**Chapter 1: Successful production of near-term, iterative forecasts of phytoplankton blooms using autoregressive linear models**

* Intro
  + to improve management of freshwater ecosystems
  + Ecosystem services provided by lentic freshwater systems are crucial to society, and yet lakes and reservoirs are increasingly threatened by eutrophication.
    - Services include cultural and aesthetic value, food production, drinking water, etc.
    - as impacts of human society on the environment continue to increase globally.
    - Broad examples of severe water quality degradation in drinking water sources --
      * Focus in on HABs (stay at HABs level when discussing water quality through the rest of the intro)
    - As this degradation increases, it has become more important than ever to not only respond to eutrophication events but to be able to anticipate them before they occur.
  + HABS are ephemeral, short-term events that have large impacts, making them critical to anticipate.
    - Maybe a paragraph about HABs ecology and modeling
    - Kara et al 2012: short-term phyto dynamics difficult to reproduce
  + However, because ecological forecasting, and especially forecasting of HABs, is a developing field, there is not yet a consensus as to the best approach for predicting future water quality. In particular, it is unclear how well data-driven empirical models are able to predict ecological dynamics. In particular, it is unclear if data-driven empirical models or numerical simulation process-based models yield more accurate forecasts (Dietze, 2017). Based on a high-level literature review, we found that empirical modeling to produce ecological forecasts was chosen in a majority of studies (65%, n=17, Table 1).
  + Empirical models run the gamut from simplistic generalized linear models (GLM) to more complex machine-learning techniques such artificial neural networks (ANN). Use this paragraph to justify using linear models (simplistic approach may be more applicable across a broad range of drinking water systems and accessible to managers varying levels of modeling expertise)
  + Uncertainties in forecasts – set up definitions here of different kinds of uncertainty in forecasts
    - Crucial to the application of forecast products in order to properly inform management decisions
    - Often unconsidered or at least unreported 🡪 for managers, uncertainty, threshold of risk, check out decision support from dietze again
  + To further assess the success of empirical models in forecasting HABs, we developed near-term (2 week), iterative forecasts using several types of autoregressive (AR) linear models. We aimed to answer three main questions:
    - 1. How well can we forecast near-term (two-week horizon) chlorophyll-a (a proxy for HABs) over a one-year time period?
    - 2. How does data assimilation outside of the training period improve forecast accuracy?
    - 3. What are the major contributions of uncertainty in our forecasts?
    - 4. How does the relative importance of sources of uncertainty change over time?
* Methods
  + Study site
    - Figure 1. Map of study site
  + Historical and sensor dataset
    - Table 2. Summary of historical and sensor datasets
  + Model development
    - Model driver selection
      * Selected 3 models within 2 AICc units
        + Model I: Chla(t) = chla(t-1) + SW mean(t) + mean flow(t)
        + Model II: Chla(t) = chla(t-1) + SW mean(t) + mean flow(t) + air temp mean(t)
        + Model III: Chla(t) = chla(t-1) + SW mean(t) + mean flow(t) + wind speed mean(t)
    - Model training over four years, 2013-2016
  + Forecasting framework (FLARE)
    - Figure 2. FLARE Workflow
    - Data assimilation (not currently doing this but am thinking of ways to incorporate incoming CTD or EXO data—may be a good conversation with RQT—could be a whole separate analysis that points to the value of automated data assimilation??)
  + Forecast assessment
    - Null-persistence model
    - Assessment metrics
  + Uncertainty analysis
    - Our model partitions five types of uncertainty which contribute to the variance in our forecasts
      * Process
      * Initial conditions
      * Discharge driver
      * Meteorological driver
      * Parameter
    - Relative contributions of uncertainty were quantified by isolating each type of uncertainty and holding that uncertainty constant while allowing all others to vary.
* Results
  + Model training
    - All three models successfully recreated observed dynamics over the historical training period of four years (R2 = 0.4, RMSE = 1.7 ug/L, Figure 3)
    - Model I (SW mean + mean flow)
      * R2 = 0.443
      * RMSE = 1.71 ug/L
    - Model II (SW mean + mean flow + air temp mean)
      * R2 = 0.449
      * RMSE = 1.71 ug/L
    - Model III (SW mean + mean flow + wind speed mean)
      * R2 = 0.446
      * RMSE = 1.70 ug/L
  + One year of forecasts, 15-Aug-2018 to 15-Aug-2019
    - Table of assessment metrics
    - Figure of time series including 95% CI and obs chl
  + Uncertainty analysis
    - Process error dominant source when averaged over the entire time period (Figure )
    - What conditions lead to relative contribution of uncertainties changing over time (Figure)
* Discussion
  + We successfully developed a near-term, iterative forecast of chl-a that produced forecasts within ~ 2 ug/L over a one-year time period. But unable to capture blooms
  + Our dominant source of uncertainty was process error, indicating that our model was missing key processes that control dynamics in phytoplankton abundance.
  + 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).
  + Considering driver variables to include in forecasts
    - Must have confidence in forecastability of driver covariates
  + Model I missed large bloom events in 2019
    - Define bloom, check ariana’s latest paper
      * When did forecast successful predict a bloom to occur, did forecast predict a bloom even if it misses the magnitude
      * Use threshold to calculate hits and misses
    - 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
      * Analysis of high-resolution EXO data to see if a signal of increased variability if noticeable before the bloom develops?
    - Copper sulfate events
      * Need to incorporate human behavior into forecasts
    - Not uncharacteristic of AR models
  + Discussion of utility of developing other types of models
    - Other empirical models (GAM, ANN, process-based, etc.)
    - Modeling averaging/ensembles
  + Applications/uses of HABs forecasts
    - What is a useful forecast?
      * What time step
        + Why are weekly chl forecasts still useful?
      * What level of accuracy (# of ug/L?)
      * What level of uncertainty
      * How to deliver information to stakeholder
  + 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