**Chapter 1: Near-term iterative hindcasts at multiple time scales improves understanding of phytoplankton dynamics**

**NOTE CHANGE ALL ALGAE TO PHYTOS**

Introduction

Ecosystem services provided by lentic freshwater systems are crucial to society, and yet lakes and reservoirs are increasingly threatened by eutrophication. On a global level, lakes and reservoirs are experiencing changes to many water quality metrics, including increases in surface water temperature (O’Reilly et al., 2015), changes to mixing regimes (CITATION), and as well as shifts in trophic status (CITATION) (others?). One increasingly common form of water quality degradation in lakes and reservoirs is eutrophication, or excessive nutrient enrichment (Smith 2003). Eutrophication often results in increases in large phytoplankton blooms, as well as increases in the presence of cyanobacteria, a type phytoplankton which are known to produce toxins harmful to human health (CITATION). Phytoplankton blooms are a significant burden to the treatment of drinking water, by increasing the need for physical, as well as chemical treatment of water before it can be delivered to citizens, and is estimated to cost $2 billion annually in the U.S. alone (Dodds et al., 2009).

Phytoplankton dynamics are governed by highly variable processes operating on different time scales in response to changing environmental conditions (CITATIONS). “Baseline” or median phytoplankton concentrations may exhibit small fluctuations on a weekly to monthly basis in response to seasonal drivers (e.g., water temperature, light availability), but quickly exhibit exponential growth on the daily scale in response to increased nutrient availability and create blooms, or large aggregations of biomass. While baseline phytoplankton concentrations are critical to daily operations of drinking water supplies during a majority of the year, rare, ephemeral phytoplankton blooms may cause disproportionate disruptions to drinking water management relative to their frequency. Given the relative importance of phytoplankton bloom events, it is crucial to drinking water management to be able to anticipate both daily fluctuations as well as large blooms.

While increasing trends in eutrophication are occurring globally, we are also seeing much more variability in phytoplankton on a daily scale (Paerl and Paul 2012, Carey et al. 2012, Ho et al., 2019). Given the global trend of water quality degradation and increase in daily variability is outside the range of historical conditions experienced by most water bodies, it is becoming increasingly difficult to manage drinking water resources in the face of uncertainty.

One tool with novel potential to improve manager’s ability to anticipate increases in phytoplankton biomass is through the use of ecological forecasts. An ecological forecast provides an estimate into the future of one or more ecological properties with quantified uncertainty around the estimate. While the field of ecological forecast is currently developing, a common step in developing operational forecasts is to produce hindcasts, or forecasts which are produced for a time period that has already occurred, but which use forecasted driver data, effectively simulating a ‘future’ state from the perspective of the model. Developing hindcasts can help in the iterative process of improving forecast models before investing in fully operational forecasting systems. Forecasting in other fields has been instrumental to society through the development of successful forecasts in meteorology, economics, agricultural science, and epidemiology (citations from prospectus). Forecasting of ecological variables is a rapidly developing field (Dietze, 2017) which has shown promise in many sub-disciplines of ecology to improve management of freshwater ecosystems (Thomas et al. 2019, others from EF class biblio).

While forecasts of phytoplankton dynamics have been increasing in recent years, they are highly variable in the time horizon being forecasted. Many are focused on sub-weekly to daily time scales, while some forecast up to two weeks in the future. Because phytoplankton are highly dynamic and respond to changes in drivers on multiple time scales (REFS), it is unclear as to what timestep is most effective at forecasting phytoplankton. Further, data availability is not consistent between daily and weekly time scales in many lakes and reservoirs. Many long-term monitoring programs of lakes and reservoirs typically take place on a weekly scale, meaning data for developing weekly forecasts may be readily available in a large number of waterbodies. Because of the design of weekly monitoring program, there are typically an abundance of potential model covariates which are collected on the same time scale as the response variable, phytoplankton. In contrast, a recent increase in the deployment of high-frequency sensors which collect data on the sub-daily scale has allowed for the development of datasets which are amenable to producing daily models of phytoplankton. However, it remains to be seen which time scale is most effective for forecasting phytoplankton within a given waterbody, and under what conditions.

In addition to the prediction of phytoplankton concentration, an important component of a forecast is the uncertainty around that prediction. Uncertainty is crucial to the application of forecast products in order to properly inform management decisions, and also as a way to iteratively improve forecast skill by addressing large sources of uncertainty (see Mary’s refs on this from the IW manuscript). Sources of uncertainty are far-reaching but can be reasonably assumed to be caused by several sources. These include but are not limited to process, parameter, initial condition, driver, and observational uncertainty (see Box 1 for definitions). Despite the need for uncertainty estimates, however, they are often unconsidered or at least unreported in current ecological forecasts.

To assess our ability to forecast phytoplankton dynamics with fully specified uncertainty, we developed near-term, iterative hindcasts of phytoplankton biomass (DEFINITION) using autoregressive (AR) linear models operating at both weekly and daily time steps. We aimed to answer two main questions:

1. How does a daily vs. weekly forecast recreate observed, near-term chlorophyll-a (a proxy for phytoplankton) over a one-year time period, during both bloom and non-bloom conditions?

2. How do the major contributions of uncertainty over time and forecast horizon vary between the daily and weekly forecasts?

|  |  |  |
| --- | --- | --- |
| Box 1. Definitions of uncertainty sources | | |
| Uncertainty type | Definition | Example |
| Process | error due to a mismatch between the structure of the model being used to forecast and the process being modelled |  |
| Parameter | uncertainty in the actual values of model parameters |  |
| Initial condition | Uncertainty in the actual initial conditions needed to start a forecast run |  |
| Driver | Uncertainty in the actual value of forecasted model co-variates |  |
| Observational | Uncertainty in the actual observed value of driver or response variables |  |

**This paragraph should dive into HABs ecology and modeling, making the case for both daily and weekly monitoring needs as a result of multiple timescales of phytoplankton growth responses**

* + - * Kara et al 2012: short-term phyto dynamics difficult to reproduce, Rigossi 2013

Methods

*Study site*

Falling Creek Reservoir (FCR) is a small (0.119 km2), eutrophic drinking water supply reservoir located in Vinton, VA. It is owned and operated by the Western Virginia Water Authority and has been monitored weekly in the summer months by our research team since 2013. FCR is a monomictic reservoir with a history of phytoplankton blooms (Gerling et al. 2014, 2016), and one major upstream tributary that has been monitored with a weir since 2013 (). LOOK UP SOME FCR PAPERS FOR OTHER THINGS TO MENTION AND WHAT TO CITE

**Figure 1. Map of study site**

* + Overview of weekly vs daily forecast development

*Historical Weekly Dataset and Model Development*

The weekly training dataset covered four years (2013-2016) and was developed from weekly measurements of chlorophyll-a, meteorological variables, as well as discharge at the major inflow to the reservoir. These were all chosen as covariates because they are either available as forecasts from National Oceanic and Atmospheric Administration (NOAA) or able to be forecasted using simple linear regression models (e.g., discharge). Chlorophyll-a was estimated by taking weekly profiles using a CTD (SeaBird SERIAL NUMBER). We measured discharge using a pressure transducer at a weir installed at the major inflow to FCR. Flow was measured every 15 minutes, and discharge to the reservoir was calculated, as described in Gerling et al. (2014). A meteorological station measuring X, Y, Z collected data every ten minutes using THESE INSTRUMENTS (serial numbers) from YYYY-present. Any weeks were data was missing (n=XX) were linearly interpolated. Variables that did not follow a normal distribution were transformed to meet the assumptions of a linear model. An autoregressive lag of chlorophyll-a was chosen at one timestep (i.e. one week) and was determined using the package ‘asta’ in R (some citation).

The weekly training dataset was limited to May to October, as this is when sampling consistently occurred on a weekly basis during 2013-2016. Using the training dataset described above (number of datapoints = XX), we fit multiple linear regression models using the function ‘dredge’ in the package ‘MuMIn’, and selected the best and most parsimonious model using AICc. Our weekly autoregressive model is as follows

Chl-a(t) = β1 + β2Chl-a(t-1) + β3SW mean(t) + β4mean flow(t) + Ɛ **(Eq. 1)**

Where the response, Chl-a(t), is the chlorophyll-a concentration at the forecasted timestep, t. Chl-a(t-1) is the autoregressive term or chlorophyll-a at the previous timestep, 1 week earlier. SW mean(t) is the mean shortwave on the forecasted timestep. Mean flow(t) is the mean discharge on the forecasted timestep. β1 , β2, β3, and β4 are parameters and Ɛ is an error term.

**Table 1. Summary of historical and sensor datasets, including data measured and length of dataset**

*Daily Dataset and Model development*

Beginning in August, 2018, an EXO sonde (serial numbers) was installed at the deepest site at FCR, measuring chlorophyll-a fluorescence at 15-minute intervals. Using the same potential covariates as listed above for the weekly model (i.e., meteorological and discharge variables), and following the same model selection protocol, we developed all possible combinations of autoregressive linear models. Because the weekly model was limited to XX number of datapoints in its training dataset, we trained our daily model on the same number of datapoints, which results in a training dataset from August 15, 2018 to December 15, 2018. Our best selected daily model is as follows

Chl-a(t) = β1 + β2Chl-a(t-1) + β3RelHum mean(t) + Ɛ  **(Eq. 2)**

Where the response, Chl-a(t), is the chlorophyll-a concentration at the forecasted timestep, t. Chl-a(t-1) is the autoregressive term or chlorophyll-a at the previous timestep, ‘n’ days earlier. RelHum mean(t) is the mean relative humidity on the forecasted timestep. β1 , β2, and β3 are parameters and Ɛ is an error term.

Forecasting framework (FLARE)

Using the linear autoregressive model described above (Eq. 1), the model was adapted to produce forecasts with uncertainty using the Forecasting Lake and Reservoir Ecosystems (FLARE) forecasting framework (Thomas et al. 2020). Using FLARE, real-time sensor data is automatically uploaded to a data repository (GitHub, link?). Sensor data includes a suite of physical, chemical, and biological data, including chlorophyll-a fluorescence measured using an EXO sonde (serial numbers and company), meteorological variables (At FCR, weather data are collected on the minute resolution from a meteorological station (with sensors measuring air temperature, wind speed, relative humidity, shortwave and longwave radiation, and precipitation; see Carey et al. 2019x) and discharge at the major inflow to the reservoir.

Something about the Bayesian framework

Number of ensembles, maybe include this in a separate section about uncertainty analysis set up

**Figure 2. FLARE Workflow**

* + Weekly forecasts
    - Exo data converted into CTD units and square root transformed inside model framework
    - Using the weekly model developed above, weekly forecasts, providing 1-week ahead and 2-week ahead forecasts, were produced every day from January 01, 2019 to December 22, 2019. The time period from August 15, 2018 to December 15, 2018 (number of datapoints = K) when the driver data was available on a daily basis was used as a spin-up period. While in forecasting mode, new driver data was assimilated weekly on Monday (in order to follow the weekly timestep), and the model was re-fit at each time step under a Bayesian framework to allow the parameter values to evolve over time. All ensemble members (n=420) and parameter values were saved for archiving at each timestep.
  + Daily forecasts
    - Using the daily model developed above, daily forecasts, from 1- to 16-days ahead, were produced from January 01, 2019 to December 22, 2019. To allow a spinup period similar to the weekly model, forecasts from December 15, 2018 to January 01, 2019 were used as a spinup period and not included in final forecast analyses. While in forecasting model, new driver data was assimilated daily when available, and the model was re-fit at each time step under a Bayesian framework to allow the parameter values and model fit to evolve over time. All ensemble members (n=420) and parameter values were saved for archiving at each timestep.
  + Forecast assessment
    - Assessing skill under different conditions
      * Because phytoplankton exhibit different growth responses to ecological conditions at different times, we analyzed our forecasts under three time periods: 1) the entire year period where forecasts were produced, 2) under nonbloom conditions, defined as X days before and after the peak observed chlorophyll-a value, and 3) under bloom conditions
    - Null-persistence model
      * We developed a null-persistence model in order to test the robustness of our forecasts. Our null model assumes that the chlorophyll-a concentration at the next timestep will be unchanged from the current timestep, with process error from the Bayesian model added. We calculated an ensemble of null models in order to compare to our ensemble forecasts by sampling 420 times (the number of ensembles) from the distribution of the process error term from the Bayesian model output at each timestep and adding this value to the observed chlorophyll-a concentration (e.g., for the 8-day forecast, the observed chlorophyll-a concentration from 8 days prior is the null model, plus the process error from the model fit, sampled for each model ensemble).
    - Forecasts and null models were assessed using RMSE, R2 (?), and CRPS.
  + Uncertainty analysis
    - We partitioned five types of uncertainty for our weekly model and four for the daily model which contribute to the variance in our forecasts
      * Process
      * Initial conditions
      * Discharge driver (weekly model only)
      * Meteorological driver
      * Parameter
    - Relative contributions of uncertainty were quantified by isolating each type of uncertainty and allowing that uncertainty vary while holding all others constant

Results

* + Do you need to have a description first of what chla did in general (patterns over time) before jumping into its forecasts? Or would that be combined with the forecast period section because the observed won’t have a separate figure- it’ll be combined with the forecasts? I think setting up what the median concentrations were, the bloom description etc would be helpful; I’m just not sure which section would be best
  + Forecasts over a one-year period
    - Both daily and weekly forecasts over one year generally capture observed chlorophyll-a dynamics (Fig. 3).
    - However, the daily forecasts were not better than a null persistence model at any timestep (Fig 4a).
    - For both the daily and weekly model, forecasts do slightly worse than the null model at 7 days (i.e., one week) ahead, but at two weeks ahead, the weekly forecast is significantly better than the null model.
      * **Figure 3. Muti-panel figure with daily and weekly models at various timesteps and observed chl** 
        1. Day 1 forecast, daily model
        2. Day 7 forecast, daily model
        3. Day 7 (i.e. week 1) forecast, weekly model
        4. Day 14 forecast, daily model?
        5. Day 14 forecast, weekly model?
  + Forecasts during nonbloom conditions
    - Under nonbloom conditions, the daily forecast is slightly better than the null until 9 days into the future (Fig 4b). However, the weekly forecast is better than the null at both 1-week and 2-week forecasts, with the forecast being much better than the null at 2-weeks ahead (Fig 4b). Interestingly, the 14-day ahead forecast of the daily model is substantially worse than the null (Fig 4b).
  + Forecasts during bloom conditions—TBD
    - **Figure 4. RMSE of forecasts at all timesteps**
    1. **RMSE over forecast horizon for daily and weekly models over whole year**
    2. **RMSE for daily and weekly models over nonbloom period only**
    3. **RMSE for daily and weekly models over bloom period only**
  + Uncertainty analysis
    - Process error dominant source when averaged over the entire time period
    - Uncertainty sources not constant over time series
    - In weekly model, during bloom, parameter uncertainty increased dramatically, indicating that parameters were not properly fitted to capture bloom dynamics
    - **Figure 5. Relative proportion of uncertainty over entire time series**
      1. **Weekly model**
      2. **Daily model**
* Discussion
  + We successfully developed a near-term, iterative forecast of chl-a that produced both daily and weekly forecasts up to 16 days ahead.
  + While daily forecasts were not an improvement over a null model over the entire time period, we did find that under nonbloom conditions the daily forecasts did slightly better than the null persistence model. Additionally, the weekly forecasts did substantially better than the null at both 1 and 2-week time horizons under nonbloom conditions. This is not surprising given than the model was calibrated to model weekly patterns which are clearly not well captured by the null model.
  + Interestingly, contrary to commonly upheld forecasting expectations/paradigms?? Probably not theory yet! (Dietze 2017) we found with our weekly model that the forecastability did not decrease with forecast horizon. This pattern holds true especially under nonbloom conditions, but is also seen over the entire year period.
    - Can you interpret this a bit more? What does this mean for forecasting phytos in general?
  + I think before launching into the uncertainty, there needs to be more discussion here about bloom forecasting vs chla forecasting, how the drivers of the daily model vs weekly model varied, and how this shows that phytoplankton respond to different drivers on different time scales, then transition to uncertainty…
  + Our dominant source of uncertainty averaged over time was process error, indicating that our model was missing key processes that control dynamics in phytoplankton abundance. This is not surprising given that due to the data latency constraints of a forecasting system, we were limited to driver variables which were also forecastable (i.e., nutrients or predator abundance are not reasonable covariates in a forecasting framework)
  + Relative contributions of uncertainty were not static throughout the year.
    - Driver uncertainty, particularly discharge driver uncertainty, increased in relative importance at various time throughout the year, especially during periods of high precipitation (e.g., fall hurricane season and spring rainfall).
      * Must have confidence in forecasts of covariates
    - Further, parameter uncertainty increased dramatically during the bloom in weekly forecasts.
    - Can you do a comparison to other forecasting analyses that partitioned uncertainty? I think that most find driver data & process error dominate, so the parameter uncertainty finding here is SUPER cool!!
  + Model I missed large bloom events in 2019
    - Do some analysis similar to the % likelihood of turnover to analyze how far ahead the model needed to recognize the bloom (because it eventually recognizes high concentrations or something about how long it took to get back into the CI of the forecast
    - Not uncharacteristic of AR models
  + Copper sulfate events
    - The model was unable to immediately anticipate this event, but does readjust after a short period of time.
  + Discussion of utility of developing other types of models
    - Other empirical models (GAM, ANN, process-based, etc.)
    - Modeling averaging/ensembles
    - Would process-based approaches do better at predicting blooms?
  + Applications/uses of HABs forecasts
    - What is a useful forecast for managers?
      * What time step
        + Why are weekly chl forecasts still useful?
      * What level of accuracy (# of ug/L?) is needed for management?
      * What level of uncertainty
      * How to deliver information to stakeholder
        + Uncertainty visualization
  + Scaling AR forecasts to other water bodies
    - GLEON
    - Using simplistic AR models helps move us toward implementation of forecasts in a diverse set of lakes and reservoirs

Figure and table list

Start to think about an authorship contribution statement? We can talk through this too