Whitney Woelmer, Master’s Prospectus Outline

Overall Introduction

1. Importance of freshwater ecosystems and the many services they provide (setting up why their water quality is so important, as well as what services are the highest priority- e.g., clean water for drinking, etc.)
2. Threats to freshwater systems in a changing world
   1. Changing climate
   2. Watershed effects
      1. Land use shifts
      2. Increasing human population
3. Understanding systems and anticipating their response is more important than ever
   1. to allow society to coexist with changing ecosystems
   2. for managers to preemptively manage/anticipate impending poor water quality events
   3. Lake Erie 2014 example, major economic loss

**Proposed Research**

*Chapter 1: Developing near-term forecasts of phytoplankton in drinking water reservoirs***Introduction**

* Forecasting is more important than ever now.
  + Why?
    - Changing world = changing ecosystems
    - Data availability (move towards publicly available data)
    - High-frequency sensor development = data deluge
  + What is forecastable
    - “Forecastable ecosystem attributes are ones for which uncertainty can be reduced to the point where a forecast reports a useful amount of information.” Clark et al 2001
      * Too much uncertainty means not easily forecastable for an applied use
      * “strong nonlinearity and stochasticity” clark
      * A few examples of things that have been or currently are being forecasted
  + Ecological forecasting currently spans numerous disciplines, with a variety of purposes. Summary of mini-lit review here, including topics and uses
    - Applied uses
      * For direct use in the management of a system
      * Ex:
      * Give summary from lit review for the % of studies doing this
    - Basic
      * For understanding how a system is/will respond to changing pressures
      * Ex:
      * Give summary from lit review for the % of studies doing this
    - Within freshwater ecosystems, we find many examples of applied forecasts (EXAMPLES HERE)
  + **Forecasting approaches**: Because the field of ecological forecasting is relatively new, there is not yet a consensus as to the best approach for predicting water quality.
    - **Empirical methods** are used in a majority (X%, n= Z) of current forecasting studies selected. Empirical approaches are popular because of their data-driven character; they are inherently suited for a particular system because they are based on past trends within that system. Input data for empirical models are commonly available through routine monitoring of a system, and these models do not require specific input data, but can be developed using a variety of driver variables. Additionally, empirical models are relatively easy and quick to develop and put to use. However, because empirical models are built on the historic conditions of a system, if future conditions are outside the realm of past conditions, models might no longer be able to capture the mechanism responsible for changes.
    - Process-based
      * Pros
        + Useful tools for simulating changes within a system

Can be used to explain underlying mechanisms

* + - * Cons
        + Require a multitude of input driver data
        + Many parameters
        + Require a lot of time and expertise to calibrate

Expertise on the model, as well as the system

* + - Further reading: Rastetter 2017
* My first chapter will focus on addressing the knowledge gap in ecological forecasting regarding how best to forecast water quality. I will produce hindcasts and near-term iterative forecasts of phytoplankton in a drinking water reservoir using both an empirical and a process-based approach. Model performance will be assessed by a suite of metrics addressing both the ability of the model to capture overall dynamics, as well as extreme events. My work will specifically address the following questions:
  + Question 1: How well can an empirical model and a process-based model hindcast observed chlorophyll dynamics over a 1-year period outside of the training period?
  + Question 2: How well (measured through quantified uncertainties/a probabilistic forecast) can an empirical model and a process-based model forecast near-term chlorophyll-a dynamics over a 16-day period?
  + Question 3: What information does an ensemble model approach provide for chlorophyll forecasting that cannot be explained from a single-model approach? near-term chlorophyll-a?
* This chapter 1 will thus**: 1) inform managers and decision-makers about which variables are most important for routine monitoring and management of water quality in the face of land use and climate change, 2) provide essential information to the freshwater research community regarding modeling approaches to forecasting water quality, and 3) expand the scientific field of ecological forecasting that will guide forecasting applications across many different ecosystems.**

Lakes and reservoirs provide a suite of critical ecosystem services, including recreation, food production, and drinking water1. Despite the importance of these services, fresh waters are increasingly threatened by rapidly changing land use and climate2, two factors that have led to the contamination of drinking water for millions globally by severe harmful algal blooms (HABs). Consequently, there is a pressing need to not only understand the current state of our freshwater ecosystems, but to predict how they will respond tomorrow, next week, and next year. Therefore, the ability to forecast the future state of our drinking water sources is of utmost importance to society and freshwater ecology as a discipline.

Because the field of ecological forecasting is relatively new, there is not yet a consensus as to the best approach for predicting water quality3. Ecologists commonly use time series statistical models to predict future water quality based on past and current conditions (e.g., today’s phosphorus load, water temperature)4, while another approach uses process-based models (built from coupled differential equations) such as the General Lake Model (GLM), a one-dimensional hydrodynamic model, to simulate different metrics of water quality5. A simplified time series model enables application in many waterbodies without the intensive monitoring technology or the extensive set of parameters required by GLM, yet GLM may give a more informed forecast with lower uncertainty, allowing managers to make decisions with more confidence. However, it remains unknown as to which of these approaches is most effective for forecasting water quality at both the single and multiple waterbody scale. Further, by developing multiple models to inform forecasts of phytoplankton, we can draw from an ensemble of possible outcomes to produce a suite of forecasts with quantified uncertainty.

**Methods**

* **Study Site**
  + Falling Creek Reservoir (FCR) is a small (~12 ha), shallow (maximum depth < 9.3m) dimictic drinking water reservoir located in southwestern Virginia. FCR is owned and operated by the Western Virginia Water Authority (WVWA). The watershed of FCR is almost entirely forested, although the reservoir continues to exhibit incidences of poor water quality as a result of historical eutrophication of the reservoir (Gerling et al 2016). The major water source to FCR comes from a single tributary which flows from Beaverdam Reservoir (BVR; Figure 1).
* **Historical and sensor dataset**
  + An extensive, routine monitoring dataset has been cultivated in FCR since 2013 in collaboration with the WVWA and Virginia Tech University. This dataset includes meteorological, physical, chemical, and biological data collected both at the deep hole of the reservoir and at the major inflow to FCR. The inflow dataset also includes discharge to the reservoir measured every 15 minutes at a weir installed at the stream site. More recently, as part of the Smart and Connected Communities (SCC) project, FCR has been outfitted with numerous high-frequency sensors to capture real-time changes in water quality. These data are streamed wirelessly to an online server and appended to the ongoing datafile. Sensor data includes meteorological, physical, and biological data.
* **Model Development (Question 1)**
  + We will develop both an empirical and a process-based model to forecast chlorophyll-a in the surface water (1.0 m) at FCR during the summer stratified period (May-October). The training period for ARIMA will be 2013-2016 and for GLM will be 2013-2017. The training period is shortened for ARIMA due to a lack of input data for the model during 2017. Both models will be trained using data from 2018.
  + Empirical: ARIMA
    - An autoregressive integrated moving average (ARIMA) generalized linear model was developed for our empirical approach to forecasting chlorophyll-a in FCR. The appropriate time step to be included as the autoregressive term will be determined by selecting the timestep with the highest Pearson’s r correlation coefficient. From a pool of 53 meteorological, physical, chemical, and biological driver variables, we will select only driver variables which are predictable in nature (e.g., meteorological and inflow variables). From within the pool of predictable driver variables, we will further select only variables which are not correlated with each other through the use of a Pearson’s correlation analysis. Using these variables, we will develop all possible ARIMA model iterations, and the best model will be selected by AICc (Akaike’s Information Criterion) and parsimony.
  + Process-based: GLM
    - We used the General Lake Model, Aquatic Ecodynamics (GLM-AED) as our process-based model. GLM-AED is a one-dimensional hydrodynamic model, etc.
* **Forecasting Framework: FLARE (Question 2)**
  + Both models will be integrated into an existing forecasting framework, Forecasting Lake and Reservoir Ecosystems (FLARE), to produce iterative near-term 16-day forecasts of chlorophyll-a.
    - Use near-term NOAA forecasts + sensor-cloud networks to pull current data and forecasts of driver variables to go into ARIMA and GLM models
  + For both models, I will sample from a distribution of parameters using Bayesian techniques to generate uncertainties for each forecast
  + These forecasts will be direct deliverables to WVWA
* **Integrated Model Averaging (Question 3)** 
  + Weighted model averaging as a way to incorporate strengths of multiple models
    - ‘‘Observation and theory get on best when they are mixed together, both helping one another in the pursuit of truth’’ (Eddington 1935).
* If model strengths between ARIMA and GLM are distinct, a weighted model average may be developed to capture the relative strengths of each model and cover as much of the variability in chlorophyll-a as possible
* TEXT TO BE INCORPORATED INTO THIS SECTION: Once both models are calibrated, I will assess how effective each model is at predicting near-term future states of water quality. For GLM, I will use 2-week weather predictions to run the calibrated model to predict chlorophyll-a. For the time series model, if the model indicates, for example, that the previous month’s P load is the most important driver of current chlorophyll-a, I will feed the model with current P loads to predict future chlorophyll-a. For both models, I will sample from a distribution of parameters using Bayesian techniques to generate uncertainties for each forecast. The forecast model outputs will then be compared to the observed high-frequency chlorophyll-a sensor data to determine which approach is most robust under different conditions. I predict that output from the time series model will have a larger uncertainty than the GLM model, given that GLM is a process-based model driven by coupled equations, rather than statistical relationships alone.
* Model assessment
  + **Model comparisons will be made using R2 and root mean squared error (RMSE) to assess model fit of chlorophyll-a to observed chlorophyll-a.**
  + Table of OBS, ARIMA, and GLM descriptors
    - Mean
      * Forecasts are said to be in consensus if they are insignificantly different from the sample mean (Gregory et al ., 2001)
    - Measure of variance (standard deviation)
    - Quantiles
    - Kurtosis (bennett et al 2013)
      * Measure of how peaked the data is
  + Table of ARIMA vs. GLM performance metrics
    - RMSE
    - R2 (variation from the 1:1 line)
    - Bias
      * SD(model)/SD(data)
      * Base R bias calc mean(observed - model)
    - Slope of the regression btw model and OBS
    - Pearson’s correlation btw model and OBS
    - Area under the curve?? (used in Araujo, Thuiller, and Pearson 2006, Climate warming and the decline of amphibians and reptiles in Europe and probably others that I wasn’t paying attention to)
    - Stow et al 2003 also reports a list of metrics used to compare
  + Figures—visual assessment
    - Plot chlorophyll over time and model predictions over time on same plot
    - Predicted vs. observed + 95% confidence interval
    - Predicted vs. observed variables with known relationship to chlorophyll
      * TP? Temp? Turbidity?, etc.

**Preliminary Results**

* Hindcasts of chlorophyll-a during training period
  + ARIMA: summer period, 2013-2016
    - Autoregressive model includes chlorophyll-a at the previous timestep, discharge to the reservoir, and shortwave radiation
    - R2 = 0.44
    - RMSE = 1.71 ug/L
    - Captures quite detailed dynamics at lower chlorophyll-a concentrations (<10 ug/L—in CTD units, maybe need to scale this for general comparison if most instruments read higher and that’s the magnitude people are used to thinking about chlorophyll?)
    - Misses some larger peaks
  + GLM: summer period, 2013-2017
    - Calibration still in progress (?)
    - R2 = 0.001
    - RMSE = 3.42 ug/L
    - Captures some larger peaks in chlorophyll-a that ARIMA does not
* Validation of models on 2018 data
  + ARIMA
    - R2 = 0.44
    - RMSE = 1.02 ug/L
  + GLM
    - Coming soon
* Iterative 16-day forecasts of chlorophyll-a
  + ARIMA
  + GLM
* Weighted model averaging

Chapter 2 Assessing the dynamics of stream-reservoir linkages across reservoirs

*Question 1: How do stream nutrient and water loads affect reservoir chla?*

*Hypothesis: something about the inflow which has the small reservoir upstream will have higher nutrient loads (Wurtsbaugh et al 2009 & 2005), or streams with lower discharge will have higher nutrient export*

***Question 4: How does nutrient processing change along a double reservoir continuum?***

***How does phytoplankton change along a double reservoir continuum?***

***How do nutrient and phytoplankton dynamics vary spatially along a double reservoir continuum?***

1. Whitney: How does nutrient processing vary across a double reservoir continuum?
2. Whitney: How do phytoplankton dynamics vary in response to nutrient availability across a double reservoir continuum?

***#What is the long-term effect of mgmt. on FW ecosystem functioning?***

Stream-lake linkages are a crucial interface which influences both the transport of nutrients and biotic responses.

Evidence supports the idea that lakes act differentially as sinks and sources of dissolved organic matter, depending on hydrologic conditions (Goodman et al 2011, Robinson et al 2007, Xu & Xu 2018).

**Hypothesis: BVR and FCR will function as sinks of DOM immediately after storm events (as measured by a decrease in the relative level of DOM along the outflow of BVR), and as a source of DOM during dry periods.**

**Background**

* Importance of watershed characteristics on waterbody condition
  + Terrestrial-aquatic linkages
  + Stream-lake linkages
    - Marcarelli & Wurtsbaugh 2009
      * Seasonality in nitrogen fixation, temporal variability in nitrogen fixation. Higher amount of fixation in lakes due to larger surface area, but higher rates in streams
    - Robinson et al 2007
      * Seasonality in nutrient availability due to weather patterns (nutrients coming from glacier melt are reduced in autumn-on)
    - Goodman et al 2011
      * Lakes act as sinks during high flow conditions and sources during low flow conditions
    - Xu & Xu 2018
      * DIC significantly decreased after passing through lake
      * Propose that CO2 outgassing is mechanism for sink behavior of lake
      * Lake functions as sink during high flow and source during low flow, consistent with other studies
    - Stachelek & Soranno 2019
      * P retention in lakes influenced by hydrologic connectivity within the whole lake watershed, but not as strongly within subwatersheds
    - Wurtsbaugh 2005
      * Can’t get access to…
    - Sadro et al 2012
      * Shows importance of landsape connectivity in explaining biogeochemical responses within snowmelt dominated landscape
    - Schmadel et al 2018
    - Jones 2010
    - Hotchkiss et al 2018., ETC
* Importance of storm events in stream dynamics
* In the new age of data availability, can we relate stream dynamics of a highly monitored stream to neighboring streams lacking data availability?

**Methods**

* Study site
  + Introduce BVR
    - 4 selected inflows (see Figure 1)
    - Relationship to FCR
      * Watershed history
      * Geographical similarity
      * Outflow to FCR
* Field sampling, summer 2019
  + Weekly (bi-monthly?) sampling of 4 major inflows to develop baseline
    - Nutrient chemistry (total and soluble nitrogen, phosphorus, and carbon)
    - Discharge (using a flowmeter)
    - Physical characteristics (dissolved oxygen, conductivity, temperature)
  + Periodic event-based sampling
    - ISCO sampler to capture high-frequency dynamics during major flow events
  + Weekly sampling at deep-hole of BVR
    - Biological, physical, and chemical sampling
  + Grab samples at strategically chosen sites along the reservoir gradient (?) (see red arrows on Figure 1)
    - CTD + nutrient grab samples

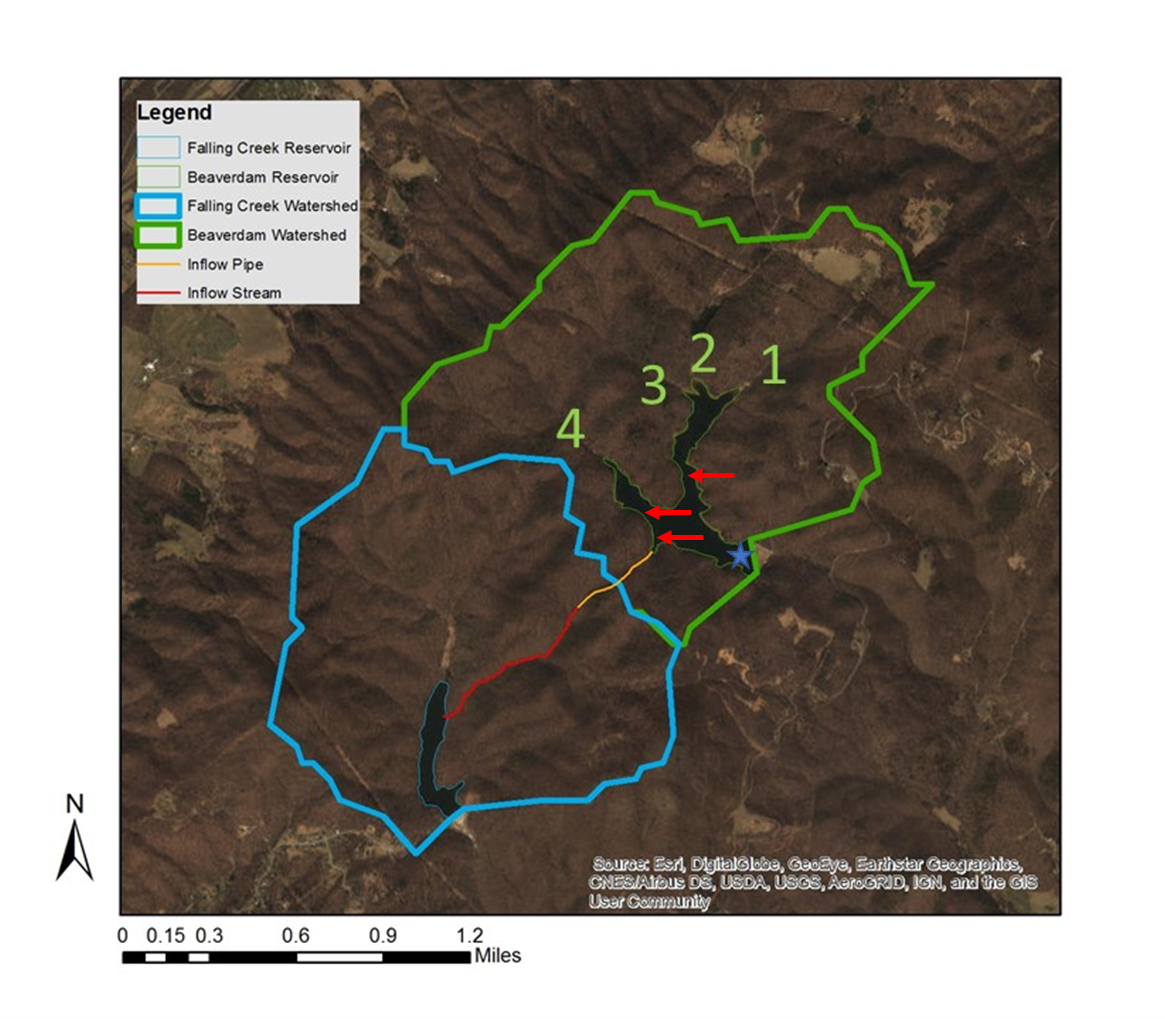


Figure 1. Map of Falling Creek Reservoir and Beaverdam Reservoir, along with watershed boundaries and proposed sampling sites

* Data analysis
  + Generalized linear models to analyze drivers of nutrient and phytoplankton dynamics along a stream-reservoir gradient

References

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