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 (emphasizing the importance of the watershed for continuity with ch 2)
      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 social and economic loss

**Proposed Research**

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

* Ecological 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
  + The current field of forecasting has substantial breadth, ranging from the well-known and widely-used weather forecasts to epidemiological forecasts of population mortality (Lee and Carter, 1992) and global Alzheimer’s trends (Brookmyer et al 2007) to forecasts of bankruptcy declaration in the United States (Shumway et al 2001).Many of these fields have been developing for decades (Shumway et al 2001), but ecological forecasting is still in its relative infancy (Dietze ref)
  + While being an emerging field of forecasting, ecological forecasting currently spans numerous disciplines, with a variety of purposes. Summary of mini-lit review here, including topics and uses: yes! Can you add in the review methods you used? E.g. search terms, which databases? What criteria were applied for if a paper was included or not, etc. I would add a table summarizing your results into the main document, in addition to the description in text here.
    - Applied uses
      * For direct use in the management of a system
      * Ex: Hazen et al 2017, Whale Watch
      * 27% of studies (n = 15) surveyed have directly applied motivations
    - Basic
      * For understanding how a system is/will respond to changing pressures for scientific purposes
      * Ex: White and Nemani 2004, forecasting soil ecology
      * 73% of studies surveyed (n = 15)
    - However, **all** of the applied studies were freshwater forecasts, illustrating not only the interest in the scientific community to understand changing freshwater systems, but the need by managers and stakeholders for probabilistic forecasts in order to cope with and adapt to changing freshwater systems.
  + **EForecasting 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 (62%, n= 16) 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 ecological important driver variables. Additionally, empirical models are relatively easy and quick to develop and implement. 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** (38%, n=16)
      * Pros
        + Useful tools for simulating changes within a system

Can be used to explain underlying mechanisms;

* + - * + Likely more generalizable because based on fundamental set of equations
      * 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?
  + Question 2: How well (assessed 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?
* This Chapter 1 will thus span the field of applied and basic forecasting science by**: 1) informing 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) providing essential information to the freshwater research community regarding modeling approaches to forecasting water quality, and 3) expanding the scientific field of ecological forecasting that will guide forecasting applications across many different ecosystems.**

**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 2).

A close up of a logo

Description automatically generated

Figure 2. Map of Falling Creek Reservoir and Beaverdam Reservoir and their watersheds

* **Historical and sensor dataset**
  + An extensive, routine monitoring dataset of water quality in FCR has been collected since 2013 in collaboration with the WVWA and Virginia Tech. 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 (add lots of lab citations here). 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, chemical, 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) when phytoplankton populations are at their highest. The training period for the empirical model will be 2013-2016 and for the process-based model will be 2013-2017. The training period is shortened for the empirical model due to a lack of input data for the model during 2017. Both models will be validated using data from 2018.
  + Empirical: ARIMA
    - We chose an autoregressive integrated moving average (ARIMA) generalized linear model for our empirical approach to forecasting chlorophyll-a in FCR. The appropriate time step to be included as the autoregressive term in ARIMA will be determined by selecting the previous measurement of chlorophyll-a with the highest Pearson’s r correlation coefficient with the current measurement of chlorophyll-a. 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 combinations with the selected driver variables, and the best model will be determined 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.
  + ADD a section here about model assessment, both for the calibration and validation periods? I think that it would be good if you could point towards what metrics you’re using with refs for assessment with a few refs backing up the why
* **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. FLARE is designed to pull 2-week weather predictions from the National Oceanic and Atmospheric Administration (NOAA) to force both ARIMA and GLM-AED. Additionally, ARIMA’s autoregressive term will be forced using our established sensor-cloud network to provide yesterday’s chlorophyll-a concentration. Lastly, inflow discharge data for both models will be estimated using an autoregressive linear model of discharge based on the previous day’s mean discharge at the major inflow to the reservoir. Driver uncertainty will be calculated for both ARIMA and GLM-AED forecasts by propagating the uncertainty associated with the NOAA weather forecasts. Parameter uncertainties for both ARIMA and GLM-AED will be calculated by sampling from a distribution of parameters within each model. Forecast effectiveness will be assessed using a suite of performance metrics calculated by comparing the forecast model outputs to the observed high-frequency chlorophyll-a sensor in FCR. Assessment metrics will be calculated for various conditions (summer stratified period, fall mixed period, winter ice period, and following storm events) to determine which approach is most robust under different conditions.
* **Integrated Model Averaging (Question 3)** 
  + In order to develop the most informative forecast of chlorophyll-a, we will develop a weighted model average to incorporate the strengths of multiple models. This will allow the forecast to capture as much variability as possible in chlorophyll-a if for example, one model does a better job of prediction chlorophyll-a at lower concentrations, while another model is more effective at capturing large bloom events. Particularly for low probability yet high impact events such as large blooms in phytoplankton, the model averaging will enable us to develop an early-warning system which will alert users when conditions that indicate a peak in phytoplankton is likely to occur.
    - ‘‘Observation and theory get on best when they are mixed together, both helping one another in the pursuit of truth’’ (Eddington 1935).
* Model assessment: Model performance will be assessed using the following metrics
  + 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)
    - 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.

Hindcasts of chlorophyll-a

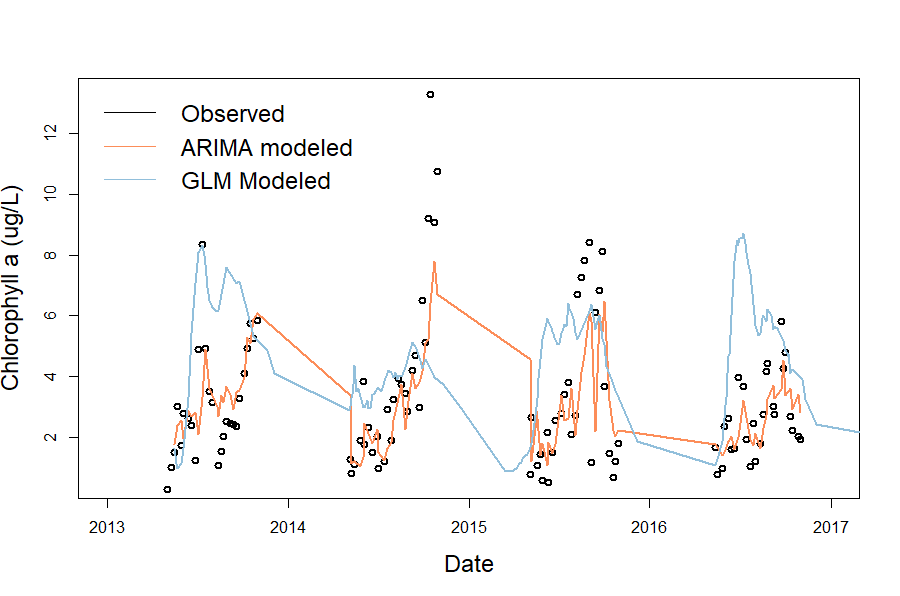
Using the summer period of 2013-2017 as our training period, we developed and calibrated both an empirical, ARIMA, and process-based model, GLM-AED. Our best-fitting ARIMA model over 2013-2016 included discharge to the reservoir and shortwave radiation. ARIMA hindcasted chlorophyll-a over 2013-2016 with an R2 = 0.44 and RMSE = 1.71 ug/L. The ARIMA model was able to successfully capture fluctuations at lower chlorophyll-a concentrations (<10 ug/L) (Figure 2). However, when chlorophyll-a reached values above ~10 ug/L, the model the model was unable to recreate these observed dynamics. GLM-AED was calibrated over 2013-2016 and hindcasted chlorophyll-a with R2 = XXX and RMSE = YYYY. While GLM-AED does not capture as much detail in chlorophyll-a dynamics, it does capture some large peaks in chlorophyll-a that ARIMA does not. 

Figure 2. Observed and modeled chlorophyll-a data over 2013-2016.

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

**Proposed timeline for Chapter 1**  
*Black boxes indicate completed tasks, gray boxes indicate proposed completion*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Fall 2018 | Spring 2019 | Summer 2019 | Fall 2019 | Spring 2020 |
| Develop ARIMA GLM-AED |  |  |  |  |  |
| Calibrate GLM-AED |  |  |  |  |  |
| Produce hindcasts |  |  |  |  |  |
| Validate model (Q1) |  |  |  |  |  |
| Create integrated model system/early-warning system (Q3??) |  |  |  |  |  |
| Adapt FLARE code for ARIMA (Q2?) |  |  |  |  |  |
| Run forecasts in FLARE |  |  |  |  |  |
| Submit manuscript, *target journal: ???* |  |  |  | ??? |  |

*Chapter 2 Assessing the dynamics of stream-reservoir linkages across a double reservoir continuum*

**Background**

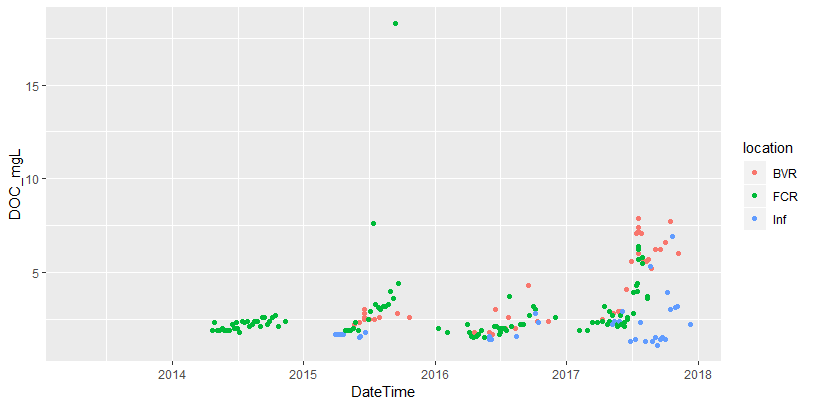
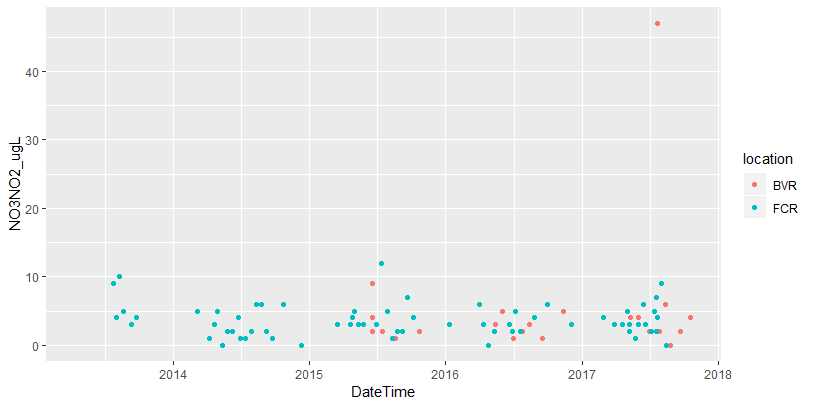
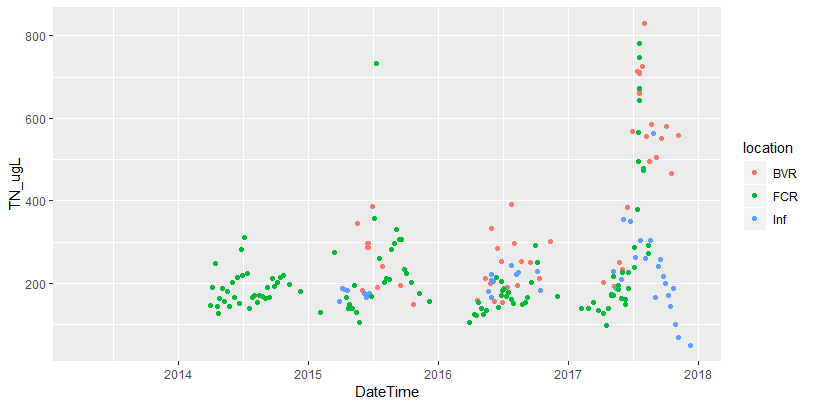
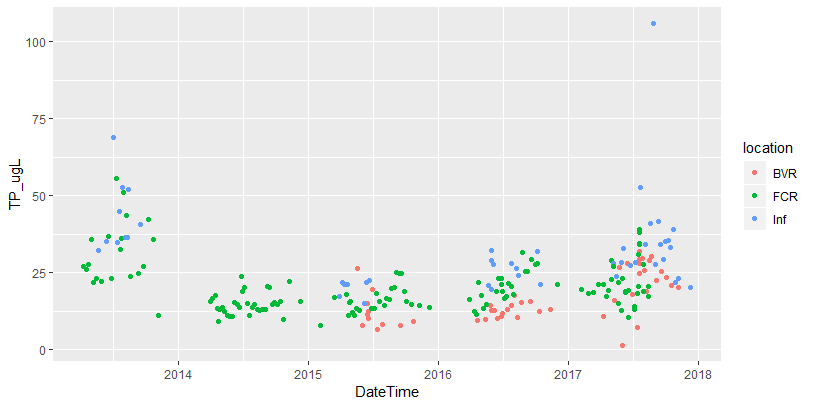
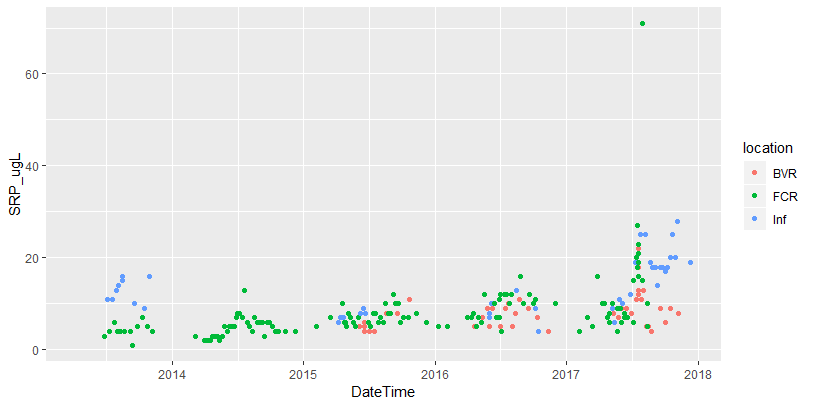
Stream-lake linkages are crucial interfaces which influence both the transport of nutrients and biotic responses. Evidence supports the idea that lakes act differentially as sinks and sources of dissolved organic matter, depending on hydrologic conditions (Goodman et al 2011, Robinson et al 2007, Xu & Xu 2018).

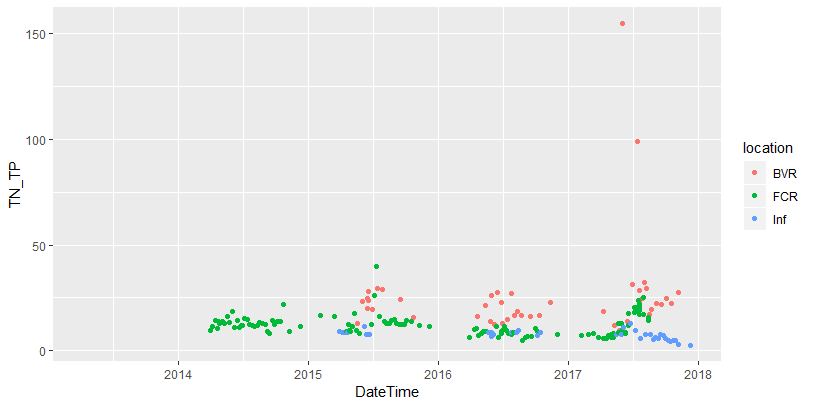
* Importance of watershed characteristics on waterbody condition
  + Terrestrial-aquatic linkages
  + Stream-lake linkages and watershed connectivity
    - Marcarelli & Wurtsbaugh 2009
      * Seasonality in nitrogen fixation, temporal variability in nitrogen fixation. Higher amount of fixation in lakes due to larger surface area, but higher rates in streams
    - Stachelek & Soranno 2019
      * P retention in lakes influenced by hydrologic connectivity within the whole lake watershed, but not as strongly within subwatersheds
        + Higher p retention in lakes with less upstream lake area (less flushing?)

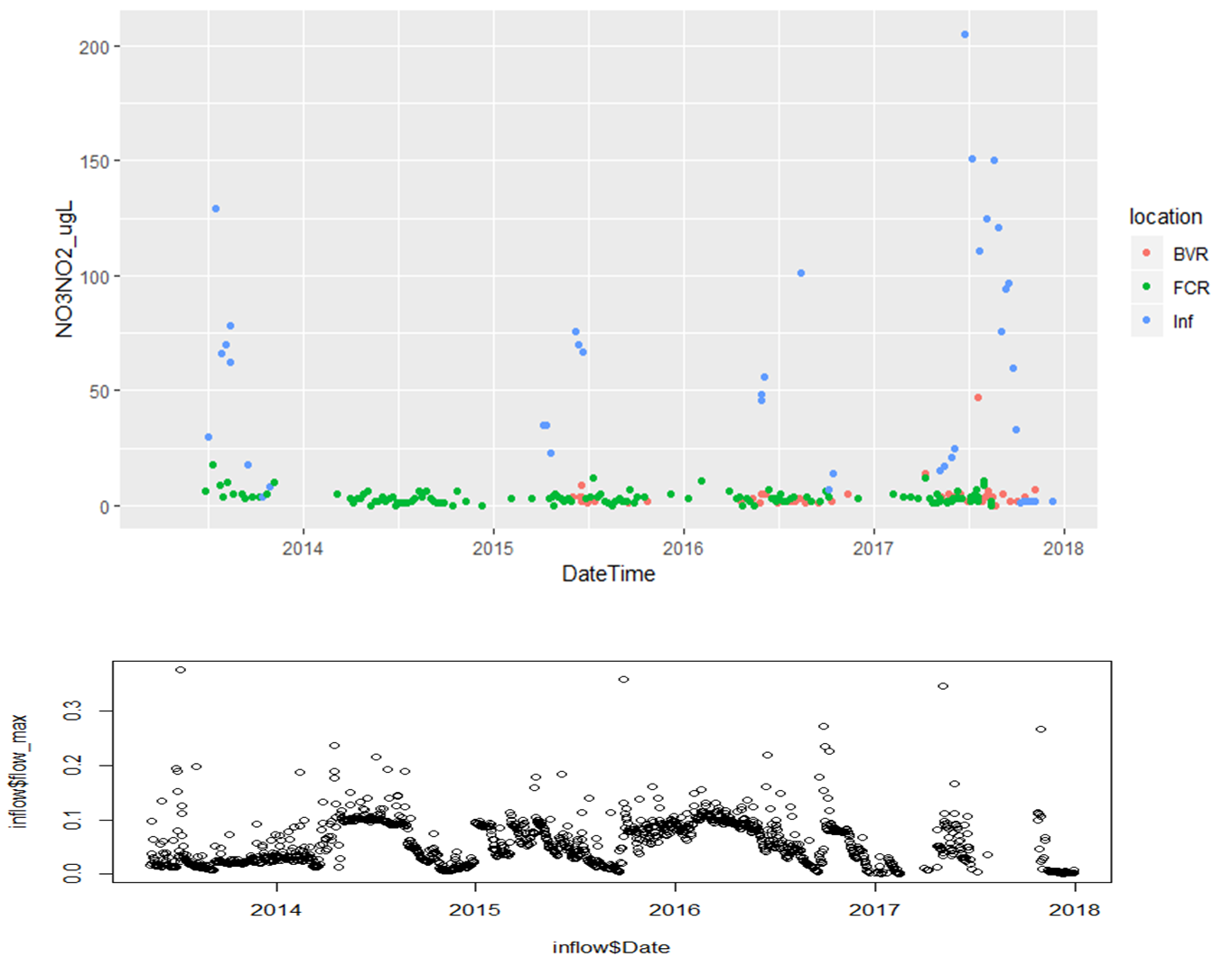
Applied to BVR-FCR system: BVR would have higher P retention than FCR if BVR outflow is flowing into FCR

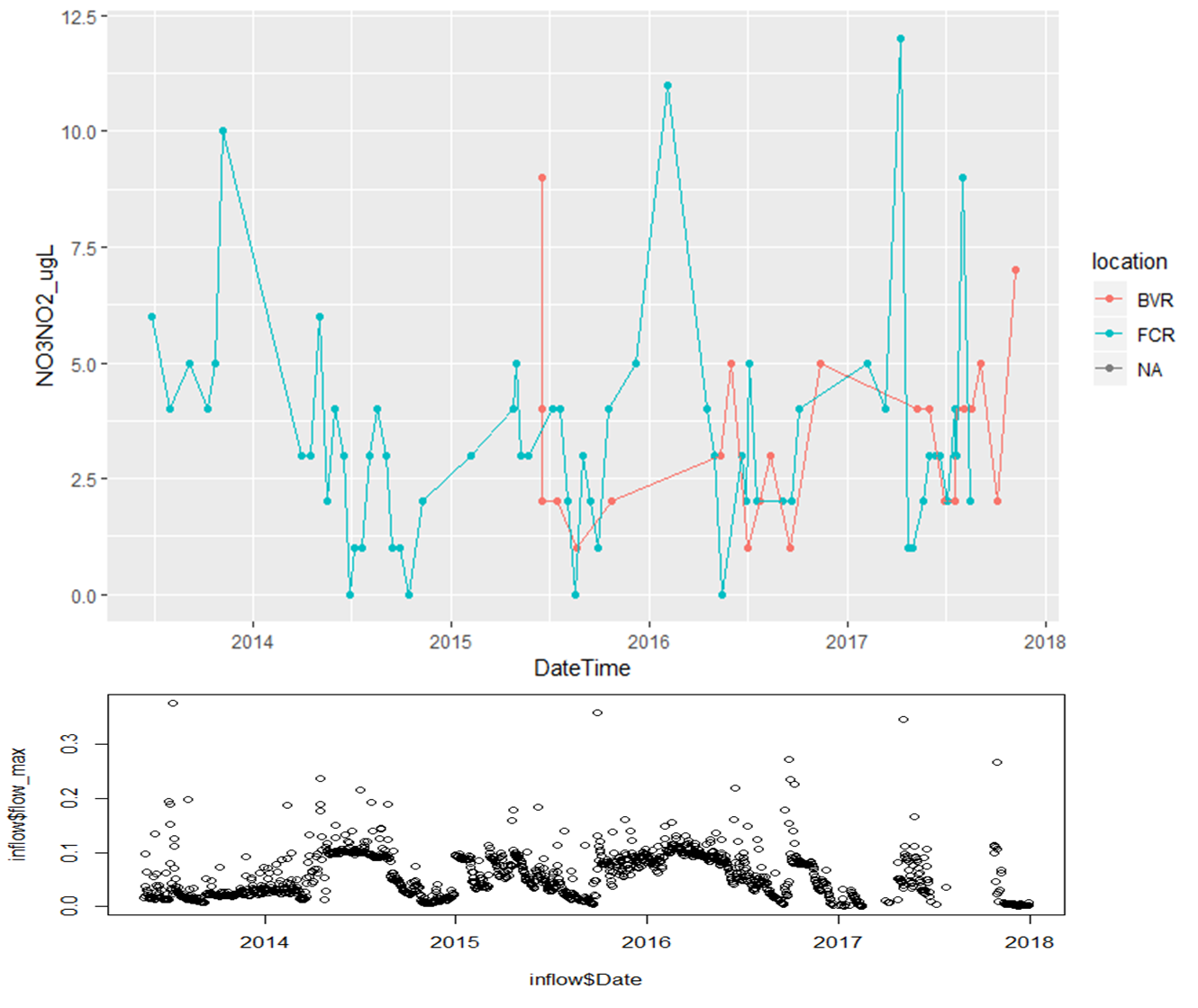
* + - Wurtsbaugh 2005
      * Can’t get access to…
    - Sadro et al 2012
      * Shows importance of landsape connectivity in explaining biogeochemical responses within snowmelt dominated landscape
      * Directly measures response of various nutrient chem variables to lake network number
    - Schmadel et al 2018
    - Jones 2010
    - Hotchkiss et al 2018
* Importance of hydrologic variance in nutrient dynamics
  + - 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
* So, in summary, some other things I’m expecting to see both lake chain/connectivity effects and residence time/hydrologic variability effects.
  + Question 1: How does hydrologic connectivity and reservoir continuum location (need some wordsmithing) influence nutrient availability and phytoplankton dynamics?
    - **Hypothesis 1: as you go along a reservoir continuum (BVR-> FCR) we expect to see:**
      * **Decreases in N:P, Nitrate**
      * **A hypothesis (that will be more informed once I’ve read more) about phytos**
  + Question 2: How does nutrient processing and resultant phytoplankton dynamics across a double reservoir continuum change in response to hydrologic events?
    - **Hypothesis 2: Reservoirs will act as sinks for nutrients in response to high flow (storm) events and source for nutrients during low flow periods**

**Preliminary Analysis of Historical BVR-Inf-FCR Data***figures made using nutrient chemistry from 0.1m at FCR and BVR + inflow grabs*

* Hypothesis 1
  1. DOC should increase from BVR -> FCR
     + Result: not a clear signal from the data we have currently
       - Inflow is generally lower but this is likely more of a per unit area problem
     + 
  2. Nitrate should decrease from BVR -> FCR (Sadro et al 2012)
     + Result: the opposite is seen in BVR -> FCR
       - Not a strong signal, but FCR appears to have slightly higher nitrate-ite levels than BVR
       - 
     + What about TN? (need to do more research to synthesize what literature says we should expect for how TN should change with increases in lake chain number or hydrologic connectivity)
       - Result: we see slightly higher levels of TN in BVR than in FCR
         * Overall, inflow TN is much lower than reservoir TN, but would be good to account for per area
       - 
  3. What about TP?? (need to read more lit to get a good handle on what we would expect the direction of this relationship to be)
     + Result: TP increases with reservoir continuum
       - FCR is consistently higher than BVR
       - Inflow is in general higher than FCR
         * Which is very interesting considering the area of FCR vs. inflow
       - Signal is similar but less clear with SRP
     + 
     + 
  4. N:P should decrease from BVR -> FCR
     + Result: we see this pretty clearly in BVR and FCR!
       - BVR consistently has higher TN:TP
       - BVR P-limited?
       - TP and SRP are also lower in BVR



* **Hypothesis 2: Reservoirs will act as sinks for nutrients in response to high flow (storm) events and source for nutrients during low flow periods**
  + A quick and dirty: Pairing nitrate levels with discharge over the same time period
    - We see huge spikes in nitrate-ite at inflow following max discharge events
    - But if you look at the figure of nitrate-ite at just FCR and BVR, we do not see increases after max discharge events (except for 2017)—I can maybe even see a decrease in nitrate-ite concentrations at FCR following major discharge events (e.g., 2014, 2015, and 2016)
  + 

****

**Methods**

* Study site
  + Beaverdam Reservoir (BVR) is a small (39 ha), shallow (maximum depth < 13m), dimictic reservoir owned and operated by the Western Virginia Water Authority (WVWA) and located in Vinton, VA. BVR has multiple inflow stream, but we have selected inflow #4 (Figure 1, above) to study, as it contributes a majority of discharge to the reservoir’s western arm. BVR has an outflow pipe that flows to Falling Creek Reservoir (FCR), a smaller (), shallower (), dimictic reservoir, also owned and operated by WVWA.
* Field sampling, summer 2019
  + Monthly sampling along the reservoir continuum
    - Sites
      * 1 (2?, TBD) stream inflow to BVR
      * X transects within left arm of BVR
      * BVR site 50
      * Z sites alone BVR-FCR inflow
      * FCR Site 20, 30, 40, 45, & 50
      * FCR spillway
    - Data collection
      * Nutrient chemistry (total and soluble nitrogen & phosphorus, and carbon)
      * Discharge (using a flowmeter)
      * Physical characteristics (dissolved oxygen, conductivity, temperature)
      * Chlorophyll-a
  + Periodic event-based sampling
    - Adaptive sampling campaign to capture seasonal, hydrological, and oxix-anoxic dynamics
    - N = 8???
* Data analysis
  + Generalized linear models to analyze drivers of nutrient and phytoplankton dynamics along a stream-reservoir gradient????