of reservoir-specific characteristics may influence the strength of longitudinal gradients along the reservoir continuum, and help explain why we observed differences in biogeochemical distributions between BVR and FCR. Reservoir morphology, the number and location of tributaries, and reservoir size may all influence the strength of biogeochemical gradients and the heterogeneity within a reservoir (Carneiro and Bini 2020). While BVR and FCR are of similar age and have similar land-use within their watersheds (Gerling et al. 2016), they differ in several other important characteristics. FCR is a single-arm reservoir with two main upstream inflows. In contrast, BVR's morphology is more complex, with two main tributaries feeding separate arms that converge downstream within the reservoir, potentially leading to additional hotspots at the confluence of the distinct parcels of water. Indeed, for some variables, especially chl-a, SRP, and on some days DOC and Peaks T and A, we did see that the confluence at B3 showed larger rates of processing (Figure 6c, i, m-o), although the pattern did not occur consistently.

While we cannot attribute differences between FCR and BVR to reservoir size given their relatively small, comparable surface areas, there are several factors which may influence the relationship between size and biogeochemical heterogeneity in reservoirs. First, the pelagic region in larger reservoirs may be less heterogeneous due to increased flow through the reservoir, leading to a flushing effect as residence time decreases (Drastik et al. 2008, Carneiro and Bini 2020). However, larger reservoirs are likely to have a higher shoreline development index, leading to more microhabitats in nearshore areas and thereby increasing the localized influence of tributaries (e.g., DelSontro et al. 2011). Lastly, larger reservoirs may be more likely to be advection-dominated systems, which may cause substantial spatial heterogeneity (Gloss et al. 1980), as delivery and uptake of nutrients in riverine regions cause distinct patterns over space. However, the influence of advective flow on surface water heterogeneity is likely to vary throughout the season with changing river and reservoir water temperatures (Gloss et al. 1980). Altogether, given the prevalence of small waterbodies ($<1 \text{ km}^2$) like BVR and FCR globally (Downing et al. 2006), a better understanding of the biogeochemical heterogeneity of small reservoir ecosystems is critical to quantifying the overall role of reservoirs in global biogeochemical cycles.

Additionally, BVR is typically much more strongly stratified than FCR (Figure S4). This difference in thermal stratification strength is likely due to differences in volume, with FCR ~ 2 m shallower and 30% smaller in surface area (difference of 0.27 km^2) than BVR. Because BVR was more strongly stratified than FCR, internal nutrient loading from sediment fluxes may have been less likely to influence surface water concentrations in BVR than in FCR, and may have prevented increases in some nutrients downstream, such as we observed in FCR for NO_3 (Figure 3j). Overall, these differences in morphology may have led to our observations of increased

spatial heterogeneity in BVR as compared to FCR (as exemplified by greater spatial heterogeneity in NH_4 , NO_3 , DOC, and SRP in BVR than FCR, Figure 3), as BVR has generally more complex morphology.

Nutrient concentrations in FCR were consistently higher than BVR. We expected to see this pattern during the periods when surface water from BVR was delivered to FCR (April to June during this study) due to both increased connectivity, leading to higher nutrients in downstream ecosystems (Wurtsbaugh et al. 2005, Sadro et al. 2012; Brown et al. 2008; Stachelek and Soranno 2019), as well as lower spatial heterogeneity in FCR due to shorter water residence times in cascading reservoirs (Barbosa et al. 1999; Drastik et al. 2008). While we did observe shorter residence times during the first three sampling dates when BVR was connected to FCR, nutrient concentrations were higher later in the season when the reservoirs were not connected. This pattern likely indicates that seasonal changes in water temperature and oxygen dynamics leading to favorable conditions for phytoplankton growth or internal nutrient loading may be more important for driving nutrient concentrations in FCR than inflow connectivity. Moving forward, examining spatial and temporal distributions of biogeochemistry across a range of reservoir ages, land-uses, morphologies, tributaries, and landscape positions will help build upon this work and inform generalizable trends in reservoir biogeochemical distributions across space and time.

Implications of time as a dominant axis of heterogeneity

Despite observing substantial spatial variation in the biogeochemistry of both BVR and FCR, heterogeneity over time was significantly higher than over space for most biogeochemical variables (Figure 5). Seasonal succession has long been considered a critical driver of variability

in freshwater ecosystems (e.g., Sommer et al. 1986; De Senerpont Domis et al. 2013). However, our study adds an important novel comparison between the relative magnitude of spatial and temporal heterogeneity, a central question in the study of reservoirs (Thornton et al. 1990). Similar to our findings, many studies have documented that spatial heterogeneity of biogeochemistry is substantial, and also highly variable over time (Gloss et al. 1980, Shaughnessy et al. 2019, Borges et al. 2008, Soares et al. 2012, Stratton and Grant 2019, Linkhorst et al. 2020).

While we cannot directly parse the mechanisms driving temporal heterogeneity in our reservoirs, a number of changing seasonal dynamics could be connected. For example, Collins et al. (2019) found that both winter precipitation, which may influence delivery of organic matter and nutrients during spring runoff (Pierson et al. 2013; Brown et al. 2008), as well as summer temperature, which is directly related to biological activities such as primary production and mineralization, were able to predict lake water quality across ~11,000 U.S. lakes, highlighting the importance of temporal variation in climate variables. Another study found that seasonal changes in temperature, nutrient availability, and hydrology also directly influenced phytoplankton communities across a reservoir continuum (Lv et al. 2014). In our study, which was focused on the summer thermally-stratified period, we were unable to examine the effect of winter precipitation, but did observe that the later months with the warmest temperatures (July and August, Figure S3) had higher concentrations of chl-a, total nutrients, and both autochthonous and allochthonous organic C indicators (Figure 3). The highest concentrations of these variables tended to occur even later in the season (September and October) after water temperatures had begun to decline, indicating that other seasonally variable drivers such as increased residence time (Figure S1) may also play a role. For example, it may be that nutrient

availability peaked in the fall months due to degradation of phytoplankton biomass, while residence times decreased, leaving nutrients to remain longer within the reservoir.

We did observe that spatial heterogeneity was higher for some variables; specifically, NH_4 , NO_3 , and DOC showed a non-significant difference between heterogeneity in space and time (Figure 5). Thus, because neither time nor space was significantly more important in the heterogeneity of these variables, we infer that drivers which change over space may be relatively more important for influencing variability in dissolved nutrients (e.g., NH_4 , NO_3 , DOC, SRP) than for total nutrients, metrics of C quality, and chl-a, which showed significantly greater heterogeneity over time than space. Consequently, the drivers of dissolved nutrient dynamics (e.g., microbial communities, microhabitats with varying temperature or oxygen) may be more variable spatially than temporally. Overall, dissolved nutrient concentrations are generally considered more variable and harder to predict than total nutrients (Robson and Dourdet 2015; Soares and Calijuri 2021), which may be related to their increased spatial variability and rapid uptake times.

Lastly, nutrient stoichiometry, i.e., the ratios of molar N to P, as well as ratios of total to dissolved fractions of these nutrients, were highly variable over time and space (Figure S5). We did not observe any significant differences between CV_{space} and CV_{time} for any of the ratios except for SRP:TP, indicating an important role of drivers over both space and time in influencing variability in stoichiometry (Figure S6). While most of these ratios show highly variable patterns, DIN:SRP tended to increase along the reservoir continuum in FCR, with the highest values found nearest the dam. This was strongly driven by increases in dissolved N (Figure 3j-k, NH_4 and NO_3) which were observed in FCR and is evidence for increased P limitation in the downstream regions of the reservoir. Despite that some individual nutrient

constituents did not exhibit consistent change spatially within our reservoirs (Figure 3), the relative availability of N and P did change, with potentially important implications for controlling phytoplankton community composition (Watson et al. 1997; Downing et al. 2001). For example, greater P-limitation downstream in FCR could result in the promotion of non-N-fixing phytoplankton upstream and N-fixing cyanobacteria downstream (Vanni et al. 2011; Hamre et al. 2018).

Importance of biogeochemical processing “hotspots”

Examining spatially explicit processing suggests that along the reservoir continuum, the stream-reservoir interface can be disproportionately important for whole-reservoir biogeochemical processing. Our spatially explicit metric of processing, $\Delta_{var}Site$, highlights that the stream-reservoir interface most consistently acts as a “hotspot” of biogeochemical processing, as shown in the largest values of $\Delta_{var}Site$ at the farthest upstream location in the reservoir for most variables (Figure 6). This finding is in line with previous conceptual frameworks (McClain et al. 2003, Bernhardt et al. 2017), field measurements (Marcé et al. 2021, Sadro et al. 2011, Ward, et al. 2022), and laboratory experiments (Lambert and Perga 2019), which also suggest that ecosystem boundaries, particularly those between stream and lake ecosystems, serve as ecosystem hotspots. This result is likely driven by multiple mechanisms. First, inflowing streams provide subsidies of nutrients and C which can be processed and transformed by organisms either via primary production or decomposition (Wetzel 2001). Second, physical processes may also play an important role. For example, if incoming stream water is much colder than reservoir surface water, the parcel of stream water may plunge to the bottom of the reservoir, resulting in delivery of organic matter and nutrients deeper within the

water column or in the sediments (Thornton et al. 1990), where they would not be measurable via our sampling method in surface waters. In FCR and BVR, it is likely that both of these mechanisms play a role in driving the importance of the stream-reservoir interface in biogeochemical processing, as the inflowing streams in FCR and BVR were 3.9°C and 4.7°C colder, respectively than the average reservoir site water temperature (Text S2, Figure S3). Altogether, our data suggest that the stream-reservoir interface is a key location of biogeochemical processing, which can inform reservoir field monitoring programs to better understand how reservoirs function as sinks or sources of biogeochemical variables.

Interestingly, we did find some exceptions to the stream-reservoir interface dominating the location of processing “hotspots.” These exceptions (i.e., large $\Delta_{var}Site$ values) occurred inconsistently at different sites throughout the reservoir. However, they occurred more often in BVR than in FCR, and in FCR during our April sampling, when discharge was highest. These results suggest that some reservoirs may be more spatially variable than others in locations of hotspots and that changes in stream inflow volume, or temporary increases in the relative importance of smaller, unmonitored tributaries along the continuum, can influence processing rates. In larger, more morphologically variable reservoirs (e.g., with multiple inflow streams along the reservoir, higher shoreline development indices), biogeochemical processing may occur less consistently in the furthest upstream stream-reservoir interface due to additional inflow streams entering the reservoir downstream or greater influence of overland runoff. Similarly, increased inflow volumes are directly connected to shorter water residence times, which may limit processing upstream and instead favor processing occurring further along the reservoir continuum where water flows have slowed (i.e., in the transitional and lacustrine regions, as we observed in April in FCR).

Comparing values of $\Delta_{var}Site$ to the traditional method of estimating whole-reservoir function, $\Delta_{var}WR$, suggests limitations of traditional whole-ecosystem mass balance approaches which tend to focus on sources into and out of the reservoir. Because most monitoring programs sample at stream sites and the deepest lacustrine sites only, current methods would provide biased estimates of reservoir functioning according to our results. For example, the estimate of whole-reservoir function, $\Delta_{var}WR$, shows that BVR was a sink of NO_3 on all sampling days. However, the reservoir was both a sink and a source of NO_3 at different locations within the reservoir on a single sampling day according to $\Delta_{var}Site$ (Figure 6). We see a similar pattern for most variables in BVR (i.e., a mismatch between $\Delta_{var}WR$ and individual values of $\Delta_{var}Site$ along the continuum), indicating that complex dynamics occur along the reservoir continuum which are ignored by measuring only one reservoir site. Identifying and monitoring reservoir biogeochemical hotspots is critical to improving our ability to predict variability in biogeochemical processing, enhancing our understanding of global biogeochemical cycles, and informing water quality management.

Strengths and limitations of study design and directions for future research

Our study provides a novel examination of spatial and temporal variability of biogeochemical concentrations and processing in reservoirs, motivating future research. First, both of our study reservoirs are relatively small and shallow. While reservoirs of similar size (less than 1 km²) make up ~99% of reservoirs globally (Lehner et al. 2011), larger reservoirs may have more complex morphology and as a result exhibit different patterns and drivers of spatial variability. Second, BVR is mesotrophic (mean TP = 15.5 µg/L, mean chl-a = 6.35 µg/L) and FCR is eutrophic (FCR mean TP = 28.7 µg/L, mean chl-a = 8.95 µg/L; following Carlson

and Simpson 1996), with algal blooms occurring regularly in FCR (Carey et al. 2022d). Research in a hyper-eutrophic lake of similar size shows that the presence of algal blooms can dramatically increase spatial heterogeneity as the blooms move from one location to another (Ortiz and Wilkinson 2021). Studies conducted in additional systems, especially those which are oligotrophic, may illuminate different patterns in spatial heterogeneity and highlight differences among systems of diverse trophic states.

Third, additional spatial data may help increase the utility of our biogeochemical processing metric, $\Delta_{var}Site$. For example, by estimating the location of the stream plunge-point (i.e., where the incoming stream water plunges into the reservoir), Marcé et al. (2021) were able to specifically pinpoint locations of processing and associated changes in dissolved organic matter composition. Because we sampled at discrete, set locations within each reservoir region irrespective of stream plunge-points, we were unable to identify the exact location where processing occurred, but rather regions within which high rates of processing likely occurred.

Fourth, because water residence time is dependent on depth and bathymetry, it is likely to change within a reservoir along the longitudinal continuum. Calculating reservoir site-specific residence times, as opposed to using the outflow of the entire reservoir, could improve our spatial processing metric. Our Eq. 1, which was based on using specific conductance to account for changing hydrology, likely resulted in more conservative estimates of $\Delta_{var}Site$ processing than what actually occurred because of small differences in specific conductance between the inflowing stream and most upstream reservoir sites. Higher values of specific conductance at stream sites resulted in lower calculations of incoming nutrient and C loads for the stream component of Eq. 1, thereby resulting in smaller values of $\Delta_{var}Site$. Standardizing specific conductance at stream sites to be equal to the observed specific conductance at the first within-

reservoir site (Figure S7) indicates that our Eq. 1 calculations likely underestimated values of $\Delta_{var}Site$ at the stream-reservoir interface, increasing our confidence in this region as a major site of biogeochemical processing.

Finally, collecting samples of biogeochemical variables throughout the water column, as opposed to just the surface, may also help inform some of the patterns we observed. For example, by sampling throughout the water column, we may be able to parse additional sources of nutrients within the reservoir, e.g., via sediment release or interflow. Altogether, a more detailed approach to sampling nutrients and C within reservoir continua will help inform the fate of biogeochemical variables and advance our understanding of how these variables change across ecosystems.

Conclusions

Our study informs future research on spatial and temporal heterogeneity across lentic ecosystems. While we focus on reservoirs which primarily release surface waters downstream, our results likely hold for reservoirs with selective depth water withdrawal, as well as naturally-formed lakes with surface water outflows. Given that these waterbodies are typically only sampled at their deepest point, we are likely missing important spatial variability influencing whole-ecosystem biogeochemical dynamics. Our work points to the need for more data examining the role of the stream-reservoir interface as “hotspots” of biogeochemical processing, as well as the need for understanding when and why these hotspots dominate whole-ecosystem dynamics. Finally, our study provides support for temporal sampling which occurs over multiple months in order to capture the full variability in reservoir functioning. Ultimately, a better

quantification of when and where biogeochemical variables are processed within reservoirs will advance our understanding of the important role reservoirs play in global biogeochemical cycles.

Acknowledgements

We thank the Western Virginia Water Authority for their long-term access to field sites and continuing support and engagement in this research. We thank many field and lab assistants, especially Rose Thai, James Maze, Jacob Wynne, Ryan McClure, Dexter Howard, Nicholas Hammond, Kait Farrell, and Bethany Bookout for their critical help in field and lab work, as well as B. Niederlehner for her tireless efforts in sample analysis. We thank the Virginia Tech Stream Team, especially Stephen Plont and Erin Hotchkiss, the Carey Lab and Reservoir Group, and the Global Lake Ecological Observatory Network for their support and feedback on this study. This work was supported by the National Science Foundation (DEB-1753639, DBI-1933016, DBI-1933102, DEB-1926050, DGE-1651272), the Fralin Life Sciences Institute, and the Western Virginia Water Authority.

Author contribution statement: WWM, AGH, and CCC co-conceived the design of the study. WMW and AGH led field sampling, with assistance from ASL, HLW, and MEL. WMW, AGH, and HLW analyzed water chemistry samples, with analytical support from DS. ASL contributed to figure design, and CCC, MEL, and DS contributed to conceptual development, which substantially improved the quality of the manuscript. WMW led the writing of the manuscript with CCC, AGH, and MEL, and all authors contributed to manuscript editing.

Availability of data and material: All data used in this study are published and available within the Environmental Data Initiative repository (Hounshell et al. 2021, Carey et al. 2020, Carey et al. 2021, Carey et al. 2022b, Carey et al. 2022c, Carey et al. 2022d, Carey et al. 2022e, Carey et

al. 2022f). All code used to download data, generate data specific to this study, and conduct analyses are available in a Zenodo repository (Woelmer, 2023).

References

- Alboukadel, K (2020) ggpubr: 'ggplot2' based publication ready plots. R package version 0.4.0.
<https://CRAN.R-project.org/package=ggpubr>
- ArcGIS Pro (Version 2.2.0) Esri Inc. 2021. <https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview>.
- Baker M, Webster J (2017) Conservative and reactive solute dynamics. In: Lamberti GA and Richard Hauer F (eds) *Methods in stream ecology Volume 2: Ecosystem Function*, 3rd edn. Academic Press, London, pp 129-145
- Barbosa FAR, Padisák J, Espindola ELG, et al (1999) The Cascading Reservoir Continuum Concept (CRCC) and its application to the river Tietê-basin, São Paulo State, Brazil. Sao Pedro, Brazil, pp 425–439
- Berberich ME, Beaulieu JJ, Hamilton TL, et al. (2020) Spatial variability of sediment methane production and methanogen communities within a eutrophic reservoir: Importance of organic matter source and quantity. *Limnol Oceanogr*, 65(6), 1336–1358.
<https://doi.org/10.1002/lno.11392>
- Bernhardt ES, Blaszczak JR, Ficken CD, et al (2017) Control points in ecosystems: moving beyond the hot spot hot moment concept. *Ecosystems* 20:665–682.
<https://doi.org/10.1007/s10021-016-0103-y>
- Borges PAF, Train S, Rodrigues LC (2008) Spatial and temporal variation of phytoplankton in two subtropical Brazilian reservoirs. *Hydrobiologia* 607:63–74.
<https://doi.org/10.1007/s10750-008-9367-3>
- Boström, B., Andersen, J.M., Fleischer, S., Jansson, M. (1988). Exchange of phosphorus across the sediment-water interface. In: Persson, G., Jansson, M. (eds) *Phosphorus in freshwater*

ecosystems. *Developments in hydrobiology*, vol 48. Springer, Dordrecht.
https://doi.org/10.1007/978-94-009-3109-1_14

Brett MT, Benjamin MM (2008) A review and reassessment of lake phosphorus retention and the nutrient loading concept. *Freshw Biology* 53:194–211. <https://doi.org/10.1111/j.1365-2427.2007.01862.x>

Brown PD, Wurtsbaugh WA, Nydick KR, et al (2008) Lakes and forests as determinants of downstream nutrient concentrations in small mountain watersheds. *Arct, Antarct, and Alp Res*, 40(3):462-469. [https://doi.org/10.1657/1523-0430\(07-052\)\[BROWN\]2.0.CO;2](https://doi.org/10.1657/1523-0430(07-052)[BROWN]2.0.CO;2)

Carey CC, Lofton ME, Woelmer WM, et al (2020) Filtered chlorophyll a time series for Beaverdam Reservoir, Carvins Cove Reservoir, Claytor Lake, Falling Creek Reservoir, Gatewood Reservoir, Smith Mountain Lake, and Spring Hollow Reservoir in southwestern Virginia, USA during 2014-2019 ver 1. Environmental Data Initiative. <https://doi.org/10.6073/pasta/4103be6062b768867e1f3b016665a35c>

Carey CC, Wander HL, Woelmer WM et al (2021) Water chemistry time series for Beaverdam Reservoir, Carvins Cove Reservoir, Falling Creek Reservoir, Gatewood Reservoir, and Spring Hollow Reservoir in southwestern Virginia, USA 2013-2020 ver 9. Environmental Data Initiative. <https://doi.org/10.6073/pasta/6343e979a970e8a2590b4a450e851dd2>

Carey CC, Hanson PC, Thomas RQ, et al (2022a) Anoxia decreases the magnitude of the carbon, nitrogen, and phosphorus sink in freshwaters. *Glob Change Biol*. <https://onlinelibrary.wiley.com/doi/10.1111/gcb.16228>

Carey CC, Wander HL, McClure RP, et al (2022b) Secchi depth data and discrete depth profiles of photosynthetically active radiation, temperature, dissolved oxygen, and pH for Beaverdam Reservoir, Carvins Cove Reservoir, Falling Creek Reservoir, Gatewood

Reservoir, and Spring Hollow Reservoir in southwestern Virginia, USA 2013-2021 ver 10. Environmental Data Initiative.
<https://doi.org/10.6073/pasta/887d8ab8c57fb8fdf3582507f3223cd6>

Carey CC, Breef-Pilz A, Howard DW, et al (2022c) Manually-collected discharge data for multiple inflow tributaries entering Falling Creek Reservoir, Beaverdam Reservoir, and Carvins Cove Reservoir, Vinton and Roanoke, Virginia, USA from 2019-2021 ver 5. Environmental Data Initiative.
<https://doi.org/10.6073/pasta/9992f38175e38a03ce5aa794eefec857>.

Carey CC, Lewis AS, McClure RP, et al (2022d) Time series of high-frequency profiles of depth, temperature, dissolved oxygen, conductivity, specific conductance, 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 2013-2021 ver 12. Environmental Data Initiative.
<https://doi.org/10.6073/pasta/c4c45b5b10b4cb4cd4b5e613c3effbd0>.

Carey CC, Breef-Pilz A, Woelmer WM, et al (2022e) 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 2018-2021 ver 6. Environmental Data Initiative.
<https://doi.org/10.6073/pasta/81c6c76f4fe22434a20aa8c00f2d4ad1>.

68 Carey CC, Breef-Pilz A, and Bookout BJ (2022f) Time series of high-frequency meteorological
69 data at Falling Creek Reservoir, Virginia, USA 2015-2021 ver 6. Environmental Data
70 Initiative. <https://doi.org/10.6073/pasta/35d8d3f9390408f12d39e44e3f03abbe>.

71 Carlson RE, Simpson J (1996) A coordinator's guide to volunteer lake monitoring methods. *N*
72 *Am Lake Manag Soc* 96:305

73 Carneiro FM, Bini LM (2020) Revisiting the concept of longitudinal gradients in reservoirs. *Acta*
74 *Limnologica Brasiliensia* 32. <https://doi.org/10.1590/s2179-975x1319>

75 Casas-Ruiz JP, Catalán N, Gómez-Gener L, et al (2017) A tale of pipes and reactors: Controls on
76 the in-stream dynamics of dissolved organic matter in rivers. *Limnol Oceanogr* 62:S85–
77 S94. <https://doi.org/10.1002/lno.10471>

78 Clow DW, Stackpoole SM, Verdin KL, et al (2015) Organic carbon burial in lakes and reservoirs
79 of the conterminous United States. *Environ Sci Technol* 49:7614–7622.
80 <https://doi.org/10.1021/acs.est.5b00373>

81 Collins SM, Yuan S, Tan PN, et al (2019) Winter precipitation and summer temperature predict
82 lake water quality at macroscales. *Water Resour Res* 55:2708–2721.
83 <https://doi.org/10.1029/2018WR023088>

84 de Eyto E, Jennings E, Ryder E, et al (2016) Response of a humic lake ecosystem to an extreme
85 precipitation event: physical, chemical, and biological implications. *Inland Waters* 6:483–
86 498. <https://doi.org/10.1080/IW-6.4.875>

87 Delontro T, Kunz MJ, Kempter, et al. (2011) Spatial heterogeneity of methane ebullition in a
88 large tropical reservoir. *Envir Sci Tech*, 45(23), 9866–9873.
89 <https://doi.org/10.1021/es2005545>

De Senerpont Domis LN, Elser JJ, Gsell AS, et al (2013) Plankton dynamics under different climatic conditions in space and time. *Freshw Biology* 58:463–482.
<https://doi.org/10.1111/fwb.12053>

Downing JA, Watson SB, McCauley E (2001) Predicting cyanobacteria dominance in lakes. *Can J Fish Aquat Sci* 58:1905–1908. <https://doi.org/10.1139/f01-143>

Downing JA, Prairie YT, Cole JJ, Duarte CM, Tranvik LJ, Striegl RG, McDowell WH, Kortelainen P, Caraco NF, Melack JM, Middelburg JJ (2006), The global abundance and size distribution of lakes, ponds, and impoundments, *Limnol Oceanogr*, 51, doi: 10.4319/lo.2006.51.5.2388.

Drastik V, Kubecka J, Tuser M, et al (2008) The effect of hydropower on fish stocks: comparison between cascade and non-cascade reservoirs. *Hydrobiologia* 609:25
<https://link.springer.com/article/10.1007/s10750-008-9393-1>.

Fellman JB, Hood E, Spencer RGM (2010) Fluorescence spectroscopy opens new windows into dissolved organic matter dynamics in freshwater ecosystems: A review. *Limnol Oceanogr* 55:2452–2462. <https://doi.org/10.4319/lo.2010.55.6.2452>

Furey PC, Nordin RN, Mazumder A (2004) Water level drawdown affects physical and biogeochemical properties of littoral sediments of a reservoir and a natural lake. *Lake Reserv Manage*, 20(4), 280–295. <https://doi.org/10.1080/07438140409354158>

Gabor R, Baker A, McKnight DM, Miller M (2014) Fluorescence indices and their interpretation, p. 303–338. In P. Coble, J. Lead, A. Baker, D. Reynolds, and R. G. M. Spencer [eds.], *Aquatic organic matter fluorescence*. Cambridge Univ. Press
<https://doi.org/10.1017/CBO9781139045452.015>

112 Gerling AB, Browne RG, Gantzer, et al. (2014) First report of the successful operation of a side
 113 stream supersaturation hypolimnetic oxygenation system in a eutrophic , shallow
 114 reservoir. *Water Res*, 67, 129–143. <https://doi.org/10.1016/j.watres.2014.09.002>
 115 Gerling AB, Munger ZW, Doubek JP, et al (2016) Whole-catchment manipulations of internal
 116 and external loading reveal the sensitivity of a century-old reservoir to hypoxia.
 117 *Ecosystems* 19:555–571. <https://doi.org/10.1007/s10021-015-9951-0>
 118 Gloss SP, Mayer LM, Kidd DE (1980) Advective control of nutrient dynamics in the epilimnion
 119 of a large reservoir. *Limnol Oceanogr*, 25(2), 219–228.
 120 <https://doi.org/10.4319/lo.1980.25.2.0219>
 121 Guillemette F, McCallister SL, del Giorgio PA (2013) Differentiating the degradation dynamics
 122 of algal and terrestrial carbon within complex natural dissolved organic carbon in
 123 temperate lakes. *J Geophys Res: Biogeosci* 118:963–973.
 124 <https://doi.org/10.1002/jgrg.20077>
 125 Hamre KD, Lofton ME, McClure RP, et al (2018) In situ fluorometry reveals a persistent,
 126 perennial hypolimnetic cyanobacterial bloom in a seasonally anoxic reservoir. *Freshw Sci*
 127 37:000–000. <https://doi.org/10.1086/699327>
 128 Harrison JA, Maranger RJ, Alexander RB, et al (2009) The regional and global significance of
 129 nitrogen removal in lakes and reservoirs. *Biogeochemistry* 93:143–157.
 130 <https://doi.org/10.1007/s10533-008-9272-x>
 131 Hayes NM, Deemer BR, Corman JR, et al (2017) Key differences between lakes and reservoirs
 132 modify climate signals: A case for a new conceptual model. *Limnol Oceanogr Lett* 2:47–
 133 62. <https://doi.org/10.1002/lol2.10036>

134 Hessen DO, Elser JJ, Sterner RW, Urabe J (2013) Ecological stoichiometry: An elementary
 135 approach using basic principles. *Limnol Oceanogr* 58:2219–2236.
 136 <https://doi.org/10.4319/lo.2013.58.6.2219>

137 Hounshell AG, Thai RH, Peeler KA, et al (2021) Time series of optical measurements
 138 (absorbance, fluorescence) for Beaverdam and Falling Creek Reservoir in Southwestern
 139 Virginia, USA 2019-2020 ver 1. Environmental Data Initiative.
 140 <https://doi.org/10.6073/pasta/d1062b14ed1df86507414afe8d45dc75>

141 Hotchkiss ER, Sadro S, Hanson PC (2018) Toward a more integrative perspective on carbon
 142 metabolism across lentic and lotic inland waters. *Limnol Oceanogr Lett* 3:57–63.
 143 <https://doi.org/10.1002/lo.2.10081>

144 Jane SF, Rose KC (2018) Carbon quality regulates the temperature dependence of aquatic
 145 ecosystem respiration. *Freshw Biology*. <https://doi.org/10.1111/fwb.13168>

146 Jennings E, Jones S, Arvola L, et al (2012) Effects of weather-related episodic events in lakes: an
 147 analysis based on high-frequency data. *Freshw Biology* 57:589–601.
 148 <https://doi.org/10.1111/j.1365-2427.2011.02729.x>

149 Kasprzak P, Shatwell T, Gessner MO, et al (2017) Extreme weather event triggers cascade
 150 towards extreme turbidity in a clear-water lake. *Ecosystems* 20:1407–1420.
 151 <https://doi.org/10.1007/s10021-017-0121-4>

152 Keller PS, Marcé R, Obrador B, & Koschorreck M (2021) Global carbon budget of reservoirs is
 153 overturned by the quantification of drawdown areas. *Nat Geosci*, 14(6), 402–408.
 154 <https://doi.org/10.1038/s41561-021-00734-z>

155 Keys TA, Caudill MF, Scott DT (2019) Storm effects on nitrogen flux and longitudinal
 156 variability in a river – reservoir system. *River Res Appl*: 1–10.
 157 <https://doi.org/10.1002/rra.3433>

158 Klausmeier CA, Litchman E, Levin SA (2004) Phytoplankton growth and stoichiometry under
 159 multiple nutrient limitation. *Limnol Oceanogr* 49:1463–1470.
 160 https://doi.org/10.4319/lo.2004.49.4_part_2.1463

161 Klippel G, Macêdo RL, Branco CWC (2020) Comparison of different trophic state indices
 162 applied to tropical reservoirs. *Lakes and Reserv: Res Manag* 25:214–229.
 163 <https://doi.org/10.1111/lre.12320>

164 Lambert T, Perga ME (2019) Non-conservative patterns of dissolved organic matter degradation
 165 when and where lake water mixes. *Aquat Sci* 81:1–13. [https://doi.org/10.1007/s00027-](https://doi.org/10.1007/s00027-019-0662-z)
 166 [019-0662-z](https://doi.org/10.1007/s00027-019-0662-z)

167 Lehner B, Liermann CR, Revenga C, et al (2011) High-resolution mapping of the world's
 168 reservoirs and dams for sustainable river-flow management. *Frontiers Ecology Environ*
 169 9:494–502. <https://doi.org/10.1890/100125>

170 Linkhorst A, Hiller C, DelSontro T, et al. (2020) Comparing methane ebullition variability
 171 across space and time in a Brazilian reservoir. *Limnol Oceanogr*, 65(7), 1623–1634.
 172 <https://doi.org/10.1002/lno.11410>

173 Linkhorst A, Paranaíba JR, Mendonça, et al. (2021) Spatially resolved measurements in tropical
 174 reservoirs reveal elevated methane ebullition at river inflows and at high productivity.
 175 *Global Biogeochem Cy*, 35(5), 1–16. <https://doi.org/10.1029/2020GB006717>

Lv H, Yang J, Liu L, et al (2014) Temperature and nutrients are significant drivers of seasonal
 shift in phytoplankton community from a drinking water reservoir, subtropical China.
 Environ Sci Pollut Res 21:5917–5928. <https://doi.org/10.1007/s11356-014-2534-3>

Maavara T, Chen Q, Van Meter K, et al (2020) River dam impacts on biogeochemical cycling.
 Nat Rev Earth Environ 1:103–116. <https://doi.org/10.1038/s43017-019-0019-0>

Maavara T, Parsons CT, Ridenour C, et al (2015) Global phosphorus retention by river
 damming. Proc National Academy Sci United States of America 112:15603–15608.
<https://doi.org/10.1073/pnas.1511797112>

Maranger R, Jones SE, Cotner JB (2018) Stoichiometry of carbon, nitrogen, and phosphorus
 through the freshwater pipe. Limnol Oceanogr Lett 3:89–101.
<https://doi.org/10.1002/lol2.10080>

Marcé R, Verdura L, Leung N (2021) Dissolved organic matter spectroscopy reveals a hot spot
 of organic matter changes at the river–reservoir boundary. Aquat Sci 83:.
<https://doi.org/10.1007/s00027-021-00823-6>

Matsuzaki SS, Kohzu A, Tsuchiya K, Shinohara R, Nakagawa M, Fukumori K, Yamaguchi H,
 Kondo NI, Kadoya T (2022). Water-level drawdowns can improve surface water quality
 and alleviate bottom hypoxia in shallow, eutrophic water bodies. Freshwater Biol,
 (October 2022), 229–244. <https://doi.org/10.1111/fwb.14020>

McClain ME, Boyer EW, Dent CL, et al (2003) Biogeochemical hot spots and hot moments at
 the interface of terrestrial and aquatic ecosystems. Ecosystems 301–312.
<https://doi.org/10.1007/s10021-003-0161-9>

McClure RP, Lofton ME, Chen S, Krueger KM, Little JC, & Carey CC (2020). The magnitude
 and drivers of methane ebullition and diffusion vary on a longitudinal gradient in a small

199 freshwater reservoir. *J Geophys Res-Bioge* 125, e2019JG005205.
 200 <https://doi.org/10.1029/2019JG005205>

201 Nilsson C, Reidy CA, Dynesius M, Revenga C (2005) Fragmentation and flow regulation of the
 202 world's large river systems. *Science* 308:405–408.
 203 <https://doi.org/10.1126/science.1107887>

204 Nürnberg GK (1988) Prediction of phosphorus release rates from total and reductant-soluble
 205 phosphorus in anoxic lake sediments. *Can J Fish Aquat Sci* 45:453–462.
 206 <https://doi.org/10.1139/f88-054>

207 Ortiz DA, Wilkinson GM (2021) Capturing the spatial variability of algal bloom development in
 208 a shallow temperate lake. *Freshw Biology* 66:2064–2075.
 209 <https://doi.org/10.1111/fwb.13814>

210 Pierson DC, Samal NR, Owens EM, et al (2013) Changes in the timing of snowmelt and the
 211 seasonality of nutrient loading: can models simulate the impacts on freshwater trophic
 212 status? *Hydrol Process* 27:3083–3093. <https://doi.org/10.1002/hyp.9894>

213 Powers SM, Tank JL, Robertson, DM (2015) Control of nitrogen and phosphorus transport by
 214 reservoirs in agricultural landscapes. *Biogeochemistry* 124:417-439.
 215 <https://doi.org/10.1007/s10533-015-0106-3>

216 Preece EP, Moore BC, Skinner, et al. (2019) A review of the biological and chemical effects of
 217 hypolimnetic oxygenation. *Lake Reserv Manage*, 35(3), 229–246.
 218 <https://doi.org/10.1080/10402381.2019.1580325>

219 Redfield AC. 1934. On the proportions of organic derivatives in sea water and their relation to
 220 the composition of plankton. *Liverpool: James Johnstone Memorial Volume*. pp 176–92.

221 Robson BJ, Dourdet V (2015) Prediction of sediment, particulate nutrient and dissolved nutrient
 222 concentrations in a dry tropical river to provide input to a mechanistic coastal water
 223 quality model. *Environ Model Softw* 63:97–108.
 224 <https://doi.org/10.1016/j.envsoft.2014.08.009>
 225 Rychtecky P, Znachor P (2011) Spatial heterogeneity and seasonal succession of phytoplankton
 226 along the longitudinal gradient in a eutrophic reservoir. *Hydrobiologia* 663:175–186.
 227 <https://doi.org/10.1007/s10750-010-0571-6>
 228 Rydin E (2000) Potentially mobile phosphorus in Lake Erken sediment. *Water Res* 34:2037–
 229 2042. [https://doi.org/10.1016/S0043-1354\(99\)00375-9](https://doi.org/10.1016/S0043-1354(99)00375-9)
 230 Sadro S, Melack JM, MacIntyre S (2011) Spatial and temporal variability in the ecosystem
 231 metabolism of a high-elevation lake: integrating benthic and pelagic habitats. *Ecosystems*
 232 14:1123–1140. <https://doi.org/10.1007/s10021-011-9471-5>
 233 Sadro S, Nelson CE, Melack JM (2012) The influence of landscape position and catchment
 234 characteristics on aquatic biogeochemistry in high-elevation lake-chains. *Ecosystems*
 235 15:363–386. <https://doi.org/10.1007/s10021-011-9515-x>
 236 Schindler DW, Curtis PJ, Parker BR, Stainton MP (1996) Consequences of climate warming and
 237 lake acidification for UV-B penetration in North American boreal lakes. *Nat* 379:705–
 238 708. <https://doi.org/10.1038/379705a0>
 239 Scott JT, Stanley JK, Doyle RD, et al (2009) River – reservoir transition zones are nitrogen
 240 fixation hot spots regardless of ecosystem trophic state. *Hydrobiologia* 625:61–68.
 241 <https://doi.org/10.1007/s10750-008-9696-2>

242 Shaughnessy AR, Sloan JJ, Corcoran MJ, Hasenmueller EA (2019) Sediments in agricultural
 243 reservoirs act as sinks and sources for nutrients over various timescales. *Water Resour*
 244 *Res* 2018WR024004. <https://doi.org/10.1029/2018WR024004>
 245 Soares LMV, Calijuri M do C (2021) Deterministic modelling of freshwater lakes and reservoirs:
 246 Current trends and recent progress. *Environ Model Softw* 144:105143.
 247 <https://doi.org/10.1016/j.envsoft.2021.105143>
 248 Soares MCS, Marinho MM, Azevedo SMOF, et al (2012) Eutrophication and retention time
 249 affecting spatial heterogeneity in a tropical reservoir. *Limnologica* 42:197–203.
 250 <https://doi.org/10.1016/j.limno.2011.11.002>
 251 Solomon CT, Jones SE, Weidel BC, et al (2015) Ecosystem consequences of changing inputs of
 252 terrestrial dissolved organic matter to lakes: current knowledge and future challenges.
 253 *Ecosystems* 18:376–389. <https://doi.org/10.1007/s10021-015-9848-y>
 254 Sommer U, Maciej Gliwicz Z, Lampert W, Duncan A (1986) The PEG-model of seasonal
 255 succession of planktonic events in fresh waters. *Arch für Hydrobiologie* 106:433–471
 256 Stachelek J, Soranno PA (2019) Does freshwater connectivity influence phosphorus retention in
 257 lakes ? *Limnol Oceanogr* 64:1586-1599. <https://doi.org/10.1002/lno.11137>
 258 Sterner, R., and J. J. Elser (2002) *Ecological stoichiometry: the biology of elements from*
 259 *molecules to the biosphere*. Princeton University Press, Princeton, NJ.
 260 Stratton LE, Haggerty R, and Grant GE (2019) The importance of coarse organic matter and
 261 depositional environment to carbon burial behind dams in mountainous environments. *J*
 262 *Geophys Res-Earth*, 124(8), 2118–2140. <https://doi.org/10.1029/2019JF005005>

263 Strauss EA, Lamberti GA (2002) Effect of dissolved organic carbon quality on microbial
 264 decomposition and nitrification rates in stream sediments. *Freshw Biology* 47:65–74.
 265 <https://doi.org/10.1046/j.1365-2427.2002.00776.x>

266 Syvitski JPM, Vorosmarty CJ, Kettner AJ, Green P (2005) Impact of humans on the flux of
 267 terrestrial sediment to the global coastal ocean. *Science* 308:376–381

268 Thornton KW, Kimmel BL, Payne FE (1990) *Reservoir Limnology--Ecological Perspectives*.
 269 John Wiley and Sons, New York

270 Torres I, Resck R, Pinto-Coelho R (2007) Mass balance estimation of nitrogen, carbon,
 271 phosphorus and total suspended solids in the urban eutrophic, Pampulha Reservoir,
 272 Brazil. *Acta Limnologica Brasiliensia* 19:79–91

273 Tranvik LJ, Downing JA, Cotner JB, et al (2009) Lakes and reservoirs as regulators of carbon
 274 cycling and climate. *Limnol Oceanogr* 54:2298–2314.
 275 https://doi.org/10.4319/lo.2009.54.6_part_2.2298

276 Triska FJ, Kennedy VC, Avanzino RJ, et al (1989) Retention and transport of nutrients in a third-
 277 order stream in northwestern California: hyporheic processes. *Ecology* 70(6): 1893-1905.
 278 <https://esajournals.onlinelibrary.wiley.com/doi/abs/10.2307/1938120>.

279 Vanni MJ, Renwick WH, Bowling AM, et al (2011) Nutrient stoichiometry of linked catchment-
 280 lake systems along a gradient of land use. *Freshw Biology* 56:791–811.
 281 <https://doi.org/10.1111/j.1365-2427.2010.02436.x>

282 Varol M, Gökot B, Bekleyen A, Şen B (2012) Spatial and temporal variations in surface water
 283 quality of the dam reservoirs in the Tigris River basin, Turkey. *Catena* 92:11–21.
 284 <https://doi.org/10.1016/j.catena.2011.11.013>

285 Vollenweider RA (1975) Input-output models - with special reference to the phosphorus loading
 286 concept in limnology. *Schweizerische Z für Hydrologie* 37:53–84.
 287 <https://doi.org/10.1007/BF02505178>
 288 Waldo S, Beaulieu JJ, Barnett W., et al. (2021) Temporal trends in methane emissions from a
 289 small eutrophic reservoir: The key role of a spring burst. *Biogeosciences*, 18(19), 5291–
 290 5311. <https://doi.org/10.5194/bg-18-5291-2021>
 291 Wang F, Maberly SC, Wang B, Liang X (2018) Effects of dams on riverine biogeochemical
 292 cycling and ecology. *Inland Waters* 8:130–140.
 293 <https://doi.org/10.1080/20442041.2018.1469335>
 294 Ward NK, Brentrup JA, Richardson DC, Weathers KC, Hanson PC, Hewett RJ, & Carey CC
 295 (2022). Dynamics of the stream – lake transitional zone affect littoral lake metabolism.
 296 *Aquat Sci*, 84(3), 1–19. <https://doi.org/10.1007/s00027-022-00854-7>
 297 Watson SB, McCauley E, Downing JA (1997) Patterns in phytoplankton taxonomic composition
 298 across temperate lakes of differing nutrient status. *Limnol Oceanogr* 42:487–495.
 299 <https://doi.org/10.4319/lo.1997.42.3.0487>
 300 Weathers KC, Strayer DL, Likens GE (2021) Fundamentals of ecosystem science. Academic
 301 Press, San Diego, USA
 302 Wetzel RG (2001) Limnology, 3rd edn. Academic Press, San Diego, USA
 303 Williamson TJ, Vanni MJ, Renwick WH (2021) Spatial and temporal variability of nutrient
 304 dynamics and ecosystem metabolism in a hyper-eutrophic reservoir differ between a wet
 305 and dry year. *Ecosystems* 24:68–88. <https://doi.org/10.1007/s10021-020-00505-8>
 306 Winslow L, Read J, Woolway R, et al (2019) rLakeAnalyzer: Lake physics tools. R package
 307 version 1.11.4.1. <https://CRAN.R-project.org/package=rLakeAnalyzer>

308 Woelmer WM (2023) wwoelmer/ReservoirContinuum: release following revisions to MS at
 309 Aquatic Sciences (AquaticSciences_MS_Feb23_v2.0). Zenodo.
 310 <https://doi.org/10.5281/zenodo.7646740>
 311 Woldeab B, Beyene A, Ambelu A, et al (2018) Seasonal and spatial variation of reservoir water
 312 quality in the southwest of Ethiopia. Environ Monit Assess 190:1
 313 <https://doi.org/10.1007/s10661-018-6527-4>
 314 Wurtsbaugh WA, Baker MA, Gross HP and Brown PD (2005) Lakes as nutrient “sources” for
 315 watersheds: a landscape analysis of the temporal flux of nitrogen through sub-alpine
 316 lakes and streams. Internationale Ver für theoretische und angewandte Limnologie:
 317 Verhandlungen, 29:2, 645-649, DOI: 10.1080/03680770.2005.11902758

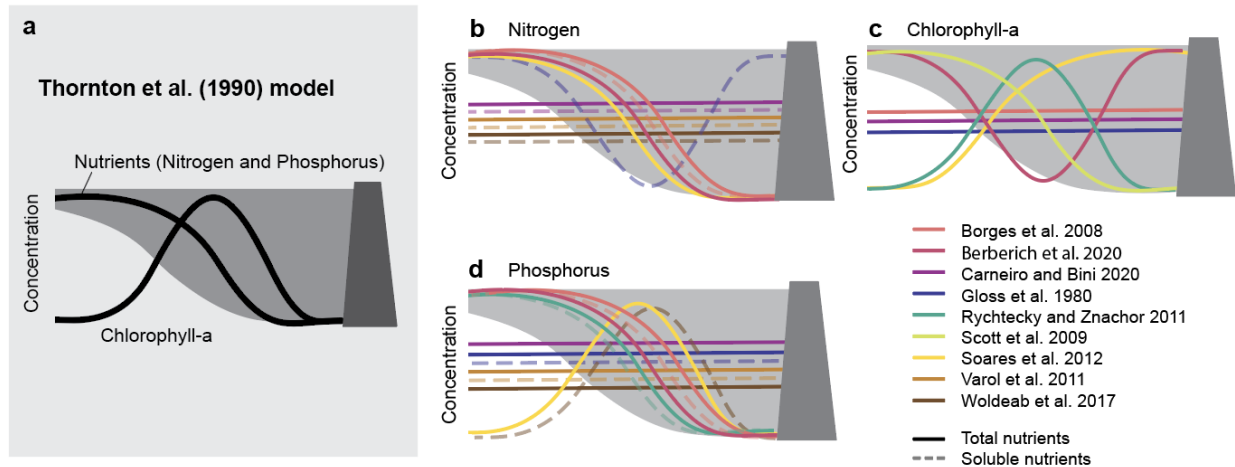


Figure 1. Conceptual diagram showing a) Thornton et al.'s (1990) model of the spatial distribution of chlorophyll-a and nutrients (N and P) along the reservoir continuum, as well as selected field studies which have examined the distribution of b) nitrogen (N), c) chlorophyll-a (chl-a), and d) phosphorus (P) in reservoirs across the globe. While Thornton et al. (1990) did not distinguish between N and P, or total and dissolved nutrient fractions, we delineate these pools here when possible (total nutrients in solid lines, dissolved nutrients in dashed lines in panels b and d). We note that patterns of reservoir longitudinal distributions are meant to be qualitative and only represent relative minimums and maximums. Other than Marcé et al. (2021), we were unable to find analyses of the distributions of allochthonous and autochthonous carbon across the riverine, transitional, and lacustrine reservoir zones in the literature to match Thornton et al. (1990)'s model

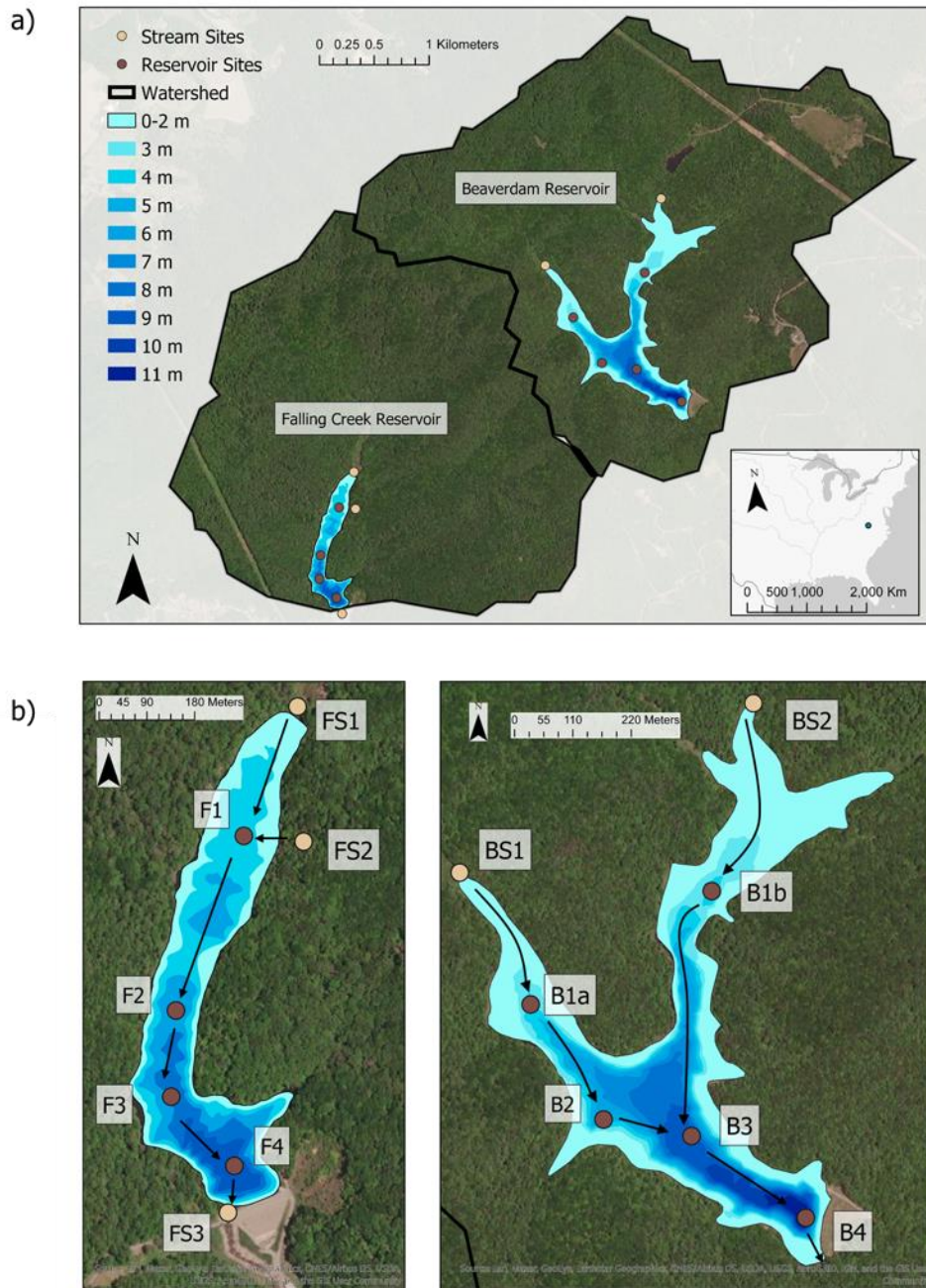


Figure 2. Map of Falling Creek Reservoir (FCR) and Beaverdam Reservoir (BVR) with a) watershed area, and locations of stream and reservoir sampling sites and b) names of sampling sites on both reservoirs' bathymetry. Arrows in panel B indicate the approximate flow of water through the stream-reservoir continuum



Figure 3. Concentrations of biogeochemical variables along the reservoir continuum starting from the most upstream reservoir site of Beaverdam Reservoir (left panels) and Falling Creek Reservoir (right panels). Each panel represents a different biogeochemical variable, and colors represent the month of sampling, from 4 (April) to 10 (October). All units are denoted in each panel header. TN = total nitrogen, TP = total phosphorus, chl-a = chlorophyll-a, NH_4 = ammonium, NO_3 = nitrate, SRP = soluble reactive phosphorus, DOC = dissolved organic carbon, T-autoch = Peak T autochthonous carbon quality, A-alloch = Peak A allochthonous carbon quality. Because BVR has two upstream sites (one within each arm), these sites are plotted according to their distance from the corresponding stream site, meaning the first point on the x-axis is B1a and the second point on the x-axis is B1b.

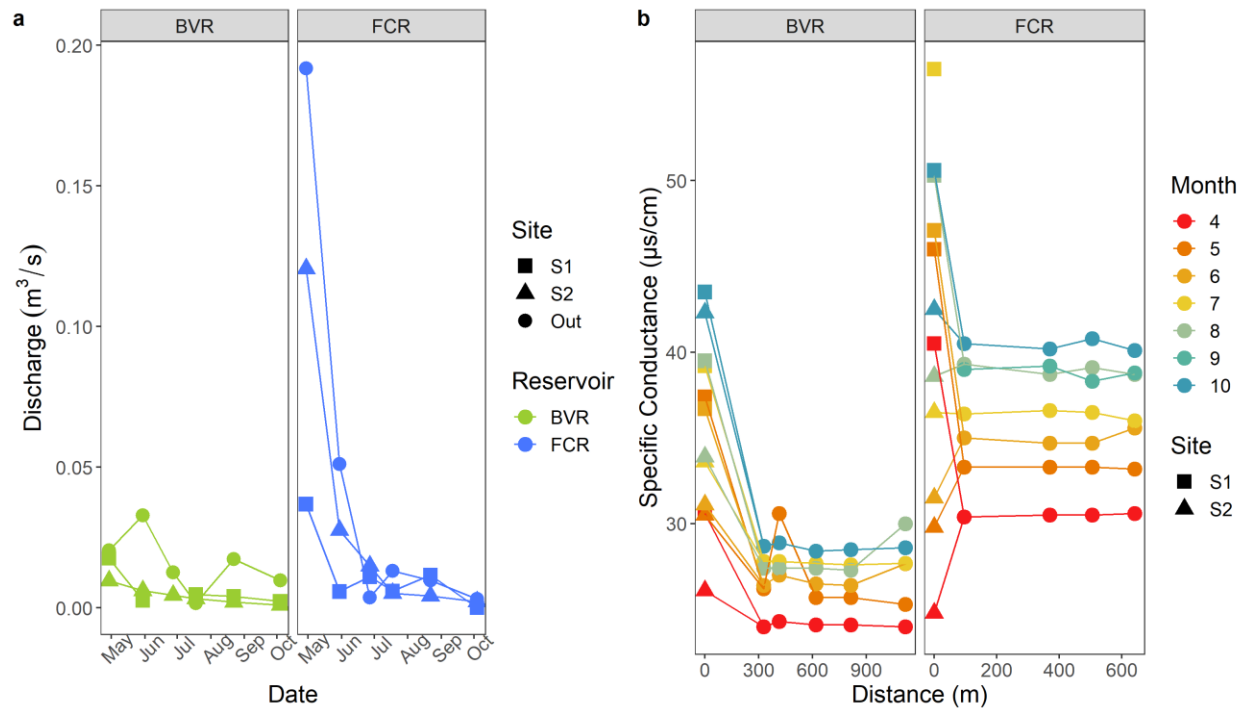


Figure 4. a) Discharge at all stream (inflow and outflow) sites on all sampling dates, with Falling Creek Reservoir (FCR) sites represented in blue and Beaverdam Reservoir (BVR) sites in green. Site S1 refers to sites FS1 or BS1, while Site S2 refers to sites FS2 or BS2, with reservoir designated by color and panel. Sites designated as ‘Out’ represent the outflow dynamics for each reservoir and were estimated at FS3 at FCR and B4 in BVR. b) Specific conductance at all sites and for all sampling dates, with colors representing the month of sampling, from 4 (April) to 10 (October). Stream sites are represented by triangles or squares, while sites within the reservoir are represented by circles. Specific conductance was not collected at BVR sites in September

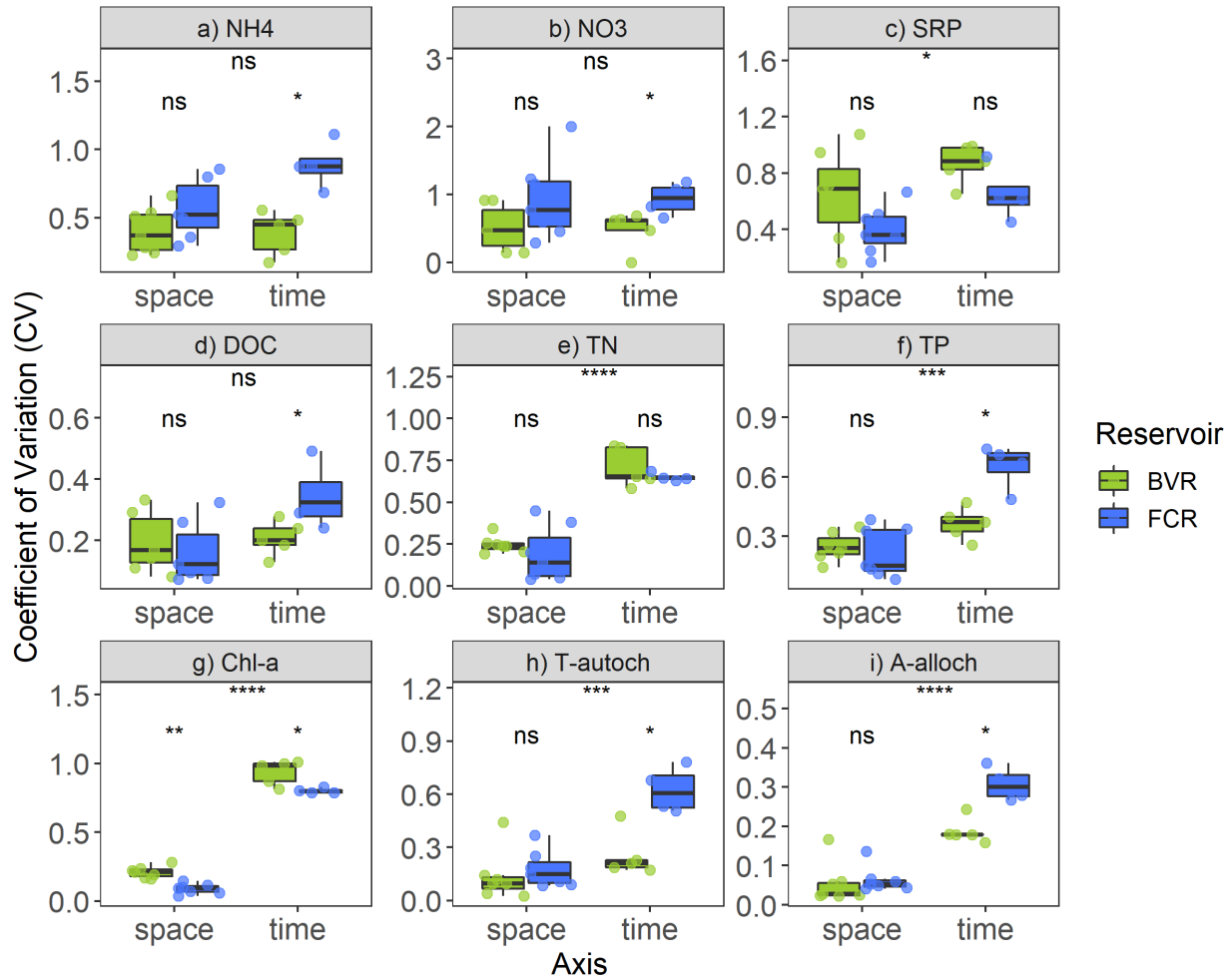
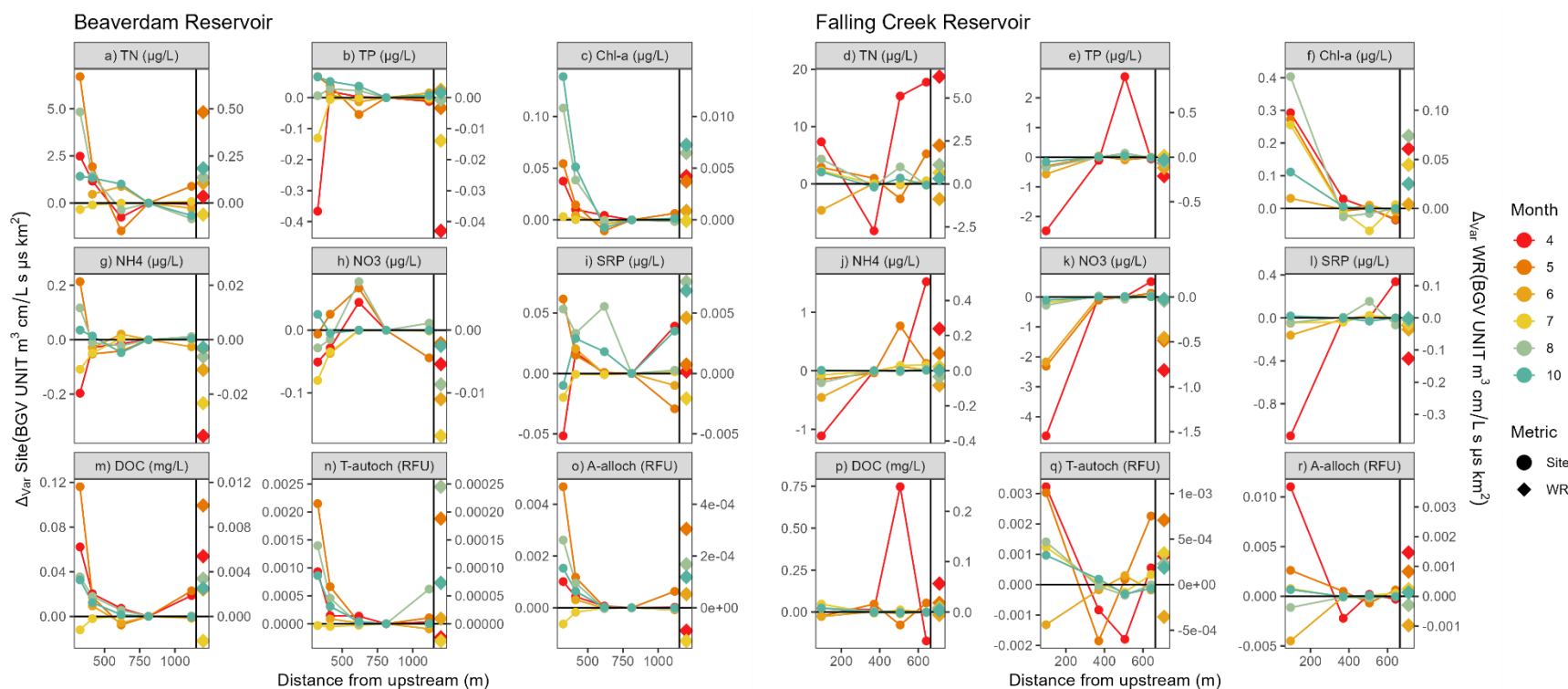


Figure 5. Heterogeneity, as measured by coefficient of variation, over space (CV_{space}) and time (CV_{time}) for BVR and FCR for total nutrients (TN and TP), chl-a, dissolved nutrients (NH₄, NO₃, SRP), dissolved organic carbon quantity (DOC), and dissolved organic carbon quality (Peak T, autochthonous, and Peak A, allochthonous). Bars above pairing represent the pairing of groups for significance testing. ‘ns’ denotes a non-significant relationship ($p > 0.05$), a single asterisk ‘*’ represents $p < 0.05$, three asterisks ‘***’ represents $p < 0.0001$, and four asterisks ‘****’ represents $p < 0.00001$.



363

364 Figure 6. The change in biogeochemical variables, Δ_{var} , along the reservoir continuum corrected by specific conductance and
 365 standardized by the area of the reservoir between sampling sites with each variable represented in individual panels. $\Delta_{var}Site$, which
 366 represents the change between sites, is shown at each sampling site along the reservoir continuum and $\Delta_{var}WR$, which represents the
 367 change from inflow stream sites and the furthest downstream site, is shown in diamond-shaped points at the furthest downstream site
 368 on the x-axis to the right of the vertical line. Note that the secondary (right) y-axis corresponds to $\Delta_{var}WR$ (large diamond points at
 369 far right of each panel) and presents the same units as $\Delta_{var}Site$ but on a smaller scale to show the range of variability in these values

370 (which are of a smaller magnitude than $\Delta_{var}Site$ due to being standardized by the area of the entire reservoir). See *Methods: Spatially*
371 *explicit processing of biogeochemical variables* for a full description of $\Delta_{var}Site$ and $\Delta_{var}WR$ (Eq. 1). Because BVR has two
372 upstream sites (one within each arm), these are plotted according to their distance from the corresponding stream site, meaning the
373 first point on the x-axis is B1a and the second point on the x-axis is B1b. Color represents the month in which the sample was
374 collected, from 4 (April) to 10 (October). ‘BGV UNIT’ stands for the individual units of each biogeochemical variable, which are
375 listed in each panel label. Samples were not collected at stream sites in September and are excluded from this analysis

Supporting Information

Chapter II: The importance of time and space in biogeochemical heterogeneity and processing along the reservoir ecosystem continuum*

Whitney M. Woelmer¹, Alexandria G. Hounshell^{1,2}, Mary E. Lofton¹, Heather L. Wander¹,
Abigail S.L. Lewis¹, Durelle Scott³, Cayelan C. Carey¹

¹Department of Biological Sciences, Virginia Tech, 926 West Campus Drive, Blacksburg, Virginia 24061, USA.

² Current Address: National Center for Coastal Ocean Science, National Oceanic and Atmospheric Administration, Beaufort, NC, 28516 USA

³Department of Biological Systems Engineering, Virginia Tech, Blacksburg, VA, 26060 USA

Contents of this file:

Texts: 4

Tables: 5

Figures: 8

References

*Published in *Aquatic Sciences* with the citation:

Woelmer, W. M., Hounshell, A. G., Lofton, M. E., Wander, H. L., Lewis, A. S., Scott, D., & Carey, C. C. (2023). The importance of time and space in biogeochemical heterogeneity and processing along the reservoir ecosystem continuum. *Aquatic Sciences*, 85(2), 66.

Text S1. Calculation of sub-watershed area for Beaverdam Reservoir

Because Beaverdam Reservoir (BVR) has several inflow streams, we estimated the area of each sub-watershed as well as the total watershed (3.65 km²) using the U.S. Geological Survey's StreamStats online interactive tool (<https://streamstats.usgs.gov/ss/>). Briefly, the StreamStats tool allows you to select a “pour point,” or the location at which you want to estimate the entire area of contributing land within the watershed, and delineates the entire watershed contributing to the pour point. For our field survey, we chose the two sub-watersheds which contribute the largest percentage of the entire watershed, which also represent both the two major east and west arms of the reservoir (B100 and B200; shaded in gray in Table S1). GPS coordinates of all pour points for all sub-watersheds are listed below in Table S1.

Text S2. Spatial and temporal variability of environmental variables.

Air temperature showed typical seasonal trends and several intermittent rainfall events occurred during the study period, none of which occurred immediately before our sampling dates (Figure S2). Surface water temperature showed a strong seasonal trend, with coolest observations during our April sampling, followed by a general warming throughout the summer, and cooler temperatures again in September and October (Figure S3). Dissolved oxygen (DO) % saturation exhibited a similar seasonal trend as water temperature in FCR, but was more variable over time in BVR.

Over space, there was a clear trend in BVR and FCR of higher temperatures and DO at in-reservoir sites than at stream sites. However, within both reservoirs, temperature and DO remained relatively constant as distance from stream inflows increased. Thermocline depth in FCR was shallower than in BVR by ~0.5-3.0 m and generally increased (i.e., deeper thermocline) as the year progressed for both reservoirs (Figure S4). Lastly, Schmidt stability, a metric of the strength of thermal stratification, was much higher in BVR than in FCR, especially in August-October (Figure S4).

Text S3. Thornthwaite-Mather Water Balance methods

We adapted the Thornthwaite-Mather Water Balance (TMWB) model to estimate daily inflow in BVR over the study period (Thornthwaite and Mather, 1957). As input to the TMWB model, we used daily precipitation and air temperature data from the meteorological station located at FCR, as well as soil characteristics for the BVR watershed from the NRCS-USDA soil survey dataset (Soil Survey Staff, NRCS-USDA 2021).

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

$$R = P - ExcessOut - Et \text{ (Eq. S1)}$$

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.

Following daily inflow calculations, we estimated outflow using observations of water level at the dam site using a mass balance approach which took into account changing water levels and thus, reservoir volume, which was adjusted approximately weekly when water level observations were available and coupled with a bathymetry map created for BVR (following Munger et al. 2019).

Text S4. Unstandardized calculation of spatially explicit processing metric

To assess the robustness of our spatially explicit mass balance approach and allow for more direct comparison to other studies which do not standardize by ecosystem area, we also calculated Δ_{var} using an unstandardized approach following Equation S1:

$$\Delta_{var} = \frac{[var]_d * Q_d}{sp\ cond_d} - \frac{[var]_u * Q_u}{sp\ cond_u} \quad \text{Eq. S1}$$

where $[var]_d$ is the concentration of a given variable at the downstream reservoir site in units of $\mu\text{g/L}$, mg/L , or RFU (relative fluorescence units) depending on the analyte (see Table S3); Q_d is the discharge at the downstream site (m^3/s), $sp\ cond_d$ is the specific conductance at the downstream reservoir site ($\mu\text{S}/\text{Cm}$), included here to account for unobserved intrusions of solutes into the reservoir which may occur between sites; $[var]_u$, Q_u , and $sp\ cond_u$ are similarly defined but at the upstream site. Q_u was calculated for each stream site on each sampling day.

Table S1. Table of GPS coordinates, sub-watershed names, and corresponding area and percent contribution for each sub-watershed within Beaverdam Reservoir's overall watershed. The shaded rows denote the sub-watersheds included in this study.

Watershed descriptor	GPS Coordinate of pour point	Area (km²)	Percent of watershed
Whole watershed	37.31321, -79.81535	3.65	NA
BS1 (west arm)	37.31951, -79.82434	0.73	20
BS1.5 (east arm, farthest west inflow)	37.32210, -79.81866	0.44	12
BS2 (east arm, central inflow)	37.32262, -79.81735	0.62	17
BS2.5 (east arm, farthest east inflow)	37.32140, -79.81417	0.62	17
BS0.75 (small southwest inflow)	37.31390, -79.82205	0.15	4
Total Percent of Watershed Estimated:			70

Table S2. GPS coordinates for all sites in this study. Refer to Figure 2 in the main text for a map of sites.

Site ID	Latitude	Longitude
BS1	37.31957	-79.8244
BS2	37.32285	-79.8172
B1a	37.31701	-79.8226
B1b	37.31921	-79.8182
B2	37.31478	-79.8209
B3	37.31447	-79.8187
B4	37.31288	-79.8159
FS1	37.30943	-79.8362
FS2	37.30761	-79.8361
F1	37.30769	-79.8371
F2	37.30534	-79.8383
F3	37.30418	-79.8383
F4	37.30325	-79.8373
FS3	37.30247	-79.8369

Table S3. Analytical chemistry limits of quantitation for total and dissolved nitrogen, phosphorus, and carbon.

Variable	Precision limit	Method	Units
NH ₄	11.8	Limit of quantitation	µg/L
NO ₃	9.3	Limit of quantitation	µg/L
SRP	6.2	Limit of quantitation	µg/L
DOC	0.411	Limit of quantitation	mg/L
TN	67.2	Limit of quantitation	µg/L
TP	13.2	Limit of quantitation	µg/L

Table S4. Summary statistics (minimum, maximum, mean) of coefficients of variation (CV) for each reservoir and variable across space and time. (chl-a = chlorophyll-a, SRP = soluble reactive phosphorus, NO₃ = nitrate, NH₄ = ammonium, TP = total phosphorus, TN = total nitrogen, A = allochthonous carbon, T = autochthonous carbon, DOC = dissolved organic carbon)

Reservoir	Variable	Min CV _{time}	Max CV _{time}	Mean CV _{time}	Min CV _{space}	Max CV _{space}	Mean CV _{space}
BVR	chl-a	0.76	0.9	0.84	0.16	0.28	0.21
FCR	chl-a	0.71	0.79	0.75	0.04	0.14	0.1
BVR	SRP	0.64	0.88	0.79	0.17	1.08	0.63
FCR	SRP	0.42	0.97	0.65	0.17	0.67	0.4
BVR	NO ₃	0	0.62	0.44	0.14	0.91	0.5
FCR	NO ₃	0.6	1.19	0.89	0.29	2	0.91
BVR	NH ₄	0.19	0.51	0.39	0.22	0.66	0.4
FCR	NH ₄	0.69	1.11	0.87	0.28	0.86	0.57
BVR	TP	0.55	0.78	0.66	0.19	0.34	0.24
FCR	TP	0.58	0.64	0.6	0.04	0.45	0.19
BVR	TN	0.24	0.44	0.34	0.14	0.35	0.25
FCR	TN	0.45	0.68	0.61	0.08	0.38	0.21
BVR	A	0.16	0.23	0.18	0.02	0.17	0.05
FCR	A	0.26	0.36	0.3	0.04	0.11	0.06
BVR	T	0.17	0.44	0.24	0.03	0.44	0.14
FCR	T	0.48	0.78	0.59	0.08	0.37	0.17

BVR	DOC	0.14	0.26	0.2	0.08	0.33	0.2
FCR	DOC	0.23	0.45	0.33	0.07	0.32	0.16

Table S5. Areas of each reservoir section used to calculate the spatially explicit processing metric, Δ_{var} Site using Eq. 1. Area between sampling points was calculated using the Calculate Geometry tool in ArcGIS Pro (2021).

Reservoir	Order within Reservoir	Area (km ²)	Description
BVR	1	0.034	Area between BS1 and B1a
BVR	2	0.103	Area between B2 and B1b
BVR	3	0.057	Area between B1a and B2
BVR	4a	0.048	Area between B2 and B3
BVR	4b	0.091	Area between B1b and B3
BVR	5	0.087	Area between B3 and B4
FCR	1	0.025	Area Between FS1, Fs2 and F1
FCR	2	0.042	Area between F1 and F2
FCR	3	0.016	Area between F2 and F3
FCR	4	0.037	Area between F3 and F4

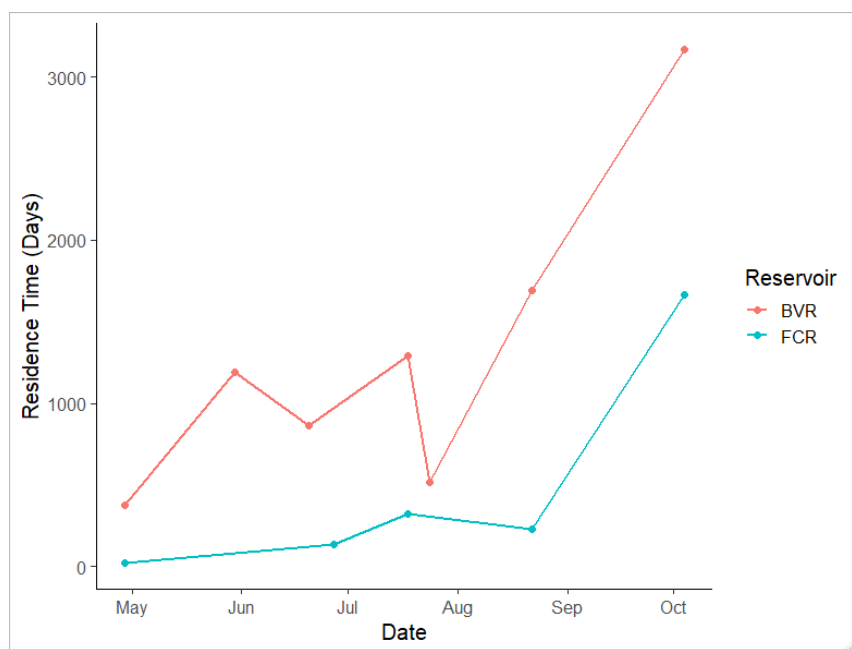


Figure S1. Residence time (days) for Beaverdam Reservoir (BVR) and Falling Creek Reservoir (FCR) calculated for each sampling day (shown as points). Residence times were calculated by dividing the reservoir volume by the incoming flow to the reservoir for a given

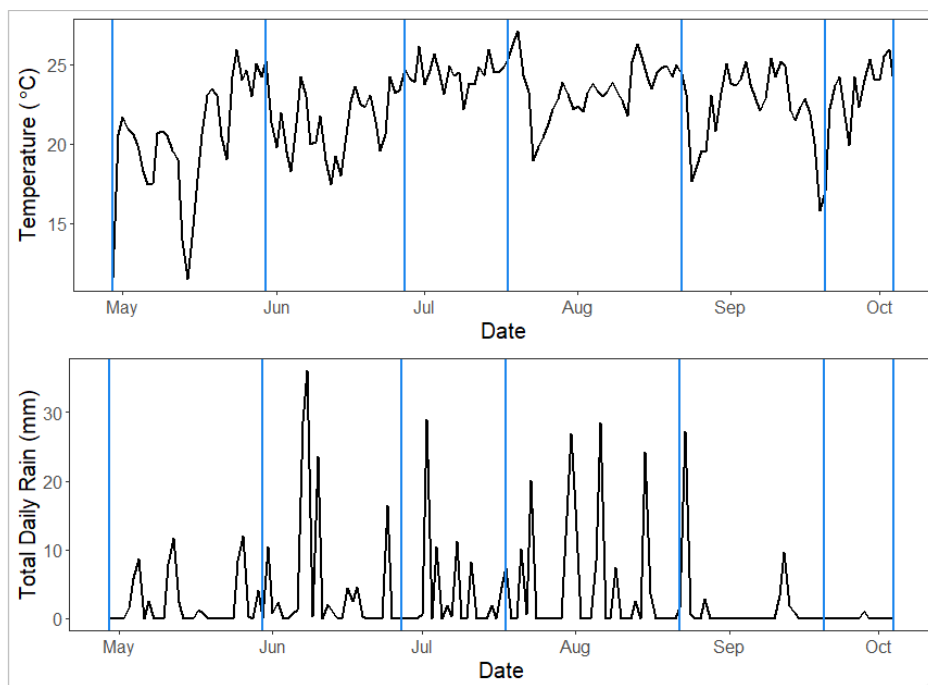


Figure S2. Daily average air temperature and cumulative daily rainfall over the entire study period, measured at the dam of Falling Creek Reservoir. Sampling dates are indicated by the vertical blue lines.

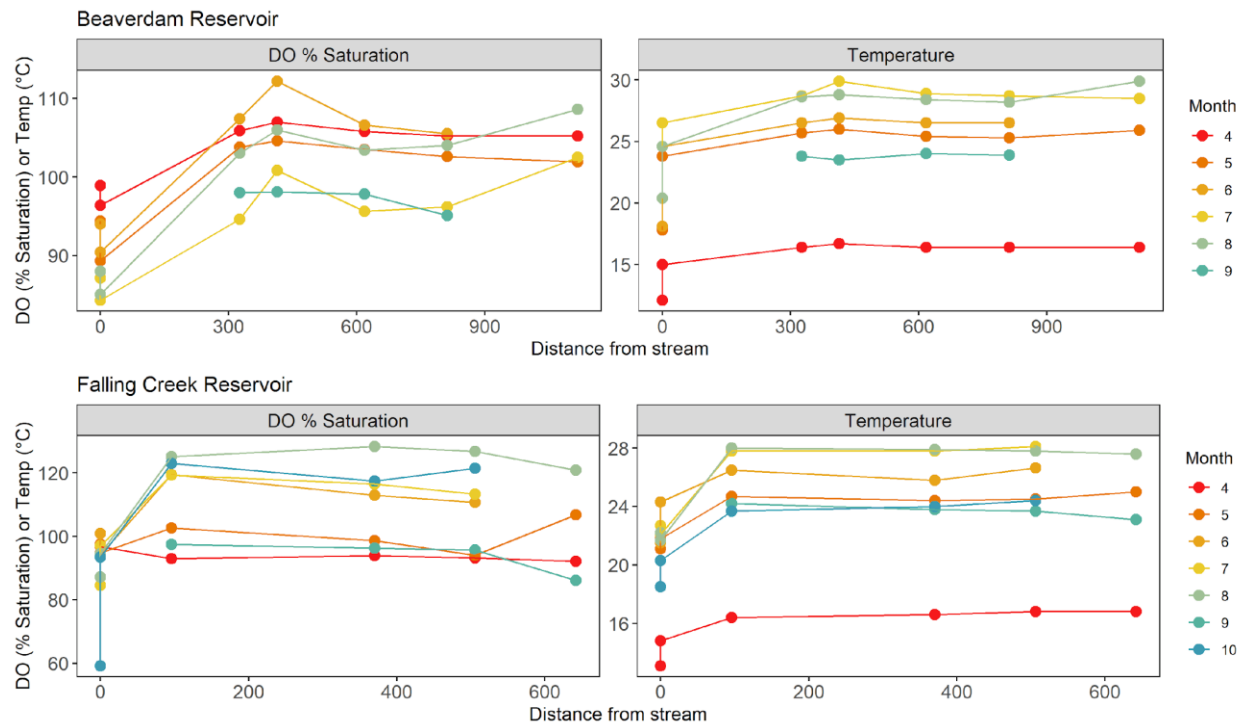


Figure S3. Surface water temperature and surface dissolved oxygen (DO) percent saturation (%) at all study sites on all sampling dates. In September, we did not visit all sites at Beaverdam Reservoir so data from that reservoir's inflow sites (BS1 and BS2), as well as B4, are absent.

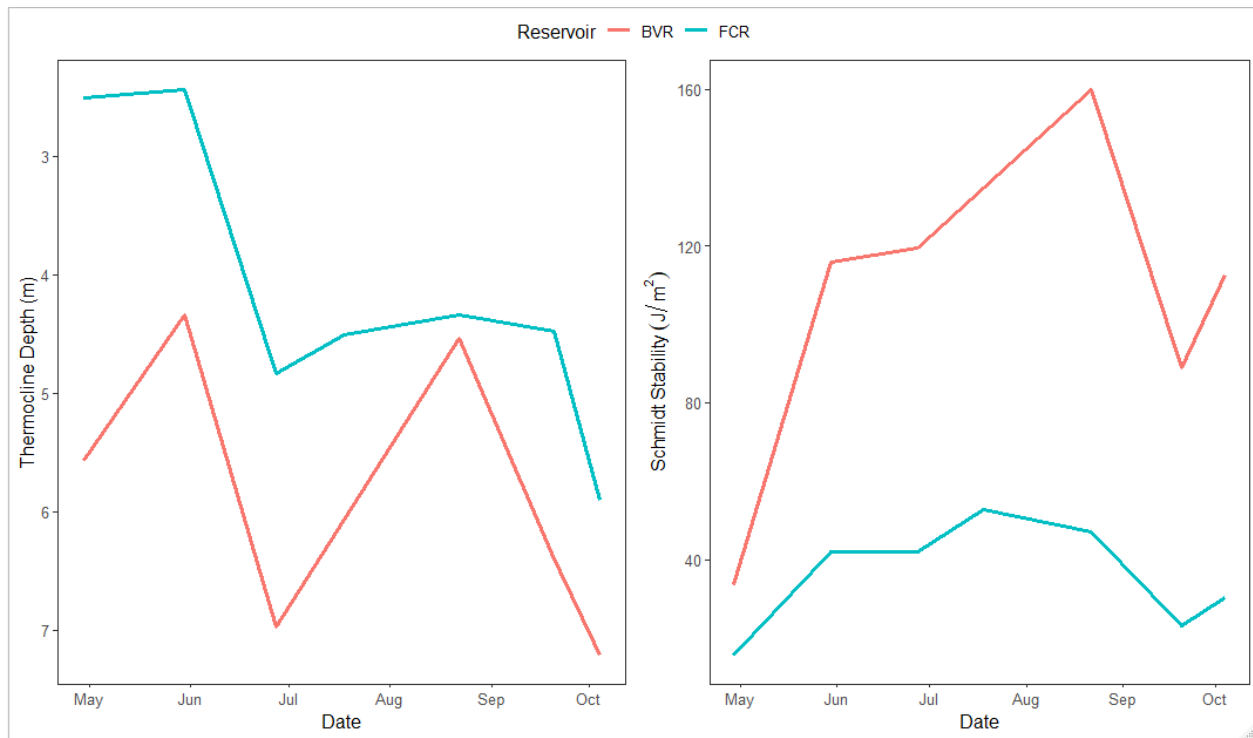


Figure S4. Time series of thermocline depth (left) and Schmidt stability (right) in Beaverdam Reservoir (red) and Falling Creek Reservoir (blue). Note that thermocline depth is plotted with the y-axis reversed, to represent the depth from the surface of the reservoir. Calculations were performed using the rLakeAnalyzer package (Winslow et al. 2019).

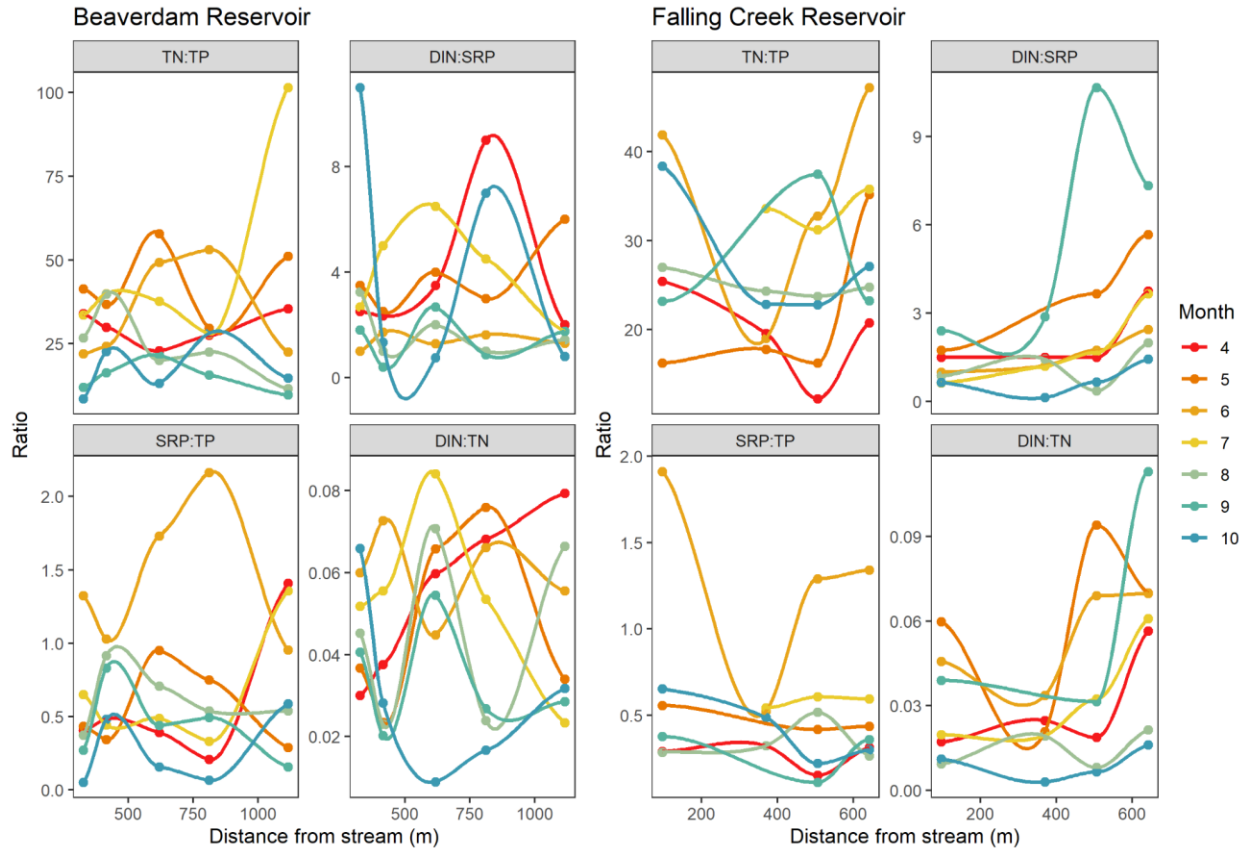


Figure S5. Nutrient stoichiometry ratios along the reservoir continuum in Beaverdam Reservoir and Falling Creek Reservoir, starting from the most upstream reservoir site in each reservoir. Colors represent the month of sampling. TN = total nitrogen; TP = total phosphorus; DIN = dissolved inorganic nitrogen consisting of the sum of NH_4 and NO_3 ; SRP = soluble reactive phosphorus.

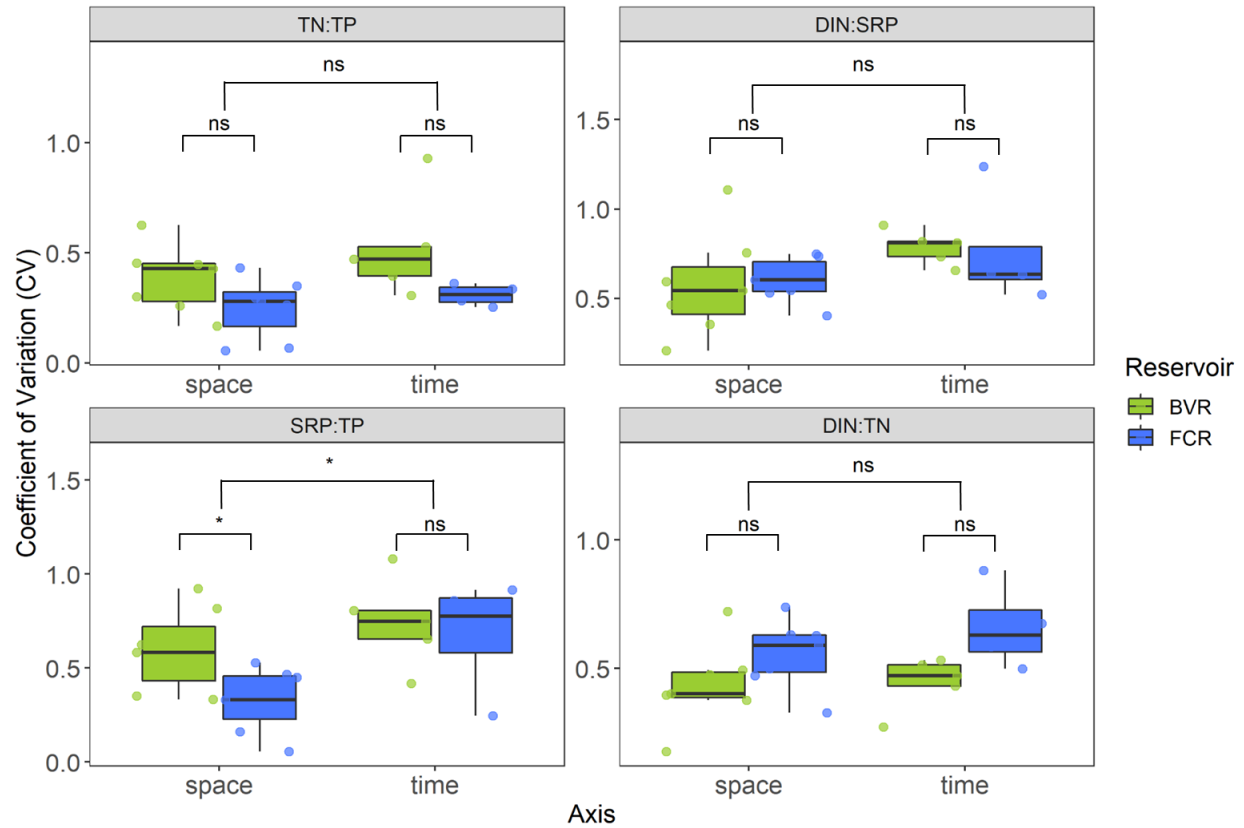


Figure S6. Heterogeneity, as measured by coefficient of variation, over space (CV_{space}) and time (CV_{time}) for nutrient stoichiometry ratios in BVR and FCR. 'ns' denotes no significant difference between groups ($p > 0.05$) and asterisks '*' denotes significant differences between groups ($p < 0.05$). TN = total nitrogen; TP = total phosphorus; DIN = dissolved inorganic nitrogen consisting of the sum of NH_4 and NO_3 ; and SRP = soluble reactive phosphorus.

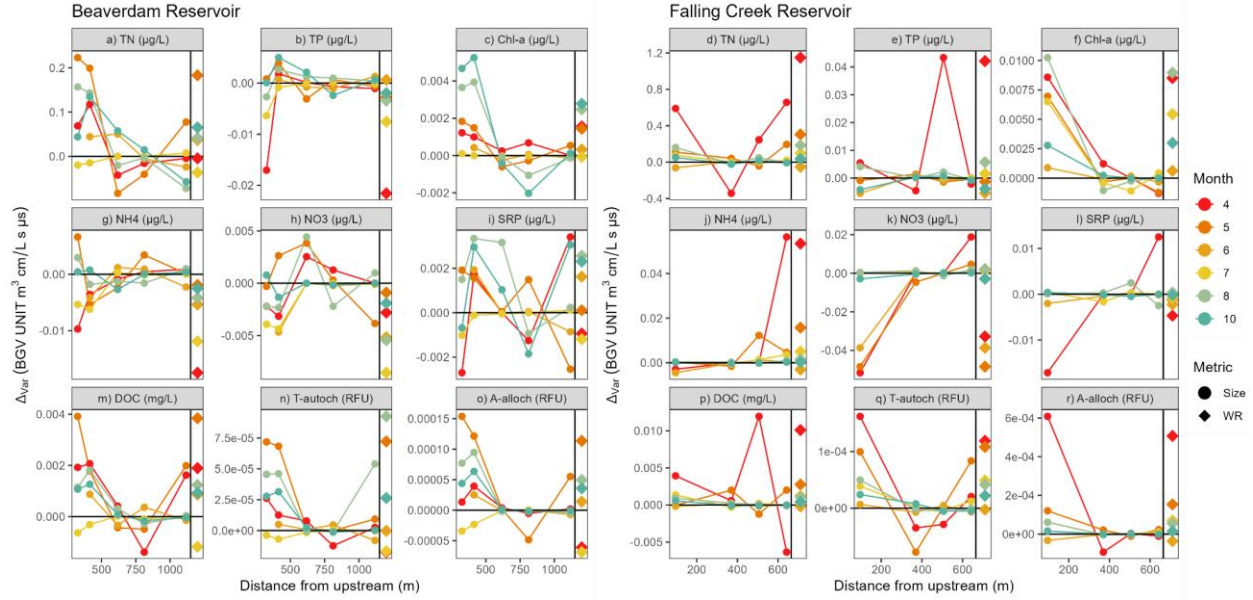


Figure S7. Δ_{var} , or the change in biogeochemical variables calculated using Eq. S1, along the reservoir continuum with each variable represented in individual panels, but with specific conductance at the stream sites standardized to the specific conductance observed at the most upstream reservoir site to examine the effect of specific conductance on the calculations. $\Delta_{var}Site$, which represents the change between sites, is plotted at the location of the downstream sampling site along the reservoir continuum and $\Delta_{var}WR$, which represents the change from inflow stream sites and the furthest downstream site, is shown in diamond-shaped points to the right of the vertical line. See *Methods: Spatially explicit processing of biogeochemical variables* for a full description of $\Delta_{var}Site$ and $\Delta_{var}WR$ (Eq. 1). Color represents the month in which the sample was collected, from 4 (April) to 10 (October). ‘BGV UNIT’ stands for the individual units of each biogeochemical variable, which are listed in each panel label. Samples were not collected at stream sites in September and are excluded from this analysis.

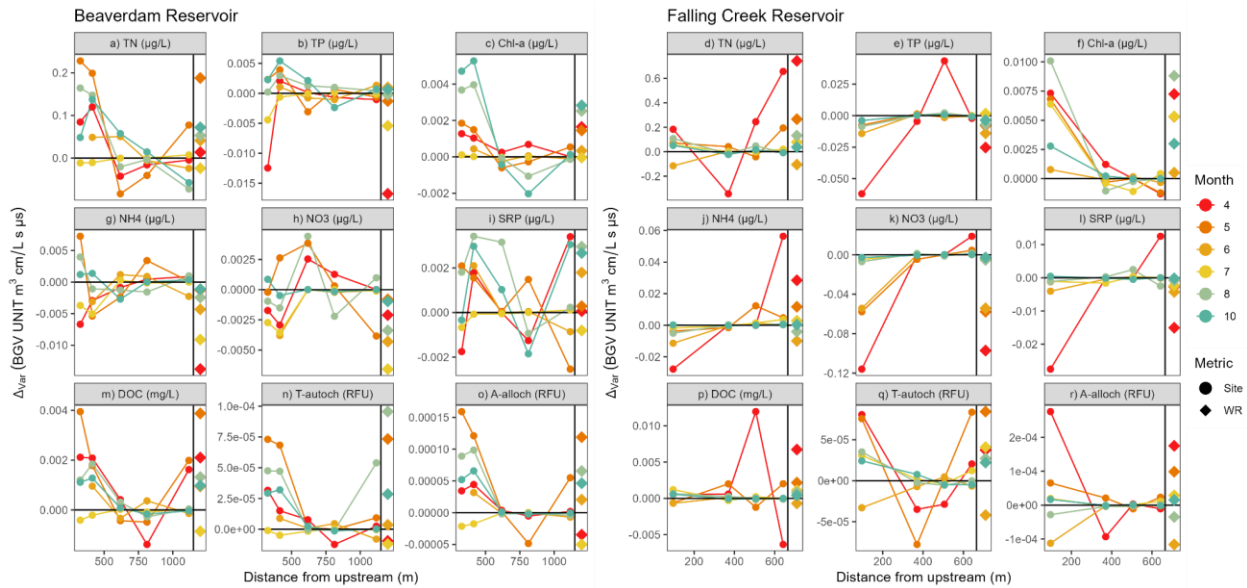


Figure S8. Δ_{var} , or the change in biogeochemical variables calculated using Eq. S1 (Text S4) not standardized by area, along the reservoir continuum (in comparison to the standardized metric shown in Figure 6). $\Delta_{var}Site$ which represents the change between sites, is shown at each sampling site along the reservoir continuum. $\Delta_{var}WR$, which represents the change from inflow stream sites and the furthest downstream site, is shown in diamond-shaped points to the right of the vertical line. Because BVR has two upstream sites (one within each arm), these are plotted according to their distance from the corresponding stream site, meaning the first point on the x-axis is B1a and the second point on the x-axis is B1b. Color represents the month in which the sample was collected, from 4 (April) to 10 (October). ‘BGV UNIT’ stands for the individual units of each biogeochemical variable, which are listed in each panel label. Samples were not collected at stream sites in September and are excluded from this analysis

References

Munger, ZW, Carey CC, Gerling AB, et al (2019) Oxygenation and hydrologic controls on iron and manganese mass budgets in a drinking-water reservoir. *Lake Reserv*

Manag, 35:3, 277-291, DOI: [10.1080/10402381.2018.1545811](https://doi.org/10.1080/10402381.2018.1545811)

Shuler, CK, and Mariner, KE (2020) Collaborative groundwater modeling: Open-source, cloud-based, applied science at a small-island water utility scale. *Environ Model Softw*, 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/>.

Thornthwaite, C. W., and J. R. Mather (1957) Instructions and tables for the computing potential evapotranspiration and the water balance. *Publ Climatol* X(3):311

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

Whitney M. Woelmer¹, R. Quinn Thomas^{1,2}, Mary E. Lofton¹, Ryan P. McClure¹, Heather L. Wander¹, Cayelan C. Carey¹

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

²Department of Forest Resources and Environmental Conservation, Virginia Tech, Blacksburg, Virginia, US

*Published in *Ecological Applications* with the citation:

Woelmer, W. M., Thomas, R. Q., Lofton, M. E., McClure, R. P., Wander, H. L., & Carey, C. C. (2022). Near-term phytoplankton forecasts reveal the effects of model time step and forecast horizon on predictability. *Ecological Applications*, 32(7), e2642.

This article is published in *Ecological Applications* as Open Access with a CC-BY-ND copyright held by the authors, giving the authors permission to republish in this dissertation.

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.

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 Roussio 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 that we are aware of which 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 km²), 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 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 are 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. ε is normally distributed random noise with mean = 0 and standard deviation of σ . Parameter (β_1 , β_2 , β_3 , and β_4), and process error standard deviation (σ) 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 σ_{add} and σ_{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 β_1 , β_2 , β_3 , and β_4 parameters by assuming normal distributions with large standard deviations:

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

Similarly, σ_{add} had an uninformative prior with a uniform distribution:

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

σ_{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, Rousso 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 σ_{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 km²) 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 (~ 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 were 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 ± 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 ± 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 ~ 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), as well as the chemical treatment and unnatural eradication of blooms at our study site, 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 Knappe 2020, Roussio 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 Roussio 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.

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:

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>

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-Aroita. 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 CO₂ 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. Tschardt. 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.
- Rajaei, 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.

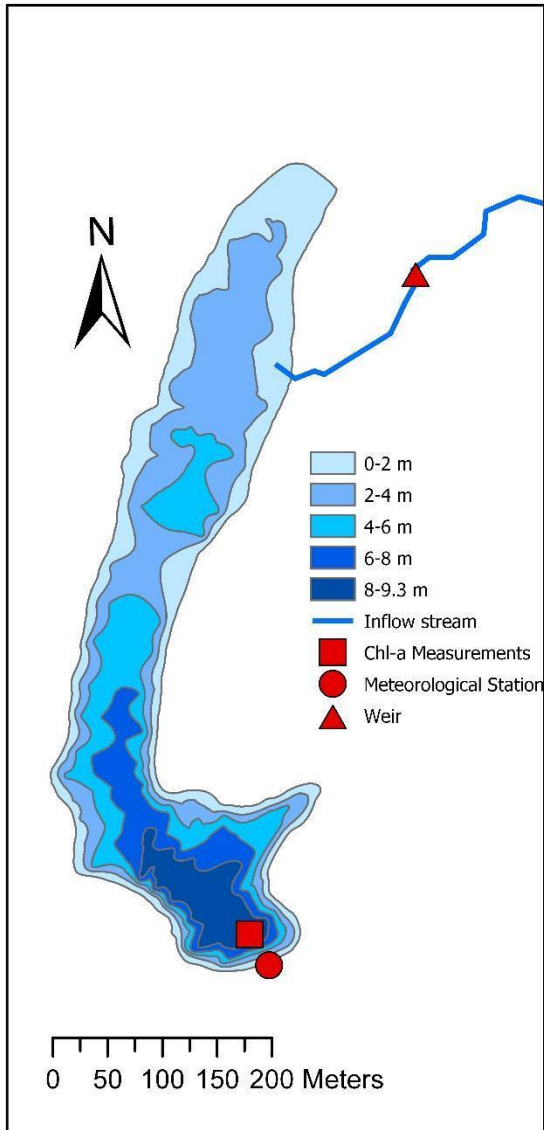


Figure 1. Map of Falling Creek Reservoir, Vinton, VA, USA (37.30°N, 79.84°W) study site showing locations of the weir, meteorological station, and chl-a sensor. Forecasts were generated for the location where chl-a measurements were collected.

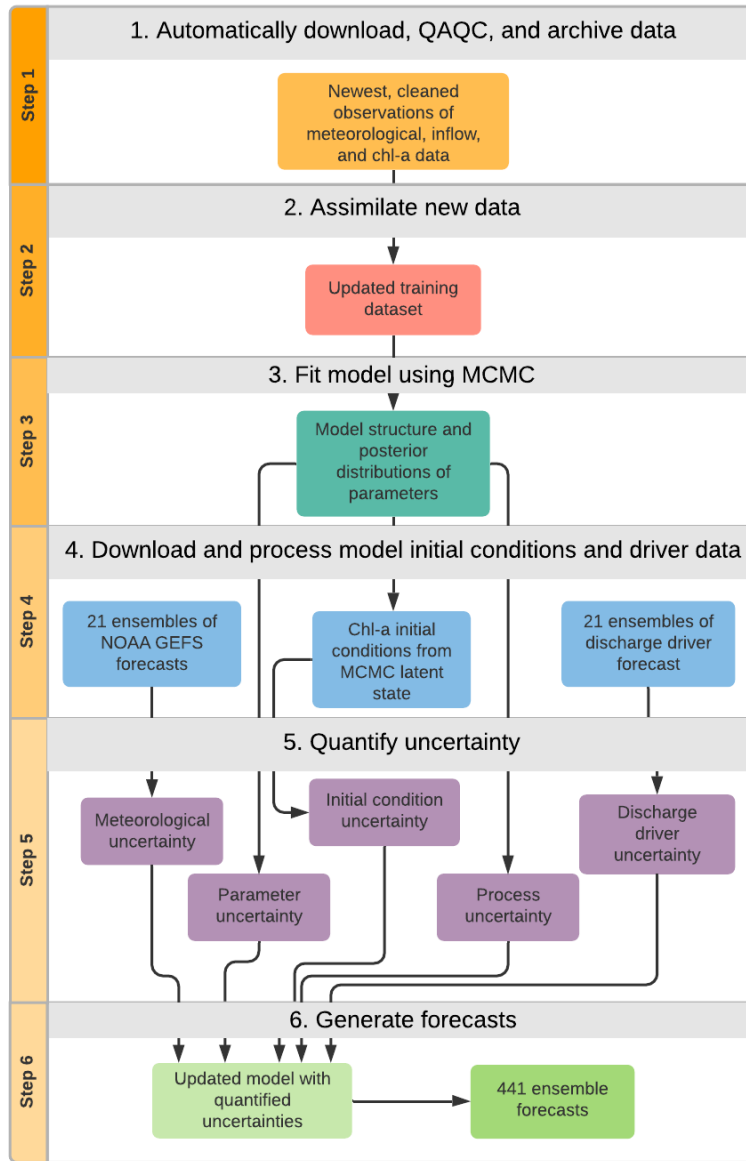


Figure 2. Summarized workflow for development of forecasts, from data acquisition and quality assurance/quality control (QA/QC) (Step 1), assimilation of new data (Step 2), re-fitting of the model with new observations (Step 3), acquisition and processing of model driver data (Step 4), quantification 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.

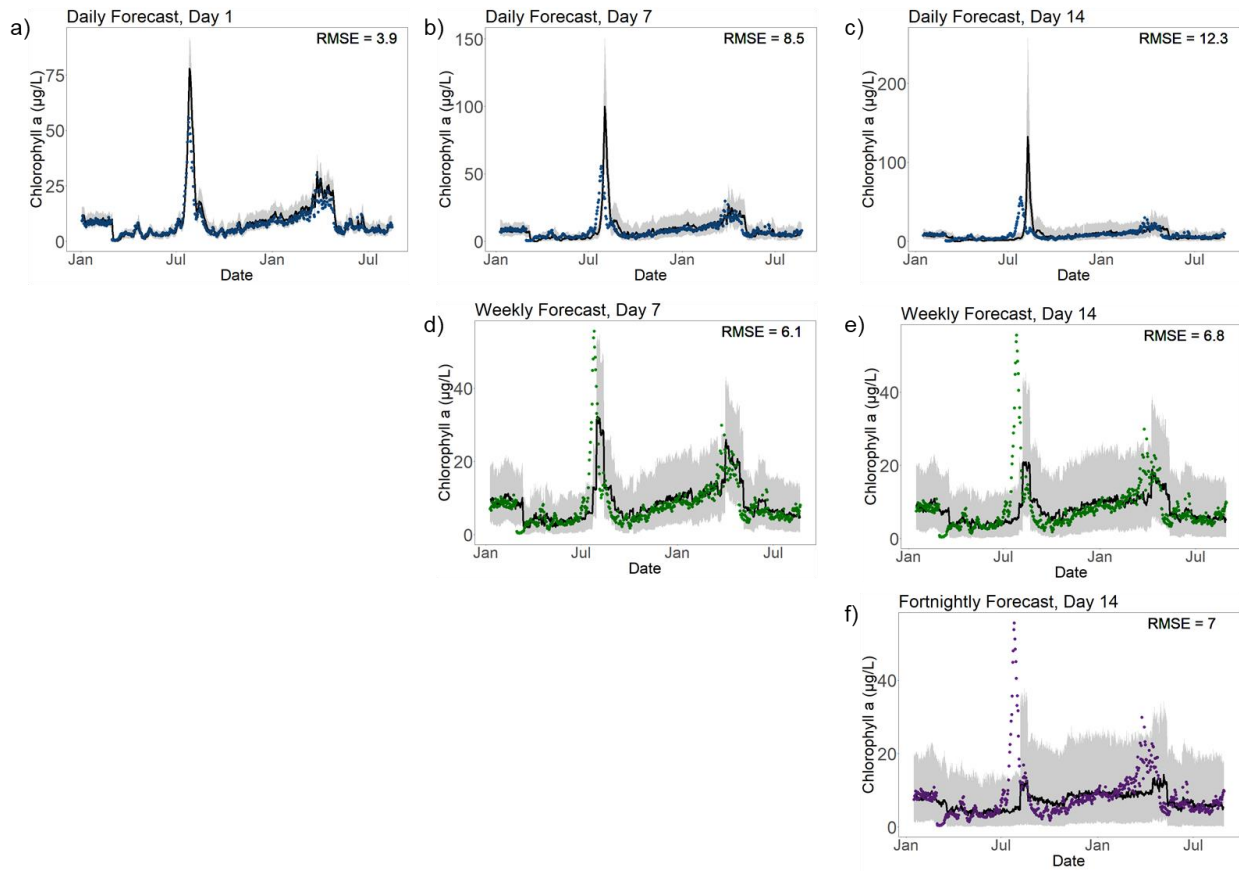


Figure 3. Forecasted (black line) and observed (points) chl-a concentrations over ~600 days from daily (blue, panels a-c), weekly (green, panel d-e), and fortnightly (purple, panel f) forecasts; the columns are grouped by forecast horizon, at 1 day (left column), 7 days (middle column), and 14 days (right 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.

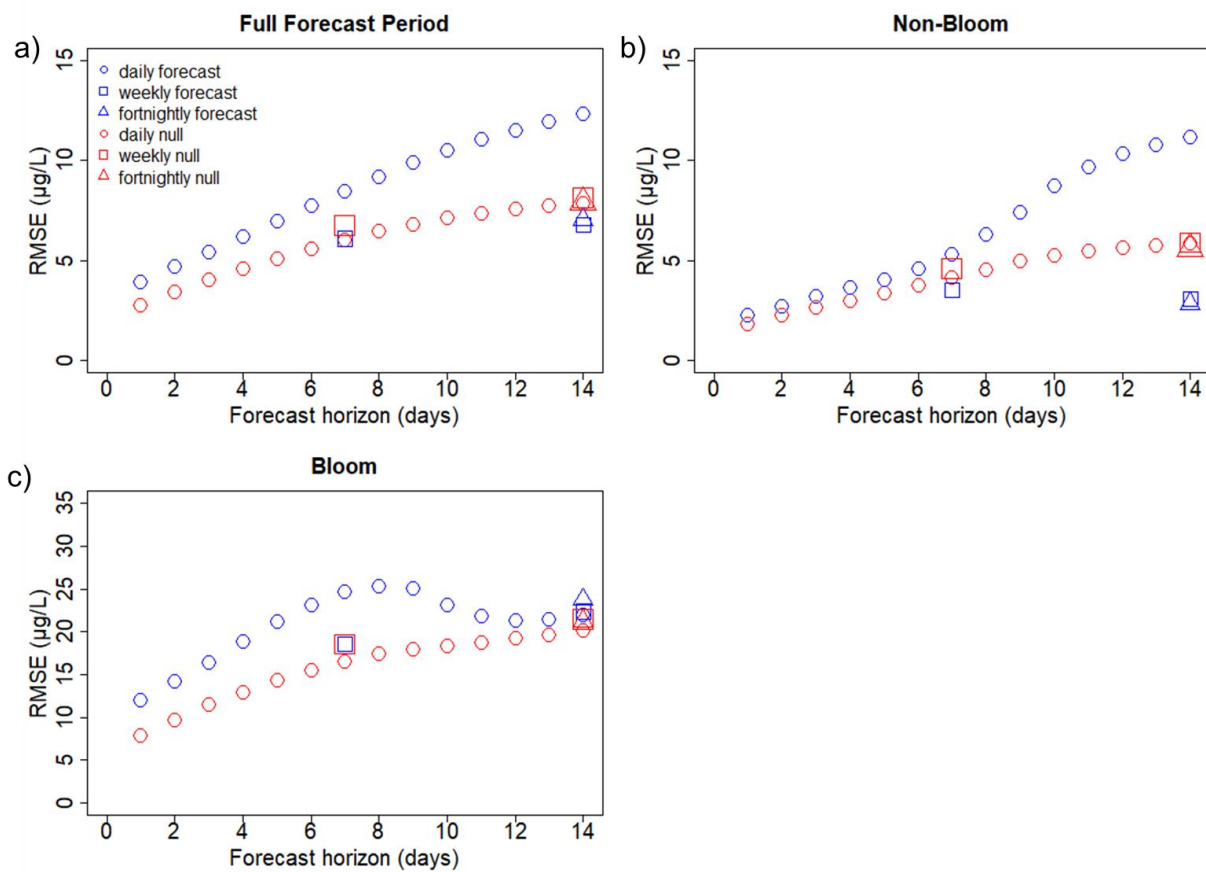


Figure 4. RMSE (root mean squared error) across 1-14 day forecast horizons for daily, weekly, and fortnightly forecasts and respective null models aggregated over a) the full ~600 day forecast period, b) non-bloom conditions only, and c) bloom conditions only. Note the different y-axis scale in panel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

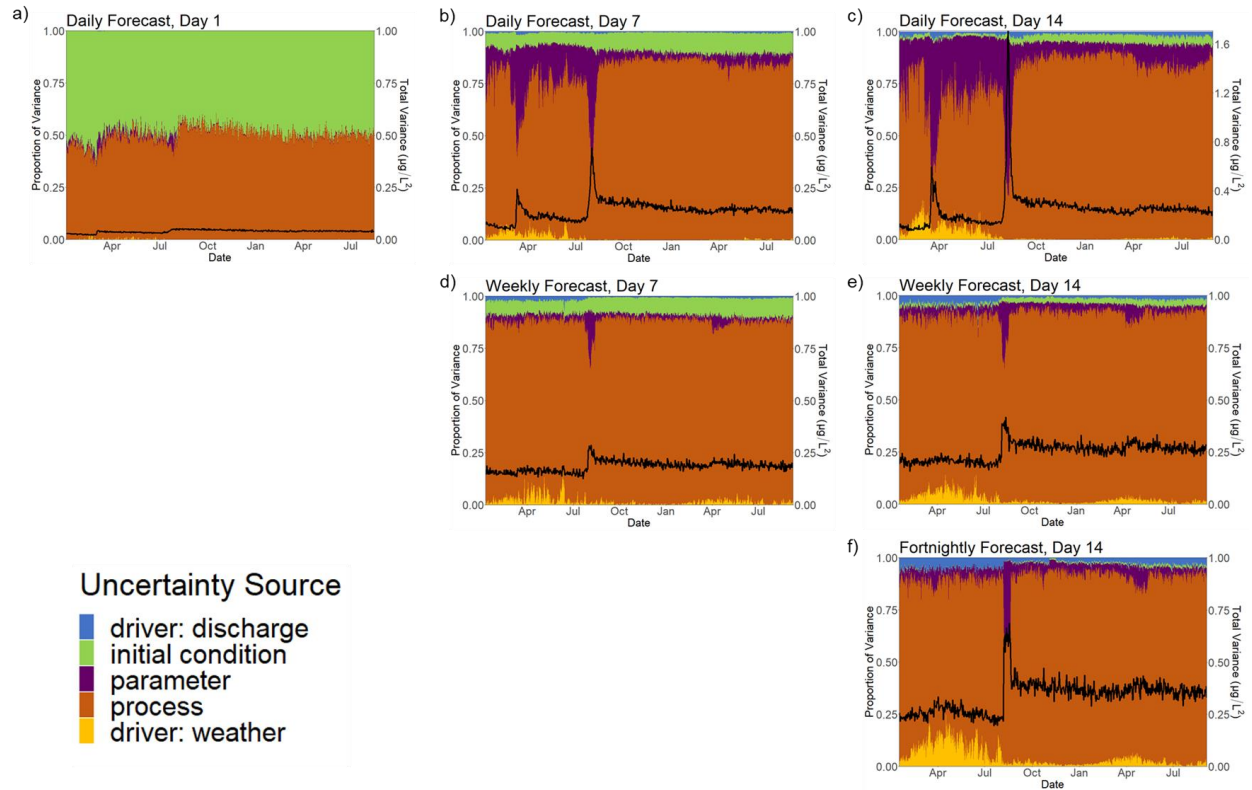


Figure 5. Relative proportion of forecast variance (uncertainty) over the forecast period (left y-axis) for a) daily forecasts at 1 day, b) daily forecasts at 7 days, c) daily forecasts at 14 days, d) weekly forecasts at 7 days, and e) weekly forecasts at 14 days, and f) fortnightly forecasts at 14 days.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

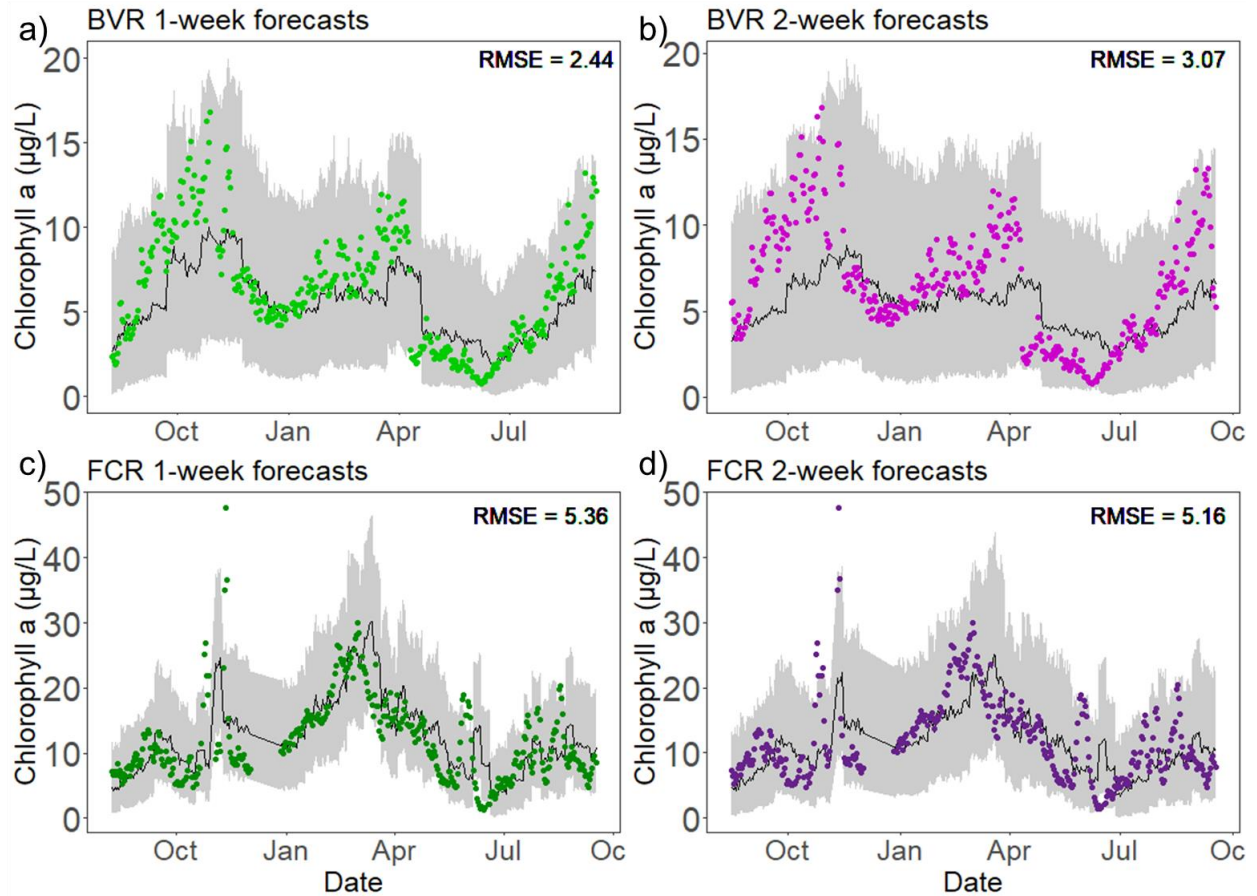


Figure 6. Forecasted (black line) and observed (points) chl-a concentrations over 1 year at Beaverdam Reservoir (panels a, b) and Falling Creek Reservoir (panels c, d) from weekly (green, panel a, c), and fortnightly (purple, panel b, d) forecasts; the columns are grouped by forecast horizon, at 7 days (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.

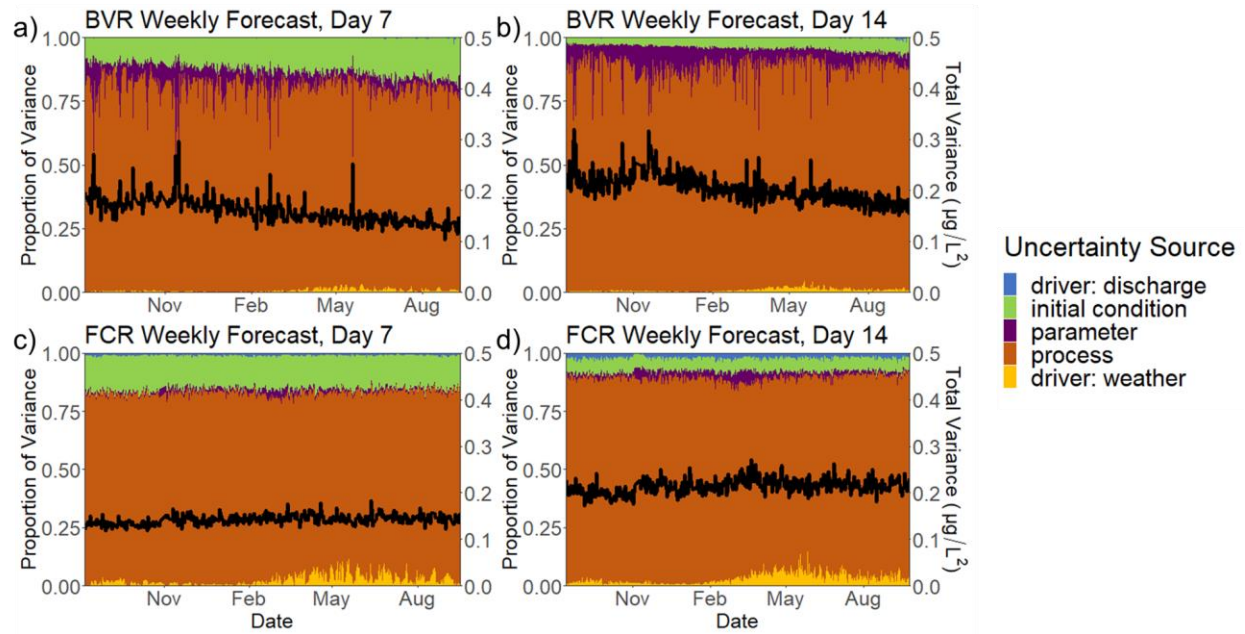


Figure 7. Relative proportion of forecast variance (uncertainty) for weekly forecasts from August 2020-August 2021. The top row shows the uncertainty analysis for Beaverdam Reservoir at 7 days (panel a) and 14 days (panel b), while the bottom row shows weekly forecasts for Falling Creek Reservoirs over 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).

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

Supporting Information

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

Whitney M. Woelmer¹, R. Quinn Thomas^{1,2}, Mary E. Lofton¹, Ryan P. McClure¹, Heather L. Wander¹, Cayelan C. Carey¹

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

²Department of Forest Resources and Environmental Conservation, Virginia Tech, Blacksburg, Virginia, US

Contents of this file:

Pages: 28

Section: 5

Tables: 10

Figures: 9

*Published in *Ecological Applications* with the citation:

Woelmer, W. M., Thomas, R. Q., Lofton, M. E., McClure, R. P., Wander, H. L., & Carey, C. C. (2022). Near-term phytoplankton forecasts reveal the effects of model time step and forecast horizon on predictability. *Ecological Applications*, 32(7), e2642.

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).

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).

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 (µg/L) for FCR	0.605	0.255	0.68	0.21
INW to Campbell Scientific discharge (m ³ /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.

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.

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 (Figure 2, Step 2)	Discharge	m ³ /s	INW Aquistar PT2X pressure sensor	2013-present	10 minutes	Weekly to monthly
Forecasts (Figure 2, Steps 2 and 4)	Discharge	m ³ /s	Campbell Scientific CS451 pressure sensor	2019-present	10 minutes	Daily
Training (Figure 2, Step 2)	Surface chl-a	µg/L	SBE19plus CTD profiler	2013-present	1 week	Weekly
Forecasts (Figure 2, Steps 2 and 4)	Surface chl-a	µg/L	EXO-Sonde	August 2018-present	15 minutes	Daily

Training (Figure 2, Step 2)	Shortwave radiation	W/m ²	NLDAS-2	2013- 2015	1 hour	N/A
Training (Figure 2, Step 2)	Shortwave radiation	W/m ²	Hukseflux NR01 4- Component Net Radiometer	2015- present	5 minutes	Daily
Forecasts (Figure 2, Step 4)	NOAA GEFS Shortwave Forecasts	W/m ²	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 ³ /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 ²)	Weekly, fortnightly	--
Mean relative humidity (%)	Daily	--
Median shortwave (W/m ²)	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 (m ³ /s)	-0.11	1																
Max flow (m ³ /s)	-0.09	0.97	1															
Min flow (m ³ /s)	-0.11	0.97	0.92	1														
Median flow (m ³ /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/m ²)	-0.13	-0.17	-0.19	-0.13	-0.1	-0.39	-0.33	1										
Mean shortwave (W/m ²)	-0.15	-0.18	-0.21	-0.15	-0.2	-0.39	-0.33	0.97	1									
Median shortwave (W/m ²)	-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/m ²)	-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/m ²)	-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	Ma x 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	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			

humidity (%)																		
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 ²)	-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 ²)	-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 Aikaike'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

Table S10. Root mean square error (RMSE) of daily and weekly forecasts over the full year of forecasts, non-bloom conditions, and bloom conditions.

Table S10.							
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

Fortnightl							
y	14'	7.0	7.9	2.7	5.6	18.8	16.9

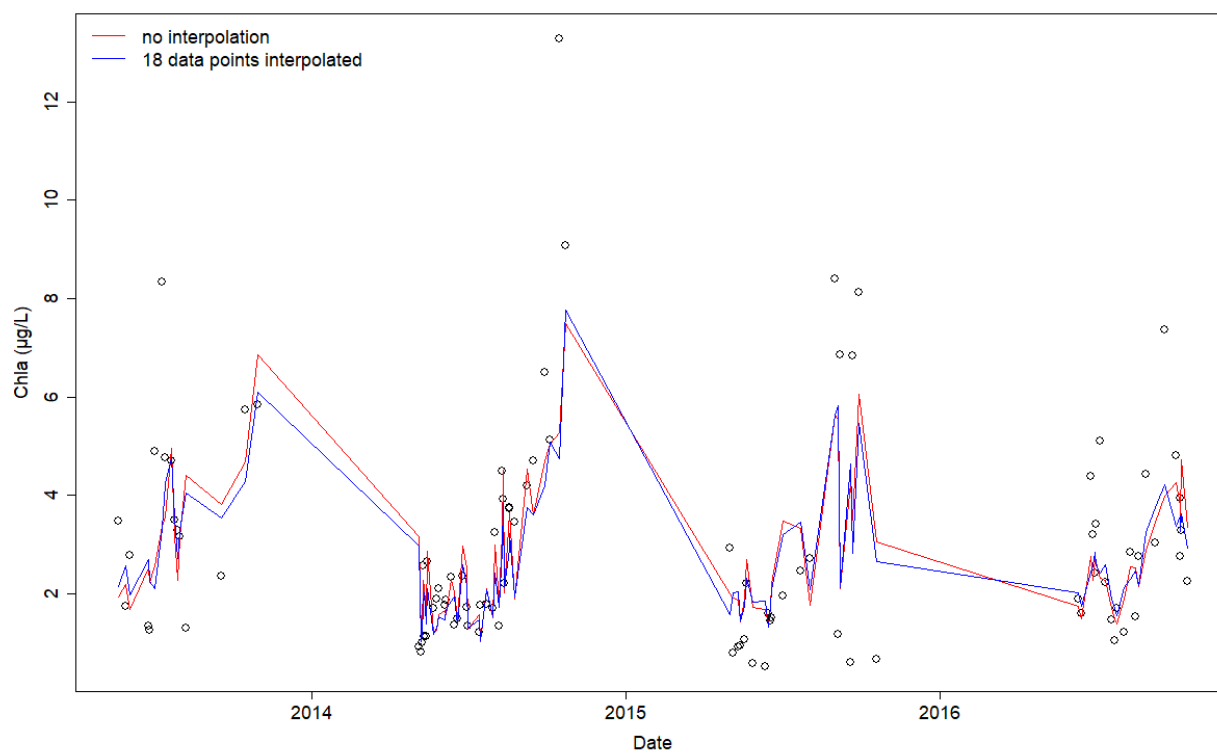


Figure S1. Observed chl-a (points) and predicted weekly chl-a using data from 2013-2016 (red line, $R^2 = 0.38$), and data from 2013-2016 including 18 interpolated datapoints (blue line, $R^2 = 0.44$).

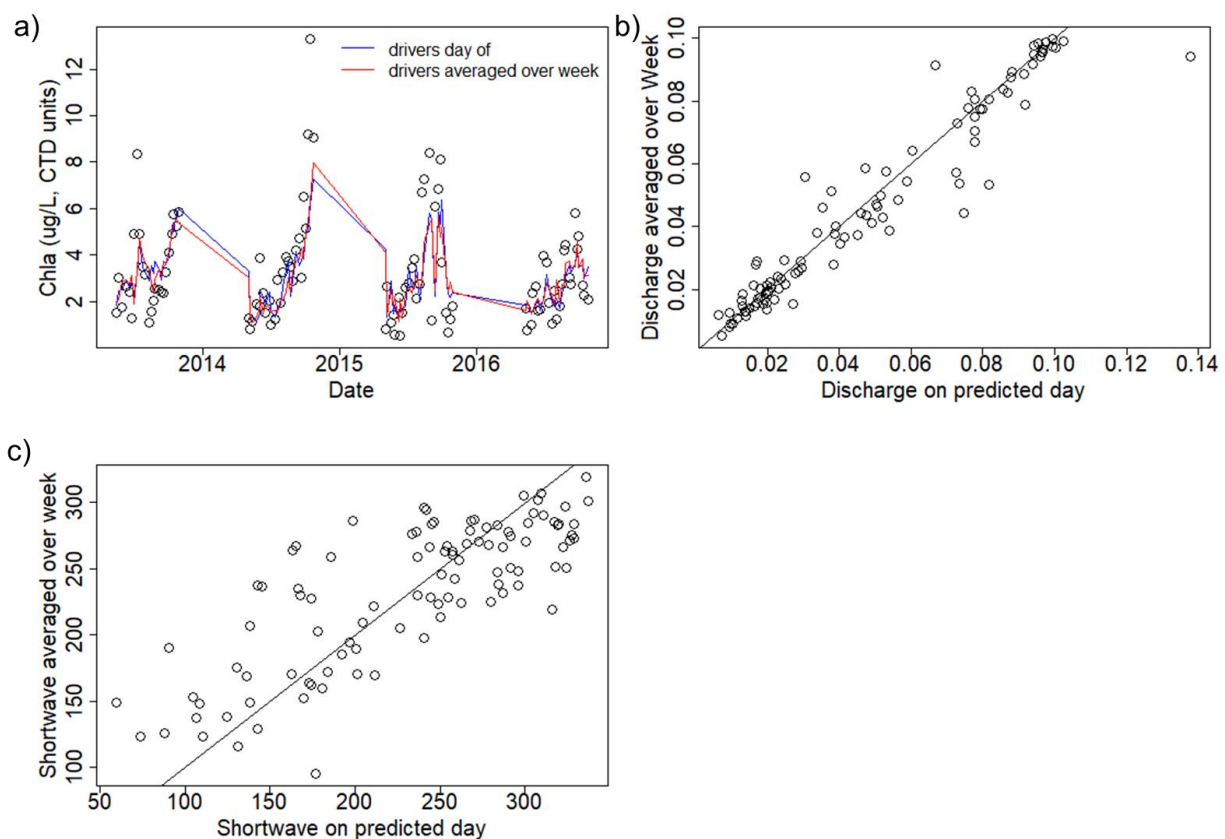


Figure S2. a) Relative performance of the same weekly linear model with mean shortwave and mean discharge covariates developed using driver variables averaged over a week period (red line) vs. driver variables derived from point estimates on the day being predicted (blue line). Observations are shown 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).

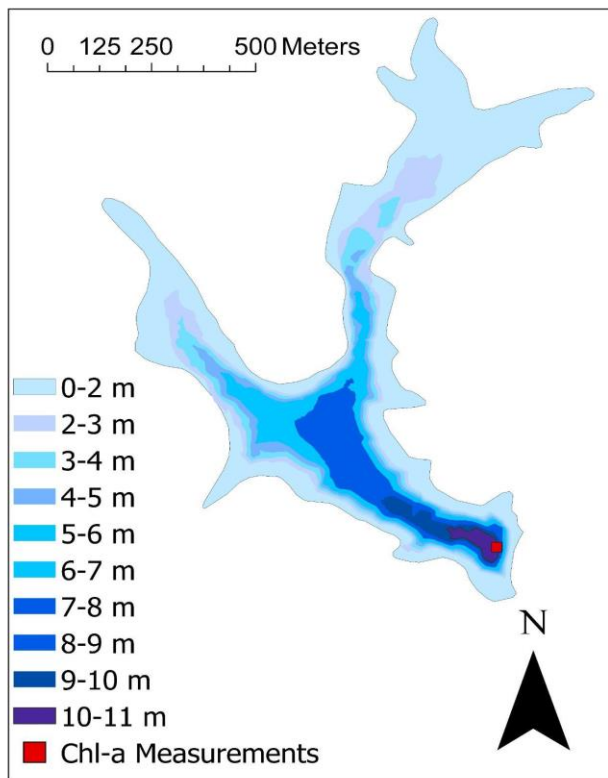


Figure S3. Map of Beaverdam Reservoir, Vinton, VA, USA (37.31°N, 79.82°W) study site showing where chl-a measurements were taken at the dam. Forecasts were generated for the location where chl-a measurements were collected.

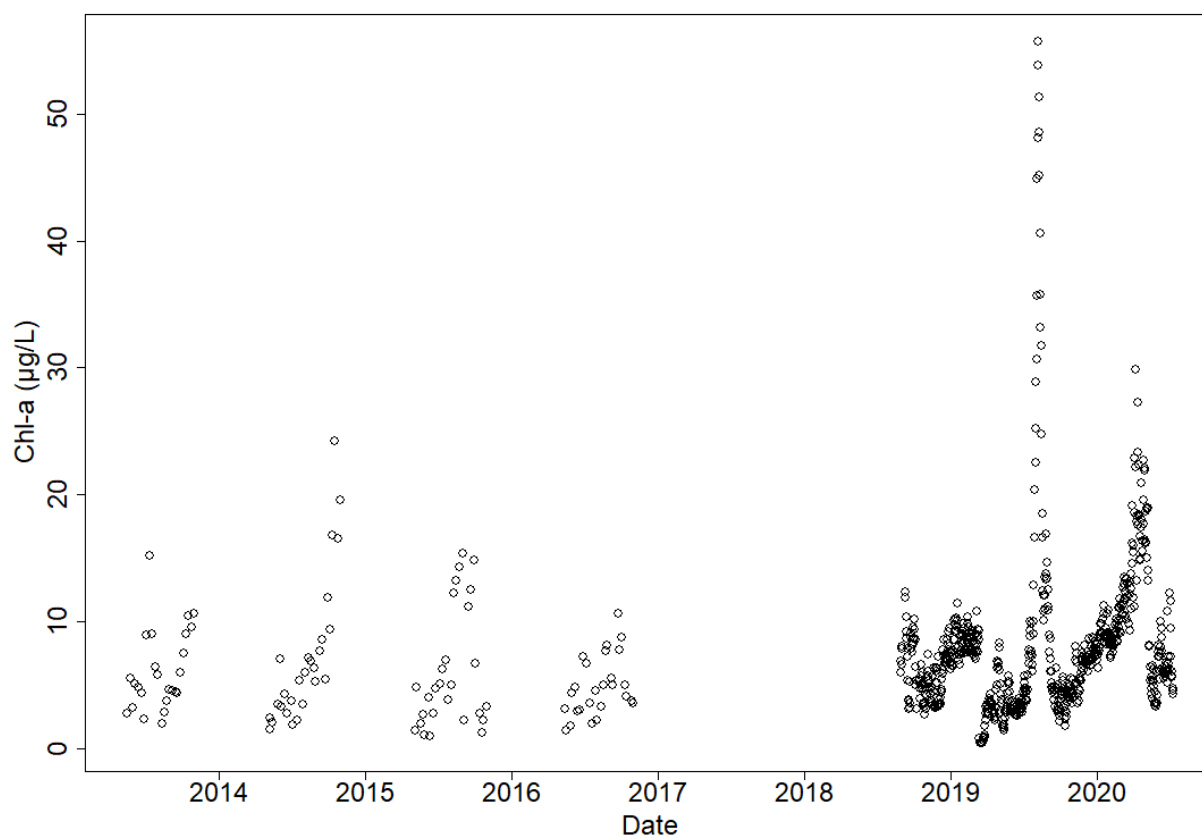


Figure S4. Observed time series of chl-a from the weekly training dataset (2013-2016) and spin-up and forecast time series for weekly and daily forecasts (Aug 2018-Dec 2019), using the harmonized CTD and EXO data, converted to EXO units (see Section S1)

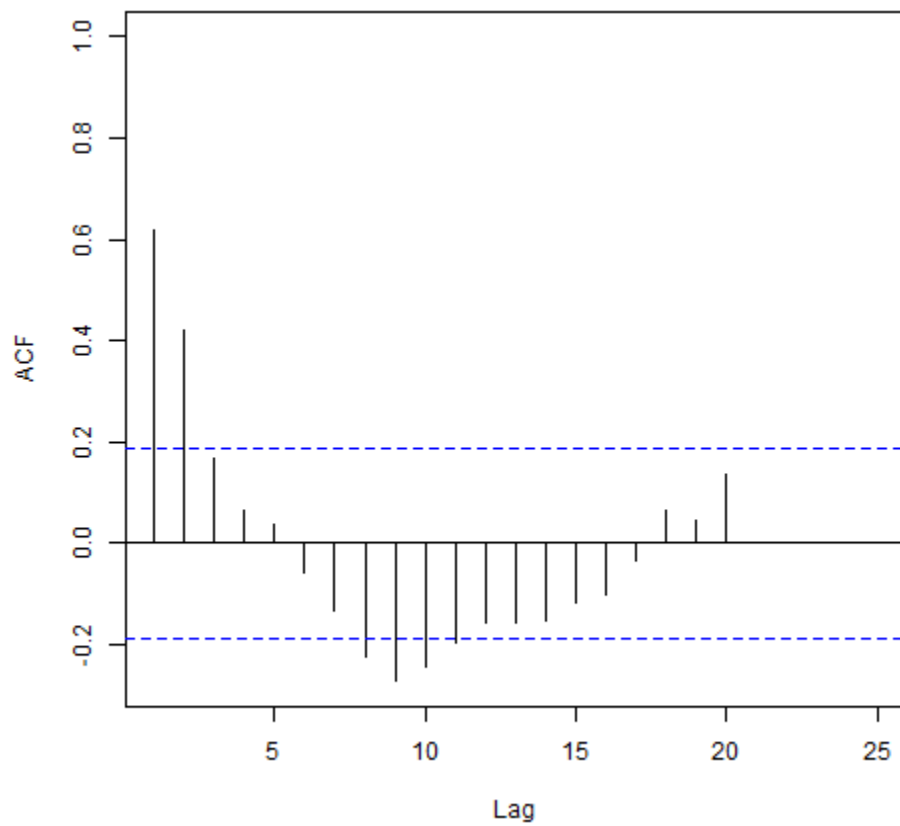


Figure S5. Auto-correlation function (ACF) for chlorophyll-a in the CTD weekly training dataset, 2013-2016.

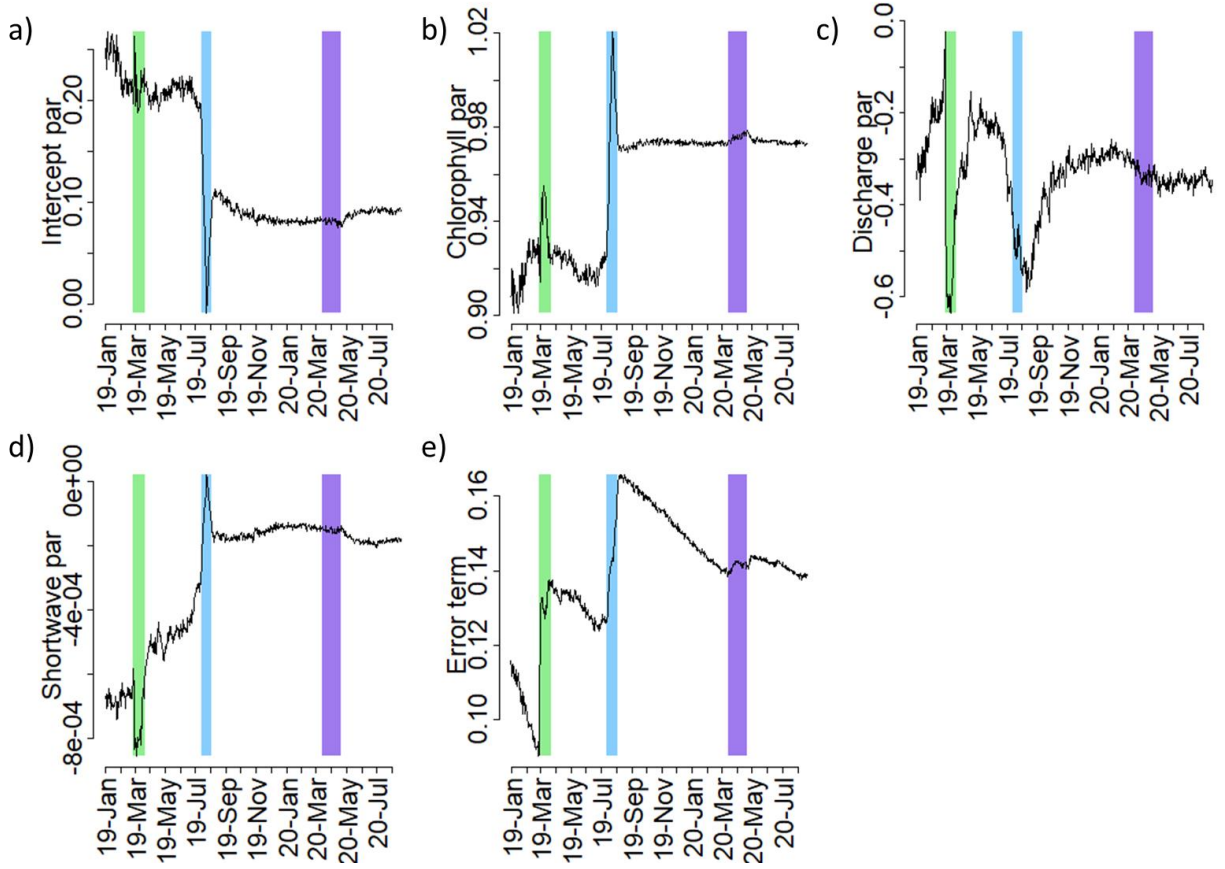


Figure S6. Parameter (par) evolution for daily forecasts throughout the forecasting period for a) the intercept parameter, b) chlorophyll lag term, c) discharge parameter, d) shortwave radiation parameter, and e) error term. See Equation 1 in the main text. The green shaded area denotes the time when 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.

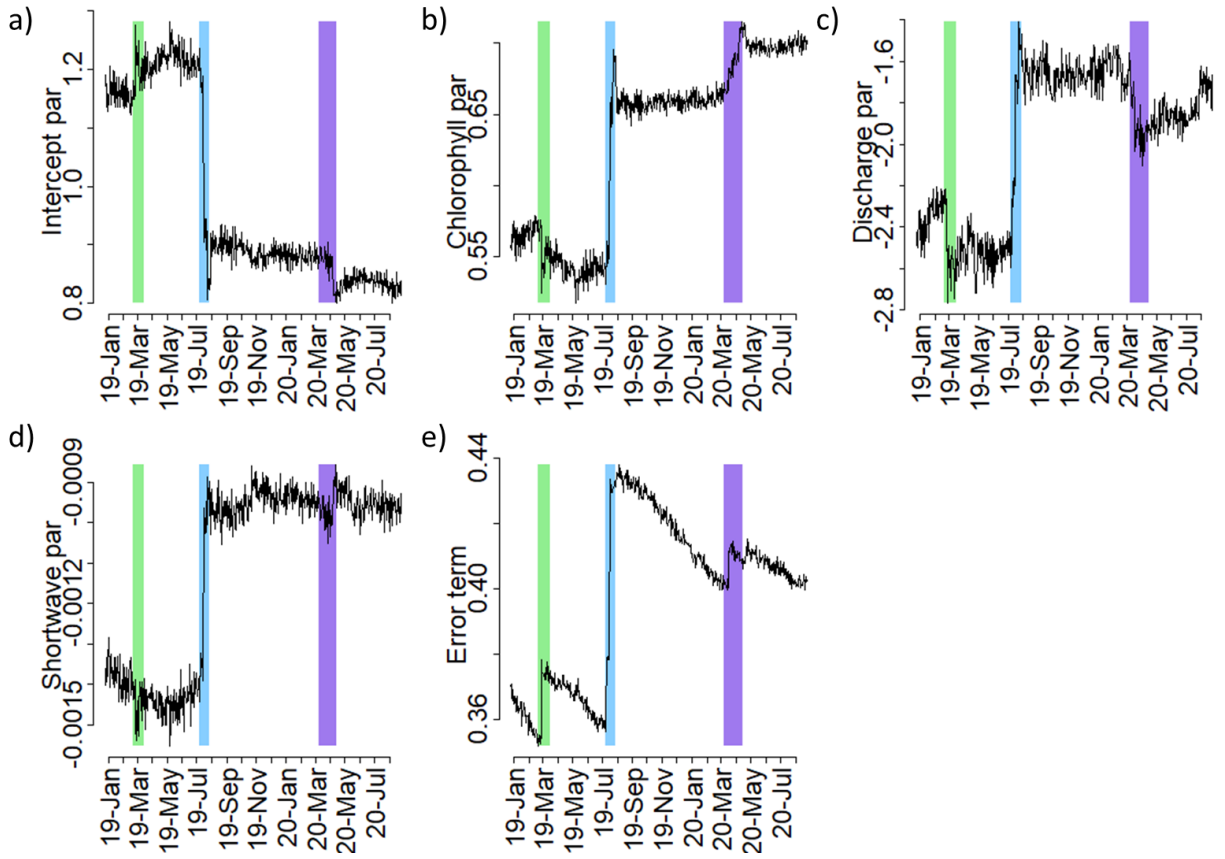


Figure S7. Parameter (par) evolution for weekly forecasts throughout the forecasting period for a) the intercept parameter, b) chlorophyll lag term, c) discharge parameter, d) shortwave radiation parameter, and e) error term. See Equation 1 in the main text. The green shaded area denotes the time when 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.

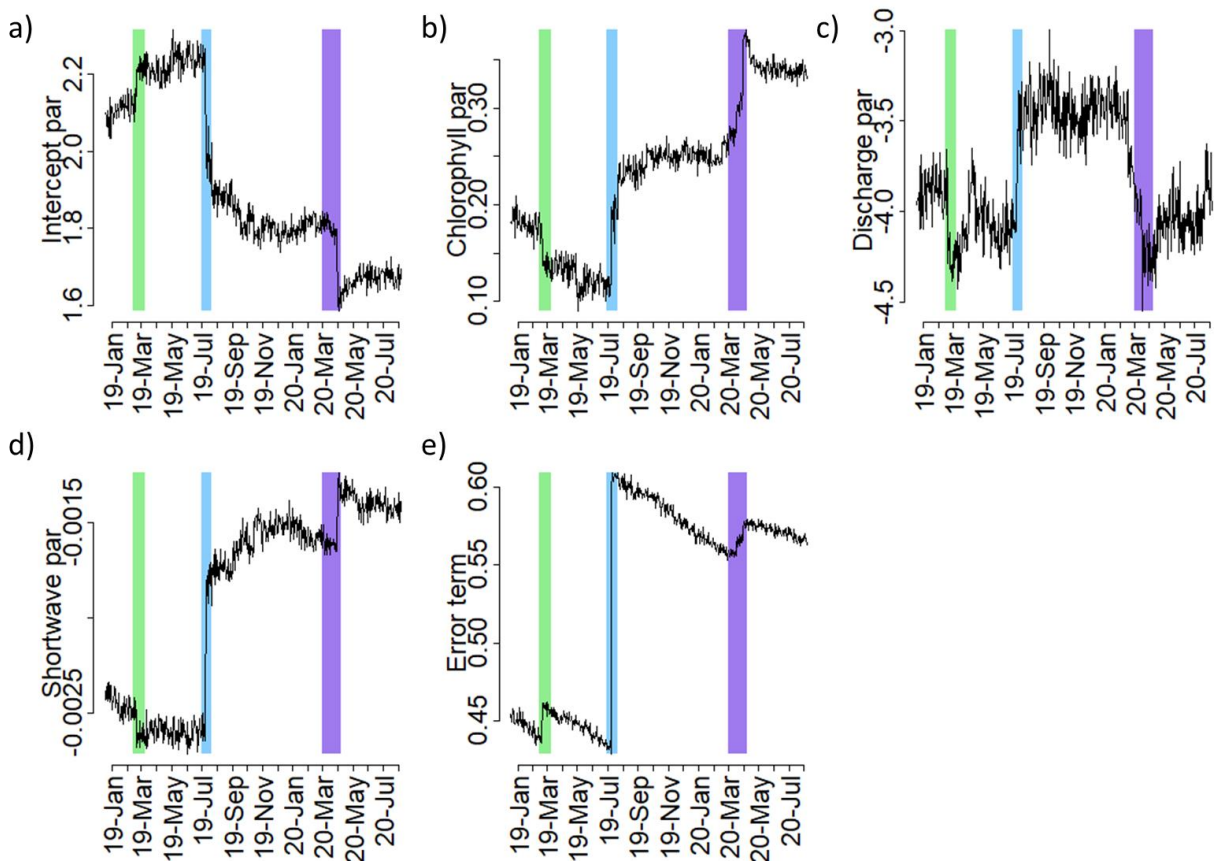


Figure S8. Parameter (par) evolution for fortnightly forecasts throughout the forecasting period for a) the intercept parameter, b) chlorophyll lag term, c) discharge parameter, d) shortwave radiation parameter, and e) error term. See Equation 1 in the main text. The green shaded area denotes the time 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.

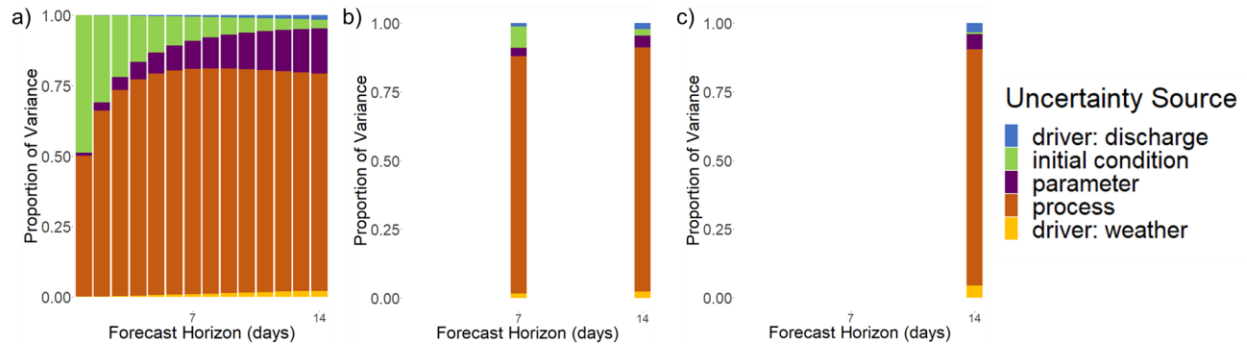


Figure S9. Proportion of variance contributed by different uncertainty sources to total forecast variance over the forecast horizon aggregated over the entire forecast period for a) daily, b) weekly, and c) fortnightly forecasts. Meteorological data, parameter, initial condition, and discharge data uncertainty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

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>
- 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)
- 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. Pollingher, and T. Zohary. 1999. Biovolume calculation for pelagic and benthic microalgae. *Journal of Phycology* 35:403–424.

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/>.

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).

**Chapter IV: Performance of near-term, iterative water temperature and dissolved oxygen
forecasts varies across time and depth***

Whitney M. Woelmer¹, R. Quinn Thomas^{1,2}, Freya Olsson¹, Bethel G. Steele³, Kathleen C.
Weathers³, Cayelan C. Carey¹

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

²Department of Forest Resources and Environmental Conservation, Virginia Tech, Blacksburg,
Virginia, USA

³Cary Institute of Ecosystem Studies, Millbrook, New York, USA

*In preparation for submission to *Ecological Informatics*

Abstract

Near-term iterative ecological forecasting has great potential for providing new insights into our ability to predict multiple ecological variables. However, variability in forecast performance across time and space has largely been unexamined across multiple ecosystem variables using a process-based modeling approach. To explore how forecast performance varies for water temperature and dissolved oxygen, two freshwater variables important for lake ecosystem functioning, we produced forecasts with uncertainty at multiple depths over two open-water seasons in Lake Sunapee, NH, USA. Our forecasting system, FLARE (Forecasting Lake And Reservoir Ecosystems), uses a 1-D coupled hydrodynamic-biogeochemical model and daily data assimilation, which updates model initial conditions of water temperature and dissolved oxygen and fits parameters over time. We assessed FLARE performance using forecast accuracy (Continuous Ranked Probability Score) and forecast skill (i.e., comparing FLARE's forecast accuracy relative to the accuracy of a climatology null model). We found that FLARE water temperature forecasts were more skillful than FLARE oxygen forecasts and could beat the null model 11 days into the future, as compared to only two days for oxygen. However, FLARE temperature forecast skill degraded more rapidly than FLARE oxygen forecast skill into the future. Across evaluation periods within different years, we observed highly variable forecast skill for both variables. Notably, despite generally lower forecast accuracy of FLARE forecasts when observational complexity was higher (assessed by permutation entropy), FLARE forecasts were more skillful than a climatology null model when the time series being forecasted was more complex. Our results emphasize that some ecosystem variables are more predictable than others and that process-based forecasts provide important information above a climatology null model, informing development of additional quantitative tools for predicting ecosystem change.

Introduction

Near-term, iterative forecasts of water quality variables have much potential for enabling managers to anticipate and mitigate change in freshwater ecosystems, which are experiencing unprecedented global change stressors (Dietze et al. 2018; Carey et al. 2022; Lofton et al. 2023). With increasing variability in water quality and ecosystem functioning due to land use and climate change (Ho and Michalak 2019; Woolway and Merchant 2019; Kraemer et al. 2021; IPCC 2023), near-term forecasts (i.e., quantitative predictions of future ecosystem states with uncertainty; Dietze 2017; Carey et al. 2022) of key freshwater ecosystem variables could improve understanding of changes in freshwater quality over day to decadal scales (Radeloff et al. 2015; Lofton et al. 2023; Lee et al. 2023). Through the near-term, iterative forecast cycle, forecasts are repeatedly produced for a range of forecast horizons, or time into the future (Dietze 2017; Moore et al. 2022). When data become available, forecasts are updated and evaluated with observations, allowing for improved predictions over time which can be integrated into decision-making frameworks (Dietze et al. 2018; Henden et al. 2020; Bodner et al. 2021).

Two foundational freshwater variables, which both control and serve as important indicators of lake ecosystem functioning, are water temperature and dissolved oxygen (Wetzel 2001). Water temperature, which is influenced by meteorological variables including air temperature, precipitation, and wind (Figure 1a), as well as other factors (Wetzel 2001), is a key regulator of thermal stratification (Figure 1b) and solubility of gasses within the water column (Figure 1e), which in turn controls dissolved oxygen concentrations (Figure 1c, d; Sánchez et al. 2007; Bhateria and Jain 2016). Both water temperature and dissolved oxygen influence many ecosystem processes including ecosystem and organismal metabolism (Caffrey 2004; Staehr et al. 2012), nutrient cycling (Sondergaard et al. 2001), and habitat availability for organisms

(Davis 1975; Jones et al. 2008; Magee et al. 2019). Consequently, understanding and forecasting these two variables is critical because shifts in lake water temperature and dissolved oxygen can indicate changes in ecosystem functioning and lake trophic state (Sánchez et al. 2007; Simões et al. 2008; Richardson et al. 2017).

As a result of their sensitivity to these interacting ecological processes, observations of water temperature and dissolved oxygen, and correspondingly forecast performance, are likely to vary over space (i.e., with depth in the water column) and time (i.e., inter- and intra-annually) in lakes. Prior studies which have assessed water temperature forecast performance at multiple depths in lakes have found that surface waters are typically more challenging to forecast accurately than bottom waters due to the strong influence of meteorological variability on surface waters and seasonal thermal stratification stabilizing bottom waters (Thomas et al. 2020, 2023; Wander et al. 2023). However, changes in dissolved oxygen forecast performance with increasing depth (i.e., a decrease or increase in performance as you go deeper into the water column) in lakes are less conclusive, with Durell et al. (2023) and Saber et al. (2020) finding that surface dissolved oxygen forecast performance was more variable than bottom water performance, while Lin et al. (2023) found no difference or slightly better forecasts in bottom waters than surface waters.

In addition to variability with depth throughout the water column, temporal variability in water quality at annual time scales is well-documented in freshwater ecosystems, with metrics of ecosystem functioning and water quality being highly variable between years (Jassby et al. 2003; Nõges and Tuvikene 2012; Carey et al. 2014; Geng et al. 2022). For example, the date of fall turnover, defined as the breakup of summer thermal stratification (Wetzel 2001), can vary from year to year, potentially driving annual differences in predictability of water temperature and

dissolved oxygen. However, the extent to which fluctuations in the forecasted variables influence performance across years and depths simultaneously remains understudied. To date, most lake forecast studies have either focused on predictions in a single year (reviewed by Lofton et al. 2023) or do not explicitly evaluate inter-annual variability in forecast performance even when forecasts are made over multiple years (Saber et al. 2020).

Lake forecasts which include both water temperature and dissolved oxygen together may be able to provide a more comprehensive perspective of changing lake ecosystem functioning and be more useful for managers than if either variable was forecasted alone. Specifically, managers may need both water temperature and dissolved oxygen forecasts simultaneously to guide decision-making, as water managers often need to optimize multiple concurrent needs (Jackson-Blake et al. 2022). For example, water temperature forecasts can guide decisions on water depth withdrawal for optimizing downstream habitat for organisms with temperature sensitivity (Kim and Choi 2021) or preparing for turnover-induced water quality degradation (Carey et al. 2022). Dissolved oxygen forecasts can guide complementary decisions to water temperature forecasts, such as optimizing downstream habitat for organisms with dissolved oxygen sensitivity (Calamita et al. 2021), informing the use of oxygenation systems for mitigating anoxia (Carey et al. 2022), or determining which water treatment applications for drinking should be implemented (Lee 2015; Li et al. 2022).

However, despite the need for both water temperature and dissolved oxygen predictions, near-term iterative ecological forecasts which simultaneously predict multiple water quality variables remain rare (Lofton et al. 2023). Of the five studies of which we are aware that do predict multiple water quality variables (e.g., water temperature, dissolved oxygen, and conductivity in one study; Saber et al. 2020), four used coupled machine learning approaches.

When input data are available, machine learning approaches provide relatively high predictive ability for nonlinear applications (Zhu et al. 2022), but offer limited ability to draw inference about underlying mechanisms (Peters et al. 2014; Lazer et al. 2014; Jia et al. 2018). In addition, all four machine learning forecasts of temperature and oxygen forecasted each variable with separate models, ignoring ecosystem interactions that occur between variables (Saber et al. 2020; Durell et al. 2022, 2023; Lin et al. 2023).

In comparison to machine learning models, process-based models are not as commonly used in forecasts of dissolved oxygen or water temperature (Durell et al. 2022; Lofton et al. 2023). This may potentially be because process-based models often require several input variables and may not provide a gain in skill relative to data-driven models to predict water quality (Jin et al. 2019; Durell et al. 2022). Predicting multiple water quality variables with a process-based model allows the forecasts to incorporate the numerous interrelated processes that control dynamics of both variables, thereby providing a more holistic representation of ecosystem changes. We are aware of only one published study which forecasted water temperature and dissolved oxygen simultaneously using a single process-based model to represent interrelated dynamics of both variables (Carey et al. 2022). However, their reservoir forecasts were only over a period of a few days and focused on scenario-based management, without a full assessment of how water temperature and dissolved oxygen forecast performance varied over time and space.

Using multiple measures of forecast performance, which assess distinct but complementary information, can provide a more holistic assessment about forecast utility than a single metric alone (Jolliffe and Stephenson, 2012). In this paper, we refer to forecast performance as a general term inclusive of multiple forecast evaluation metrics, pertaining to

both forecast *accuracy* and forecast *skill*. Forecast *accuracy* metrics compare forecasted values or distributions to observations (Hyndman and Athanasopoulos 2021), with scores often reported in native units (i.e., °C for temperature and mg/L for oxygen). In contrast, forecast *skill* extends a forecast accuracy metric by comparing accuracy of one forecast model to another model to provide a metric of relative skill. This other model is usually a null model, such as a climatology or persistence model (Jolliffe and Stephenson 2012; Pappenberger et al. 2015). Climatology null models, which make predictions based on the mean conditions for a given day over a historical period of data, often represent the “best guess” prediction of an ecosystem variable based on existing observations, and can provide useful information about how an ecosystem’s current conditions compare to historical patterns (Jolliffe and Stephenson 2012).

One factor which is likely to influence forecast performance of water temperature and oxygen is the complexity of an observational time series, with higher complexity resulting in lower forecast accuracy (Pennekamp et al. 2019). Complexity, or the amount of information contained within a time series after each permutation (Bandt and Pompe, 2002), is typically negatively correlated with forecast accuracy (Pennekamp et al. 2019), although the effect of observational complexity on forecast skill remains unknown. Forecast skill, which measures the additional information provided by a forecast over an uninformed null model’s predictions, may show a different relationship as it depends on both the forecast accuracy *and* the null model accuracy. To observe how changes in observational complexity affect forecast accuracy and skill, complexity can be measured on any individual time series and varies over time, space, and variable of interest (Bandt and Pompe, 2002).

To determine how forecast performance varied across time, space, and ecosystem variable, we produced forecasts of two water quality variables (water temperature and dissolved

oxygen) at two depths over two evaluation periods in different years and quantified the observational complexity of each of these eight time series (the factorial combination of two variables, two depths, and two years). We forecasted water temperature and dissolved oxygen simultaneously, using a process-based model. Forecast performance was assessed using both forecast *accuracy* using scores in native units (i.e., °C for temperature and mg/L for oxygen), and forecast *skill* by comparing accuracy of our process-based forecasts to forecasts from a climatology null model. Our primary research questions were: 1) How does forecast accuracy and skill compare between temperature and oxygen?, and 2) How does temperature and oxygen forecast accuracy and skill vary over time, forecast horizon, and depth?

Methods

Forecasting Framework Overview

We developed forecasts of water temperature (hereafter ‘temperature’) and dissolved oxygen (hereafter ‘oxygen’), with uncertainty, throughout the water column at a single location at Lake Sunapee, New Hampshire, USA over two open-water seasons (2021 and 2022). Forecasts were made every 0.5 m from the surface to the bottom of the lake ($Z_{\max} = 33.0$ m) using the FLARE forecasting system, but assessed only at two depths where observations were available for temperature and oxygen. FLARE is an open-source forecasting system that uses a hydrodynamic model, the General Lake Model (GLM; Hipsey et al. 2019), coupled to an aquatic ecosystem-biogeochemistry model, Aquatic EcoDynamics (AED; Hipsey et al. 2022), to make near-term, iterative forecasts. Using National Oceanic and Atmospheric Administration (NOAA) weather forecasts as driver data into FLARE, we forecasted water temperature and oxygen in Sunapee from 1 to 35 days into the future. Sensor observations of temperature and oxygen were

assimilated when available, updating model parameter values and initial conditions throughout the forecast period (see *Methods: Study Site and Observational Data* for more details on data availability). While forecasts were made throughout the full water column, our forecast evaluation in this analysis was made at 1.0 m and 10.0 m only, where sensor observations were available for both water quality variables (see *Methods: Study Site and Observational Data* for information on observational data).

FLARE was used to produce daily forecasts via the following steps: 1) obtain new sensor observations from the field (Figure 2a.1); 2) access NOAA 35-day meteorological forecasts for the lake location (Figure 2a.2); 3) assimilate new observations with the previous day's forecast using an ensemble Kalman filter (EnKF; Evensen 2009) to update model initial conditions and parameters (Figure 2a.3); 4) generate a 1-35 day-ahead ensemble forecast (i.e., simulations with $n = 200$ ensemble members to quantify the uncertainty of future predictions), which are visualized and archived (Figure 2a.4); and 5) once time has passed and new sensor observations are available, evaluate the forecast (Figure 2a.5).

Study Site and Observational Data

Lake Sunapee is a deep (33.0 m), large (16.55 km² surface area), lake in central New Hampshire, USA (43.37, -72.05, Figure 3). Lake Sunapee is oligotrophic, with long-term mean pelagic summer total phosphorus (TP) concentrations of 5 µg/L, chlorophyll-a of 1.7 µg/L, and Secchi depth of 8.6 m (Steele et al. 2023). Lake Sunapee is dimictic, i.e., it is summer-stratified from May or June to September or October and is inversely-stratified under ice cover that typically lasts from December or January until March or April (Bruesewitz et al. 2015; LSPA

and Town of Sunapee 2022). The summer thermocline depth typically ranges from 6.0 - 8.0 m (Carey et al. 2014).

Discrete observations of temperature and oxygen have been collected for every meter from the surface to the bottom of the lake monthly during the open-water (i.e., ice free) season since 1986 as part of a long-term monitoring program by the Lake Sunapee Protective Association (LSPA), using methods outlined in Steele et al. (2023). These data were collected with hand-held sensors at the deepest site of the lake (Figure 3). To complement the monthly manually-collected data, a buoy instrumented with high-frequency sensors was deployed by the LSPA at a nearby site (Figure 3) in 2007, and is usually deployed annually from ~April or May until ~October, following ice-off and ice-on. The buoy has high-frequency (10 minute) water temperature sensors deployed every meter from 0.1 m to 10.0 m, and dissolved oxygen sensors at 1.0 m and 10.0 m (LSPA et al. 2023). Both temperature and oxygen sensors were cleaned fortnightly throughout the summer season, and calibrated using standard methods before deployment in summers 2021 and 2022. Buoy sensor observations (from the surface to 10.0 m) were assimilated into forecasts on a daily basis (Figure 3, buoy site), while the monthly deep site profiles (which included measurements at depths deeper than the buoy sensors at 10.0 m; Figure 3, deep site) were assimilated when available (approximately monthly from June to September). Data from the surface to 10.0 m at both the buoy site (LSPA et al. 2023) and the deep site (Steele et al. 2023) show high agreement (Pearson correlation $r = 0.93$; Wynne et al. 2023), so we used profiles from 0.1 to 10.0 m at the buoy and 12.0 m to 33.0 m from the deep site when available.

In our two study years, the timing of the open-water buoy deployment varied substantially due to maintenance (Table S1), so forecasts were only evaluated during the periods when observations were available for both variables (temperature and oxygen) and during both

years. Following a 35-day spin-up period (described in *Methods: FLARE Configuration*), our forecast evaluation time period was from 4 August to 17 October in 2021 and 2022 (see Table S1 for a description of data and forecast duration each year). As a result, our evaluation period excluded late fall, winter, and spring dynamics, which would likely exhibit differing levels of forecast performance due to changing seasonal drivers of temperature and oxygen. However, this summer to early fall period is one of the most dynamic time periods in Lake Sunapee (Carey et al. 2014), with substantial year-to-year variability, and thus provided a rigorous test of our forecasting system to capture ecosystem dynamics. Moreover, our focus on this time period facilitates comparison with many other limnological studies which are focused on the summer and early fall (Stanley et al. 2019). Because the thermocline in Lake Sunapee over our two study summers ranged from 6.0-8.0 m (Figure S1, see *Text S3* for a description of how thermocline depth was calculated), the 10.0 m forecasts are described as “bottom water” forecasts, and the 1.0 m forecasts are described as “surface water” forecasts throughout.

2.3 *FLARE Input Data*

FLARE requires observations of state variables (in this study, water temperature and dissolved oxygen) and forecasted driver variables (in this study, air temperature, shortwave and longwave radiation, windspeed, relative humidity, and precipitation). Observations of temperature and oxygen were taken from the buoy and deep sites during the study period (LSPA et al. 2023, Figure 3) to initialize and update the model states and parameters over time. Meteorological inputs (e.g., air temperature, wind speed, etc.) to drive the GLM-AED model were obtained from the NOAA GEFS (National Oceanic and Atmospheric Administration Global Ensemble Forecasting System) 35-day meteorological forecast product (Hamill et al.

2022). NOAA GEFS 35-day forecasts are produced every six hours and were downscaled from the 6-hour to 1-hour resolution to meet the required timestep of GLM within FLARE following the methods of Thomas et al. (2020).

During the 35-day spin-up period each year (described below in *Methods: FLARE Configuration*), we used historical, not forecasted, meteorological input data (described below) to drive the model, but used forecasted weather driver data in all forecast production periods. Meteorological forecasts were downloaded from the NOAA GEFS (0.5° X 0.5°) grid cell which contains Lake Sunapee. Because the required meteorological inputs to run FLARE are not collected at Lake Sunapee, we developed an estimate of the historical meteorology using the NOAA GEFS forecasts. Historical meteorological estimates were developed by taking the first time-step of each NOAA GEFS forecast and “stacking” them to produce an estimate of the historic conditions at Lake Sunapee. This stacked meteorological data product has been used in other forecasting studies and has been shown to be a good proxy for observed meteorological conditions (e.g., Thomas et al. 2023).

2.4 FLARE Configuration

We configured the FLARE forecasts for Lake Sunapee based on previous deployments of FLARE at other lakes and reservoirs (Thomas et al. 2020, 2023; Wander et al. 2023). While previous FLARE deployments for forecasting water temperature have used the GLM hydrodynamic model alone (Thomas et al. 2020, 2023; Wander et al. 2023), forecasting oxygen required the addition of the Aquatic EcoDynamics (AED) model library (Hipsey et al. 2022). AED provides a full suite of modules for predicting multiple freshwater ecosystem state variables that can be turned on or off depending on the application and required model

complexity (Hipsey et al. 2022). Because our focus was on temperature and oxygen and Lake Sunapee is an oligotrophic lake with low nutrient availability and corresponding low primary production (Richardson et al. 2017), we used a minimized version of AED with only the sediment flux, tracer, non-cohesive, and oxygen modules included (see Hipsey et al. 2022 for more information on these modules). These modules represented the main processes determining oxygen dynamics, including atmospheric fluxes (Figure 1c), sediment-water interface fluxes controlled by biological and chemical sediment oxygen demand (Figure 1d), and temperature-based solubility of oxygen (Figure 1e). FLARE was configured for Lake Sunapee without inflows or outflows using a mass balance approach due to the lake's long residence time (~3.1 years) relative to our longest forecast horizon (35 days). This approach has been shown to adequately simulate Lake Sunapee water budgets and temperature using a range of hydrodynamic models (Wynne et al. 2023), as well as in other lakes using FLARE (e.g., Thomas et al. 2023; Wander et al. 2023).

We configured GLM-AED to run with a set of parameters selected based on the sensitivity of the forecasts to parameter-fitting in previous studies (Thomas et al. 2020, 2023; Wander et al. 2023). The parameters were set for three distinct sediment zones within the water column, corresponding to 0-10 m, 10-18 m, and 18-33 m. Configuring FLARE to run with multiple sediment zones enabled better representation of the different rates that govern sediment water-interface temperatures and sediment oxygen demand within a sediment zone. A list of parameters that were included for automated fitting over time using the EnKF, as well as initial parameter configurations, is provided in Table S2. All configuration files are available at https://github.com/wwoelmer/SUNP-forecast-code/tree/glm_aed.

Prior to the first evaluated forecast, we ran a 35-day spin-up period each year to allow for data assimilation and parameter fitting. We used a full 35 days of spin-up in both years to allow for all forecast horizons to be represented within the time period, avoiding bias towards fitting only short-horizon forecasts. Additionally, we observed that parameters evolved through the full 35-day period, emphasizing the importance of having a spin-up period of this duration (Figure S2).

FLARE was configured to quantify four different sources of uncertainty: model initial conditions, model parameter, model process, and weather driver data (see Table S3 for a full description of each source and how it was quantified). All sources of uncertainty were included in each forecast run. For each source of uncertainty, each ensemble member was drawn from a distribution determined by the mean and standard deviation as described in Table S3. We ran FLARE with 200 ensemble members to ensure a reasonable spread of uncertainty around our forecasts (Thomas et al. 2020, 2023).

2.5 Climatology Null Model

We compared our process-based FLARE forecasts of temperature and dissolved oxygen to a null model forecast to calculate forecast skill. Specifically, we developed a climatology null forecasts of water temperature and dissolved oxygen over the same forecast period. Comparing FLARE-generated forecasts to null model forecasts allowed us to quantify the relative performance across the two focal variables and provided a quantifiable estimate of the additional forecast information obtained above the null (Dietze et al. 2018; Harris et al. 2018; Lewis et al. 2022). We calculated the climatology null model for a given day of year as the mean and standard deviation of observations across all available years of data from the high-frequency

buoy prior to our study (2007-2020) collected on that specific day. Only days with three or more observations across all years were included in the climatology null model, and only data through 2020 (i.e., before this study period began) were included in the null model (Figure S3).

2.6 Forecast Evaluation

We evaluated forecast performance using two metrics: forecast accuracy and forecast skill. First, we calculated forecast accuracy using the Continuous Ranked Probability Score (CRPS; Bröcker 2012), which is based on the absolute error between the observation and each forecast ensemble member, creating a weighted estimate of the absolute error from the full forecast distribution (Hyndman and Athanasopoulos 2021) as follows:

$$CRPS(F, x) = \int_{-\infty}^{\infty} (F(u) - \mathbb{1}(u - x))^2 du \quad (\text{Eq. 1})$$

Where CRPS is a function of F , the forecasted cumulative distribution function and x , the observed outcome; $\mathbb{1}$ is a step function which equals 1 if $u \geq x$ and 0 otherwise (Gneiting et al. 2015). Forecasts were scored using the `generate_forecast_score()` function in the FLAREr package (Thomas 2023). CRPS is in native units and provides a useful metric for decision-makers because it is a quantitative estimate of the distance between the forecast and the observation. We also calculated absolute error (AE) between the mean of the forecast distribution, but found similar patterns between the AE and CRPS so focused our reporting on CRPS within the main text. AE results are provided in Figure S4. Both FLARE forecasts and climatology null forecasts were scored using CRPS.

Second, we calculated forecast skill by normalizing CRPS to facilitate the comparison of CRPS between temperature in °C and oxygen in mg/L as follows:

$$\text{Forecast Skill} = 1 - \frac{CRPS_{FLARE}}{CRPS_{Clim}} \quad (\text{Eq. 2})$$

where $CRPS_{FLARE}$ is the CRPS calculated from a single FLARE forecast and $CRPS_{clim}$ is the CRPS of the climatology null model. This unitless skill metric allowed us to compare forecasts across variables, years, and depths and serves as an indicator of the additional information content gained from the FLARE forecasts compared to the climatology model. For this forecast skill metric, values of 0 indicate that the FLARE forecast and climatology forecast were equally skillful, values above 0 indicate FLARE was more skillful than the null, and values below 0 indicate that FLARE was less skillful than the null.

Another important component of forecast skill is the degradation in forecast skill into the future (i.e., across the forecast horizon). We calculated forecast skill degradation by taking the difference between the maximum and minimum forecast skill across the 1-35 day forecast horizon. We used the maximum and minimum forecast skill, as opposed to the difference in skill between 35-day and 1-day horizons, to capture the greatest possible degradation in skill. Lastly, to capture the variability in forecast skill over each open-water period, we calculated the percentage of FLARE forecasts which outperformed the climatology null forecast within each evaluation period, based on raw CRPS scores.

2.7 Statistical analysis

To measure the complexity of an observational time series, we calculated permutation entropy (PE), with higher PE corresponding to greater complexity of a time series. We calculated PE across each of our $n=8$ observational time series (the factorial combination of two years, two depths, and two variables). Calculating PE allowed us to determine how forecast performance related to complexity in the observational time series being forecasted. While forecast accuracy metrics are typically negatively correlated to PE (Pennekamp et al. 2019), the relationship

between PE and forecast skill across water quality variables, space (i.e., depth within a lake), and time (i.e., across years) remains unknown. We calculated normalized PE using the ‘statcomp’ package in R (Sippel et al. 2019), with the ‘permutation_entropy()’ function, and an embedding distance of 4 (following Bandt and Pompe (2002), who suggest values between 3 and 7), within the ‘ordinal_pattern_distribution()’ function. We also calculated weighted PE and found that its magnitude was smaller than normalized PE, but both metrics exhibited similar patterns across years in each evaluation period and depths (Figure S5), so we report normalized PE instead of weighted PE throughout the manuscript. All statistical analyses and forecast deployments were conducted in the R statistical environment, version 4.2.2 (R Core Team 2022). Code to run the forecasts and reproduce the forecast analysis is located at https://github.com/wwoelmer/SUNP-forecast-code/tree/glm_aed. All forecast output is archived in the Zenodo repository (Woelmer et al. 2023).

Results

Inter-annual variability of observations was greater for oxygen than temperature, but variability with depth was greater for temperature than oxygen

Observations of oxygen differed more between the evaluation periods in the two study years (hereafter, years) than water temperature, but temperature exhibited greater differences across depth (Figure 4). Specifically, mean water temperature at 1.0 m was similar between years (Figure 4a, c; 21.4°C in 2021 and 21.3°C in 2022), although more variable in 2022 than 2021 (Figure 4a, c; 2022 range = 12.3°C vs. 2021 range = 9.3°C). In contrast, 10.0 m temperature in 2022 was warmer and more variable (Figure 4f, h; mean = 13.7°C, range = 8.0°C) than in 2021

(Figure 4f, h; mean = 12.7°C, range = 7.6°C). Averaged across the two evaluation periods, temperature at 1.0 m during the forecast evaluation period was 6.6°C warmer than 10.0 m.

In comparison to temperature, differences in oxygen were generally greater between years than between depths (Figure 4b, d, g, i). Oxygen at 1.0 m was slightly lower in 2022 (Figure 4d; mean = 8.1 mg/L) than in 2021 (Figure 4d; mean = 9.1 mg/L), but both years showed similar trends (Figure 4b), with a general increase in concentrations throughout the evaluation period (August – October). At 10.0 m, oxygen concentrations were much more variable in 2021 (range = 4.09 mg/L) than in 2022 (range = 1.67 mg/L), likely caused by a greater decrease in concentrations through the summer until turnover in 2021 followed by large variations in oxygen, which did not occur in 2022 (Figure 4g). Overall, the mean concentration of oxygen at 10.0 m was higher in 2022 (Figure 4i; 9.6 mg/L) than 2021 (9.1 mg/L), but 2021 had a much larger range of concentrations (4.8 mg/L) compared to 2022 (1.9 mg/L; Figure 4i). Averaged across the two years, mean oxygen concentrations were similar between depths (1.0 m = 8.6 mg/L; 10.0 m = 8.9 mg/L). Patterns of oxygen in percent saturation were similar to patterns in mg/L (Figure S6); correspondingly, we present results in mg/L throughout the manuscript for ease of interpretation.

Observational complexity, measured as PE, was higher in 2021 than 2022 for all variable-depth combinations (Figure 4e, j). Additionally, PE was generally higher for oxygen than temperature, and increased with depth for both variables (Figure 4e, j). Specifically, oxygen had higher PE than temperature for all year-depth combinations, except for 1.0 m observations in 2021, when temperature (PE = 0.78) had a slightly higher PE than oxygen (PE = 0.71; Figure 4e, j). Across years, PE was higher in 2021 than 2022 for all variable-depth combinations, but within a year, PE was always higher at 10.0 m than 1.0 m (Figure 4e, j).

Water temperature forecasts were more skillful than dissolved oxygen

Forecast accuracy across all years and depths was $<1.1^{\circ}\text{C}$ for temperature and <0.65 mg/L for oxygen (Figure 5a, b). Both variables exhibited nonlinear decreases in forecast accuracy across the 35-day forecast horizon (Figure 5a, b). Temperature forecasts had a mean CRPS of 0.27°C at 1-day ahead, decreasing to 0.53°C at 7 days, 0.93°C at 21 days, and 1.08°C at the end of the 35-day forecast horizon (Figure 5a). Oxygen forecasts similarly had a CRPS of 0.27 mg/L at 1-day ahead, decreasing to 0.48 mg/L at 7 days, with accuracy leveling out around 21 days with a CRPS of ~ 0.65 mg/L through to the end of the 35-day forecast horizon (Figure 5b).

Aggregated across all years and depths, temperature forecast skill was higher than oxygen forecasts across all horizons (Figure 5c). Forecast skill overall decreased over the forecast horizon for both variables, but the strength of this trend varied between water quality variables (Figure 5c). While forecast skill of oxygen (mean skill = -0.6) was overall worse than for temperature (mean skill = -0.08), the rate at which skill degraded over the 35-day forecast horizon was slower for oxygen and saturated at ~ 20 days into the future, when temperature showed a near linear decline across the 35-day horizon (Figure 5c). However, forecasts of both temperature and oxygen were less skillful than the null at longer forecast horizons. Specifically, aggregated temperature forecasts outperformed the climatology null at more horizons than oxygen, with temperature being more skillful up to 11 days into the future, while oxygen was more skillful than the null only 2 days into the future (Figure 5c).

Forecast performance varied by year and decreased with depth

Across years and depths, FLARE forecast accuracy varied by year and generally decreased with depth (Figure 6). For temperature, forecasts had lower CRPS scores (i.e., more accurate) in 2021 than 2022 at 1.0 m (2021 mean = 0.57 °C; 2022 mean = 0.80 °C; Figure 6a). Forecast accuracy was also more variable in 2022 than in 2021 at 1.0 m (Figure 6a). At 10.0 m, forecasts of temperature were better in 2022 (mean = 0.73 °C) than 2021 (mean = 0.92 °C; Figure 6c).

Across the forecast horizon, temperature accuracy at 1.0 m degraded more quickly in 2021, leveling out around ~15 days-ahead, while forecasts in 2022 continued to decrease in accuracy throughout the 35-day horizon (Figure 6a). This resulted in an overall increase in CRPS of 0.56 °C for 2021 and 1.1 °C for 2022 from 1- to 35-days-ahead. For both years, forecasts of temperature at 10.0 m continued to decrease in accuracy over the forecast horizon (2021 decrease = 0.9 °C; 2022 decrease = 0.68 °C). Variability across forecasts was similar between years at 10.0 m (Figure 6c). Overall, temperature forecasts were slightly worse at 10.0 m than at 1.0 m, with a decrease from 0.69°C at 1.0 m to 0.83°C at 10.0 m across all horizons and years (Figure 6a, 6c).

Accuracy across the forecast horizon was similar between years for oxygen, with forecasts in 2022 (mean = 0.49 mg/L) only slightly more accurate than forecasts in 2021 (mean = 0.43 mg/L) at 1.0 m (Figure 6b) but both years had similar accuracy at 10.0 m (2021 mean = 0.67 mg/L; 2022 mean = 0.65 mg/L; Figure 6d). Across depths, forecasts at 1.0 m degraded across the forecast horizon much more quickly than at 10.0 m, with 1.0 m forecasts leveling out around 15-days into the future with an overall decrease in accuracy of 0.35 mg/L in 2021 and 0.23 mg/L in 2022 (Figure 6b), while 10.0 m forecasts continued to worsen throughout the 35-

day forecast horizon (2021 decrease = 0.36; 2022 decrease = 0.52 mg/L; Figure 6d). Variability in oxygen forecast accuracy over the study period (i.e., the standard deviation in accuracy from August-October) was similar at 1.0 m (Figure 6b), but 10.0 m forecasts in 2021 had a much larger range in forecast accuracy at horizons shorter than 10 days-ahead than in 2022 (Figure 6d). As with temperature, accuracy of oxygen forecasts decreased with depth from 0.46 mg/L at 1.0 m to 0.66 mg/L at 10.0 m (Figure 6b, 6d). These patterns of forecast degradation across the 1-35-day horizon were generally consistent over time, with some differences across day of year but did not show any significant pattern (Figure S7).

Forecast skill (accuracy relative to a climatology null model) was different between years across all variables, and generally decreased with forecast horizon and depth (Figure 7). For temperature, forecast skill was higher in 2021 than 2022 across both depths (Figure 7a, e), with forecasts more skillful than the null up to 11 days into the future at 1.0 m and 24 days into the future at 10.0 m in 2021. Across all horizons, more than 50% of temperature forecasts performed better than the null in 2021 for both depths (Figure 7c, g). In comparison, 2022 forecasts at 1.0 m were more skillful than the null up to 8 days ahead and 11 days ahead at 10.0 m (Figure 7c). By 35 days ahead, <25% of temperature forecasts were better than the null at both 1.0 m and 10.0 m in 2022 (Figure 7c, g).

The difference in forecast skill between years varied by depth for oxygen. At 1.0 m, oxygen forecast skill was higher in 2022, with forecasts at all horizons (35 days) more skillful than the null, where no horizons in 2021 were more skillful than the null for oxygen forecasts (Figure 7b). Specifically, oxygen forecasts at 1.0 m in 2022 were better than the null 64-97% of the time, ranging across the horizon, but only 0-24% of the time in 2021 (Figure 7d). In comparison, median oxygen forecast skill at 10.0 m was similar between years with skillful

forecasts only up to 2 days into the future for both 2021 and 2022 (Figure 7f). Across the forecast horizon, forecasts were slightly better in 2021, with 54% of forecasts beating the null at 35-days-ahead in 2021, as opposed to only 10% of forecasts in 2022 (Figure 7h). Higher forecast skill in 2022 than 2021 can be attributed to substantially higher accuracy of the null model in 2022 (Figure S8), rather than poor performance of FLARE forecasts, which were generally similar between 2021 and 2022 (Figure 6b, d). Overall, the range in forecast skill across years was much greater in 2021 than 2022 at both depths, but especially at 10.0 m (Figure 7b, 7f).

Years with greater PE generally had lower forecast accuracy but were more skillful

Forecast accuracy was lower in all years with higher PE for all variable-depth combinations except temperature at 1.0 m (Figure 6 compared with Figure 4e, j). Specifically, forecasts were less accurate in 2021 than 2022 for oxygen at both 1.0 m and 10.0 m, as well as temperature at 10.0 m.

Despite lower forecast accuracy in years with greater PE (Figure 6), FLARE forecasts were generally more skillful in years with greater PE (Figure 7). Specifically, forecasts in 2021 were more skillful than forecasts in 2022 at all depths and for all variables except for oxygen at 1.0 m (Figure 7b, 7d). More skillful forecasts in 2021 for most cases corresponded to a higher PE of observations than in 2022 for all forecasts (Figure 4e, j).

Discussion

Overview

We produced forecasts of water temperature and dissolved oxygen using a process-based model with daily data assimilation. Temperature forecasts were more skillful than oxygen

forecasts overall. Specifically, temperature forecasts were more skillful than a climatology forecast 11 days into the future and oxygen forecasts 2 days into the future. We found forecast skill varied between years, and typically decreased with depth for both variables. Generally, surface forecasts were more skillful than bottom water forecasts, especially for oxygen. While accuracy of forecasts was generally worse with greater observational complexity, forecast skill relative to a climatology model was generally higher in years with more complex observations. Our results highlight that forecast performance can differ substantially across variables and over time and depth, and that null models are critical for contextualizing process-based forecast skill.

Relative skill of temperature and oxygen forecasts

We found that water temperature forecasts were more skillful than oxygen forecasts, following expectations that physical variables are more predictable than ecologically influenced (i.e., chemical and biological) variables (Figure 5c; Soares and Calijuri 2021). However, contrary to expectations, temperature skill degraded more rapidly across the forecast horizon than oxygen skill (Figure 5c). Our results suggest that forecasts of oxygen at longer forecast horizons (e.g., at 35-days into the future) are relatively more skillful than temperature forecasts at the same horizon for our forecasting application in Lake Sunapee. More work is needed to evaluate how temperature and oxygen forecast skill degrades with horizon at other times of year outside of our August-October evaluation period to determine the consistency of this result.

Despite that we found temperature to be more predictable than oxygen, we note that predictability is likely to vary by lake ecosystem. Specifically, because Lake Sunapee is an oligotrophic lake, with relatively small changes in oxygen dynamics within a year (Fig. S13; Richardson et al. 2017), oxygen forecast skill may be higher in this system than in other lakes.

Forecast skill of oxygen may be lower in eutrophic lakes which experience algal blooms and as a result can exhibit large changes in dissolved oxygen concentrations over short-term time scales (e.g., Ladwig et al. 2021).

Mechanisms for variability in forecast skill over time and space

Forecast performance of water temperature and dissolved oxygen varied between years, demonstrating that predictability of ecosystem variables can vary inter-annually. One possible explanation for inter-annual variability is differences in weather conditions, but because we did not see substantial differences in the overall magnitude or variability of the meteorological variables which drive FLARE between years (Figure S9), other processes may be influencing inter-annual predictability. In our study years, we did see differences in the timing and magnitude of rain events, including two large events in September 2022 that likely led to fall turnover in this year, which did not occur in 2021 (Figure S10). Although preliminary, this may indicate that differences in the timing of specific weather events can also influence predictability, as opposed to overall variability alone within the year. Regardless, global change will continue to alter inter- and intra-annual weather patterns, corresponding to more variable physical (Woolway and Merchant 2019; Woolway et al. 2019, 2021; Sharma et al. 2021) and chemical (Carey 2023) metrics of water quality, with important implications for the performance of lake water quality forecasts. Our results on inter-annual variability in temperature and oxygen predictability likely further extends to other important water quality variables (e.g., phytoplankton community composition and densities, fish population densities), indicating that forecast performance in one year may not be indicative of future performance.

Across temperature and oxygen, bottom water forecasts showed lower performance than surface water forecasts, at least up to 10.0 m deep. Our 10.0 m forecasts are below the thermocline in Lake Sunapee (Figure S2) and representative of the maximum depth where the buoy is located. However, Lake Sunapee is 33.0 m deep, so it is possible that our 10.0 m forecasts are not representative of dynamics at the very deepest part of the lake. Based on limited historical data comparison (Figure S11), temperature and oxygen dynamics at 10.0 m at the buoy site closely follow patterns at 15.0 m and 20.0 m at the deepest site ($R^2 = 0.76, 0.6$, respectively), but with a decrease in similarity compared to 30.0 m observations ($R^2 = 0.3$, Figure S11), although the water column below 30.0 m represents a very small proportion of the overall lake volume. These comparisons provide confidence that 10.0 m forecasts are representative of at least most of the hypolimnion in Lake Sunapee. However, this comparison is limited ($n = 180$ observations). Thus, it is possible that our finding that forecast performance decreased with depth may have been influenced by the location of our 10.0 m sensor in the upper hypolimnion. In other studies examining forecast performance with depth, a majority of studies generally found that bottom water forecasts have higher performance than surface waters for both temperature and oxygen (Saber et al. 2020; Thomas et al. 2020; Durell et al. 2023; Wander et al. 2023), with the exception of Lin et al. (2023), who found that there was little difference in oxygen forecast performance between surface and bottom layers. Our results suggest that there may be differences in predictability from surface to bottom based on waterbody, and that more examination is needed to better understand this trend across ecosystems.

Interestingly, decreases in forecast performance from the surface to the bottom were smaller for temperature than oxygen. This pattern could be due to limited representation of processes in our model configuration for forecasting deep-water oxygen dynamics. Specifically,

while our model configuration did dynamically fit a sediment oxygen flux parameter, which incorporates both biological and chemical oxygen demand at the sediments (Hipsey et al. 2022), we did not simulate the dynamics of phytoplankton and other solutes that likely altered hypolimnetic oxygen cycling. In addition, differences in inter-annual variability between oxygen and temperature may also have impacted differences in performance across depth. For example, observations of oxygen in 2021 and 2022 were anomalous from previous years (Figure S12), potentially leading to worse performance, but temperature observations were similar to historical observations (Figure S13), potentially leading to better predictions.

Implications of differential forecast skill and accuracy for forecast use and interpretation

Forecast accuracy and skill showed opposite patterns, which has important implications for how we interpret and contextualize forecast assessment holistically. These differences were especially distinct for oxygen, for which accuracy (assessed by CRPS) was quite similar between years (Figure 6), but skill (accuracy relative to a climatology forecast) was very different (Figure 7). Forecast skill integrates both FLARE forecast accuracy and climatology forecast accuracy (Joliffe and Stephenson 2012), and as such, differences in climatology accuracy can explain differences in FLARE forecast skill across years. For example, FLARE forecasts of oxygen at 1.0 m had high accuracy in 2022 (Figure 6b), but climatology forecasts had low accuracy (Figure S8), due to 0.6 mg/L lower concentrations of oxygen in 2022 relative to the historical 2007-2022 average (Figure S12 vs. Figure 6). In contrast, both FLARE and climatology forecasts of oxygen at 1.0 m had relatively high accuracy in 2021 (Figure 6b and Figure S8), but climatology forecasts were more accurate, resulting in a large proportion of FLARE forecasts being unskillful relative to the climatology forecasts.

Accuracy and skill each assess different components of forecast performance. For example, forecast skill is an intuitive measure of how well two different forecasts perform relative to each other, providing a quantification of how much more information is gained by one forecast over another (Jolliffe and Stephenson 2012). As such, assessing forecast skill may be especially useful for scoring forecasts in ecosystems which have shifting baselines, meaning historical estimates may no longer be as good at predicting the future (Pauly 1995, Daugaard et al. 2022) and indicate how much additional information can be gained from process-based model forecasts over climatology forecasts. Given the widespread impact of climate change on ecosystems globally (Bruggemann et al. 2012; Sydeman et al. 2013; Mariani et al. 2022; IPCC 2023), evaluating forecast skill relative to uninformed models provides an opportunity to robustly assess how much more information our process-based forecasts provide.

In addition, unitless skill scores allow for a quantitative comparison of forecast variables which are not in the same native units, and helps to broaden our understanding of the fundamental predictability of ecosystems. However, it is important to note that interpretation of the functional utility of forecasts can also be clouded by the use of a skill score (Wheatcroft 2019), as it is not in meaningful units familiar to managers for decision-making. Evaluating forecast performance in native units allows a user to interpret the performance of the forecast relative to their decision-making needs (e.g., we can predict oxygen within 0.6 mg/L). Altogether, incorporating both forecast accuracy and forecast skill may be more effective than a single metric when assessing forecast performance, depending on how the forecast is being used.

Overall, quantifying both the accuracy and skill of forecasts can contribute complementary information for decision-making. For example, decisions regarding specific details about an ecosystem variable likely require information provided by forecast accuracy

(e.g., how likely are hypoxic conditions over the next week, or how closely can we predict water temperature at 1.0 m). In contrast, forecast skill provides different information which is more relevant for understanding how ecosystem predictability has changed over time (e.g., how much better can we predict algal concentrations over the historical mean on this day?), or the performance across forecast models (e.g., which forecast model is more accurate in a given year or at a specific ecosystem?).

Effect of observational complexity on forecasting

Overall, we found that forecasts generally had higher skill, but lower accuracy, in years with greater observational complexity (i.e., greater permutation entropy). Improved skill in years which would otherwise be considered lower in accuracy provides a critical demonstration of the value of comparing process-based forecasts to forecasts generated by a null model over time. It is common for null models to outperform ecological forecasts under some conditions (e.g., Page et al. 2018; Mercado-Bettín et al. 2021; Woelmer et al. 2022; Wheeler et al. 2023). Our analysis adds context to this generalization, demonstrating that when observational complexity was higher, the null model was less likely to perform well, putting more value on the process-based predictions of our FLARE forecasts. This is likely due to the ability of process-based models to represent changes in state variables due to mechanistic representation of changing processes or more accurate representations of the current state of the ecosystem via data assimilation. In contrast, in years with high complexity, climatology forecasts were unable to simulate changing dynamics which varied from historical patterns. For example, oxygen at 10.0 m in 2021 had especially high observational complexity (high PE; Figure 4j), leading to relatively low accuracy of FLARE forecasts (Figure 6d) but higher skill, especially compared to 2022 with lower

complexity (Figure 7h). This may be due to FLARE's ability to represent the changing thermal dynamics throughout the water column due to weather drivers, whereas the climatology forecast was unable to represent these complexities. However, there were exceptions to the relationship between complexity and forecast performance (e.g., oxygen at 1.0 m; Figure 7), which is likely a result of the balance between FLARE accuracy and climatology accuracy (see *Discussion: Implications of differential forecast skill and accuracy for forecast use and interpretation*).

The use of PE as a metric for measuring observational complexity has been adopted in many fields, and has even been integrated into modeling approaches to improve predictability (Zhu et al. 2016; Jiang et al. 2019; Ruiz-Aguilar et al. 2021; Sudibyo et al. 2022; Yue et al. 2022). For example, forecasts have successfully predicted volcanic eruptions by calculating PE on a rolling basis and identifying an increase in PE as an early warning indicator of a major ecosystem shift (following Sudibyo et al. 2022). Additionally, forecast studies have used PE to identify specific patterns in observed data which are then used to train models and decompose trends in observations which are directly related to increased forecast performance (following Zhu et al. 2016; Jiang et al. 2019; Ruiz-Aguilar et al. 2021; Yue et al. 2022). Given the success of this method in improving predictive performance in other studies, the potential for improving forecasting techniques through the use of PE is high. Ultimately, methods which incorporate underlying observational complexity within forecast models may have the ability to improve predictions in the future.

Use of process-based modeling for forecasting water quality variables simultaneously

Our study is one of the first lake forecasting applications to include both temperature and oxygen dynamics simulated by the same process-based model, highlighting the value of this

approach to ensure that the interconnected ecosystem processes governing these two variables are integrated into forecasts. First, process-based models can represent the entire water column and simulate multiple processes that control interactions between state variables (e.g., temperature and oxygen) at each depth layer. Second, process-based models can simulate depths without observations, thus overcoming the limitations of data-driven models that require intensive datasets to make forecasts. Additionally, using a process-based lake ecosystem/water quality model such as GLM-AED provides the opportunity to add additional biogeochemical variables in the future through the inclusion of other modules (e.g., phytoplankton, nutrients, carbon, zooplankton dynamics). Lastly, simultaneously forecasting multiple water quality variables is especially valuable for managers who may need to optimize many different metrics of water quality when making holistic management decisions (following Jackson-Blake et al. 2022).

Our FLARE process-based forecasts showed high skill relative to a climatology forecast across most years, depths, and variables (Figure 7). This high skill is likely due to the ability of the process-based forecasts to represent changing dynamics in the lake ecosystem via improved process representation of important variables which influence temperature and oxygen that climatology models inherently cannot simulate. Alternatively, the improved FLARE skill could also be due to data assimilation. However, we note that forecast skill depends on the accuracy of both FLARE forecasts and the baseline model of choice, and is likely to vary based on the baseline model chosen (Murphy 1992; Mittermaier 2008). For example, while a climatology model may be a “best guess” forecast for an ecosystem with historical monitoring but no real-time data availability, a null persistence model (i.e., random walk) may provide more information due to the incorporation of autocorrelation (Hyndman and Athanasopoulos 2021).

Additionally, our climatology model is limited by the historical extent of our data (2007-2020), which may not be the most representative years for the current state of the system. However, our results indicate that as climate change continues to “push” freshwater ecosystems from previously observed ecosystem states (Janse et al. 2015; Chapra et al. 2017; Woolway and Merchant 2019; Woolway et al. 2022; Desgué-Itier et al. 2023), process-based models may become more valuable than climatology models alone.

The effect of using a process-based model, as opposed to empirical or machine learning models, on forecast performance remains unclear across variables. For water temperature, Ouellet-Proulx et al. (2017) also used a process-based model to simulate water temperature in a river system and had similar forecast accuracy (CRPS of 0.8°C up to 5-days ahead compared to our 5-day CRPS of 0.4°C); however, we note that differences in ecosystem type (e.g., river vs. lake) may also impact predictability. Similarly, other oxygen forecast studies (e.g., Saber et al. 2020; Durell et al. 2022, 2023; Lin et al. 2023) do not use the same forecast evaluation metrics as this study, limiting our ability to directly compare predictive performance, a common issue across freshwater forecasts (Lofton et al. 2023). Overall, developing standardized forecast evaluation metrics is greatly needed to enable cross-study inference (following Dietze et al. 2021).

Opportunities for expanding our understanding of water quality predictability

Results of this study point to important gaps in our understanding of predictability of freshwater forecasts broadly. First, oxygen dynamics are driven by both abiotic and biotic processes and while hydrodynamic models do well at predicting abiotic dynamics, predicting many biogeochemical processes remains a challenge. Lake Sunapee is an oligotrophic lake with

historical annual NEP (net ecosystem production), or the balance between primary production and respiration, ranging from -1 to $<1 \text{ mg O}_2 \text{ L}^{-1} \text{ day}^{-1}$ (Richardson et al. 2017), and low overall phytoplankton levels (historical mean chlorophyll-a = $1.7 \text{ }\mu\text{g/L}$; Ward et al. 2020). However, we emphasize the importance of using models that represent critical biotic processes such as epilimnetic primary production, respiration, and decomposition in other oxygen forecasting studies, especially in lakes with greater productivity. Second, both data collection limitations, as well as long data latency (i.e., the time lag to when new observations are integrated into forecasting workflows) remain as roadblocks for generating additional biogeochemical forecasts. While process-based models, such the one used in this study (GLM-AED), do include additional biogeochemical modules to improve the representation of biotic processes, in our case, limitations of phytoplankton, nutrients (e.g., nitrogen and phosphorus), and organic carbon data precluded us from including those state variables in real-time forecasting. Third, while process-based models are well-positioned for adding additional forecasted variables, including additional state variables and parameters can come at a cost, both in an inability to calibrate the model for predicting many variables (Hipsey et al. 2020), as well as potential predictive tradeoffs. Issues of parameter identifiability and over-fitting (Luo et al. 2009) especially need to be considered in forecasting frameworks which use sequential data assimilation techniques (e.g., EnKF as used in this study) to tune multiple parameters at once to multiple variables.

Conclusions

Overall, we show that forecasts of water temperature and dissolved oxygen were variable between years, demonstrating that predictability of important water quality variables for decision-making can vary substantially over time in a large oligotrophic lake. This was true

despite similar weather patterns between years, indicating that other factors, such as observational complexity, are influencing predictability. Across variables, temperature forecasts were more skillful than oxygen forecasts in Lake Sunapee, following expectation that physical variables are more predictable than chemical variables (Soares and Calijuri 2021). Finally, forecast performance was variable depending on the forecast evaluation metric used, with lower accuracy but higher skill in years with greater observational complexity. This finding emphasizes that process-based forecasts provide important information above a climatology forecast, which may continue to be more valuable over time as ecosystems are pushed further outside of historical ranges. Altogether, producing forecasts of temperature and oxygen using a process-based model provides novel insight into how forecast quality varies over time and space, especially as freshwater ecosystems continue to experience global change stressors.

Acknowledgements

We gratefully acknowledge the Lake Sunapee Protective Association, especially June Fichter, Elizabeth Harper, Geoff Lizotte, Teriko MacConnell, Sue Godin, Susie Burbidge, Nancy Brook Heckel, and Kathleen Stowell, as well as members of the Lake Sunapee community, especially Dave and Barbara Calhoun, Tim and Midge Eliassen, and John Merriman for enabling the long-term collection of data at Lake Sunapee and for their continued support of science and research at Lake Sunapee. We thank Jacob Wynne for many hours of field work, driving, and diving which enabled this research. We thank the CIBR FLARE Research team and Virginia Tech Reservoir, especially Abby Lewis, for many thoughtful discussions and helpful feedback. We thank the Ecological Forecasting Initiative (EFI), the EFI Student Association, and the Global Lake Ecological Observatory Network for useful feedback on this study, as well as forecasting and lake science more broadly. This work was funded by the Calhoun LSPA-Virginia Tech Fellowship, National Science Foundation (NSF) research grants CIBR-1933016 and CIBR-1933102, and the NSF Graduate Research Fellowship.

Authorship contributions

WMW and CCC co-developed the design of this study. RQT developed the FLARE framework, enabling forecast production. RQT and FO contributed to conceptual development, which substantially improved the quality of the manuscript. WMW, CCC, KCW, and BGS curated data from Lake Sunapee. KCW and BGS enabled access to long-term Lake Sunapee data. WMW led manuscript writing and figure development, with substantial help from CCC, FO, and RQT. All coauthors have provided feedback, edited, and approved the final version.

References

- Bandt C, Pompe B (2002) Permutation Entropy: A Natural Complexity Measure for Time Series. *Physical Review Letters* 88:4. <https://doi.org/10.1103/PhysRevLett.88.174102>
- Bhateria R, Jain D (2016) Water quality assessment of lake water: a review. *Sustain Water Resour Manag* 2:161–173. <https://doi.org/10.1007/s40899-015-0014-7>
- Bodner K, Rauen Firkowski C, Bennett JR, et al (2021) Bridging the divide between ecological forecasts and environmental decision making. *Ecosphere* 12(12): e03869 <https://doi.org/10.1002/ecs2.3869>
- Bröcker J (2012) Evaluating raw ensembles with the continuous ranked probability score. *Quarterly Journal of the Royal Meteorological Society* 138:1611–1617. <https://doi.org/10.1002/qj.1891>
- Bruesewitz DA, Carey CC, Richardson DC, Weathers KC (2015) Under-ice thermal stratification dynamics of a large, deep lake revealed by high-frequency data - Bruesewitz - 2015 - *Limnology and Oceanography* - Wiley Online Library. *Limnology and Oceanography* 60:347–359. <https://doi.org/DOI:10.1002/lno.10014>
- Bruggemann JH, Rodier M, Guillaume MMM, et al (2012) Wicked social-ecological problems forcing unprecedented change on the latitudinal margins of coral reefs: The Case of southwest Madagascar. *Ecology and Society* 17(4). <https://doi.org/10.5751/ES-05300-170447>
- Caffrey JM (2004) Factors controlling net ecosystem metabolism in U.S. estuaries. *Estuaries* 27:90–101. <https://doi.org/10.1007/BF02803563>

- Calamita E, Vanzo D, Wehrli B, Schmid M (2021) Lake Modeling Reveals Management Opportunities for Improving Water Quality Downstream of Transboundary Tropical Dams. *Water resources Research* 57:e2020WR027465.
<https://doi.org/10.1029/2020WR027465>
- Carey C (2023) Causes and consequences of changing oxygen availability in lakes. *Authorea Preprint*. <https://doi.org/10.22541/essoar.167569635.57065824/v1>
- Carey CC, Cottingham KL, Weathers KC, et al (2014) Experimental blooms of the cyanobacterium *Gloeotrichia echinulata* increase phytoplankton biomass, richness and diversity in an oligotrophic lake. *Journal of Plankton Research* 36:364–377.
<https://doi.org/10.1093/plankt/fbt105>
- Carey CC, Woelmer WM, Lofton ME, et al (2022) Advancing lake and reservoir water quality management with near-term, iterative ecological forecasting. *Inland Waters* 12:107–120.
<https://doi.org/10.1080/20442041.2020.1816421>
- Clayer F, Jackson-Blake L, Mercado-Bettín D, et al (2023) Sources of skill in lake temperature, discharge and ice-off seasonal forecasting tools. *Hydrology and Earth System Sciences* 27:1361–1381. <https://doi.org/10.5194/hess-27-1361-2023>
- Daugaard U, Munch SB, Inauen D, et al (2022) Forecasting in the face of ecological complexity: Number and strength of species interactions determine forecast skill in ecological communities. *Ecology Letters* 25:1974–1985. <https://doi.org/10.1111/ele.14070>

- Davis JC (1975) Minimal Dissolved Oxygen Requirements of Aquatic Life with Emphasis on Canadian Species: a Review. *J Fish Res Bd Can* 32:2295–2332.
<https://doi.org/10.1139/f75-268>
- Dietze MC (2017) *Ecological Forecasting*. Princeton University Press, Princeton
- Dietze MC, Fox A, Beck-Johnson LM, et al (2018) Iterative near-term ecological forecasting: Needs, opportunities, and challenges. *Proceedings of the National Academy of Sciences* 115:1424–1432. <https://doi.org/10.1073/pnas.1710231115>
- Dietze MC, Thomas RQ, Peters J, et al (2021) A Community Convention for Ecological Forecasting: Output Files and Metadata. *EcoEvoRxiv*.
<https://doi.org/10.32942/osf.io/9dgtq>
- Durell L, Scott JT, Hering AS (2023) Hybrid Forecasting for Functional Time Series of Dissolved Oxygen Profiles. *Data Science in Science* 2.
<https://doi.org/10.1080/26941899.2022.2152401>
- Durell L, Scott JT, Nychka D, Hering AS (2022) Functional forecasting of dissolved oxygen in high-frequency vertical lake profiles. 1–16. <https://doi.org/10.1002/env.2765>
- Evensen G (2009) *Data assimilation: the ensemble Kalman filter*. Springer, Berlin
- Geng M, Niu Y, Liao X, et al (2022) Inter-annual and intra-annual variations in water quality and its response to water-level fluctuations in a river-connected lake, Dongting Lake, China. *Environmental Science and Pollution Research* 29:14083–14097.
<https://doi.org/10.1007/s11356-021-16739-5>

Gneiting, T, Raftery AE, Westveld AH, and Goldman T (2005) Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Monthly Weather Review* **133**, 1098–1118, <https://doi.org/10.1175/MWR2904.1>.

Harris DJ, Taylor SD, White EP (2018) Forecasting biodiversity in breeding birds using best practices. *PeerJ* 2018:1–27. <https://doi.org/10.7717/peerj.4278>

Henden JA, Ims RA, Yoccoz NG, et al (2020) End-user involvement to improve predictions and management of populations with complex dynamics and multiple drivers. *Ecological Applications* 0:1–14. <https://doi.org/10.1002/eap.2120>

Hipsey MR, Boon C, Bruce LC, et al (2022) *AquaticEcoDynamics/glm-aed: v3.3.0*

Hipsey MR, Bruce LC, Boon C, et al (2019) A General Lake Model (GLM 3.0) for linking with high-frequency sensor data from the Global Lake Ecological Observatory Network (GLEON). *Geoscientific Model Development*; Katlenburg-Lindau 12:473–523. <http://dx.doi.org/10.5194/gmd-12-473-2019>

Hipsey MR, Gal G, Arhonditsis GB, et al (2020) A system of metrics for the assessment and improvement of aquatic ecosystem models. *Environmental Modelling & Software* 128:104697. <https://doi.org/10.1016/j.envsoft.2020.104697>

Ho JC, Michalak AM (2019) Exploring temperature and precipitation impacts on harmful algal blooms across continental U.S. lakes. *Limnology and Oceanography* 1–18. <https://doi.org/10.1002/lno.11365>

Hyndman RJ, Athanasopoulos G (2021) Forecasting: principles and practice

IPCC (2023) Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, 1st edn. Cambridge University Press

Jackson-Blake LA, Clayer F, De Eyto E, et al (2022) Opportunities for seasonal forecasting to support water management outside the tropics. *Hydrology and Earth System Sciences* 26:1389–1406. <https://doi.org/10.5194/hess-26-1389-2022>

Jassby AD, Reuter JE, Goldman CR (2003) Determining long-term water quality change in the presence of climate variability: Lake Tahoe (U.S.A.). *Can J Fish Aquat Sci* 60:1452–1461. <https://doi.org/10.1139/f03-127>

Jia X, Karpatne A, Willard J, et al (2018) Physics Guided Recurrent Neural Networks For Modeling Dynamical Systems: Application to Monitoring Water Temperature And Quality In Lakes. *arXiv Preprint*. <https://doi.org/10.48550/arXiv.1810.02880>

Jiang W, Zhou J, Xu Y, et al (2019) Multistep Degradation Tendency Prediction for Aircraft Engines Based on CEEMDAN Permutation Entropy and Improved Grey–Markov Model. *Complexity* 2019:e1576817. <https://doi.org/10.1155/2019/1576817>

Jin T, Cai S, Jiang D, Liu J (2019) A data-driven model for real-time water quality prediction and early warning by an integration method. *Environ Sci Pollut Res Int* 26:30374–30385. <https://doi.org/10.1007/s11356-019-06049-2>

- Jolliffe IT, Stephenson DB (eds) (2012) Forecast verification: a practitioner's guide in atmospheric science, 2. Wiley-Blackwell, Oxford
- Jones ID, Winfield IJ, Carse F (2008) Assessment of long-term changes in habitat availability for Arctic charr (*Salvelinus alpinus*) in a temperate lake using oxygen profiles and hydroacoustic surveys. *Freshwater Biology* 53:393–402. <https://doi.org/10.1111/j.1365-2427.2007.01902.x>
- Kim SK, Choi S-U (2021) Assessment of the impact of selective withdrawal on downstream fish habitats using a coupled hydrodynamic and habitat modeling. *Journal of Hydrology* 593:125665. <https://doi.org/10.1016/j.jhydrol.2020.125665>
- Kraemer BM, Pilla RM, Woolway RI, et al (2021) Climate change drives widespread shifts in lake thermal habitat. *Nature Climate Change* 11:521–529. <https://doi.org/10.1038/s41558-021-01060-3>
- Ladwig R, Hanson PC, Dugan HA, et al (2021) Lake thermal structure drives interannual variability in summer anoxia dynamics in a eutrophic lake over 37 years. *Hydrology and Earth System Sciences* 25:1009–1032. <https://doi.org/10.5194/hess-25-1009-2021>
- Lazer D, Kennedy R, King G, Vespignani A (2014) The Parable of Google Flu: Traps in Big Data Analysis. *Science* 343:1203–1205. <https://doi.org/10.1126/science.1248506>
- Lee C (2015) Oxidation of organic contaminants in water by iron-induced oxygen activation: A short review. *Environmental Engineering Research* 20:205–211. <https://doi.org/10.4491/eer.2015.051>

- Lee D-Y, Lee D-S, Cha Y, et al (2023) Data-driven models for predicting community changes in freshwater ecosystems: A review. *Ecological Informatics* 77:102163.
<https://doi.org/10.1016/j.ecoinf.2023.102163>
- Lewis ASL, Rollinson CR, Allyn AJ, et al (2022) The power of forecasts to advance ecological theory. *Methods in Ecology and Evolution* 14(3): 746-756. <https://doi.org/10.1111/2041-210X.13955>
- Li M, Kuang S, Kang Y, et al (2022) Recent advances in application of iron-manganese oxide nanomaterials for removal of heavy metals in the aquatic environment. *Science of The Total Environment* 819:153157. <https://doi.org/10.1016/j.scitotenv.2022.153157>
- Lin S, Pierson DC, Ladwig R, et al (2023) Multi-Model Machine Learning Approach Accurately Predicts Lake Dissolved Oxygen with Meteorological and Hydrological Input. SSRN Preprint. <https://doi.org/10.2139/ssrn.4454256>
- Lofton ME, Howard DW, Thomas RQ, Carey CC (2023) Progress and opportunities in advancing near-term forecasting of freshwater quality. *Global Change Biology* 29:1691–1714. <https://doi.org/10.1111/gcb.16590>
- LSPA (2023) Bathymetric data for Lake Sunapee, NH, USA ver 1. Environmental Data Initiative. <https://doi.org/10.6073/pasta/8a86710f25adeae3a2540b6cb14d7546>
- LSPA, Steele BG, Weathers KC (2023) Lake Sunapee Instrumented Buoy: High Frequency Water Quality Data - 2007-2022 ver 4. Environmental Data Initiative

LSPA, Sunapee Town of (2022) Ice-off dates for Lake Sunapee, Lake Sunapee Protective Association

Luo Y, Weng E, Wu X, et al (2009) Parameter Identifiability, Constraint, and Equifinality in Data Assimilation with Ecosystem Models. *Ecological Applications* 19:571–574.
<http://www.jstor.org/stable/27645995>

Magee MR, McIntyre PB, Hanson PC, Wu CH (2019) Drivers and Management Implications of Long-Term Cisco Oxythermal Habitat Decline in Lake Mendota, WI. *Environmental Management* 63:396–407. <https://doi.org/10.1007/s00267-018-01134-7>

Mariani M, Connor SE, Theuerkauf M, et al (2022) Disruption of cultural burning promotes shrub encroachment and unprecedented wildfires. *Frontiers in Ecology and the Environment* 20:292–300. <https://doi.org/10.1002/fee.2395>

Mercado-Bettín D, Clayer F, Shikhani M, et al (2021) Forecasting water temperature in lakes and reservoirs using seasonal climate prediction. *Water Research* 201:117286.
<https://doi.org/10.1016/j.watres.2021.117286>

Mittermaier, MP (2008) The potential impact of using persistence as a reference forecast on perceived forecast skill. *Weather and Forecasting* 23:1022–1031. <https://doi.org/10.1175/2008WAF2007037.1>.

Moore TN, Thomas RQ, Woelmer WM, Carey CC (2022) Integrating Ecological Forecasting into Undergraduate Ecology Curricula with an R Shiny Application-Based Teaching Module. *Forecasting* 4:604–633. <https://doi.org/10.3390/forecast4030033>

- Murphy, AH (1992) Climatology, persistence, and their linear combination as standards of reference in skill scores. *Weather and Forecasting* 7(4):692-698.
- Nõges P, Tuvikene L (2012) Spatial and annual variability of environmental and phytoplankton indicators in Lake Võrtsjärv: implications for water quality monitoring. *Estonian J Ecol* 61:227. <https://doi.org/10.3176/eco.2012.4.01>
- Ouellet-Proulx S, Chimi Chiadjeu O, Boucher MA, St-Hilaire A (2017) Assimilation of water temperature and discharge data for ensemble water temperature forecasting. *Journal of Hydrology* 554:342–359. <https://doi.org/10.1016/j.jhydrol.2017.09.027>
- Page T, Smith PJ, Beven KJ, et al (2018) Adaptive forecasting of phytoplankton communities. *Water Research* 134:74–85. <https://doi.org/10.1016/j.watres.2018.01.046>
- Pappenberger F, Ramos MH, Cloke HL, et al (2015) How do I know if my forecasts are better? Using benchmarks in hydrological ensemble prediction. *Journal of Hydrology* 522:697–713. <https://doi.org/10.1016/j.jhydrol.2015.01.024>
- Pauly D (1995) Anecdotes and the shifting baseline syndrome of fisheries. *Trends in Ecology & Evolution* 10:430. [https://doi.org/10.1016/S0169-5347\(00\)89171-5](https://doi.org/10.1016/S0169-5347(00)89171-5)
- Pennekamp F, Iles AC, Garland J, et al (2019) The intrinsic predictability of ecological time series and its potential to guide forecasting. *Ecological Monographs* 89:1–17. <https://doi.org/10.1002/ecm.1359>

- Peters DPC, Havstad KM, Cushing J, et al (2014) Harnessing the power of big data: infusing the scientific method with machine learning to transform ecology. *Ecosphere* 5(6): 1-15.
<https://doi.org/10.1890/ES13-00359.1>
- R Core Team (2022) R: A language and environment for statistical computing. R Foundation for Statistical Computing
- Radeloff VC, Williams JW, Bateman BL, et al (2015) The rise of novelty in ecosystems. *Ecological Applications* 25:2051–2068. <https://doi.org/10.1890/14-1781.1>
- Richardson DC, Carey CC, Bruesewitz DA, Weathers KC (2017) Intra- and inter-annual variability in metabolism in an oligotrophic lake. *Aquatic Sciences* 79:319–333.
<https://doi.org/10.1007/s00027-016-0499-7>
- Ruiz-Aguilar JJ, Turias I, González-Enrique J, et al (2021) A permutation entropy-based EMD–ANN forecasting ensemble approach for wind speed prediction. *Neural Comput & Applic* 33:2369–2391. <https://doi.org/10.1007/s00521-020-05141-w>
- Saber A, James DE, Hayes DF (2020) Long-term forecast of water temperature and dissolved oxygen profiles in deep lakes using artificial neural networks conjugated with wavelet transform. *Limnology and Oceanography* 65:1297–1317.
<https://doi.org/10.1002/lno.11390>
- Sánchez E, Colmenarejo MF, Vicente J, et al (2007) Use of the water quality index and dissolved oxygen deficit as simple indicators of watersheds pollution. *Ecological Indicators* 7:315–328. <https://doi.org/10.1016/j.ecolind.2006.02.005>

- Sharma S, Richardson DC, Woolway RI, et al (2021) Loss of Ice Cover, Shifting Phenology, and More Extreme Events in Northern Hemisphere Lakes. *Journal of Geophysical Research: Biogeosciences* 126:e2021JG006348. <https://doi.org/10.1029/2021JG006348>
- Simões F dos S, Moreira AB, Bisinoti MC, et al (2008) Water quality index as a simple indicator of aquaculture effects on aquatic bodies. *Ecological Indicators* 8:476–484. <https://doi.org/10.1016/j.ecolind.2007.05.002>
- Sippel S, H L, F G (2019) statcomp: Statistical Complexity and Information Measures for Time Series Analysis
- Soares LMV, Calijuri M do C (2021) Deterministic modelling of freshwater lakes and reservoirs: Current trends and recent progress. *Environmental Modelling and Software* 144:105143. <https://doi.org/10.1016/j.envsoft.2021.105143>
- Sondergaard M, Jensen PJ, Jeppesen E (2001) Retention and internal loading of phosphorus in shallow, eutrophic lakes. *TheScientificWorldJournal* 1:427–442. <https://doi.org/10.1100/tsw.2001.72>
- Staeher PA, Testa JM, Kemp WM, et al (2012) The metabolism of aquatic ecosystems: history, applications, and future challenges. *Aquat Sci* 74:15–29. <https://doi.org/10.1007/s00027-011-0199-2>
- Stanley EH, Collins SM, Lottig NR, et al (2019) Biases in lake water quality sampling and implications for macroscale research. *Limnology and Oceanography* 64:1572–1585. <https://doi.org/10.1002/lno.11136>

Steele BG, Weathers KC, Association LSP (2023) Lake-Sunapee-Protective-Association/LMP: LSPA LMP database 1986-2022 (v2) (v2023.2). Data set Zenodo.

<https://doi.org/10.5281/zenodo.8003784>

Sudibyo MRP, Eibl EPS, Hainzl S, Hersir GP (2022) Eruption Forecasting of Strokkur Geyser, Iceland, Using Permutation Entropy. *Journal of Geophysical Research: Solid Earth* 127:e2022JB024840. <https://doi.org/10.1029/2022JB024840>

Sydeman WJ, Santora JA, Thompson SA, et al (2013) Increasing variance in North Pacific climate relates to unprecedented ecosystem variability off California. *Global Change Biology* 19:1662–1675. <https://doi.org/10.1111/gcb.12165>

Thomas RQ, Figueiredo RJ, Daneshmand V, et al (2020) A Near-Term Iterative Forecasting System Successfully Predicts Reservoir Hydrodynamics and Partitions Uncertainty in Real Time. *Water Resources Research* 56:e2019WR026138. <https://doi.org/10.1029/2019WR026138>

Thomas R (2023). FLARER: Forecasting Lake and Reservoir Ecosystems_. R package version 2.2.21.

Thomas RQ, McClure RP, Moore TN, et al (2023) Near-term forecasts of NEON lakes reveal gradients of environmental predictability across the US. *Frontiers in Ecology and the Environment* 21:220–226. <https://doi.org/DOI: 10.1002/fee.2623>

Wander HL, Thomas RQ, Moore TN, et al (2023) Data assimilation experiments inform monitoring needs for near-term ecological forecasts in a eutrophic reservoir. *ESS Open Archive Preprint*. <https://doi.org/10.22541/essoar.168500255.59108131/v1>

Wetzel RG (2001) Limnology, 3rd edn. Academic Press, San Diego, USA

Wheatcroft E (2019) Interpreting the skill score form of forecast performance metrics.

International Journal of Forecasting 35:573–579.

<https://doi.org/10.1016/j.ijforecast.2018.11.010>

Wheeler K, Dietze M, LeBauer D, et al (2023) Predicting Spring Phenology in Deciduous

Broadleaf Forests: An Open Community Forecast Challenge. SSRN Preprint.

<https://doi.org/10.2139/ssrn.4357147>

Woelmer WM, Thomas RQ, Lofton ME, et al (2022) Near-term phytoplankton forecasts reveal

the effects of model time step and forecast horizon on predictability. Ecological

Applications 32:1–22. <https://doi.org/10.1002/eap.2642>

Woelmer WM, Thomas RQ, Olsson F, et al (2023) Forecasts and score summary files to

accompany the manuscript “Forecast performance of near-term, iterative water

temperature and dissolved oxygen forecasts varies across time and depth”. Dataset

Zenodo. <https://doi.org/10.5281/zenodo.8212703>

Woolway RI, Merchant CJ (2019) Worldwide alteration of lake mixing regimes in response to

climate change. Nature Geoscience 12:271–276. [https://doi.org/10.1038/s41561-019-](https://doi.org/10.1038/s41561-019-0322-x)

0322-x

Woolway RI, Sharma S, Weyhenmeyer GA, et al (2021) Phenological shifts in lake stratification

under climate change. Nature Communications 12:2318. [https://doi.org/10.1038/s41467-](https://doi.org/10.1038/s41467-021-22657-4)

021-22657-4

Woolway RI, Weyhenmeyer GA, Schmid M, et al (2019) Substantial increase in minimum lake surface temperatures under climate change. *Climatic Change* 155:81–94.

<https://doi.org/10.1007/s10584-019-02465-y>

Wynne JH, Woelmer W, Moore TN, et al (2023) Uncertainty in projections of future lake thermal dynamics is differentially driven by lake and global climate models. *PeerJ*

11:e15445. <https://doi.org/10.7717/peerj.15445>

Yue W, Liu Q, Ruan Y, et al (2022) A prediction approach with mode decomposition-recombination technique for short-term load forecasting. *Sustainable Cities and Society*

85:104034. <https://doi.org/10.1016/j.scs.2022.104034>

Zhu G, Hunter J, Jiang Y (2016) Improved Prediction of Dengue Outbreak Using the Delay Permutation Entropy. In: 2016 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData). pp 828–832. <https://doi.org/10.1109/iThings-GreenCom-CPSCom-SmartData.2016.172>

Zhu M, Wang J, Yang X, et al (2022) Eco-Environment & Health A review of the application of machine learning in water quality evaluation. *Eco-Environment & Health* 1:107–116.

<https://doi.org/10.1016/j.eehl.2022.06.001>

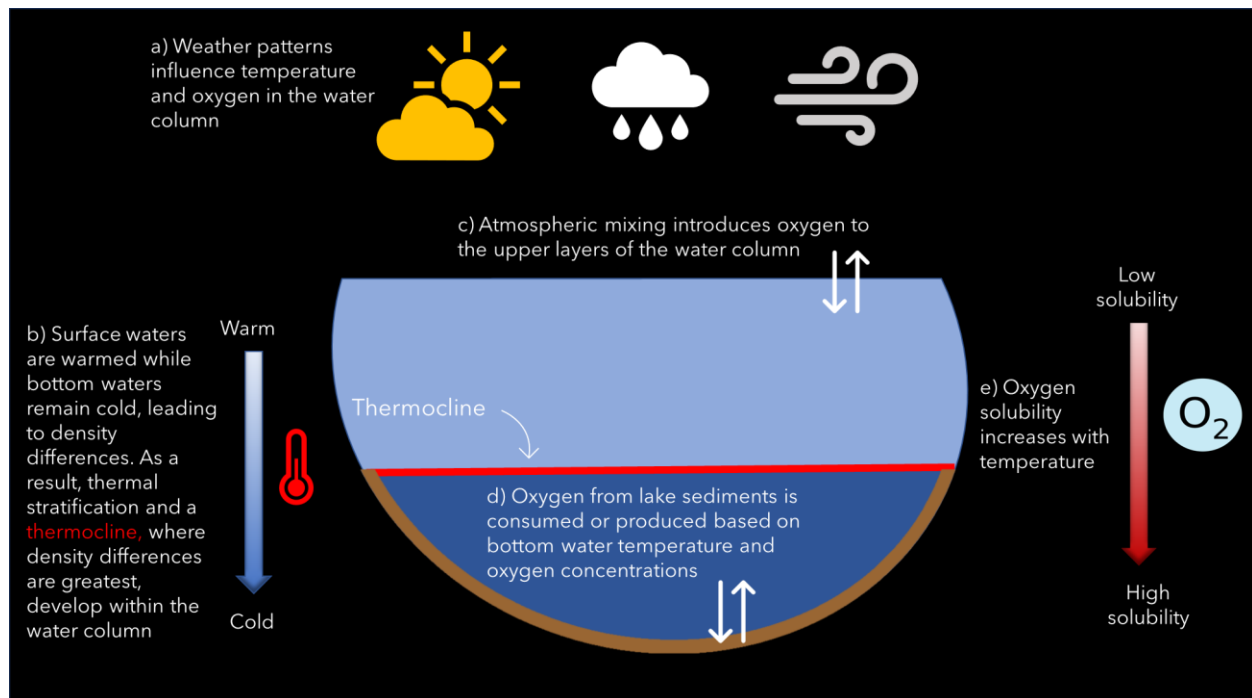


Figure 1. Conceptual diagram of the processes represented in this study which are important for driving changes in water temperature and dissolved oxygen.

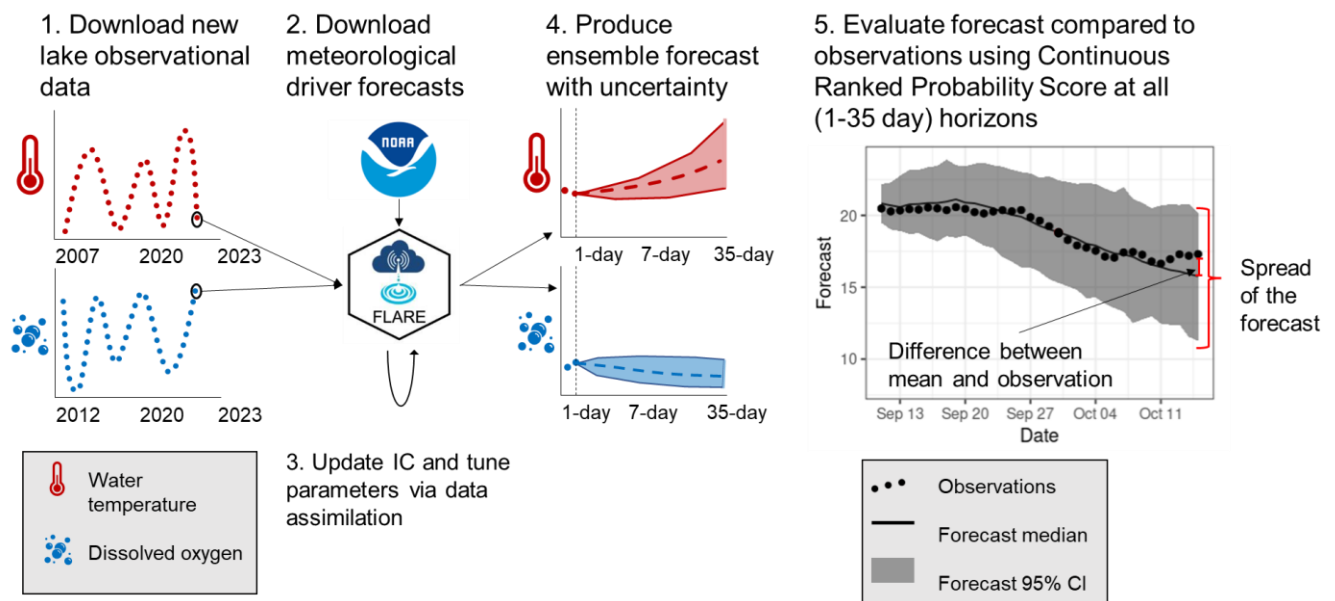


Figure 2. Conceptual representation of the FLARE (Forecasting Lakes And Reservoir Ecosystems) framework which shows the daily steps of 1) downloading lake observational data and appending the new data to the long-term observational dataset, 2) downloading meteorological forecasts from NOAA needed as driver data for the lake model, 3) updating model initial conditions (IC) and tuning parameters via an ensemble Kalman filter (EnKF), 4) producing a forecast of 1 to 35-days-ahead with uncertainty, and 5) evaluating the forecast. CI = confidence interval. Awaiting permission from NOAA to use logo for publication. Use of the emblem or logo does not imply an endorsement by NOAA/NWS.

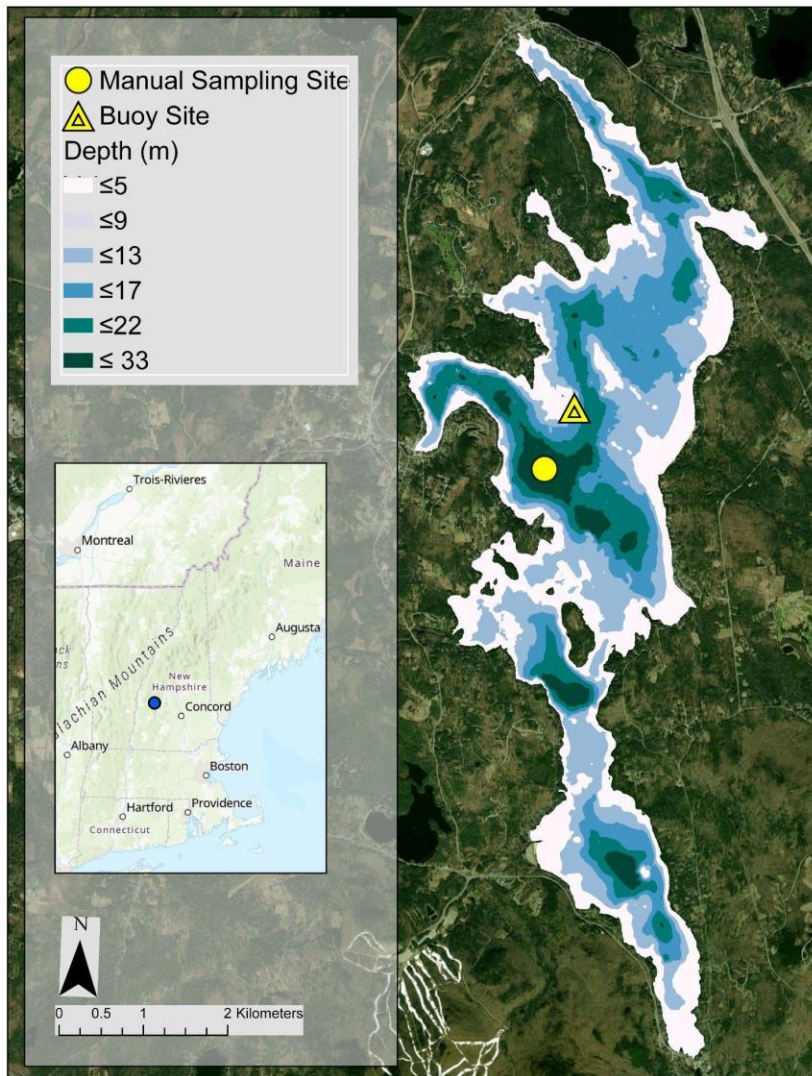


Figure 3. Bathymetric map of Lake Sunapee, NH, USA (43.37°, -72.05°; (LSPA 2023), showing the location of the manually-collected data sampling site (circle) and the sensed buoy site (triangle).

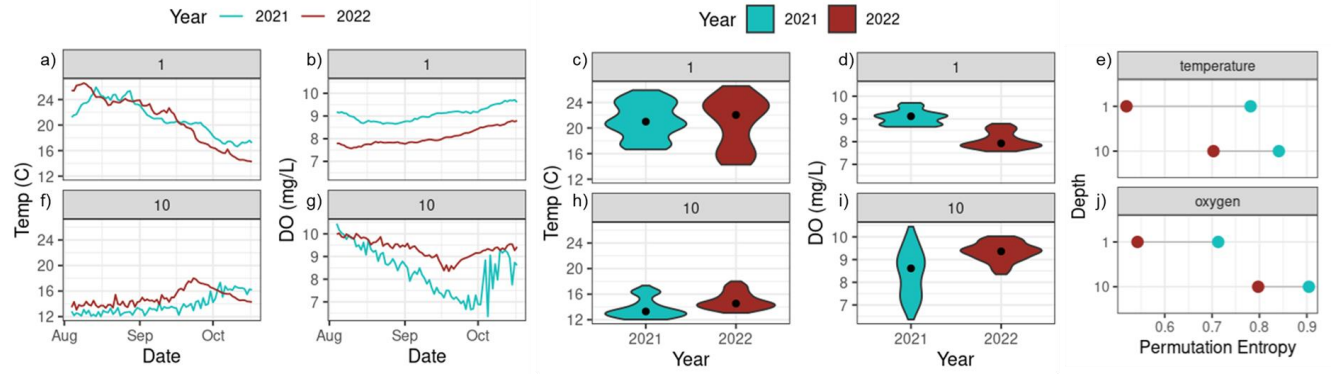


Figure 4. High-frequency observations of temperature and oxygen from the buoy in Lake Sunapee over the study period (August - October) in 2021 and 2022 showing the time series of observations for a) temperature at 1.0 m, b) oxygen at 1.0 m, violin plots of observations for c) temperature at 1.0 m, d) oxygen at 1.0 m, and e) permutation entropy of temperature at 1.0 m and 10.0 m in both years. Observations at 10.0 m are shown in the bottom row with the time series of e) temperature at 10.0 m, and f) oxygen at 10.0 m, and violin plots of g) temperature at 10.0 m, and h) oxygen at 10.0 m, as well as j) permutation entropy of oxygen at 1.0 m and 10.0 m in both years.

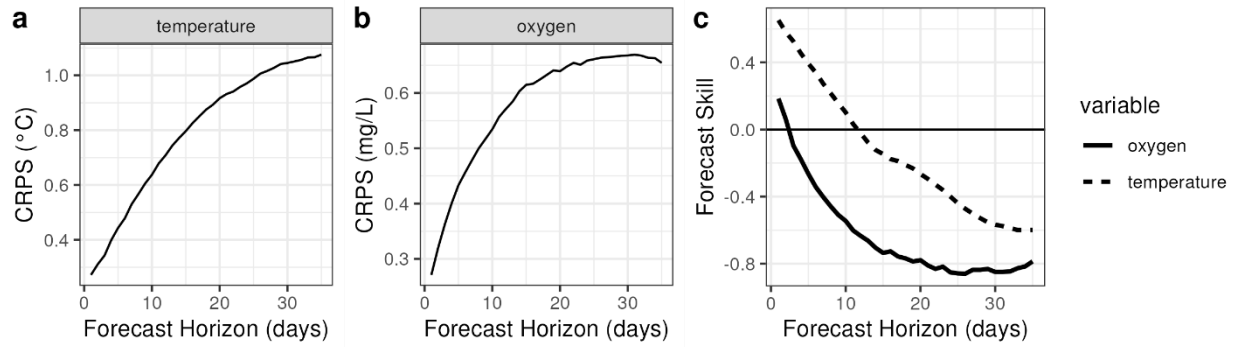


Figure 5. Forecast accuracy of a) water temperature ($^{\circ}\text{C}$), and b) dissolved oxygen (mg/L) and c) forecast skill (unitless; calculated using Eq. 2) of both temperature and oxygen. All three of these metrics were calculated by aggregating over the full study period and across all depths at forecast horizons from 1-35 days into the future (note the difference in y-axes). In panels a and b, more accurate forecasts are represented by lower CRPS values. In Panel c, forecasts less than zero are less skillful than the null model.

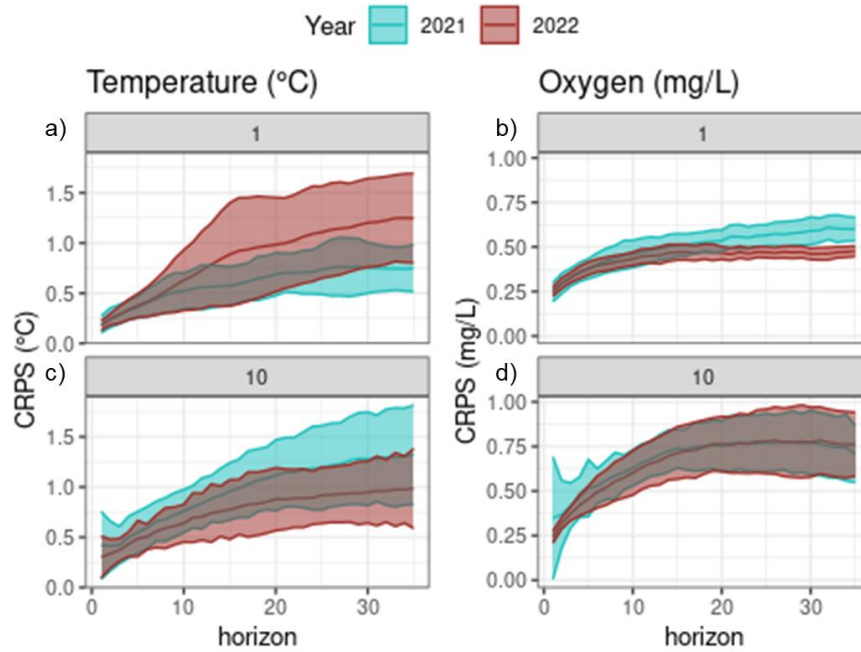


Figure 6. Forecast accuracy of a) temperature at 1.0 m, b) oxygen at 1.0 m, c) temperature at 10.0 m, and d) oxygen at 10.0 m as measured by the Continuous Ranked Probability Score (CRPS).

The mean across all days within a given year is shown by the central solid line, with the shaded area representing the standard deviation across all days within a given year. Higher CRPS values indicate lower forecast accuracy. Note the difference in y-axis.

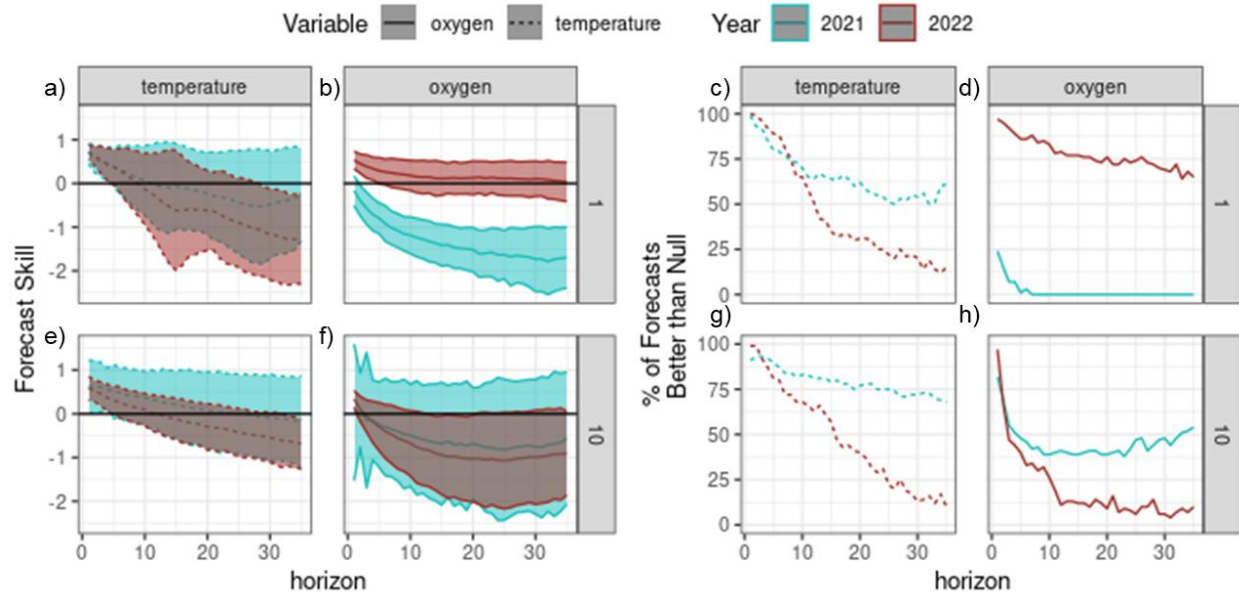


Figure 7. Forecast skill across the two evaluation periods for a) temperature at 1.0 m, b) oxygen at 1.0 m, as well as the percent of forecasts which beat the null for c) temperature and 1.0 m, and d) oxygen and 1.0 m. Forecasts from 10.0 m are shown in the bottom row for e) temperature at 10.0 m, f) oxygen at 10.0 m, and percent of forecasts better than the null for g) temperature at 10.0 m, and h) oxygen at 10.0 m.

Supporting Information

Chapter IV: Performance of near-term, iterative water temperature and dissolved oxygen forecasts varies across time and depth*

Whitney M. Woelmer¹, R. Quinn Thomas^{1,2}, Freya Olsson¹, Bethel Steele³, Kathleen C. Weathers³, Cayelan C. Carey¹

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

²Department of Forest Resources and Environmental Conservation, Virginia Tech, Blacksburg, Virginia, USA

³Cary Institute of Ecosystem Studies, Millbrook, New York, USA

Contents of this file:

Texts: 3

Tables: 3

Figures: 11

*In preparation for submission to *Ecological Informatics*

Text S1. Configuration of initial condition and observational uncertainty

Initial condition uncertainty was specified in the 'observations_config.csv' file in FLARE (Table S2). For both temperature and dissolved oxygen, the standard deviation of the observations was derived from the reported instrument accuracy for each variable. Temperature was measured every meter from the surface to 10.0m depth using a TS210 thermistor string, which reports a +/- 0.075C reading accuracy. We rounded this value to the nearest decimal, resulting in +/- 0.1C accuracy. Dissolved oxygen was measured at 1.0m and 10.0m using an EXO Sonde2 and an RDO Pro-X Process Optical Dissolved Oxygen Probe, respectively. Both sensors report a measurement accuracy of 0.1 mg/L when readings range between 0 and 20 mg/L, which was the case for this study period. Because FLARE operates in units of mmol/m³ of dissolved oxygen, we report the measurement accuracy of 0.1 mg/L in mmol/m³, which corresponds to 3.125 mmol/m³.

Text S2. Configuration of process uncertainty

The standard deviation of the normal distribution of process uncertainty was specified in the ‘states_config.csv’ and ‘depth_model_sd.csv’ files in https://github.com/wwoelmer/SUNP-forecast-code/tree/glm_aed. For temperature, the standard deviation of this distribution was calculated every meter from the surface to 10.0 m as the standard deviation of the residuals in an AR-1 lag model during May 2018. This resulted in a ratio of 1:9 between the initial condition standard deviation (0.1) and process standard deviation (0.9). Given that forecasts of water temperature have shown skillful performance in previous studies, we maintained the same ratio of initial condition: process standard deviation to estimate our process standard deviation for oxygen (i.e., 3.125: 28.125). It was necessary to maintain this ratio consistently between temperature and oxygen due to the inherent mechanisms embedded within the EnKF which rely on the relative amount of process and observational uncertainty.

Text S3. Description of thermocline depth calculation

We calculated thermocline depth in each year of study period (2021 and 2022) using all available buoy observations from 0.1 m to 10.0 m. We used the “thermo.depth()” function in the ‘RLakeAnalyzer’ package (Winslow et al. 2019) within R, which uses a maximum density gradient to define the thermocline depth.

Table S1. Details of data availability, spin-up length, and forecast date ranges for all years forecasted in this study.

Year	First date with data availability for all variables	Spin-up date range	Forecast date range	Date range analyzed in this study	Number of forecast days analyzed
2021	2021-06-30	06-30 – 08-04	08-04 – 10-19	08-04 – 10-17	74
2022	2022-04-27	04-27 – 06-01	06-01 – 10-17	08-04 – 10-17	74

Table S2. List of parameter values included in the Ensemble Kalman Filter (EnKF) parameter calibration in FLARE. The zones represent zones within the water column with increasing numbers corresponding to depth above the lake bottom (e.g., Zone 1 is the first zone on the bottom of the lake).

Parameter name	Parameter description	Initial parameter value	Initial parameter lower bound	Initial parameter Upper bound	Parameter lower bound	Parameter upper bound	Inflation factor	Perturbation factor	Units
sed_temp_mean	Temperature at sediment-water interface in Zone 1	12	5	15	-100	100	1.04	1.2	deg_C
sed_temp_mean	Temperature at sediment-water interface in Zone 2	15	10	20	-100	100	1.04	1.5	deg_C
sed_temp_mean	Temperature at sediment-water interface in Zone 3	15	10	20	-100	100	1.04	1.5	deg_C

lw_factor	Longwave scaling factor	1	0.9	1.1	-10	10	2	0.09	unitless
Fsed_oxy	Sediment oxygen demand in Zone 1	-40	-40	0	-100	10	1.04	4	mmol_m3_d ay
Fsed_oxy	Sediment oxygen demand in Zone 2	-40	-30	0	-100	10	1.04	4	mmol_m3_d ay
Fsed_oxy	Sediment oxygen demand in Zone 3	-40	-30	0	-100	10	1.04	4	mmol_m3_d ay

Table S3. Definitions and descriptions of uncertainty sources included in this study. The method of uncertainty quantification is described. For sources whose initial configuration is different from the quantification throughout the forecast run, that is described separately below.

UC Source	Definition	Method
Initial Condition	Uncertainty in the starting conditions of your forecast state variable, often due to inability to measure the state of the system at time t	Each ensemble member is assigned a different initial condition value taken from a distribution around the initial observation. The standard deviation of the distribution was chosen based on the accuracy of the instrument used to measure each forecast state variable (See Text S1 for detailed description of each forecast state variable IC configuration)
Parameter	Uncertainty in the true values of model parameters	Each ensemble member is assigned a different parameter value randomly sampled from a normal distribution of parameter values determined via EnKF data assimilation. A detailed description of the parameter values and ranges for parameter tuning is found in Table S2. Initial parameter values were determined from calibration of GLM-AED outside of the forecasting framework
Process	Uncertainty in the model's ability to represent the processes which determine fluctuations in forecast state variables	Each ensemble member is assigned a different value for process uncertainty drawn from a normal distribution with a mean of zero and standard deviation determined based on historical performance of a random walk. See Text S2 for a full description of how the standard deviation was determined.
Weather driver data	Uncertainty in the values of the forecasted weather driver variables	30 individual weather trajectories for each weather driver variable were used in our ensemble forecasts. Weather ensembles were determined by NOAA GEFS

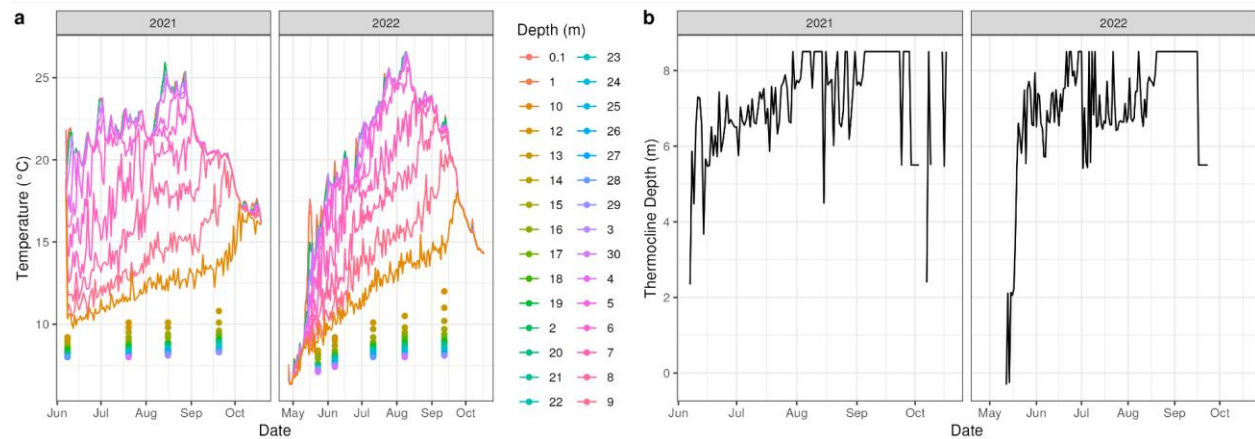


Figure S1. Lake Sunapee high-frequency buoy observations (lines) and monthly profiles from depths > 10 m (points) in 2021 and 2022 of a) water temperature profiles from the surface to 30.0m and b) calculated thermocline depth using these profiles

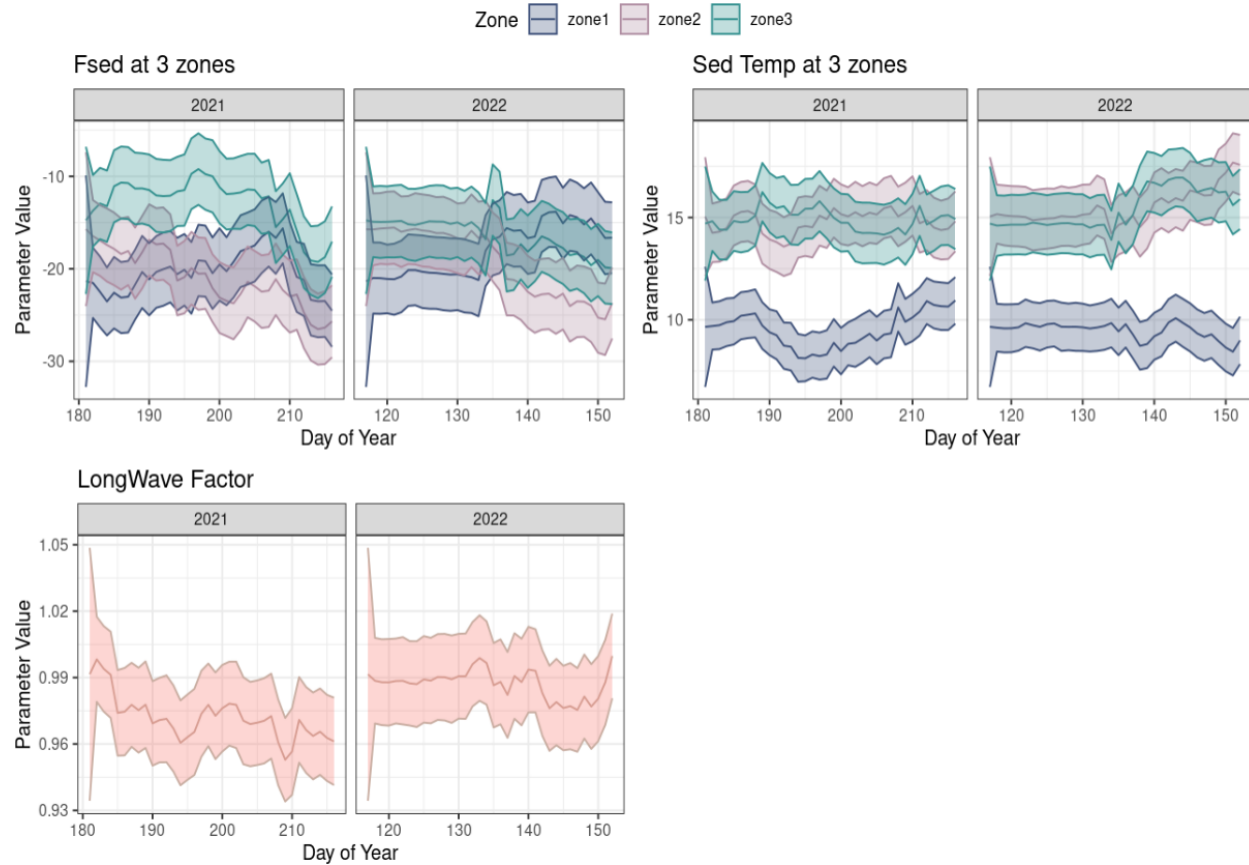


Figure S2. Evolution of parameter values of the spin-up period in the two focal years of this study (2021 and 2022). Note that the spin-up time period started on different dates in each year but covered 35 days total. See Table S2 for a description of each parameter.

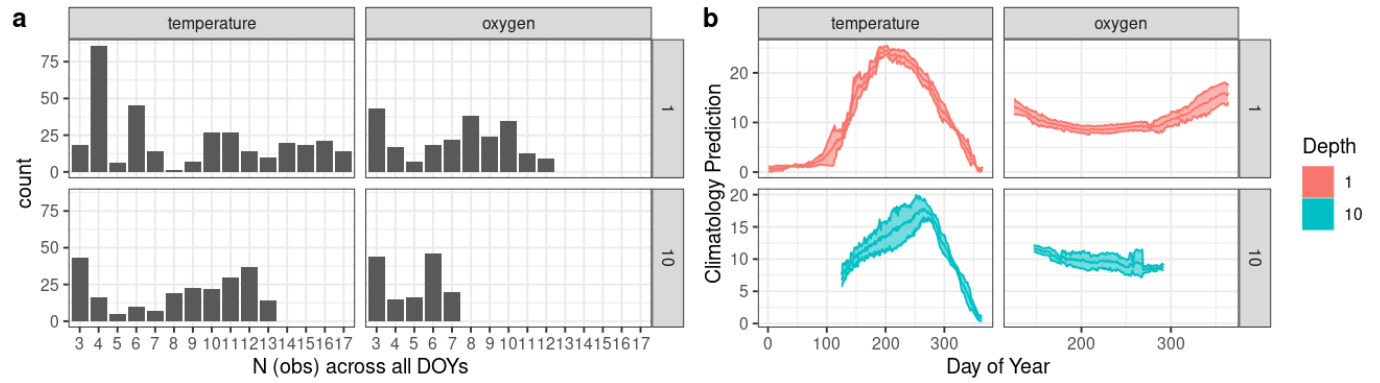


Figure S3. Climatology null calculated as the mean and standard deviation across all available data from 2007-2022. Only days of year with three or more observations on that date were included. Sensors were not all deployed over the same time frame since the buoy was first deployed in 2007, resulting in differences in data availability across days of year for each variable and depth.

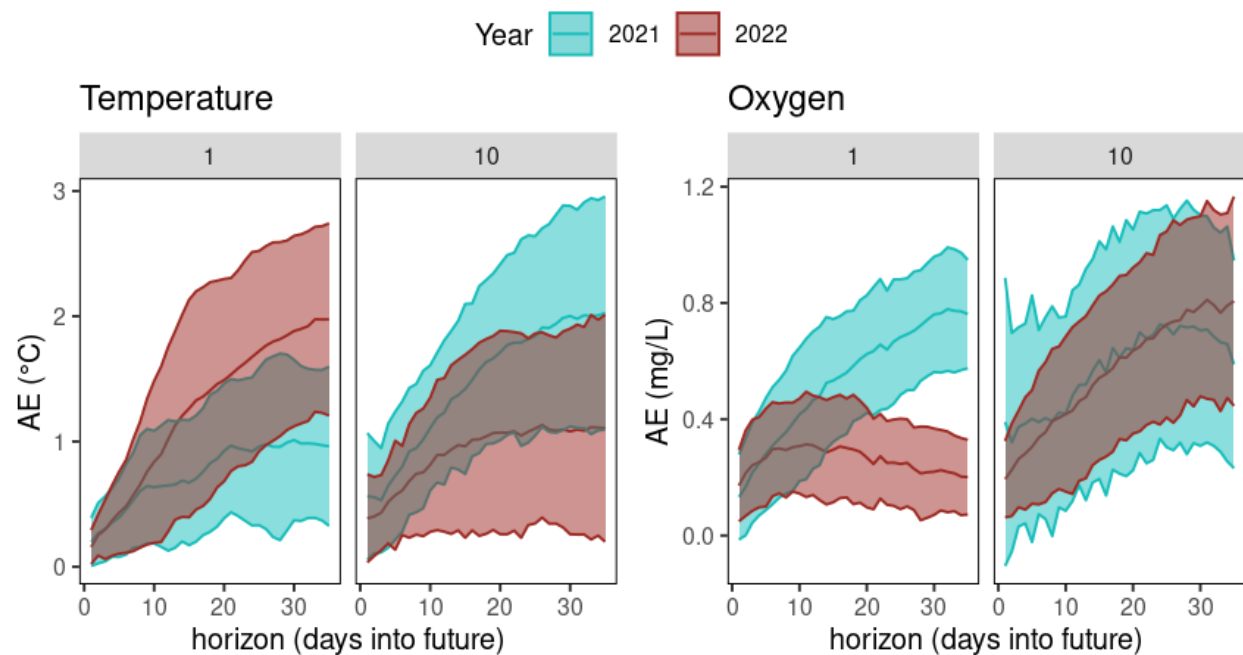


Figure S4. Absolute Error (AE) of temperature and oxygen forecasts in 2021 and 2022 at Lake Sunapee, NH at 1.0 m and 10.0 m depth across a 1-35 day forecast horizon.

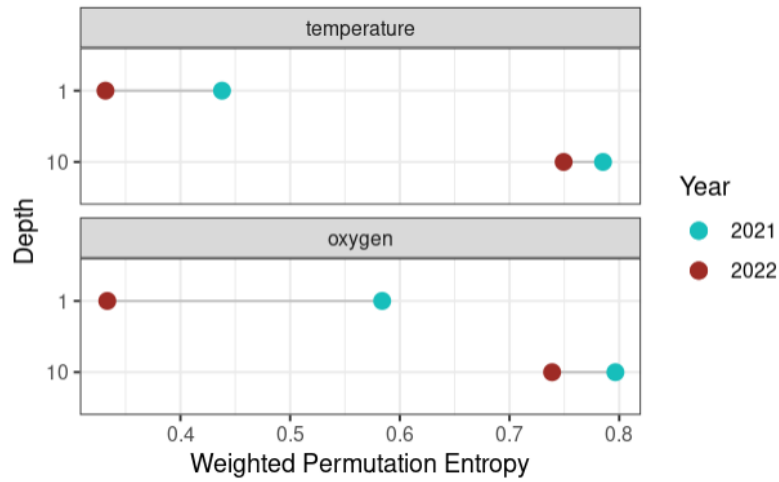


Figure S5. Weighted permutation entropy across years (2021 and 2022), depths (1.0 m and 10.0 m) and variables (water temperature and dissolved oxygen) in this study.

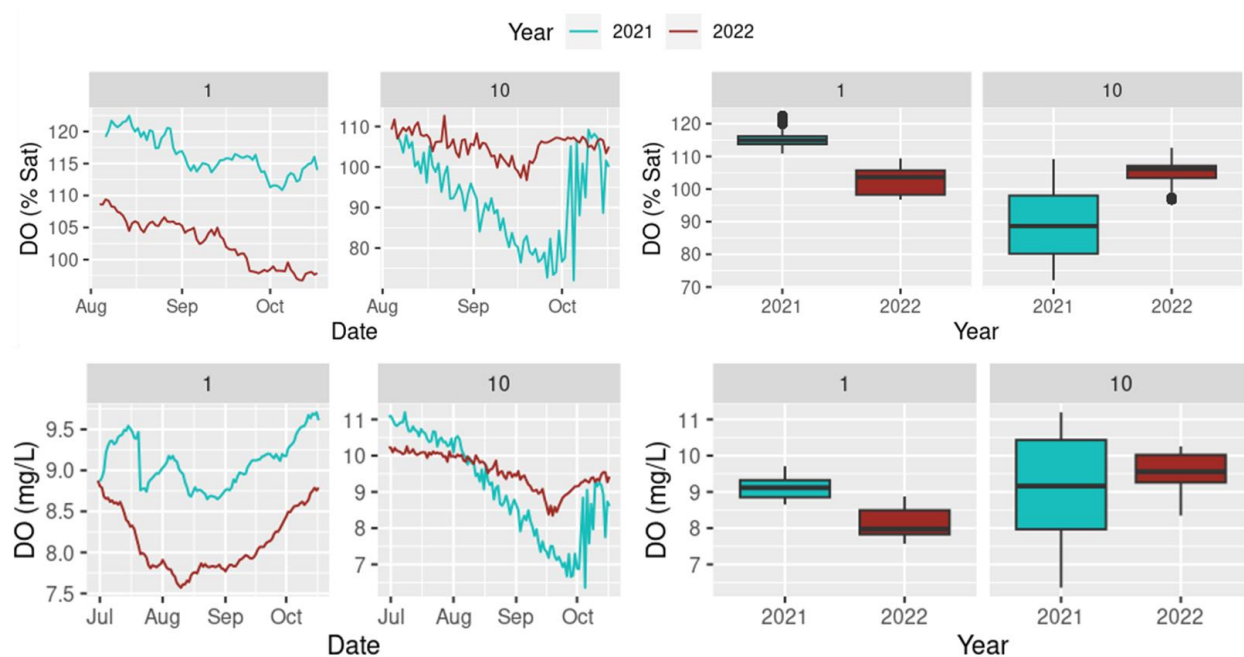


Figure S6. Comparison of dissolved oxygen in percent saturation and mg/L over the focal years in this study (2021 and 2022)

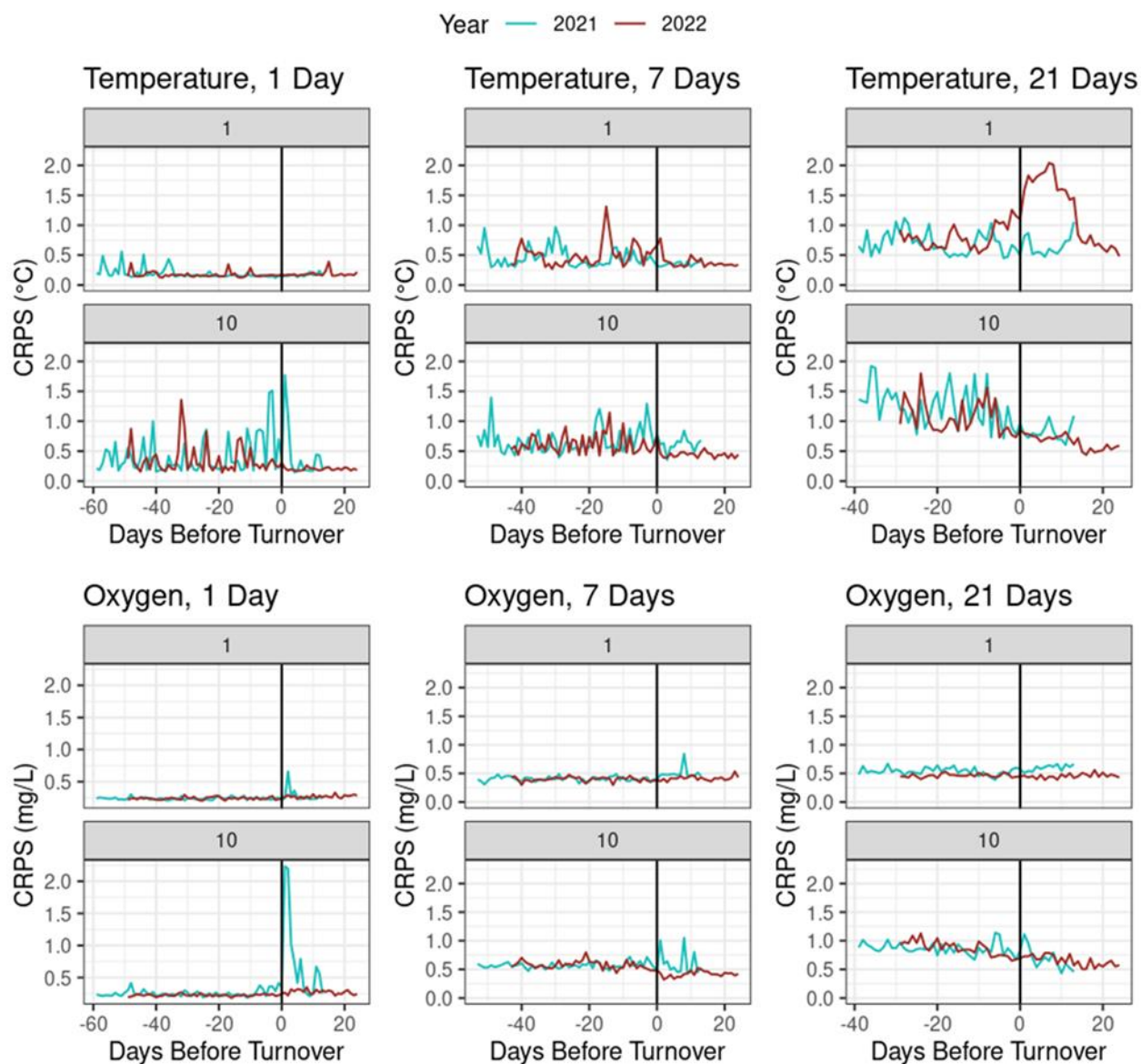


Figure S7. Forecast performance as Continuous Ranked Probability Score (CRPS) of forecasts at select (1-day, 7-day, and 21-day) horizons across the full forecast time period (August 04 to October 17). Dates are shown relative to the day of turnover in each year (2021-10-04 and 2022-09-23). CRPS is shown across years, depth, and variable for each horizon.

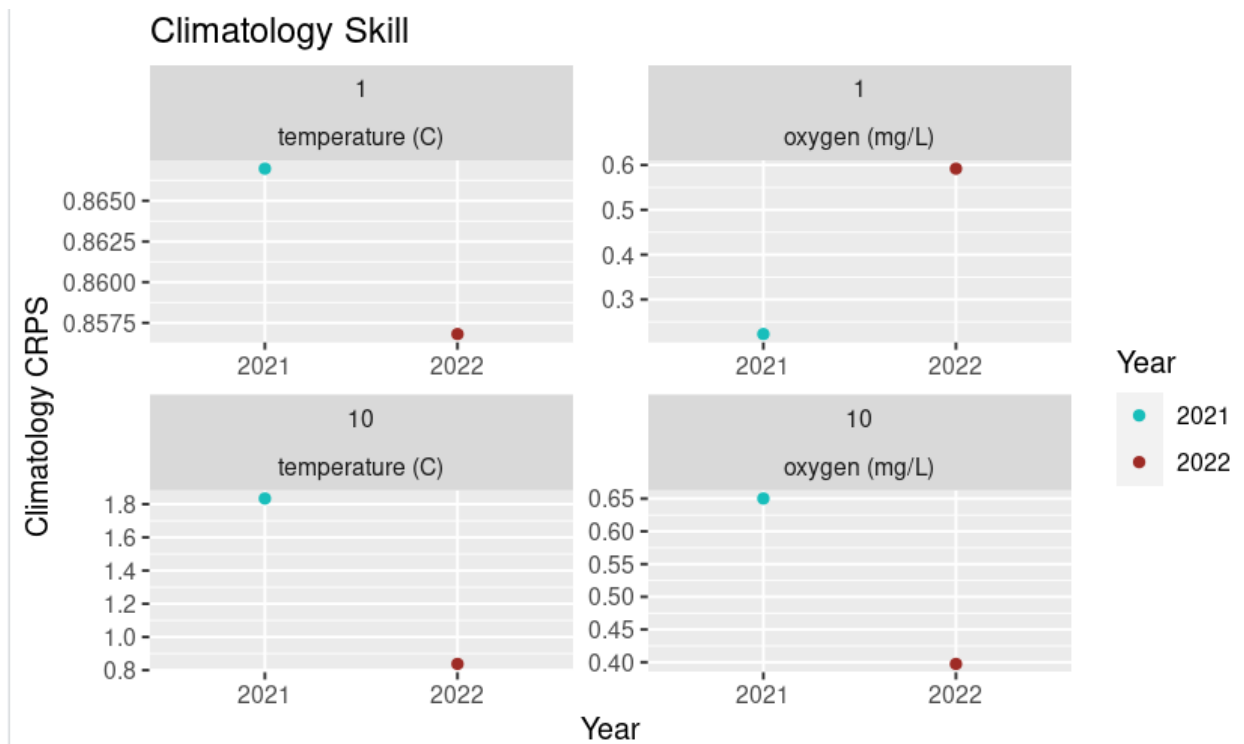


Figure S8. Forecast accuracy of climatology model across years, depths, and variables. The climatology null was calculated across all historical years with available observations (Fig. S1) and was evaluated against observations in 2021 and 2022, resulting in different skill scores for each year.

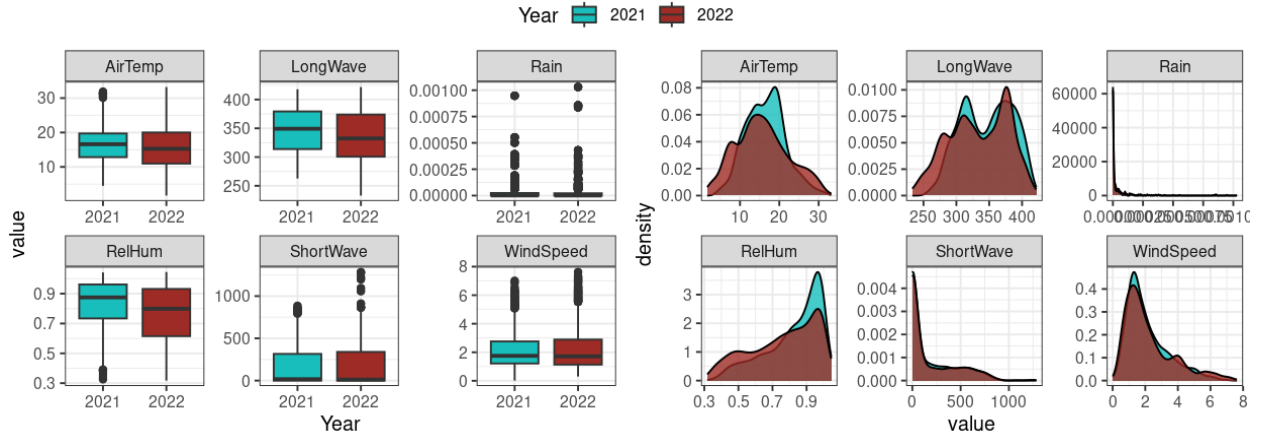


Figure S9. Meteorological variables over the two years in our study at Lake Sunapee, NH, USA.

AirTemp = air temperature in $^{\circ}\text{C}$; LongWave = longwave radiation in W m^{-2} ; Rain = precipitation in mm/day; RelHum = relative humidity in decimal percentage; ShortWave = shortwave radiation in W m^{-2} ; WindSpeed = wind speed in m s^{-1} .

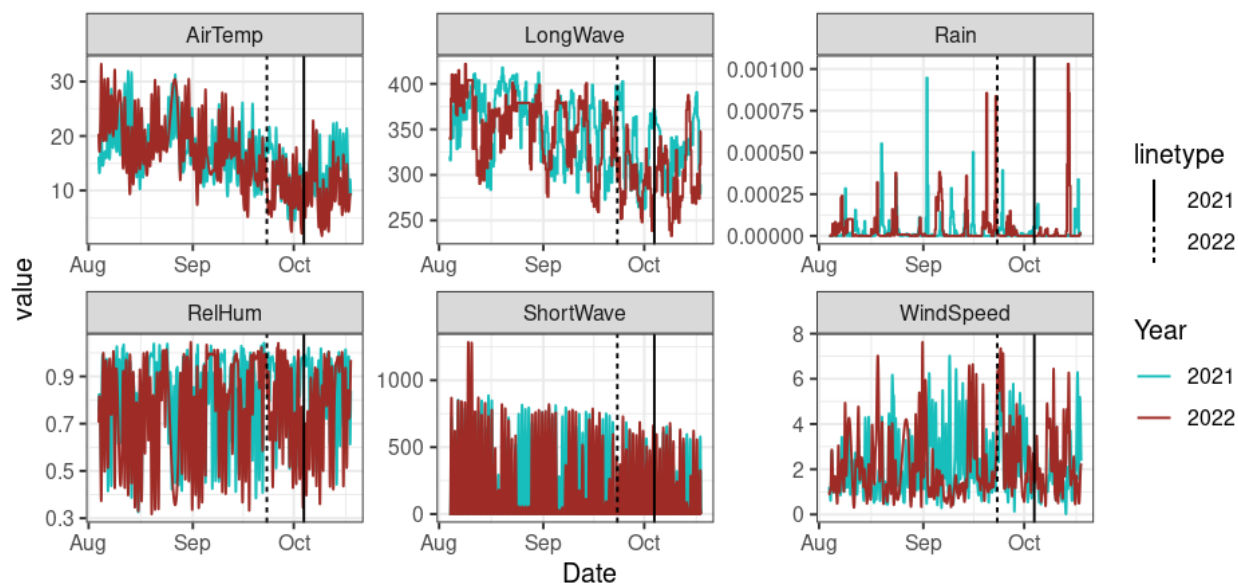


Figure S10. Time series of meteorological variables which were used to drive FLARE forecasts at Lake Sunapee, NH in 2021 and 2022. The date of turnover is indicated by the vertical line.

AirTemp = air temperature in $^{\circ}\text{C}$; LongWave = longwave radiation in W m^{-2} ; Rain = precipitation in mm/day; RelHum = relative humidity in decimal percentage; ShortWave = shortwave radiation in W m^{-2} ; WindSpeed = wind speed in m s^{-1} .

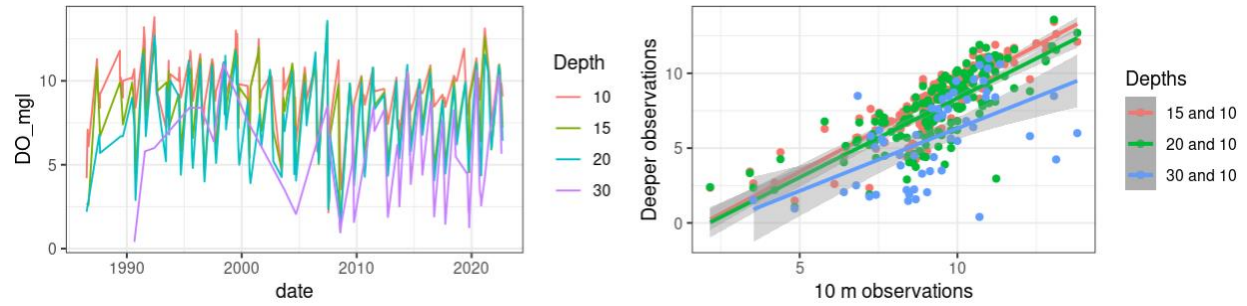


Figure S11. Comparison of deepwater observations at 10.0 m, 15.0 m, 20.0 m, and 30.0 m from the deep site at Lake Sunapee, NH.

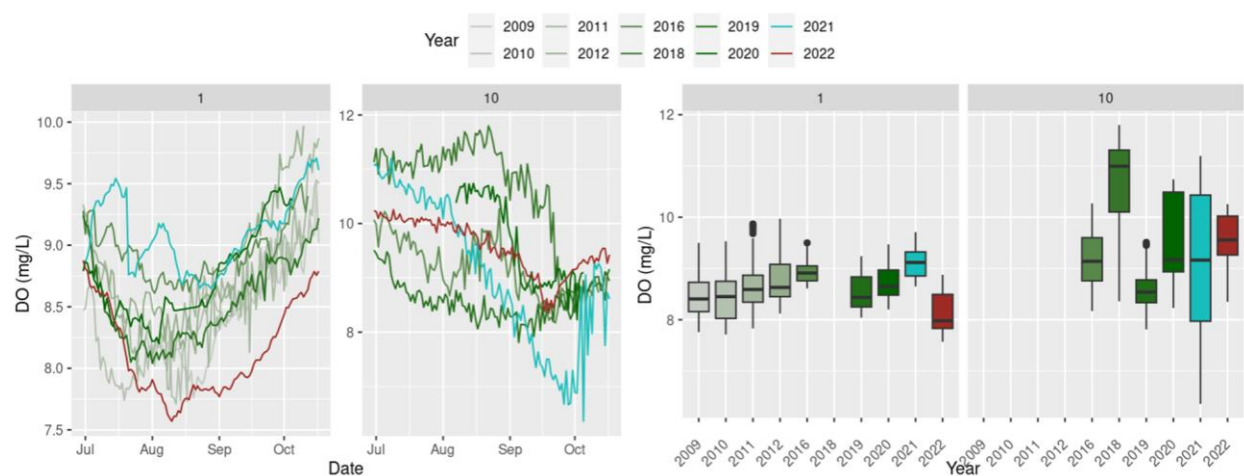


Figure S12. Comparison of historical observations of dissolved oxygen and focal years in this study (2021 and 2022). Years with incomplete observations were removed (2007, 2008, 2013, 2014, 2017).

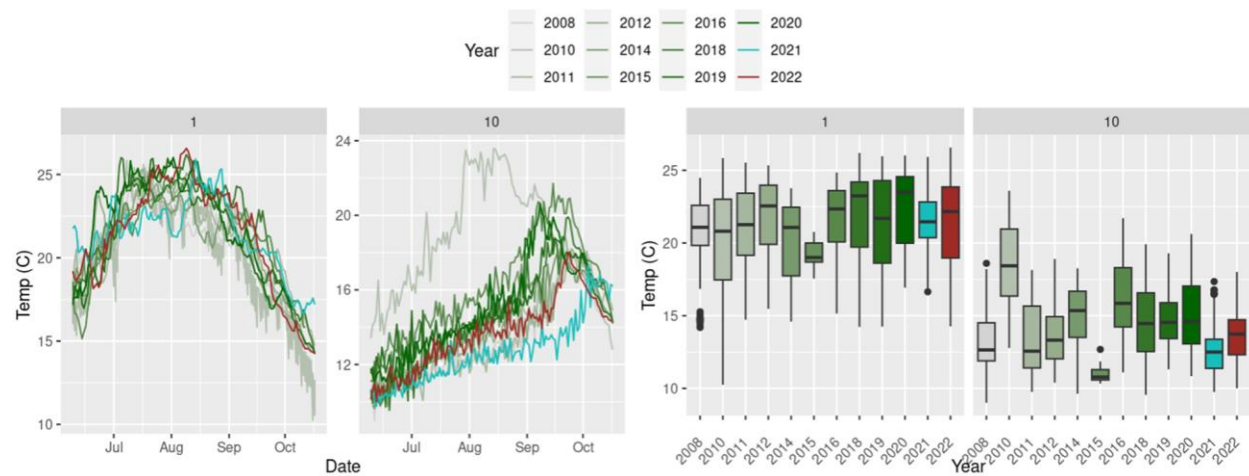


Figure S13. Comparison of historical observations of water temperature and focal years in this study (2021 and 2022). Years with incomplete observations were removed (2007, 2009, 2013, 2017).

**Chapter V: Embedding communication concepts in forecasting training increases students’
understanding of ecological uncertainty***

Whitney M. Woelmer¹, Tadhg N. Moore^{1,2}, Mary E. Lofton¹, R. Quinn Thomas^{1,2}, and Cayelan C. Carey¹

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

²Department of Forest Resources and Environmental Conservation, Virginia Tech, Blacksburg, VA, USA

*Published in *Ecosphere* with the citation:

Woelmer, W. M., Moore, T.N., Lofton, M. E., Thomas, R. Q., & Carey, C. C. (2023).

Embedding communication concepts in forecasting training increases students’
understanding of ecological uncertainty. *Ecosphere*, 14(8), e4628.

This article is published in *Ecosphere* as Open Access with a CC-BY-ND copyright held by the authors, giving the authors permission to republish in this dissertation.

Abstract

Communicating and interpreting uncertainty in ecological model predictions is notoriously challenging, motivating the need for new educational tools which introduce ecology students to core concepts in uncertainty communication. Ecological forecasting, an emerging approach to estimate future states of ecological systems with uncertainty, provides a relevant and engaging framework for introducing uncertainty communication to undergraduate students, as forecasts can be used as decision support tools for addressing real-world ecological problems and are inherently uncertain. To provide critical training on uncertainty communication and introduce undergraduate students to the use of ecological forecasts for guiding decision-making, we developed a hands-on teaching module within the Macrosystems EDDIE (Environmental Data-Driven Inquiry and Exploration; MacrosystemsEDDIE.org) educational program. Our module used an active learning approach by embedding forecasting activities in an R Shiny application to engage ecology students in introductory data science, ecological modeling, and forecasting concepts without needing advanced computational or programming skills. Pre- and post-module assessment data from >250 undergraduate students enrolled in ecology, freshwater ecology, and zoology courses indicate that the module significantly increased students' ability to interpret forecast visualizations with uncertainty, identify different ways to communicate forecast uncertainty for diverse users, and correctly define ecological forecasting terms. Specifically, students were more likely to describe visual, numeric, and probabilistic methods of uncertainty communication following module completion. Students were also able to identify more benefits of ecological forecasting following module completion, with the key benefits of using forecasts for prediction and decision-making most commonly described. These results show promise for introducing ecological model uncertainty, data visualizations, and forecasting into undergraduate

ecology curricula via software-based learning, which can increase students' ability to engage and understand complex ecological concepts.

Introduction

Communicating uncertainty in ecological models is a pressing challenge across ecology, motivating the need for new educational tools to train students in understanding and interpreting uncertainty in model predictions. Uncertainty in ecological model predictions is inherent across ecological disciplines, ranging across population and community ecology models (e.g., Halpern et al. 2006, Bird et al. 2021), disease ecology models (e.g., Briggs et al. 2009, McClintock et al. 2010), landscape ecology models (e.g., Wu et al. 2006, Lechner et al. 2012), and ecosystem models (e.g., Link et al. 2012, Melbourne-Thomas et al. 2012). Sources of uncertainty in ecological models include uncertainty in model parameter estimates, initial conditions, and the underlying processes being modeled (Dietze 2017). Combined together, these sources of uncertainty can have important implications for interpreting model results, as well as their utility in decision-making (e.g., Berthet et al. 2016, Cheong et al. 2016). However, uncertainty is rarely communicated or is communicated poorly (Boukhelifa and Duke 2009, Hullman 2020), hindering the use of model output for both advancing ecological understanding and decision-making (Joslyn and Savelli 2010, Milner-Gulland and Shea 2017). This is likely because uncertainty is a difficult concept for most individuals to understand (Belia et al. 2005), as well as to mathematically quantify and represent graphically with visualizations (Spiegelhalter et al. 2011, Potter et al. 2012, Bonneau et al. 2015). Given low levels of visualization literacy in both the general and scientific population (Maltese et al. 2015), educational tools to improve communication of ecological model uncertainty are critically needed.

Ecological forecasting provides a powerful framework for teaching students uncertainty communication and data science skills, which are increasingly needed for 21st century careers (Rieley, 2018, Vought and Droegemeier, 2020). Ecological forecasts, which are future, out-of-

sample model predictions of ecological variables with quantified uncertainty (Table 1), can serve as useful decision support tools for a variety of users (Tulloch et al. 2020, Bodner et al. 2021). Because of the utility of forecasts in both informing decision-making and the testing of ecological theory (Dietze et al. 2018, Lewis et al. 2022a, Carey et al. 2022), ecological forecasting is a rapidly growing sub-field of ecology (Lewis et al. 2022b).

Many near-term (day to decade ahead) ecological forecasts are developed using the iterative forecasting cycle (Lewis et al. 2022b), which has the potential to teach students foundational ecological forecasting concepts (Moore et al. 2022a). The iterative, near-term forecasting cycle consists of multiple steps, which parallel the scientific method: 1) make a hypothesis about an ecological phenomena, 2) develop or apply an ecological model which represents that hypothesis (e.g., a population-based model to predict bird abundance), 3) quantify uncertainty around predictions, 4) generate a forecast with uncertainty, 5) communicate the forecast to users, 6) assess the forecast with observations, and 7) update the forecast with new data (Dietze et al. 2018, Moore et al. 2022a). Altogether, teaching this iterative framework in ecology courses could improve student understanding of complex ecological concepts (Selutin and Lebedeva 2017), as well as uncertainty visualization skills.

Developing educational materials on ecological forecast communication and interpretation presents several unique challenges. First, forecasts are inherently uncertain, yet they are needed to guide environmental management decisions, making it critical to properly communicate the uncertainty associated with forecast predictions (Berthet et al. 2016). Second, while there are numerous studies on visualizing data uncertainty (Olston and Mackinlay 2002, Potter et al. 2012, Smith Mason et al. 2017, Wiggins et al. 2018), little consensus has emerged as to the best approach for visualizing forecast uncertainty for both end user comprehension and

decision support. Third, it has been well-documented that different approaches to visualizing uncertainty result in varying levels of comprehension by users (Ramos et al. 2013, Cheong et al. 2016, McKenzie et al. 2016, Kinkeldey et al. 2017). Altogether, these challenges emphasize the need for thoughtful representation of uncertainty in forecasts, as well as the need for educational materials that teach students how to interpret and develop forecast visualizations for decision support applications.

Several pedagogical methods may be useful for incorporating uncertainty visualization skills into introductory ecological forecasting education. First, having students create their own visualizations has been shown to improve data visualization literacy (Huron et al. 2014, Börner et al. 2016, 2019, Alper et al. 2017). Second, teaching students how to produce a range of visualizations for the same forecast using a toolbox of different visualization styles may enable them to communicate their forecast to a broader range of users, as well as adapt their visualizations for different user needs. For example, teaching students how to communicate uncertainty in a single forecast using multiple methods (e.g., representing uncertainty with numbers, words, icons, and graphs such as maps or time series; *sensu* Spiegelhalter et al. 2011) can help illustrate the multitude of ways uncertainty can be visualized and build students' ability to interpret diverse forecast visualizations. Third, teaching students to communicate forecast uncertainty using thresholds which are directly meaningful for decision-making has proven utility in uncertainty communication (Kox et al. 2018). For example, communicating a forecast of the abundance of an endangered species as a forecast index (e.g., the likelihood of encountering that endangered species at a site) may be a more effective communication style for some forecast users by placing forecast output in a decision-making context (see Table 1 for definitions). Fourth, emphasizing the importance of identifying forecast users and specifically

the decisions which could be made with forecasts could increase the relevance of ecological forecasting for students. Presenting ecological concepts in culturally and societally relevant contexts is known to stimulate student engagement (Cid and Pouyat 2013, Vance-Chalcraft and Jelks 2022, Henri et al. 2022), and can lead to more collaborative and effective research and management broadly within the scientific community (Armitage et al. 2009, Cvitanovic et al. 2013).

In addition to the pedagogical approaches above, integrating the concepts of decision science (e.g., through structured decision-making or decision use cases, see Table 1 for definitions; Clemen and Reilly 2004, Gregory et al. 2012) may help students better understand the needs of different forecast users, and correspondingly lead to improved forecast visualizations. Current ecological forecasting teaching materials have largely been methodology-focused, omitting application and communication components (Willson et al. 2022). This focus on methods skill-building, while very valuable, may fail to engage students who have yet to master the computational and quantitative skills needed for forecast development. For example, while development of methodological skill-building is critical to training the next generation of ecological forecasters, the technical skills associated with uncertainty quantification or data assimilation (for example), may be initially intimidating to ecology students who have not yet had substantial mathematical coursework. Initially learning about forecasting through the lens of data visualization and decision science, which robustly demonstrates the utility of ecological forecasting, can help provide students with motivation to master the technical quantitative skills needed for this emerging discipline.

To introduce students to key concepts in uncertainty visualization and communication in the context of using near-term ecological forecasts for real-world decision-making, we developed

a 3-hour teaching module, “Using Ecological Forecasts to Guide Decision-Making,” as part of the Macrosystems EDDIE (Environmental Data-Driven Inquiry and Exploration; MacrosystemsEDDIE.org) program. The module entailed a short introductory lecture, three scaffolded, hands-on forecasting activities embedded within an online interactive tool built using an R Shiny application (Chang et al. 2022) and discussion questions. Instructors were also provided with a pre-module student handout which included suggested readings and discussion questions to provide students with background information before beginning hands-on module activities. To test the effectiveness of our interactive teaching module on students’ ability to learn uncertainty communication and foundational ecological forecasting concepts within a decision support framework, we conducted pre- and post-module assessment surveys. We analyzed the student assessment data to determine how completion of the module affected: 1) students’ ability to interpret and communicate uncertainty in forecast visualizations, and 2) students’ understanding of foundational ecological forecasting concepts.

Methods

Module Overview

We designed Macrosystems EDDIE (Environmental Data-Driven Inquiry and Exploration; MacrosystemsEDDIE.org) Module 8 “Using Ecological Forecasts to Guide Decision-Making” to teach students uncertainty communication and foundational ecological forecasting concepts within a decision support framework (Figure 1). This is the 8th module in the Macrosystems EDDIE teaching module series (Carey et al. 2020, Hounshell et al. 2021, Moore et al. 2022a). Specifically, the module activities encompassed a range of decision support concepts and applications, such as structured decision-making through role-playing and

identification of forecast user needs. The version of the module used for this study is archived and available for download from Woelmer et al. (2022a, 2022b). All module materials are publicly available for use and are iteratively updated following user feedback; the most recent version of the module can be accessed at:

https://serc.carleton.edu/eddie/teaching_materials/modules/module8.html. Our assessment focused on measuring student understanding of uncertainty communication and foundational ecological forecasting as two important yet currently overlooked concepts within undergraduate ecology curricula (Willson et al. 2022).

This module, following the Macrosystems EDDIE pedagogical framework (Carey et al. 2020), consisted of a suite of three self-contained, scaffolded activities (Activities A, B, and C) which can be adapted to meet the needs of individual lecture or laboratory classes. The three activities taught students different ways to visualize forecasts (Activity A); how uncertainty in forecast visualizations can influence decision-making (Activity B); and how to create visualizations of probabilistic ecological forecasts tailored to a specific user (Activity C). All Macrosystems EDDIE modules follow the 5E Instructional Model (Bybee et al. 2006), which uses activities to enable engagement, exploration, explanation, elaboration, and evaluation. This module, as well as other Macrosystems EDDIE modules, are primarily geared towards the undergraduate level but can also be applied in graduate-level courses (e.g., Moore et al. 2022a).

Because uncertainty interpretation and communication are not commonly integrated into undergraduate ecology education (Willson et al. 2022), this module introduced students to a broad suite of methods currently applied in visualization and decision science throughout the module activities. These methods include: 1) creating one's own visualizations (Huron et al. 2014, Alper et al. 2017, Börner et al. 2016, Börner et al. 2019), 2) visualizing uncertainty in

multiple ways (*sensu* Spiegelhalter et al. 2011), 3) using meaningful thresholds for decision-making (Kox et al. 2018), 4) identifying forecast users to increase engagement and relevance (Cid and Pouyat 2013, Henri et al. 2022, Vance-Chalcraft and Osborne Jelks, 2022), and 5) considering forecast user decision needs to guide visualization development (Raftery 2016). These four methods are all explicitly taught in this module, with some components included in the three other Macrosystems EDDIE forecasting modules. Specifically, both Module 5: Introduction to Ecological Forecasting (Moore et al. 2022b) and Module 6: Understanding Uncertainty in Ecological Forecasts (Moore et al. 2021) incorporate multiple methods of visualizing uncertainty and Module 7: Using Data to Improve Ecological Forecasts (Lofton et al. 2022) teaches students to consider meaningful thresholds for management decision-making.

Our module assessment (described below) focused on two learning objectives (LOs) taught throughout the module activities. The two LOs were: LO1) describe what ecological forecasts are and how they are used (Activity A, B, C); and LO2) identify different ways to represent uncertainty in a visualization (Activity A, B, C). In addition to LO1 and LO2, this module included four additional LOs for instructors: LO3) identify the components of a structured decision (Activity B); LO4) discuss how forecast uncertainty relates to decision-making (Activity A, B, C); LO5) match forecast user needs with different levels of forecasting decision support (Activity A, C); and LO6) create visualizations tailored to specific forecast users (Activity C). The activities within the module were designed to meet all six LOs, with several activities targeting multiple LOs (Appendix S1: Table S1). Our focus on LO1 and LO2 for the assessment was motivated by the importance of increasing representation of foundational ecological forecasting and uncertainty communication concepts, respectively, in undergraduate ecology curricula (Appendix S1: Table S1).

The module was taught and assessed with >250 undergraduate students in upper-level ecology, freshwater ecology, or zoology courses across four institutions (Table 2). The module was designed for students with little to no background in ecological modeling, ecological forecasting, or data visualization, although a general background in ecology would be helpful. All content knowledge needed to complete the module (e.g., defining an ecological model, an ecological forecast, or a structured decision-making process) is provided in the introductory PowerPoint presentation and R Shiny application. While the timing of when to introduce this module will vary among instructors and courses, in our experience, instructors have most often used this module at the end of the semester to emphasize the impact of ecology broadly for informing decision-making.

Detailed module description

The module included an introductory PowerPoint lecture, a suite of three activities embedded within an R Shiny application accessed in a web browser, and discussion questions. First, the PowerPoint presentation (~20 minutes) introduced students to the key concepts taught in the module, including a general introduction to ecological forecasting and a case study of an ecological forecasting application with visualization examples. Instructor notes for each slide were provided, as well as an ‘Introduction to R Shiny’ guide for students and instructors who were not previously familiar with using R Shiny applications.

For the case study within the introductory PowerPoint lecture, students were given an example of a forecast of the future distribution of the invasive spongy moth (*Lymantria dispar*) and introduced to different types of forecast users and corresponding decisions that different forecast users could make, as well as different ways of visualizing the same forecast for

individual forecast users' decision use cases (Table 1). For example, a homeowner deciding whether to treat the oak trees on their property to prevent spongy moth invasion might benefit from a forecast index visualizing the percent likelihood of spongy moth colonization in a particular location. In contrast, a natural resource manager deciding where to prioritize conservation efforts of a native competitor of spongy moth might prefer a map of spongy moth densities and associated uncertainty across the region. Through the case study, students were shown a range of visualization types that can be altered to suit different decision use cases. Students were taught about how uncertainty can be represented and communicated using several methods, including numbers, words, icons, and graphs. For example, using the same forecast, uncertainty could be communicated with numbers ('22% chance of a spongy moth outbreak'), words ('low risk of spongy moth outbreak'), an icon (showing a 'traffic light' symbol indicating 'green' for low risk), or a graph (a map of the likelihood of an outbreak across a region) (Appendix S1: Figure S2). Within these four categories, students were taught how to communicate forecast output (e.g., the density of spongy moths in a given area, see Table 1 for an example), which uses output directly from a forecast model. In addition, they were taught to communicate using a forecast index, which is forecast output that is translated into an index based on some threshold which is meaningful to decision-making (e.g., the likelihood of a spongy moth outbreak; Table 1, Appendix S1: Figure S2).

Second, following the presentation, students were instructed to access the module via the R Shiny application and work through the module activities A, B, and C with a partner. R Shiny is an interactive tool built within the R coding environment that allows users to interact with complex data through a simple web browser interface (Chang et al. 2021, Kasprzak et al. 2021), increasing the ease of use. Applications developed using R Shiny have been proven effective at

teaching students challenging topics in a variety of educational settings (e.g., Fawcett 2018, Moore et al. 2022a). All module activities were designed to meet one or more of the module LOs (Appendix S1: Table S1).

Within the Shiny app, students first completed Activity A, “Explore ecological forecast visualizations and decision use,” in which they individually selected an ecological forecast from a curated list of current forecasting systems (Appendix S1: Table S2), answered several embedded questions about how their selected forecast is visualized and how it can be used, and then compared their answers with their partner. Through these activities, students directly addressed LO1 (‘Describe what ecological forecasts are and how they are used’) by analyzing forecasts and identifying forecast users and LO2 (‘Identify different ways to represent uncertainty in a visualization’) by analyzing how or whether their forecast visualizes uncertainty.

In Activity B, “Make decisions using an ecological forecast,” students completed an in-depth case study in which they role-played as resource managers and made decisions about optimizing multiple objectives using two different forecast visualizations (Figure 2A). The use of role-playing as an active form of learning has documented success in education, especially in science education (Howes and Cruz 2009), but has not been tested in ecological forecasting education specifically. Students were given a case study in which they were asked to role-play as water managers and make decisions about whether or not they should allow a swimming race in a drinking water reservoir given different forecasts of potentially toxic algal blooms occurring at the time of the race (see Appendix S1: Text S1 for a full description of the case study scenario).

As part of Activity B, students were taught to use structured decision-making techniques to apply their management objectives for the drinking water case study. Specifically, students were taught the ProACT structured decision-making tool (see Table 1 for definition, e.g.,

Hammond et al. 2002, Hemming et al. 2022). With a goal of optimizing four different management objectives identified using the PrOACT tool (Figure 2A.3), students created hypotheses about how to manage the drinking water reservoir each day as the forecasts were iteratively updated over time (Figure 2A.1; Appendix S1: Figure S3). They completed this objective twice, using forecast visualizations which represented uncertainty using two different methods (Figure 2A.1, 2A.2). Students were encouraged to work through this activity independently and consult with their partner as needed. Finally, students answered questions about how different forecast visualizations influenced their ability to make decisions about managing the reservoir. The culminating discussion of Activity B asked students to discuss how they might improve or alter the visualizations for their decision needs as a water resource manager. Students addressed LO1 in Activity B by using ecological forecasts to make decisions, and LO2 by making decisions using different types of uncertainty visualizations.

In Activity C, “Create a customized visualization of an ecological forecast for a forecast user,” students worked individually to choose a different forecast user that was not a drinking water manager (e.g., a swimmer) of the same drinking water forecast they used in Activity B (Figure 2B-C). Students identified a decision to be made by their forecast user (e.g., whether or not to go swimming in a lake based on an algae threshold). Based on the decision that they identified, students created a customized forecast visualization for their user. Additionally, students explored the underlying forecast distribution by examining the mean, median, and upper ranges of the forecast to better understand the uncertainty underlying the forecast. Lastly, students compared their visualizations with their partner, who chose a different forecast user. This Activity C advanced student understanding of LO1 by connecting the forecast to a variety of potential users. By comparing across forecast users, students were also encouraged to think

about how different users might benefit from different types of visualizations (Figure 2B and 2C), contributing to their understanding of LO2.

At the end of Activity C (as well as between completion of each activity, time permitting), instructors were guided to bring the student pairs back together for a full group discussion and answer any remaining questions. A list of discussion questions for the instructor to use as prompts was provided for each Activity in the Instructor Manual. For example, to recap Activity A, instructors could ask students to discuss how they were able to tell whether visualizations included uncertainty and if there were some types of visualizations that made it more or less difficult to recognize and interpret forecast uncertainty. For Activity B, instructors could ask students to present their decisions in the case study and explain how the trade-offs among their management objectives influenced their decision-making. Lastly, for Activity C, instructors could ask students to discuss the visualization that they chose for their forecast user and how it related to their forecast user's decision needs, as well as what they would do if they had to create a visualization which served multiple forecast user needs.

Instructional Information and Accessibility

As noted above, all materials for teaching this module are publicly available. The teaching materials (introductory lecture, introduction to R Shiny guide, pre-module student handout, and instructor's manual) are archived in the Environmental Data Initiative repository (Woelmer et al. 2023a). The R Shiny application code is archived in the Zenodo repository (Woelmer et al. 2022) and can be accessed at <http://module8.MacrosystemsEDDIE.org>. To access the R Shiny application, students only need an internet connection and a web browser. For students without consistent access to an internet connection, the R Shiny code can be

downloaded from Zenodo (Woelmer et al. 2022) and run locally on a computer using R and RStudio software.

Macrosystems EDDIE Module 8 is designed to be taught either in-person, virtually, or in a hybrid modality. The Instructor's Manual includes instructions on best practices for facilitating the module in each modality. For example, we recommend using Zoom breakout rooms with four students per room (two pairs of students) if teaching virtually, or having students sit together in pairs if teaching in-person.

To increase accessibility for users, the module includes alternative (alt) text descriptions of all images. In addition, text throughout the module was adjusted to have sufficient levels of contrast for improved readability, following Moore et al. (2022a).

Module Assessment

To assess the effects of our module on student learning and answer our research questions, we administered pre- and post-module assessment surveys to undergraduate students before and after module completion, respectively. A detailed description of the development of the module assessment questionnaire can be found in Text S2. In total, we tested the module in four undergraduate courses at four different universities with $N = 314$ consenting students and 7 unique instructors (Table 2). Not all students completed every question so the number of responses per assessment question varied. All students who completed the assessment were undergraduates in their second year or later and were enrolled in General Ecology, Zoology, or Freshwater Ecology courses. Because the module was taught across a variety of institutions, course types, classroom formats, and student experience levels, we were not able to control for these variables in our design, and thus focused our analysis on the total pool of consenting

students who completed the module. Instructors were recruited via personal communication, participation in conference workshops, or through an email listserv. The module was taught both virtually and in-person (Table 2), though the majority of students (92%) completed the module with in-person instruction.

As described above, the goal of the assessment was to measure the effects of the module on students' ability to understand foundational ecological forecasting concepts (LO1) and uncertainty communication (LO2; Figure 1, Table 3). We grouped the questions by LOs, resulting in three questions which measured foundational ecological forecasting concepts (LO1) and five questions which measured uncertainty communication concepts (LO2, Figure 1).

The assessment included multiple-choice and qualitative, open-ended questions (Table 3). Pre- and post-surveys were identical and administered via an online, secure portal run by the Science Education Research Center at Carleton College. All students and faculty consented to participate in the study per our Institutional Review Board (IRB) protocols (Virginia Tech IRB 19-669 and Carleton College IRB 19-20 065).

Analysis of assessment surveys

We analyzed multiple-choice and qualitative assessment questions from the pre- and post-module surveys. Multiple-choice questions (Q1-2, 5-9) were scored by whether students selected the correct answer. Qualitative questions (Q3-4) were scored using a rubric developed by two Macrosystems EDDIE coordinators, following a standardized two-step process (see Appendix S1: Text S3 for methodology), based on the rubric methodology of Moore et al. (2022a) and Miles et al (2020). A detailed description of the coding criteria for both Q3 and Q4 is included in Appendix S1: Tables S5 and S6, respectively. We also screened answers to Q4

(Table 3) for the presence of three keywords related to uncertainty communication ('icon', 'color', and 'forecast output/index'). We recorded whether the keywords were present or absent in student responses but did not consider responses correct unless students also explained how the keywords were used to communicate uncertainty.

To determine the overall performance within and across LO1 (foundational ecological forecasting) and LO2 (uncertainty communication), we calculated the percent correct within each LO (i.e., resulting in a score for LO1 and LO2) for each student. For the two qualitative questions, which included multiple open-ended responses, student responses were considered 'correct' if they identified at least one benefit of ecological forecasting (Q3) and at least one way of communicating uncertainty (Q4). For Q3, we also calculated the number of students who identified each of the categories of benefits for decision-making (e.g., management or policy, understanding of ecological systems or models)

For all assessment questions, we calculated the percentage of student responses given for each question within the pre- and post-assessment responses and compared performance. We used paired Wilcoxon signed-rank tests to analyze the differences between pre- and post-survey responses on both multiple-choice and qualitative questions as well as the grouped categories. Due to varying class sizes, instruction, student experience levels, and teaching modalities across the four institutions, all data were pooled and analyzed together. Statistical significance was defined as $p < 0.05$. All analyses were conducted in R version 4.2.1 (R Core Team, 2022).

Results

Our assessment data indicate that student understanding of foundational ecological forecasting and uncertainty communication concepts increased after module completion (Figures 3, 4). Specifically, students identified significantly more ways to communicate uncertainty in a

forecast, and were significantly more likely to identify ‘decision-making’ and ‘prediction’ as important benefits of ecological forecasts (Figure 4). Across the two LOs, students scored higher in foundational ecological forecasting concepts after completing the module, but showed strong growth in understanding both ecological forecasting and uncertainty communication concepts from pre- to post-module surveys (Figure 5).

Student understanding of uncertainty communication

Students were more likely to correctly identify and describe multiple ways to communicate uncertainty in forecast visualizations after completion of the module (Figures 3, 4b, 4d; Table 4). The percent of students able to correctly distinguish among different ways to visualize forecast uncertainty increased from 33% pre-module to 60% after module completion ($p < 0.001$; Figure 3e, Table 4). In addition, students were significantly more likely to identify and interpret differences between two visualizations that had varying representations of uncertainty after module completion (33% of students pre-module vs. 49% post-module; $p < 0.001$; Figure 3d, Table 4). We also observed post-module increases in the percent of students who correctly interpreted a forecast visualization (49% pre-module; 52% post-module) and matched a forecast visualization with a forecast user decision need (42% pre-module; 49% post-module), but these increases were not statistically significant ($p = 0.34$ and 0.28 , respectively; Figure 3c, 3f; Table 4).

Students also showed increased comprehension of uncertainty communication after module completion in our qualitative assessment. When asked to describe two different ways to communicate uncertainty, the number of correct answers students provided increased from an average of 0.2 ± 0.5 on the pre-survey (S.D.) to 1.1 ± 0.8 on the post-survey ($p < 0.001$; Figure

4d, Table 4). Specifically, the number of students who identified numeric, visual, or probabilistic methods to visualize uncertainty increased significantly after module completion (all $p < 0.001$; Figure 4b), while student responses which included text or multiple predictions increased, but not significantly (both $p = 0.18$; Figure 4b, Table 4). The number of students who identified ‘numeric’ methods to visualize uncertainty (e.g., ‘standard deviation’, ‘statistical confidence intervals’, ‘intervals’, or ‘ranges’) increased from 8% on the pre-survey to 29% on the post-survey ($p < 0.001$; Figure 4b, Table 4). Additionally, student responses which included ‘visual’ descriptions of uncertainty (e.g., ‘boxplots’, ‘shaded area around a line’, ‘error bars’) increased from 8% to 41% ($p < 0.001$; Figure 4b, Table 4), while responses including ‘probabilistic’ methods to visualize uncertainty (e.g., ‘percentage likelihood’, ‘probability of exceeding a certain threshold’) increased from 2% to 15% from the pre-survey to the post-survey ($p < 0.001$; Figure 4b, Table 4).

Several keywords were significantly more prevalent in post-module than pre-module responses to the question about identifying uncertainty communication methods (Q4, Table 3). None of our keywords were identified in the pre-survey responses, but 4% ($n = 9$) of students included the word ‘icon’ and 13% ($n = 31$) of students included ‘color’ in their answers when asked to identify ways of communicating uncertainty in the post-survey. In addition, 11% ($n = 27$) of students named ‘forecast output’ or ‘forecast index’ in their post-survey responses. However, we note that of the 27 students who listed ‘forecast index/output’ as a way to visualize uncertainty in their post-module response, only three students correctly described these terms in the context of uncertainty communication. For example, one student explicitly described what they meant by a ‘forecast index/output’ (‘Forecasts can be visualized through figures that show forecast output, which have direct information on it, and forecast index, which contains a

meaningful threshold that is based off what decision is being made”), demonstrating a deeper understanding than students who mentioned forecast index/output without a definition (e.g., “You can visualize uncertainty with a forecast index and a forecast output”).

Student understanding of foundational ecological forecasting

We found that student understanding of foundational ecological forecasting concepts increased after module completion (Figures 3, 4; Table 4). Students were significantly more likely to correctly define an ecological forecast after completing the module, with an increase from 27% pre-module to 78% of students answering correctly post-module ($p < 0.001$; Figure 3a). Student understanding of how forecast uncertainty changes over time also increased from 52% to 58% after module completion, although this increase was not statistically significant ($p = 0.16$; Figure 3b, Table 4).

We saw a significant increase in the total number of benefits of ecological forecasts identified by students after module completion ($p < 0.001$; Figure 4c, Table 4). Certain benefits were more likely to be mentioned than others in the student responses. Specifically, students were more likely to identify how forecasting can be used for facilitating decision-making (which was included in 13% of pre-module responses and 35% of post-module responses), and predicting a future event (included in 47% of pre-module and 59% of post-module responses; Figure 4a, Table 4). Correspondingly, the number of “I don’t know” responses to the question about forecast benefits significantly decreased from 38% to 8% after module completion (Figure 4a, Table 4). Students also identified other benefits of ecological forecasting in both the pre- and post-survey (e.g., benefits related to management or policy, as well as increased understanding of

ecological systems or models), although these themes were not significantly more prevalent in student responses after module completion (Figure 4a).

Overall, student responses to the open-ended, qualitative question about forecast benefits showed an expansion in their understanding of how ecological forecasts can be used. Several students provided fairly simplistic answers about the benefits of forecasts in their pre-module responses and then more nuanced and complex answers in their post-module responses. For example, one student answered “Limiting the effects of climate change” in their pre-module response as a forecast benefit, while their response after module completion showed a more in-depth understanding of forecast applications: “Ecological forecasting allows people to better understand how environmental conditions will change and how that will impact them.” Another student similarly wrote in their pre-survey response to the forecast benefit question, “Maybe it’ll show the effects of climate change in response to what we do now?” and then expanded to “They [forecasts] can assist in planning future events as well as management or conservation for at risk areas” in their post-survey response.

Performance across ecological forecasting and uncertainty communication learning objectives

Student performance improved from pre- to post-module in correctly answering questions on both foundational ecological forecasting concepts (LO1) and uncertainty communication (LO2; Figure 5). The increase in performance was stronger for students who scored lower on the pre-survey (Figure 5b). Students were more likely to correctly answer questions on ecological forecasting concepts than questions on uncertainty communication concepts on the pre-module assessment (Figure 5a). Both LOs showed strong growth after module completion, with many students who answered zero questions correctly on the pre-survey answering all questions

correctly on the post-survey in both categories (Figure 5b). More students scored 100% (all LO-specific questions answered correctly) on foundational ecological forecasting concepts than on uncertainty communication both before and after module completion, though we note that the number of questions in the two categories differed (Figure 1).

Discussion

Our results indicate that completion of a 3-hour module can significantly improve undergraduate ecology students' understanding of uncertainty communication and ecological forecasting. While the percentage of correct answers increased for all assessment questions after module completion, students were more likely to perform higher on foundational ecological forecasting concepts than uncertainty communication concepts prior to module completion. Higher initial performance on ecological forecasting questions may be because students were more familiar with ecological forecasting concepts relative to uncertainty communication concepts before completing the module. However, students showed more growth in describing multiple ways to communicate uncertainty than in identifying benefits of ecological forecasting (Figure 4c, 4d). Below, we explore the implications of our results for undergraduate education in uncertainty communication, ecological forecasting, and ecology broadly.

Improved uncertainty communication skills and implications for visualization literacy

Students identified significantly more ways to communicate uncertainty following module completion, indicating that the module introduced students to a toolbox of approaches for developing and understanding uncertainty in ecological visualizations. Before completing our module, the majority (85%) of students were unable to describe any ways to communicate

uncertainty, while after module completion 72% were able to describe one or more ways to communicate uncertainty (Figure 4d). Being able to identify and describe multiple methods for uncertainty communication is an important skill, as the method used to visualize uncertainty can have a substantial effect on user comprehension and decision-making (Nadav-Greenberg et al. 2008, Ramos et al. 2013, Cheong et al. 2016, McKenzie et al. 2016, Kinkeldey et al. 2017). For example, using summary visualizations (e.g., boxplots) can decrease users' cognitive load and increase the speed of decision-making, but are more likely to lead to misinterpretation (Ruginski et al. 2016, Correll et al. 2018). In contrast, ensemble-based visualizations (i.e., forecast visualizations that show all possible model outputs) may provide users with more information about the whole spread of uncertainty, but viewers may overweight certain ensemble members, leading to inconsistent decision-making (Padilla et al. 2017a). Given that there is no single "best" visualization method for uncertainty communication due to differences in decision-making needs (Spiegelhalter et al. 2011), the ability to create a variety of visualization options and adapt visualizations based on forecast user feedback is critical for developing effective uncertainty visualizations.

All of the methods for uncertainty communication included in the students' post-module responses are aligned with current state-of-the-art practices for uncertainty communication in visualization science. "Visual" and "numeric," the two uncertainty communication methods most commonly reported by students in post-module responses (Figure 4b), mirror the two key uncertainty representation techniques ("visualization" and "quantification") identified in a recent review (Kamal et al. 2021). Probabilistic methods, which were also significantly more common in student responses following module completion, can decrease cognitive load and increase use and understanding of visualizations for decision support (Kox et al. 2018). Some students ($n = 8$)

also identified “text” as a useful method for uncertainty communication (Figure 4b), but “text” was almost always (7/8 students) reported in addition to another form of communication (e.g., “visual,” “numeric,” or “probability”), following literature which shows that text is most useful for explaining and providing context for visualizations (Carr et al. 2018). Additionally, common keywords throughout student responses included “color” and “icon.” Thoughtful use of color palettes (e.g., by using discrete rather than continuous color palettes) has been shown to be a powerful tool in representing ranges of uncertainty (Padilla et al. 2017b, Correll et al. 2018). Similarly, the use of icons or symbols has been shown to improve user understanding and usability of decision support tools in diverse settings (Galesic et al. 2009, Garcia-Retamero et al. 2010, Zikmund-Fisher et al. 2014, Kamal et al. 2021), potentially by decreasing the cognitive load required to interpret the communication. While neither text, icons, or numbers alone are typically most effective in scientific communication of complex ideas (Larkin and Simon 1987, Tait et al. 2010), the student post-module responses are reflective of a common theme in the visualization literature that using multiple communication forms increases user comprehension and confidence in decision-making (Fagerlin et al. 2005, Spiegelhalter et al. 2011). Ultimately, the module increased students’ ability to communicate uncertainty using multiple approaches, a key skill for developing decision support tools for forecast users.

Our results demonstrate that the module shows promise for increasing visualization literacy and introducing much needed skills in uncertainty communication to undergraduate ecology students. Most students who completed this module had little to no prior experience with uncertainty communication, but showed substantial improvement in performance after module completion, indicating that a 3-hour module can help build these critical skills (Figure 5). Students’ lack of previous exposure to uncertainty communication is likely because

undergraduate ecology classes do not currently include uncertainty communication and visualization literacy topics as often as, e.g., ecological modeling and prediction (Willson et al. 2022). Because the *communication* of uncertainty is just as important as the *quantification* of uncertainty in forecasts for ensuring that forecast visualizations guide end users' decision-making, it is critical that science communication and visualization science, including incorporation of end user decision needs in visualization development, are included in ecological forecasting training (Robinson et al. 2012, Schwartz et al. 2017, Eisenhauer et al. 2021). Overall, given the importance of uncertainty communication not only in ecology, but across scientific disciplines broadly (e.g., medicine, meteorology, economics; Tait et al. 2010, Ferstl et al. 2017, Wesslen et al. 2022), improving students' ability to interpret and produce uncertainty visualizations may help enable student participation in a variety of scientific disciplines. Moreover, providing students with improved visualization literacy and uncertainty communication skills will yield a more data-literate population, regardless of students' future careers. This module provides an important first step for incorporating visualization literacy coursework across undergraduate curricula broadly and initiating training in critical visualization interpretation and communication skills.

Increase in student understanding of foundational ecological forecasting concepts

In addition to expanding students' uncertainty communication skills, completion of the 3-hour module improved students' understanding of foundational ecological forecasting concepts. Following module completion, students were significantly more likely to correctly define an ecological forecast as a future prediction of environmental conditions with uncertainty (Figure 3a; Appendix S1: Table S1). Overall, developing a common definition of "forecast" is important

for furthering the field of ecological forecasting, as having common definitions enables meaningful discourse on topics within and across disciplines, providing a scaffold for interdisciplinary work to address complex socioecological problems (Lélé and Norgaard 2005, Robinson et al. 2012). Given that ecological forecasting is an emerging field (Woelmer et al. 2021), codifying definitions in training materials enables undergraduate ecology students to more effectively discuss and learn forecasting topics.

Student understanding of the benefits of ecological forecasting also significantly increased after module completion. Specifically, we found a significant increase in the number of students who identified ‘decision-making’ and ‘prediction’ as benefits, but saw only a minimal increase for ‘policy,’ and decreases for ‘management’ and ‘understanding.’ Since ‘decision-making’ and ‘prediction’ were emphasized throughout the module, it is unsurprising that these two benefits of ecological forecasting were most commonly provided in student responses. However, the small decrease in responses related to management and understanding is surprising, and may indicate that the module did not sufficiently focus on the benefits of forecasts for management or ecological understanding. For example, while Activity B used a management-centered role-playing example, the activity was primarily focused on the effect of visualization type on decision-making, rather than how forecasts could be integrated into management workflows. Similarly, the module did not emphasize how ecological forecasts can advance understanding of ecosystems and testing of ecological theory or be integrated into policy-making decisions, leaving an opportunity to bolster these forecast applications in future iterations of the module. Alternatively, it is possible that students who identified “decision-making” as a benefit of forecasts may have had policy or management decisions in mind, but not specifically stated this.

We note that students' ability to correctly identify that uncertainty should increase the further into the future a forecast is made (Q2) showed only marginal growth (Figure 3b), leaving room for improvement in teaching this foundational concept of ecological forecasting (*sensu* Dietze et al. 2018). To complement this module and provide additional training in foundational ecological forecasting concepts, we suggest pairing this module with other Macrosystems EDDIE modules (Module 5: Introduction to Ecological Forecasting, Moore et al. 2022b; Module 6: Understanding Uncertainty in Ecological Forecasts, Moore et al. 2021; or Module 7: Using Data to Improve Ecological Forecasts, Lofton et al. 2022).

Integration of decision support concepts into ecology curricula

Integrating applied decision-making concepts into ecological forecasting and uncertainty communication lessons heeds a widespread call to make ecological research more societally relevant (e.g., Belovsky et al. 2004, Ruhl et al. 2022). Training that incorporates components of translational ecology (e.g., science communication, end user engagement, structured decision-making, multidisciplinary training) has long been recommended for ecologists at all career stages (Robinson et al. 2012, Schwartz et al. 2017, Eisenhauer et al. 2021), but resources targeted at the undergraduate level have been lacking (Bakermans and Pfeifer 2018). Our module aims to close this gap by incorporating complex water management decision-making scenarios with multiple end users to engage students in solving real-world problems, while also developing visualization literacy skills. While we did not quantitatively assess the engagement of students in the module, we received open-ended positive feedback from many students. Students reported that the module was enjoyable and important ("I really enjoyed looking at decision analysis in an ecology class;" "I think it's very important to talk about this in science classes"). Additionally,

student responses suggest that they found that the module was interesting (“This was informative and a really interesting way for me to realize the actual impacts of ecological forecasting”) and novel (“It was helpful because I came in with no information”).

Updating ecology curricula to incorporate materials on visualization literacy and translational ecology concepts (such as this module) could be beneficial for students. From our experience with the Macrosystems EDDIE program, we identified two common pathways for implementing this module: 1) targeting new instructors looking for ready-made materials as they develop syllabi for the first time, and 2) targeting experienced instructors who are looking to update existing syllabi with new, easy-to-use materials on visualization and data science topics. For the second case of revising established courses, it is inevitable that incorporating new material comes with the trade-off of removing other content. We note that these substitutions are likely to be highly variable from instructor to instructor and across different courses (i.e., introductory ecology vs. more specialized ecology courses), and should be determined at the discretion of the instructor. Some instructors have used this module as a culminating activity at the end of the semester to emphasize the utility of ecology as a discipline for informing environmental decision-making broadly. To further develop students’ knowledge of visualization and decision science, we suggest supplementing this module with additional coursework or courses focused on science communication, data science and visualization, and/or applied ecology material. Ultimately, introducing students to applications of ecological forecasting for real-world decision-making may help recruit students to work in this subdiscipline, as well as highlight the importance of using ecology to produce actionable tools to address societal problems (e.g., Enquist et al. 2017). Many ecological forecasters have already begun to integrate decision support and uncertainty communication components into their forecasts by making

forecasts which are actionable, useful, and targeted towards forecast users (e.g., Gerst et al. 2019, Turner et al. 2020, Jackson-Blake et al. 2022). Our experience with this module indicates that even a short (3-hour) exposure to decision support and uncertainty communication concepts can increase students' understanding of potential applications and benefits of ecological forecasting.

Module caveats and opportunities for future use

The impact of this module on student learning of decision science and uncertainty communication skills could be improved in several aspects. First, during module development, we intentionally introduced students to a single method of structured decision-making (PROACT) and a limited number of uncertainty communication methods (i.e., visual, numeric, probabilistic) to provide a simplified introduction to the decision and visualization sciences. While students showed successful understanding of the PrOACT tool (Appendix S1: Figure S4), the addition of other decision support components, including solutions-oriented decision-making theory (Deitrick and Wentz 2015) or additional methods of structured decision-making (Gregory et al. 2012), would increase students' breadth of understanding of decision science. Inclusion of a broader variety of decision support concepts could also lead to improved performance on the decision-related questions (e.g., Figure 3c, 3f). Second, allowing students more control over visualizations within the Shiny app (e.g., additional visualization or personalization options, inclusion of R-based coding activities) would likely increase students' visualization literacy (Huron et al. 2014, Alper et al. 2017, Börner et al. 2016, Börner et al. 2019). Third, due to time constraints, our module asked students to imagine what type of forecast visualization would best meet different end users' needs, rather than asking them to actively engage in co-development of visualizations with different forecast users, which would likely increase the utility of the

visualization (Raftery 2016, Padilla et al. 2017b, Gerst et al. 2019). Fourth, shifting the focus of the case study in Activity B to be customizable for specific, nearby ecosystems which are directly relevant to students' everyday lives could potentially increase engagement and student learning (Cid and Pouyat 2013, Henri et al. 2022, Vance-Chalcraft and Osborne Jelks, 2022). For example, students living in areas where wildfires are common may be more engaged analyzing a case study presenting a decision-making scenario on wildfire forecasts. While we recognize the value in including additional content on uncertainty communication, decision science, and ecological forecasting, we note that expanding the module may make it less feasible for instructors to add into their ecology curricula.

Several caveats should be considered when interpreting the assessment results from our module. First, we used a pre- and post-module methodology because instructors were unable to divide their classes into treatment and control groups for instruction. Second, there were many factors which were not held constant across the classrooms that tested our module, including student experience level, instructor experience level, classroom size, institutional familiarity with forecasting and others that could influence the effect of the module on individual student learning. Third, due to the length of the module and limitations of our assessment survey, our analysis provides only a limited understanding of students' knowledge gain. Future longer-term assessments are needed to assess student growth over a longer duration of time. Lastly, while we did not collect student responses to embedded questions within the module for analysis, a future analysis which relates student answers to within-module questions could provide valuable insight on student engagement with the material and performance on post-assessment responses.

Conclusions

Communication of model uncertainty is of paramount importance for advancing the utility of ecological research findings for decision-making. Our teaching module provides an introduction to concepts and skills needed for ecology students to increase their visualization literacy, engage in data science applications, and develop decision support tools. Introduction to ecological forecasting concepts at an early educational stage, including an improved understanding of the importance of ecological forecasting for societal benefit, is increasingly necessary for training the next generation of predictive ecologists to meet both European (Nativi et al. 2021) and U.S. government agency directives (Vought and Droegemeier, 2020; Arsenault et al. 2020; NOAA, 2022; CDC 2022). Moreover, by teaching ecological forecasting and uncertainty communication skills via a real-world decision-making scenario, this module helps to emphasize the relevance and lower the barrier of entry to ecology. Through an R Shiny interface that is easy to implement for educators in a range of classroom experience levels, this 3-hour, adaptable module fills a critical gap in undergraduate ecology curricula. By introducing students to uncertainty communication and ecological forecasting early in their careers, this module can help train the next generation of ecologists to conduct societally relevant research and tackle pressing ecological challenges.

Acknowledgements

We are grateful to the students who participated in this module and their instructors, especially Kait Farrell, Hilary Dugan, Kait Reinl, Sheila Lyons-Sobaski, Matt Aiello-Lammens, and graduate teaching assistants, who provided constructive feedback which substantially improved the quality of the module. We thank Melissa Kenney for her invaluable feedback on module activities, as well as the Carey and Schreiber Labs for their feedback throughout module development, especially Caroline Bryant, Arpita Das, and Jacob Wynne, who completed pilot versions of this module. We thank Kait Farrell and Alex Hounshell for their contributions to the Macrosystems EDDIE program, as well as Kristin O’Connell, Monica Bruckner, Ashley Carlson, Cailin Huyck Orr, Ellen Iverson, and the Science Education Resource Center at Carleton College for coordinating website development and survey administration. Lastly, we thank members of the Ecological Forecasting Initiative and the Global Lake Ecological Observatory Network for providing support and feedback on module content and fostering community around these topics. Our study was financially supported by NSF DEB-1926050, DBI-1933016, and DEB-1926388.

Author Contributions

CCC developed the Macrosystems EDDIE program. CCC and RQT expanded the EDDIE program for forecasting and acquired funding for the project. WMW and CCC conceived of the study design. WMW led the development of module materials, with substantial guidance from RQT, CCC, and TNM. TNM assisted with R Shiny app development, as well as administration of module assessment. WMW, TNM, and MEL co-analyzed qualitative assessment questions. WMW led the analysis of module assessment results and manuscript writing, with significant

input from CCC and MEL. All authors edited and approved of the final manuscript.

References

- Alper, B., N. H. Riche, F. Chevalier, J. Boy, and M. Sezgin. 2017. Visualization literacy at elementary school. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. ACM, Denver Colorado USA. Pages 5485–5497.
<http://dx.doi.org/10.1145/3025453.3025877>
- Armitage, D. R., R. Plummer, F. Berkes, R. I. Arthur, A. T. Charles, I. J. Davidson-Hunt, A. P. Diduck, N. C. Doubleday, D. S. Johnson, M. Marschke, P. McConney, E. W. Pinkerton, and E. K. Wollenberg. 2009. Adaptive co-management for social-ecological complexity. *Frontiers in Ecology and the Environment* 7: 95–102.
<http://www.jstor.org/stable/25595062>
- Arsenault, K. R., Shukla, S., Hazra, A., Getirana, A., McNally, A., Kumar, S. V., ... Verdin, J. P. 2020. The NASA hydrological forecast system for food and water security applications. *Bulletin of the American Meteorological Society*, 101(7): E1007–E1025.
<https://doi.org/10.1175/BAMS-D-18-0264.1>
- Bakermans, M. H., and Pfeifer, G. 2018. A model for translational science in undergraduate classrooms. *Frontiers in Ecology and the Environment*, 16(6): 319–321.
<https://doi.org/10.1002/fee.1920>
- Belia, S., F. Fidler, J. Williams, and G. Cumming. 2005. Researchers misunderstand confidence intervals and standard error bars. *Psychological Methods* 10: 389–396.
<https://doi.org/10.1037/1082-989X.10.4.389>

- Belovsky, G. E., D. B. Botkin, T. A. Cowl, K. W. Cummins, J. F. Franklin, M. L. Hunter, A. Joern, D. B. Lindenmayer, J. A. MacMahon, C. R. Margules, and J. M. Scott. 2004. Ten suggestions to strengthen the science of ecology. *BioScience* 54: 345–351.
<https://doi.org/10.1007/s10531-005-2631-1>
- 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.
<https://doi.org/10.1051/e3sconf/20160718005>
- Bird, J. P., B. K. Woodworth, R. A. Fuller, and J. D. Shaw. 2021. Uncertainty in population estimates: A meta-analysis for petrels. *Ecological Solutions and Evidence* 2: 1–13.
<https://doi.org/10.1002/2688-8319.12077>
- Bodner, K., C. Rauen Firkowski, J. R. Bennett, C. Brookson, M. Dietze, S. Green, J. Hughes, J. Kerr, M. Kunegel-Lion, S. J. Leroux, E. McIntire, P. K. Molnár, C. Simpkins, E. Tekwa, A. Watts, and M. J. Fortin. 2021. Bridging the divide between ecological forecasts and environmental decision making. *Ecosphere* 12: e03869.
<https://doi.org/10.1002/ecs2.3869>
- Bonneau, G., H. Hege, C. R. Johnson, M. M. Oliveira, K. C. Potter, P. Rheingans, and T. Schultz. 2015. Chapter 1: Overview and state-of-the-art of uncertainty visualization in *Scientific Visualization*, pages 3-27. Springer.
- Börner, K., A. Bueckle, and M. Ginda. 2019. Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. *Proceedings of the National Academy of*

Sciences of the United States of America 116: 1857–1864.

<https://doi.org/10.1073/pnas.1807180116>

Börner, K., A. Maltese, R. N. Balliet, and J. Heimlich. 2016. Investigating aspects of data visualization literacy using 20 information visualizations and 273 science museum visitors. *Information Visualization* 15: 198–213.

<https://doi.org/10.1177/1473871615594652>

Boukhelifa, N., and D.J. Duke. Uncertainty visualization - why might it fail? 2009. In: Conference on Human Factors in Computing Systems - Proceedings (April), pp. 4051–4056. <https://doi.org/10.1145/1520340.1520616>

Briggs, D. J., C. E. Sabel, and K. Lee. 2009. Uncertainty in epidemiology and health risk and impact assessment. *Environmental Geochemistry and Health* 31: 189–203.

<https://doi.org/10.1007/s10653-008-9214-5>

Bybee, R. W., J. A. Taylor, A. Gardner, P. V. Scotter, J. C. Powell, A. Westbrook, and N. Landes. 2006. The BSCS 5E Instructional Model: Origins, effectiveness, and applications. Colorado Springs, CO, USA.

Carey, C. C., K. J. Farrell, A. G. Hounshell, and K. O’Connell. 2020. Macrosystems EDDIE teaching modules significantly increase ecology students’ proficiency and confidence working with ecosystem models and use of systems thinking. *Ecology and Evolution* 10: 12515–12527. <https://doi.org/10.1002/ece3.6757>

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, and R. Q. Thomas. 2022. Advancing lake and reservoir water quality management with near-term, iterative ecological forecasting. *Inland Waters* 12: 107–120. <https://doi.org/10.1080/20442041.2020.1816421>
- Carr, R. H., B. Montz, K. Semmens, K. Maxfield, S. Connolly, P. Ahnert, R. Shedd, and J. Elliott. 2018. Major risks, uncertain outcomes: Making ensemble forecasts work for multiple audiences. *Weather and Forecasting* 33: 1359–1373. <https://doi.org/10.1175/WAF-D-18-0018.1>
- Centers for Disease Control and Prevention. 2022. CDC Launches New Center for Forecasting and Outbreak Analytics. Press Release. 19 April 2022. <https://stacks.cdc.gov/view/cdc/116460>
- Chang, W., J. Cheng, J. J. Allaire, C. Sievert, B. Schloerke, Y. Xie, J. Allen, J. McPherson, A. Dipert, B. Borges. 2022. shiny: Web Application Framework for R. <https://shiny.rstudio.com/>
- Cheong, L., S. Bleisch, A. Kealy, K. Tolhurst, T. Wilkenning, and M. Duckham. 2016. Evaluating the impact of visualization of wildfire hazard upon decision-making under uncertainty. *International Journal of Geographical Information Science* 30: 1377–1404.
- Cid, C. R., and R. V. Pouyat. 2013. Making ecology relevant to decision making: the human-centered, place-based approach. *Frontiers in Ecology and the Environment* 11: 447–448. <https://doi.org/10.1890/1540-9295-11.8.447>
- Clemen, R. T., and T. Reilly. 2004. Making hard decisions with decision tools suite. 1st edition. Cengage Learning, Pacific Grove, Calif.

Correll, M., D. Moritz, and J. Heer. 2018. Value-suppressing uncertainty palettes. Conference on Human Factors in Computing Systems - Proceedings 2018-April. 1–11.

<https://doi.org/10.1145/3173574.3174216>

Cvitanovic, C., S. K. Wilson, C. J. Fulton, G. R. Almany, P. Anderson, R. C. Babcock, N. C.

Ban, R. J. Beeden, M. Beger, J. Cinner, K. Dobbs, L. S. Evans, A. Farnham, K. J.

Friedman, K. Gale, W. Gladstone, Q. Grafton, N. A. J. Graham, S. Gudge, P. L. Harrison,

T. H. Holmes, N. Johnstone, G. P. Jones, A. Jordan, A. J. Kendrick, C. J. Klein, L. R.

Little, H. A. Malcolm, D. Morris, H. P. Possingham, J. Prescott, R. L. Pressey, G. A.

Skilleter, C. Simpson, K. Waples, D. Wilson, and D. H. Williamson. 2013. Critical research needs for managing coral reef marine protected areas: Perspectives of academics and managers. *Journal of Environmental Management* 114: 84–91.

<https://doi.org/10.1016/j.jenvman.2012.10.051>

Deitrick, S., and E. A. Wentz. 2015. Developing implicit uncertainty visualization methods motivated by theories in decision science. *Annals of the Association of American Geographers* 105(3): 531–551. <https://doi.org/10.1080/00045608.2015.1012635>

Dietze, M. C. 2017. Ecological Forecasting. Princeton: Princeton University Press.

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.

<https://doi.org/10.1073/pnas.17102311>

- Eisenhauer, E., Williams, K. C., Margeson, K., Paczuski, S., Hano, M. C., and Mulvaney, K. 2021. Advancing translational research in environmental science: The role and impact of social sciences. *Environmental Science and Policy*, 120: 165–172.
<https://doi.org/10.1016/j.envsci.2021.03.010>
- Enquist, C. A. F., Jackson, S. T., Garfin, G. M., Davis, F. W., Gerber, L. R., Littell, J. A., ... Shaw, M. R. 2017. Foundations of translational ecology. *Frontiers in Ecology and the Environment*, 15(10): 541–550. <https://doi.org/10.1002/fee.1733>
- Fagerlin, A., C. Wang, and P. A. Ubel. 2005. Reducing the influence of anecdotal reasoning on people’s health care decisions: Is a picture worth a thousand statistics? *Medical Decision Making* 25: 398–405. <https://doi.org/10.1177/0272989X05278931>
- Fawcett, L. 2018. Using interactive Shiny applications to facilitate research-informed learning and teaching. *Journal of Statistics Education* 26: 2–16.
<https://doi.org/10.1080/10691898.2018.1436999>
- Ferstl, F., M. Kanzler, M. Rautenhaus, and R. Westermann. 2017. Time-hierarchical clustering and visualization of weather forecast ensembles. *IEEE Transactions on Visualization and Computer Graphics* 23: 831–840. <https://doi.org/10.1109/TVCG.2016.2598868>
- Galesic, M., R. Garcia-Retamero, and G. Gigerenzer. 2009. Using icon arrays to communicate medical risks: Overcoming low numeracy. *Health Psychology* 28: 210–216.
<https://doi.org/10.1037/a0014474>

- Garcia-Retamero, R., M. Galesic, and G. Gigerenzer. 2010. Do icon arrays help reduce denominator neglect? *Medical Decision Making* 30: 672–684.
<https://doi.org/10.1177/0272989X10369000>
- Gerst, M. D., M. A. Kenney, A. E. Baer, A. Speciale, J. F. Wolfinger, J. Gottschalck, S. Handel, M. Rosencrans, and D. Dewitt. 2019. Using visualization science to improve expert and public understanding of probabilistic temperature and precipitation outlooks. *Weather, Climate, and Society* 12: 117-133. <https://doi.org/10.1175/WCAS-D-18-0094.1>
- Gregory, R., L. Failing, M. Harstone, G. Long, T. McDaniels, and D. Ohlson. 2012. Structured decision making: A practical guide to environmental management choices. John Wiley and Sons.
- Halpern, B. S., H. M. Regan, H. P. Possingham, and M. A. McCarthy. 2006. Accounting for uncertainty in marine reserve design. *Ecology Letters* 9: 2–11. <https://doi.org/10.1111/j.1461-0248.2005.00827.x>
- Hammond, J. S., R. L. Keeney, and H. Raiffa. 2002. Smart choices: A practical guide to making better decisions. Crown Business, New York, NY.
- Hemming, V., A. E. Camacalang, M. S. Adams, M. Burgman, K. Carbeck, J. Carwardine, I. Chadès, L. Chalifour, S. J. Converse, L. N. K. Davidson, G. E. Garrard, R. Finn, J. R. Fleri, J. Huard, H. J. Mayfield, E. M. Madden, I. Naujokaitis-Lewis, H. P. Possingham, L. Rumpff, M. C. Runge, D. Stewart, V. J. D. Tulloch, T. Walshe, and T. G. Martin. 2022. An introduction to decision science for conservation. *Conservation Biology* 1–16.
<https://doi.org/10.1111/cobi.13868>

- Henri, D. A., L. M. Martinez-Levasseur, J. F. Provencher, C. D. Debets, M. Appaqaq, and M. Houde. 2022. Engaging Inuit youth in environmental research: Braiding Western science and Indigenous knowledge through school workshops. *The Journal of Environmental Education* 53: 261–279. <https://doi.org/10.1080/00958964.2022.2125926>
- Hounshell, A. G., K. J. Farrell, and C. C. Carey. 2021. Macrosystems EDDIE teaching modules increase students' ability to define, interpret, and apply concepts in macrosystems ecology. *Education Sciences* 11(8): 382. <https://doi.org/10.3390/educsci11080382>
- Howes, E., and B. Cruz. 2009. Role-playing in science education: an effective strategy for developing multiple perspectives. *Journal of Elementary Science Education* 21: 33–46.
- Hullman, J. 2020. Why authors don't visualize uncertainty. *IEEE Transactions on Visualization and Computer Graphics* 26: 130–139. <https://doi.org/10.1109/TVCG.2019.2934287>
- Huron, S., S. Carpendale, A. Thudt, A. Tang, and M. Mauerer. 2014. Constructive visualization. Pages 433–442 Proceedings of the 2014 Conference on Designing Interactive Systems. Association for Computing Machinery, New York, NY, USA.
- Jackson-Blake, L. A., F. Clayser, E. De Eyto, A. S. French, M. D. Frías, D. Mercado-Bettín, T. Moore, L. Puértolas, R. Poole, K. Rinke, M. Shikhani, L. Van Der Linden, and R. Marcé. 2022. Opportunities for seasonal forecasting to support water management outside the tropics. *Hydrology and Earth System Sciences* 26: 1389–1406. <https://doi.org/10.5194/hess-26-1389-2022>

- Joslyn, S., and S. Savelli. 2010. Communicating forecast uncertainty: Public perception of weather forecast uncertainty. *Meteorological Applications* 17: 180–195.
<https://doi.org/10.1002/met.190>
- Kamal, A., P. Dhakal, A. Y. Javaid, V. K. Devabhaktuni, D. Kaur, J. Zaiantz, and R. Marinier. 2021. Recent advances and challenges in uncertainty visualization: a survey. *Journal of Visualization* 24: 861–890. <https://doi.org/10.1007/s12650-021-00755-1>
- Kasprzak, P., L. Mitchell, O. Kravchuk, and A. Timmins. 2020. Six years of Shiny in research -- Collaborative development of web tools in R. *The R Journal* 12(2): 20–42.
<https://doi.org/10.32614/RJ-2021-004>
- Kinkeldey, C., A. M. MacEachren, M. Riveiro, and J. Schiewe. 2017. Evaluating the effect of visually represented geodata uncertainty on decision-making: systematic review, lessons learned, and recommendations. *Cartography and Geographic Information Science* 44: 1–21. <https://doi.org/10.1080/15230406.2015.1089792>
- Kox, T., H. Kempf, C. Lüder, R. Hagedorn, and L. Gerhold. 2018. Towards user-orientated weather warnings. *International Journal of Disaster Risk Reduction* 30: 74–80.
<https://doi.org/10.1016/j.ijdrr.2018.02.033>
- Larkin, J. H., and H. A. Simon. 1987. Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science* 11: 65–100. <https://doi.org/10.1111/j.1551-6708.1987.tb00863.x>
- Lechner, A. M., W. T. Langford, S. A. Bekessy, and S. D. Jones. 2012. Are landscape ecologists addressing uncertainty in their remote sensing data? *Landscape Ecology* 27: 1249–1261.
<https://doi.org/10.1007/s10980-012-9791-7>

- Lélé, S., and Norgaard, R. B. (2005). Practicing Interdisciplinarity. *BioScience* 55(11): 967–975.
[https://doi.org/10.1641/0006-3568\(2005\)055\[0967:PI\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2005)055[0967:PI]2.0.CO;2)
- Lewis, A. S. L., C. R. Rollinson, A. J. Allyn, J. Ashander, S. Brodie, C. B. Brookson, E. Collins, M. C. Dietze, A. S. Gallinat, N. Juvigny-Khenafou, G. Koren, D. J. McGlinn, H. Moustahfid, J. A. Peters, N. R. Record, C. J. Robbins, J. Tonkin, and G. M. Wardle. 2022a. The power of forecasts to advance ecological theory. *Methods in Ecology and Evolution* 14: 746-756. <https://doi.org/10.1111/2041-210X.13955>
- 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. 2022b. Increased adoption of best practices in ecological forecasting enables comparisons of forecastability. *Ecological Applications* 32: e2500. <https://doi.org/10.1002/eap.2500>
- Link, J. S., T. F. Ihde, C. J. Harvey, S. K. Gaichas, J. C. Field, J. K. T. Brodziak, H. M. Townsend, and R. M. Peterman. 2012. Dealing with uncertainty in ecosystem models: The paradox of use for living marine resource management. *Progress in Oceanography* 102: 102–114. <https://doi.org/10.1016/j.pocean.2012.03.008>
- Lofton, M.E., T.N. Moore, Thomas, R.Q., and C.C. Carey. 20 September 2022. Macrosystems EDDIE: Using Data to Improve Ecological Forecasts. Macrosystems EDDIE Module 7, Version 1. <https://macrosystemseddieshinyapps.io/module7>.
- Maltese, A., J. Harsh, and D. Svetina. 2015. Data visualization literacy: investigating data interpretation along the novice-expert continuum. *Journal of College Science Teaching* 45: 84.

- McClintock, B. T., J. D. Nichols, L. L. Bailey, D. I. MacKenzie, W. L. Kendall, and A. B. Franklin. 2010. Seeking a second opinion: Uncertainty in disease ecology. *Ecology Letters* 13: 659–674. <https://doi.org/10.1111/j.1461-0248.2010.01472.x>
- McKenzie, G., M. Hegarty, T. Barrett, and M. Goodchild. 2016. Assessing the effectiveness of different visualizations for judgments of positional uncertainty. *International Journal of Geographical Information Science* 30: 221–239. <https://doi.org/10.1080/13658816.2015.1082566>
- Melbourne-Thomas, J., S. Wotherspoon, B. Raymond, and A. Constable. 2012. Comprehensive evaluation of model uncertainty in qualitative network analyses. *Ecological Monographs* 82: 505–519. <https://doi.org/10.1890/12-0207.1>
- Miles, M. B., A. M. Huberman, J. Saldana. 2020. *Qualitative data analysis: A methods sourcebook*, 4th ed.; SAGE Publications Inc.: Thousand Oaks, CA, USA.
- Milner-Gulland, E. J., and K. Shea. 2017. Embracing uncertainty in applied ecology. *Journal of Applied Ecology*. 54:2063–2068. <https://doi.org/10.1111/1365-2664.12887>
- Moore, T. N., Carey, C.C. and Thomas, R. Q. 13 October 2021. Macrosystems EDDIE: Understanding Uncertainty in Ecological Forecasts. Macrosystems EDDIE Module 6, Version 1. <http://module6.macrosystemseddie.org>.
- Moore, T. N., R. Q. Thomas, W. M. Woelmer, and C. C. Carey. 2022a. Integrating ecological forecasting into undergraduate ecology curricula with an R Shiny application-based teaching module. *Forecasting* 4: 604–633. <https://doi.org/10.3390/forecast4030033>

- Moore, T.N., C.C. Carey, and R.Q. Thomas. 2022b. Macrosystems EDDIE Module 5: Introduction to Ecological Forecasting (Instructor Materials) ver 3. Environmental Data Initiative. <https://doi.org/10.6073/pasta/1da866a2eb79be84195e785a4370010c>
- Nadav-Greenberg, L., S. L. Joslyn, and M. U. Taing. 2008. The effect of weather forecast uncertainty visualization on decision making. *Journal of Cognitive Engineering and Decision Making* 2: 24-47
- Nativi, S., Mazzetti, P., and Craglia, M. 2021. Digital ecosystems for developing digital twins of the earth: The destination earth case. *Remote Sensing* 13(11): 1–25. <https://doi.org/10.3390/rs13112119>
- National Oceanic and Atmospheric Administration. 2022. Strategic Plan or Fiscal Year 2022-2026. https://www.noaa.gov/sites/default/files/2022-06/NOAA_FY2226_Strategic_Plan.pdf
- Olston, C., and J. D. Mackinlay. 2002. Visualizing data with bounded uncertainty. Pages 37–40 IEEE Symposium on Information Visualization, 2002. INFOVIS IEEE Comput. Soc, Boston, MA, USA. <https://doi.org/10.1109/INFVIS.2002.1173145>
- Padilla, L. M., I. T. Ruginski, and S. H. Creem-Regehr. 2017a. Effects of ensemble and summary displays on interpretations of geospatial uncertainty data. *Cognitive Research: Principles and Implications* 2: 1–16. <https://doi.org/10.1186/s41235-017-0076-1>
- Padilla, L., P. S. Quinan, M. Meyer, and S. H. Creem-Regehr. 2017b. Evaluating the impact of binning 2D scalar fields. *IEEE Transactions on Visualization and Computer Graphics* 23: 431–440. <https://doi.org/10.1109/TVCG.2016.2599106>

- Potter, K., P. Rosen, and C. R. Johnson. 2012. From quantification to visualization: A taxonomy of uncertainty visualization approaches. *IFIP Advances in Information and Communication Technology* 377: 226–247.
- R Core Team. 2022. R: A Language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Raftery, A. E. 2016. Use and communication of probabilistic forecasts. *Statistical Analysis and Data Mining* 9: 397–410. <https://doi.org/10.1002/sam.11302>
- Ramos, M. H., S. J. Van Andel, and F. Pappenberger. 2013. Do probabilistic forecasts lead to better decisions? *Hydrology and Earth System Sciences* 17: 2219–2232. <https://doi.org/10.5194/hess-17-2219-2013>
- Rieley, M. 2018. Big data adds up to opportunities in math careers : Beyond the numbers: U.S. Bureau of Labor Statistics. <https://www.bls.gov/opub/btn/volume-7/big-data-adds-up.htm>.
- Robinson, P., Genskow, K., and Shaw, B. 2012. Barriers and opportunities for integrating social science into natural resource management: lessons from National Estuarine Research Reserves. *Environmental Management* 998–1011. <https://doi.org/10.1007/s00267-012-9930-6>
- Ruginski, I. T., A. P. Boone, L. M. Padilla, L. Liu, N. Heydari, H. S. Kramer, M. Hegarty, W. B. Thompson, D. H. House, and S. H. Creem-Regehr. 2016. Non-expert interpretations of hurricane forecast uncertainty visualizations. *Spatial Cognition and Computation* 16: 154–172. <https://doi.org/10.1080/13875868.2015.1137577>

- Ruhl, N., P. Crumrine, J. Oberle, C. Richmond, S. Thomas, and S. Wright. 2022. Harnessing the Four-Dimensional Ecology Education Framework to redesign an introductory ecology course in a changing higher education landscape. *Ecosphere* 13:e03857. <https://doi.org/10.1002/ecs2.3857>
- Selutin, V. D., and E. V. Lebedeva. 2017. Teaching probability theory and forecasting-based mathematical statistics to Bachelors of economics. *Advances in Social Science, Education and Humanities Research* 97: 264–268. <https://doi.org/10.2991/cildiah-17.2017.46>
- Smith Mason, J., D. Retchless, and A. Klippel. 2017. Domains of uncertainty visualization research: a visual summary approach. *Cartography and Geographic Information Science* 44: 296–309. <https://doi.org/10.1080/15230406.2016.1154804>
- Spiegelhalter, D., M. Pearson, and I. Short. 2011. Visualizing uncertainty about the future. *Science* 333: 1393–1400. <https://doi.org/10.1126/science.1191181>
- Schwartz, M. W., Hiers, J. K., Davis, F. W., Garfin, G. M., Jackson, S. T., Terando, A. J., ... Brunson, M. W. 2017. Developing a translational ecology workforce. *Frontiers in Ecology and the Environment* 15(10): 587–596. <https://doi.org/10.1002/fee.1732>
- Tait, A. R., T. Voepel-Lewis, B. J. Zikmund-Fisher, and A. Fagerlin. 2010. The effect of format on parents' understanding of the risks and benefits of clinical research: A comparison between text, tables, and graphics. *Journal of Health Communication* 15: 487–501. <https://doi.org/10.1080/10810730.2010.492560>

Tulloch, A. I. T., V. Hagger, and A. C. Greenville. 2020. Ecological forecasts to inform near-term management of threats to biodiversity. *Global Change Biology* 00: 1-13.

<https://doi.org/10.1111/gcb.15272>

Turner, S. W. D., W. Xu, and N. Voisin. 2020. Inferred inflow forecast horizons guiding reservoir release decisions across the United States. *Hydrology and Earth System*

Sciences 24: 1275–1291. <https://doi.org/10.5194/hess-24-1275-2020>

Vance-Chalcraft, H. D., and N. O. Jelks. 2022. Community-engaged learning to broaden the impact of applied ecology: A case study. *Ecological Applications* e2768.

<https://doi.org/10.1002/eap.2768>

Vought, R.T., and K.K. Droegemeier. 2020. “M-20-29: Fiscal Year (FY) 2022 Administration Research and Development Budget Priorities and Cross-Cutting Actions.”

<https://www.whitehouse.gov/wp-content/uploads/2020/08/M-20-29.pdf>.

Wesslen, R., A. Karduni, D. Markant, and W. Dou. 2022. Effect of uncertainty visualizations on myopic loss aversion and the equity premium puzzle in retirement investment decisions.

IEEE Transactions on Visualization and Computer Graphics 28: 454–464.

<https://doi.org/10.1109/TVCG.2021.3114692>

Wiggins, A., A. Young, and M. A. Kenney. 2018. Exploring visual representations to support data re-use for interdisciplinary science. *Proceedings of the Association for Information*

Science and Technology 55: 554–563. <https://doi.org/10.1002/pr2.2018.14505501060>

Willson, A.M., H. Gallo, J.A. Peters, A. Abeyta, N. Bueno Watts, C.C. Carey, T.N. Moore, G.

Smies, R.Q. Thomas, W.M. Woelmer, and J.S. McLachlan. 2022. Assessing

opportunities and inequities in undergraduate ecological forecasting education.

<https://doi.org/10.5281/zenodo/7702393>

Woelmer, W. M., Bradley, L. M., Haber, L. T., Klinges, D. H., Lewis, A. S. L., Mohr, E. J., ...

Willson, A. M. 2021. Ten simple rules for training yourself in an emerging field. *PLoS Computational Biology* 17(10): 1–12. <https://doi.org/10.1371/journal.pcbi.1009440>

Woelmer, W.M., R.Q. Thomas, T.N. Moore, and C.C. Carey. 2023. Macrosystems EDDIE

Module 8: Using Ecological Forecasts to Guide Decision-Making (Instructor Materials) ver 4. Environmental Data Initiative.

<https://doi.org/10.6073/pasta/8bf4a076433f0e9f74f1d764d5bd4c3f>

Woelmer, W.M., T.N. Moore, R.Q. Thomas, and C.C. Carey. 2022. Macrosystems EDDIE

Module 8: Using Ecological Forecasts to Guide Decision-Making (R Shiny application) (v1.1). Zenodo. <https://doi.org/10.5281/zenodo.7074674>

Wu, J., K. B. Jones, H. Li, and O. L. Loucks. 2006. Scaling and uncertainty analysis in ecology. Methods and applications. Springer, New York.

Zikmund-Fisher, B. J., H. O. Witteman, M. Dickson, A. Fuhrel-Forbis, V. C. Kahn, N. L. Exe,

M. Valerio, L. G. Holtzman, L. D. Scherer, and A. Fagerlin. 2014. Blocks, ovals, or people? Icon type affects risk perceptions and recall of pictographs. *Medical Decision Making* 34(4): 443–453. <https://doi.org/10.1177/0272989X1351170>

Table 1. Glossary of uncertainty communication and ecological forecasting terms taught in Macrosystems EDDIE Module 8: “Using Ecological Forecasts to Guide Decision-Making” as well as examples of how each term is applied to real-world, near-term forecasting.

Term	Definition	Example
Ecological forecast	A prediction of a future event with uncertainty	A forecast of the distribution and density of the invasive spongy moth for 1 month into the future which includes uncertainty
Forecast index	A forecast output that translated into thresholds which are meaningful for decision-making	22% chance of spongy moth outbreak in a given location
Forecast output	Future predictions with uncertainty generated using a model	Spongy moth density is 24 individuals/km ² ± 4 individuals/km ²
Forecast user	Anyone who can use a forecast to gain understanding or make a decision	Scientist studying oak tree populations, Homeowner, etc.
Forecast decision use	A specific way in which a forecast is used to inform a decision	A forecast of the density of invasive spongy moth guiding a decision about buying moth insecticide
Forecast decision use cases	Categories of forecast users defined by their decision use needs (adapted from Raftery 2016)	• Casual user: Users who do not require probabilistic forecasts; e.g., a park visitor interested in which areas within the park are affected by spongy moth • Practitioner: Users who need an overall idea of uncertainty; e.g., homeowner deciding to protect oak trees on their land in an area affected by spongy moth • Decision analyst: Users who require detailed information about uncertainty; e.g., a natural resource manager deciding which area of a park to treat for spongy moth invasion
Structured decision-making	A formalized method of analyzing a decision by dissecting its components	PrOACT is a structured decision-making tool which guides users through identifying and analyze the following components of a decision: <u>P</u> roblem <u>O</u> bjective <u>A</u> lternatives <u>C</u> onsequences <u>T</u> rade-offs

Table 2. Summary of courses which participated in assessment of Macrosystems EDDIE Module 8 “Using Ecological Forecasts to Guide Decision-Making.”

Institution	Course Level	Class Name	Carnegie Code	Number of students enrolled	Number of unique instructors	Mode
Albion College	Upper-level undergraduate	Ecology	Baccalaureate College: Arts and Sciences Focus	8	1	In-person
University of Georgia	Upper-level undergraduate	Ecology Lab/ Honors Ecology	Doctoral/ Research University	250	7	In-person
Virginia Tech	Upper-level undergraduate	Freshwater Ecology	Doctoral/ Research University	31	1	In-person
University of Wisconsin-Madison	Upper-level undergraduate	Zoology	Doctoral/ Research University	25	1	Virtual

Note: Carnegie codes are categories of universities based on educational and research activity at each institution (<https://carnegieclassifications.acenet.edu>).

Table 3. Selected assessment questions and their corresponding learning objectives for
Macrosystems EDDIE Module 8: “Using Ecological Forecasts to Guide Decision-Making.”

Assessment Question	Learning Objective (LO)	Assessment question short name	Question style
Q1. Which of the following statements best describes an ecological forecast?	LO1: EF Concepts	Define	MC
Q2. When an ecological forecast is generated, how does the uncertainty of the forecast change as it predicts conditions further into the future?	LO1: EF Concepts	Uncertainty	MC
Q3. List what you think are some benefits of ecological forecasting.	LO1: EF Concepts	Forecast benefits	Q
Q4. Describe two different ways uncertainty can be visualized in a forecast.	LO2: UC Communication	Communication type	Q
Q5. Which of the following is the best description of the forecast presented in Figure 1A?	LO2: UC Communication	Interpret	MC
Q6. The ecological forecasts in Figure 1A and 1B present information differently. Which of the following is true?	LO2: UC Communication	Contrast	MC
Q7. Which of the following is an example of a forecast index, as opposed to a forecast output)?	LO2: UC Communication	Distinguish	MC
Q8. You have been hired as a marine resource manager tasked with deciding which region of the world you should prioritize for coral reef conservation based on coral reef stress. The forecaster you are working with is trying to develop a visualization that can help you make your conservation decision. Which of the following visualization options would be best?	LO2: UC Communication	Connect to user	MC

Note: The full list of possible answers for multiple choice questions are listed in Appendix S1:

Table S3, and thematic bins used for scoring the qualitative questions are listed in Appendix S1:

Tables S5 and S6. EF = ecological forecasting. UC = uncertainty. LO1: EF Concepts = Describe what ecological forecasts are and how they are used; LO2: UC Communication = Identify different ways to represent uncertainty in a visualization. The assessment question short names are used to refer to specific questions throughout the text and in Figures 3 and 4. MC = multiple-choice, Q = qualitative.

Table 4. Summary statistics of pre- and post-module assessment questions for assessment questions in this study (Q1-8).

Question	Two-Tailed p-value	Test Statistic	Effect size	Pre-Module Correct Responses	Post-Module Correct Responses	Units for Correct Responses	<i>n</i>
Q1: Define	<0.001	725	0.67	28	80	Percent correct (%) across all students	240
Q2: Uncertainty	0.16	1628	0.09	55	60	Percent correct (%) across all students	240
Q3: Forecast benefits	<0.001	1606	0.12	0.82 ± 0.79	1.13 ± 0.69	Number correct (Mean $\pm$ 1 SD) per student	208
Q4: Communication method	<0.001	238	1.09	0.4 ± 1.02	2.52 ± 1.79	Number correct (Mean $\pm$ 1 SD) per student	240
Q5: Interpret	0.34	1716	0.06	50	53	Percent correct (%) across all students	240
Q6: Contrast	<0.001	1751	0.22	33	47	Percent correct (%) across all students	240
Q7: Distinguish	<0.001	1633.5	0.39	33	61	Percent correct (%) across all students	240
Q8: Connect to user	0.28	2392	0.07	46	51	Percent correct	221

						(%) across all students	

Note: Test statistics, p-values, and effect sizes are for paired, two-sided Wilcoxon signed-rank tests. Significant p-values ($p < 0.05$) are shown in bold.

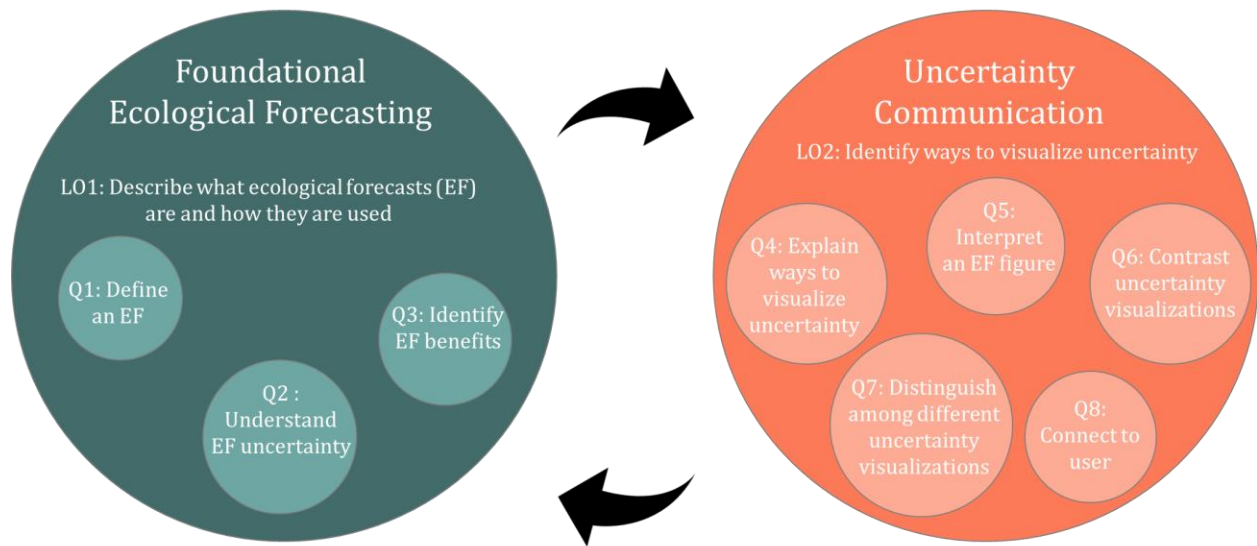


Figure 1. Conceptual diagram of the two learning objectives (LOs) associated with uncertainty communication and foundational ecological forecasting concepts taught in Macrosystems EDDIE Module 8: “Using Ecological Forecasts to Guide Decision-Making.” Within the larger circles, each bubble shows the corresponding assessment questions which tested the effectiveness of the module in meeting these LOs. Descriptions of the assessment questions are in Table 3 and the full assessment questions can be found in Appendix S1: Tables S1-2. Ecological forecast is abbreviated as EF.

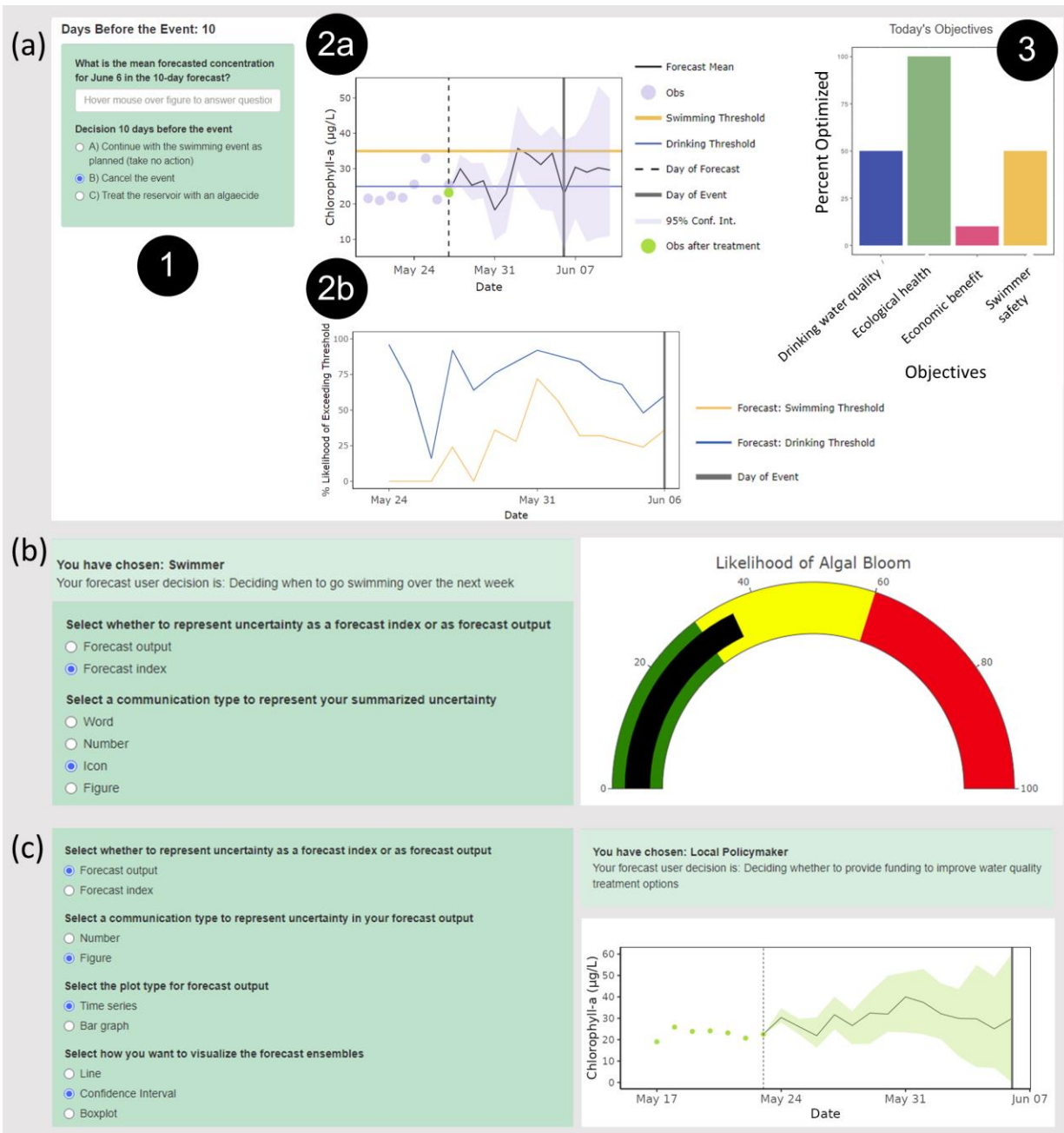


Figure 2. Screenshots of activities from the module R Shiny application showing A) Activity B, in which students (1) make decisions about how to manage a drinking water reservoir, using two different forecast visualizations (2a and 2b, which were shown separately to students within the application), while weighing the consequences on multiple objectives, such as maintaining good drinking water quality, preserving ecological health, maximizing economic benefit, and ensuring

swimmer safety (3). In Activity C, students chose a forecast user and customized a visualization for that particular users' decision needs; for example, (B) shows a visualization and decision chosen for a swimmer, while (C) shows a visualization and decision chosen for a local policymaker.

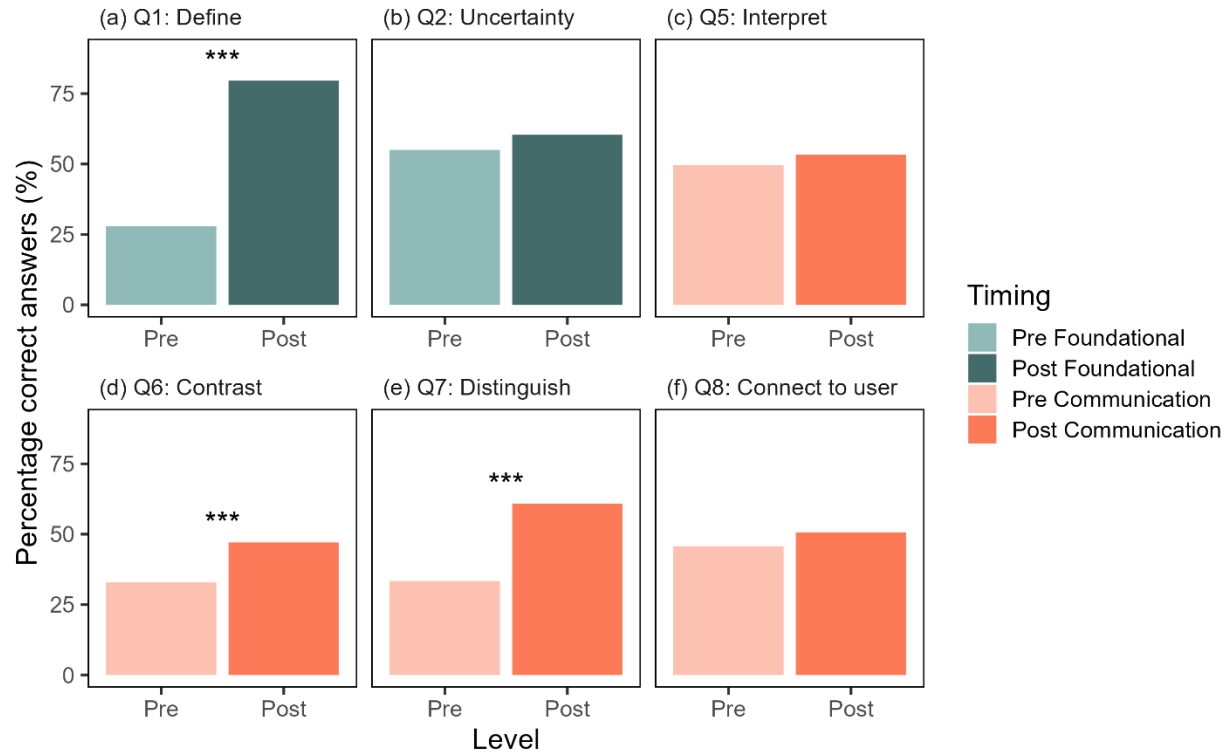


Figure 3. Percentage of students who answered multiple-choice questions correctly in the pre- and post-module surveys. Asterisks (***) indicate a statistically significant difference between the pre- and post-module survey according to a Wilcoxon signed-rank test ($p < 0.001$). Colors of the bars correspond to LO1: Foundational ecological forecasting (EF) concepts (green) and LO2: Uncertainty Communication (orange; Figure 1, Table 3). A description of questions can be found in Table 3 and Appendix S1: Table S3.

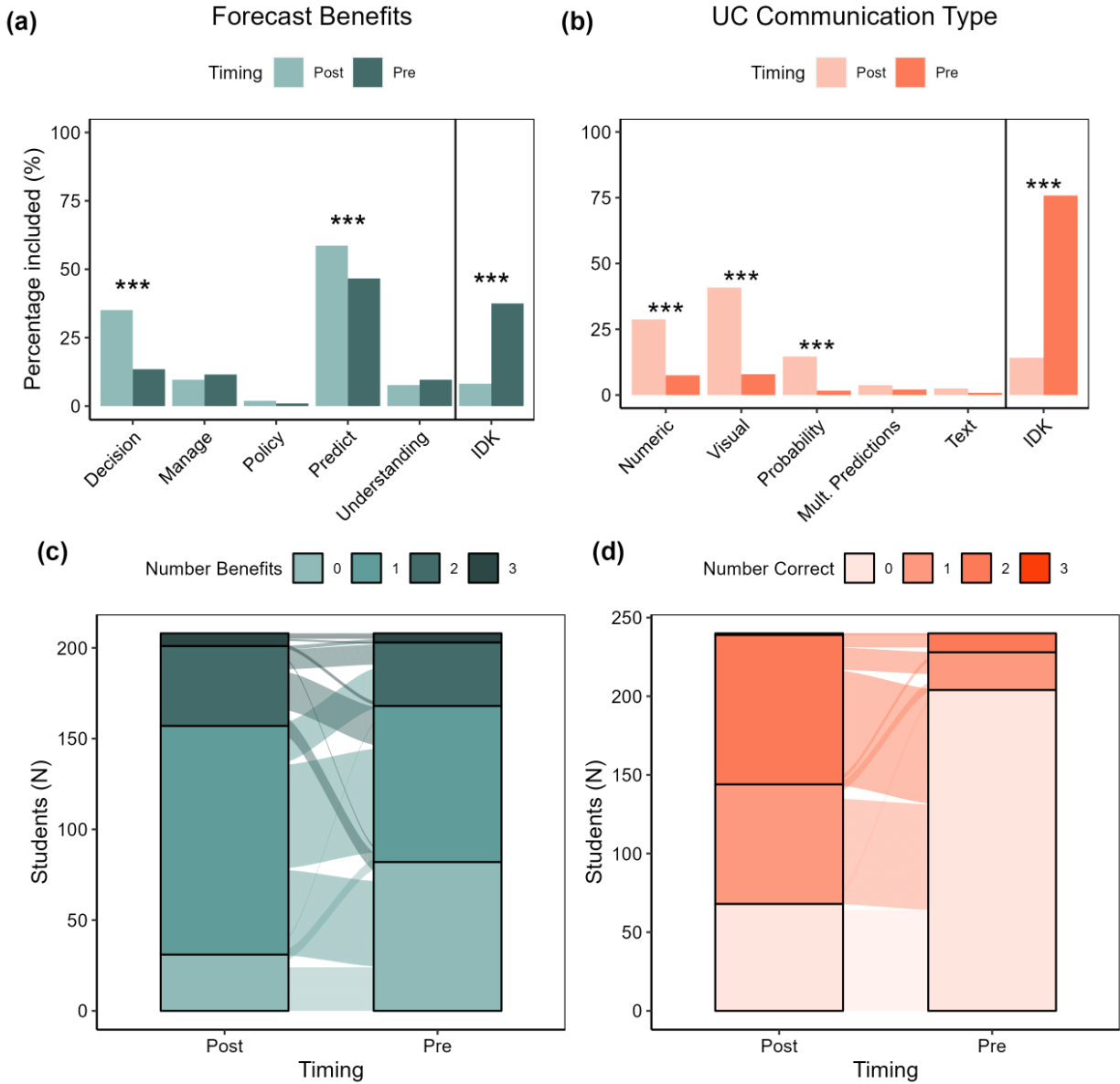


Figure 4. Percentage of students who identified a) different benefits of ecological forecasting (Q3), and b) different ways to visualize uncertainty (Q4) in the pre- and post-module responses. These responses correspond to the c) total number of benefits identified by individual students in Q3, and d) the total number of ways to visualize uncertainty identified by individual students in Q4. Students were also given the option to state “I don’t know,” represented as “IDK” in panels a and b. The categories listed on the x-axis in panels a and b were determined through the

methods outlined in Appendix S1: Text S1 and are listed in Table 4, respectively. Asterisks (***) indicate a significant difference ($p < 0.001$) between the pre- and post-survey responses according to paired Wilcoxon signed-rank tests. A full description of questions can be found in Appendix S1: Table S4.

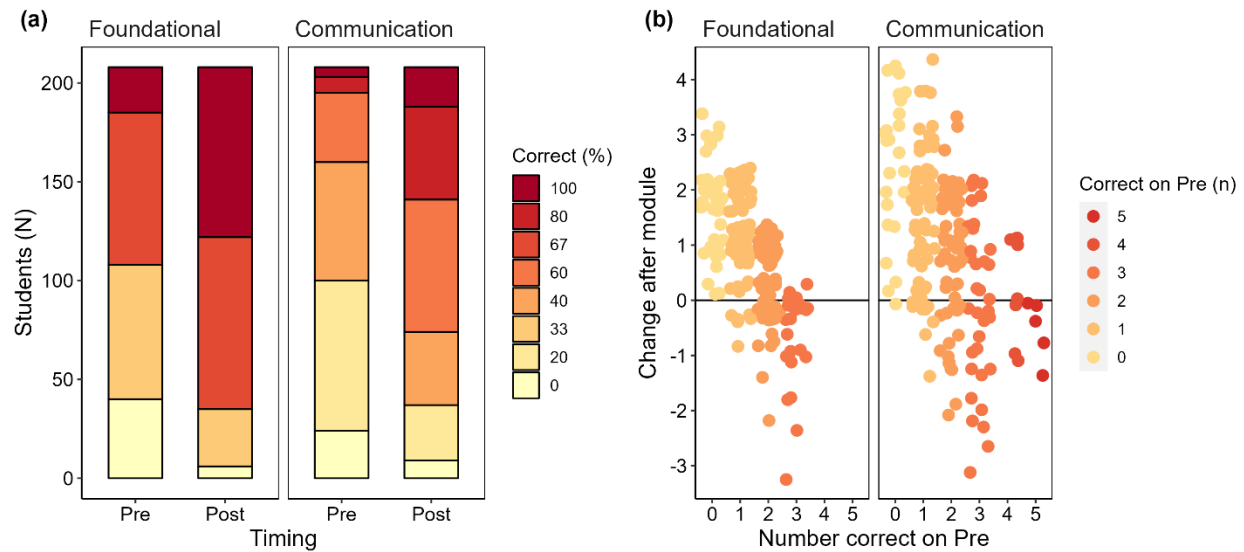


Figure 5. Pre- and post-survey results across the two LOs showing a) the total percent correct for all students across each category, and b) the change in the number of correct answers for each student after module completion relative to each student's percent correct before taking the module. Color in (b) corresponds to the number correct on the pre-survey only and points are jittered to improve legibility of individual points. Note that the number of questions corresponding to each LO varied, with three questions assessing foundational ecological forecasting concepts and five questions assessing uncertainty communication concepts. The percentage of correct answers are standardized to the total number of questions per LO in (a), while the number of questions answered correctly is shown in (b). "Foundational" corresponds to LO1: Ecological Forecasting Concepts, and "Communication" corresponds to LO2: Uncertainty Communication.

Supporting Information

Chapter V: Embedding communication concepts in forecasting training increases students' understanding of ecological uncertainty*

Whitney M. Woelmer¹, Tadhg N. Moore^{1,2}, Mary E. Lofton¹, R. Quinn Thomas^{1,2}, and Cayelan C. Carey¹

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

²Department of Forest Resources and Environmental Conservation, Virginia Tech, Blacksburg, VA, USA

Contents of this file:

Pages: 19

Texts: 3

Tables: 6

Figures: 4

References: 1

*Published in *Ecosphere* with the citation:

Woelmer, W. M., Moore, T.N., Lofton, M. E., Thomas, R. Q., & Carey, C. C. (2023).

Embedding communication concepts in forecasting training increases students' understanding of ecological uncertainty. *Ecosphere*, 14(8), e4628.

Text S1. *Full description of the scenario given to students in Activity B.*

You and your partner are drinking water managers for the primary drinking water source reservoir for your city, which has ~300,000 residents. In 14 days, a major international swimming competition, with an estimated ~2,000 participants, will be held in the reservoir. As a result of hosting this event, the water utility will receive a direct bonus of >\$3 million dollars from the city. However, the water utility is concerned about the reservoir's water quality at this time of year, as historically there have been major algal blooms, which typically require expensive water treatment techniques. Without treatment, algal blooms can pose potential health issues to aquatic life, swimmers, and the city's drinking water, which means you may have to cancel the swimming event if an algal bloom occurs. During the swimming competition, the water utility must continue to use the reservoir as a drinking water source, and so must balance the needs of maintaining safe water quality for drinking and swimming, ecological health of the reservoir, as well as the economic costs and benefits of treatment or holding the event.

Text S2. Description of the development and piloting of assessment questions.

Assessment questions were designed to measure student understanding of the two main learning objectives assessed in this study (LO1 and LO2). We used both multiple choice and qualitative questions to provide diverse methods of measuring student learning (Johnson and Christensen, 2019). We piloted assessment questions in a two-pronged approach. First, we led a targeted focus group with three undergraduate students who were taught the introductory PowerPoint lecture and independently completed the module, including answering the embedded module questions and completing pre- and post-assessment questions. We then met with the students as a small group and discussed components of the module and assessment questions that were confusing or misleading, and modified the module and assessment questions correspondingly.

Second, after revisions from the focus group were implemented, we piloted the module in three courses. These courses included: 1) a mixed graduate and undergraduate class on Ecological Forecasting with 16 students; 2) an undergraduate class on Research and Statistical Methods with 10 students; and 3) a conference workshop with 21 graduate students. During each of the pilots, the module was taught by one of the module co-authors following the same guidelines provided to the external instructors who participated in this study. Specifically, following the instructor introductory PowerPoint lecture, students completed the module during class in small groups. Pre-assessment questionnaires were given to students within one week before the class, and post-assessment questionnaires were completed within one week following the class. Students were given extra credit for completing the pre- and post-assessment questionnaires to incentivize participation. We analyzed the assessment responses, identified any questions that had either very high or very low rates of correctness to prevent ceiling or floor effects,

respectively, and revised assessment questions correspondingly. We also further identified and refined components of the module that were confusing to students and instructors after the three pilots. Finally, after the pre-testing and revisions were completed, we disseminated the module to the external instructors who participated in our study.

Text S3. Description of method used to analyze qualitative assessment responses.

To score qualitative assessment responses, we followed a multi-step procedure which allowed us to develop a codebook, or standardized list of accepted answers, for categorizing student responses (following Miles et al. 2020). During the whole procedure, all identifying student, course, and survey timing (e.g., pre- or post-module) information was hidden and all responses were randomly ordered based on a random number generator.

First, two Macrosystems EDDIE coordinators independently reviewed student responses to each qualitative question and recorded emerging themes in the student responses. The coordinators then met to combine their emerging themes into a master codebook of accepted responses to the assessment questions. Second, the same two coordinators used the master codebook to independently grade an initial subset of responses (10% of responses) and then compared responses. They then met to resolve any differences and further refined the codebook to better characterize the relevant categories (bins) for student responses. Following this agreement on both the codebook categories and scoring process, the same two coordinators scored the remaining assessment responses.

Table S1. All module learning objectives (LOs) and associated assessment questions and module activities.

Learning Objective	Assessment Question(s)	Module Activity
LO1: Describe what ecological forecasts (EFs) are and how they are used	Q1, Q2, Q3	• Activity A: Explore existing EFs • Activity B: Use an ecological forecast to make decisions • Activity C: Explore the underlying numbers in an ecological forecast and identify forecast users
LO2: Identify ways to visualize uncertainty	Q4, Q5, Q6, Q7, Q8	• Activity A: Examine existing EFs visualizations • Activity B: Make decisions using different forecast visualizations • Activity C: Make different visualizations of the same forecast
LO3: Identify the components of a structured decision	Q9	• Activity B: Complete structured decision-making using the ProACT framework
LO4: Discuss how forecast uncertainty relates to decision-making	Q8	• Activity A: Identify forecast users of existing forecasts • Activity B: Role-play as a decision maker with different visualizations • Activity C: Customize a visualize for specific decision needs for different forecast users

LO5: Match forecast user needs with different levels of forecasting decision support	Q8	● Activity A: Identify decision use cases of forecast users of existing forecasts ● Activity C: Customize visualizations for a student-chosen forecast user and associated decision
LO6: Create visualizations tailored to specific forecast users	Q8	● Activity C: Customize visualizations for a student-chosen forecast user and associated decision

Table S2. List of ecological forecasts used in Activity A of Module 8 to introduce students to existing examples of ecological forecasts and their visualizations.

Forecast	Webpage	Brief Summary
Naturecast Phenology Forecasts	https://phenology.naturecast.org/	This project produces spatial forecasts of phenological events such as bud burst (when flowers first bloom each year) or leaf out (when leaves first appear) for a variety of plant species around the US, produced by ecologists at the University of Florida. The forecast presented here shows predictions for the date of leaf out across the U.S. for the maidenhair tree (<i>Ginkgo biloba</i>).
Phenology Monitoring at the Morton Arboretum	https://www.rolliecolony.com/phenology	This project forecasts the probability of bud burst (when flowers first bloom each year) of various types of tree species at the Morton Arboretum in Illinois, USA using forecasted weather data. The forecast presented here is for bud burst of <i>Quercus marcoparpa</i> , or bur oak, at the Morton Arboretum.
Portal Forecast	https://portal.naturecast.org/	This project forecasts desert rodent abundances using a long-term historical dataset. The forecast presented here shows total rodent abundance, which includes species like the northern pygmy mouse (<i>Baiomys taylor</i>), tawny-bellied cotton rat (<i>Sigmodon fulviventer</i>), and the silky pocket mouse (<i>Perognathus flavus</i>), among many others.
EcoCast	https://coastwatch.pfeg.noaa.gov/ecocast/	EcoCast is a fisheries sustainability tool developed by NOAA (National Oceanic and Atmospheric Administration) that helps fishers and managers evaluate how to allocate fishing effort to optimize the sustainable harvest of target fish while minimizing bycatch of protected or threatened animals. The forecast presented here shows areas which are predicted to be more or less suitable for minimizing fisheries bycatch and maximizing catch of desired fish species.
Smart & Connected Water Systems	https://smartreservoir.org/forecasts/	This project develops forecasts of water temperature every meter from the surface to the bottom of Falling Creek Reservoir in Virginia. The forecasts are made using high-frequency sensors, weather forecasts, and ecosystem models. The project aims to improve the management of drinking water supply lakes and reservoirs.
GrassCast	https://grasscast.unl.edu/	GrassCast is an experimental grassland productivity forecast for the Great Plains region which uses a nearly 40-year historical dataset combined with seasonal precipitation forecasts to predict how rangeland productivity will differ from past trends. The forecast presented here shows the predicted percent change in Annual Net Primary Productivity (ANPP) compared to the historical ANPP across the Great Plains region of the U.S.

Coral Reef Watch	https://coralreefwatch.noaa.gov/satellite/bleachingoutlook_cfs/outlook_cfs.php	The National Oceanic and Atmospheric Administration's (NOAA) Coral Reef Watch uses remote sensing and models to produce forecasts of coral reef environments worldwide. The forecast presented here shows the global outlook for Coral Reef Bleaching Heat Stress based on NOAA's Climate Forecast System Version 2 (CFSv2).
USA-NPN Pheno Forecast	https://www.usanpn.org/data/forecasts/	The USA-National Phenology Network's Pheno Forecast delivers short-term (6 day) threshold-based forecasts of phenological events in plants and pest insects. The forecast presented here shows predictions of emergence of the Emerald Ash Borer, an invasive species, based on Growing Degree Days (GDD), which is a cumulative estimate of temperature over the course of year which influences plant and animal growth.
Atlantic Sturgeon Risk of Encounter	http://robots.coe.udel.edu/shiny/sturgeon/	This forecast is developed for mature Atlantic sturgeon using historic telemetry observations matched to date, bathymetry, and sea surface temperature and ocean color from NASA's MODIS AQUA satellite.
Ecocaster	https://eco.bigelow.org/	This project produces ecosystem forecasts informed by citizen reports. They are launching daily forecasts for ticks, jellyfish, and more using data provided by citizens across the country.

Table S3. Multiple-choice assessment questions for pre- and post-module assessment questions, with correct answers are indicated in bold.

Q1. Which of the following statements best describes an ecological forecast? a. An estimate of future environmental conditions which can be influenced by human actions b. A prediction of changes in ecosystems in response to extreme weather events c. A prediction of future environmental conditions with uncertainty d. A projection of how ecological communities will change in response to long-term climate change e. An informed guess of how ecosystems will change over time using historical data
Q2. When an ecological forecast is generated, how does the uncertainty of the forecast change as it predicts conditions further into the future? For example, if a forecast was generated today for ecological conditions over the next two weeks, how would the uncertainty in the first week of the forecast compare to the uncertainty in the second week of the forecast? a. Stays the same over time b. Increases over time c. Decreases over time d. Changes randomly over time e. It is unknown how it will change over time
Q5. Which of the following is the <i>best</i> description of the forecast presented in Figure 1A? (refer to Figure S1 in this supplement) a. On June 16, surface temperature will be 24.2°C and bottom temperature will be 14.5°C b. Surface temperature will likely increase while bottom temperature will likely stay the same c. Surface and bottom temperatures will increase and decrease in the future d. There is no chance that surface temperature will decrease, and bottom temperature will not increase above 15°C e. It is highly likely that surface temperature will be above 25°C by June 20
Q6. The ecological forecasts in Figure 1A and 1B present information differently. Which of the following is true? (refer to Figure S1 in this supplement) a. Figure 1A only represents uncertainty over time, while Figure 1B represents uncertainty over space and time b. Both Figure 1A and Figure 1B represent uncertainty with color c. Figure 1A represents uncertainty as a metric of forecast output, while Figure 1B represents uncertainty in raw forecast output d. Figure 1A shows how uncertainty changes as you go further into the future but Figure 1B does not e. Figure 1A shows how the forecast relates to past conditions while Figure 1B does not

Q7. Which of the following is an example of a forecast index (as opposed to a forecast output)? a. A forecast of invasive species abundance four months in the future b. A forecast of the likelihood that you will encounter an endangered species in a forest c. A forecast of grassland above-ground biomass for the next summer d. A forecast of the oxygen concentration in a lake over the next 16 days with uncertainty e. A forecast of leaf density in a forest one week in the future
Q8. You have been hired as a marine resource manager tasked with deciding which region of the world you should prioritize for coral reef conservation based on coral reef heat stress. The forecaster you are working with is trying to develop a visualization that can help you make your conservation decision. Which of the following visualization options would be best? a. A bar graph showing the likelihood of forecasted coral reef heat stress across the globe b. A map of forecasted percent change in ocean temperatures from historical average c. A map showing forecasted categories of the level of severity of heat stress across the globe d. A times series (line graph) showing the global average of coral reef heat stress leading up to the present e. A map showing the amount of uncertainty around heat stress level forecasts across the globe
Q9. Which of the following is NOT a component of structured decision-making? a. Frame the decision problem b. Specify the decision objectives c. Rank the values associated with your objectives d. Examine alternative options e. Compare consequences of your alternative options

Table S4. Qualitative questions from pre- and post-module assessments which were graded using the method outlined in Text S1.

Q3: List what you think are some benefits of ecological forecasting. If you don't know any benefits, please write "I don't know" instead of leaving the response blank.
Q4: Describe two different ways uncertainty can be visualized in a forecast. If you don't know the answer, please write "I don't know" instead of leaving the response blank. If you can only think of one answer, list just one.

Table S5. Coding criteria used for Q3 (List some benefits of ecological forecasting) thematic bins. Bins were developed using the method described in Text S1.

Bin name	Description	Grading Criteria
IDK	I don't know	Presence or absence
Decision	Decisions, decision-making, take action, intervention, adapt, respond, implement, alter	Presence or absence
Manage	Management, conservation, preservation, sustainability	Presence or absence
Policy	Policy, policy-making, regulation, laws	Presence or absence
Predict	Predict, anticipate, foresee, identify risk, assess, possible, inform, provide insight	Presence or absence
Understanding	Improve understanding/models, aid in research/studying, test hypotheses/models, understand predictability, provide information/data	Presence or absence

Table S6. Coding criteria used for Q4 (describe two different ways uncertainty can be visualized) thematic bins. Bins were developed using the method described in Text S1.

Bin name	Description	Grading criteria
IDK	I don't know	Correct or incorrect
Visual	A visual representation of uncertainty which implies use of a figure (e.g., boxplot, confidence intervals, error bars, shaded area, etc.)	Correct or incorrect
Numeric	A numeric representation of uncertainty which implies use of numbers or statistics (e.g., maximum or minimum, range, standard deviation, intervals, etc.)	Correct or incorrect
Probability	A representation of uncertainty that includes probability, percent likelihood, risk assessment	Correct or incorrect
Multiple predictions	A way to represent	Correct or incorrect

	uncertainty using multiple predictions (e.g., multiple trend lines, multiple different data plots, multiple numbers for each day, etc.)	
Text statement	A representation of uncertainty with text (e.g., a disclaimer, text on the side, figure title explains uncertainty)	Correct or incorrect
Forecast index/output	Use of a forecast index or forecast output to represent uncertainty	Presence or absence
Stated/explain for index/output	For each index/output response, we recorded whether the response explained the concept or simply stated it	Correct or incorrect
Icon	Mention of icons, symbol for representing uncertainty	Presence or absence
Bar graph	Mention of bar graph for	Presence or absence

	representing uncertainty	
Color	Mention of color for representing uncertainty	Presence or absence
Number correct	The total number of correct answers summed across visual, numeric, probability, multiple predictions, text statement	Sum of correct answers

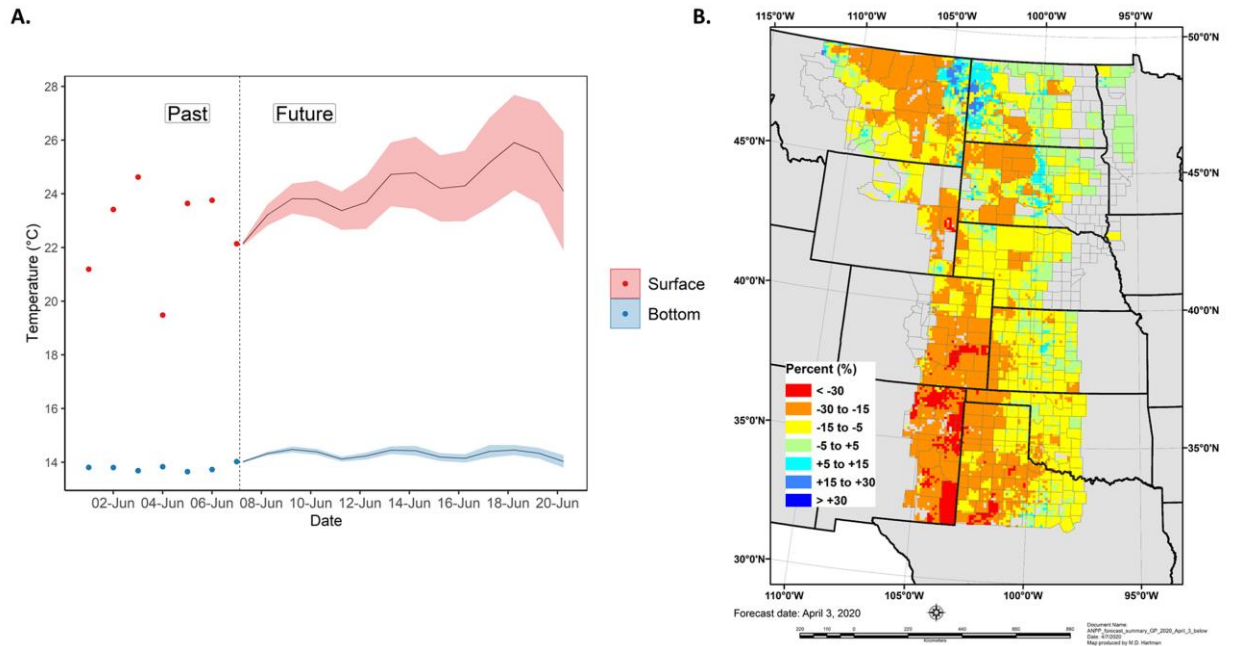


Figure S1. The figure accompanying Questions 5 and 6 on the pre- and post-module assessment survey. Forecast visualizations of: A) water temperature forecasts in Blue Lake; and B) grassland biomass in the Great Plains region of the United States. Panel A shows a forecast generated for surface (red) and bottom (blue) temperatures in Blue Lake for 14 days into the future. Points represent measurements from the last 7 days, the black line represents the mean forecast estimate, and the shaded polygon represents the probabilistic forecasts of future temperatures. Panel B shows a forecast for the upcoming year's growing season. It is a forecast of the percent change in future grassland biomass in the U.S. Great Plains region compared to the historical (1982–2019) average.

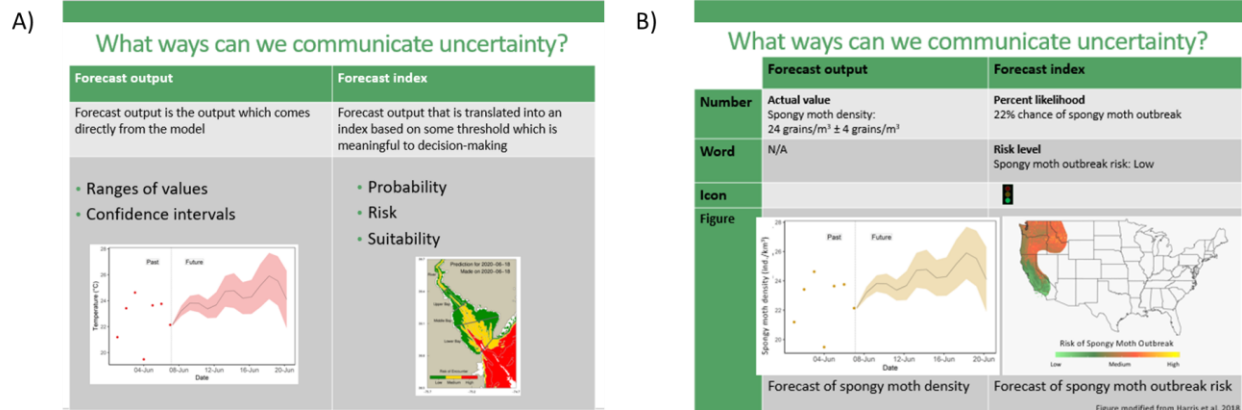


Figure S2. Screenshot from introductory PowerPoint lecture at beginning of module explaining A) the concept of a forecast output and a forecast index, and B) different ways to communicate uncertainty using numbers, words, icons, and figures as either forecast output or forecast index.

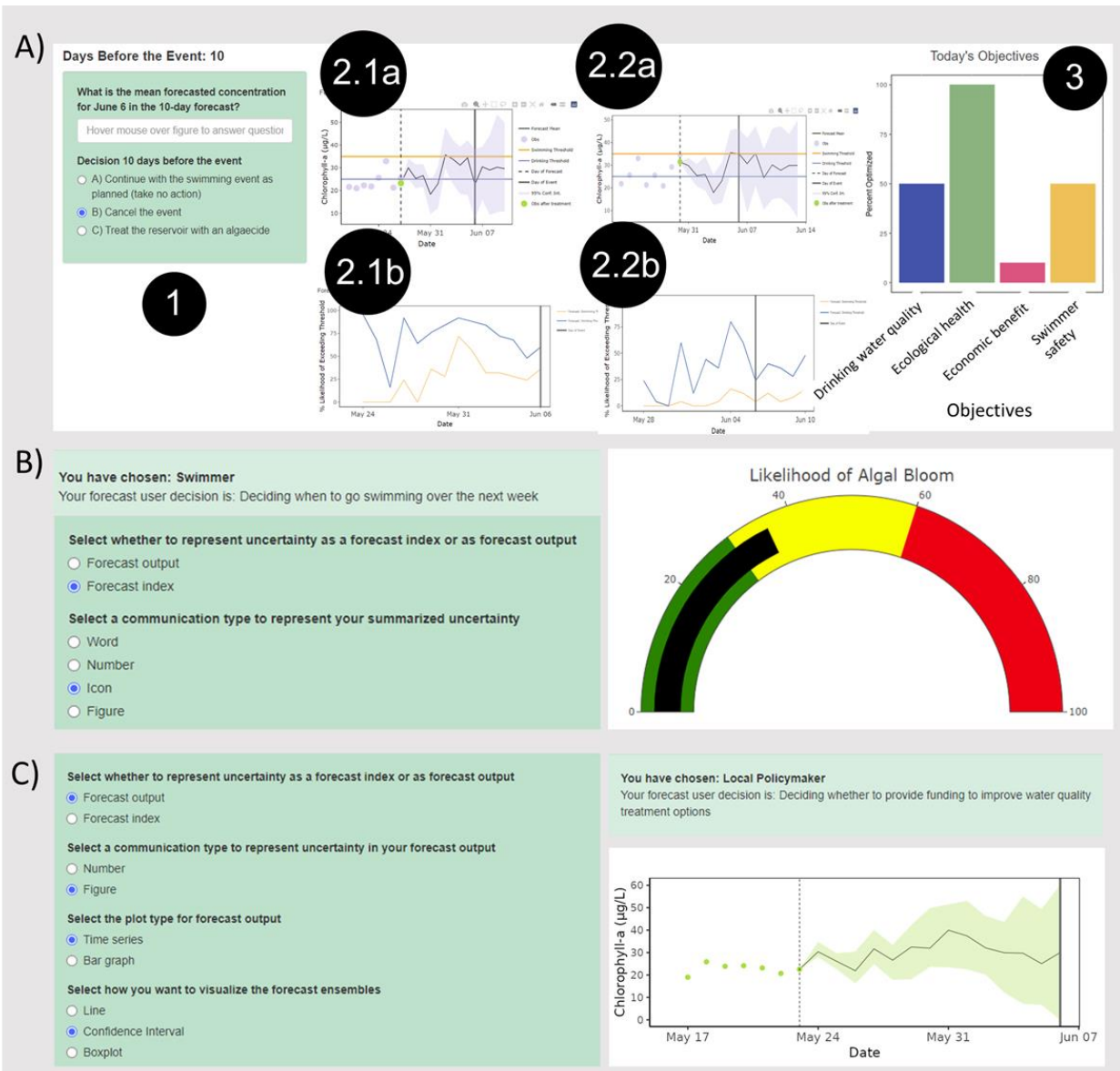


Figure S3. Screenshots of activities from the module R Shiny application showing A) Activity B, where students (1) make decisions about how to manage a drinking water reservoir, using two different forecast visualizations (2a and 2b), while weighing the consequences on multiple objectives (3). The transition from 2.1 to 2.2 shows how forecasts update over time and following students' management decisions. In Activity C, students choose a forecast user and customize a visualization for that particular users' decision needs showing B) a visualization and

decision chosen for a swimmer, as opposed to C) a visualization and decision chosen for a local policymaker.

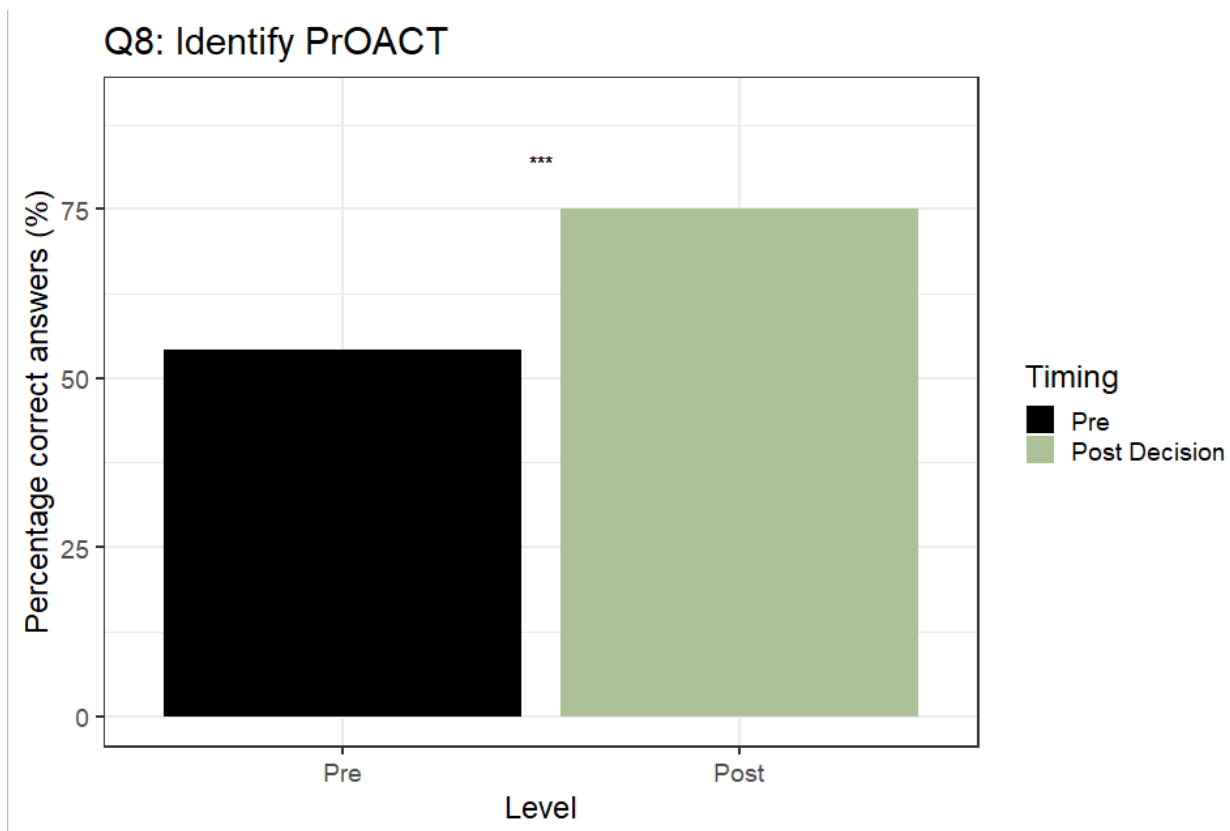


Figure S4. Results from pre- and post-assessment Question 9 which asked students ($n = 240$) to identify the components of PrOACT, a structured decision-making technique ($p < 0.001$). Based on Wilcoxon signed-rank test, this question had a test statistic of 845.5, and effect size of 0.34, with 54% of student responses correct on the pre-module and 75% of student responses correct on the post-module.

References

- Johnson, R.B. and Christensen, L., 2019. Educational research: Quantitative, qualitative, and mixed approaches. Sage publications.
- Miles, M. B., A. M. Huberman, J. Saldana. 2020. *Qualitative data analysis: A methods sourcebook*, 4th ed.; SAGE Publications Inc.: Thousand Oaks, CA, USA.

Chapter VI: Synthesis

Overview and overarching research question

In an era of accelerating climate and land use change, developing forecasts to serve as decision support tools is increasingly valuable, yet requires both improved quantification of ecosystem variability, as well as proper communication of forecast tools to guide decision-making. My dissertation combined ecosystem monitoring, ecological forecasting, and development of educational materials, with the ultimate goal of addressing the question, “**How can we build, apply, and communicate ecosystem understanding in freshwater ecological forecasts?**” Below, I identify several cross-chapter insights and provide recommendations and ideas for potential future work embedded within each section. The three major themes which emerged from my dissertation (Integrating ecosystem understanding with ecological forecasting; Ecological forecasts as a testbed for ecological understanding; and Ecological forecasting as a catalyst for end user engagement and iterative forecast production) synthesize knowledge gained across chapters, with a final conclusion section below.

I answer my overarching Ph.D. dissertation research question in complementary ways across the four chapters, which follow the conceptual framework (Figure 1) outlined in Chapter I for developing accurate, useful forecasts. In Chapter 2, I show that intensive cross-ecosystem monitoring enables improved quantification of spatial and temporal heterogeneity of biogeochemistry in reservoirs (Figure 1a). Moreover, better quantification of variability in water quality informs the design of long-term ecosystem monitoring programs, constrains uncertainty in observations and ecological processes, and enables forecast model development (Figure 1b). In Chapters 3 and 4, I developed ecological forecasts of water quality variables over time and space, leveraging data from long-term monitoring programs at three freshwater ecosystems

(Figure 1c). In Chapter 3, I used a Bayesian statistical model to quantify uncertainty in forecasts of reservoir phytoplankton concentrations, an important metric of water quality for informing decision-making. My analysis showed that while bloom prediction decreased forecast accuracy, baseline phytoplankton concentrations were relatively accurate up to a week ahead, and process model representation contributed most to total forecast uncertainty. In Chapter 4, I employed a process-based forecasting framework to predict the interrelated dynamics of water temperature and oxygen throughout a lake water column over two years. My results showed that inter-annual variability in forecast skill was high, and that forecast skill was dependent on the level of complexity in the observations being forecasted. Altogether, production of these forecasts provide a starting point for developing decision support tools (Figure 1d). In Chapter 5, to improve the utility of decision support tool development, I developed an educational module to teach concepts of visualization and decision science to improve the communication of forecasts for decision-making (Figure 1e). Based on student assessment surveys, I found that undergraduate students who completed the module exhibited significantly greater understanding of the importance of forecasting and an improved ability to communicate uncertainty. Ultimately, thoughtful communication of ecological forecasts has the potential to engage a broader audience in the value of preserving freshwater ecosystems (Figure 1f). Each chapter of my dissertation provides complementary information on both ecosystem variability and the value of ecosystem understanding for decision-making, improving our ability to adapt to changing freshwater ecosystems.

Integrating ecosystem understanding with ecological forecasting

Having a foundational understanding of variability in water quality over time and space is critical for informing forecast development. In my dissertation, Chapters 2 and 3 both analyzed long-term monitoring datasets from Falling Creek Reservoir (FCR) and Beaverdam Reservoir (BVR) to provide complementary insight into the predictability of biogeochemistry in these two reservoirs. For example, identifying the importance of relatively high temporal heterogeneity and low spatial heterogeneity in chlorophyll-a concentrations in the two reservoirs in Chapter 2 supported our research team's decision to forecast chlorophyll-a near the dam in Chapter 3, without needing additional monitoring of chlorophyll-a upstream. However, the identification of process model uncertainty as the largest source of total forecast uncertainty in Chapter 3 suggests that a better incorporation of the high temporal heterogeneity of chlorophyll-a and other important driving variables (e.g., nutrients) in forecast models may help improve predictions of phytoplankton. For example, monitoring chlorophyll-a at a finer temporal scale (e.g., minute to sub-minute) may be necessary for better identifying shifts in variability leading to bloom development (Dakos et al. 2012; Shi et al. 2018; Luo et al. 2022), in addition to finer scale monitoring of model covariates which are known to drive phytoplankton dynamics (e.g., nitrogen and phosphorus concentrations; Carpenter 2008; Paerl et al. 2014).

Results from Chapter 2 and 3 both indicate that specific locations upstream within my focal study reservoirs may also require additional monitoring. Specifically, Chapter 2 identified the stream-reservoir interface as a critical transition point for many biogeochemical variables, ultimately driving whole-reservoir functioning. To improve understanding of changes in phytoplankton dynamics along the full reservoir continuum (and ultimately at the dam where water withdrawal pipes are located), it may be informative to more regularly monitor changes in

biogeochemistry at the stream-reservoir interface. The importance of inflow dynamics for downstream chlorophyll-a is further supported by the inclusion of discharge in the best-fitting model used for forecasting phytoplankton dynamics at the dam in Chapter 3.

Understanding spatial heterogeneity in reservoirs also has implications for naturally-formed lake ecosystems. While Lake Sunapee, the focus of Chapter 4, is an oligotrophic lake (Richardson et al. 2017), phytoplankton blooms periodically develop in nearshore areas (Carey et al. 2014). These littoral sites have also been shown to have higher metabolic rates than pelagic sites, potentially serving as an early indicator of shifting trophic states in Lake Sunapee (Ward et al. 2022). Altogether, these studies, together with Chapter 2's finding on the importance of stream-reservoir interface for determining whole-ecosystem functioning, indicate that transitional regions between freshwater ecosystems (i.e., stream-reservoir *and* stream-lake interfaces) may be of critical importance for shifting ecosystem functioning (Mackay et al. 2011; Hotchkiss et al. 2018).

I suggest several avenues for future work to deepen our understanding of spatial and temporal heterogeneity. First, expanding field studies similar to those in Chapter 2 to quantify the spatial and temporal heterogeneity across many lakes and reservoirs will enable cross-ecosystem comparisons. Specifically, analyzing longitudinal trends in biogeochemistry across multiple reservoir use-types (e.g., reservoirs built for hydropower flood control, water supply, etc.), ecoregion (e.g., temperate, tropical, boreal, etc.), size, and age, can help better quantify patterns in biogeochemistry observed along the reservoir continuum, enabling a more robust assessment of when and why longitudinal trends follow or deviate from Thornton et al.'s (1990) proposed reservoir continuum concept. Second, quantifying the importance of the stream-reservoir interface by a) identifying more precisely the stream-water plunge point location in

different reservoirs and determining how the location of this site changes seasonally, b) quantifying the contribution of this location to whole-ecosystem biogeochemical budgets, and c) examining differences in the importance of the stream interface on whole-ecosystem functioning in lakes vs. reservoirs would all greatly expand upon my dissertation research. Specifically, I recommend that future work on predicting changing trophic dynamics or phytoplankton blooms in Lake Sunapee focus on better understanding the dynamics of the stream-lake transition zone to better constrain the role of this “control point” on overall ecosystem functioning (following Ward et al. 2022; Bernhardt et al. 2017). Third, studies with higher-scale temporal resolution (e.g., sub-minute to hourly) to quantify changes in temporal heterogeneity in phytoplankton in the two reservoir sites studied in Chapters 2 and 3 may help improve bloom predictions. Lastly, expanding on both the temporal forecasting work in Chapter 3 and the knowledge gained on spatial variability in Chapter 2, I would suggest developing spatial forecasts of biogeochemical variables at multiple sites in the two reservoirs.

Ecological forecasts as a testbed for ecological understanding

Ecological forecasts of water quality provide information about predictability across a gradient of physical, chemical, and biological water quality variables. Lake modeling studies have shown that predictability should decrease along this gradient, with biological variables being the most challenging to predict relative to physical or chemical variables (*sensu* Soares and Calijuri 2021). In Chapters 3 and 4, I compared physical, chemical, and biological water quality forecasts to different null models, quantifying how well our forecasts predicted relative to an uninformed estimate. I found that biological (chlorophyll-a) forecasts were most skillful, beating a persistence null up to 14 days ahead, followed by physical (water temperature) forecasts,

beating a climatology null 11 days ahead, and lastly, chemical (dissolved oxygen) forecasts, beating a climatology null only 2 days ahead.

Variability from the expected decrease in predictability from physical, chemical, biological likely depends on a number of factors. First, the higher skill of chlorophyll-a forecasts I observed relative to water temperature and oxygen may be due to the difference in the null model type I used in the two analyses (persistence models vs. climatology models), which could be informed by comparing forecasts to multiple null models for a given variable (following Pappenberger et al. 2015). Second, differences in the model time step across the studies is an important driver of the model's skill. Specifically, the temperature and oxygen forecasts in Chapter 4 were generated on a daily timestep, and while the weekly and fortnightly chlorophyll-a forecasts in Chapter 3 performed well one and two-weeks ahead, daily forecasts of chlorophyll-a were never more skillful than the null model. This may indicate that forecasts of temperature and oxygen developed using models on weekly and fortnightly timesteps may also be more skillful at if models were developed with longer timesteps. Third, I used differing datasets for model training based on the availability of local data: specifically, the chlorophyll-a model was trained on ~100 weekly observations vs. ~400 daily observations for the temperature/oxygen model, which may have influenced their relative forecast skill. Finally, there may be inherent differences in the predictability of the reservoir study systems where the chlorophyll-a was forecasted vs. the glacially-formed lake where temperature and oxygen was forecasted.

Forecast development can also provide cross-ecosystem understanding. First, within Chapter 3, I showed that the same chlorophyll-a model developed at FCR could be applied to a similar nearby reservoir, BVR, indicating that waterbodies with similar meteorological and geomorphic conditions may exhibit some similarities in predictability. When scaling our

forecasting system from small Virginia reservoirs to Lake Sunapee, a much larger naturally-formed lake, our research team hypothesized that the importance of inflow dynamics would be less due to Sunapee's longer residence time, and as a result, chose to configure our process-based model without inflows at Lake Sunapee when forecasting the deepest site of the lake. However, a multi-basin lake like Lake Sunapee, which has complex morphometry and no uni-directional flow, may actually have more spatial variability (similar to what Ortiz and Wilkinson 2021 found in a small lake) than small run-of-river reservoirs like FCR and BVR, which have directional flow and short residence times, so upstream to dam conditions are more directly connected. Additionally, because FCR is eutrophic, with high chlorophyll-a concentrations and intermittent blooms, while Lake Sunapee is oligotrophic, with low and relatively consistent chlorophyll-a dynamics, I expected that chlorophyll-a dynamics would be easier to predict at Lake Sunapee than in FCR, but did not develop chlorophyll-a forecasts at Lake Sunapee to test this expectation. Instead, I focused forecasts at Lake Sunapee on dissolved oxygen, as it has been shown to be an indicator of changing metabolic activity and trophic state in Lake Sunapee (Richardson et al. 2017; Ward et al. 2022), and changes in lake productivity are of concern for the Lake Sunapee community. Altogether, a robust testing of differences in predictability of variables across ecosystems through standardized development of forecasts using the same model configuration would expand our understanding of how forecasting models can be applied across ecosystems.

In all three forecasts of water quality (temperature, oxygen, chlorophyll-a), I found differences in predictability between forecasts of median or “baseline” concentrations and rare, yet important, events. For example, turnover-induced mixing in the fall at Lake Sunapee resulted in relatively lower forecast skill of both temperature and oxygen during and immediately after turnover onset. Similarly, forecasts of chlorophyll-a at FCR had relatively high performance

under baseline (i.e., non-bloom) conditions, but poor performance at predicting ephemeral bloom dynamics. Onset of hypoxia is likely another limnological phenomenon which may be harder to predict than baseline oxygen conditions, and likely varies with ecosystem characteristics (e.g., mixing regimes; (Lin et al. 2023)). However, because Lake Sunapee does not experience anoxia at the lowest depth where fixed buoy sensor observations are available (10.0 m), we were unable to test this prediction. Despite discrete events being more difficult to predict than median concentrations, they remain some of the most important types of forecasts needed by managers and the public for informing decision-making, and thus it is imperative to improve their accuracy.

Differences across model types can also result in differences in predictability. In my dissertation, I used two distinct modeling approaches to forecast several water quality variables, each with their own strengths and weaknesses. The differing levels of complexity between the models used to forecast in Chapters 3 vs. 4 provide unique opportunities to better understand fundamental ecological processes. Using a Bayesian empirical model in Chapter 3, I was able to robustly quantify and partition different sources of uncertainty and make inferences about how different sources of uncertainty impact forecast performance. Additionally, in Chapter 3, I used model selection to identify important drivers of chlorophyll-a dynamics (inflow dynamics and shortwave radiation at the reservoir site). While model selection allowed me to make inferences about the importance of the selected drivers for chlorophyll-a dynamics, the direct mechanism by which driver variables influence chlorophyll-a cannot be parsed from the linear model structure. Additionally, because the model only included two covariates, additional information regarding how other potential driving variables may be important to chlorophyll-a (e.g., hypolimnetic water temperature or nutrient availability) is missing from our forecasting framework. In

contrast, the process-based model used in Chapter 4 to predict water temperature and dissolved oxygen can enable more deeply examining underlying mechanisms behind temperature and oxygen dynamics (e.g., changes in parameter dynamics), and to make predictions throughout the water column, despite only having observations at a few depths. Overall, quantifying model complexity remains a difficult task and requires rethinking some of the axes upon which we typically assign complexity, as axes of complexity are not consistently defined throughout the literature (e.g., model structure: Ward et al. 2014; number of covariates: Chevalier and Knappe 2020; number of processes represented: Geary et al. 2020). Consistently defining axes of model complexity and developing a catalog of forecasts with a diverse array of model types may provide the best approach for beginning to disentangle the role that model complexity may play in determining ecological predictability.

Analyzing dynamics in forecast uncertainty across time, space, and variables can help explain differences in predictability. Uncertainty dynamics were highly variable over time for forecasts of chlorophyll-a at FCR and BVR in Chapter 3, and especially sensitive to the development of blooms. Following this, developing forecasts of temperature and oxygen with a full uncertainty partitioning similar to that used for chlorophyll-a in Chapter 3 would provide novel information about mechanisms underlying forecast performance and what sources need to be constrained to improve forecast performance (following Dietze 2017). First, this would allow us to examine how uncertainty dynamics vary over time, especially in response to discrete events like turnover or onset of anoxia and across different water quality variables. Additionally, following Wynne et al. (2023)'s finding that surface and bottom water temperature projections through the end of the century had distinct uncertainty dynamics, a full uncertainty partitioning

analysis of near-term forecasts of temperature could help us better understand how uncertainty dynamics change with depth at shorter time scales.

I suggest several areas to expand upon my dissertation to deepen the knowledge gained from applying ecosystem understanding through forecasting. First, continued production of water quality forecasts for a) a single variable with multiple models, b) multiple variables using a single model, and c) the same forecast model scaled across ecosystems would provide invaluable insight into fundamental predictability across time, space, and water quality variable complexity. As a starting point, to address (a), I suggest developing additional forecasts of chlorophyll-a at FCR using multiple modeling approaches (e.g., process-based models, null models, machine learning models) to complement the statistical forecasts developed in Chapter 3. To address (b), I suggest expanding upon the capabilities of GLM-AED within FLARE to produce forecasts of additional biogeochemical variables, especially chlorophyll-a and other important water quality variables for decision-making. Lastly, to address (c), I suggest developing the same model from Chapter 3 (autoregressive linear model) at Lake Sunapee for chlorophyll-a, or using the previously configured GLM-AED model at FCR to compare to Lake Sunapee. Additionally, to address the paradox of predicting rare, yet important, events (e.g., phytoplankton blooms), high-frequency monitoring to collect as much information as possible about antecedent and driving conditions for these events would be useful, as well as help identify early warning signals preceding these events. Additionally, I propose a targeted modeling approach which would focus on categorizing training data into “event” and “non-event” datasets to develop two separate models, e.g., to predict bloom or non-bloom dynamics, respectively. This type of ensemble modeling approach would leverage information developed from both datasets to have two different models designed to predict either blooms or baseline dynamics, with the identification

of a trigger (e.g., early warning indicator; *sensu* Dakos et al. 2012; Shi et al. 2018; Luo et al. 2022) indicating when the forecasts should switch from being generated with the non-bloom model to the bloom model, and vice versa. Finally, developing a catalog of forecasts across variables, years, and sites would enable a more robust analysis of the effect of observational variability on forecast skill, expanding the preliminary analysis which I began in Chapter 4.

Ecological forecasting as a catalyst for end user engagement and iterative forecast production

To realize the potential of forecasts for decision-making and environmental conservation, end users must be engaged throughout the process of forecast development. Importantly, co-production begins from early stages of the relationship between scientists and end users and includes data sharing, dedicated efforts to discuss needs for decision support tools, iterative development of end products, and thoughtful design of the end products themselves, including the design of model structure and/or final visualizations (Nichols et al. 2007; Cook et al. 2014; Parrott 2017; Henden et al. 2020). Through engagement throughout the research process, rather than just at the product development stage, the potential for relevant and usable decision support tools is highly increased (Henden et al. 2020).

Forecasts at both FCR and Lake Sunapee were developed with specific end users in mind. At FCR, nuisance phytoplankton blooms have become an issue for the treatment plant in clogging filters, resulting in treatment of the reservoir with copper sulfate, an algaecide with adverse ecological effects (Kirici 2017; Leal et al. 2018), on numerous occasions. As a result, development of phytoplankton bloom forecasts were of great interest to the Western Virginia Water Authority. Following forecast development, we found that limitations to implementing the forecast into reservoir management decision-making workflows included limitations of the

forecasts' accuracy (specifically in predicting bloom events), as well as difficulty integrating forecasts into the preexisting management computer interface which managers were familiar with for decision-making. Extending this work into the future would require new co-design of forecast products, with close collaboration with managers who would use the forecasts, to determine the optimal type of information included, the style of the visualization, as well as the delivery for truly effective integration into their well-established decision-making workflows.

Water quality forecasts at Lake Sunapee included a much broader group of end users. We began forecast development by prioritizing existing public and lake association interests and decision-making workflows, which included improving an existing live visualization on the Lake Sunapee Protective Association's website of real-time lake data. While phytoplankton blooms do sometimes occur at Lake Sunapee (Carey et al. 2014) they are typically in near-shore regions and not present at the buoy site, where the real-time data are collected. As a result, we prioritized identifying forecast variables that could be predicted at the buoy site and were also of interest to end users. For Lake Sunapee Protective Association staff and scientists interested in the lake, we wanted to produce a metric of trophic state change in Lake Sunapee, and identified dissolved oxygen as a proxy for understanding changing trophic states and ecosystem health (Jane et al. 2021). However, given that the general public is currently unused to receiving and interpreting forecasts of dissolved oxygen, we also considered ways to use water temperature forecasts to help inform recreational users of Lake Sunapee, including homeowners, boaters, divers, and swimmers. Consequently, our team decided on the development of a recreational swimming suitability index based on combined water and air temperature forecasts that emerged from Chapter 4 (following Kox et al. 2018). We prototyped an initial version of this forecast index visualization with Lake Sunapee Protective Association staff and a small group of community

members at a visualization workshop in May 2023. We iteratively incorporated their feedback and future work will include further iteration to continue improving the visualization for the general Lake Sunapee community.

In addition to co-production of forecast tools, I also suggest targeted research on the effects of visualization on user comprehension and decision-making. This would expand upon ideas developed in Chapter 5 to not only develop visualizations catered to specific end user needs but to test their effect of comprehension and decision-making. One possible avenue could be to undertake a structured experimental approach with a group of Lake Sunapee forecast users with diverse interests, and develop a set of distinct visualization types for a single forecast variable. Groups of different users at Lake Sunapee could be primed with a visualization workshop using Macrosystems EDDIE Module 8 to standardize participant knowledge of forecasting, visualization, and decision science, followed by a controlled experiment in which each user would receive a different visualization and the effect of the visualization on decision-making would be tested. Through dedicated efforts to improve co-production and relationships with end users, as well as robust experimental approaches to better understand the effects of uncertainty visualization on decision-making, we can work towards improving the utility of forecasts for decision-making.

My work in engaging both forecast users and students of ecology has the potential to have broad impacts on ecology and ecosystem science. For example, by partnering directly with forecast users to design forecast visualizations (e.g., with the Lake Sunapee Protective Association or with the Western Virginia Water Authority), I worked to increase user understanding and interpretation of forecasts. In addition, through the development of an educational module in Chapter 5, concepts of ecological forecasting, visualization, and decision

science have reached nearly 1,000 students across the United States to date. By providing faculty with independent teaching materials that can be taught outside of our study, the module will continue to reach students in the future. Overall, these training materials are critical for helping students see the connection between ecology and real-world applications, with the potential to both increase their engagement in the classroom, as well as their investment in further environmental conservation and/or scientific careers.

Conclusions

Ultimately, freshwater ecosystems are socio-ecological systems which are inherently connected to human decision-making and societal impacts (Liu et al. 2007; Cobourn et al. 2018). As decision support tools are increasingly integrated into decision-making workflows, ecosystem services will inevitably be impacted. For example, incorporating forecasts of Pacific sardine biomass within United States harvest guidelines have led to increased sardine biomass and fisheries yield (Tommasi et al. 2017), demonstrating both the value of using forecasts for decision-making and the potential feedbacks from using forecasts on ecosystem and economic health. The combination of ecosystem monitoring following intervention with decision support tools developed from ecological forecasts follows my dissertation conceptual framework (Figure 1) and emphasizes the iterative nature of this process. While I started my dissertation with the original research question of, “**How can we build, apply, and communicate ecosystem understanding in freshwater ecological forecasts?**”, my dissertation research has culminated in a new driving research question, “**How does the use of ecological forecasts as decision support tools influence water quality?**” Through the development of the proposed work above,

components of this question can be addressed by iteration of the conceptual framework representing my dissertation work (Figure 1).

My dissertation provides an example of coupling foundational understanding of lake and reservoir ecosystems with ecological forecasting to inform and educate about freshwater ecosystems. I demonstrate that temporal heterogeneity is often greater than spatial heterogeneity in two small reservoirs, and that the stream-reservoir interface can play a disproportionately important role in determining whole-reservoir functioning. Through long-term quantification of variability in water quality, coupled with diverse modeling approaches, I developed forecasts of water temperature, dissolved oxygen, and chlorophyll-a to begin developing decision support tools for decision-making frameworks. Lastly, my work shows the value of integrating visualization and decision science concepts into undergraduate curricula, demonstrating that students can be quickly introduced to these valuable concepts and exhibit increased interest and engagement in science through the lens of ecological forecasting.

Altogether, this work has important implications for improving human-freshwater interactions in an era of global change. By producing useful and actionable decision support tools through ecological forecasts (Figure 1e), we can increase the utility of ecology and the ability to make decisions for improving freshwater resource conservation (Figure 1f). Those decisions will ultimately impact the state of freshwater ecosystems over space and time, which can be measured (Figure 1a) and quantified (Figure 1b), and new knowledge can be integrated into iterative ecological forecasting (Figure 1c) to develop new decision support tools (Figure 1d), which can be communicated for diverse users (Figure 1e). Ultimately, considering the two-way direction between ecosystems and human decision-making is critical for understanding how ecosystems have changed and will continue to change into the future. My dissertation provides a

framework for better understanding ecosystem variability, testing ecosystem understanding through forecasts, and enabling decision-making through communication of forecasts, with the aim of improving the management of, and engagement with, freshwater resources for generations.

References

- Bernhardt ES, Blaszczyk JR, Ficken CD, et al (2017) Control Points in Ecosystems: Moving Beyond the Hot Spot Hot Moment Concept. *Ecosystems* 20:665–682.
<https://doi.org/10.1007/s10021-016-0103-y>
- Carey CC, Cottingham KL, Weathers KC, et al (2014) Experimental blooms of the cyanobacterium *Gloeotrichia echinulata* increase phytoplankton biomass, richness and diversity in an oligotrophic lake. *Journal of Plankton Research* 36:364–377.
<https://doi.org/10.1093/plankt/fbt105>
- Carpenter SR (2008) Phosphorus control is critical to mitigating eutrophication. *Proceedings of the National Academy of Sciences USA* 105:11039–11040.
<https://doi.org/10.1073/pnas.0806112105>
- Chevalier M, Knappe J (2020) The cost of complexity in forecasts of population abundances is reduced but not eliminated by borrowing information across space using a hierarchical approach. *Oikos* 129:249–260. <https://doi.org/10.1111/oik.06401>
- Cobourn KM, Carey CC, Boyle K, et al (2018) From concept to practice to policy: modeling coupled natural and human systems in lake catchments. *Ecosphere* 9:e02209.
<https://doi.org/10.1002/ecs2.2209>
- Cook CN, Inayatullah S, Burgman MA, et al (2014) Strategic foresight: how planning for the unpredictable can improve environmental decision-making. *Trends in Ecology & Evolution* 29:531–541. <https://doi.org/10.1016/j.tree.2014.07.005>

Dakos V, Carpenter SR, Brock WA, et al (2012) Methods for detecting early warnings of critical transitions in time series illustrated using simulated ecological data. PLoS ONE 7:.

<https://doi.org/10.1371/journal.pone.0041010>

Dietze MC (2017) Ecological Forecasting. Princeton University Press, Princeton

Geary WL, Bode M, Doherty TS, et al (2020) A guide to ecosystem models and their environmental applications. Nature Ecology and Evolution 4:1459–1471.

<https://doi.org/10.1038/s41559-020-01298-8>

Henden JA, Ims RA, Yoccoz NG, et al (2020) End-user involvement to improve predictions and management of populations with complex dynamics and multiple drivers. Ecological

Applications 0:1–14. <https://doi.org/10.1002/eap.2120>

Hotchkiss ER, Sadro S, Hanson PC (2018) Toward a more integrative perspective on carbon metabolism across lentic and lotic inland waters. Limnology and Oceanography Letters

3:57–63. <https://doi.org/10.1002/lol2.10081>

Jane SF, Hansen GJA, Kraemer BM, et al (2021) Widespread deoxygenation of temperate lakes.

Nature 594:66–70. <https://doi.org/10.1038/s41586-021-03550-y>

Kirici M (2017) Toxic effects of copper sulphate pentahydrate on antioxidant enzyme activities and lipid peroxidation of freshwater fish *Capoeta umbla* (Heckel, 1843) tissues. Applied

Ecology & Environmental Research 15:1685–1696.

https://doi.org/10.15666/aeer/1503_16851696

- Kox T, Kempf H, Lüder C, et al (2018) Towards user-orientated weather warnings. *International Journal of Disaster Risk Reduction* 30:74–80. <https://doi.org/10.1016/j.ijdrr.2018.02.033>
- Leal PR, Moschini-Carlos V, López-Doval JC, et al (2018) Impact of copper sulfate application at an urban Brazilian reservoir: A geostatistical and ecotoxicological approach. *Science of The Total Environment* 618:621–634. <https://doi.org/10.1016/j.scitotenv.2017.07.095>
- Lin S, Pierson DC, Ladwig R, et al (2023) Multi-Model Machine Learning Approach Accurately Predicts Lake Dissolved Oxygen with Meteorological and Hydrological Input. *SSRN Preprint*. <https://doi.org/10.2139/ssrn.4454256>
- Liu J, Dietz T, Carpenter SR, et al (2007) Complexity of Coupled Human and Natural Systems. *Science* 317:1513–1516. <https://doi.org/10.1126/science.1144004>
- Luo L, Lan J, Wang Y, et al (2022) A Novel Early Warning System (EWS) for Water Quality, Integrating a High-Frequency Monitoring Database with Efficient Data Quality Control Technology at a Large and Deep Lake (Lake Qiandao), China. *Water* 14:602. <https://doi.org/10.3390/w14040602>
- Mackay EB, Jones ID, Folkard AM, Thackeray SJ (2011) Transition zones in small lakes: the importance of dilution and biological uptake on lake-wide heterogeneity. *Hydrobiologia* 678:85–97. [https://doi.org/DOI 10.1007/s10750-011-0825-y](https://doi.org/DOI%2010.1007/s10750-011-0825-y)
- Nichols JD, Runge MC, Johnson FA, Williams BK (2007) Adaptive harvest management of North American waterfowl populations: a brief history and future prospects. *Journal of Ornithology* 148:343–349. <https://doi.org/10.1007/s10336-007-0256-8>

- Ortiz DA, Wilkinson GM (2021) Capturing the spatial variability of algal bloom development in a shallow temperate lake. *Freshwater Biology* 66:2064–2075.
<https://doi.org/10.1111/fwb.13814>
- Paerl HW, Gardner WS, McCarthy MJ, et al (2014) Algal blooms: Noteworthy nitrogen. *Science* 346:175–175. <https://doi.org/10.1126/science.346.6206.175-a>
- Pappenberger F, Ramos MH, Cloke HL, et al (2015) How do I know if my forecasts are better? Using benchmarks in hydrological ensemble prediction. *Journal of Hydrology* 522:697–713. <https://doi.org/10.1016/j.jhydrol.2015.01.024>
- Parrott L (2017) The modelling spiral for solving ‘wicked’ environmental problems: guidance for stakeholder involvement and collaborative model development. *Methods in Ecology and Evolution* 8:1005–1011. <https://doi.org/doi:10.1111/2041-210X.12757>
- Richardson DC, Carey CC, Bruesewitz DA, Weathers KC (2017) Intra- and inter-annual variability in metabolism in an oligotrophic lake. *Aquatic Sciences* 79:319–333.
<https://doi.org/10.1007/s00027-016-0499-7>
- Shi B, Wang P, Jiang J, Liu R (2018) Applying high-frequency surrogate measurements and a wavelet-ANN model to provide early warnings of rapid surface water quality anomalies. *Science of The Total Environment* 610–611:1390–1399.
<https://doi.org/10.1016/j.scitotenv.2017.08.232>
- Soares LMV, Calijuri M do C (2021) Deterministic modelling of freshwater lakes and reservoirs: Current trends and recent progress. *Environmental Modelling and Software* 144:105143.
<https://doi.org/10.1016/j.envsoft.2021.105143>

Thornton KW, Kimmel BL, Payne FE (1990) *Reservoir Limnology--Ecological Perspectives*.
John Wiley and Sons, New York

Tommasi D, Stock CA, Pegion K, et al (2017) Improved management of small pelagic fisheries through seasonal climate prediction. *Ecological Applications* 27:378–388.
<https://doi.org/10.1002/eap.1458>

Ward EJ, Holmes EE, Thorson JT, Collen B (2014) Complexity is costly: A meta-analysis of parametric and non-parametric methods for short-term population forecasting. *Oikos* 123:652–661. <https://doi.org/10.1111/j.1600-0706.2014.00916.x>

Ward NK, Brentrup JA, Richardson DC, et al (2022) Dynamics of the stream – lake transitional zone affect littoral lake metabolism. *Aquatic Sciences* 84:1–19.
<https://doi.org/10.1007/s00027-022-00854-7>

Wynne JH, Woelmer W, Moore TN, et al (2023) Uncertainty in projections of future lake thermal dynamics is differentially driven by lake and global climate models. *PeerJ* 11:e15445. <https://doi.org/10.7717/peerj.15445>

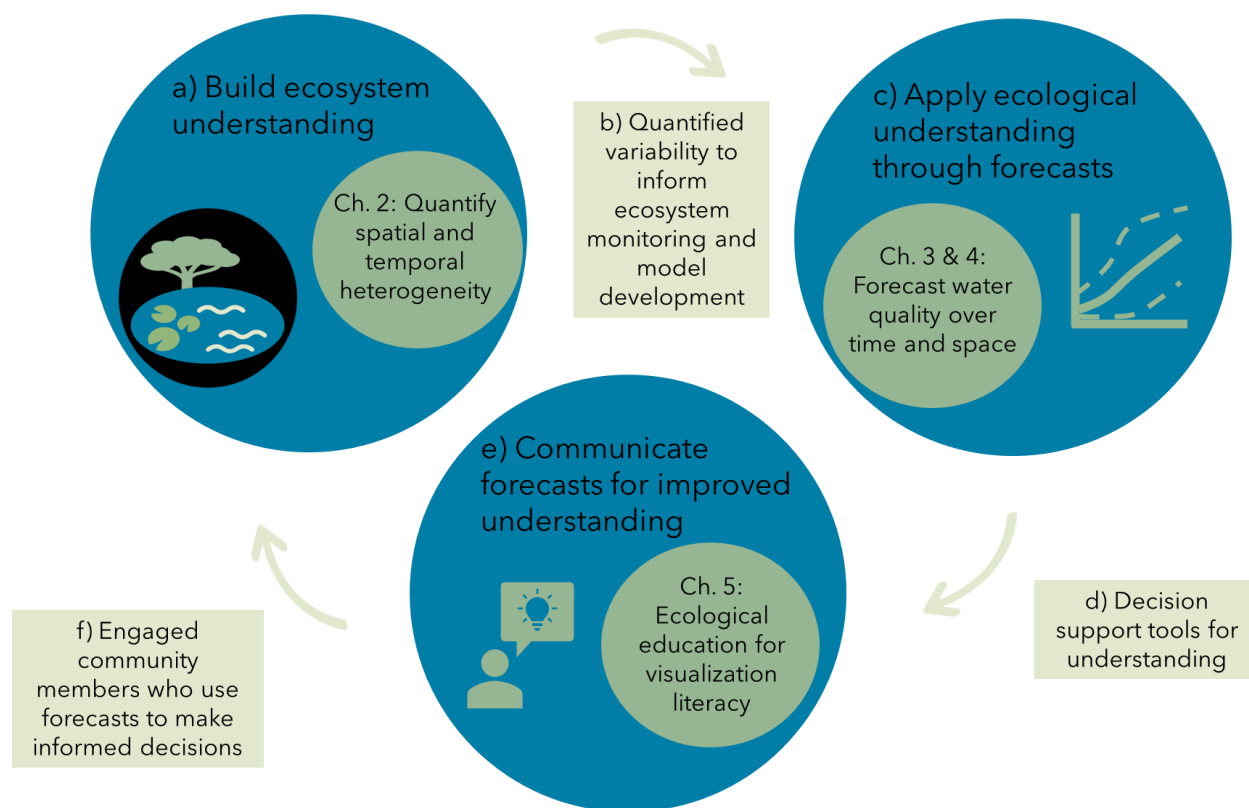


Figure 1. Iterative conceptual framework that integrates my four dissertation chapters (Ch.) to build, apply, and communicate ecological understanding via forecasting.