**A near-term, iterative forecasting workflow to predict reservoir methane ebullition rates and characterize forecast uncertainty**

Ryan P. McClure1, R. Quinn Thomas2, Mary E. Lofton1, Whitney M. Woelmer1 and Cayelan C. Carey1

1Department of Biological Sciences, Virginia Tech, Blacksburg, VA, 24061, USA

2Department of Forest Resources and Environmental Conservation, Virginia Tech, Blacksburg, VA, 24061, USA

*Correspondence to*: Ryan P. McClure (ryan333@vt.edu)

**Abstract.** Near-term, iterative ecological forecasting with data assimilation has shown great promise for improving our understanding of ecological dynamics, but to date has rarely been applied to predicting biogeochemical variables. Most ecological forecasting workflows use automated, high-frequency sensor data to iteratively update predictions in near-real time as new observations become available. However, many ecosystem-level biogeochemical processes cannot be quantified reliably with automated sensors and instead depend on manually collected samples and laboratory analysis, representing a major gap in our ability to predict some ecosystem-level processes. One important biogeochemical process that near-term, iterative ecological forecasts could potentially improve estimates of – yet is challenging to both monitor and predict – is freshwater methane (CH4) ebullition (CH4 bubbling from the sediments). We developed a forecasting and data assimilation workflow that generated real-time CH4 ebullition rate forecasts and then partitioned forecast uncertainty among different sources (parameters, model process, initial conditions, and driver data). We applied the forecasting system to a small eutrophic reservoir for a summer season and found that the workflow successfully predicted an ensemble of future CH4 ebullition rates that encompassed out-of-sample observations, with initial conditions and parameter uncertainty dominating total forecast uncertainty. Despite the high variability of CH4 ebullition emissions in freshwater ecosystems, our iterative forecasting workflow demonstrates that ebullition in shallow zones within a freshwater reservoir can be successfully forecasted and the sources of uncertainty can be characterized to better improve the forecasts, suggesting that ecological forecasting is a novel approach to improve our predictions of freshwater ebullition rates in the future.

# 1 Introduction

Near-term, iterative ecological forecasting with data assimilation is a novel approach to improve our understanding and quantification of ecosystem processes (Dietze et al. 2018). Here, we define an ecological forecast as a prediction with specified uncertainties of the future state of an ecosystem in the near-term (days to months) that is iteratively updated as new data become available (Clark et al. 2001, Dietze 2017*a*). Similarly, we define data assimilation as iteratively incorporating observations as they are collected into a forecast model by updating model states and parameter estimates (Luo et al. 2011). The iterative process of making a forecast, collecting observations to compare with the forecast, and then updating forecast models via data assimilation before making a new forecast is referred to as a forecast cycle (Dietze 2017*a*). Iterative ecological forecasting can substantially improve our understanding of ecosystem processes because it embodies the scientific method: our hypotheses about ecosystem functioning are instantiated in a model, which generates predictions of the future that are then tested with observed data when they become available (Dietze et al. 2018). Thus, near-term, iterative ecological forecasting creates a model-data feedback loop that evaluates how effectively a model predicts future ecosystem states, with the forecasts evolving as the ecosystem experiences different environmental conditions.

Because of the need to constantly update forecasts with new data as they become available, most iterative ecological forecasts have been generated for variables that can be measured with high-frequency sensors providing data in real-time or near real-time, thereby enabling rapid forecast updating on minute to daily time scales (e.g., Dietze 2017*b*, Barachinni et al. 2020, Nickl et al. 2020). For example, Dietze (2017*b*) used automated, high-frequency eddy-flux covariance sensor estimates of carbon fluxes in a deciduous forest to iteratively forecast net ecosystem exchange. Data were collected by sensors on 30-minute intervals and transmitted wirelessly through automated cyberinfrastructure to update future forecasts. Similar automated workflows have been implemented to forecast other variables that can be monitored with sensors, such as lake chlorophyll *a* (Page et al. 2018), water temperature (Baracchini et al. 2020, Thomas et al. 2020), and stream discharge (Ouellet-Proulx et al. 2017). However, many ecosystem-level biogeochemical states and processes (e.g., nutrient and carbon concentrations and fluxes) are still quantified using manual data collection, not sensors (Marcé et al. 2016). Additionally, because technology for automated high-frequency monitoring of many ecosystem processes is not affordable or fully developed (e.g., Porter et al. 2009, Marcé et al. 2016), there is a major need to develop new forecasting workflows which integrate manually collected data streams for iterative forecasting.

One biogeochemical process that near-term, iterative ecological forecasts could potentially improve estimates of **–** yet is challenging to both monitor and predict **–** is freshwater methane (CH4) ebullition, or bubble fluxes of CH4 from organic-rich sediments to the waterbody's surface. Freshwater ecosystems emit large quantities of CH4 to the atmosphere (currently estimated between 117 and 212 Tg CH4 yr-1; Saunois et al. 2020). Among the different types of freshwater CH4 emissions, estimates of ebullition are considered one of the most uncertain in regional CH4 budgets because of ebullition's high spatial and temporal variability within and among freshwater ecosystems (Wik et al. 2016, Saunois et al. 2020). Spatially, ebullition rates can vary substantially at sites located just a few meters apart (DelSontro et al. 2011, Beaulieu et al. 2016, Wik et al. 2016, McClure et al. 2020, Linkhorst et al. 2020). Temporally, CH4 ebullition rates within the same waterbody can vary substantially from daily to interannual time scales (Burke et al., 2019, Männistö et al. 2019, Linkhorst et al. 2020). Because of the high spatiotemporal variation of ebullition, it is extremely challenging to both monitor and predict, which is critically important for upscaling estimates (Wik et al. 2016). As a result, freshwater ebullition remains unaccounted for in the current global CH4 budget (Saunois et al. 2020).

Because freshwater CH4 ebullition is a highly variable flux for global emission estimates (Maher et al. 2019, Saunois et al. 2020), new methods are needed to evaluate how predictable CH4 ebullition emission rates are over shorter time scales (Butman et al. 2018, Saunois et al. 2020). Near-term iterative forecasting with data assimilation may be a particularly valuable method for estimating and evaluating CH4 ebullition predictability because this approach has been shown to successfully predict other CH4 emission pathways. For example, Nickl et al. (2020) developed iterative forecasts of CH4 emissions to the atmosphere from coal mining regions in Poland that predicted CH4 plumes up to four days into the future at the 7 km2 scale. Their CH4 emission forecast workflow used Earth system models that were updated with remote sensing data and airborne sampling quantifying atmospheric concentrations of CH4 (Nickl et al. 2020).

In addition to improving CH4 ebullition predictions, near-term iterative forecasting using data assimilation and partitioning sources of uncertainty in the forecasts also has the potential to improve our basic understanding of the drivers of CH4 ebullition dynamics. There are multiple variables which drive freshwater CH4 ebullition rates at different spatial and temporal scales, including water temperature (DelSontro et al. 2016, Davidson et al. 2018, McClure et al. 2020), hydrostatic pressure and water table depth (Harrison et al. 2016, Liu et al. 2019), primary productivity (West et al. 2016), water turbulence (Joyce and Jewel 2002), sediment organic matter and organic matter quality (Wik et al. 2018, Zhou et al. 2019), plant abundance (Davidson et al. 2019), and zooplankton bioturbation (Bezerra et al. 2020). The relative importance of these predictor variables can vary substantially over days (Maher et al. 2019, Bezerra et al. 2020), seasons (DelSontro et al. 2016, McClure et al. 2020), and years (Männistö et al. 2019, Linkhorst et al. 2020). Developing iterative, near-term forecasts in which new observations regularly update a forecasting model could provide useful information on how the importance of predictors change under different environmental conditions, as indicated by the evolution of the model parameters and model performance over time.

In contrast to atmospheric CH4 plume estimates spanning large regions (e.g., Nickl et al. 2020), freshwater CH4 ebullition rates are usually estimated individually for single ecosystems (e.g., lake or wetland) and rarely are monitored using continuous remote sensing or live-streaming sensor data, necessitating an iterative forecasting workflow using manually collected data and empirical models. While there have been recent advances in monitoring CH4 emissions using automated CH4 ebullition rate sensors (Delwiche et al. 2017, Maher et al. 2019) and remote sensing (Engram et al. 2020), the vast majority of lake and reservoir CH4 ebullition emissions are still measured with passive traps that are deployed in one waterbody, retrieved manually, and later analyzed on a gas chromatograph (GC) in the laboratory (e.g., Klapstein et al. 2014, Beaulieu et al. 2016, Männistö et al. 2019). Consequently, to advance iterative forecasting of manually sampled CH4 ebullition, new iterative forecasting and data assimilation workflows must be developed.

As a result, we developed the first (to our knowledge) iterative forecast workflow for near-term CH4 ebullition rates in a freshwater ecosystem, using manually collected ebullition data. We used sequential data assimilation to update a time series ebullition model's states, parameters, initial conditions, and driver data on each daily model time step. Here, we refer to sequential data assimilation as a method in which newly acquired observations are used to refit a model before generating a new forecast (Dietze 2017*a*).

We developed and tested our CH4 ebullition rate forecast workflow in a small, eutrophic freshwater reservoir in southwestern Virginia, USA (Figure 1). Our objectives were: 1) develop an iterative, near-term forecasting workflow for manually collected CH4 ebullition rates; 2) assess the performance of forecasted future CH4 ebullition rates in comparison to observations and a persistence null model; and 3)partition the sources of uncertainty in our forecast workflow as a means to improve future CH4 ebullition rate forecasts. Altogether, our goal was to use near-term iterative forecasting to advance our understanding of CH4 ebullition dynamics, specifically, to both quantify the predictability of ebullition as well as examine how its drivers change over time.

## 2 Methods

## 2.1 Site description

We developed near-term, iterative CH4 ebullition rate forecasts in Falling Creek Reservoir (FCR, Figure 1) in real-time in summer 2019. FCR is a small (0.119 km2), shallow (Zmax = 9.3 m), eutrophic, drinking water reservoir located in southwestern Virginia, USA (37.30°N, 79.84°W). FCR is owned and operated by the Western Virginia Water Authority as a drinking water supply and is located in a completely forested watershed (Gerling et al. 2016).

The forecasting system leveraged two years of CH4 ebullition monitoring data collected at FCR. In 2017, four ebullition traps were deployed and monitored weekly from 8 May until 24 October along a shallow upstream transect, as well as 12 other sites in the reservoir (McClure et al. 2020, Figure 1). In 2018, two ebullition traps were deployed and monitored weekly from 7 May to 29 October at the same shallow upstream transect, as close as possible to the two middle traps from the transect in 2017 (Traps 2 and 3 in Figure 1).

In 2019, the year of this forecasting study, we redeployed all four traps as close as possible to their original four locations in 2017. Our first forecasting cycle began on 27 May, with forecast evaluation between 27 May to 7 November, a 164-day forecast period. FCR’s water level was managed by the Water Authority and did not experience large fluctuations during 2017-2019 (Carey et al. 2020*a*).

## 2.2 Forecast model development

A previous study by McClure et al. (2020) analyzed the May to October 2017 historical monitoring data and quantified the rates and drivers of CH4 ebullition at multiple sites in FCR. That study demonstrated that ≥60% of the total reservoir-wide CH4 ebullition was emitted from the shallow upstream transect in FCR during the ice-free period, with a strong positive relationship between ebullition and sediment-water interface (SWI) temperatures (Figure 1; McClure et al. 2020). That analysis developed an auto-regressive (AR) time series model (Eqn. 1) that predicted weekly CH4 ebullition aggregated across the four traps at the upstream transect (Supporting figures, Figure S1; McClure et al. 2020):

(1)

where (mg CH4 m-2 d-1) is the predicted CH4 ebullition rate at the upstream transect, (mg CH4 m-2 d-1) is the observed CH4 ebullition rate at the previous time step (the AR term), (°C) is the water temperature averaged from measurements at the SWI below the upstream transect between each sampling date, and is the stochastic error term; see McClure et al. 2020. is the intercept term, is the parameter governing the effect of the autoregressive term, and is the parameter governing the effect of SWI temperature. In the McClure et al. (2020) analysis, these three parameters were fixed throughout the May to October 2017 period, at = -6.58±1.27 (1 S.E.), = 0.26±0.05, and = 0.38±0.05. As our focus in this forecasting analysis was on the development of an iterative forecast workflow, not model selection, we used this same AR model structure to iteratively forecast future CH4 ebullition rates in FCR in summer 2019.

Because we were generating forecasts in real-time, we needed forecasted covariate data (i.e., future SWI temperatures) to generate forecasts of future CH4 ebullition rates of all four traps on the transect (Figure 1). Thus, we leveraged the existing Forecasting Lake And Reservoir Ecosystems (FLARE; Thomas et al. 2020) water temperature forecasting framework and infrastructure deployed in FCR to generate predictions of future water temperatures that became the driver data for our CH4 ebullition rate forecasts. FLARE is composed of three components: water quality and meteorological sensors that wirelessly stream data, a data assimilation algorithm that uses sensor observations to update the predictions and re-calibrate the model’s parameters daily, and an ensemble forecast algorithm that generates an ensemble water temperature forecasts at 29 depths on 0.33 m increments from the surface to sediments at the deepest site of the reservoir (the dam site) (Thomas et al. 2020). The hydrodynamic-water temperature model in FLARE (General Lake Model or GLM; Hipsey et al. 2019) uses NOAA GEFS meteorological forecasts as driver data. We refer interested readers to Thomas et al. (2020) and <https://smartreservoir.org/user_guide/> for a detailed description of FLARE and tutorial on the FLARE forecasting system.

Our forecasts predicted future CH4 ebullition rates at a daily time-step over a 10-day forecast horizon (the time into the future forecasts extended) using FLARE’s forecasted water temperatures between 2 and 3m, which encompassed the same depths as the sediments below all four upstream ebullition traps. Because the temperature forecasts from FLARE were generated ~700m downstream from the ebullition traps (Figure 1), we developed a model that estimated the upstream SWI temperatures at the ebullition transect. Details on the model that scaled FLARE’s forecasted water temperatures to the SWI temperatures upstream can be found in supporting information A. Thus, to generate CH4 ebullition rate forecasts we input the water temperature forecasts from FLARE into our temperature scaling model (Supporting Information A), in which the estimated water temperatures from the scaling model were then used as driver data for a state-spaced forecast model (see below) developed to predict future CH4 ebullition rates.

## 2.3 State-space forecast model structure for CH4 ebullition rates

We used a Hierarchical Hidden Markov model (henceforth referred to as a HHM model; Dietze 2017a) with forecasted water temperatures derived from FLARE as a covariate to generate future CH4 ebullition rates and quantify the contribution of different sources of uncertainty. Our HHM model used a Markov Chain Monte Carlo (MCMC) analyses to estimate the uncertainty that naturally arises from modeling the process uncertainty and the uncertainty of the observations of the process (Hobbs and Hooten, 2015). For every forecast, we conducted an MCMC analysis and then used the posterior predictive distributions of the estimated out-of-sample ebullition rates, model parameters, model process error, the model initial conditions, and driver data to produce forecasts of CH4 ebullition rates 10-days into the future. Importantly, this approach allowed us to quantify the contribution of each uncertainty source in our model throughout all iterative forecast cycles (parameters, process, initial conditions, driver data).

Our HHM model included a process model depicted by Equation 2 that mirrored the original ebullition model in Eqn. 1:

(2)

where is the predicted future ebullition rate, is the current ebullition rate at time *f*, is the additive process error, are the model parameters, and represents the forecasted covariate (scaled water temperature from FLARE) used to drive the model. In the above equation, the subscript *f* is the day (0 to 10) into the 10-day forecast horizon. The index *t* (as in ) is different from *f* because it represents the number of days since the initiation of the forecasts and increases each day throughout a multiyear forecast period, while *f* is reset to 0 each day before initiating a new 1 to 10-day forecast.

Next, because our observations in Equation 2 are unable to account for the true states of the process model, we developed independent observation models that estimated the unobserved latent states:

(3)

(4)

where in Equation 3 and in Equation 4 represent the unobserved latent state of CH4 ebullition rates and forecasted water temperatures, respectively. and represent the mean of the ebullition rates at the transect and 210 forecasted water temperature ensembles while and represent the standard deviation of the mean of the transect level ebullition rates and forecasted water temperature from FLARE, respectively.

For our MCMC analyses, we ran three MCMC chains with a burn-in period of 5,000 iterations and a sample size of 20,000 iterations. Prior distributions for the process model (Equation 2) parameters () were pulled from a normal distribution shown in Equation 5.

(5)

Process error depicted in Equation 2 was estimated using Equation 6, where is drawn from a uniform distribution defined by Equation 7.

(6)

(7)

The predicted posterior output from the MCMC distributions for all parameters, process error, and latent states, were saved from each forecast to quantify uncertainty of the forecasts on each day through the forecasting period.

In addition to our HHM model (Equations 2 - 7), we also developed a similar Hidden Markov model that assumed the CH4 ebullition rates would remain similar into the future with propagated uncertainty (henceforth referred to as “persistence null model”). The persistence null mode was depicted as:

(8)

where is the contribution of process uncertainty to total forecast uncertainty and was randomly drawn from a uniform distribution defined by Equations 6 and 7.

The MCMC chains from each forecast were assessed for convergence using the potential scale reduction factor of the Gelman-Rubin statistic (). Finally, we used the ‘rjags’ and ‘R2jags’ package (Plummer 2019) within the R statistical environment for the model development (R Core Development Team 2021). Importantly, the temperature scaling model that estimated the upstream SWI temperatures based off of FLARE’s forecasted water temperatures was similarly structured as our HHM model described above and can be found in Supporting Information A. Finally, the jags model code that was used for the temperature scaling model, the HHM model to forecast CH4 ebullition rates, and the persistence null model to also forecast CH4 ebullition rates can be found in Supporting Information B.

## 2.3 Field data collection

Field sampling to update the forecast model with sequential data assimilation happened every 5-9 days during summer 2019 (Figure 2). Our iterative forecast workflow encompassed data collection, which was used for sequential data assimilation

We sampled the CH4 ebullition rates at the traps following McClure et al. (2020). The ebullition samples from each trap were extracted across a septum stopper using a needle attached to a 10-mL syringe. We injected 10 mL of ebullition gas into a 20-mL crimped top glass vial that was pre-filled with saturated salt brine solution. A secondary exit syringe extracted the salt brine solution as the sample was injected to generate 10 mL of gas headspace in the vial. If enough gas sample was available, two replicates were collected from each trap on a sampling day. The vials were stored upside down until analysis, so the remaining 10 mL of salt brine solution acted as a barrier to prevent any gas from escaping. We extracted any remaining gas from each ebullition trap using a 30 mL syringe and summed the total volume of ebullition gas collected each week.

We analyzed the manually collected ebullition gas from the traps for its CH4 concentration using a Shimadzu Nexus-2030 Gas Chromatography – Flame Ionization Detector (GC-FID; Shimadzu Corporation; Kyoto, Japan) within 24 hours of collection. The gas samples were analysed using a gas chromatograph coupled with a flame ionization detector (GC-FID) within 24 hours of collection (following McClure et al. 2018, 2020). We determined the directly observed CH4 ebullition rate (Et) as follows:

(9)

where was the volume of ebullition collected in the trap (liters), is the CH4 concentration of the gas (mg CH4 L−1), is the duration of time the trap was deployed (in days), and is the cross-sectional area of the funnel (0.26 m2). As with the forecasted ebullition rates, the observed ebullition rate therefore represents an integration of the total gas captured per unit area during . Following McClure at al. (2020), we calculated the daily ebullition rate separately for each trap every week and then averaged the rates from the four traps within the transect to determine a mean daily transect ebullition rate.

## 2.4 Forecast evaluation

We tested how well the CH4 ebullition rate forecasts of our HHM and persistence null model at the transect performed against out-of-sample observations via hindcasts (forecasts generated in the past) between 27 May and 7 November. First, we calculated forecast bias to evaluate the performance of our HHM and persistence null models to observations. Bias was assessed using the difference of the mean forecasted CH4 ebullition rates that corresponded with the observed CH4 ebullition rate, following Equation 10:

(10)

where was the posterior mean of the CH4 ebullition rate forecasts and was the observed transect-level mean CH4 ebullition rates. Importantly, was determined separately for both the HHM model and persistence null model forecasts to evaluate forecast performance.

In addition to , we then used Nash-Sutcliffe efficiency (NSE) to evaluate the performance of the HHM and persistence null model’s CH4 ebullition rate forecasts to compare our out-of-sample observations to the forecasts aggregated between 27 May and 7 November. The NSE is a normalized metric that evaluates a model’s performance relative to the observed time series to evaluate how well the forecast predicted the observed time series (Nash and Sutcliffe 1970). We calculated the NSE coefficient (Nash and Sutcliffe 1970) as follows:

(11)

where was the mean of the CH4 ebullition rate forecast ensembles, was the observed mean daily CH4 ebullition rates, and is the average of all the observed mean daily CH4 ebullition rates. NSE values range between -∞ to 1, where 1 indicates a perfect score (i.e., the model perfectly recreates observations), a value of 0 indicates that the model predictions are as accurate as the mean of the observations, and a value <0 indicates that the model performs worse than the observed mean (Nash and Sutcliffe 1970, Moriasi et al. 2007).

## 2.5 Forecast variance and uncertainty partitioning

We quantified the total daily forecast uncertainty in the CH4 ebullition forecasts each day of all 24 forecast cycles. Total forecast uncertainty is a valuable metric of predictability and provides insight to the overall predictability of an ecosystem process (Petchey et al. 2015, Dietze 2017*a*, Dietze 2017*b*). Additionally, we partitioned the relative contributions of four different sources of uncertainty including model parameter, model process, driver data, and initial conditions (ebullition latent state) uncertainty (Dietze 2017*b*) daily among all 24 forecast cycles. Partitioning uncertainty sources can help inform specific improvements to be made to forecasting systems (Dietze 2017*b*, Petchey et al. 2015), and may be particularly useful for developing future forecasts of biogeochemical processes that still rely on manual data collection, like CH4 ebullition rates (following Carey et al. 2021).

We used a One-At-a-Time Sensitivity (OATS) analysis to determine the relative contribution of four uncertainty sources to total uncertainty during each forecast cycle evaluated. An OATS analysis holds all sources of uncertainty at their mean except for one, and then numerically evaluates the sensitivity of the forecast to that specific source of uncertainty (Dietze 2017*a*). First, we quantified the total CH4 ebullition rate forecast variance using the var() function in the R statistical environment R (R Core Development Team, 2020). Next, we quantified the relative contribution of parameters (), model process (), driver data (), and initial conditions () uncertainty to the total CH4 ebullition rate forecast uncertainty. Finally, the relative proportion of each source of uncertainty was determined by dividing the total variance of each isolated source of uncertainty by the sum of the variance of all uncertainty sources.

## 2.5 Forecast archiving and reproducibility

All forecasting driver data (EDI REFS) and code needed to predict CH4 ebullition rates are archived at the Environmental Data Initiative (EDI) repository and Zenodo ([XXX](https://github.com/ryanmclake/CH4cast)) and run in the R statistical environment (R Core Development Team 2021).

## 3 Results

## 3.1 Near-term iterative CH4 ebullition forecasts

We successfully developed near-term, iterative CH4 ebullition rate forecasts that predicted future CH4 ebullition rates at FCR’s upstream transect (Figure 3). For the HHM model, the out-of-sample CH4 ebullition rate observations were within the 90% percentiles of the forecast ensemble members for 23 of the 24 forecast cycles evaluated between 27 May and 7 November 2019 (Figure 3A). In comparison, while the persistence null forecasts encompassed the observations within the 90% percentile of its ensemble members over all 24 forecast cycles, the null forecasts had substantially higher 90% posterior interval quantile maximum and minimum ranges (Figure 3B).

## 3.2 Observed upstream transect CH4 ebullition rates

We observed high temporal variation in mean daily CH4 ebullition rates at FCR’s upstream transect during our 164-day forecasting period (Figure S2). Daily ebullition rates at the upstream increased from 0.8 mg CH4 m-2 d-1 to 14 mg CH4 m-2 d-1 from 27 May though 24 June (Figure S2). Between 24 June and 15 July, daily CH4 ebullition rates from the transect continued to increase until the peak maximum CH4 ebullition rate for the forecast period was observed on 15 July (49 mg CH4 m-2 d-1). After 15 July and until 16 October, the observed CH4 ebullition rates ranged between 13 and 42 mg CH4 m-2 d-1. After 16 October, CH4 ebullition rates dropped to ≤4.6 mg CH4 m-2 d-1 for the remainder of the forecasting period.

## 3.3 CH4 ebullition forecast performance evaluation

The CH4 ebullition forecasts using the HHM model performed worse than the persistence null among the 24 forecast cycles evaluated (Figure 4). The NSE of the persistence null model was 0.69, higher than the NSE of the HHM model (0.38). It is important to note, however, that although the persistence null model performed better than the HHM model, the NSE of both models were higher than 0, indicating that both model predictions were more accurate than the mean of the observations.

In addition, the forecast bias of all 24 evaluated forecast cycles showed that the persistence null model performed better than HHM model (Figure 4). The mean bias of all 24 forecast cycles for the persistence null model was 0.02 while the mean bias of the HHM model was 0.57, indicating that our HHM model overestimated future ebullition rates compared to the persistence null model (Figure 5A).

Despite the overall overestimate from HHM model, a time series of the forecast bias between both models showed that there were forecast cycles when the HHM model outperformed the persistence null model (Figure 5B). For example, the persistence null model underestimated the forecasted ebullition rates between 27 May and 11 August, the same period there was a consistent increase in the transect-level ebullition rates, and then consistently overestimated the forecasted ebullition rates as they gradually decreased between 11 August and the final forecast on 7 November. Conversely, the HHM model overestimated ebullition rates throughout most of the forecast period but then closely matched the forecasted ebullition rates between 11 September and 30 October, the same time period when temperatures declined and the reservoir experienced seasonal turnover, which occurred on 23 October.

## 3.4 Forecast horizon uncertainty

Although the persistence null model outperformed our HHM model based on forecast bias and NSE evaluation metrics, the HHM model had lower overall uncertainty at each daily time step across our 10-day forecast horizon (Figure 6). By the final forecast on 7 November, the HHM forecasts had substantially reduced in total forecast variance. Among all HHM model forecasts evaluated, the forecast variance one day into the future ranged had a mean of 0.5 ln(mg CH4 m-2 d-1) (range = 0.3-1.2 ln(mg CH4 m-2 d-1)) (Figure 6A). However, two days into the future, the mean forecast variance among all forecasts evaluated doubled to 1.1 ln(mg CH4 m-2 d-1), with a range between 0.75 and 2.0 ln(mg CH4 m-2 d-1). By the end of the 10-day forecast horizon, the mean forecast variance among all forecasts evaluated increased to 2.2 ln(mg CH4 m-2 d-1), with a range between 1.9 and 3.2 ln(mg CH4 m-2 d-1).

Conversely, the persistence null model started with high total forecast variance one day into the future, with a mean of 1.4 ln(mg CH4 m-2 d-1) and range between 0.5 and 7 ln(mg CH4 m-2 d-1) (Figure 6B). Like the HHM model, the total variance increased with the number of days into the forecast horizon and by 10-days ahead, the mean variance among all forecasts evaluated increased to 4 ln(mg CH4 m-2 d-1) with a range between 3 and 11 ln(mg CH4 m-2 d-1), substantially higher than the forecasts.

## 3.5 HHM model uncertainty partitioning

Aggregated across the forecasting period, we found that initial condition uncertainty contributed the lowest proportion to total variance on the first day of the forecast but then became the largest source of uncertainty as there were no “true” observations of ebullition to update the HHM model within the 10-day forecast horizon (Figure 6A). Thus, initial condition uncertainty accounted for the largest source of uncertainty across all 24 forecast cycles, contributing a mean of 56% to total forecast uncertainty (Figure 6B). While initial condition uncertainty was low on the first day of all 24 10-day forecast horizons evaluated, parameter and process uncertainty contributed the most uncertainty on the first and second days of the forecasts until initial conditions of the unobserved ebullition rates became the largest source of uncertainty (Figure 6A). When evaluated across all 24 forecast cycles, parameter uncertainty was the next highest contribution behind initial conditions with a mean of 30% variance followed by model process uncertainty at 12% variance (Figure 6B). Driver data uncertainty was low throughout all 24 forecast cycles and contributed the remaining 2% to forecast uncertainty among all 24 forecast cycles (Figure 6B).

There were also a few forecast cycles that did not match the general pattern observed within all 24 forecast cycles. For example, the forecast cycles that corresponded with 17 June to 27 June, 24 June to 4 July, 15 July to 25 July, and 27 September to 7 October had low overall initial condition uncertainty but parameter and process uncertainty contributed ~80% to total uncertainty for these two forecast cycles (Figure 6A). ADD IN SEVERAL SENTENCES HERE ABOUT WHAT WAS SIMILAR ABOUT THESE WEEKS AND HOW THEY DIFFERED FROM THE OTHER WEEKS Additionally, the final six forecast cycles that all occurred between 2 October and 7 November, the contribution to parameter uncertainty was less than 25% while the contributions from process uncertainty and initial conditions increased.

## 3.6 HHM model parameter evolution

Finally, we observed shifts in the temperature covariate model parameter estimates during our forecast evaluation cycle (Figure 8). The range for , the parameter associated with the autoregressive ebullition term in our model (Equation XX), increased throughout the evaluated forecast period. Starting with an ensemble mean of 0.85 ± 0.04 on 27 May and ending the period with an ensemble mean of 0.88 ± 0.02. Some type of transition adverb here, the range for , the parameter associated with the temperature term in our model, started at 0.11 ± 0.03 and then decreased to 0.06 ± 0.01 by the final forecast on 7 November. Finally, the mean model process error () associated with our model, decreased from 0.82 ± 0.13 to 0.65 ± 0.07 by the final forecast evaluated on 7 November.

## 4 Discussion

Our CH4 ebullition forecasting workflow demonstrates that near-term, iterative ecological forecasts with data assimilation can occur without automated sensor systems and provides evidence that CH4 ebullition rates can be predicted on a 10-day forecast horizon using this approach (Figure 3). Because many biogeochemical states and fluxes are still quantified using manual data collection in both aquatic and terrestrial ecosystems, creating near-term, iterative ecological forecasting workflows that use manually collected observations for data assimilation like ours may improve our understanding of their predictability with time. Although a persistence null model performed better than an AR model with temperature as a driver (Figure 4), this approach suggests that iteratively forecasting CH4 ebullition with data assimilation in freshwater ecosystems can improve model performance with time (Figure 5) and respond to changing environmental conditions with time (Figure 6 and 7). Additionally, iterative forecasting using hierarchical models allows the user to partition sources of model uncertainty and provide critical information where the model could be improved or whether the model is appropriate (Figure 6). Altogether, the iterative improvement in our CH4 ebullition rate forecasts throughout a 164-day forecast period (Figure 5), the partitioned uncertainty of our evaluated forecast cycles (Figure 6), and the changes in parameters over the study period (Figure 7) indicate that near-term, iterative forecasting can improve our quantification and understanding of this biogeochemical process.

## 4.1 Predictability of CH4 ebullition rates

Contrary to the expectation that CH4 ebullition is difficult to predict at varying scales (Deemer et al. 2016, Wik et al. 2016, Maher et al. 2019, Saunois et al. 2020), our near-term, iterative ecological forecasts with data assimilation showed that CH4 ebullition at this specific transect in FCR waspredictable up to 10-days into the future. While we note that our forecasts were for just one site of a small reservoir, other studies have also recently developed models that were able to successfully predict ebullition on different temporal and spatial scales, from the mesocosm (Davidson et al. 2018) to whole-ecosystem (Nachimuthu et al. 2016, Walter-Anthony 2013), and on temporal scales varying from hours (Bezerra et al. 2020) to years (Männistö et al. 2019). While these studies generated prediction models after data were collected, not updating their models and predictions in real-time, this previous work in combination with ours suggests that freshwater CH4 ebullition may be more predictable as a biogeochemical process than previously thought. As a start, developing models that include additional forecasted covariates (e.g., chlorophyll *a*, nutrients, pressure; DelSontro et al. 2016, Harrison et al., 2016, West et al. 2016, Davidson et al. 2018) and increase in complexity over our AR forecast model are likely needed to predict CH4 ebullition at other spatial and temporal scales than we present here.

The reduction in total forecast variance over the forecasting period (Figure 5) further demonstrates the value of the iterative forecasting for improving our understanding of CH4 ebullition predictability. Following Petchey et al. (2015), we assessed predictability by using the metric of total forecast variance over the forecasting period. Other studies that followed this approach found that total forecast variance can decrease with time if the chosen prediction model is appropriate (Luo et al. 2011, Petchey et al. 2015, Dietze 2017*b*). The decrease in the daily total forecast variance over our forecast period suggests that our AR model with temperature as a covariate was appropriate for forecasting CH4 ebullition even though it did not perform as well as the persistence null model. For other near-term, iterative forecasting applications of CH4 ebullition, rapid decreases in total forecast variance can be used to check if a chosen predictive model is appropriate (following Dietze 2017b).

## 4.1 CH4 ebullition basic knowledge gain

Importantly, the evolution of model parameters throughout the forecasting period (Figure 7) highlights the ability of our forecasting framework to improve our understanding of how CH4 ebullition rates can respond to seasonal environmental conditions. For example, an increase in the observed ebullition rates (Figure S2) between 1 July and 8 July was concomitant with an increase in the SWI temperature parameter and decrease in the AR parameter (Figure 6) as the reservoir warmed, indicating that temperature is likely an important predictor of CH4 ebullition rates during periods of sediment warming periods while previously observed ebullition rates are less so. However, between 15 and 22 July, the SWI temperature parameter estimate decreased and AR term parameter estimate increased and then remained stationary until 11 October. This shift between the parameter estimates followed by the stationary trend were concomitant with the seasonal increase and decrease in the observed CH4 ebullition rates (Figure S2), indicating that the AR term in our model was the important forecast variable for CH4 ebullition rates throughout the middle of the forecast period when seasonal trends are observed. Finally, after 11 October, the SWI temperature parameter substantially increased again while the AR term parameter decreased and then increased to a similar value that was observed on the first 1 July forecast. This dynamic change in both parameter estimates highlight how SWI temperature is likely most important for predicting CH4 ebullition in spring and late autumn, when there are large changes occurring in SWI temperature week to week as water at the sediment water interface cools during fall mixing. Conversely, the AR term was most important during mid-summer when there are consistent weekly autocorrelated changes in ebullition.

ADD PARAGRAPH HERE ON HOW THE RELATIONSHIP OF CH4 AND TEMP THAT THE MODEL SHOWS VIA PARAMETER EVOLUTION CHANGES OUR UNDERSTANDING OF THE DRIVERS OF CH4 EBULLITION

## 4.1 Uncertainty partitioning provides insight on forecasting system performance

Partitioning different sources of uncertainty (initial conditions, parameter, model process, and driver data; Dietze 2017*b*) in our CH4 ebullition rate forecasts provides valuable information on how best to improve future forecasts that operate on a similar time scale (Figure 5). Process uncertainty was the largest contributor to total forecast uncertainty during the entire forecasting period, followed by driver data uncertainty, then initial conditions uncertainty, and then parameter uncertainty (Figure 5). Process uncertainty encompasses the ability of the model to recreate ecosystem dynamics (Dietze 2017*b*). The dominance of model process uncertainty during the middle of the forecasting period (when the SWI temperature parameter estimate was lowest and the AR term parameter estimate was highest) suggests that additional environmental variables that have been shown to be important predictors of CH4 ebullition in other studies (e.g., chlorophyll *a*, sediment organic quality, water-level/hydrostatic pressure; West et al. 2016, Beaulieu et al. 2018, Wik et al. 2018) into our model would potentially reduce the model process uncertainty and improve the CH4 ebullition forecasts. However, it is important to also recognize that increasing the number of forecasted predictors into our model may reduce model process uncertainty but also increase uncertainty in other areas like driver data and parameter uncertainty, especially if there is a lot of uncertainty in forecasted the driver variables which can be challenging to generate and likely increase total uncertainty for the CH4 ebullition forecast.

We also observed an increase in the relative contribution of driver data uncertainty in the forecasting period (Figure 5). The driver data uncertainty in the CH4 ebullition forecasts was due to both uncertainty in the forecasted water temperatures from FLARE and the SWI temperature scaling model (Figure S5), with most of the uncertainty likely from FLARE’s forecasted water temperatures. Our results are similar to other work who found that driver data uncertainty can contribute a large part of total forecast uncertainty. For example, Thomas et al. (2020) found that meteorological forecasts greater than 2 days into the future contributed the largest proportion to total uncertainty in the water temperature forecasts in the reservoir’s surface waters in the FLARE forecasting system. Similarly, Dietze (2017*b*) found that meteorological driver data uncertainty dominated forest net ecosystem exchange forecasts at longer forecast horizons. The larger driver data uncertainty in our CH4 ebullition rate forecasts is likely a result of propagated meteorological uncertainty from the FLARE forecasting system into the SWI temperature forecasts. Additionally, developing CH4 ebullition forecasts over shorter forecast horizons (e.g. < 2-days) may decrease driver data uncertainty overall but likely increase uncertainty in other areas.

## 4.1 Forecasting system limitations and improvements

There are limitations of our near-term, iterative CH4 ebullition rate forecasts. As noted above, there are many other potential predictors of CH4 ebullition and other model structures that could be integrated into our iterative workflow for forecasting CH4 ebullition in other freshwater ecosystems at different temporal and spatial scales. Because our forecasts used only one statistical model developed for FCR in earlier work (McClure et al. 2020), we were unable to quantify model selection uncertainty by using different models in our forecasting workflow (e.g., process-based ebullition models; Peltola et al. 2018), which is an important step for future work. Testing different CH4 ebullition models at the weekly time scale in FCR and varying temporal and spatial scales may further improve our understanding of this biogeochemical process.

Additional improvements to our forecasting system include the addition of automated sensors and other statistical methods into the forecasting workflow. Recent technological improvements in automated ebullition sensors hold great potential for advancing the future of CH4 ebullition forecasting at daily or subdaily scales (e.g., Varadharajan et al. 2010, Delwiche et al. 2015, Maher at al. 2019). Similarly, alternate data assimilation methods (e.g., ensemble Kalman filters (e.g., Page et al. 2018, Thomas et al. 2020) or particle filters (Rowe et al. 2016)), could be useful for future applications of this work that uses more complex models.

## 4.1 Scaling CH4 ebullition forecasting to other sites

Although we designed near-term, iterative CH4 ebullition forecasts specifically for FCR at weekly forecast horizons, our forecasting iterative workflow could be adapted for predicting CH4 ebullition in other freshwater ecosystems. Here, we provide a list of minimum requirements to begin forecasting CH4 ebullition rates from other lakes and reservoirs:

**1) A CH4 ebullition field monitoring program.** The foundation of near-term, iterative forecasting is a coupled model-data feedback loop (Dietze 2017*a*). Thus, developing a field monitoring program that routinely samples ebullition from the waterbody of interest is necessary to forecast ebullition. Initial observations are needed to develop and train a forecast model, as we did in this study using observations collected in summer 2017 by McClure et al. (2020) to develop a model that was trained for summer 2019 forecasts using 27 May to 24 June 2019 data. Additionally, data from a monitoring program can help prioritize CH4 ebullition spatial hotspots for modeling and forecasting, as we did in this study by focusing on the upstream transect site that emitted most of the reservoir-wide CH4 ebullition in 2017. Finally, it is ideal if the monitoring data collection can be collected *within* the maximum time horizon of the forecast driver data, to enable forecast evaluation. In our case, the NOAA GEFS maximum horizon is 16 days, so we had a 16-day window from when the forecast was generated to collect new observations to evaluate forecast skill.

**2) A model to generate future predictions of CH4 ebullition that can be updated as observations become available.** Models are the second essential component of the model-data feedback loop in an iterative forecast workflow. While we used a relatively simple statistical model for this study (McClure et al. 2020), any model type can be applied as long as its drivers can also be forecasted, and the model can iteratively update as new observations become available. Multiple model types can be used: e.g., time series models, process-based models, and even neural networks (Abbasi et al. 2020) can all be used to develop CH4 ebullition forecasts.

**3) Forecasted driver data.** Depending on model structure, real-time ebullition forecasts may need forecasted driver data or an autoregressive term to predict CH4 ebullition at the next time step. This, indeed, may be a hurdle to begin forecasting CH4 ebullition. In our case study, we had access to FLARE’s existing infrastructure, which generated daily forecasts of water temperature that were scaled to become driver data for our model (Eqn. 2). Although FLARE can be deployed to other lakes and reservoirs (Thomas et al. 2020), other waterbodies may not have the cyberinfrastructure or temperature sensors needed to run that forecasting system. However, in lieu of having a forecasting system like FLARE, weather forecasts by themselves can be integrated into a CH4 ebullition forecast workflow if the ebullition prediction model is driven by meteorological variables like barometric pressure or shortwave radiation (Tokida et al. 2007, Wik et al. 2014, Peltola et al. 2018). For example, Tokida et al. (2007) showed how falling atmospheric pressure triggered substantial increases in CH4 ebullition from natural wetlands and Wik et al. (2014) associated seasonal CH4 ebullitive fluxes from thermokarst lakes with shortwave radiation. Thus, a forecast workflow for CH4 ebullition is possible, even if the site has no pre-existing forecasting infrastructure.

## 5 Conclusions

We were able generate successful forecasts of CH4 ebullition rates, a highly variable biogeochemical process in freshwater ecosystems, using using forecasted SWI temperature data and manually collected CH4 ebullition rates (Figure 3a). Sequential data assimilation in each forecast cycle substantially improved model performance and the forecasts performed better than forecasts without data assimilation and a persistence null model (Table 1, Figure 3 and 4). Finally, data assimilation decreased total forecast variance and uncertainty partitioning showed how the drivers of total forecast uncertainty changed during the forecast period, which improved and changed the parameter estimates through time (Figures 4 – 6). 2 SENTENCES ON WHAT YOU LEARNED The success of our case study suggests that iterative ecological forecasting with data assimilation may be a valuable approach for better quantifying biogeochemical processes that are difficult to monitor with automated sensors simultaneously while providing insight to their predictability.

## 6 Acknowledgements

*Data and Code Availability*: All data used in this manuscript will be available in the Environmental Data Initiative repository. The code, directions, and data to run the CH4 ebullition rate forecasts are available on GitHub (McClure et al. 2020): <https://github.com/ryanmclake/CH4cast>.

*Author Contributions:* RPM and CCC developed the original research idea.RPM developed the forecasting system framework with substantial input on structure and code development from RQT. WMW and MEL contributed to code development, field data collection, and model testing. RPM wrote the manuscript with CCC and RQT. All authors provided feedback and approved the final version. The authors declare no conflicts of interest.

*Funding and Support*: This work was supported by the U.S. National Science Foundation (CNS-1737424, DEB-1753639, DBI-1933016, DEB-1926050); the Virginia Tech Global Change Center; and the Fralin Life Sciences Institute. We thank the Smart Reservoir and FCR Carbon teams for their helpful feedback during this project and the Western Virginia Water Authority for their long-term support and access to the reservoir. We thank Ashley Mickens for her critical help with data collection and GC-FID analyses. Additionally, we thank Bethany Bookout, Alexandria Hounshell, Abby Lewis, Heather Wander, Rachael Corrigan, James Maze, Dexter Howard, Jacob Wynne, and Nick Hammond for their help with field data collection and Vahid Daneshmand and Renato Figueiredo for their development and maintenance of the FLARE forecast system. We also thank Paul Hanson, Erin Hotchkiss, Madeline Schreiber, and GLEON colleagues for their feedback on manuscript development. Finally, as always, we thank Bobbie Niederlehner for helping with the GC-FID and laboratory analysis methods development.

# References

Aman, A. A. and Bman, B. B.: The test article, J. Sci. Res., 12, 135–147, doi:10.1234/56789, 2015.

Aman, A. A., Cman, C., and Bman, B. B.: More test articles, J. Adv. Res., 35, 13–28, doi:10.2345/67890, 2014.

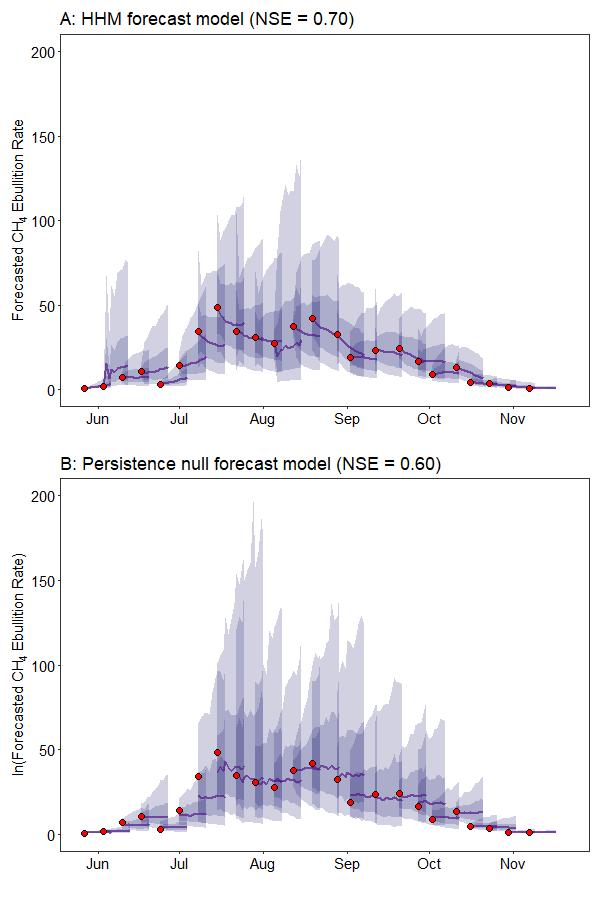


Figure 3: The logo of Copernicus Publications.