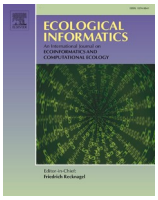
Ecological Informatics 83 (2024) 102825

Contents lists available at ScienceDirect

Ecological Informatics

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



Process-based forecasts of lake water temperature and dissolved oxygen outperform null models, with variability over time and depth

Whitney M. Woelmer [a,*], R. Quinn Thomas [a,b], Freya Olsson [a], Bethel G. Steele [c], Kathleen C. Weathers [c], Cayelan C. Carey [a]

[a] Department of Biological Sciences, 926 West Campus Drive, Virginia Tech, Blacksburg, VA 24061, USA
[b] Department of Forest Resources and Environmental Conservation, 310 West Campus Drive, Virginia Tech, Blacksburg, VA 24061, USA
[c] Cary Institute of Ecosystem Studies, Millbrook, NY 12545, USA

ARTICLE INFO

Keywords:
Baseline model
Climatology
Ecological forecasting
Forecast skill
Persistence
Water quality

ABSTRACT

Near-term iterative ecological forecasting has great potential for providing new insights into our ability to predict multiple ecological variables. However, true, out-of-sample probabilistic forecasts remain rare, and variability in forecast performance has largely been unexamined in process-based forecasts which predict multiple ecosystem variables. To explore how forecast performance varies for water temperature and dissolved oxygen, two freshwater variables important for lake ecosystem functioning, we produced probabilistic forecasts 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 process model, which we assessed relative to both climatology and persistence null models to quantify how much information process-based FLARE forecasts provide over null models across varying environmental conditions. We found that FLARE water temperature forecasts were always more skillful than FLARE oxygen forecasts. Specifically, temperature forecasts outperformed both null models up to 11 days into the future, as compared to only two days for oxygen. Across different years, we observed variable forecast skill, with performance generally decreasing with depth for both variables. Overall, all temperature forecasts and surface oxygen, but not deep oxygen, forecasts were more skillful than at least one null model >80 % of the forecasted period, indicating that our process-based model was able to reproduce the dynamics of these two variables with greater reliability than the null models. However, process-based oxygen forecasts from deeper waters were less skillful than both null models during a majority of the forecasted period, which suggests that deep-water oxygen dynamics are dominated by autocorrelation and seasonal change, which are inherently captured by the null forecasts. Our results highlight that forecast performance varies among lake water quality metrics and that process-based forecasts can provide important information in conjunction with null models in varying environmental conditions. Altogether, these process-based forecasts can be used to develop quantitative tools which inform our understanding of future ecosystem change.

1. 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 (Carey et al., 2022; Dietze et al., 2018; Lofton et al., 2023). With increasing variability in water quality and changes in ecosystem functioning due to land use and climate change (Ho and Michalak, 2019; IPCC, 2023; Kraemer et al., 2021; Woolway and Merchant, 2019), 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 (Lee et al., 2023; Lofton et al., 2023; Radeloff et al., 2015). Through the near-term, iterative forecast cycle, forecasts are repeatedly produced for a range of forecast horizons, or periods of time into the future (Dietze, 2017). When monitoring data become available, forecasts are updated and evaluated with observations, allowing for iterative, improved predictions over time. These predictions can then be integrated into decision-making frameworks (Bodner et al., 2021; Dietze et al., 2018;

* Corresponding author.
E-mail address: wwoelmer@vt.edu (W.M. Woelmer).

https://doi.org/10.1016/j.ecoinf.2024.102825
Received 16 January 2024; Received in revised form 9 September 2024; Accepted 9 September 2024
Available online 11 September 2024

Henden et al., 2020).

Two foundational freshwater variables, which both control and serve as important indicators of lake ecosystem functioning, are water temperature and dissolved oxygen (Jones and Smol, 2023). Water temperature, which is influenced by meteorological variables including air temperature, precipitation, and wind, as well as other factors (Jones and Smol, 2023), is a key regulator of thermal stratification and solubility of gases within the water column, which in turn controls dissolved oxygen concentrations (Bhateria and Jain, 2016; Sánchez et al., 2007). 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 the dynamics of 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 dos Simões et al., 2008; Richardson et al., 2017).

Lake forecasts that include both water temperature and dissolved oxygen together may be able to provide a more comprehensive perspective of changing lake ecosystem functioning, than forecasts of temperature or oxygen alone. Specifically, managers may need both water temperature and dissolved oxygen forecasts simultaneously to guide decision-making, as they often need to optimize multiple concurrent needs (Jackson-Blake et al., 2022a). For example, both water temperature and dissolved oxygen forecasts could guide decisions on the depth of water withdrawal necessary for optimizing downstream habitat for organisms with temperature and oxygen sensitivities (Calamita et al., 2021; Kim and Choi, 2021). Additionally, temperature forecasts could provide advance notice of mixing events (Carey et al., 2022), which would be complementary to dissolved oxygen forecasts of hypolimnetic anoxia. Together, these two forecasts could inform the use of oxygenation systems for mitigating anoxia (Carey et al., 2022), and aid in determining whether to implement chemical water treatment applications (Lee, 2015; Li et al., 2022).

Although near-term iterative ecological forecasts that predict both water temperature and dissolved oxygen simultaneously remain rare, the majority of those which currently exist use coupled machine learning approaches (Lofton et al., 2023). 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 ecological processes (Jia et al., 2018; Lazer et al., 2014; Peters et al., 2014). Of the machine learning studies which predict multiple water quality variables simultaneously, all have generated forecasts of temperature and oxygen from separate temperature and oxygen models (Durell et al., 2022; Lin et al., 2023; Saber et al., 2020). As a result, these studies lack explicit representation of the ecosystem interactions that occur between the two variables, which would otherwise be captured by explicitly defined interactions between state variables in a process model.

In comparison to machine learning models, process-based models are not as commonly used to produce forecasts of dissolved oxygen or water temperature (Durell et al., 2022; Lofton et al., 2023). This may be because process-based models often require numerous input variables, extensive expert-based calibration, and may not provide a gain in skill relative to machine learning models to predict water quality (Durell et al., 2022; Jin et al., 2019). 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 (sensu Cuddington et al., 2013). We are aware of only one published study that 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 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 depth.

As a result of sensitivity to numerous 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 between surface and deeper lake layers in the water column) in lakes are less conclusive. Specifically, Durell et al. (2023) and Saber et al. (2020) found 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 (Carey et al., 2014; Geng et al., 2022; Jassby et al., 2003; Nõges and Tuvikene, 2012). For example, the date of fall turnover, defined as the breakup of summer thermal stratification (Jones and Smol, 2023), can vary from year to year, potentially driving annual differences in predictability of water temperature and dissolved oxygen. 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).

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). Here 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., ^{o}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, provide useful information about how an ecosystem's current conditions compare to historical patterns (Jolliffe and Stephenson, 2012) and are often the "best guess" in an ecosystem that is dominated by longer-term seasonal dynamics. In contrast, null persistence models, which predict that a variable will stay the same as the most recent observation across the forecast horizon, are valuable reference models (Mittermaier, 2008), which may have high skill in forecasting variables with little variation across the forecast horizon, but are less skilled at capturing variables that exhibit dynamics with substantial directional fluctuations over time (Olsson et al., 2024).

Examining the forecast skill of process-based models relative to both climatology and persistence null models can provide useful insight about the additional information process-based models provide. For example, if process-based forecasts are more skillful than climatology forecasts (which best predict dynamics dominated by historical seasonal patterns) or persistence forecasts (which best predict dynamics dominated by autocorrelation), we can make inferences about the mechanisms instantiated within the process-based model that govern dynamics in the target forecast variables beyond seasonality and autocorrelation. However, this comparison across multiple null forecasts and variables remains unexamined in water quality forecasts (Lofton et al., 2023).

Comparing process-based forecasts to null models has been highlighted as an important best practice in ecological forecasting (Lewis et al., 2022), and has great potential for furthering our understanding of what drives forecast successes and failures (Dietze et al., 2018; Lewis et al., 2023; Olsson et al., 2024).

To determine how forecast performance varied across time, space, and ecosystem variable, we produced forecasts of water temperature and dissolved oxygen at two depths over two forecast periods in different 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., ^{o}C for temperature and mg/L for oxygen), and forecast *skill* by comparing accuracy of our process-based forecasts to forecasts from both climatology and persistence null models. Our primary research questions were: 1) How does process-based forecast performance compare between water temperature and oxygen?, and 2) How does process-based forecast skill vary over time, depth, and forecast horizon, relative to climatology and persistence null forecasts?

2. Materials and methods

2.1. Forecasting framework overview

We developed forecasts of water temperature (hereafter 'temperature') and dissolved oxygen (hereafter 'oxygen') with uncertainty for Lake Sunapee, New Hampshire, USA. These forecasts were generated for depths throughout the entire water column at one central lake location over two open-water (i.e., ice-free) 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 (Thomas et al., 2020), with forecast evaluation focusing on only two depths (1.0 m and 10.0 m), where high-frequency observations were available for both temperature and oxygen at a high temporal resolution (see *Methods: Study Site and Observational Data* for information on observational data). FLARE is an open-source forecasting system that uses a 1-dimensional 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 daily near-term, iterative, probabilistic forecasts. Using National Oceanic and Atmospheric Administration (NOAA) weather forecasts as driver data into FLARE, we forecasted 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).

FLARE was used to produce daily forecasts via the following steps: 1) obtain new sensor observations from the field (Fig. 1a.1); 2) access NOAA 35-day meteorological forecasts for the focal lake location (Fig. 1a.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 (Fig. 1a.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 (Fig. 1a.4); and 5) once time has passed and new sensor observations are available, evaluate the forecast using multiple metrics, including comparisons to null model forecasts (Fig. 1a.5).

2.2. Study site and observational data

Lake Sunapee is a deep (33.0 m), medium-sized (16.55 km^2 surface area) lake in central New Hampshire, USA (43.37, −72.05, Fig. 2). 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 Sunapee, 2022). The summer thermocline depth has been documented to range from 6.0 to 8.0 m (Fig. A.1, Carey et al., 2014).

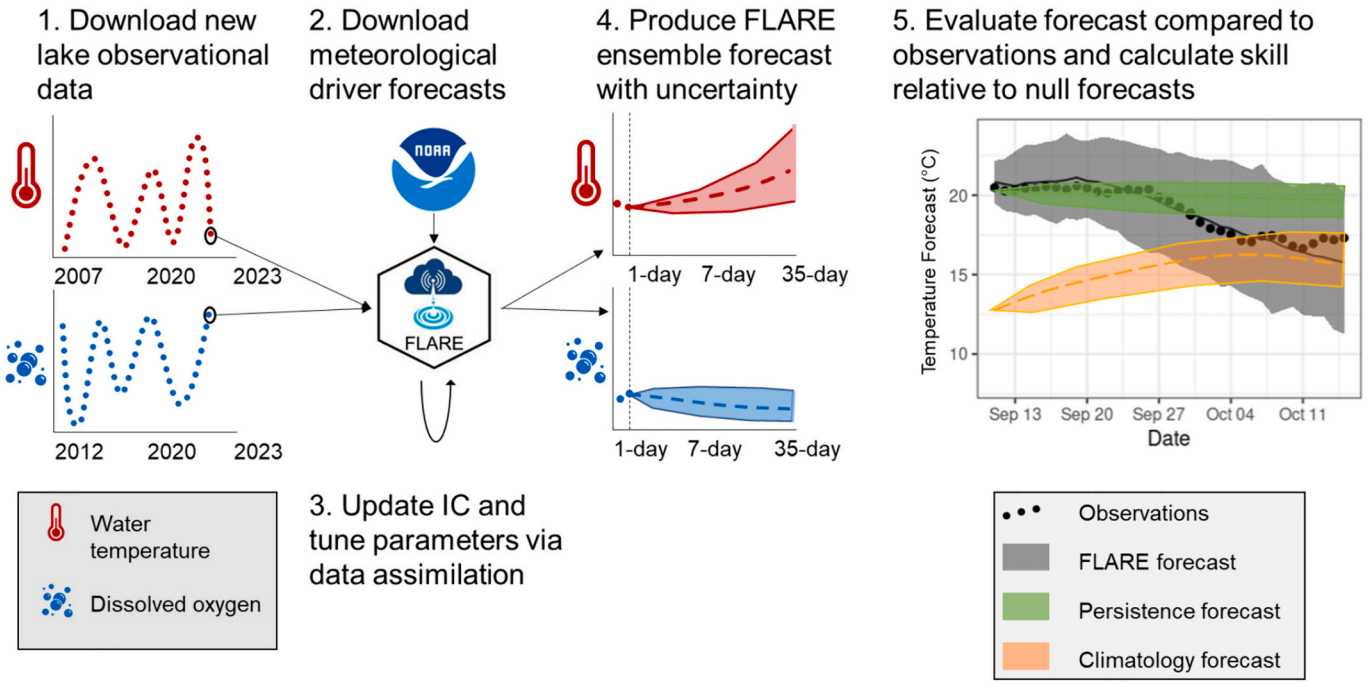


Fig. 1. Conceptual representation of the FLARE (Forecasting Lakes And Reservoir Ecosystems) framework which shows the daily steps of 1) downloading lake observational data and appending the new data to the long-term observational dataset, 2) downloading meteorological forecasts from NOAA needed as driver 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. Use of the emblem/logo does not imply an endorsement by NOAA/NWS.

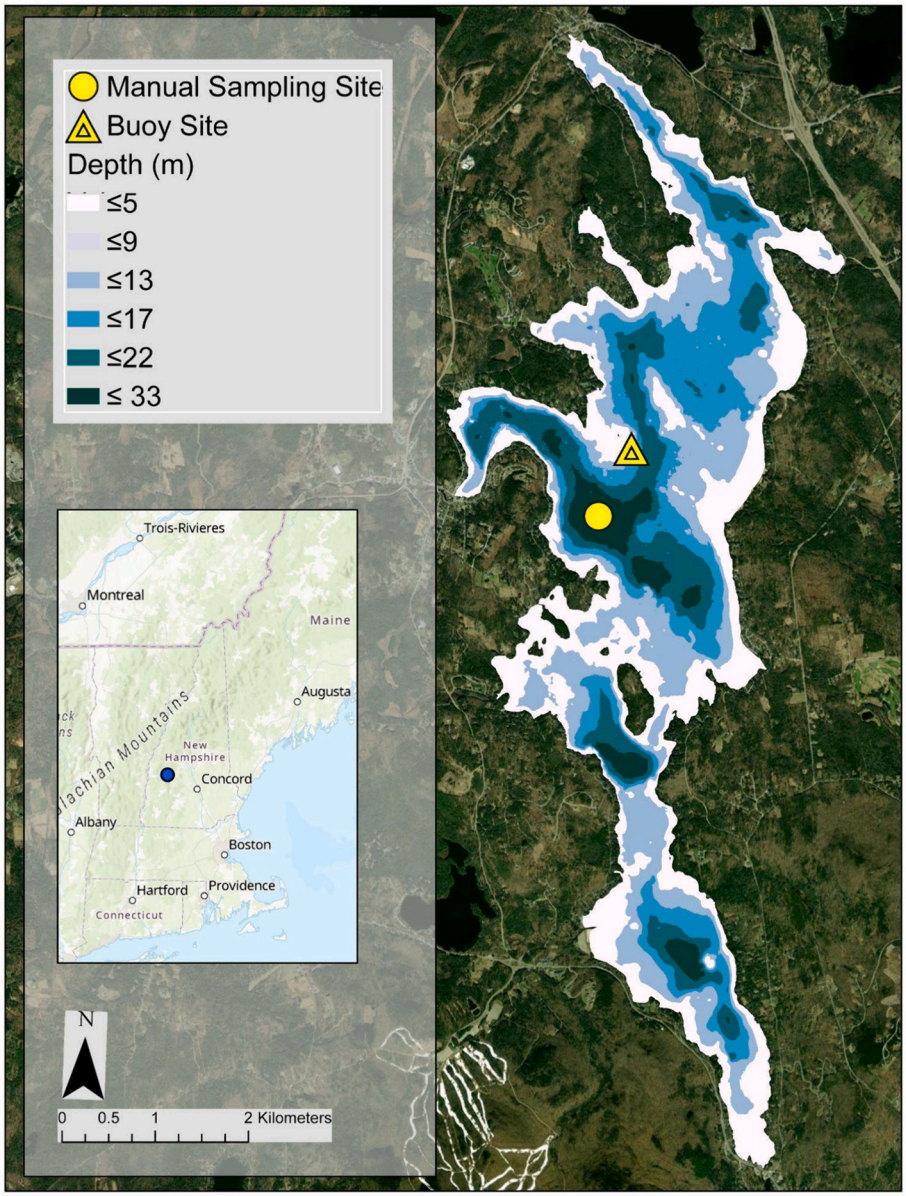


Fig. 2. Bathymetric map of Lake Sunapee, NH, USA (43.37°, −72.05°) showing the location of the manually-collected data sampling site (circle) and the sensored buoy site (triangle). Data from LSPA et al. (2023).Discrete observations of temperature and oxygen have been collected monthly for every meter from the surface to the bottom of the lake during the open-water (i.e., ice free) season since 1986 as part of a long-term monitoring program conducted 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. To complement the monthly manually-collected data, a buoy instrumented with high-frequency sensors was deployed by the LSPA at a nearby site (Fig. 2) in 2007. The buoy is usually deployed annually from ~April or May until ~October, following ice-off and ice-on. The buoy has high-frequency (10-min) 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). The 10-min measurements of water temperature and dissolved oxygen were aggregated into daily averages for use in this study. Observations of temperature were placed in 1.0 m bins to account for marginal differences in thermistor depths among years, with the value given by the top of the bin (e.g., 1.0 m are observations of temperature at depths $\geq$1.0 and $<$ 2.0 m). 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 (Fig. 2, buoy site), while the monthly deep site profiles (which included measurements at depths deeper than the buoy sensors at 10.0 m; Fig. 2, 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 nearby deep site (Fig. 2; Steele et al., 2023) show high agreement (Pearson correlation $r = 0.93$; Wynne et al., 2023), so we combined temperature and oxygen 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. A single hypsometric curve representing the bathymetry from the lake's surface to its deepest site was used as an input for FLARE (LSPA, 2023, Fig. 2).

In our two study years, the timing of the open-water buoy deployment varied substantially due to maintenance (Table A.1), 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 A.1 for a description of data and forecast duration each year). As a result, our forecast 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).

2.3. FLARE input data

FLARE requires observations of state variables (in this study, water temperature and dissolved oxygen), forecasted driver variables (in this study, air temperature, shortwave and longwave radiation, windspeed, relative humidity, and precipitation), and hypsography. Observations of temperature and oxygen were taken from the buoy and deep sites during the study period (LSPA et al., 2023, Fig. 2) 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 is a state-of-the-art weather product commonly used to drive ecological forecasts with demonstrated skill in recreating meteorological observations to within <1.6 °C RMSE (Hamill et al., 2022). NOAA GEFS 35-day forecasts are produced every six hours and were downscaled from the 6-h to 1-h resolution to meet the required timestep of GLM within FLARE following the methods of Thomas et al. (2020). Hypsography was provided by the Lake Sunapee Protective Association at the deepest site of the lake (LSPA, 2023).

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. Since new NOAA GEFS forecasts were generated every 6 h, there were four values for each day (the 0-h horizon for states and the 0-6 h horizon for fluxes in each forecast) that were interpolated to an hourly product following the downscaling methods outlines in Thomas et al. (2020). 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 and ensures a seamless transition between historical and forecasted weather 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; Solomon et al., 2013), we used a 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, sediment-water interface fluxes controlled by biological and chemical sediment oxygen demand, and temperature-based solubility of oxygen. 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 model parameters that were selected based on the sensitivity of the forecasts to parameter-fitting in previous studies (Thomas et al., 2020, 2023; Wander et al., 2023). The model 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 across sediment zones. 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 A.2. All configuration files are available at Woelmer et al. (2024). Additional parameters which were not included in automated fitting were fixed based on a previously calibrated version of GLM-AED at Lake Sunapee (Ward et al., 2020), with calibration from 2007 to 2015 and validation from 2015 to 2020 for an RMSE (root mean square error) of <2 °C water temperature and 2 mg/L dissolved oxygen.

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 all forecast horizons to be represented within the time period, avoiding bias towards fitting only short-horizon forecasts and acclimating the forecasts to the conditions of that year. We observed that parameters evolved through the full 35-day period, emphasizing the importance of having a spin-up period of this duration (Fig. A.2).

FLARE was configured to quantify four different sources of forecast uncertainty: model initial conditions, model parameter, model process, and weather driver data which varied across our 200 ensemble members (see Table A.3 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 A.3. We ran FLARE with 200 ensemble members to ensure a reasonable spread of uncertainty around our forecasts (following Thomas et al., 2020, 2023).

2.5. Null model forecasts

We compared our process-based FLARE forecasts of water temperature and dissolved oxygen to two null model forecasts (a climatology null and a persistence null) to calculate forecast skill. 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 process-based forecast information obtained above the nulls (Dietze et al., 2018; Harris et al., 2018; Lewis et al., 2022). Both climatology and persistence forecasts of water temperature and dissolved oxygen were developed for the same forecast period as the FLARE process-based forecasts. First, 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 (Fig. A.3). Second, the persistence null model was calculated as the last observation for a given water quality variable and depth with random noise added at each forecast horizon across the 35-day forecast horizon. Random noise was calculated following a random walk model with the error term drawn from a distribution of residuals from the historical fit of the persistence forecast and added to the observation to produce the next day's

persistence forecast. Persistence forecasts were generated using the RW (random walk) function within the fable R package (version 0.3.2; O'Hara-Wild et al., 2022).

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:

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

Where CRPS is a function of F, the forecasted cumulative distribution function and x, the observed outcome; 1 is a step function which equals 1 if $u \geq x$, and otherwise equals 0 (Gneiting et al., 2005). Forecasts were scored using the *scoringRules* package in R (Jordan et al., 2019). 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 that also evaluates forecast precision. Lower values of CRPS indicate better performance. We also calculated absolute error (AE) between the mean of the forecast distribution and the observation, but found similar patterns between the AE and CRPS so focused our reporting on CRPS within the main text. AE results are provided in Fig. A.4. Both FLARE forecasts and null forecasts (climatology and persistence) were scored using CRPS.

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

$$Forecast\ Skill = 1 - \frac{CRPS_{FLARE}}{CRPS_{Null}} \quad (2)$$

where $CRPS_{FLARE}$ is the CRPS calculated from a single FLARE forecast and $CRPS_{Null}$ is the CRPS of the null model, either the climatology or the persistence null. 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 a given null model. For this forecast skill metric, values of 0 indicate that the FLARE forecast and null 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. We calculated forecast skill across both variables (temperature and oxygen), depths (1.0 and 10.0 m), and for each null model. We denote climatology-based forecast skill as $Skill_{Climatology}$ and persistence-based forecast skill as $Skill_{Persistence}$.

Another important component of forecast skill is the change 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 each null forecast within each forecast period.

We also evaluated the calibration of forecast confidence intervals using reliability plots (Bröcker and Smith, 2007). These plots were calculated by the number of forecasts which fell within specific confidence intervals, ranging from the 10th to the 90th confidence intervals across all forecast horizons and for both water quality variables and both depths. Reliability refers to the statistical agreement of forecast probabilities and observed frequencies of events (Gneiting et al., 2007; Schepen et al., 2016) and these plots show the degree to which a predicted distribution matches the underlying distribution of the data.

Lastly, we evaluated the ability of FLARE forecasts to predict the date of fall turnover (following Thomas et al., 2020), by calculating the percent of forecast ensemble members that predicted turnover on each day leading up to the observed date of turnover. We defined turnover as the first day on which there was less than 1 °C difference in temperature between the surface (0.1 m) and bottom-most temperature measurement (in our case, 10.0 m given a lack of deeper sensor measurements). We identified the earliest forecast to predict the true date of turnover.

All statistical analyses and forecast deployments were conducted in the R statistical environment, version 4.2.2 (R Core Team, 2022). All forecast output is archived in the Zenodo repository (Woelmer et al., 2023), as well as code to run the forecasts and reproduce the forecast analysis (Woelmer et al., 2024).

3. Results

3.1. Variability with depth was greater for temperature than oxygen, while inter-annual variability was greater for oxygen than temperature

In both years, surface temperature generally decreased, while bottom temperature generally increased throughout the open-water season, following expected seasonal patterns (Fig. 3). Oxygen patterns also followed expected seasonal patterns, with decreases in bottom oxygen and small increases in surface oxygen throughout the season (Fig. 3). Turnover occurred on October 4 in 2021 and was associated with a marked increase in bottom oxygen variability. In contrast, in 2022, turnover occurred on September 23 and was followed by a sharp decrease in bottom water temperature and an increase in bottom water oxygen.

Observations of water temperature during the forecast period differed more between the two study years than dissolved oxygen, but exhibited greater differences across depth (Fig. 3). Specifically, mean water temperature was only 0.1 °C different between years at 1.0 m (Fig. 3a, c; 21.4 °C in 2021 and 21.3 °C in 2022), and up to 1 °C different between years at 10.0 m (Fig. 3e, g; 12.7 °C in 2021 and 13.7 °C in 2022). However, there were large differences across depth for water temperature in both years, with temperature at 1.0 m during the forecast period 6.6 °C warmer than 10.0 m (Fig. 3a, e). In comparison to temperature, differences in oxygen were generally greater between years than between depths (Fig. 3b, d, f, h). At 1.0 m, oxygen was 1 mg/L lower in 2022 than in 2021 (Fig. 3b, d; 9.1 mg/L in 2021; 8.1 mg/L in 2022). At 10.0 m mean oxygen was 0.5 mg/L higher in 2022 (Fig. 3f, h; 9.1 mg/L in 2021, 9.6 mg/L in 2022) and much more variable in 2021 (range = 4.09 mg/L) than in 2022 (range = 1.67 mg/L; Fig. 3h) at 10.0 m. Averaged across the two years, mean oxygen concentrations were similar between depths (Fig. 3h; 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 concentration (mg/L) (Fig. A.5); correspondingly, we present results in mg/L throughout the manuscript for ease of interpretation.

3.2. Accuracy of FLARE-forecasted water temperature and dissolved oxygen across the forecast period

Forecast accuracy evaluated using CRPS across all years and depths was <1.1 °C for temperature and < 0.6 mg/L for oxygen (Fig. 4a, b). Both variables exhibited nonlinear decreases in forecast accuracy across the 35-day forecast horizon (Fig. 4a, 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 (Fig. 4a). Oxygen forecasts similarly had a CRPS of 0.3 mg/L at 1-day ahead, decreasing to 0.5 mg/L at 7 days, with accuracy leveling out around 21 days with a CRPS of ~0.6 mg/L through to the end of the 35-day forecast horizon (Fig. 4b).

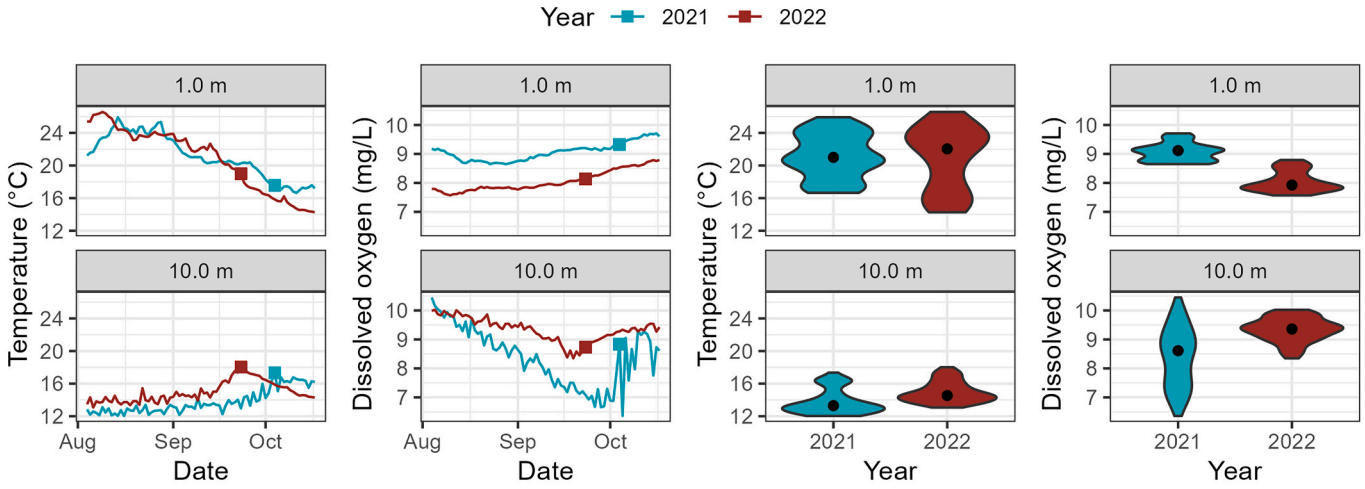


Fig. 3. High-frequency observations of daily water temperature and dissolved oxygen from the buoy in Lake Sunapee over the forecast period (August–October) in 2021 and 2022 showing the time series of observations for a) temperature at 1.0 m, b) oxygen at 1.0 m, violin plots of observations for c) temperature at 1.0 m, d) oxygen at 1.0 m. 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.Turnover is represented by the square points on the time series in panels a, b, e, and f. Circle points on violin plots represents the median of the observations.

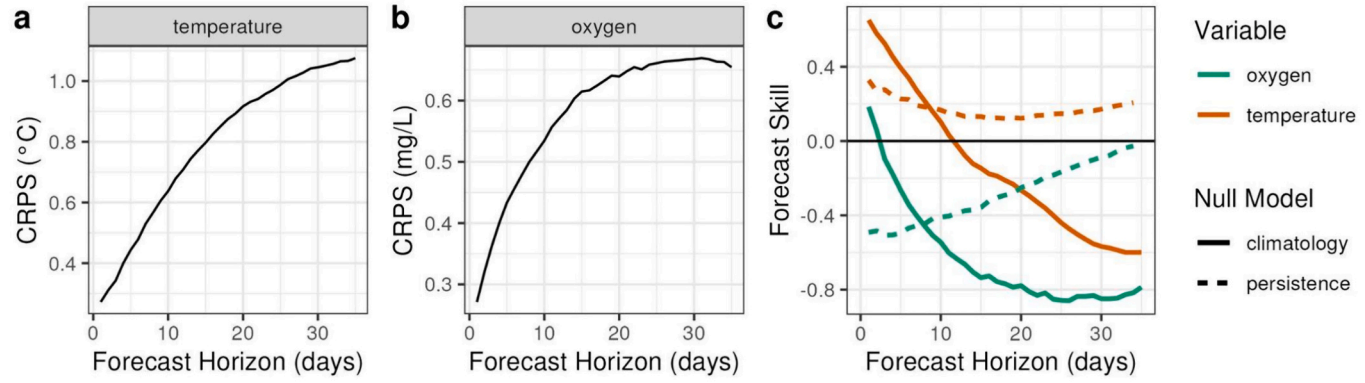


Fig. 4. FLARE forecast accuracy of a) water temperature (°C), and b) dissolved oxygen (mg/L) and c) forecast skill (unitless; calculated using Eq. (2)) of both temperature and oxygen. These metrics were calculated by aggregating mean forecast skill over the forecast period and across all depths at forecast horizons from 1 to 35 days into the future (note the difference in y-axes). In panels a and b, more accurate FLARE forecasts are represented by lower CRPS values. In Panel c, values less than zero indicate a less skillful FLARE forecast than the null model.

3.3. Water temperature forecasts were more skillful than dissolved oxygen

Across the forecast period, forecast skill was higher for temperature than for oxygen across all horizons, regardless of which null forecast was used to calculate skill (Fig. 4c). When comparing FLARE and climatology forecasts (hereafter, $Skill_{Climatology}$), $Skill_{Climatology}$ overall decreased over the forecast horizon for both variables, but the strength of this pattern varied between water quality variables (Fig. 4c). Specifically, aggregated temperature forecasts outperformed the climatology null at more horizons than oxygen, with temperature being skillful up to 11 days into the future, while oxygen was only skillful 2 days into the future (Fig. 4c). While $Skill_{Climatology}$ of oxygen (mean skill = −0.6) was overall worse than for temperature (mean skill = −0.08), the rate at which $Skill_{Climatology}$ degraded over the 35-day forecast horizon was slower for oxygen and saturated at ~20 days into the future, while temperature showed a near-linear decline across the 35-day horizon (Fig. 4c). However, forecasts of both temperature and oxygen were not skillful (i.e., climatology forecasts performed better than FLARE forecasts) at longer forecast horizons.

Forecast skill relative to a persistence forecast (hereafter, $Skill_{Persistence}$) showed different patterns from $Skill_{Climatology}$. First, $Skill_{Persistence}$ of FLARE forecasts increased over the forecast horizon, especially for oxygen forecasts, indicating that the value of FLARE forecasts over persistence forecasts improved with time into the future (Fig. 4c). Additionally, temperature forecasts were always skillful relative to the persistence forecast, across the full 35-day horizon, whereas FLARE oxygen forecasts were never more skillful, although skill at the 35-day-horizons was near zero (Fig. 4c).

The uncertainty around the water temperature forecasts was generally better calibrated than for oxygen forecasts, as determined by reliability plots (Fig. A.6). Specifically, oxygen forecasts tended to be underconfident (i.e., uncertainty was too large), whereas temperature forecast uncertainty was well calibrated at mid-forecast horizons (14–21 days ahead), but overconfident (i.e., uncertainty was too small) before ~14 days and underconfident after ~21 days.

Overall, FLARE forecasts were generally more skillful than climatology null forecasts at short horizons and persistence null forecasts at longer horizons. The trade-off in skill between climatology and persistence (i.e., where the two lines intersect on Fig. 4c) occurred at 9-days ahead for temperature and 7-days ahead for oxygen. This intersection highlights the forecast horizon at which autocorrelation (via null persistence) and seasonality (via null climatology) each dominate

processes in these two water quality variables.

3.4. Forecast skill varies by year and depth

Forecast skill (accuracy relative to a null model) was different between years across all variables and null models (climatology or persistence), and generally worse at 10.0 m than 1.0 m (Fig. 5). For $Skill_{Climatology}$, FLARE temperature forecasts were more skillful in 2021 than 2022 across both depths (Fig. 5a, 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. 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 (Fig. 5c).

$Skill_{Climatology}$ varied by depth and year for oxygen (Fig. 5b, f). At 1.0 m, $Skill_{Climatology}$ for oxygen was higher in 2022, with forecasts at all horizons (1–35 days-ahead) more skillful than the climatology forecasts, while no horizons in 2021 exhibited skillful oxygen forecasts (Fig. 5b). In comparison, at 10.0 m, median oxygen $Skill_{Climatology}$ was similar between years, with skillful forecasts up to only two days into the future for both 2021 and 2022 (Fig. 5f). Higher $Skill_{Climatology}$ in 2022 than 2021 can be attributed to substantially higher accuracy of the climatology null model in 2022 (Fig. A.7), rather than low accuracy of FLARE forecasts, which were generally similar between 2021 and 2022 (Fig. A.8b, d). Overall, the range in $Skill_{Climatology}$ across years was greater in 2021 than 2022 at both depths, but especially at 10.0 m (Fig. 5b, 6f).

Persistence-based forecast skill ($Skill_{Persistence}$) also varied by year and depth for temperature and oxygen, with $Skill_{Persistence}$ generally decreasing with depth (Fig. 5c-d, g-h). For temperature, $Skill_{Persistence}$ was worse in 2022 than 2021 at 1.0 m, although most forecasts were skillful in both years for temperature at 1.0 m, across all horizons and forecasted days. $Skill_{Persistence}$ was lower at 10.0 m, with forecasts in 2022 more frequently skillful than forecasts in 2021 across all forecast horizons.

For oxygen, $Skill_{Persistence}$ was overall lower than for temperature, but also varied by year and showed greater skill at 1.0 m than 10.0 m. Across both 1.0 m and 10.0 m depths, forecasts in 2021 were more skillful than in 2022. At 1.0 m, all forecasts in 2021 were more skillful than the persistence after 1-day forecasts, and after 15 days into the future in 2022, indicating that FLARE forecasts provide significant additional information over the persistence forecast at 1.0 m oxygen (Fig. 5d). In contrast, at 10.0 m, forecasts were almost always less skillful than a persistence null (Fig. 5h), with very few forecasts skillful at this depth, although 2021 forecasts were more skillful than 2022 forecasts across the forecast horizon.

We focused our evaluation on 1.0 and 10.0 m forecasts because these are the depths with both temperature and oxygen high-frequency observations. However, we also evaluated water temperature forecast accuracy, as measured by CRPS, across all available sensor depths (every meter from the surface to 10.0 m). In 2021, forecasts were more accurate at surface depths than deeper in the water column, with a similar mean CRPS of <0.75 °C across all depths from the surface to 6.0 m (approximate depth of the summer thermocline), but worse performance deeper in the water column from 7.0 m to 10.0 m, with CRPS ranging from 1 °C and above (Fig. A.9). In contrast, in 2022, forecasts from 8.0 m to 10.0 m were the best performing especially at longer horizons, with a mean CRPS <0.8 °C, while forecasts from the surface to 6.0 m all had a mean CRPS >0.8 °C (Fig. A.9). Across 2021 and 2022, forecast performance generally decreased with forecast horizon. Additionally, forecasts identified the date of turnover up to four days in advance in 2021 but not in 2022 (Fig. A.10).

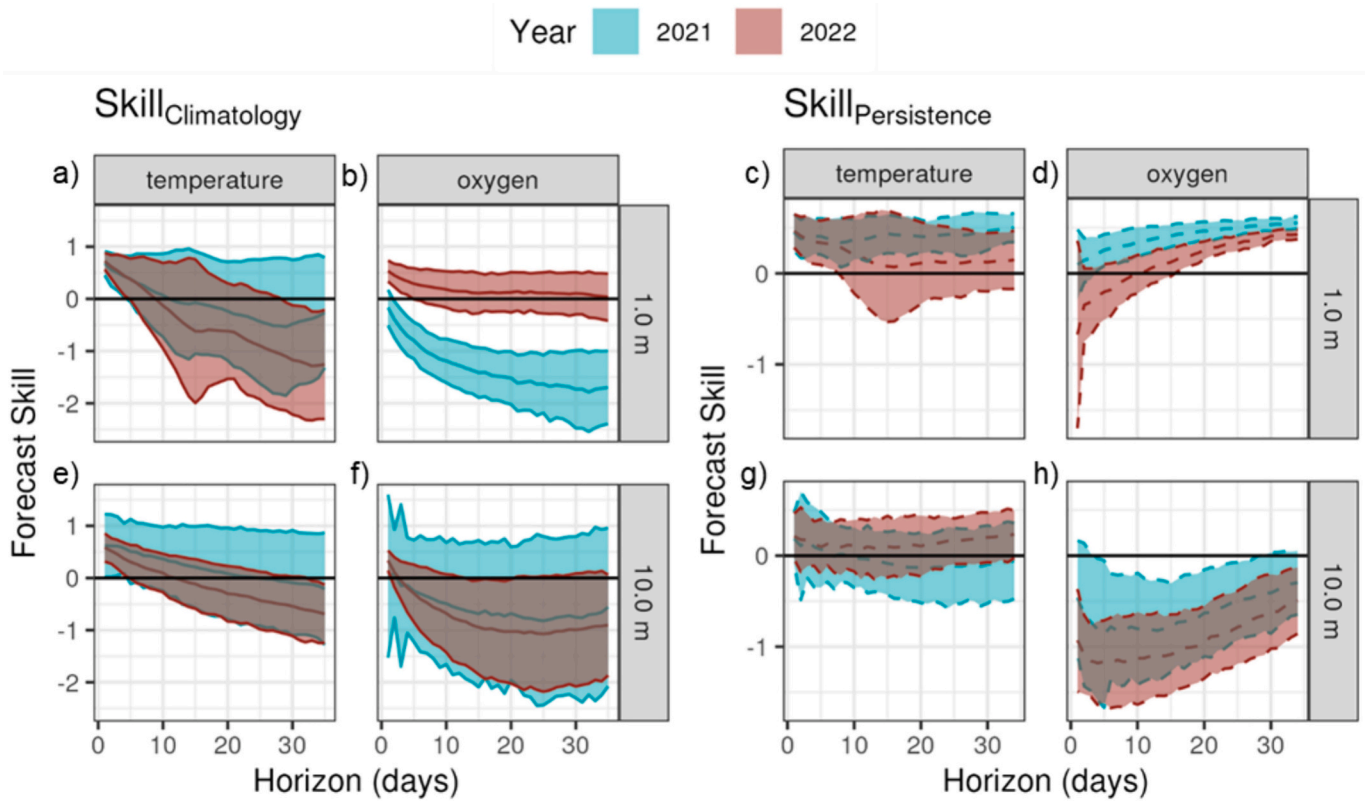


Fig. 5. Forecast skill across the forecast period of August–October in 2021 and 2022. The top row shows surface (1.0 m) $Skill_{Climatology}$ of a) water temperature at 1.0 m, b) dissolved oxygen at 1.0 m and $Skill_{Persistence}$ c) of water temperature at 1.0 m and d) dissolved oxygen at 1.0 m. In the bottom row, skill of 10.0 m forecasts are shown with $Skill_{Climatology}$ of e) temperature at 10.0 m, f) oxygen at 10.0 m and $Skill_{Persistence}$ of g) temperature 10.0 m and h) oxygen at 10.0 m.

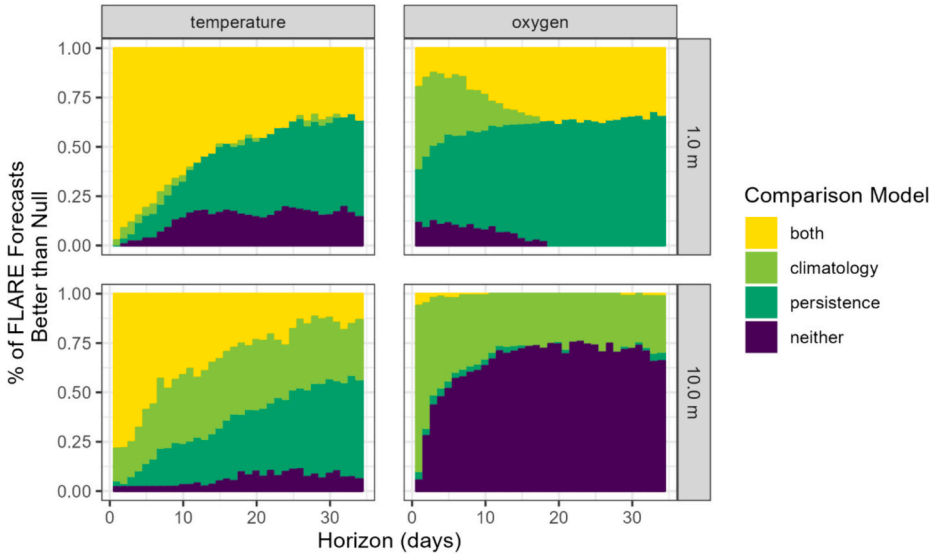


Fig. 6. Percent of FLARE forecasts at 1 to 35-day horizons (aggregated over the entire forecasting period) which were more skillful than both null models (yellow, listed as 'both' in the figure legend), only the climatology null model (lime green), only the persistence null model (teal green), or when FLARE did not perform better than either null model (purple, listed as 'neither' in the figure legend) for a) water temperature at 1.0 m, b) dissolved oxygen at 1.0 m, c) temperature at 10.0 m, and d) oxygen at 10.0 m. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.5. Trade-offs in forecast skill depend on choice of null model

FLARE forecasts outperformed at least one null model for a majority of forecasts across both water quality variables and depth and across the full forecast horizon (Fig. 6). Temperature forecasts outperformed both null models nearly 100 % of the time in 1-day ahead forecasts, with only 14 % of forecasts outperforming neither null model by the end of the 35-day forecast horizon (Fig. 6a). When FLARE forecasts did not outperform both null models, FLARE forecasts at 1.0 m were more likely to outperform persistence forecasts than climatology forecasts across all forecast horizons. At 10.0 m, temperature forecasts were also highly skillful, outperforming at least one null model 98 % of the time in 1-day ahead forecasts, and decreasing only to 94 % of the time by the end of the forecast horizon (Fig. 6c). In contrast, 10.0 m forecasts were more likely to be skillful relative to the climatology null, with ~25 % of forecasts beating climatology throughout the horizon, as compared to only 1–5 % of forecasts outperforming the climatology alone at 1.0 m.

In contrast to temperature, the performance of FLARE oxygen forecasts relative to a null was highly variable with depth. At 1.0 m, the FLARE oxygen forecasts were highly skillful from 1 to 35-days ahead, with 89 % of 1-day forecasts outperforming at least one null model, and increasing to 100 % by the end of the forecast horizon. Overall, FLARE oxygen forecasts at 1.0 m were more likely to outperform a persistence forecast than a climatology forecast, especially at longer horizons. FLARE oxygen forecasts at 10.0 m were highly skillful at short horizons, with 95 % outperforming at least one null model at the 1-day horizon. However, 10.0 m FLARE oxygen forecasts were less skillful at longer forecast horizons, with 66 % outperforming neither null model by the end of the 35-day forecast horizon. However, FLARE forecasts of oxygen at 10.0 m were more likely to outperform a climatology null than a persistence null by the end of the forecast horizon, with 28 % outperforming a climatology, and only 5 % outperforming the persistence null. Across both temperature and oxygen, FLARE forecasts at the surface (1.0 m) were more likely to outperform the persistence forecast than climatology forecast (Fig. 6a, b). The opposite pattern was observed in deeper forecasts (10.0 m), which were more likely to outperform a climatology forecast than a persistence forecast (Fig. 6c, d).

4. Discussion

4.1. Overview

Forecasts of water temperature and dissolved oxygen using a process-based model with daily data assimilation were highly accurate, predicting dynamics within ~1 °C and ~ 1 mg/L, respectively, up to 35-days into the future. Temperature forecasts were more skillful than oxygen forecasts overall, with temperature forecasts being skillful at least 11 days into the future and oxygen forecasts only two 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. Forecast skill also varied based on the null model being compared, with FLARE forecasts outperforming a larger percentage of climatology forecasts at shorter horizons, and a larger percentage of persistence forecasts at longer horizons. Overall, temperature forecasts were more likely than oxygen to outperform both persistence and climatology forecasts, indicating that FLARE forecasts of temperature provide more process information above autocorrelation and seasonal dynamics than for forecasts of oxygen. 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.

4.2. Why the differences in relative skill of temperature and oxygen forecasts?

We found that water temperature forecasts were more skillful than oxygen forecasts according to both of our null models. This result follows the expectation that physical variables are more predictable than chemical or biological variables (Fig. 4c; Soares and do Calijuri, 2021). This is likely a result of the number and complexity of processes that impact water quality variables like oxygen, especially the non-linear dynamics and feedback mechanisms which may be increasing from physical to chemical to biological processes. However, by the end of the 35-day horizon, the difference in skill between temperature and oxygen for both $Skill_{Persistence}$ and $Skill_{Climatology}$ was minimal, with $Skill_{Persistence}$ for oxygen indicating skill may continue to increase with longer horizons (Fig. 4c). This result may indicate that at horizons longer than 35

days, FLARE oxygen forecast skill would continue to improve relative to FLARE temperature forecasts. We hypothesize that increased oxygen skill would occur at longer horizons as the importance of autocorrelation (captured by the persistence null) and seasonal variability (captured by the climatology null) decreases and FLARE process representation of oxygen dynamics dominate. This would be the case if, at longer horizons, oxygen is dominated by mechanistic drivers of dynamics, which would be represented by FLARE, and result in an increase in FLARE skill as autocorrelation (i.e. persistence) and seasonality (i.e., climatology) decrease in importance. However, we cannot currently test this hypothesis due to limitations in the maximum forecast horizon available for the weather forecasts which drive FLARE forecasts. Similarly, it may be possible that the calibration of the uncertainty confidence intervals played a role in the relative skill of the water quality variables (Fig. A.6), but we would need additional summers of forecasts to definitively examine this pattern (see below).

Differential skill across water quality variables is not uncommon. For example, Peng et al. (2019) found that process-based dissolved oxygen forecasts were less skillful (relative to a persistence forecast) than forecasts of total nitrogen and phosphorus. In contrast, another study found that machine-learning derived forecasts of total phosphorus and cyanobacterial concentrations were more skillful than a climatology forecast, but not forecasts of aggregate measures of chlorophyll-*a* or lake color (Jackson-Blake et al., 2022b). Within the existing literature, we were unable to find direct comparisons of water temperature and dissolved oxygen forecast skill. Across these studies and our own, patterns in skill from physical to biological forecast variables remain unclear, likely due to interactions with external drivers (e.g., weather patterns) and differences in the importance of lake processes represented in models across ecosystems (e.g., from eutrophic to oligotrophic systems). Ultimately, while it is not uncommon for ecological forecasts to be less skillful than null models (Page et al., 2018; Mercado-Bettín et al., 2021; Woelmer et al., 2022; Wheeler et al., 2023), more detailed exploration of forecast performance across ecosystem variables, as well as across seasons, is needed.

While we found that forecasts of water temperature were more skillful than forecasts of oxygen in Lake Sunapee, forecast performance 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. 3, Fig. A.11; 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 with higher productivity and algal blooms that cause large changes in dissolved oxygen concentrations over short-term time scales (e.g., Ladwig et al., 2021). Additionally, higher inter-annual variability in oxygen (Fig. 3) may have also led to lower predictability from year to year for oxygen as compared to temperature. Ultimately, these results emphasize the value of calculating forecast skill relative to multiple baseline models and the need for similar studies across a gradient of productivity to help inform our understanding of lake ecosystem predictability more broadly.

4.3. Mechanisms for variability in forecast performance over time and depth

Forecast performance (i.e., accuracy and skill) of water temperature and dissolved oxygen varied between years (Fig. 5, Fig. A.8), demonstrating that predictability of ecosystem variables can vary inter-annually. One possible explanation for inter-annual variability is differences in weather conditions, which can directly influence temperature and oxygen in lakes. However, we did not see substantial differences in the overall magnitude or variability of the meteorological variables which drive FLARE between years (Fig. A.12). Rather, there were differences in the timing and magnitude of rain events, including two large events in September 2022 which likely led to fall turnover that year, yet no such similar occurrences in 2021 (Fig. A.13). It is likely that differences in the timing of specific weather events can also influence predictability of in-lake conditions, in addition to overall in-lake variability. Global change will continue to alter inter- and intra-annual weather patterns, corresponding to more variable physical (Sharma et al., 2021; Woolway et al., 2019, 2021; Woolway and Merchant, 2019) and chemical (Carey, 2023) conditions, which has important implications for the performance of lake water quality forecasts. Across temperature and oxygen, bottom water forecasts at 10.0 m showed lower performance than surface water forecasts at 1.0 m. The majority of studies examining forecast performance across multiple depths have generally found that bottom water forecasts have higher performance than surface waters for both temperature and oxygen (Durell et al., 2023; Saber et al., 2020; Thomas et al., 2020; 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. It is possible that this divergence from our findings may have been influenced by the location of our 10.0 m sensor in the upper hypolimnion (Fig. A.1), rather than at the deepest point of Lake Sunapee (33.0 m). However, a historical data comparison shows that 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 (Pearson correlation $r = 0.89$ and 0.87, respectively; $n = 180$ observations, Fig. A.14). There is a decrease in similarity between the 10.0 m and 30.0 m observations ($r = 0.55$, Fig. A.14), although we note that 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. Thus, our results suggest that differences in forecast performance between surface and bottom layers may vary across waterbodies based on other factors (e.g., waterbody type, morphometry, productivity, mixing dynamics) and that more examination is needed to better understand this vertical pattern across ecosystems.

Interestingly, decreases in forecast performance between the surface and 10.0 m 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 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 or other solutes that can also alter hypolimnetic oxygen cycling. In addition, differences in inter-annual variability between oxygen and temperature may have also impacted differences in performance across depth. For example, observations of oxygen in 2021 and 2022 followed different patterns than in previous years (Fig. A.11), potentially leading to worse model performance, while temperature observations showed similar patterns to historical years (Fig. A.15), potentially leading to better predictions.

4.4. Differences in forecast performance metrics across years

Forecast skill integrates both FLARE forecast accuracy and null (persistence and climatology) forecast accuracy (Jolliffe and Stephenson, 2012). As such, differences in null forecast 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 (Fig. A.8b), but climatology forecasts had low accuracy (Fig. A.7), due to 0.6 mg/L lower concentrations of oxygen in 2022 relative to the historical 2007–2022 average (Fig. A.14). In contrast, both FLARE and climatology forecasts of oxygen at 1.0 m had relatively high accuracy in 2021 (Fig. A.8 and Fig. A.7), but climatology forecasts were more accurate, resulting in a large proportion of unskillful FLARE forecasts 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, where historical estimates may no longer be as good at predicting the future (Daugaard et al., 2022; Pauly, 1995). Given the widespread impact of climate change on ecosystems globally (Bruggemann et al., 2012; IPCC, 2023; Mariani et al., 2022; Sydeman et al., 2013), evaluating forecast skill relative to uninformed models provides an opportunity to robustly assess how much more information process-based forecasts provide.

By comparing a novel forecasting approach to multiple null forecasts, we can disentangle what types of information these forecasts provide above each type of null. For example, persistence null forecasts directly represent the autocorrelation in observed dynamics and climatology null forecasts represent expected seasonal patterns in observed dynamics. As a result, in this study, when our process-based FLARE forecasts were more skillful than the persistence forecast, as in a large proportion of oxygen forecasts at 1.0 m (Fig. 6b), we can infer that FLARE forecasts are providing information about seasonal expectations and/or process representation above the climatology forecast. This is likely because process-based models such as FLARE can represent the entire water column and simulate multiple processes that control interactions between state variables (e.g., temperature and oxygen) in each water column layer (e.g., Hipsey et al., 2019; Li et al., 2022). In contrast, forecasts which were more skillful than a climatology forecast, such as the temperature and oxygen forecasts at 10.0 m, provide useful information about the role of autocorrelation or process representation in observed dynamics. Forecasts which were more skillful than both persistence and climatology null forecasts indicate that FLARE process-based forecasts were directly incorporating information about process representation which was not included in either null forecast. This was evident in a majority of temperature forecasts at both depths, and ~ 25 % of oxygen forecasts at 1.0 m. However, oxygen forecasts at 10.0 m were largely less skillful than either null forecast, indicating that dynamics of this variable were primarily dominated by autocorrelation and expected seasonal patterns, which were relatively accurate at this depth (mean persistence CRPS = 0.4 mg/L, climatology CRPS = 0.5 mg/L).

Overall, quantifying both the accuracy and skill of forecasts can contribute complementary information for decision-making. Decisions regarding specific details about an ecosystem variable likely require information provided by forecast accuracy. For example, 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 alternate information which is more relevant for understanding how ecosystem predictability has changed over time or performance across forecast models. For example, how much better can we predict algal concentrations over the historical mean on a given day? Or, which forecast model is more accurate in a given year or at a specific ecosystem?. In addition, unitless skill scores allow for a quantitative comparison of forecast variables which are not in the same native units, which 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. Overall, this study emphasizes the importance of using both of these forecast performance metrics, depending on decision-making needs.

4.5. Opportunities for expanding our understanding of water quality predictability

Results of this study point to important gaps in our understanding of predictability of freshwater ecosystems. First, oxygen dynamics are driven by both abiotic and biotic processes (Marce et al., 2023) motivating the need for future studies to use more complex oxygen models for forecasting. While hydrodynamic models do well at predicting abiotic dynamics, predicting many biogeochemical processes remains a challenge (Soares and do Calijuri, 2021). Lake Sunapee is an oligotrophic lake with low annual NEP (net ecosystem production) that ranges from −1 to <1 mg O^2 L^{-1} day^{-1} (Richardson et al., 2017), and low overall phytoplankton levels (historical mean chlorophyll-a = 1.7 μg/L; Ward et al., 2020), justifying our use of an oxygen model that did not include phytoplankton. Looking ahead, future studies generating oxygen forecasts for lakes with greater productivity would benefit from using models that represent critical biotic processes governing oxygen dynamics (e.g., epilimnetic primary production, respiration, and decomposition). 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, especially for forecast variables which show high autocorrelation, as high-frequency sensors may be especially critical for informing these model states. Third, while process-based models are well-positioned for including more 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 while simultaneously fitting models to multiple variables.

4.6. Conclusions

Overall, we show that forecast performance of water temperature and dissolved oxygen was variable between years, demonstrating that predictability of important water quality variables can vary substantially over time in a large oligotrophic lake. This variability was notable despite similar average weather patterns between years, indicating that specific events such as storms or other factors are likely influencing predictability. Across variables, temperature forecasts were more skillful than oxygen forecasts in Lake Sunapee, following expectations that physical variables are more predictable than chemical variables like oxygen, which are influenced by nonlinear dynamics and feedbacks from biogeochemical processes, even in an oligotrophic lake. Finally, FLARE forecast performance varied between metrics of forecast accuracy (e.g., CRPS) and forecast skill (relative to a null model). Specifically, we found low FLARE accuracy (e.g., CRPS) in some cases but high skill when compared to null persistence or climatology forecasts. This finding emphasizes that process-based forecasts provide important information above null model forecasts, and may result in increasing value of process-based forecasts 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 performance varies over time and depth, especially as freshwater ecosystems continue to experience global change stressors.

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.

Funding

This work was supported by the Calhoun LSPA-Virginia Tech Fellowship, National Science Foundation (NSF) research grants DBI-1933016, EF-1926050, EF-2318861, and an NSF Graduate Research

Fellowship.

CRediT authorship contribution statement

Whitney M. Woelmer: Writing – original draft, Visualization, Project administration, Investigation, Formal analysis, Conceptualization. **R. Quinn Thomas:** Writing – review & editing, Software, Methodology. **Freya Olsson:** Writing – review & editing, Methodology, Investigation. **Bethel G. Steele:** Writing – review & editing, Data curation. **Kathleen C. Weathers:** Writing – review & editing, Data curation. **Cayelan C. Carey:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Data availability

All forecast output and required input data (water quality observational data and meteorological driver data) are available at Woelmer et al. (2023; https://doi.org/10.5281/zenodo.10127798). All code and forecast configuration files to reproduce this analysis are available at Woelmer et al. (2024; https://doi.org/10.5281/zenodo.12694838).

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 fieldwork, driving, and diving which enabled this research. We thank the CIBR FLARE Research team and Virginia Tech Reservoir Group, 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. Many thanks to Erin Hotchkiss and Paul Hanson for helpful reviews and feedback which substantially improved the quality of this work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecoinf.2024.102825.

References

Bhateria, R., Jain, D., 2016. Water quality assessment of lake water: a review. Sustain. Water Resour. Manage. 2, 161–173. https://doi.org/10.1007/s40899-015-0014-7.
Bodner, K., 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. Q. J. R. Meteorol. Soc. 138, 1611–1617. https://doi.org/10.1002/qj.1891.
Bröcker, J., Smith, L.A., 2007. Increasing the reliability of reliability diagrams. Weather Forecast. 22 (3), 651–661. https://doi.org/10.1175/waf993.1.
Bruesewitz, D.A., et al., 2015. Under-ice thermal stratification dynamics of a large, deep lake revealed by high-frequency data. Limnol. Oceanogr. 60, 347–359. https://doi.org/10.1002/lno.10014.
Bruggemann, J.H., et al., 2012. Wicked social-ecological problems forcing unprecedented change on the latitudinal margins of coral reefs: the case of Southwest Madagascar. Ecol. Soc. 17 (4) https://doi.org/10.5751/ES-05300-170447.
Caffrey, J.M., 2004. Factors controlling net ecosystem metabolism in U.S. estuaries. Estuaries 27, 90–101. https://doi.org/10.1007/BF02803563.
Calamita, E., et al., 2021. Lake modeling reveals management opportunities for improving water quality downstream of transboundary tropical dams. Water Resour. Res. 57, e2020WR027465 https://doi.org/10.1029/2020WR027465.
Carey, C., 2023. Causes and consequences of changing oxygen availability in lakes. Inland Waters. https://doi.org/10.1080/20442041.2023.2239110.
Carey, C.C., et al., 2014. Experimental blooms of the cyanobacterium *Gloeotrichia echinulata* increase phytoplankton biomass, richness and diversity in an oligotrophic lake. J. Plankton Res. 36, 364–377. https://doi.org/10.1093/plankt/fbt105.
Carey, C.C., 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.
Cuddington, K., et al., 2013. Process-based models are required to manage ecological systems in a changing world. Ecosphere 4 (2), 20. https://doi.org/10.1890/ES12-00178.1.
Daugaard, U., et al., 2022. Forecasting in the face of ecological complexity: number and strength of species interactions determine forecast skill in ecological communities. Ecol. Lett. 25, 1974–1985. https://doi.org/10.1111/ele.14070.
Davis, J.C., 1975. Minimal dissolved oxygen requirements of aquatic life with emphasis on Canadian species: a review. J. Fish. Res. Board Can. 32, 2295–2332. https://doi.org/10.1139/f75-268.
Dietze, M.C., 2017. Ecological Forecasting. Princeton University Press, Princeton.
Dietze, M.C., et al., 2018. Iterative near-term ecological forecasting: needs, opportunities, and challenges. PNAS 115, 1424–1432. https://doi.org/10.1073/pnas.1710231115.
dos Simões, F.S., et al., 2008. Water quality index as a simple indicator of aquaculture effects on aquatic bodies. Ecol. Indic. 8, 476–484. https://doi.org/10.1016/j.ecolind.2007.05.002.
Durell, L., et al., 2022. Functional forecasting of dissolved oxygen in high-frequency vertical lake profiles. Environmetrics 1–16. https://doi.org/10.1002/env.2765.
Durell, L., Scott, J.T., Hering, A.S., 2023. Hybrid forecasting for functional time series of dissolved oxygen profiles. Data Sci. Sci. 2, 1. https://doi.org/10.1080/26941899.2022.2152401.
Evensen, G., 2009. Data Assimilation: The Ensemble Kalman Filter. Springer, Berlin.
Geng, M., 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. Environ. Sci. Pollut. Res. 29, 14083–14097. https://doi.org/10.1007/s11356-021-16739-5.
Gneiting, T., et al., 2005. Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. Mon. Weather Rev. 133, 1098–1118. https://doi.org/10.1175/MWR2904.1.
Gneiting, T., Balabdaoui, F., Raftery, A.E., 2007. Probabilistic forecasts, calibration and sharpness. J. R. Stat. Soc. Ser. B Stat Methodol. 69, 243–268.
Hamill, T.M., et al., 2022. The reanalysis for the global ensemble forecast system, version 12. Mon. Weather Rev. 150 (1), 59–79. https://doi.org/10.1175/MWR-D-21-0023.1.
Harris, D.J., Taylor, S.D., White, E.P., 2018. Forecasting biodiversity in breeding birds using best practices. PeerJ 2018, 1–27. https://doi.org/10.7717/peerj.4278.
Henden, J.A., et al., 2020. End-user involvement to improve predictions and management of populations with complex dynamics and multiple drivers. Ecol. Appl. 0, 1–14. https://doi.org/10.1002/eap.2120.
Hipsey, M.R., 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). Geosci. Model Dev. Katlenburg-Lindau 12, 473–523. https://doi.org/10.5194/gmd-12-473-2019.
Hipsey, M.R., et al., 2020. A system of metrics for the assessment and improvement of aquatic ecosystem models. Environ. Model Softw. 128, 104697 https://doi.org/10.1016/j.envsoft.2020.104697.
Hipsey, M.R., Boon, C., Bruce, L.C., et al., 2022. AquaticEcoDynamics/glm-aed: v3.3.0.
Ho, J.C., Michalak, A.M., 2019. Exploring temperature and precipitation impacts on harmful algal blooms across continental U.S. lakes. Limnol. Oceanogr. 1–18. https://doi.org/10.1002/lno.11365.
Hyndman, R.J., 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, L.A., et al., 2022a. Opportunities for seasonal forecasting to support water management outside the tropics. Hydrol. Earth Syst. Sci. 26, 1389–1406. https://doi.org/10.5194/hess-26-1389-2022.
Jackson-Blake, L.A., et al., 2022b. Seasonal forecasting of lake water quality and algal bloom risk using a continuous Gaussian Bayesian network. Hydrol. Earth Syst. Sci. 26, 3103–3124. https://doi.org/10.5194/hess-26-3103-2022.
Jassby, A.D., Reuter, J.E., Goldman, C.R., 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., et al., 2018. Physics guided recurrent neural networks for modeling dynamical systems: application to monitoring water temperature and quality in lakes. arXiv. https://doi.org/10.48550/arXiv.1810.02880. Preprint.
Jin, T., et al., 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, I.T., Stephenson, D.B. (Eds.), 2012. Forecast Verification: A practitioner's Guide in Atmospheric Science, 2. Wiley-Blackwell, Oxford.
Jones, I.D., Smol, J.P., 2023. Wetzel's Limnology, 4th edn. Academic Press, San Diego, USA. https://doi.org/10.1016/CS2019-0-04412-3.
Jones, I.D., Winfield, I.J., 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. Freshw. Biol. 53, 393–402. https://doi.org/10.1111/j.1365-2427.2007.01902.x.
Jordan, A., et al., 2019. Evaluating probabilistic forecasts with scoring rules. J. Stat. Softw. 90 (12), 1–37. https://doi.org/10.18637/jss.v090.i12.

Kim, S.K., Choi, S.-U., 2021. Assessment of the impact of selective withdrawal on downstream fish habitats using a coupled hydrodynamic and habitat modeling. J. Hydrol. 593, 125665 https://doi.org/10.1016/j.jhydrol.2020.125665.
Kraemer, B.M., et al., 2021. Climate change drives widespread shifts in lake thermal habitat. Nat. Clim. Chang. 11, 521–529. https://doi.org/10.1038/s41558-021-01060-3.
Ladwig, R., et al., 2021. Lake thermal structure drives interannual variability in summer anoxia dynamics in a eutrophic lake over 37 years. Hydrol. Earth Syst. Sci. 25, 1009–1032. https://doi.org/10.5194/hess-25-1009-2021.
Lazer, D., et al., 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. Environ. Eng. Res. 20, 205–211. https://doi.org/10.4491/eer.2015.051.
Lee, D.-Y., et al., 2023. Data-driven models for predicting community changes in freshwater ecosystems: a review. Eco. Inform. 77, 102163 https://doi.org/10.1016/j.ecoinf.2023.102163.
Lewis, A.S.L., et al., 2022. Increased adoption of best practices in ecological forecasting enables comparisons of forecastability. Ecol. Appl. 32 (2), e2500 https://doi.org/10.1002/eap.2500.
Lewis, A.S.L., et al., 2023. The power of forecasts to advance ecological theory. Methods Ecol. Evol. 14 (3), 746–756. https://doi.org/10.1111/2041-210X.13955.
Li, M., et al., 2022. Recent advances in application of iron-manganese oxide nanomaterials for removal of heavy metals in the aquatic environment. Sci. Total Environ. 819, 153157 https://doi.org/10.1016/j.scitotenv.2022.153157.
Lin, S., 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, M.E., et al., 2023. Progress and opportunities in advancing near-term forecasting of freshwater quality. Glob. Chang. Biol. 29, 1691–1714. https://doi.org/10.1111/gcb.16590.
LSPA, 2023. Bathymetric data for Lake Sunapee, NH, USA ver 1. Environ. Data Initiative. https://doi.org/10.6073/pasta/8a86710f25adeae3a2540b6cb14d7546.
LSPA, Sunapee, T., 2022. Ice-off dates for Lake Sunapee, NH, USA, 1869-2022 ver 1. Environ. Data Initiative. https://doi.org/10.6073/pasta/3c60c873c18c2d4811a084831b3f375a (Accessed 2023-11-29).
LSPA, Steele, B.G., Weathers, K.C., 2023. Lake Sunapee Instrumented Buoy: High Frequency Water Quality Data - 2007-2022 ver 4. Environmental Data Initiative. https://doi.org/10.6073/pasta/8a86710f25adeae3a2540b6cb14d7546.
Luo, Y., et al., 2009. Parameter identifiability, constraint, and equifinality in data assimilation with ecosystem models. Ecol. Appl. 19, 571–574. http://www.jstor.org/stable/27645995.
Magee, M.R., et al., 2019. Drivers and management implications of long-term Cisco Oxythermal habitat decline in Lake Mendota, WI. Environ. Manag. 63, 396–407. https://doi.org/10.1007/s00267-018-01134-7.
Marce, R., et al., 2023. Chapter 11 – Oxygen. In: Jones, Smol (Eds.), Wetzel's Limnology, 4th ed. Academic Press, pp. 237–274.
Mariani, M., et al., 2022. Disruption of cultural burning promotes shrub encroachment and unprecedented wildfires. Front. Ecol. Environ. 20, 292–300. https://doi.org/10.1002/fee.2395.
Mercado-Bettín, D., et al., 2021. Forecasting water temperature in lakes and reservoirs using seasonal climate prediction. Water Res. 201, 117286 https://doi.org/10.1016/j.watres.2021.117286.
Mittermaier, M.P., 2008. The potential impact of using persistence as a reference forecast on perceived forecast skill. Weather Forecast. 23, 1022–1031. https://doi.org/10.1175/2008WAF2007037.1.
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.
O'Hara-Wild, M., Hyndman, R., Wang, E., 2022. Fable: Forecasting Models for Tidy Time Series. R package version 0.3.2. Retrieved from. https://cran.r-project.org/package=fable.
Olsson, F., Moore, T.N., Carey, C.C., Breef-Pilz, A., Thomas, R.Q., 2024. A multi-model ensemble of baseline and process-based models improves the predictive skill of near-term Lake forecasts. Water Resour. Res. 60 (3), e2023WR035901.
Page, T., et al., 2018. Adaptive forecasting of phytoplankton communities. Water Res. 134, 74–85. https://doi.org/10.1016/j.watres.2018.01.046.
Pappenberger, F., et al., 2015. How do I know if my forecasts are better? Using benchmarks in hydrological ensemble prediction. J. Hydrol. 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 Ecol. Evol. 10, 430. https://doi.org/10.1016/S0169-5347(00)89171-5.
Peng, Z., et al., 2019. Development and evaluation of a real-time forecasting framework for daily water quality forecasts for Lake Chaohu to Lead time of six days. Sci. Total Environ. 687, 218–231. https://doi.org/10.1016/j.scitotenv.2019.06.067.
Peters, D.P.C., 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, V.C., et al., 2015. The rise of novelty in ecosystems. Ecol. Appl. 25, 2051–2068. https://doi.org/10.1890/14-1781.1.
Richardson, D.C., et al., 2017. Intra- and inter-annual variability in metabolism in an oligotrophic lake. Aquat. Sci. 79, 319–333. https://doi.org/10.1007/s00027-016-0499-7.
Saber, A., James, D.E., Hayes, D.F., 2020. Long-term forecast of water temperature and dissolved oxygen profiles in deep lakes using artificial neural networks conjugated with wavelet transform. Limnol. Oceanogr. 65, 1297–1317. https://doi.org/10.1002/lno.11390.
Sánchez, E., et al., 2007. Use of the water quality index and dissolved oxygen deficit as simple indicators of watersheds pollution. Ecol. Indic. 7, 315–328. https://doi.org/10.1016/j.ecolind.2006.02.005.
Schepen, A., Zhao, T., Wang, Q.J., Zhou, S., Feikema, P., 2016. Optimising seasonal streamflow forecast lead time for operational decision making in Australia. Hydrol. Earth Syst. Sci. 20, 4117–4128.
Sharma, S., et al., 2021. Loss of ice cover, shifting phenology, and more extreme events in northern Hemisphere Lakes. J. Geophys. Res. Biogeosci. 126, e2021JG006348 https://doi.org/10.1029/2021JG006348.
Soares, L.M.V., do Calijuri, M.C., 2021. Deterministic modelling of freshwater lakes and reservoirs: current trends and recent progress. Environ. Model Softw. 144, 105143 https://doi.org/10.1016/j.envsoft.2021.105143.
Solomon, C.T., et al., 2013. Ecosystem respiration: drivers of daily variability and background respiration in lakes around the globe. Limnol. Oceanogr. 58, 849–866.
Sondergaard, M., Jensen, P.J., Jeppesen, E., 2001. Retention and internal loading of phosphorus in shallow, eutrophic lakes. Sci. World J. 1, 427–442. https://doi.org/10.1100/tsw.2001.72.
Staehr, P.A., 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, E.H., et al., 2019. Biases in lake water quality sampling and implications for macroscale research. Limnol. Oceanogr. 64, 1572–1585. https://doi.org/10.1002/lno.11136.
Steele, B.G., Weathers, K.C., Association LSPA, 2023. Lake-Sunapee-Protective-Association/LMP: LSPA LMP database 1986-2022 (v2) (v2023.2). Data set Zenodo. https://doi.org/10.5281/zenodo.8003784.
Sydeman, W.J., et al., 2013. Increasing variance in North Pacific climate relates to unprecedented ecosystem variability off California. Glob. Chang. Biol. 19, 1662–1675. https://doi.org/10.1111/gcb.12165.
Thomas, R.Q., et al., 2020. A near-term iterative forecasting system successfully predicts reservoir hydrodynamics and partitions uncertainty in real time. Water Resour. Res. 56, e2019WR026138 https://doi.org/10.1029/2019WR026138.
Thomas, R.Q., et al., 2023. Near-term forecasts of NEON lakes reveal gradients of environmental predictability across the US. Front. Ecol. Environ. 21, 220–226. https://doi.org/10.1002/fee.2623.
Wander, H.L., et al., 2023. Data assimilation experiments inform monitoring needs for near-term ecological forecasts in a eutrophic reservoir. In: ESS Open Archive Preprint. https://doi.org/10.22541/essoar.168500255.59108131/v1.
Ward, N.K., et al., 2020. Differential responses of maximum versus median chlorophyll-a to air temperature and nutrient loads in an oligotrophic lake over 31 years. Water Resour. Res. 56 (7), e2020WR027296.
Wheatcroft, E., 2019. Interpreting the skill score form of forecast performance metrics. Int. J. Forecast. 35, 573–579. https://doi.org/10.1016/j.ijforecast.2018.11.010.
Wheeler, K., 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, W.M., et al., 2022. Near-term phytoplankton forecasts reveal the effects of model time step and forecast horizon on predictability. Ecol. Appl. 32, 1–22. https://doi.org/10.1002/eap.2642.
Woelmer, W.M., et al., 2023. Forecasts, score summary files, target observational data, and meteorological driver files to accompany the manuscript "process-based forecasts of lake water temperature and dissolved oxygen outperform null models, with variability over time and depth" [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10127798.
Woelmer, W.M., et al., 2024. Wwoelmer/SUNP_fcasts_temp_DO_MS: EcoInformatics Resubmission July 2024 (v1.2). Zenodo. https://doi.org/10.5281/zenodo.12694838.
Woolway, R.I., Merchant, C.J., 2019. Worldwide alteration of lake mixing regimes in response to climate change. Nat. Geosci. 12, 271–276. https://doi.org/10.1038/s41561-019-0322-x.
Woolway, R.I., et al., 2019. Substantial increase in minimum lake surface temperatures under climate change. Clim. Chang. 155, 81–94. https://doi.org/10.1007/s10584-019-02465-y.
Woolway, R.I., et al., 2021. Phenological shifts in lake stratification under climate change. Nat. Commun. 12, 2318. https://doi.org/10.1038/s41467-021-22657-4.
Wynne, J.H., 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., et al., 2022. Eco-Environment & Health a review of the application of machine learning in water quality evaluation. Eco-Environ. Health 1, 107–116. https://doi.org/10.1016/j.eehl.2022.06.001.