Water Resources Research

RESEARCH ARTICLE
10.1029/2019WR026138

A Near-Term Iterative Forecasting System Successfully Predicts Reservoir Hydrodynamics and Partitions Uncertainty in Real Time

R. Quinn Thomas[1], Renato J. Figueiredo[2], Vahid Daneshmand[2], Bethany J. Bookout[3], Laura K. Puckett[1], and Cayelan C. Carey[3]

[1]Department of Forest Resources & Environmental Conservation, Virginia Tech, Blacksburg, VA, USA, [2]Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA, [3]Department of Biological Sciences, Virginia Tech, Blacksburg, VA, USA

Key Points:
- We created a real-time iterative lake water temperature forecasting system that uses sensors, data assimilation, and hydrodynamic modeling
- Our water temperature forecasting system quantifies uncertainty in each daily forecast and is open source
- Sixteen-day future forecasted temperatures were within 1.4°C of observations over 16 months in a reservoir case study

Supporting Information:
- Supporting Information S1
- Supporting Information S2
- Supporting Information S3
- Supporting Information S4
- Supporting Information Table S1
- Supporting Information Table S2
- Supporting Information Table S3
- Supporting Information Figure S1
- Supporting Information Figure S2

Correspondence to:
R. Q. Thomas,
rqthomas@vt.edu

Citation:
Thomas, R. Q., Figueiredo, R. J., Daneshmand, V., Bookout, B. J., Puckett, L. K., & Carey, C. C. (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

Received 10 AUG 2019
Accepted 27 AUG 2020
Accepted article online 1 SEP 2020

Author Contributions:
Conceptualization: R. Quinn Thomas, Renato J. Figueiredo, Cayelan C. Carey
Data curation: R. Quinn Thomas, Vahid Daneshmand, Cayelan C. Carey
(continued)

©2020. The Authors.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Abstract Freshwater ecosystems are experiencing greater variability due to human activities, necessitating new tools to anticipate future water quality. In response, we developed and deployed a real-time iterative water temperature forecasting system (FLARE—Forecasting Lake And Reservoir Ecosystems). FLARE is composed of water temperature and meteorology sensors that wirelessly stream data, a data assimilation algorithm that uses sensor observations to update predictions from a hydrodynamic model and calibrate model parameters, and an ensemble-based forecasting algorithm to generate forecasts that include uncertainty. Importantly, FLARE quantifies the contribution of different sources of uncertainty (driver data, initial conditions, model process, and parameters) to each daily forecast of water temperature at multiple depths. We applied FLARE to Falling Creek Reservoir (Vinton, Virginia, USA), a drinking water supply, during a 475-day period encompassing stratified and mixed thermal conditions. Aggregated across this period, root mean square error (RMSE) of daily forecasted water temperatures was 1.13°C at the reservoir's near-surface (1.0 m) for 7-day ahead forecasts and 1.62°C for 16-day ahead forecasts. The RMSE of forecasted water temperatures at the near-sediments (8.0 m) was 0.87°C for 7-day forecasts and 1.20°C for 16-day forecasts. FLARE successfully predicted the onset of fall turnover 4–14 days in advance in two sequential years. Uncertainty partitioning identified meteorology driver data as the dominant source of uncertainty in forecasts for most depths and thermal conditions, except for the near-sediments in summer, when model process uncertainty dominated. Overall, FLARE provides an open-source system for lake and reservoir water quality forecasting to improve real-time management.

1. Introduction

As a result of human activities, ecosystems around the globe are increasingly changing (Stocker et al., 2013; Ummenhofer & Meehl, 2017), making it challenging for resource managers to consistently provide vital ecosystem services (West et al., 2009). In particular, managers of freshwater ecosystems, which have been more degraded than any other ecosystem on the planet (Millennium Ecosystem Assessment, 2005), are seeking new tools to anticipate future change and ensure clean water for drinking, fisheries, irrigation, industry, and recreation (Brookes et al., 2014).

In response to this need, near-term, real-time iterative ecological forecasting has emerged as a solution to provide stakeholders, managers, and policy-makers crucial information about future ecosystem conditions (Clark et al., 2001; Dietze et al., 2018; Luo et al., 2011). Here, we define a real-time iterative forecast as a prediction of future ecosystem states with quantified uncertainty, generated from models that can be continually updated with new data as they become available (adapted from Clark et al., 2001; Dietze, 2017a, 2018; Luo et al., 2011). Multiple statistical approaches, such as Bayesian state-space modeling, particle filters, and ensemble filters, are used to estimate and propagate the different sources of uncertainty that contribute to the total uncertainty in a forecast (e.g., driver data, initial conditions, model process, and parameters; Clark et al., 2008; Dietze, 2017a; Ouellet-Proulx, Chimi Chiadjeu, et al., 2017). Fully specifying all of these uncertainty sources provides both an assessment of confidence in a forecast for managers as they interpret the forecasts for decision-making, and valuable information for researchers about how to improve forecasts (Berthet et al., 2016; Dietze, 2017a; Morss et al., 2008).

Formal analysis: R. Quinn Thomas, Laura K. Puckett
Funding acquisition: R. Quinn Thomas, Renato J. Figueiredo, Cayelan C. Carey
Investigation: Vahid Daneshmand, Bethany J. Bookout
Methodology: R. Quinn Thomas, Renato J. Figueiredo, Cayelan C. Carey
Project administration: Cayelan C. Carey
Resources: Renato J. Figueiredo, Cayelan C. Carey
Software: R. Quinn Thomas, Renato J. Figueiredo, Vahid Daneshmand, Laura K. Puckett
Supervision: R. Quinn Thomas, Renato J. Figueiredo, Cayelan C. Carey
Writing - original draft: R. Quinn Thomas, Cayelan C. Carey
Writing – review & editing: R. Quinn Thomas, Renato J. Figueiredo, Vahid Daneshmand, Bethany J. Bookout, Laura K. Puckett, Cayelan C. Carey

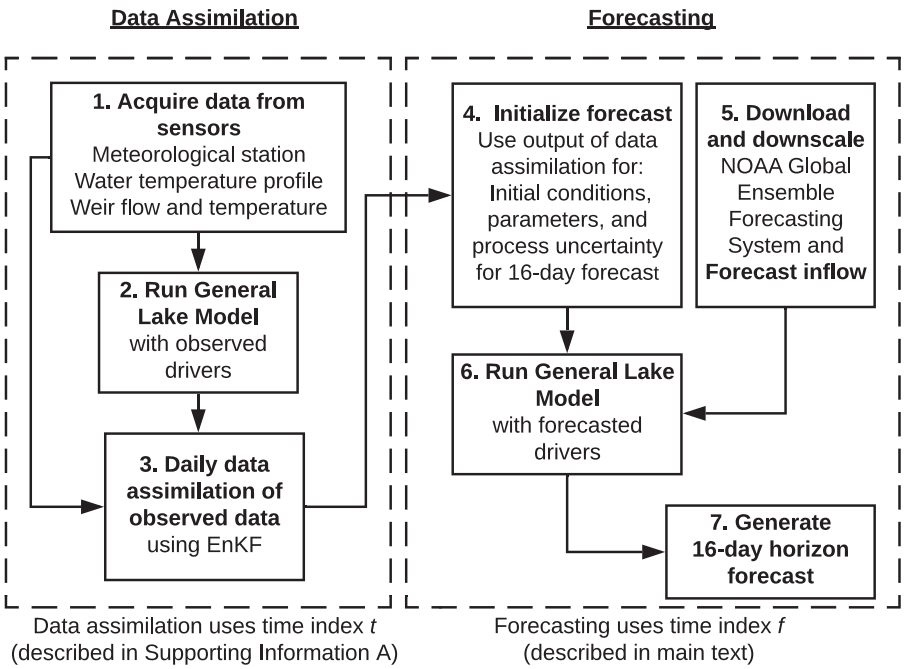

Figure 1. A simplified conceptual diagram describing the major data assimilation and forecasting steps in the daily iterative forecasting cycle for the FLARE (Forecasting Lake And Reservoir Ecosystems) forecasting system. The forecasting cycle has two key stages in which data are assimilated each day (left dashed box) to produce the initial conditions for a 16-day forecast (right dashed box). EnKF = ensemble Kalman filter; NOAA = National Oceanic and Atmospheric Administration.

Real-time forecasts of water temperature with fully specified uncertainties are particularly valuable for managers that oversee drinking water supply lakes and reservoirs, as waterbody temperatures can be very dynamic due to meteorological forcing, management, and seasonality (e.g., Klug et al., 2012; Mi et al., 2019; Schmidt et al., 2018; Sharma et al., 2015). Water temperature is closely related to many water quality metrics and defines the physical conditions for multiple ecological processes, including microbial and algal growth, dissolved oxygen saturation, the release of chemical constituents from sediments into the water column, and habitat suitability for organisms, such as fish (Butcher et al., 2015; Carey et al., 2012; Delpla et al., 2009; Jöhnk et al., 2008). As a result, depth profiles of temperature data are used to determine withdrawal depths for water treatment, schedule dam releases for optimizing the temperature of downstream flows, and manage lake and reservoir water quality (Çalışkan & Elçi, 2008; Casamitjana et al., 2003; Hague & Patterson, 2014; Huang et al., 2011; Pike et al., 2013; Weber et al., 2017).

Water temperature depth profiles also determine the strength of thermal stratification, i.e., if there are discrete epilimnetic (surface) and hypolimnetic (bottom) layers or isothermal (fully mixed) conditions (Read et al., 2011). When waterbodies transition from stratified to mixed conditions during the onset of fall turnover, nutrients and metals that have accumulated in the hypolimnion during the summer are mixed throughout the water column, decreasing water quality (Cooke et al., 2005; Effler & Matthews, 2008). The strength of thermal stratification in reservoirs can also modify the response of downstream rivers to climate change (Cheng et al., 2020). Consequently, real-time iterative forecasts of water temperature profiles would allow managers to preemptively respond to impending poor water quality during fall turnover and other episodic events (e.g., storms) that alter water temperature and thermal stratification.

Here, we introduce a forecasting system (FLARE, Forecasting Lake And Reservoir Ecosystems) that generates automated 16-day water temperature forecasts in real time (Figure 1).

FLARE is composed of (1) water temperature and meteorology sensors deployed in a lake or reservoir that wirelessly stream data, (2) a data assimilation algorithm that uses sensor observations to update water temperature predictions from a hydrodynamic model and to calibrate model parameters, and (3) an ensemble-

based forecasting algorithm to generate forecasts that quantify the sources of forecast uncertainty. FLARE quantifies uncertainty from driver data (i.e., the uncertainty in future weather or inflow forecasts that are needed to run the hydrodynamic model), initial conditions (i.e., the uncertainty observed in water temperatures on the first day of the forecast), model process (i.e., the capacity of a calibrated model to reproduce observations), and model parameters (following Dietze, 2017a). The forecasting system samples from these sources of uncertainty to generate probability distributions for water temperature at multiple depths and can generate probability distributions of hydrodynamic events such as the occurrence of fall turnover.

FLARE builds off of previous studies that have developed forecasting systems with data assimilation to predict water temperature (e.g., Bal et al., 2014; Baracchini, Wuest, & Bouffard, 2020; Caissie et al., 2016; Hague & Patterson, 2014; Huang et al., 2011; Mestekemper et al., 2010; Ouellet-Proulx, Chimi Chiadjeu, et al., 2017; Pike et al., 2013) in four key ways. First, FLARE was built using all open-source software and model components and is completely automated via its cyberinfrastructure: once set up, the forecasting system automatically retrieves real-time observations for data assimilation, runs models, and triggers the delivery of daily forecasts and system status updates. Second, unlike many real-time water forecasting systems that depend on governmental agencies for the collection of data used for assimilation (e.g., Ouellet-Proulx, Chimi Chiadjeu, et al., 2017; Pike et al., 2013), FLARE forecasts can be run and updated with water temperature data commonly collected by local water managers. These features enable scalability to a range of waterbodies without existing infrastructure. Third, FLARE daily forecasts extend up to 16 days in the future, which allows for the evaluation of forecast performance at a longer forecast horizon than reported in the literature by other water temperature forecasting systems, which have forecast horizons ranging from 3 days (Caissie et al., 2016) to 10 days (Hague & Patterson, 2014). Fourth, FLARE propagates and partitions multiple sources of forecast uncertainty, which are updated over time, allowing for the evaluation of both forecast accuracy and the reliability of uncertainty estimation. Despite the importance of quantifying multiple uncertainty sources, few water resource forecasting studies quantify more than one or two sources of uncertainty and when they do, they typically only include initial conditions uncertainty (via data assimilation) and meteorological uncertainty (via ensemble weather forecasts) (e.g., Baracchini, Wuest, & Bouffard, 2020; Komatsu et al., 2007; Ouellet-Proulx, Chimi Chiadjeu, et al., 2017; Ouellet-Proulx, St-Hilaire, & Boucher, 2017; Page et al., 2018). Furthermore, they rarely partition the relative contributions of the individual sources of uncertainty to the total forecast uncertainty (but see Ouellet-Proulx, Chimi Chiadjeu, et al., 2017).

We set up FLARE to generate automated, probabilistic water temperature forecasts for a drinking water reservoir over 16 months (475 days) to address the following questions: (1) How does forecasting performance, as evaluated against observations and a null persistence model, differ between thermally stratified versus mixed conditions, two key regimes of lake thermal dynamics? (2) How well does the forecasting system predict the onset of fall turnover? and (3) What are the contributions of different sources of uncertainty to the forecasts, and how do they vary between thermally stratified versus mixed conditions? Based on root mean square error (RMSE) values reported from non-forecasting applications of lake hydrodynamic models (e.g., Bruce et al., 2018), we expected model process to be an important source of forecast uncertainty (Hipsey et al., 2020). Following Dietze (2017b), we expected meteorological driver data to also be an important contributor of forecast uncertainty, which would increase throughout the 16-day forecast horizon.

2. Methods

We developed a forecasting system (FLARE) that predicts water temperature at a daily time step for a set of specified depths in a lake or reservoir over a 16-day time horizon (Figure 1). The system uses data assimilation of sensor observations to generate both (1) initial conditions and parameters for a physics-based hydrodynamic model and (2) uncertainty estimates for a forecast that extends 16 days in the future. Each day, the data assimilation is advanced one daily time step before launching the forecast of future conditions. We describe the forecasting methods below and the data assimilation methods in Supporting Information S1.

2.1. Hydrodynamic Model

FLARE simulated reservoir hydrodynamics with the General Lake Model (GLM), a one-dimensional (1-D) vertical stratification model (Hipsey et al., 2019). We used GLM because (1) the model has successfully reproduced observed water temperature profiles in lakes around the world with varying mixing regime, climate, and morphology (Bruce et al., 2018); (2) GLM is an open-source, community-developed model and thus

scalable to other waterbodies; and (3) GLM has low computational needs, enabling many model ensemble members to be run quickly and efficiently, a requirement for real-time iterative forecasting. GLM runs at a daily time step with subdaily (hourly) meteorological drivers and daily mean inflow drivers. In FLARE, we used the GLM output for the hour at the end of the daily step (e.g., 12:00:00).

We set all but three highly sensitive parameters equal to the values reported in the default GLM version 3.1.0a model (Hipsey et al., 2019), and FLARE updated the highly sensitive parameters at each model time step. The three parameters that had automated tuning using data assimilation were selected using a sensitivity analysis described in Supporting Information S1 and were: one parameter that scales the incoming shortwave radiation (sw_factor) and two parameters defining the sediment temperatures in the hypolimnion (zone1temp, 5–9.3 m) and epilimnion (zone2temp: 0–5 m). For driver data, GLM requires hourly meteorological data on downwelling shortwave radiation (W m^{-2}), downwelling longwave radiation (W m^{-2}), air temperature (°C), wind speed (m s^{-1}), relative humidity (%), and precipitation (m day^{-1}) as well as daily rates of water inflow (m^3 day^{-1}) and inflow temperature (°C) and daily rates of outflow (m^3 day^{-1}). The full GLM configuration is provided in Supporting Information S2.

2.2. Ensemble Forecasting Approach

The FLARE system uses an ensemble approach to numerically simulate and propagate forecast uncertainty into the future. The ensemble forecasting is based on Equation 1:

$$x^i_{f+1} = G\left(x^i_f, \alpha^i_t, D^i_f\right) + \varepsilon \tag{1}$$

where G() is the GLM model that requires a vector of water temperatures at the modeled depths as initial conditions (x^i_f), a vector of the three calibrated model parameters (α^i_t), and a set of model drivers (D, which encompassed weather, inflows, and outflows). The index i in Equation 1 represents the ith ensemble member and the subscript f is the day (0 to 16) in the 16-day forecast horizon. The index t is the number of days since the initiation of forecasting and data assimilation (i.e., 0 = the day the automated forecasting system is deployed). The index t is different from f because t increases each day throughout a multiyear forecasting period, while f is reset to 0 each day before initiating a new 16-day forecast. The ε represents the contribution of process uncertainty (i.e., errors due to an imperfect representation of physical processes in a model; Dietze, 2017a) to total forecast uncertainty and is randomly drawn from a multivariate normal distribution with a mean of 0 and a covariance matrix of Σ_t. The covariance matrix (Σ_t) represents the uncertainty at each depth (the diagonals of Σ_t) and the covariance of uncertainty between depths (the off-diagonals of Σ_t). Σ_t is calculated from the residuals between the ensemble mean and the observed temperatures before data are assimilated. The subscript t on Σ_t signifies that the covariance matrix is determined by data assimilation of historical observations and is constant over each day f in the 16-day forecast horizon. Equation 1 is applied at the daily time step and the process uncertainty is added each day of the forecast.

Each 16-day forecast requires initial conditions of the water temperature at each modeled depth ($x^i_{f=0} = x^i_t$, where t is the number of days since the automated forecasting system is deployed) and model parameters (α^i_t) on day $f = 0$ of the forecast. The variance in x_t and α_t across ensemble members when the forecast is initialized represents the contribution of initial conditions and parameters, respectively, to the total forecast uncertainty. To generate these initial conditions while also calibrating the three focal parameters, FLARE assimilates temperature sensor observations from the previous day into the GLM using an ensemble Kalman filter (EnKF; Evensen, 2003, 2009). We used data assimilation rather than simply specifying the initial conditions from the observations because the EnKF (1) enables the generation of initial conditions when sensor data were not available (e.g., during maintenance), (2) mechanistically interpolates water temperature between sensor depths, (3) enables the automated calibration of the three highly sensitive model parameters, and (4) generates a historical data product of water temperature with spatial and temporal gap filling. The EnKF method of data assimilation is well suited for nonlinear mechanistic models like the GLM and enables ensemble-based forecasts of future states (Baracchini, Chu, et al. 2020; Clark et al., 2008; Dietze, 2017a; Page et al., 2018). Our implementation of the EnKF with state augmentation to calibrate parameters (Supporting Information S1) follows Zhang et al. (2017).

To quantify the contribution of future meteorological conditions to the total forecast uncertainty, each of the i water temperature ensemble members is assigned one of the ensemble meteorological forecast members from the National Oceanic and Atmospheric Administration Global Ensemble Forecasting System (NOAA GEFS) throughout the 16-day forecast horizon. NOAA GEFS provides ensemble forecasts with 21 members, with each member representing slightly modified versions of the model that result in different predictions of future weather conditions at a 16-day time horizon. Each day that a forecast is generated, the meteorological driver data required by the GLM from NOAA GEFS are downloaded for the 12:00:00 UTC forecast using the rNOMADS package (Version 2.4.2) in R (Bowman & Lees, 2015).

Additionally, our forecasts include the uncertainty from statistically downscaling the gridded NOAA GEFS meteorological forecasts to local meteorological conditions. First, we downscaled the NOAA GEFS forecasts to a local site using a linear relationship between the NOAA GEFS forecast and observed meteorology. Next, we added random noise to each meteorological variable for each day of each NOAA GEFS ensemble member. This random noise was generated by a multivariate normal distribution that described the covariance in the relationship between the observed meteorology and NOAA GEFS forecasted meteorology for the meteorological variables required by GLM. By adding random error from this multivariate normal distribution to each member of the 21-member NOAA GEFS, we generated an ensemble of meteorological drivers that represented both NOAA GEFS forecast and downscaling uncertainty (see Supporting Information S3 for more detail about the downscaling methods).

In addition to meteorological driver uncertainty, we include driver uncertainty from future daily inflow discharge and temperature of the primary tributary to the lake or reservoir. For simplicity, we assume here that there is only one primary inflow, but additional inflows could be easily added in applications to other waterbodies. Our approach requires sensor observations of inflow discharge and temperature as covariates in time series models to create forecasted inflow driver data; however, other modeling approaches using precipitation or air temperature could also be substituted to generate the inflow driver data (e.g., Benyahya et al., 2007; Ward et al., 2020). Given that the discharge and water temperature tomorrow is likely to be dependent on today's inflow and we had inflow sensor data available, we used a simple first-order autoregressive time series model that predicts tomorrow's inflow discharge (I^i_{f+1}) as function of today's inflow (I^i_f), today's daily precipitation P^i_f, and random error (ε) that is normally distributed with a mean of 0 and a standard deviation of σ^{inflow}_{rate}:

$$I^i_{f+1} = \beta_0 + \beta_1 I^i_f + \beta_2 P^i_f + \varepsilon \tag{2}$$

in which β_0, β_1,and β_2 are regression coefficients. The σ^{inflow}_{rate} is estimated using the residuals from the regression fit. Equation 2 is iteratively applied through the 16-day forecast horizon for each downscaled meteorological driver ensemble member.

Similar to inflow, we forecast future inflow water temperature (T) as a function of today's water temperature (T^i_f), today's air temperature (A^i_f), and random error (ε) that is normally distributed with a mean of 0 and a standard deviation of σ^{temp}_{rate}:

$$T^t_{f+1} = \beta_3 + \beta_4 T^i_f + \beta_5 A^i_f + \varepsilon \tag{3}$$

β_3, β_4, and β_5 are regression coefficients. σ^{temp}_{rate} is estimated using the residuals from the regression fit. After inflow rate is calculated for each ensemble member, the outflow for that ensemble member is set to equal the inflow to maintain constant water levels.

2.3. Application of Forecasting System

2.3.1. Site Description

We applied and evaluated FLARE at Falling Creek Reservoir (FCR), a dimictic, eutrophic reservoir located in Vinton, Virginia, USA (37.30°N, 79.84°W). FCR is a shallow (maximum depth = 9.3 m, mean depth = 4 m), small (surface area = 0.119 km^2) reservoir (Gerling et al., 2016). The lake generally exhibits summer thermal stratification from May to October and is ice covered for short durations during December to March (Carey, 2019). FCR is primarily fed by one upstream tributary and was maintained at full pond

throughout this study by the Western Virginia Water Authority (WVWA), who own and manage the reservoir as a drinking water supply (Gerling et al., 2014, 2016).

We used high-frequency sensors to monitor FCR water temperature (Carey, Bookout, Woelmer, et al., 2020) and meteorology (Carey, Bookout, Lofton, et al., 2020) and measure the inflow discharge rate and water temperature of the primary tributary entering FCR through a weir (Carey, Hounshell, Lofton, et al., 2020). Descriptions of the sensor array and methods for real-time wireless transfer of data to cloud storage are in Supporting Information S4 and Supporting Information Table S1.

2.3.2. Description of Forecasting Analysis

Our application of the forecasting system at FCR was from 11 July 2018 to 15 December 2019. There were no 16-day forecasts produced between 11 July and 27 August 2018 in order to use observed meteorology and data assimilation during this period to estimate the Σ_t process uncertainty covariance matrix and for the automated tuning of the three GLM parameters (see Supporting Information S1). Starting on 28 August 2018, 16-day forecasts were produced each day until 15 December 2019. We forecasted water temperature on 0.33-m depth intervals starting at the surface through 9.0 m at the sediments. This resulted in 28 model states of water temperature depths in the x matrix (Equation 1). We had sensor observations for 11 of the 28 modeled depths (0.1, 1.0, 1.6, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, and 9.0 m; Supporting Information S4).

In the data assimilation and forecasting steps of the daily workflow (Figure 1), we used $N = 441$ ensemble members. Even though fewer ensemble members have been shown to be effective when applying the EnKF in other studies (e.g., Loos et al., 2020), our ensemble size was selected to adequately sample forecasted meteorology drivers to include uncertainty from both NOAA GEFS and its statistical downscaling. In the forecasting step for this application (Figure 1), we generated 21 members of our downscaling ensemble for each of the 21 NOAA GEFS ensemble members (following Supporting Information S3, Supporting Information Table S2, and Supporting Information Table S3), resulting in a total ensemble size of $N = 441$. The same ensemble size was applied to the data assimilation step to match the forecasting step.

We estimated the coefficients for the inflow discharge and temperature forecasting models in Equations 2 and 3 using data from 1 January 2017 to 12 July 2018 ($\beta_0 = 0.001$, $\beta_1 = 0.948$, $\beta_2 = 0.358$, $\sigma_{rate}^{inflow} = 0.001$, $\beta_3 = 0.203$, $\beta_4 = 0.942$, $\beta_5 = 0.043$, $\sigma_{temp}^{inflow} = 0.943$). Both regressions had high R^2 (Equation 2: $R^2 = 0.98$; Equation 3: $R^2 = 0.93$). In the forecast, each of the 441 meteorological ensembles were used to separately calculate inflow discharge and temperature. Following Equations 2 and 3, random noise was added to each ensemble member to represent uncertainty in forecasting inflow. We set the inflow discharge and temperature at the beginning of each 16-day forecast ($f = 0$) directly from the sensors located on the primary tributary entering FCR.

Beginning 28 August 2018, 16-day horizon forecasts were produced each day until the end of the forecasting period on 15 December 2019. The 475 daily forecasts generated over this period included summer stratified, fall mixed, under-ice, and spring mixed conditions in the reservoir (Carey, Bookout, Woelmer, et al., 2020). Our iterative daily forecasting cycle (Figure 1) entailed: (1) retrieving the previous 24 hours of meteorological, inflow, and water temperature data from the reservoir sensor network, (2) advancing the model states (i.e., the water temperatures at the modeled depths) one day using the observed meteorology and inflows as drivers to the GLM model, (3) using the EnKF to assimilate observed water temperature and update states and parameters, and (4) initiating a 16-day forecast using updated states and parameters as initial conditions and parameters, which started at 12:00:00 UTC of the current day. As a result of the daily data assimilation within the forecasting cycle, parameter distributions and process uncertainty (Σ_t) were updated through each iteration of the forecasting cycle. Once a 16-day forecast was launched from the initial conditions set by data assimilation, the parameters did not change over its 16-day horizon.

We also used a persistence null model to forecast water temperatures from 28 August 2018 to 15 December 2019 as a baseline for evaluating forecast skill. We used a persistence null model rather than a climatology null model because we lacked the multiple years of daily water temperature observations that would be necessary to generate historical averages. Our null model assumed water temperature did not change over the 16-day horizon based on Equation 4:

$$x_{f+1}^{i} = x_{f}^{i} + \varepsilon \tag{4}$$

where ε is represents the contribution of process uncertainty to total forecast uncertainty and is randomly drawn from a multivariate normal distribution with a mean of zero and a covariance matrix of Σ_t^{null}. We

applied the same data assimilation methods used to determine the initial conditions ($x^i_{f=0} = x^i_t$) and process uncertainty covariance matrix (Σ^{null}_t) in the forecast (Supporting Information S1) to the persistence null forecast.

2.3.3. Evaluation of Forecasts

We used multiple metrics to evaluate forecast performance. First, we calculated RMSE and the mean Continuous Ranked Probability Score (CRPS) for the forecasted water temperature (F) and the observed water temperature (y) at each day in the 16-day forecast horizon. CRPS is analogous to mean absolute error and is used with probability distributions produced by ensemble forecasts (Gneiting et al., 2005). CRPS is defined as:

$$CRPS(F, y) = \int_{-\infty}^{\infty} \left(F(u) - 1_{u \geq y}\right)^2 du \tag{5}$$

where $1_{u \geq y}$ is a step function that takes the value 1 if $u \geq y$ and the value 0 otherwise (Gneiting et al., 2005). We used the "verification" package in R (NCAR-Research Applications Laboratory, 2015) to calculate CRPS. The forecast skill was assessed by comparing the CRPS of the forecast ($CRPS_{forecast}$) and the null model ($CRPS_{null}$) using a skill score:

$$CRPS\ Skill = 1 - \frac{CRPS_{forecast}}{CRPS_{null}} \tag{6}$$

Thus, we report three CRPS values: $CRPS_{forecast}$, $CRPS_{null}$, and CRPS forecast skill, which was a normalized value in which 0 indicated a forecast that performed the same as the null, one was a perfect forecast in that it predicted observations with no error, and values less than one indicated a forecast that performed worse than the null.

Second, bias was assessed using the absolute difference between the mean forecasted water temperature (f_t) and the observed water temperature (o_t) at each day in the 16-day forecast, following Equation 7:

$$Bias = \frac{1}{N}\sum_{t=1}^{N}(f_t - o_t) \tag{7}$$

Third, the reliability of the confidence intervals was also assessed by calculating the total number of observations contained in a specific confidence interval (i.e., a reliability diagram; Bröcker & Smith, 2007). We considered a forecast to be well-calibrated if its 90% confidence intervals contained 90% of the observations over the 475 days of forecasts. The confidence interval would be overconfident if fewer than 90% of observations were in this interval and would be underconfident if more than 90% observations were in this interval. We calculated the proportion of observations in the 10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th, and 90th confidence intervals for 1-, 7-, and 16-day time horizons.

In our evaluation, we averaged RMSE, CRPS, bias, and reliability of confidence intervals for each day for the entire water column, as well as individually report forecast performance statistics for the near-surface (1.0 m), middle of the water column (5.0 m), and near-sediments (8.0 m) within the 16-day forecast horizon over the 475 daily forecasts. We primarily focused on 1.0 and 8.0 m for forecast evaluation to capture contrasting thermal conditions in the water column; however, we note that performance at 5.0 m was similar, as shown below. In addition, we classified each of the 475 days as either thermally stratified or mixed to compare relative forecast performance between the two thermal regimes. We applied the turnover criterion of McClure et al. (2018): days with water temperatures at the 1.0 and 8.0 m at 12:00:00 UTC (the beginning of each forecast's daily time step) that were <1°C different were categorized as mixed, and days with water temperatures at 1.0 and 8.0 m that were ≥1°C different were categorized as thermally stratified. Given that there was only intermittent ice cover on the reservoir for $N = 11$ days total during the forecasting period and stable inverse thermal stratification never occurred (Carey, 2019), days with ice were classified as mixed.

In addition, we evaluated how missing temperature observations (e.g., due to potential sensor failure) influence forecast performance by simulating 1-week gaps in water temperature observations throughout the forecasting period. For the 16 forecasts generated on the first day of each month between 28 August 2018 to 15 December 2019, we selectively removed the preceding week's data from data assimilation. We chose

a 1-week data gap because it allowed us to quantify the performance of data assimilation at a site without automated temperature sensors that instead had, e.g., a weekly manual sampling regime with a temperature profiler or handheld temperature sensor. For each of the 16 forecasts, we compared the CRPS values from the forecast, which used the previous week's data for data assimilation, to the forecast which had the previous week's data selectively removed.

Finally, we evaluated the ability of the forecasting system to predict the day that fall turnover was first observed at the reservoir in both years. In the forecasts prior to turnover, we calculated the proportion of ensemble members that predicted the 1.0- and 8.0-m water temperatures to be <1°C different on each day in the forecast. We expected this probability to increase for the day of turnover relative to other days as the onset of turnover approached.

2.3.4. Partitioning of Uncertainty

We isolated the contribution of each uncertainty source by removing the contributions of all other uncertainty sources in a set of scenarios run for each day over the forecasting period. Initial condition uncertainty was removed from the scenarios by initializing all ensemble members at $f = 0$ (Equation 1) with the mean from all ensemble members. Parameter uncertainty was removed by assigning all 441 ensemble members to have the same ensemble means for each of the three automatically calibrated parameters. Process uncertainty was removed by excluding the error term (ε) in Equation 1. NOAA GEFS meteorological driver data uncertainty was removed by using the ensemble mean from the 21-member NOAA GEFS ensemble instead of each individual NOAA GEFS ensemble member. Meteorology statistical downscaling uncertainty was removed by not sampling from the multivariate normal distribution describing the variance and covariance in downscaling error among meteorology variables. Inflow driver data uncertainty was removed by not sampling from the normally distributed uncertainty in Equations 2 and 3.

In our analysis, we compared the variance contributed by each isolated uncertainty source and summed the uncertainty sources to calculate the total forecast variance separately each day in the 16-day horizon. To examine how the uncertainty contributions varied over time, we partitioned the uncertainty of the forecasts generated on the first day of each month, resulting in 16 uncertainty-partitioned forecasts during 28 August 2018 to 15 December 2019. We separately averaged the partitioned uncertainty for the $N = 9$ forecasts generated during thermally stratified conditions and $N = 7$ forecasts generated during mixed conditions.

3. Results

3.1. Observational Water Temperature Data

Falling Creek Reservoir exhibited summer thermal stratification from the beginning of the monitoring period on 11 July 2018 until the onset of fall turnover occurred on 22 October 2018 and then remained mixed until thermal stratification began to set up on 12 March 2019, as defined by a difference of >1°C between 1.0 and 8.0 m (Figure 2a). The reservoir remained thermally stratified throughout the summer until the onset of the second fall turnover on 24 October 2019.

We observed similar thermal dynamics between the 2 years: observed water temperatures at the reservoir's near-surface (1.0 m) peaked on 17 July (28.1°C) in 2018 and on 21 July (29.3°C) in 2019. In both years, fall turnover was preceded by rapidly cooling surface water temperatures, which decreased from 22.6 to 14.7°C at 1.0 m in the 14 days prior to 22 October 2018 and from 20.0 to 14.6°C in the 14 days prior to 24 October 2019. In the mixed periods of 2018 and 2019 (Figure 2b), the 1.0- and 8.0-m depth temperature sensors recorded water temperatures at 12:00:00 UTC that had a mean difference of 0.23°C (±0.55°C, 1 S.D).

3.2. GLM Model and EnKF Performance

The GLM model was able to successfully reproduce observed patterns in reservoir water temperature following data assimilation. With daily assimilation of observed water temperatures using the EnKF and observed (not forecasted) meteorological and inflow drivers, the RMSE for the water temperature predicted by GLM at 1.0- and 8.0-m depth was 0.06°C and 0.07°C, respectively, during the period when data assimilation occurred (11 July 2018 to 15 December 2019; Figure 3). Data assimilation substantially improved GLM predictions, as the RMSE for predicted water temperatures at 1.0 and 8.0 m using observed drivers but without data assimilation was 1.71°C and 1.56°C, respectively (Figure 3). Data assimilation also removed a warming bias at 1 m and reduced the rate of warming through the spring and summer of 2019 at 8.0 m to more closely align

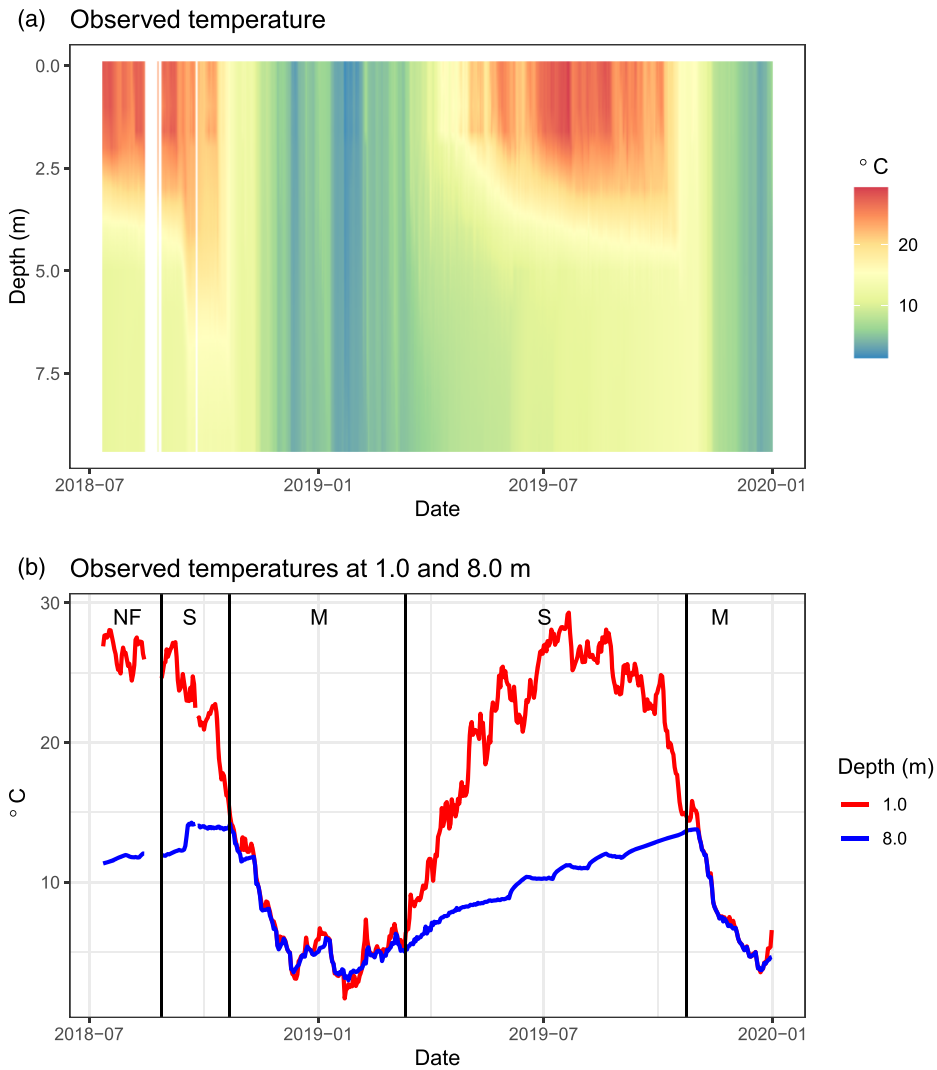

Figure 2. Observed water temperature (in °C) depth profiles in Falling Creek Reservoir during the forecasting period (28 August 2018 to 15 December 2019). (a) A heat map showing the observed temperatures through the water column for the forecasting period based on water temperature sensor observations measured every 10 min at 0.1-, 1.0-, 1.6-, 2.0-, 3.0-, 4.0-, 5.0-, 6.0-, 7.0-, 8.0-, and 9.0-m depths. White values denote days with missing water temperature sensor data as a result of sensor maintenance (15–27 August and 9 September 2018). (b) A comparison of the observed water temperature for two depths (1.0 m: red, 8.0 m: blue). NF (No Forecasts) denotes the period in the analysis when data assimilation was performed but no forecasts were generated. The S and M labels refer to the Stratified and Mixed periods, respectively.

model predictions with observed data (Figure 3). The three tuned GLM parameters were well constrained by data assimilation (Supporting Information Figure S1), and their values varied following expected seasonal patterns.

3.3. Forecast Performance

Every day between 28 August 2018 and 15 December 2019, the forecasting system generated 16-day forecasts of water temperature for the entire water column (see Figure 4 as an example). In general, forecast accuracy was high throughout the 16-day forecast horizon, with the mean bias in forecasted water temperatures within 0.30°C for 1.0 m and 0.22°C for 8.0 m of observed temperatures throughout the 475 daily forecasts, averaged across all forecast time horizons and ensemble members (Figure 5a and Table 1). RMSE was

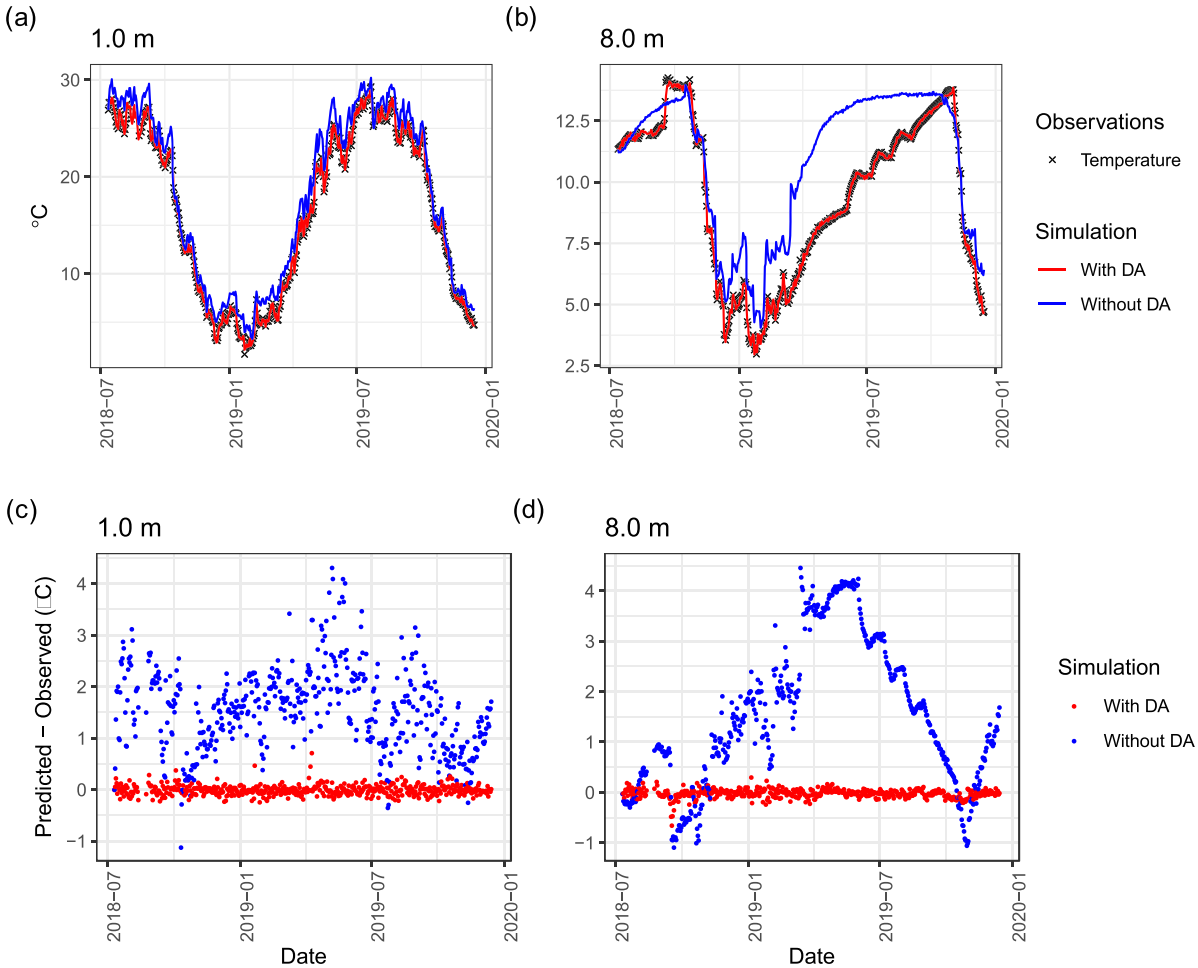

Figure 3. Mean predictions of water temperature (in °C) at 1.0 (a) and 8.0 m (b) with (red lines) and without (blue lines) data assimilation (DA) of in situ observations of temperature measured at 10 depths. All simulations used observed meteorological and inflow drivers. Panels (c) and (d) show the bias in predictions at 1.0 and 8.0 m, respectively, with and without data assimilation.

0.22–0.40°C lower for 8.0 m forecasts than for 1.0 m forecasts aggregated across horizons and ensemble members, ranging from 0.30°C for 8.0 m at a 1-day horizon to 1.62°C for 1.0 m at a 16-day horizon (Table 1).

Forecast error as indicated by the $CRPS_{forecast}$ increased as the forecast horizon increased, though it was notable that the $CRPS_{forecast}$ was consistently better than the null persistence model ($CRPS_{null}$) throughout the 16-day horizon for 1.0 m temperatures (Figure 5b). This resulted in a CRPS forecast skill score for the 16-day time horizon of 0.53 when averaged across all forecasts. At 8.0-m depth, the null persistence model ($CRPS_{null}$) performed better than the forecasts at the 1- and 7-day time horizons and only marginally worse at the 16-day time horizon. This resulted in a lower CRPS forecast skill score of 0.14 at 8.0 m when averaged across all forecasts (Table 1).

Forecast confidence intervals at 1.0-m depth were well calibrated (exactly 90% of the observations were in the 90% confidence intervals; Table 1) and generally fell along the 1:1 line between the percentage of observations within a specific forecast confidence interval (Figure 5c). Forecast confidence intervals were less well calibrated at 8.0 m (56–76% of observations were within the 90% interval) and underestimated the uncertainty across all time horizons (Table 1 and Figure 5d). Averaged across all depths, the forecast confidence intervals were smaller than expected (79–85% of observations fell within the 90% confidence intervals; Table 1).

Forecasting performance was generally similar between the thermally stratified and mixed periods, with a few key exceptions. At 1.0 m, the CRPS forecast skill was consistently greater (up to 0.3 skill score) and bias was lower in stratified than mixed conditions, regardless of forecast time horizon (Table 1). In contrast, at 8.0 m, CRPS

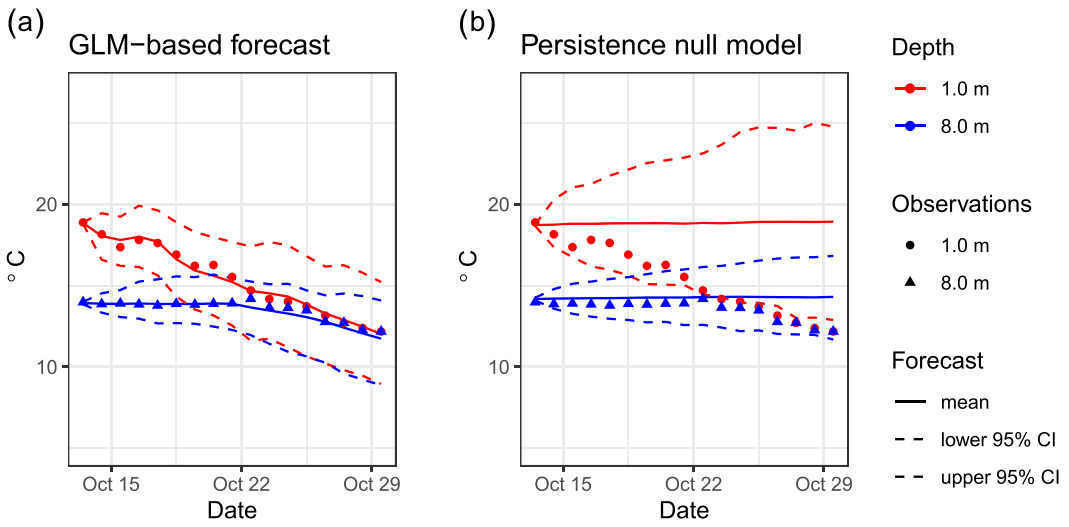

Figure 4. An example of a 16-day water temperature (in °C) forecast (initiated on 13 October 2018) at the near-surface (1.0 m) and near-sediment (8.0 m) depths by the (a) General Lake Model (GLM) and (b) the persistence null model. The forecast uncertainty in both panels is compared to the observations from the water temperature sensors at 1.0- and 8.0-m depths.

forecast skill was slightly greater at 1- and 7-day horizons but there was no difference in forecast skill between mixed and stratified conditions at a 16-day horizon (Table 1). Finally, the confidence intervals were better calibrated at 8.0 m for the mixed period (85% of observations in the 90% confidence interval at the 16-day time forecast horizon) than the stratified period (36% of observations in the 90% confidence interval; Table 1).

The forecasts were sensitive to missing sensor data during the week prior to a forecast being generated but the overall effect on forecast performance was minimal (Supporting Information Figure S2). Aggregated over 16 forecasts generated monthly throughout the forecasting period, the mean $CRPS_{forecast}$ on the first day of the forecast horizon was ~0.2°C higher at both 1.0 and 8.0 m when the preceding week of water temperature observations was selectively removed, in comparison to forecasts for the same time periods with no data gaps. At 1.0 m depth, this difference in mean $CRPS_{forecast}$ diminished to <0.1°C after the first 6 days of the forecast horizon. At 8.0 m, the difference in mean $CRPS_{forecast}$ decreased over the forecast horizon but remained ≥0.15°C at the end of the 16-day forecast horizon (Supporting Information Figure S2).

The forecasts successfully predicted the onset of fall turnover that occurred on 22 October 2018 up to 14 days in advance (Figure 6a). At 14 days prior to 22 October 2018, the predicted chance of turnover occurring on that day emerged above 50% for the first time (Figure 6a). Confidence that turnover would occur on 22 October increased 8 days prior to turnover, when its predicted chance exceeded 75%. By 19 October (three days prior to turnover), the predicted chance of turnover on 22 October increased up to 81%, 31% higher than any of the other potential days.

The forecasts successfully predicted the onset of fall turnover that occurred on 24 October 2019 up to 4 days in advance (Figure 6b). The predicted chance of turnover occurring that day increased from 43% 5 days prior to turnover to 63% 4 days prior to turnover. The predicted chance of turnover occurring on 24 October remained above 50% as the day of turnover approached. However, the confidence in turnover date was lower in 2019 than 2018, potentially because the reservoir temporarily restratified immediately after the initial onset of mixed conditions in 2019 (Figure 2).

3.4. Uncertainty Partitioning

Forecast uncertainty varied over time and depth. In the thermally stratified period, the total forecast uncertainty was approximately four times higher for forecasts at 1.0 than 8.0 m, with striking differences in the relative importance of different sources to total forecast uncertainty (Figure 7). For both depths, model process uncertainty was the most important contributor to total uncertainty over the first 3 days in a 16-day forecast horizon during stratified conditions. After 3 days, the contribution of meteorological driver uncertainty dominated at 1.0-m depth (>50% of uncertainty) while process uncertainty remained the dominant source of

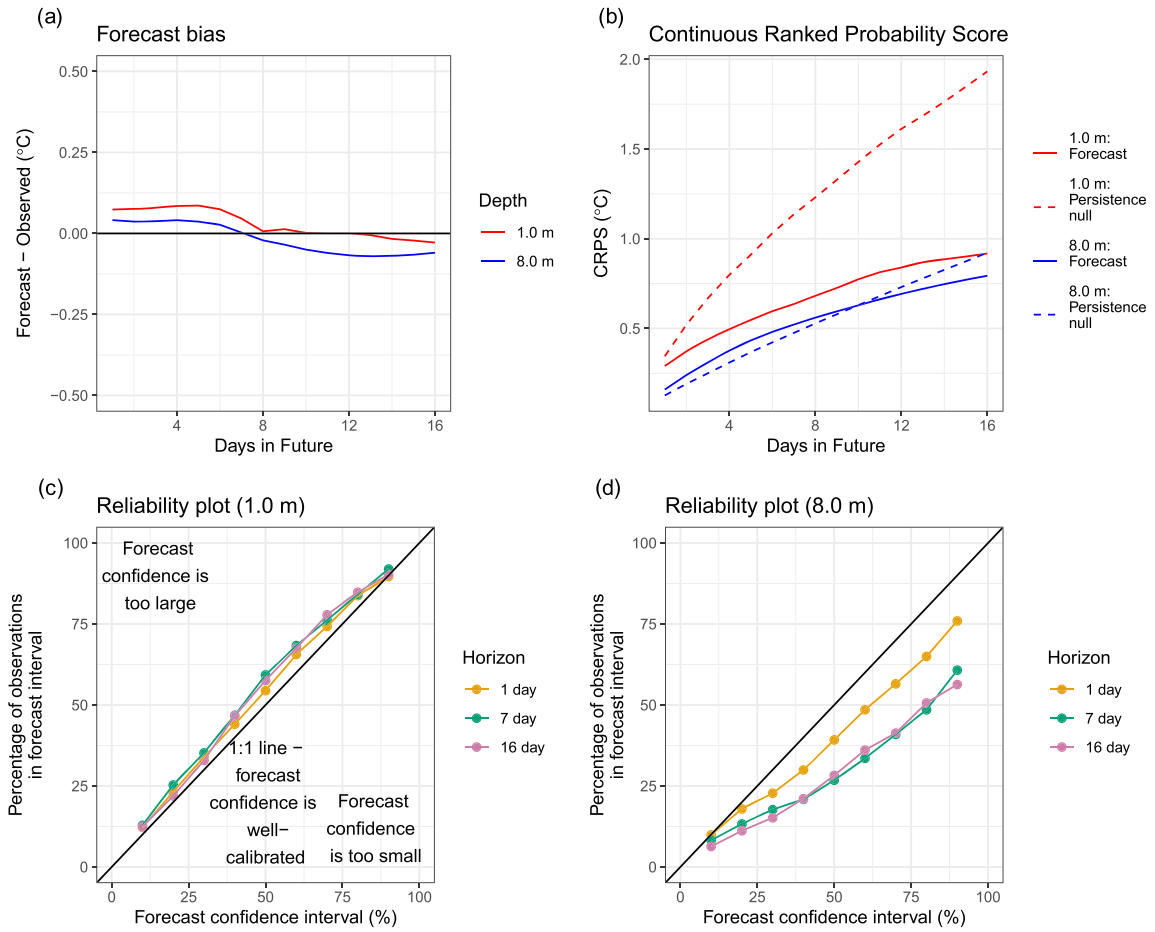

Figure 5. Water temperature forecasts for 1.0 (red) and 8.0 m (blue) had low bias, high forecast skill (relative to a null persistence model), and well-calibrated confidence intervals. (a) The difference between the forecasted and observed water temperatures (bias) (in °C) at 1.0 and 8.0 m averaged across all forecasts generated during the 475-day forecasting period. (b) The Continuous Ranked Probability Score (CRPS, a metric for evaluating ensemble forecasting that is analogous to Mean Absolute Error) of the forecasts at the two depths averaged across the entire forecasting period in comparison to a null persistence model. The null model assumed that water temperature did not change over the 16-day forecast horizon. The percentage of (c) 1.0- and (d) 8.0-m observations during the forecasting period that were within the forecast confidence interval that is specified on the *x* axis are shown. For example, the 90% forecast confidence interval encompasses the 5th to 95th percentile of the 441 forecast ensemble members and ideally should contain 90% of water temperature observations. Values for 1-, 7-, and 16-day forecast horizons for the 1.0-m depth are shown.

uncertainty at 8.0-m depth throughout the 16-day forecast horizon (>95% of uncertainty). At the near-surface, the statistical downscaling of the meteorological driver data was a more important uncertainty source than the uncertainty due to the NOAA meteorological forecast itself throughout the 16-day horizon. Parameters, inflow driver data, and initial conditions were not important uncertainty sources at either depth.

In contrast to the thermally stratified period, the total uncertainty and the relative contribution of the different sources were similar between the near-surface (1.0 m) and near-sediment (8.0 m) depths during mixed conditions (Figure 7). In the mixed period, process uncertainty was the most important source for the first 5 days of the forecast before the relative importance of the combined meteorological downscaling and forecasts dominated. Similar to the stratified period, parameters, inflow driver data, and initial conditions were minimal sources of uncertainty during mixed conditions.

4. Discussion

Overall, FLARE was able to forecast water temperature on average within 1.4°C (RMSE) at all depths of the reservoir over a 16-day horizon during a 475-day period that encompassed both thermally stratified and mixed conditions (Table 1). Importantly, the forecasting system was able to predict both observed

Table 1
Forecast Evaluation Metrics for the Entire Forecasting Period ("All Conditions"), During Thermally Stratified Conditions Only, and During Mixed Conditions Only

Forecast time horizon	Metric	All conditions (n = 475 days)				Stratified conditions (n = 282 days)		Mixed conditions (n = 193 days)	
		1.0 m	5.0 m	8.0 m	All depths	1.0 m	8.0 m	1.0 m	8.0 m
1 day	RMSE (°C)	0.52	0.36	0.30	0.44	0.56	0.23	0.45	0.39
	$CRPS_{forecast}$ (°C)	0.29	0.20	0.16	0.23	0.32	0.12	0.25	0.22
	$CRPS_{null}$ (°C)	0.34	0.20	0.13	0.23	0.40	0.09	0.26	0.18
	CRPS forecast skill (unitless)	0.16	0.00	−0.26	−0.07	0.21	−0.3	0.04	−0.23
	Bias (°C)	0.07	0.05	0.04	0.03	0.07	−0.03	0.09	0.15
	CI reliability (%)	90	89	76	85	93	74	85	79
7 day	RMSE (°C)	1.13	0.99	0.87	1.04	0.98	0.70	1.32	1.07
	$CRPS_{forecast}$ (°C)	0.64	0.48	0.52	0.56	0.58	0.46	0.72	0.61
	$CRPS_{null}$ (°C)	1.14	0.64	0.47	0.77	1.26	0.28	0.95	0.76
	CRPS forecast skill (unitless)	0.44	0.25	−0.10	0.20	0.54	−0.66	0.24	0.20
	Bias (°C)	0.04	0.09	0.00	0.05	−0.16	−0.27	0.34	0.39
	CI reliability (%)	92	91	61	82	99	46	82	82
16 day	RMSE (°C)	1.62	1.30	1.20	1.40	1.68	1.13	1.53	1.30
	$CRPS_{forecast}$ (°C)	0.92	0.65	0.79	0.80	0.95	0.83	0.87	0.75
	$CRPS_{null}$ (°C)	1.93	1.19	0.92	1.40	2.11	0.56	1.68	1.45
	CRPS forecast skill (unitless)	0.53	0.45	0.14	0.39	0.55	−0.48	0.48	0.49
	Bias (°C)	−0.03	0.12	−0.06	0.03	−0.45	−0.56	0.57	0.66
	CI reliability (%)	90	90	56	79	94	36	85	85

Note. Metrics are calculated for all depths averaged together ("all depths"), as well as reported individually for the near-surface (1.0 m), middle of the water column (5.0 m), and at the near-sediments (8.0 m). The metrics include: RMSE (root mean square error between observations and forecasted temperatures); $CRPS_{forecast}$ (Continuous Ranked Probability Score for the FLARE forecast); $CRPS_{null}$ (CRPS for the persistence null forecast); CRPS forecast Skill (skill score for the FLARE forecasting based on the comparison between the $CRPS_{forecast}$ and $CRPS_{null}$); bias (the difference between observations and forecasted temperatures); and confidence interval (CI) reliability (the percentage of observations within the 90% confidence intervals).

temperatures and identify the date of fall turnover at least four, and up to 14, days in advance. In general, forecasting system performance was similar between stratified and mixed periods (Table 1) and did not strongly degrade when a week of data was not available for data assimilation prior to forecast generation (Supporting Information Figure S2). These results suggest that the system is likely robust in a range of reservoir conditions, though additional forecasts are needed to provide a full assessment of FLARE performance. Generally, 1-D hydrodynamic models used for hindcasting aim to predict water temperature within an RMSE of 2°C (e.g., Bruce et al., 2018), so the level of accuracy associated with FLARE future forecasts over 475 days at Falling Creek Reservoir exceeds expectations (RMSE = 1.4°C aggregated across all periods and all depths; Table 1). Below, we examine the drivers of forecast uncertainty, performance of turnover forecasts, applications for scaling FLARE to other waterbodies, and potential improvements that could improve the use of FLARE for water management.

4.1. Forecast Uncertainty

We found that process uncertainty was the most important source of uncertainty early in the 16-day forecast but that meteorological driver data uncertainty dominated by the end of the forecasting period for all but near-sediment depths during the thermally stratified period (Figure 7). This finding is intuitive, as the epilimnion is more sensitive to meteorological forcing than the hypolimnion, and during mixed conditions, both surface and near-sediment depths exhibit similar dynamics. We note that process uncertainty did not decline through the forecast horizon, rather, the total uncertainty grew while the relative importance of process uncertainty decreased. By definition, process uncertainty is not attributed to an individual component of a model but a model's overall inability to represent complex interactions occurring in the natural environment (Dietze, 2017a). Improvements to the GLM model structure, such as improved sediment heating formulations, could decrease this uncertainty source, particularly in shallower waterbodies; however, there is likely an irreducible component of process uncertainty inherent to using a 1-D model to simulate a 3-D waterbody.

The contribution of the NOAA ensemble forecast uncertainty to total forecast uncertainty was overall less than the contribution of uncertainty from downscaling the coarse-scale NOAA forecast to the local site.

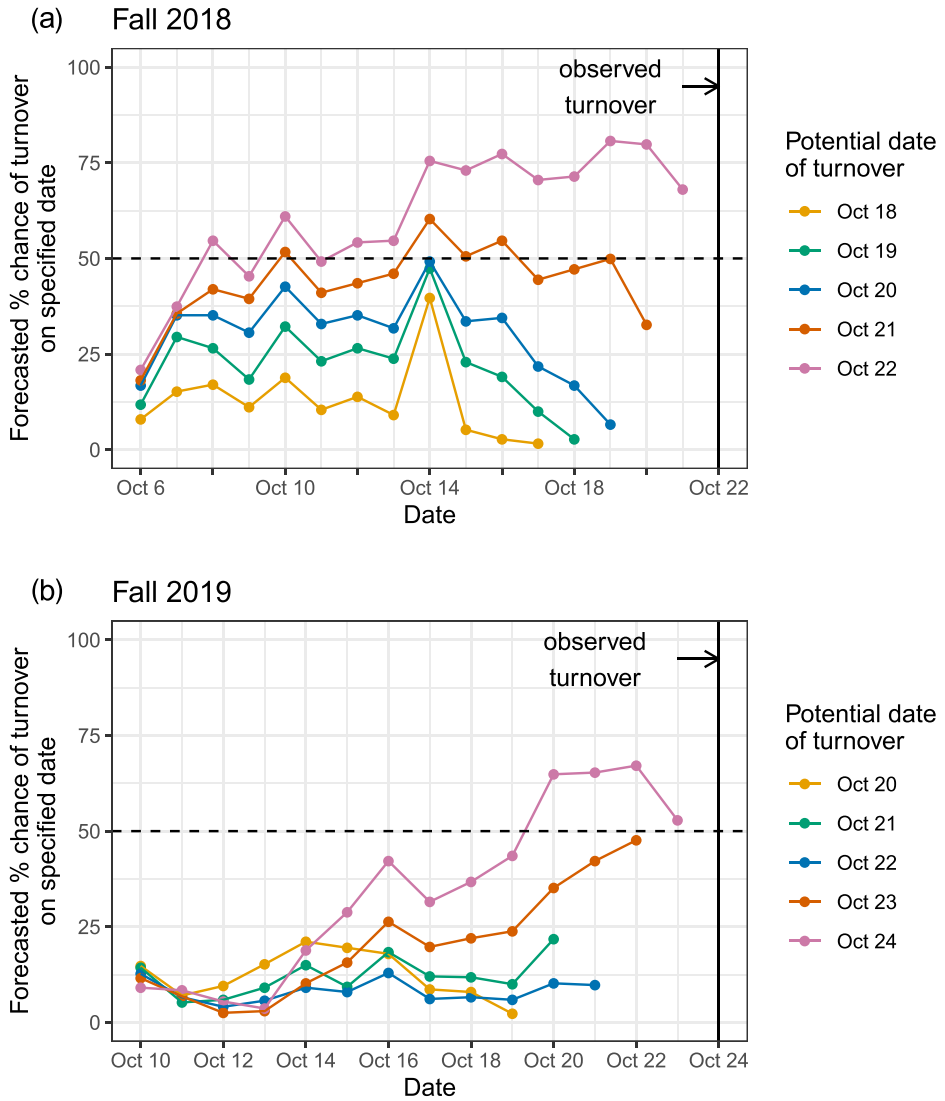

Figure 6. The forecasting system was able to anticipate the day of fall turnover (a) ~14 days in advance in 2018 and (b) ~4 days in advance in 2019 (as defined by when the percent chance first exceeded 50%). Turnover is defined as the first day in autumn when water temperatures at 1.0 and 8.0 m were <1°C at 12:00:00 UTC. Each panel shows how the percent chance of turnover for a particular forecasted day updated daily until turnover was observed.

This downscaling was based on relationships between the NOAA GEFS forecasted variables and the meteorological data measured at the reservoir (see Supporting Information S3). This highlights that future work should focus on applying more advanced meteorological downscaling methods, such as neural networks (Kumar et al., 2012), to build better relationships between the NOAA forecasts and the local meteorological station. Our uncertainty results are comparable to Dietze (2017b), who partitioned the uncertainty in forest carbon flux forecasts over 16-day horizons and similarly found that NOAA ensemble driver data uncertainty dominated total forecast uncertainty at the end of the 16-day horizon, though that study did not include NOAA forecast downscaling uncertainty. Dietze (2017b) also found that process uncertainty was more important than the meteorological driver data uncertainty early in the 16-day forecast horizon. Overall, while that study and ours are just two examples of forecast uncertainty partitioning, they provide insight from very different ecosystems (reservoir water temperatures vs. forest carbon fluxes) and together suggest that meteorological driver data are an important contributor of uncertainty in near-term iterative forecasts. Moreover, our study indicates that future forecasting

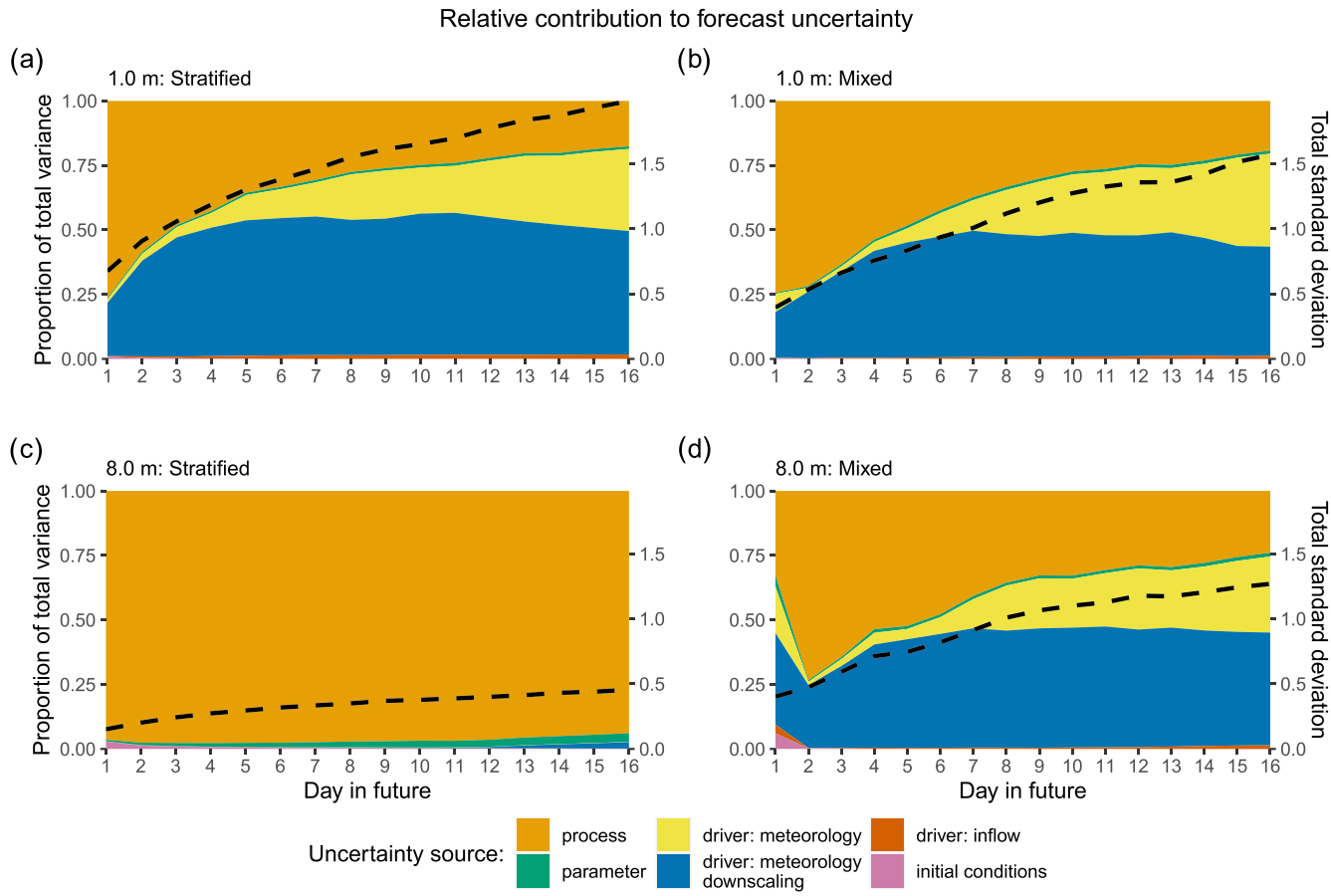

Figure 7. The relative contribution of the individual sources of uncertainty (left axis) to the total forecast uncertainty (right axis, black dashed line) varies through the 16-day forecast horizon and between stratified (left column) and mixed (right column) conditions at both 1.0-m depth (top row) and 8.0-m depth (bottom row). The relative contributions and total forecast variance for each day are averaged across 16 forecasts: $N = 9$ during summer stratification (a, c) and $N = 7$ during mixed conditions (b, d). Two depths are shown: 1.0 m (a and b); and 8.0 m (c and d).

applications should investigate the relative role of meteorological forecasts versus their downscaling when quantifying total forecast uncertainty.

Throughout the forecasting period, inflow driver data, parameters, and initial conditions were minimal sources of uncertainty, regardless of forecast horizon. During the forecasting period, the reservoir exhibited a median water residence time of 145 days (Carey, Hounshell, Lofton, et al., 2020; Gerling et al., 2014), so it was surprising that the inflow driver data did not contribute more to total uncertainty. However, the low residual error in our inflow forecast model (Equation 2; $R^2 = 0.98$) suggests that our autoregressive time series model was able to successfully capture variability in inflow dynamics throughout the forecasting period. Inflow driver data may contribute more uncertainty for waterbodies with multiple inflows and higher discharge than we observed in our case study (e.g., Roberts et al., 2018), and for applications that use other modeling approaches to generate forecasted inflow discharge and water temperature when inflow sensor data are not available. Parameter uncertainty and initial conditions uncertainty were similarly minor, indicating that the data assimilation with state updating and automated parameter tuning that was built into the iterative forecasting cycle was able to constrain these sources of forecast uncertainty in our application.

4.2. Turnover Forecasts

FLARE was able to identify the onset of fall turnover ~14 days in advance in 2018 and approximately 4 days in advance in 2019. This determination was made by identifying the first day when the probability of turnover exceeded 50% (Figure 6). This difference in performance between years may be due to differing underlying hydrodynamic and meteorological drivers of fall turnover. In 2018, the surface waters of the reservoir

had rapidly cooled (by ~8°C) in the 14 days prior to the onset of turnover in response to cold night air temperatures (down to 3°C) that continued after turnover. The day prior to turnover onset in 2018 coincided with high wind speeds, up to 9.8 m s^{-1} (Carey, Bookout, Lofton, et al., 2020), which persisted after turnover and promoted continuous mixing once turnover started. In contrast, in 2019, the reservoir restratified later on the same day that the turnover criterion was first met due to warm daytime air temperatures (up to 18.7°C on 24 October) and wind speeds were approximately half of what was observed in 2018. The reservoir did not continuously exhibit isothermal conditions in 2019 until 2 November, 8 days after the turnover criterion was initially met. The McClure et al. (2018) criterion of turnover was only met for 17 hr on 24 October 2019, making it extremely challenging for the forecasting system to correctly predict turnover on that particular day. Here, we used a previously defined criteria for turnover; other definitions of turnover could be used with different water temperature thresholds or time periods (e.g., the temperature difference between 1- and 8-m depth must be <1°C and maintained for 48 consecutive hours, following Woolway et al., 2014). How ever chosen, it is important that the turnover criterion be defined a priori before turnover forecasts are generated.

4.3. Future Applications of FLARE

We anticipate that FLARE may be of interest to managers of lakes and reservoirs where forecasts of water temperature can provide decision support. Many reservoirs, including FCR, have dynamic water extraction schedules, which are informed by water temperature depth profiles and the magnitude of thermal stratification (Mi et al., 2019). Thus, knowing water temperature depth profiles a priori provides managers with valuable information for optimizing the temperature of downstream releases, water quality management, and in situ management prior to water treatment (Caissie et al., 2016; Huang et al., 2011; Mi et al., 2019; Pike et al., 2013; Weber et al., 2017).

Importantly, managers could use their current water temperature sensors for setting up and running FLARE to generate forecasts for their waterbody of interest. All software components of the FLARE system are open-source and the automated parameter tuning of the GLM model as part of the EnKF can assist model setup. To deploy FLARE for an application similar to the FCR case study described here, a water utility would need water temperature sensors and an inflow discharge sensor connected to the internet, meteorology data, and knowledge of the waterbody's bathymetry to set up boundary conditions in a hydrodynamic model. While our application of FLARE at FCR used the GLM model, our forecasting framework is flexible to accommodate other lake and reservoir hydrodynamic models. To be able to run in FLARE, the model would need to be set up for iterative forecasting and daily data assimilation, which requires running the model for 1 day, adjusting the model states and parameters based on sensor observations, and then restarting the model from the adjusted states for the next day's run (Figure 1).

4.4. Limitations and Potential Improvements

While the FLARE forecasting system presented here was able to predict water temperature over the 16-day time horizon with low bias and an error substantially lower than a null persistence model, there are two important potential improvements. First, FLARE is using a physics-based, 1-D hydrodynamic model, so improvements to the physical parameterization and numerical solution could improve forecast skill, though we note that a 3-D model may be more time- and compute-intensive. Alternatively, thermal stratification models with a more simple structure (e.g., the 2-layer model in Niemeyer et al., 2018) than GLM that leverage global datasets on storage-area-depth (Yigzaw et al., 2018, 2019) could be used to evaluate the level of complexity required for water temperature forecasting across many waterbodies of varying size. Furthermore, forecasting approaches that combine physics-based models with machine learning (e.g., Jia et al., 2019; Read et al., 2019) could potentially reduce forecast uncertainty by leveraging the power of mechanistic models and machine learning methods. However, machine learning-based methods must be able to fully quantify forecast uncertainty to be comparable to our process model-based approach (see Daw et al., 2020). Second, our parameter uncertainty may be underestimated because we only included three parameters in the EnKF data assimilation, though we note that our sensitivity analysis determined those three parameters to be the most important for water temperature predictions within the GLM (Supporting Information S1). While the ability to estimate parameters using EnKF is well-established, the EnKF method is not specifically designed to estimate parameter distributions like Bayesian Monte Carlo Markov Chain methods (Dietze, 2017a). A current limitation to implementing a Bayesian Monte Carlo Markov Chain approach is the computation time to execute the GLM simulation within a daily iterative forecasting

cycle. Future work that uses emulators of GLM may be able to speed computation and allow for more robust estimation of the joint distribution model of parameters that represent both prior knowledge and observed data (e.g., Fer et al., 2018).

We note four limitations of our application of FLARE at FCR. First, our implementation included only one inflow, in which we used a first-order autoregressive model to forecast future inflow discharge and water temperature. Further work is needed to evaluate FLARE in waterbodies with more inflows. For those waterbodies, FLARE could be linked to a watershed hydrology model to mechanistically link precipitation and temperature forecasts to inflow driver data. Second, we set outflow equal to inflow to maintain a constant water level based on observations, which does not occur in many lakes and reservoirs. Future development of FLARE could explicitly assimilate observations of water level to constrain forecasts in more dynamic systems. Third, our data assimilation in FLARE used high-frequency sensors measuring water temperature at 1-m depth intervals at 10-min resolution, an observational capacity at FCR that may not be available for all waterbodies. While we found that a 1-week data gap did not substantially degrade performance (Supplemental Information Figure S2), further evaluation of FLARE should quantify how forecast quality degrades as the vertical and temporal resolution of observations are reduced to reflect common temperature monitoring approaches (e.g., monthly temperature depth profile measurements with a handheld sensor, minute to hourly temporal observation at only one depth, and remote sensing measurements of surface temperature). Finally, our application was limited to a period with minimal ice cover; further work is needed to determine FLARE performance for FCR during longer ice-covered periods and other waterbodies that experience ice for many consecutive months.

4.5. Conclusions

Overall, our study demonstrates the utility of FLARE's daily iterative forecasting cycle (Figure 1) for forecasting lake and reservoir water temperatures with fully specified uncertainties that can be applied to other waterbodies. FLARE builds the foundation for future water quality data assimilation and forecasting because multiple ecosystem models can easily be coupled to the GLM hydrodynamic model (e.g., Hipsey et al., 2013; Snortheim et al., 2017), enabling predictions of dissolved oxygen, algal blooms, and biogeochemical cycling with uncertainty. Importantly, FLARE provides a method for partitioning uncertainty in forecasts that identifies how to prioritize future research to increase confidence in forecasts. Given the pressing need for tools to anticipate the increasing variability of freshwater ecosystems, near-term iterative forecasting systems such as FLARE provide the ability to anticipate future change for stakeholders, managers, and policy-makers.

Conflict of Interest

The authors declare no conflicts of interest.

Data Availability Statement

All data used in this study are available in the Environmental Data Initiative repository (Carey, 2019; Carey, Bookout, Lofton, et al., 2020; Carey, Bookout, Woelmer, et al., 2020; Carey, Hounshell, Lofton, et al., 2020) and through NOAA forecast archives (https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-ensemble-forecast-system-gefs). Code for FLARE (Thomas et al., 2020) and the sensor gateways (Daneshmand et al., 2020) can be found in the Zenodo repository.

Acknowledgments

This work was supported by the U.S. National Science Foundation (CNS-1737424, DEB-1753639, DBI-1933016, DEB-1926050, and DEB-1926388); the Virginia Tech Global Change Center; and Fralin Life Sciences Institute. We thank our Ecological Forecasting seminar students, the Smart Reservoir team, and Ecological Forecasting Initiative (EFI) colleagues for helpful feedback on the project; the Western Virginia Water Authority for their long-term support and access to field sites; and Mary Lofton, Ryan McClure, and Whitney Woelmer for their critical help in the field.

References

Bal, G., Rivot, E., Bagliniere, J. L., White, J., & Prevost, E. (2014). A hierarchical Bayesian model to quantify uncertainty of stream water temperature forecasts. *PLoS ONE*, *9*(12), e115659. https://doi.org/10.1371/journal.pone.0115659

Baracchini, T., Chu, P. Y., Šukys, J., Lieberherr, G., Wunderle, S., Wüest, A., & Bouffard, D. (2020). Data assimilation of in situ and satellite remote sensing data to 3D hydrodynamic lake models: A case study using Delft3D-FLOW v4.03 and OpenDA v2.4. *Geoscientific Model Development*, *13*(3), 1267–1284. https://doi.org/10.5194/gmd-13-1267-2020

Baracchini, T., Wuest, A., & Bouffard, D. (2020). Meteolakes: An operational online three-dimensional forecasting platform for lake hydrodynamics. *Water Research*, *172*, 115529. https://doi.org/10.1016/j.watres.2020.115529

Benyahya, L., Caissie, D., St-Hilaire, A., Ouarda, T. B. M. J., & Bobée, B. (2007). A review of statistical water temperature models. *Canadian Water Resources Journal*, *32*(3), 179–192. https://doi.org/10.4296/cwrj3203179

Berthet, L., Piotte, O., Gaume, É., Marty, R., & Ardilouze, C. (2016). *Operational forecast uncertainty assessment for better information to stakeholders and crisis managers paper presented at 3rd European conference on flood risk management (FLOODrisk 2016)*. Lyon, France: EDP Sciences.

Bowman, D. C., & Lees, J. M. (2015). Near real time weather and ocean model data access with rNOMADS. *Computers & Geosciences*, *78*, 88–95. https://doi.org/10.1016/j.cageo.2015.02.013

Bröcker, J., & Smith, L. A. (2007). Increasing the reliability of reliability diagrams. *Weather and Forecasting*, *22*(3), 651–661. https://doi.org/10.1175/waf993.1

Brookes, J. D., Carey, C. C., Hamilton, D. P., Ho, L., van der Linden, L., Renner, R., & Rigosi, A. (2014). Emerging challenges for the drinking water industry. *Environmental Science & Technology*, *48*(4), 2099–2101. https://doi.org/10.1021/es405606t

Bruce, L. C., Frassl, M. A., Arhonditsis, G. B., Gal, G., Hamilton, D. P., Hanson, P. C., et al. (2018). A multi-lake comparative analysis of the General Lake Model (GLM): Stress-testing across a global observatory network. *Environmental Modelling & Software*, *102*, 274–291. https://doi.org/10.1016/j.envsoft.2017.11.016

Butcher, J. B., Nover, D., Johnson, T. E., & Clark, C. M. (2015). Sensitivity of lake thermal and mixing dynamics to climate change. *Climatic Change*, *129*(1–2), 295–305. https://doi.org/10.1007/s10584-015-1326-1

Caissie, D., Thistle, M. E., & Benyahya, L. (2016). River temperature forecasting: Case study for Little Southwest Miramichi River (New Brunswick, Canada). *Hydrological Sciences Journal*, *62*(5), 683–697. https://doi.org/10.1080/02626667.2016.1261144

Çalışkan, A., & Elçi, Ş. (2008). Effects of selective withdrawal on hydrodynamics of a stratified reservoir. *Water Resources Management*, *23*(7), 1257–1273. https://doi.org/10.1007/s11269-008-9325-x

Carey, C. C. (2019). *Ice cover data for Falling Creek reservoir, Vinton, Virginia, USA for 2013-2019*. Environmental Data Initiative. https://doi.org/10.6073/pasta/eecfa7ba08f9ee233ece6df0e9cbcb3a

Carey, C. C., Bookout, B. J., Lofton, M. E., & McClure, R. P. (2020). *Time series of high-frequency meteorological data at Falling Creek Reservoir, Virginia, USA 2015–2019*. Environmental Data Initiative. https://doi.org/10.6073/pasta/ea47ae493c7025d61245287649895e60

Carey, C. C., Bookout, B. J., Woelmer, W. M., & Lewis, A. S. L. (2020). *Time series of high-frequency sensor data measuring water temperature, dissolved oxygen, conductivity, specific conductivity, total dissolved solids, chlorophyll a, phycocyanin, and fluorescent dissolved organic matter at discrete depths in Falling Creek Reservoir, Virginia, USA in 2018–2019*. Environmental Data Initiative. https://doi.org/10.6073/pasta/b888ac006ef4ca601f63e2703d7476b9

Carey, C. C., Hounshell, A. G., Lofton, M. E., Bookout, B. J., Corrigan, R. S., Gerling, A. B., et al. (2020). *Discharge time series for the primary inflow tributary entering Falling Creek Reservoir, Vinton, Virginia, USA 2013-2020*. Environmental Data Initiative. https://doi.org/10.6073/pasta/b888ac006ef4ca601f63e2703d7476b9

Carey, C. C., Ibelings, B. W., Hoffmann, E. P., Hamilton, D. P., & Brookes, J. D. (2012). Eco-physiological adaptations that favour freshwater cyanobacteria in a changing climate. *Water Research*, *46*(5), 1394–1407. https://doi.org/10.1016/j.watres.2011.12.016

Casamitjana, X., Serra, T., Colomer, J., Baserba, C., & Pérez-Losada, J. (2003). Effects of the water withdrawal in the stratification patterns of a reservoir. *Hydrobiologia*, *504*(1–3), 21–28. https://doi.org/10.1023/b:Hydr.0000008504.61773.77

Cheng, Y., Voisin, N., Yearsley, J. R., & Nijssen, B. (2020). Reservoirs modify river thermal regime sensitivity to climate change: A case study in the Southeastern United States. *Water Resources Research*, *56*. https://doi.org/10.1029/2019WR025784

Clark, J. S., Carpenter, S. R., Barber, M., Collins, S., Dobson, A., Foley, J. A., et al. (2001). Ecological forecasts: An emerging imperative. *Science*, *293*(5530), 657–660. https://doi.org/10.1126/science.293.5530.657

Clark, M. P., Rupp, D. E., Woods, R. A., Zheng, X., Ibbitt, R. P., Slater, A. G., et al. (2008). Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model. *Advances in Water Resources*, *31*(10), 1309–1324. https://doi.org/10.1016/j.advwatres.2008.06.005

Cooke, G. D., Welch, E., Peterson, B. S., & Nichols, S. A. (2005). *Restoration and management of lakes and reservoirs*. Boca Raton, FL: CRC Press.

Daneshmand, V., Thomas, R. Q., Bookout, B. J., Carey, C. C., & Figueiredo, R. J. (2020). *Sensor Gateway Code for Forecasting Lake and Reservoir Ecosystems (FLARE) (Version v1.0)*. Zenodo. https://doi.org/10.5281/zenodo.3862907

Daw, A. D., Thomas, R. Q., Carey, C. C., Read, J. S., Appling, A. P., & Karpatne, A. (2020). Physics-guided architecture (PGA) of neural networks for quantifying uncertainty in Lake temperature Modeling. In *Proceedings of the 2020 SIAM International Conference on Data Mining* (pp. 532–540). Cincinnati, Ohio: Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611976236.60

Delpla, I., Jung, A. V., Baures, E., Clement, M., & Thomas, O. (2009). Impacts of climate change on surface water quality in relation to drinking water production. *Environment International*, *35*(8), 1225–1233. https://doi.org/10.1016/j.envint.2009.07.001

Dietze, M. C. (2017a). *Ecological forecasting*. Princeton: Princeton University Press.

Dietze, M. C. (2017b). Prediction in ecology: A first-principles framework. *Ecological Applications*, *112*(1), 6252–6213. https://doi.org/10.1002/eap.1589

Dietze, M. C., Fox, A., Beck-Johnson, L. M., Betancourt, J. L., Hooten, M. B., Jarnevich, C. S., et al. (2018). Iterative near-term ecological forecasting: Needs, opportunities, and challenges. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(7), 1424–1432. https://doi.org/10.1073/pnas.1710231115

Effler, S. W., & Matthews, D. A. (2008). Implications of redox processes for the rehabilitation of an urban lake, Onondaga Lake, New York. *Lake and Reservoir Management*, *24*(2), 122–138. https://doi.org/10.1080/07438140809354056

Evensen, G. (2003). The Ensemble Kalman Filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, *53*(4), 343–367. https://doi.org/10.1007/s10236-003-0036-9

Evensen, G. (2009). *Data Assimilation*. Berlin, Heidelberg: Springer.

Fer, I., Kelly, R., Moorcroft, P. R., Richardson, A. D., Cowdery, E. M., & Dietze, M. C. (2018). Linking big models to big data: Efficient ecosystem model calibration through Bayesian model emulation. *Biogeosciences*, *15*(19), 5801–5830. https://doi.org/10.5194/bg-15-5801-2018

Gerling, A. B., Browne, R. G., Gantzer, P. A., Mobley, M. H., Little, J. C., & Carey, C. C. (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. https://doi.org/10.1016/j.watres.2014.09.002

Gerling, A. B., Munger, Z. W., Doubek, J. P., Hamre, K. D., Gantzer, P. A., Little, J. C., & Carey, C. C. (2016). Whole-catchment manipulations of internal and external loading reveal the sensitivity of a century-old reservoir to hypoxia. *Ecosystems*, *19*(3), 555–571. https://doi.org/10.1007/s10021-015-9951-0

Gneiting, T., Raftery, A. E., Westveld, A. H., & Goldman, T. (2005). Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Monthly Weather Review*, *133*(5), 1098–1118. https://doi.org/10.1175/mwr2904.1

Hague, M. J., & Patterson, D. A. (2014). Evaluation of statistical river temperature forecast models for fisheries management. *North American Journal of Fisheries Management*, *34*(1), 132–146. https://doi.org/10.1080/02755947.2013.847879

Hipsey, M. R., Bruce, L. C., & Hamilton, D. P. (2013). *Aquatic Ecodynamics (AED) model library science manual*. Perth, Australia: University of Western Australia. http://aed.see.uwa.edu.au/research/models/AED/Download/AED_ScienceManual_v4_draft.pdf

Hipsey, M. R., Bruce, L. C., Boon, C., Busch, B., Carey, C. C., Hamilton, D. P., 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*, *12*(1), 473–523. https://doi.org/10.5194/gmd-12-473-2019

Hipsey, M. R., Gal, G., Arhonditsis, G. B., Carey, C. C., Elliott, J. A., Frassl, M. A., et al. (2020). A system of metrics for the assessment and improvement of aquatic ecosystem models. *Environmental Modelling & Software*, *128*. https://doi.org/10.1016/j.envsoft.2020.104697

Huang, B., Langpap, C., & Adams, R. M. (2011). Using instream water temperature forecasts for fisheries management: An application in the Pacific Northwest1. *JAWRA Journal of the American Water Resources Association*, *47*(4), 861–876. https://doi.org/10.1111/j.1752-1688.2011.00562.x

Jia, X., Willard, J., Karpatne, A., Read, J., Zwart, J., Steinbach, M., & Kumar, V. (2019). Physics guided RNNs for Modeling dynamical systems: A case study in simulating Lake temperature profiles. In *Proceedings of the 2019 SIAM International Conference on Data Mining* (pp. 558–566). Philadelphia, PA: Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611975673.63

Jöhnk, K. D., Huisman, J. E. F., Sharples, J., Sommeijer, B. E. N., Visser, P. M., & Stroom, J. M. (2008). Summer heatwaves promote blooms of harmful cyanobacteria. *Global Change Biology*, *14*(3), 495–512. https://doi.org/10.1111/j.1365-2486.2007.01510.x

Klug, J. L., Richardson, D. C., Ewing, H. A., Hargreaves, B. R., Samal, N. R., Vachon, D., et al. (2012). Ecosystem effects of a tropical cyclone on a network of lakes in northeastern North America. *Environmental Science & Technology*, *46*(21), 11,693–11,701. https://doi.org/10.1021/es302063v

Komatsu, E., Fukushima, T., & Harasawa, H. (2007). A modeling approach to forecast the effect of long-term climate change on lake water quality. *Ecological Modelling*, *209*(2–4), 351–366. https://doi.org/10.1016/j.ecolmodel.2007.07.021

Kumar, J., Brooks, B.-G. J., Thornton, P. E., & Dietze, M. C. (2012). Sub-daily statistical downscaling of meteorological variables using neural networks. *Procedia Computer Science*, *9*, 887–896. https://doi.org/10.1016/j.procs.2012.04.095

Loos, S., Shin, C. M., Sumihar, J., Kim, K., Cho, J., & Weerts, A. H. (2020). Ensemble data assimilation methods for improving river water quality forecasting accuracy. *Water Research*, *171*, 115343. https://doi.org/10.1016/j.watres.2019.115343

Luo, Y., Ogle, K., Tucker, C., Fei, S., Gao, C., LaDeau, S., et al. (2011). Ecological forecasting and data assimilation in a data-rich era. *Ecological Applications*, *21*(5), 1429–1442. https://doi.org/10.1890/09-1275.1

McClure, R. P., Hamre, K. D., Niederlehner, B. R., Munger, Z. W., Chen, S., Lofton, M. E., et al. (2018). Metalimnetic oxygen minima alter the vertical profiles of carbon dioxide and methane in a managed freshwater reservoir. *Science of the Total Environment*, *636*, 610–620. https://doi.org/10.1016/j.scitotenv.2018.04.255

Mestekemper, T., Windmann, M., & Kauermann, G. (2010). Functional hourly forecasting of water temperature. *International Journal of Forecasting*, *26*(4), 684–699. https://doi.org/10.1016/j.ijforecast.2009.10.010

Mi, C., Sadeghian, A., Lindenschmidt, K.-E., & Rinke, K. (2019). Variable withdrawal elevations as a management tool to counter the effects of climate warming in Germany's largest drinking water reservoir. *Environmental Sciences Europe*, *31*(1), 1–15. https://doi.org/10.1186/s12302-019-0202-4

Millennium Ecosystem Assessment (2005). *Ecosystems and human well-being: Synthesis*. Washington, DC: Island Press.

Morss, R. E., Demuth, J. L., & Lazo, J. K. (2008). Communicating uncertainty in weather forecasts: A survey of the U.S. *Public, Weather and Forecasting*, *23*(5), 974–991. https://doi.org/10.1175/2008waf2007088.1

NCAR - Research Applications Laboratory (2015). *verification: Weather Forecast Verification Utilities (R package version 1.42)*. CRAN. https://CRAN.R-project.org/package=verification

Niemeyer, R. J., Cheng, Y., Mao, Y., Yearsley, J. R., & Nijssen, B. (2018). A thermally stratified reservoir module for large-scale distributed stream temperature models with application in the Tennessee River basin. *Water Resources Research*, *54*, 8103–8119. https://doi.org/10.1029/2018WR022615

Ouellet-Proulx, S., Chimi Chiadjeu, O., Boucher, M.-A., & 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

Ouellet-Proulx, S., St-Hilaire, A., & Boucher, M.-A. (2017). Water temperature ensemble forecasts: Implementation using the CEQUEAU model on two contrasted river systems. *Watermark*, *9*(7). https://doi.org/10.3390/w9070457

Page, T., Smith, P. J., Beven, K. J., Jones, I. D., Elliott, J. A., Maberly, S. C., et al. (2018). Adaptive forecasting of phytoplankton communities. *Water Research*, *134*, 74–85. https://doi.org/10.1016/j.watres.2018.01.046

Pike, A., Danner, E., Boughton, D., Melton, F., Nemani, R., Rajagopalan, B., & Lindley, S. (2013). Forecasting river temperatures in real time using a stochastic dynamics approach. *Water Resources Research*, *49*, 5168–5182. https://doi.org/10.1002/wrcr.20389

Read, J. S., Hamilton, D. P., Jones, I. D., Muraoka, K., Winslow, L. A., Kroiss, R., et al. (2011). Derivation of lake mixing and stratification indices from high-resolution lake buoy data. *Environmental Modelling and Software*, *26*(11), 1325–1336. https://doi.org/10.1016/j.envsoft.2011.05.006

Read, J. S., Jia, X., Willard, J., Appling, A. P., Zwart, J. A., Oliver, S. K., et al. (2019). Process-guided deep learning predictions of lake water temperature. *Water Resources Research*, *55*, 9173–9190. https://doi.org/10.1029/2019WR024922

Roberts, D. C., Forrest, A. L., Sahoo, G. B., Hook, S. J., & Schladow, S. G. (2018). Snowmelt timing as a determinant of lake inflow mixing. *Water Resources Research*, *54*, 1237–1251. http://doi.org/10.1002/2017WR021977

Schmidt, S. R., Gerten, D., Hintze, T., Lischeid, G., Livingstone, D. M., & Adrian, R. (2018). Temporal and spatial scales of water temperature variability as an indicator for mixing in a polymictic lake. *Inland Waters*, *8*(1), 82–95. https://doi.org/10.1080/20442041.2018.1429067

Sharma, S., Gray, D. K., Read, J. S., O'Reilly, C. M., Schneider, P., Qudrat, A., et al. (2015). A global database of lake surface temperatures collected by in situ and satellite methods from 1985–2009. *ScientificData*, *2*, 150008. https://doi.org/10.1038/sdata.2015.8

Snortheim, C. A., Hanson, P. C., McMahon, K. D., Read, J. S., Carey, C. C., & Dugan, H. A. (2017). Meteorological drivers of hypolimnetic anoxia in a eutrophic, north temperate lake. *Ecological Modelling*, *343*, 39–53. https://doi.org/10.1016/j.ecolmodel.2016.10.014

Stocker, F. T., Qin, D., Plattner, G. K., Tignor, M., Allen, S. K., Boschung, J., et al. (2013). *Climate change 2013: The physical science basis. Contribution of Working Group I to the fifth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Thomas, R. Q., Figueiredo, R. J., Daneshmand, V., Puckett, L. K., & Carey, C. C. (2020). Forecasting Lakes and Reservoir Ecosystems (FLARE) (Version v1.11). Zenodo. https://doi.org/10.5281/zenodo.3862905

Ummenhofer, C. C., & Meehl, G. A. (2017). Extreme weather and climate events with ecological relevance: A review. *Philosophical Transactions of the Royal Society B*, *372*(1723), 20,160,135–20,160,113. https://doi.org/10.1098/rstb.2016.0135

Ward, N. K., Steele, B. G., Weathers, K. C., Cottingham, K. L., Ewing, H. A., Hanson, P. C., & Carey, C. C. (2020). Differential responses of maximum versus median chlorophyll-a to air temperature and nutrient loads in an oligotrophic lake over 31 years. *Water Resources Research*, *56*, e2020WR027296. https://doi.org/10.1029/2020WR027296

Weber, M., Rinke, K., Hipsey, M. R., & Boehrer, B. (2017). Optimizing withdrawal from drinking water reservoirs to reduce downstream temperature pollution and reservoir hypoxia. *Journal of Environmental Management*, *197*, 96–105. https://doi.org/10.1016/j.jenvman.2017.03.020

West, J. M., Julius, S. H., Kareiva, P., Enquist, C., Lawler, J. J., Petersen, B., et al. (2009). U.S. natural resources and climate change: Concepts and approaches for management adaptation. *Environmental Management*, *44*(6), 1001–1021. https://doi.org/10.1007/s00267-009-9345-1

Woolway, R. I., Maberly, S. C., Jones, I. D., & Feuchtmayr, H. (2014). A novel method for estimating the onset of thermal stratification in lakes from surface water measurements. *Water Resources Research*, *50*(6), 5131–5140. https://doi.org/10.1002/2013WR014975

Yigzaw, W., Li, H. Y., Demissie, Y., Hejazi, M. I., Leung, L. R., Voisin, N., & Payn, R. (2018). A new global storage-area-depth data set for Modeling reservoirs in land surface and earth system models. *Water Resources Research*, *54*, 10,372–10,386. https://doi.org/10.1029/2017WR022040

Yigzaw, W., Li, H. Y., Fang, X., Leung, L. R., Voisin, N., Hejazi, M. I., & Demissie, Y. (2019). A multilayer reservoir thermal stratification module for earth system models. *Journal of Advances in Modeling Earth Systems*, *11*(10), 3265–3283. https://doi.org/10.1029/2019ms001632

Zhang, H., Hendricks Franssen, H.-J., Han, X., Vrugt, J. A., & Vereecken, H. (2017). State and parameter estimation of two land surface models using the ensemble Kalman filter and the particle filter. *Hydrology and Earth System Sciences*, *21*(9), 4927–4958. https://doi.org/10.5194/hess-21-4927-2017